AN ENERGY STABLE AND POSITIVITY-PRESERVING SCHEME FOR THE MAXWELL-STEFAN DIFFUSION SYSTEM

XIAOKAI HUO, HAILIANG LIU, ATHANASIOS E. TZAVARAS, AND SHUAIKUN WANG

Abstract. We develop a new finite difference scheme for the Maxwell-Stefan diffusion system. The scheme is conservative, energy stable and positivity-preserving. These nice properties stem from a variational structure and are proved by reformulating the finite difference scheme into an equivalent optimization problem. The solution to the scheme emerges as the minimizer of the optimization problem, and as a consequence energy stability and positivity-preserving properties are obtained.

1. Introduction

Cross diffusion occurs in multicomponent systems, such as ionic liquids, wildlife populations, gas mixtures, tumor growth, etc [13, 16]. In these multicomponent systems, the diffusion happens not only in the direction from high concentration to low concentration, but also in the opposite direction due to cross diffusion. In such cases, diffusion can not be described by Fick’s diffusion law and the Maxwell-Stefan diffusion model can be used instead. The Maxwell-Stefan model assumes the friction between two components is proportional to their difference in velocity and molecular fractions. It is widely used in modeling multicomponent systems.

In this work, we consider the Maxwell-Stefan diffusion system for a $n$-component mixture on the torus $T^d$, which reads for $i = 1, \ldots, n$,

$$
\begin{align*}
\partial_t \rho_i + \nabla \cdot (\rho_i v_i) &= 0, \\
- \sum_{j=1}^{n} b_{ij} \rho_j (v_i - v_j) &= \nabla \log \rho_i - \frac{1}{\sum_{j=1}^{n} \rho_j} \sum_{j=1}^{n} \rho_j \nabla \log \rho_j, \\
\sum_{j=1}^{n} \rho_j v_j &= 0.
\end{align*}
$$

(1)

(2)

(3)

Here $x \in T^d$, $\rho_i = \rho_i(x,t)$ and $v_i = v_i(x,t)$ are the density and velocity of the $i$-th component. The initial conditions are taken to be

$$
\rho_i(x,0) = \rho_{i0}(x), \ i = 1, \ldots, n,
$$

and we assume that

$$
\rho_{i0}(x) > 0, \quad \text{and} \quad \sum_{j=1}^{n} \rho_{j0}(x) = 1 \quad \text{for} \ x \in T^d.
$$

(4)

Solutions of (1) conserve the total mass $\partial_t \sum_{i=1}^{n} \rho_i + \nabla \cdot \sum_{i=1}^{n} \rho_i v_i = 0$. Condition (3) imposes that the average velocity of the mixture is $v_{\text{av}} \equiv 0$ and thus the total density $\sum_{i=1}^{n} \rho_i$ is conserved at each $x \in T^d$. Hypothesis (4) then fixes the total mass to

$$
\sum_{j=1}^{n} \rho_j(x,t) = 1, \quad \text{for} \ x \in T^d, \ t > 0.
$$

(5)

2000 Mathematics Subject Classification. 35K55, 35Q79, 65M06, 35L45.

Key words and phrases. Finite difference, Maxwell-Stefan systems, cross-diffusion, Positivity-preserving, Energy dissipation.
Accordingly, (1)-(3) reduces to
\[ \partial_t \rho_i + \nabla \cdot (\rho_i v_i) = 0, \]
\[ \nabla \rho_i = -\sum_{j=1}^{n} b_{ij} \rho_i \rho_j (v_i - v_j), \]
for \( i = 1, \ldots, n \), which is the usual form of the Maxwell-Stefan diffusion system.

The system (1)-(3) can be obtained as the high-friction limit of the multicomponent Euler equations [10].
\[ \partial_t \rho_i + \nabla \cdot (\rho_i v_i) = 0, \]
\[ \partial_t (\rho_i v_i) + \nabla \cdot (\rho_i v_i v_j) + \rho_i \nabla \frac{\delta F(\rho)}{\delta \rho_i} = -\frac{1}{\varepsilon} \sum_{j=1}^{n} b_{ij} \rho_i \rho_j (v_i - v_j), \]
when the total momentum (or the mean velocity) is zero. Here the energy functional
\[ F(\rho) = \sum_{i=1}^{n} \int_{\mathbb{R}^d} \rho_i(x) \log \rho_i(x) dx. \]

It was proved in [10] that, when the total momentum is zero, the system (8) converges to (1)-(3) in the high-friction limit \( \varepsilon \to 0 \). Moreover, (1)-(3) can be regarded as a gradient flow for \( F(\rho) \).

This raises the following question: Given densities \( \rho^0 = (\rho^0_i)^{n}_{i=1}, \rho^i = (\rho^i_1)^{n}_{i=1} \), with \( \sum_i \rho^i_i = \sum_i \rho^0_i = 1 \), consider the minimization problem
\[ \min_{(\rho,v) \in K} \int_{0}^{1} \int_{\mathbb{T}^{d}} \sum_{i,j=1}^{n} \frac{1}{4} b_{ij} \rho_i \rho_j (v_i - v_j)^2 \] dx dt \]
over the set
\[ K = \left\{ \rho = (\rho_1, \ldots, \rho_n), v = (v_1, \ldots, v_n) : \partial_t \rho_i + \nabla \cdot (\rho_i v_i) = 0, \quad i = 1, \ldots, n, \quad \sum_{j=1}^{n} \rho_j v_j = 0, \quad \rho_i(0, x) = \rho^0_i(x), \quad \rho_i(1, x) = \rho^1_i(x) \right\}. \]

The problem (10) as the minimum of the frictional work is motivated by the well-known characterization of the Wasserstein distance in a one-component fluid obtained by Benamou-Brenier [1]. The study of this question will be given in a forthcoming work. The minimization (10) and the gradient structure of (1)-(3) detailed in [10], motivate us to use the work of friction as a building block for a numerical scheme of variational provenance – in the spirit of the well known JKO scheme [12] – in order to exploit the gradient structure of the Maxwell-Stefan system. This connection is pursued in the present work.

In this paper, we develop a new implicit-explicit finite difference scheme for the Maxwell-Stefan system (1)-(3) and prove that the scheme is energy dissipating and positivity preserving, for arbitrary time step and spatial meshes. The scheme in one dimension takes the form:
\[ \frac{\rho^k_i - \rho^{k+1}_i}{\Delta t} + d_h (\rho^k_i v^k_i) = 0, \]
\[ -\sum_{j=1}^{n} b_{ij} \rho^k_j (v^{k+1}_i - v^{k+1}_j) = D_h \log \rho^{k+1}_i - \frac{1}{\sum_{j=1}^{n} \rho^k_j} \sum_{j=1}^{n} \rho^k_j D_h \log \rho^{k+1}_j, \]
\[ \sum_{j=1}^{n} \rho^k_j v^{k+1}_j = 0 \]
(for the \( d \)-dimensional case the reader is referred to Section [4]). The subscript \( i \) refers to the \( i \)-th component and takes values \( i = 1, \ldots, n \), while the superscript \( k \) refers to the \( k \)-th time step. The equations (11)-(13) are computed at spatial grid points \( \ell \) or \( \ell + \frac{1}{2} \) of staggered lattices in a way precised in Section [2]. The parameter \( \Delta t \) is the time step and \( h \) is the mesh size. The operators
$d_h, D_h$ are central difference operators, in one dimension, defined by
\[
(d_hf)_i^\ell = \frac{f_{i+1/2}^\ell - f_{i-1/2}^\ell}{h}, \quad (D_hf)_i^{\ell+1/2} = \frac{f_{i+1}^\ell - f_i^\ell}{h},
\]
where $\ell = \{1, \ldots, N\}$, $N$ the number of mesh intervals, and we set $(f_i)^{\ell+1/2} = \frac{1}{2}(f_i + f_{i+1})$.

The scheme is induced by a spatial discretization of the constrained optimization problem (cf. [38])
\[
\min_K \left\{ \int_{\Omega} \Delta t \frac{1}{2} \sum_{i,j=1}^n b_{ij} \rho_i^k \rho_j^k |u_i - u_j|^2 dx + \int_{\Omega} \sum_{j=1}^n \rho_j \log \rho_j dx \right\},
\]
where the set $K$ is defined to be
\[
K = \left\{ (\rho, v) : \rho > 0, \frac{\rho_i - \rho_i^k}{\Delta t} + \nabla \cdot (\rho_i^k u_i) = 0, \sum_{i=1}^n \rho_i^k u_i = 0 \right\}.
\]

The approach is motivated by the JKO-scheme [12] and the Benamou-Brenier interpretation of the Wasserstein distance [1], the latter suggesting an alternate variational scheme for nonlinear Fokker-Planck equations espoused in [17]. The novelty here is (i) that the limiting problem is a coupled parabolic system and (ii) that the mechanical friction is a complex interaction among the different components (see [2]) that is only captured in bulk by the dissipation functional [10]. Nevertheless, this suffices in capturing the detailed interaction.

We show that there exists a discrete energy function which dissipates along time iterations, and that the numerical solutions for the densities generated by the scheme [11]-[13] preserve the positivity of the initial densities. The proof uses variational arguments and is based on the reformulation of the finite difference scheme as an equivalent optimization problem. An interesting feature is the role played by an elliptic operator $\mathcal{L}_b$ defined in [19] and the induced dual norm [20]. The reader familiar with the Wasserstein distance will recognize analogies with duality induced norms [20] [21] [22] appearing in the theory of nonlinear Fokker-Planck equations and induced by the metric tensor generating the Wasserstein metric.

A large literature [2] [3] [7] [8] [13] [14] employing diverse techniques has provided a basic theory for the Maxwell-Stefan system [11]-[3]. The existence of global weak solutions is established in [15], while local existence of strong solutions was shown in [2] [8]. Explicit finite difference schemes were developed in [3] [7]. An implicit Euler Galerkin scheme was developed in [14] for the Maxwell-Stefan system coupled with a Poisson equation. The scheme was also shown to satisfy a discrete entropy inequality. However, the property of preserving positivity has not been investigated in the above works. The present work provides a connection between finite difference schemes and variational minimization problems. This approach is quite robust and we expect that, once the theory for the continuous problem [15] is further developed, it will lead to theoretical results for more complicated schemes such as finite elements.

Recently there has been a growing interest in developing energy stable and/or positivity-preserving numerical schemes for nonlinear diffusion equations [4] [5] [6] [11] [15] [19] [23]. Positivity-preserving schemes for the Poisson-Nernst-Planck systems were developed in [18] [19], where the maximum principle was used to show the non-negativity of the scheme. A series of diffusion equations satisfying a gradient flow structure was considered in [5] [6] [9] [23], where energy-stable schemes were developed for the Cahn-Hilliard equations, with positivity-preserving properties proved in [4] [6] via optimization formulations. The technique was also used in [11] to prove the positivity and energy-stability properties for a scheme associated to the quantum diffusion equation. Our approach extends such works to a setting of systems that are gradient flows by exploiting the frictional dissipation natural to the Maxwell-Stefan system.

The structure of the paper is as follows: in Section 2 we give the details of the numerical scheme and show that it conserves the total mass and is consistent. In Section 3 we first prove that the numerical scheme is equivalent to an optimization problem, in Theorem 1, and then show the energy stability and positivity-preserving properties in Theorem 2. We provide the multidimensional scheme in Section 4 and show that similar properties also hold. Finally, we give some numerical examples to verify the proved properties.
2. The scheme

2.1. Notations. We use notations from [21]. We define the following two grids on the torus \( \mathbb{T} = [0, L] \) with mesh size \( h = L/N \), where \( N \) is the number of mesh intervals:

\[
C := \{ h, 2h, \ldots, L \}, \quad E := \left\{ \frac{h}{2}, \frac{3h}{2}, \ldots, \left( N - \frac{1}{2} \right)h \right\}.
\]

We define the discrete \( N \)-periodic function spaces as

\[
C_{\text{per}} := \{ f : C \to \mathbb{R} \}, \quad E_{\text{per}} := \{ f : E \to \mathbb{R} \}.
\]

Here we call \( C_{\text{per}} \) the space of cell centered functions and \( E_{\text{per}} \) the space of edge centered functions. We use \( f_t \) to denote the value of function \( f \) at grid point \( x_t = \ell h \). We also define the subspace \( \hat{C}_{\text{per}} := \left\{ f : f \in C_{\text{per}}, \sum_{\ell=1}^N f_\ell = 0 \right\} \). We can extend the above definitions to vector value functions.

For example, we define \( \hat{C}^n_{\text{per}} \) by

\[
\hat{C}^n_{\text{per}} := \{ f = (f_1, \ldots, f_n) : f_\ell \in C_{\text{per}}, \ell = 1, \ldots, n \}.
\]

The spaces \( C^n_{\text{per}}, \hat{C}^n_{\text{per}} \) are defined the same way. The discrete gradients \( D_h \) and \( d_h \) are defined in [14]. We define the average of the function values of nearby points by

\[
\hat{f}_{\ell + \frac{1}{2}} = \frac{f_\ell + f_{\ell + 1}}{2}, \quad \text{if } f \in C_{\text{per}}, \quad \text{and} \quad \hat{f}_t = \frac{f_{t + \frac{1}{2}} + f_{t - \frac{1}{2}}}{2}, \quad \text{if } f \in E_{\text{per}}.
\]

The inner products are defined by \( \langle f, g \rangle := h \sum_{\ell=1}^N f_\ell g_\ell, \forall f, g \in C_{\text{per}}, \quad \text{and} \quad [f, g] := h \sum_{\ell=1}^N f_{\ell + \frac{1}{2}} g_{\ell + \frac{1}{2}}, \forall f, g \in E_{\text{per}}. \)

They can also be extended on \( C^n_{\text{per}} \) and \( \hat{C}^n_{\text{per}} \) with

\[
\langle f, g \rangle := h \sum_{\ell=1}^n f_\ell g_\ell, \forall f, g \in C^n_{\text{per}}, \quad [f, g] := h \sum_{\ell=1}^n f_{\ell + \frac{1}{2}} g_{\ell + \frac{1}{2}}.
\]

We also take the following notation:

\[
\langle f \rangle := h \sum_{\ell=1}^N f_\ell, \quad f \in C_{\text{per}}, \quad [f] := h \sum_{\ell=1}^N f_{\ell + \frac{1}{2}}, \quad f \in E_{\text{per}}.
\]

Suppose \( f \in C_{\text{per}} \) and \( \phi \in E_{\text{per}} \), the following summation-by-parts formula holds:

\[
\langle f, d_h \phi \rangle = -[D_h f, \phi].
\]

Next, we introduce a norm on \( \hat{C}^{n-1}_{\text{per}} \). Let \( \Phi \) be a \( (n-1) \times (n-1) \) symmetric, positive definite matrix, with \( \Phi_{ij} \in E_{\text{per}}, i, j = 1, \ldots, n - 1 \). We introduce the operator \( \mathcal{L}_\Phi \) on \( \hat{C}^{n-1}_{\text{per}} \) defined by

\[
\mathcal{L}_\Phi f := -d_h (\Phi D_h f) = \left( -\sum_{j=1}^{n-1} d_h (\Phi_{ij} D_h f_j) \right), \quad \forall f \in \hat{C}^n_{\text{per}}.
\]

For any \( g \in \hat{C}^{n-1}_{\text{per}} \), let \( f \) be determined by \( g = \mathcal{L}_\Phi f \), we define the norm

\[
\|g\|^2_{\hat{C}^{n-1}_{\per}} := [D_h f, \Phi D_h f].
\]

2.2. The scheme. The scheme [11]-[13] is written in the component form as follows:

\[
\frac{\rho_{ij, \ell+\frac{1}{2}}^{k+1} - \rho_{ij, \ell+\frac{1}{2}}^{k}}{\Delta t} = \frac{1}{h} \left( \rho_{ij, \ell+\frac{1}{2}}^{k} \left( \rho_{ij, \ell+\frac{1}{2}}^{k} - \rho_{ij, \ell-\frac{1}{2}}^{k} \right) - b_{ij} \left( \rho_{ij, \ell+\frac{1}{2}}^{k} - \rho_{ij, \ell+\frac{1}{2}}^{k} \right) \right),
\]

\[
\sum_{j=1}^{n} b_{ij} \rho_{ij, \ell+\frac{1}{2}}^{k} = \log \rho_{ij, \ell+\frac{1}{2}}^{k+1} - \log \rho_{ij, \ell+\frac{1}{2}}^{k+1},
\]

\[
\sum_{j=1}^{n} \rho_{ij, \ell+\frac{1}{2}}^{k} = 0.
\]
subject to initial data
\[ \rho_{i,0}^0 = \rho_{0i}(x_i), \quad i = 1, \ldots, n, \quad \ell = 1, \ldots, N. \] (24)

The scheme \([21]-[23]\) is an implicit-explicit finite difference scheme. It can be obtained formally by discretizing the system \([1]-[3]\).

Next we study the conservation properties of the scheme. First we show that, at each grid point, the total mass is preserved.

**Lemma 1.** Suppose the solutions to the scheme \([11]-[13]\) are positive for \(k \geq 1\). Then the total mass at each grid point is conserved, i.e.

\[ \sum_{i=1}^{n} \rho_{i,\ell}^k = \sum_{i=1}^{n} \rho_{i,\ell}^0, \quad \ell = 1, \ldots, N \text{ and } k \geq 1. \] (25)

**Proof.** From equations \([21]\) and \([23]\), we have for \(\ell = 1, \ldots, N,

\[ \sum_{i=1}^{n} \rho_{i,\ell}^{k+1} = \sum_{i=1}^{n} \rho_{i,\ell}^k - \Delta t \sum_{i=1}^{n} d_h (\rho_{i,\ell}^k v_{i,\ell}^{k+1}) \]

Next, we show that for each component, the mass is conserved, i.e. the summation over grid points is conserved. The following lemma holds.

**Lemma 2.** Suppose the solutions to the scheme \([11]-[13]\) are positive for any \(k \geq 1\). Then the mass for each component is conserved, i.e.,

\[ \sum_{\ell=1}^{N} \rho_{i,\ell}^k = \sum_{\ell=1}^{N} \rho_{i,\ell}^0, \quad i = 1, \ldots, n, \quad k \geq 1. \] (26)

**Proof.** From \([21]\), we get

\[ \sum_{\ell=1}^{N} \rho_{i,\ell}^{k+1} - \sum_{\ell=1}^{N} \rho_{i,\ell}^k = - \Delta t \sum_{\ell=1}^{N} d_h (\rho_{i,\ell}^k v_{i,\ell}^{k+1}) \]

Iterating in \(k\) we obtain \([26]\). \(\square\)

### 2.3. The scheme in \(n-1\) components

We consider first the solvability of the algebraic system \([2]-[3]\) under the hypothesis \(b_{ij} > 0\). Since summing the equations \([2]\) in \(i = 1, \ldots, n\) equals zero, these \(n\) equations are not independent. One easily checks that for \(\rho_{i} > 0\) the homogeneous system

\[ - \sum_{j=1}^{n} b_{ij} \rho_{j} (v_{i} - v_{j}) = 0 \]

has only the trivial solution \(v_{1} = \cdots = v_{n}\). Hence the null space has dimension one. The solution of \([2]-[3]\) is given by the following lemma.

**Lemma 3.** Let \(\rho_{i}(x, t) > 0, \ x \in \mathbb{T}^d, t > 0, \ i = 1, \ldots, n, \) and suppose that \(b_{ij} > 0\) and \(b_{ij} = b_{ji}, \) for \(i \neq j\) and \(i, j = 1, \ldots, n.\) Then the algebraic system \([2], [3]\) has a unique solution that is explicitly expressed by

\[ \rho_{i} v_{i} = - \sum_{j=1}^{n-1} D_{ij} \nabla (\log \rho_{j} - \log \rho_{n}), \quad i = 1, \ldots, n - 1, \text{ and } \rho_{n} v_{n} = - \sum_{i=1}^{n-1} \rho_{i} v_{i}, \]

where

\[ D_{ij} = D_{ij}(p) = \sum_{s,m=1}^{n-1} Q^{-T}_{is} B_{sm}^{-1} Q^{-1}_{mj}, \quad i,j = 1, \ldots, n-1, \] (27)
Using (13) we get
\[ B_{ij} = B_{ij}(\rho) = \delta_{ij} \sum_{m=1}^{n} b_{im} \rho_i \rho_m - b_{ij} \rho_i \rho_j, \quad (28) \]

\[ Q_{ij} = Q_{ij}(\rho) = \frac{1}{\rho_i} \delta_{ij} + \frac{1}{\rho_n}, \quad (29) \]

\[ (Q^{-1})_{ij} = Q_{ij}^{-1}(\rho) = \delta_{ij} \rho_n - \frac{\rho_i \rho_j}{\sum_{j=1}^{n} \rho_j}. \quad (30) \]

For \( \rho > 0 \), \( B \) is diagonally dominant and thus invertible. We note that \( Q^T = Q \) and that by a direct computation \( QQ^{-1} = Q^{-1}Q = I \), where \( Q^{-1} \) is determined by (30); hence, \( Q \) is also invertible. The proof can be found in [10] or [25]. A similar formula is established for the numerical scheme [11]–[13].

**Lemma 4.** Assume \( b_{ij} > 0 \) and \( b_{ij} = b_{ji} \) for \( i \neq j \) and \( i, j = 1, \ldots, n \). Suppose \( \rho^{k + 1}_i > 0 \) for \( i = 1, \ldots, n, \ell = 1, \ldots, N \). The solutions of (12)–(13) are calculated by the explicit formula
\[ \rho^{k + 1}_i v^{k+1}_i = - \sum_{j=1}^{n-1} \hat{D}_{ij} D_h (\log \rho^{k+1}_j - \log \rho^{k+1}_n), \quad i = 1, \ldots, n - 1, \quad (31) \]

and \( \rho^{k + 1}_n v^{k+1}_n = - \sum_{i=1}^{n-1} \hat{D}^{k+1}_i v^{k+1}_i \). Here
\[ \hat{D}_{ij} = \sum_{s, m=1}^{n-1} (\hat{Q}^k)_{is} (\hat{B}^k)_{sm} (\hat{Q}^k)_{jm}^{-1}, \quad (32) \]

and \( \hat{Q}_{ij} = Q_{ij}(\rho^k), \hat{B}_{ij} = B_{ij}(\rho^k), (\hat{Q}^k)_{ij}^{-1} = Q_{ij}^{-1}(\rho^k) \) are the corresponding matrices [25]–[30] with \( \rho_i \) replaced by \( \hat{\rho}_i^k \).

Notice that formulas (31) hold at each grid point \( \ell + 1/2 = 3/2, \ldots, N/2 + 1 \) (or \( 1/2 \)); to simplify the notation, we do not write the subscript \( \ell + 1/2 \).

**Proof.** Multiplying (12) by \( \hat{\rho}_i^k \) gives
\[ \rho^k_i D_h \log \rho^{k+1}_i - \frac{\rho^k_i}{\sum_{j=1}^{n} \hat{\rho}_j^k} \sum_{j=1}^{n} \hat{\rho}_j^k D_h \log \rho^{k+1}_j = - \sum_{j=1}^{n} b_{ij} \hat{\rho}_i^k \rho_j^k (v^{k+1}_i - v^{k+1}_j), \]

which is rewritten as
\[ \sum_{j=1}^{n} \left( \delta_{ij} \hat{\rho}_i^k - \frac{\rho^k_i \rho^k_j}{\sum_{j=1}^{n} \hat{\rho}_j^k} \right) D_h \log \rho^{k+1}_j = - \sum_{j=1}^{n} \left( \delta_{ij} \sum_{m=1}^{n} b_{im} \hat{\rho}_i^k \rho_m^k - b_{ij} \rho^k_i \hat{\rho}_j^k \right) v^{k+1}_j. \quad (33) \]

Setting \( \hat{B}^k_{ij} = B_{ij}(\hat{\rho}^k) = \delta_{ij} \sum_{m=1}^{n} b_{im} \hat{\rho}_i^k \rho_m^k - b_{ij} \hat{\rho}_i^k \hat{\rho}_j^k \), the right side of (33) is expressed as
\[ - \sum_{j=1}^{n} \hat{B}^k_{ij} v^{k+1}_j = - \sum_{j=1}^{n} \hat{B}^{k+1}_{ij} v^{k+1}_j - \hat{B}^k_{in} v^{k+1}_n = - \sum_{j=1}^{n-1} \hat{B}^k_{ij} (v^{k+1}_j - v^{k+1}_n). \quad (34) \]

Using (13) we get
\[ - \sum_{j=1}^{n-1} \hat{B}^k_{ij} (v^{k+1}_j - v^{k+1}_n) = - \sum_{j=1}^{n-1} \hat{B}^k_{ij} (v^{k+1}_j + \frac{1}{\hat{\rho}_n^k} \sum_{m=1}^{n-1} \hat{\rho}_m^k v^{k+1}_m) \]
\[ = - \sum_{j=1}^{n-1} \hat{B}^k_{ij} \sum_{m=1}^{n-1} \left( \frac{1}{\hat{\rho}_n^k} \delta_{jm} + \frac{1}{\hat{\rho}_m^k} \right) \hat{\rho}_m^k v^{k+1}_m = - \sum_{j=1}^{n-1} \hat{B}^k_{ij} \hat{Q}^k_{jm} \hat{\rho}_m^k v^{k+1}_m, \quad (35) \]
where $\hat{Q}_{jm} = Q_{jm}(\hat{\rho}) = \frac{1}{\rho_m} \delta_{jm} + \frac{1}{\rho_n}$. By direct calculation it is shown that $\hat{Q}_{jm}$ is invertible with inverse $(\hat{Q}^k)^{-1}_{ij} = \left( \delta_{ij} \hat{\rho}_i - \frac{\hat{\rho}_i^k \hat{\rho}_j^k}{\sum_{j=1}^{p} \hat{\rho}_j^k} \right)$. The left side of (33) is rewritten for $i \neq n$ as

$$
\sum_{j=1}^{n} \left( \delta_{ij} \hat{\rho}_i^k - \frac{\hat{\rho}_i^k \hat{\rho}_j^k}{\sum_{j=1}^{p} \hat{\rho}_j^k} \right) D_h \log \rho_{j}^{k+1}
= \sum_{j=1}^{n-1} (\hat{Q}^k)^{-1}_{ij} D_h \log \rho_{j}^{k+1} - \frac{\hat{\rho}_i^k (\sum_{j=1}^{n} \hat{\rho}_j^k - \sum_{j=1}^{n-1} \hat{\rho}_j^k)}{\sum_{j=1}^{n} \hat{\rho}_j^k} D_h \log \rho_{n}^{k+1}
= \sum_{j=1}^{n-1} (\hat{Q}^k)^{-1}_{ij} D_h (\log \rho_{j}^{k+1} - \log \rho_{n}^{k+1}).
$$

This leads to expressing (33) as

$$
\sum_{j=1}^{n-1} (\hat{Q}^k)^{-1}_{ij} D_h (\log \rho_{j}^{k+1} - \log \rho_{n}^{k+1}) = - \sum_{j,m=1}^{n} \hat{B}^k_{ij} \hat{Q}_{jm} \hat{\rho}_m^{k} \hat{v}_m^{k+1}.
$$

Since $\hat{B}^k$ and $\hat{Q}^k = (\hat{Q}^k)^T$ are invertible, we conclude that (31) holds.

We adopt the notation

$$
\hat{f} = (f_1, \ldots, f_{n-1}) \text{ for } f = (f_1, \ldots, f_n).
$$

With Lemma 4, the scheme (11)-(13) can be written as

$$
\frac{\hat{\rho}_{i}^{k+1} - \hat{\rho}_{i}^{k}}{\Delta t} = -d_h \left( \hat{B}^k \hat{D}_h \left( \frac{1}{h} \frac{\partial F_h}{\partial \hat{\rho}} (\hat{\rho}_{i}^{k+1}) \right) \right),
$$

where

$$
F_h = F_h(\hat{\rho}) := \left( \sum_{j=1}^{n} \rho_j \log \rho_j \right) + \left( \frac{1}{n} \sum_{i=1}^{n} \rho_i \right) \log \left( \frac{1}{n} \sum_{i=1}^{n} \rho_i \right).
$$

2.4. Consistency. Let $(P, V)$ be the exact smooth solution of the equations (1)-(2) in the space $P, V \in C^1_{p,x}(0,T \times \mathbb{T})$. The values at grid points are $\hat{P}_{i}^{k} := P_i(x_i, k\Delta t), V_{i}^{k} := V_i(x_i, k\Delta t)$. The local truncation errors are defined by

$$
\tau_i^1 = \frac{P_{i}^{k+1} - P_{i}^{k}}{\Delta t} + d_h (\hat{P}_{i}^{k} V_{i}^{k+1}),
$$

$$
\tau_i^2 = D_h \log P_{i}^{k+1} + \frac{1}{\sum_{n=1}^{n} \hat{P}_{j}^{k}} \sum_{j=1}^{n} \hat{P}_{j}^{k} \log P_{i}^{k+1} + \sum_{j=1}^{n} b_{ij} \hat{P}_{i}^{k} (V_{i}^{k+1} - V_{j}^{k+1}),
$$

$$
\tau_i^3 = \sum_{i=1}^{n} \hat{P}_{i}^{k} V_{i}^{k+1}.
$$

We have the following lemma.

Lemma 5. Suppose the solutions $(P, V)$ to the system (1)-(3) are smooth in time and space, with $P, V \in C^1_{p,x}$ and $P_i(x, t) > 0$ for $x \in \mathbb{T}$ and $t > 0$ and for any $i = 1, \ldots, n$. Suppose $(P, V)$ satisfies the condition (4). Then the local truncation errors satisfy

$$
|\tau_i^1|, |\tau_i^2|, |\tau_i^3| \leq C(\Delta t + h^2).
$$

Here $C > 0$ is a positive constant depending on $(P, V)$.

An elementary verification is deferred to Appendix A.
3. Optimization formulation

3.1. Formulation via an optimization problem. In this section, we give an optimization formulation of the scheme \((11)-(13)\). We recall that the system \((1)-(3)\) can be written as the gradient flow of the energy functional \((9)\), see \([10]\). Consider the minimization problem

\[
\rho^{k+1} = \arg\min_{\rho \geq 0} \left\{ \frac{1}{\Delta t} \int_{\Omega_d} \sum_{i,j=1}^{n} \frac{1}{4} b_{ij} \rho^k_i \rho^k_j (w_i - w_j)^2 dx + F(\rho) \right\},
\]

with \(F(\rho)\) defined in \([9]\), subject to the constraints

\[
\rho_i - \rho^k_i + \nabla \cdot (\rho^k_i w_i) = 0, \quad i = 1, \ldots, n, \quad \text{and} \quad \sum_{i=1}^{n} \rho^k_i w_i = 0.
\]

The idea is to calculate minimizers of the free energy penalized by the work consumed by friction. The variational scheme is related to the Jordan-Kinderlehrer-Otto scheme \([12]\), an analogy due to the Benamou-Brenier interpretation \([1]\) of the Monge-Kantorovich mass transfer problem. There is however one important difference, as the frictional dissipation is more elaborate in the multi-component mixture situation.

The minimizers of the above constraint problem can be calculated by considering the min-max augmented Lagrangian

\[
\min_{\rho, w} \max_{\alpha, \beta} L(\rho, w, \alpha, \beta) = \frac{1}{\Delta t} \int_{\Omega_d} \sum_{i,j=1}^{n} \frac{1}{4} b_{ij} \rho_i^k \rho_j^k (w_i - w_j)^2 + \sum_{j=1}^{n} \rho_j \log \rho_j \, dx + \int_{\Omega_d} \alpha \sum_{i=1}^{n} \rho^k_i w_i \, dx + \int_{\Omega_d} \sum_{i=1}^{n} (\beta_i (\rho_i - \rho^k_i) - \nabla \beta_i \cdot (\rho^k_i w_i)) \, dx,
\]

Computing the variational derivatives gives:

\[
\frac{\delta L}{\delta \rho_i} = 0 \quad \text{implies} \quad \log \rho_i + 1 + \beta_i = 0,
\]

\[
\frac{\delta L}{\delta w_i} = 0 \quad \text{implies} \quad \frac{1}{\Delta t} \sum_{j=1}^{n} b_{ij} \rho_i^k \rho_j^k (w_i - w_j) + \alpha \rho_i^k - \rho_i^k \nabla \beta_i = 0,
\]

\[
\frac{\delta L}{\delta \alpha} = 0 \quad \text{implies} \quad \sum_{i=1}^{n} \rho_i^k w_i = 0,
\]

\[
\frac{\delta L}{\delta \beta_i} = 0 \quad \text{implies} \quad \rho_i - \rho_i^k + \nabla \cdot (\rho_i^k w_i) = 0.
\]

Taking \(v_i = w_i/\Delta t\), we get

\[
\frac{\rho_i^{k+1} - \rho_i^k}{\Delta t} + \nabla \cdot (\rho_i^k v_i^{k+1}) = 0,
\]

\[
- \sum_{j=1}^{n} b_{ij} \rho_i^k \rho_j^k (v_i^{k+1} - v_j^{k+1}) = \rho_i^k \nabla \log \rho_i^{k+1} - \sum_{j=1}^{n} \rho_j^k \nabla \log \rho_i^{k+1}.
\]

The latter corresponds to an implicit-explicit discretization in time of the system \((1)-(3)\).

Next we will give details of the optimization formulation for the fully discretized scheme \((11)-(13)\).

We prove the following theorem.

**Theorem 1.** Assume \(b_{ij} > 0\) and \(b_{ij} = b_{ji}\) for \(i \neq j\) and \(i, j = 1, \ldots, n\). Given \(\rho^k \in C_{\text{per}}\) with \(\rho^k > 0\). There exists \(\delta_0 > 0\) such that \(\rho^{k+1} > 0\) is a solution of the numerical scheme \((11)-(13)\) if
and only if it is a minimizer of the optimization problem:

\[
\rho^{k+1} = \arg\min_{(\rho, w) \in K_d} \left\{ J = \frac{1}{4\Delta t} \sum_{i,j=1}^{n} b_{ij} \rho_i^k \rho_j^k (w_i - w_j)^2 \right\} + F_h(\rho),
\]

where \(F_h(\rho) = (\sum_{i=1}^{n} \rho_i \log \rho_i)\), and

\[
K_d = \left\{ (\rho, w) : \rho \in \mathcal{C}_\text{per}^n, w \in \mathcal{E}_\text{per}^n, \rho_i, \rho_i - \rho_i, \rho_i, \rho_i + d_h(\rho_i^k w_i) \right\}
\]

for any \(0 < \delta \leq \delta_0\).

We first prove a lemma that will be used later in the proof.

**Lemma 6.** Suppose \(\Phi\) is a \((n-1) \times (n-1)\) symmetric positive definite matrix, with \(\Phi_{ij} \in \mathcal{C}_\text{per}\) for \(i, j = 1, \ldots, n-1\). Suppose \(\phi \in \mathcal{C}_\text{per}^{n-1}\) is bounded in \(L^\infty\) satisfying \(\|\phi\|_{L^\infty} \leq M\), where \(\|\cdot\|_{L^\infty}\) is defined by

\[
\|\phi\|_{L^\infty} := \max_{i=1, \ldots, n-1} |\phi_i|.
\]

Then the following estimate holds

\[
\|L_{\Phi}^{-1} \phi\|_{L^\infty} \leq \frac{C_M}{\lambda_{\min}} h^{-\frac{\gamma}{2}} (n-1)^{\frac{\gamma}{2}},
\]

where \(C > 0\) is a constant independent of \(h\), \(\lambda_{\min}\) the minimum of all eigenvalues of \(\Phi\):

\[
\lambda_{\min} = \min_{\ell=1, \ldots, N} \left\{ \lambda_\ell : \lambda_\ell \text{ is the eigenvalue of } (\Phi_{ij, \ell+\frac{1}{2}})_{(n-1) \times (n-1)} \right\}.
\]

**Proof.** Since \(\|\phi\|_{L^\infty} \leq M\),

\[
|\phi_i|_{L^2}^2 = h \sum_{i=1, \ldots, n-1} |\phi_i_{\ell}|^2 = h \sum_{i=1, \ldots, n-1} |M|^2 \leq (n-1) h N |M|^2 = (n-1) L |M|^2.
\]

Set \(g = \phi \in \mathcal{C}_\text{per}^{n-1}\), and \(f = L_{\Phi}^{-1} g\) in (26), we get

\[
\|\phi\|_{L^2}^2 = |D_h f, \Phi D_h f|.
\]

Since \(\Phi\) is positive definite so its minimum eigenvalues \(\lambda_{\min} > 0\), we get

\[
\lambda_{\min} |D_h f|_{L^2}^2 \leq |D_h f, \Phi D_h f| = -(f, d_h(\Phi D_h f)) = (f, \phi) \leq \|f\|_{L^2} \|\phi\|_{L^2}.
\]

The use of the discrete Poincaré inequality gives \(\|f\|_{L^2} \leq C_P \|D_h f\|_{L^2}\). Therefore, we get

\[
\|D_h f\|_{L^2} \leq \frac{C_P}{\lambda_{\min}} \|\phi\|_{L^2}.
\]

Using an inverse inequality leads to

\[
\|f\|_{L^\infty} \leq C_1 h^{-\frac{\gamma}{2}} \|D_h f\|_{L^2} \leq \frac{C_1 C_P}{\lambda_{\min}} h^{-\frac{\gamma}{2}} L^\frac{\gamma}{2} M (n-1)^{\frac{\gamma}{2}} \leq \frac{C_M}{\lambda_{\min}} h^{-\frac{\gamma}{2}} (n-1)^{\frac{\gamma}{2}}.
\]

**Proof of Theorem 3.** The proof is divided into three steps. In the first two steps, we prove that the optimization problem (38) has a unique interior minimizer and, in the last step, we prove that this minimizer is equivalent to the solution of the numerical scheme (11)–(13).

**Step 1. Existence of the optimization problem.** First we show existence for the optimization problem (38) for any \(\delta > 0\). Notice that the objective function \(J\) in (38) is convex in \(w\) but it is not strictly convex. However, we can rewrite the optimization problem by using the first \(n-1\) components of \(w\) and get an equivalent convex optimization problem. We introduce

\[
W = (W_1, \ldots, W_n), \quad W_i = \rho_i^k w_i, \quad i = 1, \ldots, n,
\]

and so \(\sum_{i=1}^{n} W_i = 0\). We adopt the notation (36) and define \(\tilde{W} = (W_1, \ldots, W_{n-1})\). We have the following lemma.
Lemma 7. The following formula holds:
\[
I(\hat{W}) := \frac{1}{2} \sum_{i,j=1}^{n} b_{ij} \hat{\rho}_{ij}^{k}(w_i - w_j)^2 = \hat{W}^{T}(\hat{Q}^{k})^{T} \hat{B}^{k} \hat{Q}^{k} \hat{W} = \hat{W}^{T}(\hat{B}^{k})^{-1} \hat{W}.
\] (39)

For \( \hat{B}^{k} \succ 0 \), the function \( I : \mathbb{R}^{n-1} \to \mathbb{R}^{+} \) is strictly convex.

Proof. By the assumption that \( b_{ij} \) is symmetric, the following formula holds
\[
\frac{1}{2} \sum_{i,j=1}^{n} b_{ij} \hat{\rho}_{ij}^{k}(w_i - w_j)^2 = \sum_{j=1}^{n} w_j \sum_{i=1}^{n} b_{ij} \hat{\rho}_{ij}^{k}(w_i - w_j).
\]
Recalling (34), (35), we also have
\[
\sum_{j=1}^{n-1} \hat{B}_{ij}^{k} \hat{Q}_{jm} \hat{\rho}_{jm} w_m = \sum_{i,j,m=1}^{n-1} \hat{B}_{ij}^{k} \hat{Q}_{jm} \hat{\rho}_{jm} w_m = \hat{W}^{T}(\hat{Q}^{k})^{T} \hat{B}^{k} \hat{Q}^{k} \hat{W}.
\]

Therefore,
\[
\frac{1}{2} \sum_{i,j=1}^{n} b_{ij} \hat{\rho}_{ij}^{k}(w_i - w_j)^2 = \sum_{j=1}^{n-1} \hat{B}_{ij}^{k} \hat{Q}_{jm} \hat{\rho}_{jm} w_m.
\]

Notice that \( \hat{B}^{k} \) is a symmetric strictly diagonally dominant matrix with positive diagonal entries since \( \hat{\rho}^{k} > 0 \) and thus is positive definite. Because of this and since \( \hat{Q}^{k} \) is non-singular, we have
\[(\hat{Q}^{k})^{T} \hat{B}^{k} \hat{Q}^{k} \] is positive definite.

Therefore, (39) is a convex function of \( \hat{W} \). \( \square \)

We also need a lemma on the convexity of the discretized energy function \( F_{h}(\hat{\rho}) \), defined by (37) that incorporates the constraint \( \sum_{i=1}^{n} \hat{\rho}_{i} = 1 \).

Lemma 8. The energy function \( F_{h} = F_{h}(\hat{\rho}) \) is a convex function of \( \hat{\rho} \).

Proof. Considering the function
\[
f = \sum_{i=1}^{n-1} \rho_{i} \log \rho_{i} + \rho_{n} \log \rho_{n}, \quad \rho_{n} = 1 - \sum_{i=1}^{n-1} \rho_{i},
\]
we have
\[
\frac{\partial f}{\partial \rho_{i}} = \log \rho_{i} + 1 - (\log \rho_{n} + 1) = \log \rho_{i} - \log \rho_{n}, \quad \frac{\partial^2 f}{\partial \rho_{i} \partial \rho_{j}} = \frac{1}{\rho_{i}} \delta_{ij} + \frac{1}{\rho_{n}}.
\]
Since for any \( z \in \mathbb{R}^{n-1} \) and \( z \neq 0 \),
\[
\sum_{i,j=1}^{n-1} \frac{\partial^2 f}{\partial \rho_{i} \partial \rho_{j}} z_{i} z_{j} = \sum_{i,j=1}^{n-1} \left( \frac{1}{\rho_{i}} \delta_{ij} + \frac{1}{\rho_{n}} \right) z_{i} z_{j} = \sum_{i=1}^{n-1} \frac{1}{\rho_{i}} z_{i}^{2} + \frac{1}{\rho_{n}} \left( \sum_{i=1}^{n-1} z_{i} \right)^{2} > 0,
\]
the function \( f \) is a convex function of \( \hat{\rho} \). Therefore, \( F_{h}(\hat{\rho}) \) is convex in \( \hat{\rho} \). \( \square \)
Using Lemmas 7 and 8 we deduce that the optimization problem (38) is equivalent to
\[
\min_{(\tilde{\rho}, \tilde{W}) \in \tilde{K}_\delta} \left\{ J = \frac{1}{2\Delta t} \left[ \tilde{W}^T (\tilde{Q}^k)^T \tilde{B}^k \tilde{Q}^k \tilde{W} \right] + F_h(\tilde{\rho}) \right\},
\] (40)
where
\[
\tilde{K}_\delta = \{(\tilde{\rho}, \tilde{W}) : \tilde{\rho} \in \mathcal{C}_{\text{per}}^{n-1}, \tilde{W} \in \mathcal{C}_{\text{per}}^{n-1}, \rho_{i,\ell} \geq \delta, \sum_{i=1}^{n-1} \rho_{i,\ell} \leq 1 - \delta \text{ and } \rho_{i,\ell} - \rho_{i,\ell}^k + d_h(W_i)_\ell = 0, \forall i = 1, \ldots, n-1, \ell = 1, \ldots, N\}.
\]
Due to the above lemmas, the objective function \( J \) is a convex function of \( \tilde{W} \) and \( \tilde{\rho} \) (note that \((\tilde{Q}^k)^T \tilde{B}^k \tilde{Q}^k \) is a fixed matrix determined from the previous step). The domain \( \tilde{K}_\delta \) is affine in \( \tilde{W} \) and it is convex and bounded in \( \tilde{\rho} \). The optimization problem (40) has a unique minimizer according to standard optimization theory [4]. Since the problems (38) and (40) are equivalent, there also exists a unique solution to the optimization problem (38).

**Step 2. The minimizer does not touch the boundary.** Next, we show that there exists a constant \( \delta_0 > 0 \) such that the solution of the optimization problem (38) could not touch the boundary of \( K_\delta \) for \( \delta \leq \delta_0 \). Recall that on the set \( \tilde{K}_\delta \),
\[
\rho_i - \rho_i^k + d_h(W_i)_\ell = 0.
\]
Hence, if we set
\[
\tilde{W} = \tilde{D}^k D_h \tilde{f}, \quad \tilde{g} = \tilde{\rho} - \tilde{\rho}^k \in \mathcal{C}_{\text{per}}^{n-1},
\]
then according to the definition (20),
\[
\left[ \tilde{W}^T (\tilde{Q}^k)^T \tilde{B}^k \tilde{Q}^k \tilde{W} \right] = [(D_h \tilde{f})^T \tilde{D}^k D_h \tilde{f}] = \|\tilde{\rho} - \tilde{\rho}^k\|^2_{C_{\text{per}}^{n-1}}.
\] (41)
Therefore, the optimization problem (40) is equivalent to
\[
\min_{\tilde{\rho} \in \tilde{K}_\delta} \left\{ J = \frac{1}{2\Delta t} \|\tilde{\rho} - \tilde{\rho}^k\|^2_{C_{\text{per}}^{n-1}} + F_h(\tilde{\rho}) \right\},
\] (42)
over the set
\[
\hat{K}_\delta = \left\{ \tilde{\rho} : \tilde{\rho} - \tilde{\rho}^k \in \mathcal{C}_{\text{per}}^{n-1}, \rho_{i,\ell} \geq \delta, \sum_{i=1}^{n-1} \rho_{i,\ell} \leq 1 - \delta, \forall i = 1, \ldots, n-1, \ell = 1, \ldots, N \right\}.
\]
Recall the notation \( \tilde{\rho} = (\rho_1, \ldots, \rho_{n-1}) \) stands for the vector of the first \( n - 1 \) densities which are computed at the grid points \( l = 1, \ldots, N \). The density \( \rho_n \) appears in the formulation (42) only indirectly through the constraint (5). Also, \( \tilde{\rho} - \tilde{\rho}^k \in \mathcal{C}_{\text{per}}^{n-1} \) means \( \sum_{\ell=1}^{N} (\rho_{i,\ell} - \rho_{i,\ell}^k) = 0 \) for any \( i = 1, \ldots, n-1 \).

Let \( \tilde{\rho}^* \in \hat{K}_\delta \) be a minimizer of the optimization problem (42). We will show that \( \tilde{\rho}^* \) does not lie on the boundary of \( \hat{K}_\delta \). If it lies on the boundary:

(i) either \( \rho_{i,\ell}^* = \delta \) for some \( i = 1, \ldots, n-1 \) at some grid point \( \ell \),

(ii) or \( \sum_{i=1}^{n-1} \rho_{i,\ell}^* = 1 - \delta \) at some grid point \( \ell \).

First consider the case (i). Suppose that \( \tilde{\rho}^* \) touches the boundary at the grid point \( \ell_0 \) for the \( i_0 \)-th component, that is
\[
\rho_{i_0,\ell_0}^* = \delta.
\] (43)
We calculate the directional derivative of the objective function $J$ at $\bar{\nu}^*$ along the direction $\{\nu : \nu \in \mathbb{R}^{(n-1) \times N}\}$ with $\bar{\nu}^* + su \in \tilde{K}_\delta$ as

$$
d \frac{d}{ds} J(\bar{\nu}^* + su) \bigg|_{s=0} = \frac{d}{ds} \bigg|_{s=0} \left( \frac{1}{2\Delta t} \|\bar{\nu}^* + su - \bar{\rho}^k\|^2 + F_h(\bar{\nu}^* + su) \right)
$$

$$
= \frac{1}{\Delta t} \left\langle \mathcal{L}^{-1}_{D\delta}(\bar{\nu}^* - \bar{\rho}^k), \nu \right\rangle + \sum_{i=1}^{n-1} \left( \log \rho_i^* + 1 - \log \left( 1 - \sum_{i=1}^{n-1} \rho_i^* \right) - 1, \nu_i \right)
$$

We divide into the following two cases:

(a) \[ \sum_{i=1}^{n-1} \rho_i^* f_0 \geq \frac{1}{2} \]

(b) \[ \sum_{i=1}^{n-1} \rho_i^* f_0 < \frac{1}{2} \]

**Case (a) and (b).** Suppose $\{\rho_{i,\ell}^* \}_{i=1}^{n-1}$ achieves its maximum at the $i_1$-th component while $\{\rho_{i_0,\ell}^* \}_{\ell=1}^{N}$ achieves its maximum at $\ell_1$. Define $\nu$ by

$$
\nu_{i,\ell} = \begin{cases} 
1, & \text{for } i = i_0, \ell = \ell_0, \\
-1, & \text{for } i = i_1, \ell = \ell_0, \\
-1, & \text{for } i = i_0, \ell = \ell_1, \\
1, & \text{for } i = i_1, \ell = \ell_1, \\
0, & \text{otherwise.}
\end{cases}
$$

Taking a variation in this direction, (44) becomes

$$
\frac{1}{\Delta t} \left\langle \mathcal{L}^{-1}_{D\delta}(\bar{\nu}^* - \bar{\rho}^k), \nu \right\rangle + \sum_{i=1}^{n-1} \left( \log \rho_i^* + 1 - \log \left( 1 - \sum_{i=1}^{n-1} \rho_i^* \right) - 1, \nu_i \right)
$$

Since $\{\rho_{i,\ell}^* \}_{i=1}^{n-1}$ achieves its maximum for the $i_1$-th component, in the case (b) $\sum_{i=1}^{n-1} \rho_i^* f_0 \geq \frac{1}{2}$ implies

$$
\rho_{i_1,\ell_0}^* \geq \frac{1}{2(n-1)}.
$$

Since $\{\rho_{i_0,\ell}^* \}_{\ell=1}^{N}$ achieves its maximum at the grid point $\ell_1$ and $\bar{\nu}^* - \bar{\rho}^k \in \mathcal{C}_{per}^{n-1}$,

$$
\rho_{i_0,\ell_1}^* \geq \frac{1}{N} \sum_{\ell=1}^{N} \rho_{i,\ell}^* = \frac{1}{N} \sum_{\ell=1}^{N} \rho_{i,\ell}^* \geq \frac{m}{N}
$$

where $m$ is set to be $m := \min_{i \in \{1, \ldots, n-1\}} \left\{ \frac{1}{N} \sum_{\ell=1}^{N} \rho_{i,\ell}^k \right\}$. Moreover, for $\bar{\nu}^* \in \tilde{K}_\delta$ the constraint $\sum_{i=1}^{n-1} \rho_i^* f_0 \leq 1 - \delta$ implies

$$
\rho_{i_1,\ell_1}^* < 1.
$$

Next, we show that for $\delta$ satisfying

$$
\delta \leq \min \left\{ \frac{m}{2hN}, \frac{1}{4(n-1)} \right\},
$$

(49)
if $s > 0$ is selected sufficiently small and $\nu$ as above we have $\tilde{\rho}^* + s\nu \in \hat{K}_\delta$. Indeed,

$$
\rho^*_{i,\ell} + s = \delta + s \geq \delta, \quad \rho^*_{i,\ell_1} + s \geq \delta + s,
$$

$$
\rho^*_{i,\ell} - \nu \leq \frac{m}{hN} - \nu \leq \rho^*_{i,\ell_0} - \nu \leq \frac{1}{2(n-1)} - \nu \leq \delta,
$$

$$
\sum_{i=1}^{n-1} (\rho^*_{i,\ell_0} + s\nu_{i,\ell_0}) = \sum_{i=1}^{n-1} (\rho^*_{i,\ell_1} + s\nu_{i,\ell_1}) = \sum_{i=1}^{n-1} \rho^*_{i,\ell_1} \leq 1 - \delta,
$$

and note that (47) still holds in the present setting. Using (43), (b), (47), and the inequality

$$
\lambda_k \leq \sum_{i=1}^{n} \nu_{i,\ell} = \sum_{i=1}^{n} \nu_{i,\ell_0} = \lambda_{k_{\min}}\Delta t h^{-\frac{1}{2}} (n-1)^{\frac{1}{2}} + \log \delta - \log \frac{1}{2(n-1)} - \log \frac{m}{hN} + \log 1.
$$

Here $\lambda_{k_{\min}}$ is the minimum eigenvalue of $\hat{D}^k$. Taking

$$
\delta_0 \leq \min \left\{ \frac{m}{4(n-1)hN} e^{-\frac{4C}{\lambda_{k_{\min}}} h^{-\frac{1}{2}} (n-1)^{\frac{1}{2}}}, \frac{m}{2hN}, \frac{1}{4(n-1)} \right\},
$$

we have for $\delta \leq \delta_0$, $\tilde{\rho}^* + s\nu \in \hat{K}_\delta$ and

$$
\frac{1}{h} \frac{d}{ds} J(\tilde{\rho}^* + s\nu) \bigg|_{s=0} \leq -\log 2 < 0.
$$

This contradicts the assumption that $\tilde{\rho}^*$ is a minimizer, and so the situation (a) cannot occur.

**Case (b) and (c).** Again $\rho_{i,\ell_0} = \delta$ and suppose now that $\{\rho^*_{i,\ell}\}_{i=1}^{n}$ achieves its maximum at the $\ell_1$-th grid point. We take

$$
\nu_{i,\ell} = \begin{cases} 
1, & \text{for } i = i_0, \ell = \ell_0, \\
-1, & \text{for } i = i_0, \ell = \ell_1, \\
0, & \text{otherwise}, 
\end{cases}
$$

and note that (47) still holds in the present setting. Using (43), (b), (47), and the inequality

$$
1 - \sum_{i=1}^{n-1} \rho^*_{i,\ell_1} \leq 1 - (n-1)\delta \leq 1,
$$

we obtain

$$
\frac{1}{h} \frac{d}{ds} J(\tilde{\rho}^* + s\nu) \bigg|_{s=0} = \frac{1}{\Delta t} (C_{\hat{D}^k}^{-1}(\tilde{\rho}^* - \hat{\rho}^k))_{i_0,\ell_0} + \log \rho^*_{i_0,\ell_0} - \log \left(1 - \sum_{i=1}^{n-1} \rho^*_{i,\ell_0}\right)

- \frac{1}{\Delta t} (C_{\hat{D}^k}^{-1}(\tilde{\rho}^* - \hat{\rho}^k))_{i_0,\ell_1} - \log \rho^*_{i_0,\ell_1} + \log \left(1 - \sum_{i=1}^{n-1} \rho^*_{i,\ell_1}\right)

\leq \frac{4C}{\lambda_{k_{\min}}} h^{-\frac{1}{2}} (n-1)^{\frac{1}{2}} + \log \delta - \log \frac{1}{2} - \log \frac{m}{hN} + \log 1

\leq \frac{4C}{\lambda_{k_{\min}}} h^{-\frac{1}{2}} (n-1)^{\frac{1}{2}} + \log \delta - \log \frac{m}{2hN}.
$$

Taking

$$
\delta_0 \leq \min \left\{ \frac{m}{4hN} e^{-\frac{4C}{\lambda_{k_{\min}}} h^{-\frac{1}{2}} (n-1)^{\frac{1}{2}}}, \frac{m}{2hN} \right\}
$$

leads to $\tilde{\rho}^* + s\nu \in \hat{K}_\delta$ and

$$
\frac{1}{h} \frac{d}{ds} J(\tilde{\rho}^* + s\nu) \bigg|_{s=0} = -\log 2 < 0,
$$

which contradicts the hypothesis that $\tilde{\rho}^*$ is a minimizer; so the situation (b) cannot occur.
Taking now (44) and using (53), (54), (55), Lemma 6, and the inequality \( \delta \leq 1/2 \), equation (46) holds, that is

\[
\rho_{n,\ell_0}^* \geq \frac{1 - \delta}{n - 1} \geq \frac{1}{2(n - 1)}.
\]

Setting \( \rho_{\min} := \min_{\ell_1, \ldots, n} \rho_{\ell_1, \ell}^k > 0 \), we have \( \sum_{i=1}^{n-1} \rho_{i,\ell} = 1 - \rho_{n,\ell}^k \leq 1 - \rho_{\min}^k \). Since \( \tilde{\rho}^* - \tilde{\rho}^k \in \mathcal{C}_{\text{per}}^{n-1} \), we have

\[
\sum_{\ell=1}^{n} \sum_{i=1}^{n-1} \rho_{i,\ell}^* = \sum_{\ell=1}^{n} \sum_{i=1}^{n-1} \rho_{i,\ell}^k \leq N(1 - \rho_{\min}^k).
\]

Suppose \( \left\{ \sum_{i=1}^{n-1} \rho_{i,\ell}^* \right\}_{\ell=1}^{N} \) achieves its minimum at the grid point \( \ell_1 \). Then using (53) it follows for \( \delta \leq \frac{1}{2} \rho_{\min}^k \),

\[
\sum_{i=1}^{n-1} \rho_{i,\ell_1}^* \leq \frac{1}{N-1} \sum_{\ell \neq \ell_0} \sum_{i=1}^{n-1} \rho_{i,\ell}^*
\]

\[
= \frac{1}{N-1} \left( \sum_{\ell=1}^{n} \sum_{i=1}^{n-1} \rho_{i,\ell}^* - \sum_{i=1}^{n-1} \rho_{i,\ell_0}^* \right)
\]

\[
\leq \frac{1}{N-1} \left( N(1 - \rho_{\min}^k) - (1 - \delta) \right)
\]

\[
\leq 1 - \frac{N \rho_{\min}^k - \delta}{N - 1}
\]

\[
\leq 1 - \frac{2N - 1}{2(N - 1)} \rho_{\min}^k.
\]

Taking now

\[
\nu_{i,\ell} = \begin{cases} -1, & \text{for } i = i_0, \ \ell = \ell_0, \\ 1, & \text{for } i = i_0, \ \ell = \ell_1, \\ 0, & \text{otherwise}, \end{cases}
\]

into (44) and using (53), (54), (55), Lemma 6 and the inequality \( \rho_{i,\ell_1}^* \leq 1 - \delta \leq 1 \) we obtain

\[
\frac{1}{h \Delta t} \frac{d}{ds} J(\tilde{\rho}^* + s \nu) \bigg|_{s=0}
\]

\[
= -\frac{1}{\Delta t} (\mathcal{L} \rho_{\min}^k (\tilde{\rho}^* - \tilde{\rho}^k))_{i_0,\ell_0} - \log \rho_{i_0,\ell_0} + \log \left( 1 - \sum_{i=1}^{n-1} \rho_{i,\ell_0}^* \right)
\]

\[
+ \frac{1}{\Delta t} (\mathcal{L} \rho_{\min}^k (\tilde{\rho}^* - \tilde{\rho}^k))_{i_0,\ell_1} + \log \rho_{i_0,\ell_1} - \log \left( 1 - \sum_{i=1}^{n-1} \rho_{i,\ell_1}^* \right)
\]

\[
\leq \frac{4C}{\lambda_{\min} \Delta t} h^{\frac{1}{2}} (n - 1)^{\frac{1}{2}} - \log \frac{1}{2(n - 1)} + \log \delta + \log 1 - \log 2N - 1 \rho_{\min}^k.
\]

Taking

\[
\delta_0 \leq \min \left\{ \frac{(2N - 1) \rho_{\min}^k}{8(N - 1)(n - 1)} e^{\frac{4C}{\lambda_{\min}} h^{\frac{1}{2}} (n - 1)^{\frac{1}{2}}} \rho_{\min}^k, \frac{1}{2} \rho_{\min}^k, \frac{1}{4(n - 1)} \right\},
\]

(56)
we see that for $\delta < \delta_0$ the above inequality becomes negative. In addition,

$$\rho_{i_0, t_0}^* - s \geq \frac{1}{2(n-1)} - s \geq \delta, \quad \rho_{i_0, t_1}^* + s \geq \delta + s \geq \delta,$$

$$\sum_{i=1}^n \rho_{i, t_0}^* - s = 1 - \delta - s \leq 1 - \delta, \quad \sum_{i=1}^{n-1} \rho_{i, t_1}^* + s \leq 1 - \frac{2N-1}{N-1} \delta + s \leq 1 - \delta,$$

imply that for $\delta < \delta_0$ the variation $\rho^* + sv \in \mathring{K}_\delta$ for sufficiently small $s > 0$. This contradicts the assumption that $\rho^*$ is a minimizer and thus case (ii) cannot occur.

In summary, setting $\delta_0$ to be the minimum among (50), (52) and (56) we conclude that (i) and (ii) cannot occur. Consequently, for $\delta \leq \delta_0$, the minimizer to the optimization problem (42), or equivalently (38), does not occur at the boundary.

Step 3. The equivalence with the numerical scheme. Any interior minimizer $\tilde{\rho}^*$ of (42) must satisfies

$$\left\langle \frac{\partial J}{\partial \rho}(\tilde{\rho}^*), \nu \right\rangle = 0,$$

for any $\nu \in C_{\text{per}}^{n-1}$ which is its tangent space, i.e., (44) equals zero. Due to the arbitrary choice of $\nu$, we get

$$\frac{1}{\Delta t} L_{\mathring{D}_{h}}^{-1}(\tilde{\rho}^* - \rho^*_k) + \log \tilde{\rho}_i^* - \log \left(1 - \sum_{j=1}^n \rho_j^*\right) = C_i,$$

with $C_i, i = 1, \ldots, n \in 1$ being constants, from which it follows that for $i = 1, \ldots, n - 1$,

$$\frac{\rho_i^* - \rho_k^*}{\Delta t} = -L_{\mathring{D}_{h}} \left(\log \tilde{\rho}^* - \log \left(1 - \sum_{j=1}^n \tilde{\rho}_j^*\right)\right)_i = \sum_{j=1}^{n-1} d_h(\tilde{D}_{ij} D_h (\log \tilde{\rho}_j^* - \log \rho_j^*)).$$

By Lemma 3 $\tilde{\rho}$ satisfies the numerical scheme (11)-(13).

Conversely, assume $\rho_{k+1}^* > 0$ is a solution of the numerical scheme (11)-(13), we can reverse the above calculation with $C_i = 0$ to show that (57) holds, which together with the fact that the convex optimization problem (42) has a unique interior minimizer, implies that $\rho_{k+1}^*$ is also the minimizer of (42), or equivalently of (38).

3.2. Properties of the scheme. The positivity-preserving and energy stability properties of the scheme follow directly from Theorem 1.

Theorem 2. Assume $\rho^1$ defined in (23) is positive, the solution of the numerical scheme (11)-(12) then satisfies

1. (Positivity-preserving) $\rho^k > 0$ for any $k \geq 1$,
2. (Unconditionally energy stability) the inequality

$$F_h(\rho^k) + \|\rho^k - \rho^{k-1}\|_{L_{\mathring{D}_{h}}^{-1}}^2 \leq F_h(\rho^{k-1})$$

holds for any $k \geq 1$.

Proof. 1. Starting from $\rho_0$, we apply Theorem 1 recursively to obtain

$$\rho^k \in K_{\delta_k},$$

for some constant $\delta_k$ that is chosen for each step by the minimum among (50), (52) and (56). This yields for every $k$,

$$\rho^k \in \bigcap_{k=1}^\infty K_{\delta_k} \subset K_0 \setminus \{0\},$$

so that $\rho^k > 0$.

2. Since the solution of the numerical scheme (11)-(13) is the minimizer of (42), we have

$$J(\rho^{k+1}) \leq J(\rho^k),$$

which is (58).
4. Multidimensional case

The scheme can be generalized to the multidimensional case and similar properties can be established. Before we present the multi-dimensional scheme, we introduce some notations following [24]. Consider two multidimensional grids defined by

\[ C^d := \mathbb{C} \times \cdots \times \mathbb{C}, \quad E^d := \mathbb{E} \times \cdots \times \mathbb{E}, \quad s = 1, \ldots, d, \]

and the functions on them

\[ C^d_{\text{per}} := \{ f : C^d \to \mathbb{R} \}, \quad E^d_{x,s,\text{per}} := \{ f : E^d_{x,s} \to \mathbb{R} \}, \quad \mathcal{E}^d := \left\{ f : \bigcup_{s=1}^d E^d_{x,s} \to \mathbb{R} \right\}, \]

as well as the vector functions, \((C^d_{\text{per}})^n := \{ f = (f_1, \ldots, f_n) : f_i \in C^d_{\text{per}} \text{ for } i = 1, \ldots, n \}, \quad (\mathcal{E}^d)^n := \{ f = (f_1, \ldots, f_n) : f_i \in \mathcal{E}^d_{\text{per}} \text{ for } i = 1, \ldots, n \}\). We also define the space \((C^d_{\text{per}})^n := \{ f \in (C^d_{\text{per}})^n : \sum_{\ell \in (1, \ldots, N)^d} f_{i, \ell} = 0, i = 1, \ldots, n \}\) for \(d \geq 1\).

We introduce the finite difference operators \(D_h : C^d_{\text{per}} \mapsto C^d_{\text{per}}\) and \(d_h : \mathcal{E}^d_{\text{per}} \mapsto C^d_{\text{per}}\) as

\[ D_h f_{\ell_1, \ldots, \ell_d} = \frac{f_{\ell_1, \ldots, \ell_d+1, \ldots, \ell_d} - f_{\ell_1, \ldots, \ell_d, \ldots, \ell_d}}{h}, \]

and

\[ d_h f_{\ell_1, \ldots, \ell_d} := \sum_{s=1}^d f_{\ell_1, \ldots, \ell_s+1, \ldots, \ell_d} - f_{\ell_1, \ldots, \ell_s, \ldots, \ell_d}. \]

We also define for \( f \in C^d_{\text{per}}, \quad \hat{f}_{\ell_1, \ldots, \ell_s+1, \ldots, \ell_d} = \frac{f_{\ell_1, \ldots, \ell_s+1, \ldots, \ell_d} + f_{\ell_1, \ldots, \ell_s, \ldots, \ell_d}}{2}, \quad s = 1, \ldots, d, \) so that \( \hat{f} \in \mathcal{E}^d_{\text{per}} \). We define the inner products

\[ \langle f, g \rangle := h^d \sum_{i=1}^n \sum_{\ell \in (1, \ldots, N)^d} f_{i, \ell} g_{i, \ell}, \quad \forall f, g \in (C^d_{\text{per}})^n, \]

\[ [f, g] := h^d \sum_{i=1}^n \sum_{\ell_1, \ldots, \ell_n=1}^N f_{i, \ell_1, \ldots, \ell_s+\frac{1}{2}, \ldots, \ell_d} g_{i, \ell_1, \ldots, \ell_s+\frac{1}{2}, \ldots, \ell_d}, \quad \forall f, g \in (C^d_{\text{per}})^n. \]

The following summation-by-parts formula holds for any \( f \in (C^d_{\text{per}})^n \) and \( \phi \in (\mathcal{E}^d_{\text{per}})^n \),

\[ \langle f, d_h \phi \rangle = -[D_h f, \phi]. \]

Next we define a norm on \((C^d_{\text{per}})^n)^{n-1}\). Suppose \( \Phi \) is a \((n-1) \times (n-1)\) symmetric positive definite matrix, with \( \Phi_{ij} \in \mathcal{E}^d_{\text{per}} \). We introduce the following operator

\[ \mathcal{L}_\Phi f = -d_h (\Phi_{ij} D_h f_j), \]

where the multiplication \( \Phi_{ij} D_h f_j \) is taken elementwise on the grid points. For any \( g \in (C^d_{\text{per}})^n \), let \( f \) be determined by \( g = \mathcal{L}_\Phi f \), we define the following norm

\[ \| g \|^2_{\mathcal{L}_\Phi^{-1}} := [D_h f, \Phi D_h f]. \]
With the above notations, the numerical scheme for the system (1)-(2) is
\[ \frac{\rho_i^{k+1} - \rho_i^k}{\Delta t} + d_h(\rho_i^k v_i^{k+1}) = 0, \]
\[ D_h \log \rho_i^{k+1} - \frac{1}{\sum_{i=1}^n \rho_i^k} \sum_{j=1}^n \rho_j^k D_h \log \rho_i^{k+1} = -\sum_{j=1}^n b_{ij} \rho_j^k (v_i^{k+1} - v_j^{k+1}), \]
subject to initial data
\[ \rho_{i,\ell}^0 = \rho_{i_0}(x_\ell), \quad i = 1, \ldots, n, \quad \ell = \{1, \ldots, N\}^d. \]
All properties proved for the one dimensional case carry over the \( d \)-dimensional case. The following theorem holds.

**Theorem 3.** Suppose \( \rho^0 > 0 \). The solution of the numerical scheme (60)-(62) satisfies

1. (Conservation of mass.) For \( k \geq 1 \),
   \[ \sum_{i=1}^n \rho_{i,\ell}^k = \sum_{i=1}^n \rho_{i,\ell}^0, \quad \text{for all} \ \ell \in \{1, \ldots, d\}^N, \]
   and
   \[ \sum_{\ell \in \{1, \ldots, d\}^N} \rho_{i,\ell}^k = \sum_{\ell \in \{1, \ldots, d\}^N} \rho_{i,\ell}^0, \quad \text{for all} \ i = 1, \ldots, n. \]

2. (Positivity-preserving.) For \( k \geq 1 \),
   \[ \rho^k > 0. \]

3. (Unconditional energy stability.) For \( k \geq 1 \), the following inequality holds:
   \[ F_h(\rho^k) + \|\rho^k - \rho^{k-1}\|_{\rho^k}^2 \leq F_h(\rho^{k-1}), \]
   where \( F_h(\rho) := (\sum_{i=1}^n \rho_i \log \rho_i) \).

The proof of this result is similar, and therefore deferred to Appendix A.

5. Numerical Examples

We numerically validate our theoretical findings using numerical examples in both one and two dimensions.

5.1. One dimension. We consider the numerical example on the unit torus \( T = [0,1] \) and take the initial condition similar as in [3] as

\[ \rho_{10}(x) = \begin{cases} 
0.8, & \text{for } 0 \leq x < 0.25 \\
1.6(0.75 - x), & \text{for } 0.25 \leq x < 0.5, \\
1.6(x - 0.25), & \text{for } 0.5 \leq x < 0.75, \\
0.8, & \text{for } 0.75 \leq x < 1, 
\end{cases} \]
\[ \rho_{20}(x) = 1 \times 10^{-4}, \]
\[ \rho_{30}(x) = 1 - \rho_{10}(x) - \rho_{20}(x). \]

We take the parameter \((b_{ij})_{n \times n}\) in the model to be
\[ b_{12} = b_{13} = \frac{1}{0.833}, \quad b_{23} = \frac{1}{0.168}. \]

The mesh size is taken to be \( h = 0.01 \) and time step \( \Delta t = 0.001 \).

Here we calculate for 500 time steps and the solutions reach equilibrium. The solution over time and the solution at \( x = 0.5 \) are plotted in Figure 5.1. In our numerical test we observe that the variations of the mass defined in Lemma 1 and Lemma 2 are of size \( 10^{-12} \sim 10^{-13} \), which confirms the mass conservation results. The energy function \( F_h(\rho) \) and the minimum value of \( \rho \) are plotted in Figure 5.2. Theorem 2 is verified. We fix \( \Delta t = 0.01 \) and calculate from \( h = 0.01 \) to \( h = 0.2 \) with 8 values in equally distributed logarithmically. We plot the numerical error at \( t = 0.5 \) with respect
to the real solution $\rho_0(x) = (0.7, 0.0001, 0.299)$ in Figure 5.3. The fitted curve showed that the scheme is approximately of order $h^2$. We also keep $h = 0.01$ fixed and compute the numerical error with $\Delta t$ ranging from 0.001 to 0.1. The result is plotted in Figure 5.3. We see that the numerical error is approximately linear in $\Delta t$.

5.2. Two dimensions. We take

\[
\rho_{10}(x, y) = \begin{cases} 
\frac{\sqrt{(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2}}{2} + \frac{1}{10}, & \text{for } \sqrt{(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2} \leq \frac{1}{8}, \\
\frac{3}{5}, & \text{otherwise},
\end{cases}
\]

$\rho_{20}(x, y) = 1 \times 10^{-4},$

$\rho_{30}(x, y) = 1 - \rho_{10}(x, y) - \rho_{20}(x, y).$
The mesh size is taken to be $h = 0.05$ and time step $\Delta t = 0.001$. We calculate for 500 time steps. The energy and minimum values are shown in Figure 5.4. We can see that the energy is decaying and the minimum values are all positive.

**References**

[1] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
[2] Dieter Bothe. On the Maxwell-Stefan approach to multicomponent diffusion. In *Parabolic problems*, pages 81–93. Springer, 2011.
[3] Laurent Boudin, Bérénice Grec, and Francesco Salvarani. A mathematical and numerical analysis of the Maxwell-Stefan diffusion equations. *Discrete and Continuous Dynamical Systems-Series B*, 17(5):1427–1440, 2012.
[4] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
[5] Wenbin Chen, Cheng Wang, Xiaoming Wang, and Steven M Wise. Positivity-preserving, energy stable numerical schemes for the Cahn-Hilliard equation with logarithmic potential. *Journal of Computational Physics: X*, 3:100031, 2019.

---

**Figure 5.4.** Energy (left) and Minimum value (right)
The following estimate holds

\[ C > 0 \] for any \( 0 < \delta \leq \delta_0 \).

The proof follows a similar strategy as the proof of Theorem 1 for the one dimensional case. We establish a multidimensional version of Lemma 8.

**Lemma 9.** Suppose \( \Phi \) is a \((n-1) \times (n-1)\) symmetric positive definite matrix, with \( \Phi_{ij} \in \mathcal{C}_{\text{per}}^d \).

Suppose \( \phi \in (\mathcal{C}_{\text{per}}^d)^{n-1} \) satisfies \( \| \phi \|_{L^\infty} \leq M \),

\[
\| \phi \|_{L^\infty} := \max_{i=1,\ldots,n, s=1,\ldots,d} |\phi_{i_1\ldots i_n}|.
\]

The following estimate holds

\[
\| \mathcal{L}_{\Phi}^{-1} \phi \|_{L^\infty} \leq \frac{CM}{\lambda_{\min}} h^{-\frac{1}{2}} (n-1)^{\frac{1}{2}},
\]

where \( C > 0 \) depends only on the domain, \( \lambda_{\min} \) is the minimum of the eigenvalues of \( \Phi \) over all grid points:

\[ \lambda_{\min} = \min_{\ell_s = 1, \ldots, N} \left\{ \lambda_{\ell_1, \ldots, \ell_d + \frac{1}{2}, \ldots, \ell_d} \text{ the eigenvalue of } (\Phi_{ij}, \ell_1, \ldots, \ell_d, \frac{1}{2}, \ldots, \ell_d) \right\} \]

**Proof.**

\[
\| \phi \|_{L^2}^2 := h^d \sum_{i=1,\ldots,n-1} \sum_{\ell_i, s=1,\ldots,N} |\phi_{i_1\ldots i_n}|^2
\]

\[
\leq h^d \sum_{i=1,\ldots,n-1} \sum_{\ell_s = 1,\ldots,N} \sum_{\ell_i = 1,\ldots,N} |M|^2 \leq (n-1)h^d N^d |M|^2 = (n-1)L^d |M|^2.
\]
Let $g = \phi$ and $f = L_\Phi^{-1} g$ in (63), the norm satisfies
\[ \lambda_{\min} \| D_h f \|_{L^2}^2 \leq \| D_h f, \Phi D_h f \|_{L^2} = - \langle f, d_h (\Phi D_h f) \rangle = - \langle f, i \phi \rangle \leq \| f \|_{L^2} \| \phi \|_{L^2} \leq C_P \| f \|_{L^2} \| \phi \|_{L^2}, \]
according to the discrete Poincaré inequality. Therefore, we get
\[ \| D_h f \|_{L^2} \leq \frac{C_P}{\lambda_{\min}} \| \phi \|_{L^2}. \]

Using an inverse inequality in (64) leads to
\[ \| f \|_{L^2} \leq C_1 h^{-\frac{1}{2}} \| D_h f \|_{L^2} \leq C_1 C_P h^{-\frac{1}{2}} L^2 M (n - 1)^{\frac{1}{2}} \leq \frac{CM}{\lambda_{\min}} h^{-\frac{1}{2}} (n - 1)^{\frac{1}{2}}. \]

Now we prove Theorem 4.

**Proof.** In a fashion similar to the proof of the one dimensional case, there exists a unique solution to the optimization problem (64) for any $\delta > 0$. This follows from the same argument with notations replaced by the multidimensional version. To prove that the minimizer of (64) does not touch the boundary of $K_\delta$, we use the equivalent optimization problem
\[ \min_{\hat{\rho} \in \hat{K}_\delta} \left\{ J = \frac{1}{2\Delta t} \| \hat{\rho} - \hat{\rho}^k \|_{L^2}^2 + F_h (\hat{\rho}) \right\}, \tag{65} \]
over the set
\[ \hat{K}_\delta = \left\{ \hat{\rho} : \hat{\rho} - \hat{\rho}^k \in (\tilde{C}_{\delta})^{n-1} : \rho_{i,\ell} \geq \delta, \sum_{i=1}^{n-1} \rho_{i,\ell} \leq 1 - \delta, \forall i = 1, \ldots, n - 1, \ \ell \in \{1, \ldots, N\}^d \right\}. \]

Assume the minimizer touches the boundary of $\hat{K}_\delta$ at the grid point $\ell^0 = (\ell_1^0, \ldots, \ell_d^0)$ for the $i_0$-th component, i.e.
\[ \rho^*_{i_0,\ell_1^0, \ldots, \ell_d^0} = \delta. \tag{66} \]

Next we consider the following two cases:

(a) \[ \sum_{i=1}^{n-1} \rho^*_{i,\ell^0} \geq \frac{1}{2}. \]

(b) \[ \sum_{i=1}^{n-1} \rho^*_{i,\ell^0} < \frac{1}{2}. \]

First consider the case [a]. We also suppose \( \rho^*_{i,\ell} \) achieves its maximum at the $i_1$-th component, and \( \rho^*_{i,\ell} \) achieves its maximum at $\ell = (\ell_1^1, \ldots, \ell_d^1)$. We calculate the directional derivative of the objective function (65) along the direction
\[ \nu_{i,\ell_1, \ldots, \ell_d} = \begin{cases} 1, & \text{for } i = i_0, \ \ell_s = \ell_s^0, \ \forall s = 1, \ldots, d, \\ -1, & \text{for } i = i_1, \ \ell_s = \ell_s^0, \ \forall s = 1, \ldots, d, \\ -1, & \text{for } i = i_0, \ \ell_s = \ell_s^1, \ \forall s = 1, \ldots, d, \\ 1, & \text{for } i = i_1, \ \ell_s = \ell_s^1, \ \forall s = 1, \ldots, d, \\ 0, & \text{otherwise}, \end{cases} \]
and we get
\[
\frac{1}{h^d} \frac{d}{ds} J(\tilde{\rho}^* + sv) \bigg|_{s=0} = \frac{1}{\Delta t} (L^{-1}_D(\tilde{\rho}^* - \tilde{k}))_{i_0, \ell_0} - \frac{1}{\Delta t} (L^{-1}_D(\tilde{\rho}^* - \tilde{k}))_{i_1, \ell_1} - \frac{1}{\Delta t} (L^{-1}_D(\tilde{\rho}^* - \tilde{k}))_{i_0, \ell_1} + \frac{1}{\Delta t} (L^{-1}_D(\tilde{\rho}^* - \tilde{k}))_{i_1, \ell_0} + \log \rho^*_{i_0, \ell_0} - \log \rho^*_{i_1, \ell_1} - \log \rho^*_{i_0, \ell_1} + \log \rho^*_{i_1, \ell_0}. \tag{67}
\]

Since \(\rho^*_{i_1, \ell_0}\) is the maximum point and the assumption \(\rho^*_{i_1, \ell_0} \geq \frac{1}{2}\),
\[
\rho^*_{i_1, \ell_0} \geq \frac{1}{2(n-1)}.	ag{68}
\]

Since \(\rho^*_{i_0, \ell_1}\) is the maximum point and
\[
\sum_{\ell \in \{1, \ldots, d\}^N} \rho^*_{i_0, \ell} = \sum_{\ell \in \{1, \ldots, d\}^N} \rho^*_{i_1, \ell},
\]
we have
\[
\rho^*_{i_0, \ell_1} \geq \frac{m}{h^d N^d}, \tag{69}
\]
where \(m\) is set to be
\[
m = \min_{\{i = 1, \ldots, n-1\}} \left\{ h^d \sum_{\ell \in \{1, \ldots, N\}^d} \rho^*_{i, \ell} \right\}.
\]

In order to guarantee \(\tilde{\rho}^* + sv \in \tilde{K}_\delta\), we assume
\[
\delta \leq \frac{m}{2h^d N^d}
\]
so that \(\rho^*_{i_0, \ell_1} - s \geq \frac{G}{m N^d} - s \geq \delta\) for small \(s\). One can check for other components and get \(\tilde{\rho}^* + sv \in \tilde{K}_\delta\) for \(\delta \leq \frac{1}{4(n-1)}\). We also have
\[
\rho^*_{i_1, \ell_1} \leq 1 - \delta < 1.
\]

Taking the above inequality and \(\textbf{(60), (68)-(69)}\) into \(\textbf{(67)}\) and applying Lemma \(\textbf{3}\) leads to
\[
\frac{1}{h^d} \frac{d}{ds} J(\tilde{\rho}^* + sv) \bigg|_{s=0} \leq \frac{8C}{\lambda^k_{\text{min}, \Delta t}} h^{-\frac{1}{2}} (n-1)^{\frac{1}{2}} + \log \rho^*_{i_0, \ell_0} - \log (\rho^*_{i_1, \ell_1} - \rho^*_{i_0, \ell_1}) + \log (\rho^*_{i_1, \ell_0} + \rho^*_{i_0, \ell_1}).
\]

where \(\lambda^k_{\text{min}}\) is the minimum eigenvalue of \(\hat{D}^k\). Taking
\[
\delta_0 \leq \min \left\{ \frac{m}{4(n-1)h^d N^d}, \frac{m}{2h^d N^d} \right\} \frac{\log \frac{m}{2h^d N^d} + 1}{4(n-1)} \tag{70}
\]
leads to
\[
\frac{1}{h^d} \frac{d}{ds} J(\tilde{\rho}^* + sv) \bigg|_{s=0} \leq - \log 2 < 0,
\]
which contradicts to the assumption that \(\tilde{\rho}^*\) is a minimizer.

Next we consider the case \(\textbf{(B)}\). We also suppose \(\{\rho^*_{i_0, \ell}\}_{\ell \in \{1, \ldots, N\}^d}\) achieves its maximum at \(\ell = \ell^1 = (\ell^1_1, \ldots, \ell^1_d)\). We take
\[
\nu_{i_0, \ell_1, \ldots, \ell_d} = \begin{cases} 
1, & \text{for } i = i_0, \quad \ell_s = \ell^1_s, \quad \forall s = 1, \ldots, d, \\
-1, & \text{for } i = i_1, \quad \ell_s = \ell^0_s, \quad \forall s = 1, \ldots, d, \\
0, & \text{otherwise},
\end{cases}
\]
and use (66), (1), (69) to get
\[
\frac{1}{h^d} \left. \frac{d}{ds} \left( J(\tilde{\rho}^* + s\nu) \right) \right|_{s=0} = \frac{1}{\Delta t} (L_{Dk}^{-1}(\tilde{\rho}^* - \tilde{\rho}^k))_{i_0,\ell_0} - \frac{1}{\Delta t} (L_{Dk}^{-1}(\tilde{\rho}^* - \tilde{\rho}^k))_{i_0,\ell_1} + \log \rho_{i_0,\ell_0}^* - \log \rho_{i_0,\ell_1}^* + \log \left( 1 - \sum_{i=1}^{n-1} \rho_{i_0,\ell_1}^* \right)
\]
\[
\leq \frac{4C}{\lambda_{\min}^k} h^{-\frac{d}{2}} (n-1)^{\frac{d}{2}} + \log \delta - \log \left( \frac{1}{2} \right) - \log \frac{m}{h^dN^d} + \log 1,
\]
Taking
\[
\delta_0 \leq \min \left\{ \frac{m}{4h^dN^d} e^{-\frac{4C}{\lambda_{\min}^k} h^{-\frac{d}{2}} (n-1)^{\frac{d}{2}}} : \frac{m}{2h^dN^d} \right\}
\]
leads to
\[
\left. \frac{1}{h^d} \frac{d}{ds} \left( J(\tilde{\rho}^* + s\nu) \right) \right|_{s=0} = -\log 2 < 0,
\]
which contradicts to the assumption that \( \tilde{\rho}^* \) is a minimizer, and so the situation (b) cannot occur.

On the other hand, we suppose \( \tilde{\rho}^* \) touches the other boundary with
\[
\sum_{i=1}^{n-1} \rho_{i,e}^* = 1 - \delta.
\]
Suppose \( \rho_{i,\ell}^* \) achieves its maximum at \( i_0, \ell_0 \), then
\[
\rho_{i_0,\ell_0}^* \geq \frac{1 - \delta}{n - 1} \geq \frac{1}{2(n-1)},
\]
for \( \delta \leq \frac{1}{2} \).

Since \( \rho^* - \rho^k \in (\mathcal{C}^d)_{\text{per}}^{n-1} \), we have
\[
\sum_{\ell \in \{1,\ldots,N\}^d} \sum_{i=1}^{n-1} \rho_{i,\ell}^* \leq \sum_{\ell \in \{1,\ldots,N\}^d} \sum_{i=1}^{n-1} \rho_{i,\ell}^k \leq N^d(1 - \rho_{\min}^k)
\]
with
\[
\rho_{\min}^k = \min_{\ell \in \{1,\ldots,N\}^d} \rho_{i,\ell}^k.
\]
Suppose \( \sum_{i=1}^{n-1} \rho_{i,\ell}^* \) achieves its minimum at \( \ell_1 \), then we have
\[
\sum_{i=1}^{n-1} \rho_{i,\ell_1}^* \leq \frac{1}{N^d - 1} (N^d(1 - \rho_{\min}^k) - (1 - \delta))
\]
\[
\leq 1 - \frac{N^d \rho_{\min}^k - \delta}{N^d - 1} \leq 1 - \frac{2N^d - 1}{2(N^d - 1)} \rho_{\min}^k.
\]
if \( \delta \leq \frac{1}{2} \rho_{\min}^k \).

We take
\[
\nu_{i,\ell} = \begin{cases} 
-1, & \text{for } i = i_0, \ell = \ell_0, \\
1, & \text{for } i = i_0, \ell = \ell_1, \\
0, & \text{otherwise},
\end{cases}
\]
and use the above inequality together with (72), (73) to obtain
\[ \frac{1}{h^2} \frac{d}{ds} J(\hat{\rho}^* + sv) \bigg|_{s=0} \]
\[ = -\frac{1}{\Delta t} (C_{Dk}^{-1}(\hat{\rho}^* - \hat{\rho}^k))_{i_0, \ell_0} - \log \rho_{i_0, \ell_0} + \log \left( 1 - \sum_{i=1}^{n-1} \rho_{i, \ell_0} \right) \]
\[ + \frac{1}{\Delta t} (C_{Dk}^{-1}(\hat{\rho}^* - \hat{\rho}^k))_{i_0, \ell_1} + \log \rho_{i_0, \ell_1} - \log \left( 1 - \sum_{i=1}^{n-1} \rho_{i, \ell_1} \right) \]
\[ \leq \frac{4C}{\lambda_{\text{min}}^2} \frac{h}{\Delta t} + \frac{\rho_{i_0, \ell_0}^k}{n(n-1)^{1/2}} - \log \left( 1 + \frac{1}{h^2} \right) + \log \delta + \log 1 - \frac{2N^d - 1}{2(N^d - 1)} \rho_{i, \ell_0}^k. \]

Taking
\[ \delta_0 \leq \min \left\{ \frac{(2N^d - 1)\rho_{i, \ell_0}^k}{8(N^d - 1)(n-1)^{1/2}} e^{-\frac{4C}{\lambda_{\text{min}}^2} \frac{h}{\Delta t}}, \frac{1}{\rho_{i, \ell_0}^k} \frac{1}{8(N^d - 1)} \right\} \]
leads to
\[ \frac{1}{h^2} \frac{d}{ds} J(\hat{\rho}^* + sv) \bigg|_{s=0} \leq -\log 2 < 0, \]
which contradicts to the assumption that \( \hat{\rho}^* \) is a minimizer.

We conclude that there exists a \( \delta_0 \), which can be chosen to be the smaller value of (70), (71) and (74) that only depends on \( h, \Delta t, \rho^k \) and the domain, such that the minimizer of (64) cannot touch the boundary.

To prove the equivalence of the numerical scheme with the minimizer of the optimization problem (64), we follow Step 3 of the proof of Theorem 1 for the one dimensional case. We omit the details here.

Theorem 3 is then proved in a fashion similar to the proof of Theorem 2.

**Appendix B. Proof of consistency**

Here we present detailed calculations of the truncation error defined by
\[ \tau_i^1 = \frac{P_{i+1}^k - P_i^k}{\Delta t} + d_h(\hat{P}_i^k V_{i+1}^k), \]
\[ \tau_i^2 = D_h \log P_i^k + 1 \sum_{j=1}^{n} b_{ij} \hat{P}_j^k (V_{i+1}^k - V_j^k), \]
\[ \tau_i^3 = \frac{n}{i} \hat{P}_i^k V_{i+1}^k. \]

We first calculate \( \tau_i^1 \).
\[ \tau_i^1 = \frac{P_{i+1}^k - P_i^k}{\Delta t} + d_h(\hat{P}_i^k V_{i+1}^k) \]
\[ = \frac{P_{i+1}^k - P_i^k}{\Delta t} + \frac{1}{h} \left( \hat{P}_i^k \frac{V_{i+1}^k - V_{i+1}^k}{\Delta t}, \hat{P}_i^k \frac{V_{i+1}^k - V_{i+1}^k}{\Delta t} \right) \]
\[ = \frac{P_{i+1}^k - P_i^k}{\Delta t} + \frac{1}{h} \left( \frac{V_{i+1}^k - V_i^k}{\Delta t}, \frac{V_{i+1}^k - V_i^k}{\Delta t} \right). \]

The terms in the above equation can be calculated using Taylor’s expansion as
\[ P_{i+1}^k = P_i^k + \partial_i P_i^k \Delta t + O(\Delta t^2), \]
\[ P_i^k = P_i^k \pm \frac{h}{2} \partial_i P_i^k + \frac{h^2}{2} \partial_i^2 P_i^k + O(h^3), \]
\[ V_{i+1}^k = V_i^k \pm \frac{h}{2} \partial_i V_i^k + \Delta t \partial_i V_i^k + \frac{h^2}{2} \partial_i^2 V_i^k + \frac{1}{2} \Delta t^2 V_i^k + \frac{1}{2} \Delta t^2 V_i^k. \]
The terms \(\tau^1\) into the previous equation leads to

\[
\tau_{i,\ell}^1 \approx \partial_t P_{i,\ell}^k - \frac{1}{2h} \left( 2h \partial_x P_{i,\ell}^k \left( V_{i,\ell}^k + \Delta t \partial_t V_{i,\ell}^k + \frac{1}{4} h^2 \partial_{xx} V_{i,\ell}^k + \frac{1}{2} \Delta t^2 V_{i,\ell}^k \right) \right) \\
- \frac{1}{2h} \left( 2P_{i,\ell}^k + \frac{1}{2} h^2 \partial_{xx} P_{i,\ell}^k \right) \left( h \partial_x V_{i,\ell}^k + h \Delta t \partial_{xt} V_{i,\ell}^k \right) \\
+ O(\Delta t + h^2 + \Delta th + \Delta t^2 + \Delta t^3) \\
= (\partial_t P - \partial_x (PV))_{i,\ell}^k + O(\Delta t + h^2 + \Delta t^2 + \Delta th + \Delta t^3). 
\]

The terms \(\tau^2\) and \(\tau^3\) can be also approximated again using the Taylor expansion. The results are

\[
\tau_{i,\ell+\frac{1}{2}}^2 \approx \frac{\partial_x P_{i,\ell}^k}{P_{i,\ell}^k} + \sum_{j=1}^{n} b_{ij} P_{j,\ell}^k (V_{j,\ell}^k - V_{j,\ell}^k) + \frac{h}{2} \left[ \frac{\partial_{xx} P_{i,\ell}^k}{P_{i,\ell}^k} - \frac{(\partial_x P_{i,\ell}^k)^2}{P_{i,\ell}^k} \right] \\
- \sum_{i=1}^{n} \left( \frac{(\partial_x P_{i,\ell}^k)^2}{P_{i,\ell}^k} + \partial_{xx} P_{i,\ell}^k - \frac{(\partial_x P_{i,\ell}^k)^2}{P_{i,\ell}^k} \right) \\
+ \sum_{j=1}^{n} b_{ij} (\partial_x P_{j,\ell}^k (V_{j,\ell}^k - V_{j,\ell}^k) + P_{j,\ell}^k (\partial_x V_{j,\ell}^k - \partial_x V_{j,\ell}^k)) + O(\Delta t + h^2) \\
= 0 + \frac{h}{2} \partial_x \left( \frac{\partial_x P_{i,\ell}^k}{P_{i,\ell}^k} - \sum_{j=1}^{n} b_{ij} P_{j,\ell}^k (V_{j,\ell}^k - V_{j,\ell}^k) \right) + O(\Delta t + h^2) \\
= O(\Delta t + h^2).
\]

\[
\tau_{i,\ell+\frac{1}{2}}^3 \approx \sum_{i=1}^{n} P_{i,\ell}^k V_{i,\ell}^k + \frac{1}{2} h \sum_{i=1}^{n} \partial_x (P_{i,\ell}^k V_{i,\ell}^k) + \Delta t \sum_{i=1}^{n} P_{i,\ell}^k \partial_t V_{i,\ell}^k + O(\Delta t^2 + h^2) \\
= O(\Delta t + h^2).
\]

In summary, we conclude the result stated in Lemma \[\text{Lemma 5}\] i.e., there exists \(C > 0\) depending on \((P, V)\) so that

\[
|\tau_{i,\ell}^1|, |\tau_{i,\ell+\frac{1}{2}}^2|, |\tau_{i,\ell+\frac{1}{2}}^3| \leq C(\Delta t + h^2).
\]