A PRIMAL-DUAL FORWARD-BACKWARD SPLITTING METHOD FOR CROSS-DIFFUSION GRADIENT FLOWS WITH GENERAL MOBILITY MATRICES*

YUNHONG DENG† AND CHAOZHEN WEI‡

Abstract. In this work, we construct a primal-dual forward-backward (PDFB) splitting method for computing a class of cross-diffusion systems that can be formulated as gradient flows under transport distances induced by matrix mobilities. By leveraging their gradient flow structure, we use minimizing movements as the variational formulation and compute these cross-diffusion systems by solving the minimizing movements as optimization problems at the fully discrete level. Our strategy to solve the optimization problems is the PDFB splitting method outlined in our previous work [21]. The efficiency of the proposed PDFB splitting method is demonstrated on several challenging cross-diffusion equations from the literature.

Key words. Cross-diffusion gradient flows, structure-preserving methods, minimizing movement schemes, optimal transport

AMS subject classifications. 35A15, 47J25, 47J35, 49M29, 65K10, 76M30

1. Introduction. Cross-diffusion systems arise in a variety of applications involving multi-species or multi-component systems including, to name just a few, chemistry [33, 22], biology [55, 20], thin film dynamics [31, 38, 57], and multi-component flows [24, 54, 48]. In this work, we consider a general cross-diffusion system of the following form [56, 2],

$$(1.1) \qquad \partial_t \boldsymbol{\mu} = \nabla \cdot \left[M(\boldsymbol{\mu}) \nabla \frac{\delta \mathcal{E}[\boldsymbol{\mu}]}{\delta \boldsymbol{\mu}} \right] \quad \text{in } \Omega,$$

with a non-flux boundary condition,

$$\left[M(\boldsymbol{\mu}) \nabla \frac{\delta \mathcal{E}[\boldsymbol{\mu}]}{\delta \boldsymbol{\mu}} \right]_\alpha \cdot \mathbf{n} = 0 \text{ on } \partial \Omega, \quad \forall 1 \leq \alpha \leq n,$$

where $\boldsymbol{\mu} = [\mu_\alpha]_{\alpha=1}^n \in \mathbb{R}^{n \times 1}$ is a vector field representing the population densities or concentrations of n type of species, $\Omega \subset \mathbb{R}^d$ is a regular bounded domain, and $\mathbf{n}$ is the unit outer normal vector of Ω. In particular, we consider the concentration-dependent mobility for general cross-diffusion systems

$$M(\boldsymbol{\mu}) = [M_{\alpha\beta}(\boldsymbol{\mu})]_{\alpha,\beta=1}^n \in \mathbb{R}^{n \times n},$$

which is a (symmetric) positive-definite matrix field. By denoting $[\,\cdot\,]_\alpha$ the α-th row of the matrix, the gradient operator ∇ acting on a vector field $\boldsymbol{\mu} = [\mu_\alpha]_{\alpha=1}^n \in \mathbb{R}^{n \times 1}$, and the divergence operator $\nabla \cdot$ acting on a matrix field $\mathbf{u} = [u_{\alpha,\beta}]_{\alpha,\beta=1}^{n,d} \in \mathbb{R}^{n \times d}$, are

†School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA, and School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China (deng0335@umn.edu).

‡School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China (cwei4@uestc.edu.cn).

hereinafter defined row-wise operations, i.e.,

$$\nabla\boldsymbol{\mu}(x) = \begin{bmatrix} \dfrac{\partial\mu_1(x)}{\partial x_1} & \cdots & \dfrac{\partial\mu_1(x)}{\partial x_d} \\ \vdots & & \vdots \\ \dfrac{\partial\mu_n(x)}{\partial x_1} & \cdots & \dfrac{\partial\mu_n(x)}{\partial x_d} \end{bmatrix}, \quad \nabla\cdot\mathbf{u}(x) = \sum_{\beta=1}^{d} \begin{bmatrix} \dfrac{\partial u_{1,\beta}(x)}{\partial x_\beta} \\ \vdots \\ \dfrac{\partial u_{n,\beta}(x)}{\partial x_\beta} \end{bmatrix},$$

where $x = (x_1, \cdots, x_d)$. Consider the free energy functional $\mathcal{E}[\boldsymbol{\mu}]$ in a general form given by

$$(1.2) \qquad \mathcal{E}[\boldsymbol{\mu}] = \int_\Omega U(\boldsymbol{\mu}(x))\,\mathrm{d}x + \int_\Omega V(x)\cdot\boldsymbol{\mu}(x)\,\mathrm{d}x + \sum_{\alpha=1}^{n} \int_\Omega \frac{\varepsilon^2}{2}|\nabla\mu_\alpha(x)|^2\,\mathrm{d}x,$$

where the Fréchet derivative of $\mathcal{E}[\boldsymbol{\mu}]$ is given by

$$(1.3) \qquad \frac{\delta\mathcal{E}[\boldsymbol{\mu}]}{\delta\boldsymbol{\mu}} = \left[\frac{\delta\mathcal{E}[\boldsymbol{\mu}]}{\delta\mu_\alpha}\right]_{\alpha=1}^{n} = \left[\frac{\partial U(\boldsymbol{\mu}(x))}{\partial\mu_\alpha} + V_\alpha(x) - \varepsilon^2\Delta\mu_\alpha(x)\right]_{\alpha=1}^{n},$$

where $U : \mathrm{dom}(U) \subseteq \mathbb{R}^{n\times 1} \to \mathbb{R}$ represents the internal energy, and $V = [V_\alpha]_{\alpha=1}^{n} : \Omega \to \mathbb{R}^{n\times 1}$ represents the confinement potential. The last term in (1.2) represents the Dirichlet energy usually considered in phase separation problems, and when it is considered, we impose an additional Neumann boundary condition $\nabla\mu_\alpha \cdot \mathbf{n} = 0$ on $\partial\Omega$ for all α.

Efficient numerical methods for cross-diffusion systems have become a popular topic in modern computational physics. To reflect the fundamental properties of the physical system (1.1), it is important to construct numerical methods that preserve key structural properties, such as energy dissipation

$$\mathcal{E}[\boldsymbol{\mu}(t + \Delta t)] \le \mathcal{E}[\boldsymbol{\mu}(t)] \text{ for all } \Delta t > 0,$$

and bounded preservation

$$0 \preceq_{\mathcal{S}_+} M(\boldsymbol{\mu}(t)),$$

along the flow (1.1). For example, Bailo, Carrillo, and Hu [2] construct a finite volume method via combining carefully chosen numerical fluxes and techniques for hyperbolic conservation laws, which preserves the natural bound of the numerical solution, and guarantees the energy dissipation at the fully discrete level. Sun, Carrillo, and Shu [56] proposed a discontinuous Galerkin method which preserves the energy dissipation structure by choosing appropriate numerical fluxes, which also ensures the positivity of the numerical solution via a suitable scaling limiter.

Our approach to computing (1.1) is based on the gradient flow framework from optimal transport theory developed over the past few decades [39, 50, 15, 1, 58]. This framework allows us to interpret equations of the form (1.1) as gradient flows and to characterize their dynamics via the minimizing movement scheme [39]. For general cross-diffusion systems (1.1), the gradient flow structure was established by Mielke [47], Liero and Mielke [42], who formulated (1.1) as a gradient flow with respect to a transport metric induced by the matrix mobility. This transport metric was further studied in [43]. Here we summarize their construction in the following scheme, which generalizes several minimizing movement schemes [14, 16] originating from the seminal work [39] developed by Jordan, Kinderlehrer, and Otto.

DEFINITION 1.1 (Time-discrete scheme). Given $\tau > 0$ and an initial value $\boldsymbol{\mu}_0$, we define the time-discrete solution $\boldsymbol{\mu}_\tau$ by $\boldsymbol{\mu}_\tau(t) = \boldsymbol{\mu}^k$ when $(k-1)\tau < t \leq k\tau$, where $\boldsymbol{\mu}^0 = \boldsymbol{\mu}_0$ and $(\boldsymbol{\mu}^k)$ is defined iteratively via a sequence of optimization problems:

$$\text{(1.4)} \qquad \boldsymbol{\mu}^{k+1} = \arg\min_{\boldsymbol{\mu}} \tau \mathcal{E}[\boldsymbol{\mu}] + \frac{1}{2} D(\boldsymbol{\mu}, \boldsymbol{\mu}^k)^2,$$

where $D(\boldsymbol{\mu}, \boldsymbol{\nu})$ is a transport distance defined as follows,

$$\text{(1.5)} \qquad D(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\boldsymbol{\omega}, \mathbf{u}} \left\{ \int_0^1 \int_\Omega \mathbf{u} : M(\boldsymbol{\omega})\mathbf{u} \, \mathrm{d}x\mathrm{d}t : \boldsymbol{\omega}(0) = \boldsymbol{\mu} \text{ and } \boldsymbol{\omega}(1) = \boldsymbol{\nu} \right.$$

$$\left. \partial_t \boldsymbol{\omega} + \nabla \cdot \left[M(\boldsymbol{\omega})\mathbf{u} \right] = 0 \text{ in } \Omega \times (0,1) \right\}^{\frac{1}{2}}.$$

where $\mathbf{u} : \mathbf{v} = \sum_{\alpha,\beta=1}^{n,d} u_{\alpha,\beta} v_{\alpha,\beta}$ is the Frobenius inner product between two matrices $\mathbf{u} = [u_{\alpha,\beta}]_{\alpha,\beta=1}^{n,d}$ and $\mathbf{v} = [v_{\alpha,\beta}]_{\alpha,\beta=1}^{n,d}$.

We will also provide a derivation of this minimizing movement scheme in Section 2. Substantial progress has been made in the rigorous analysis of the scheme (1.4). For $n = 1$, the existence of a unique solution $\boldsymbol{\mu}_\tau$ and its convergence as $\tau \to 0$ to the corresponding gradient flow have been established for Wasserstein gradient flows [39, 1, 7] and for general concave mobilities [44]. For coupled Wasserstein gradient flows where, $M(\boldsymbol{\mu}) = \mathrm{diag}(\boldsymbol{\mu})$, the minimizing movement scheme has been rigorously analyzed [40, 23]. For general cross-diffusion gradient flows, Zinsl and Matthes [61] proved the existence of a unique solution $\boldsymbol{\mu}_\tau$ and its convergence to (1.1) in the one-dimensional case ($d = 1$).

In this work, we compute (1.1) by solving minimizing movement scheme (1.4) at a discrete level as an optimization problem. There are several advantages to construct numerical solvers based on (1.4). For example, it guarantees energy dissipation and bounded preservation at the time discrete level (see Theorem 3.5 below). Several progresses in numerical methods have been made in this topic in recent years, for example, Wasserstein gradient flows [6, 14], Wasserstein-like gradient flows with nonlinear mobilities [16, 21], and reaction-diffusion systems with diagonal matrix mobilities [28, 27].

This general form (1.4) where $M(\boldsymbol{\mu})$ is non-diagonal can be non-trivial to compute. Current computational methods for cross-diffusion systems based on this framework are mainly restricted to the case where $M(\boldsymbol{\mu})$ is diagonal [12, 13, 28, 27]. One technical difficulty in the general form (1.4) is the fact that the mobility of each species are coupled. Thanks to our previous work [21], we can overcome this technical difficulty by using the primal-dual forward-backward (PDFB) splitting method [21], where one can decouple the mobility matrix through auxiliary dual variables.

The rest of the paper is organized as follows. In Section 2, we review the gradient flow structure of the cross-diffusion system (1.1) and derive the corresponding minimizing movement scheme. In Section 3, we construct a staggered grid approximation of the variational problem (1.4), and derive a saddle point form that can be solved by operator splitting methods. In Section 4, we discuss the PDFB splitting method [21] to solve the saddle point form via optimization solvers. In Section 5, we present the numerical results of the PDFB splitting method on challenging numerical examples from the literature. Section 7 is our conclusion and future directions.

2. Gradient flow structure. In this section, we define the gradient flow structure of (1.1) via an energy dissipation inequality (EDI) form of gradient flows (see, for example, [26, 25, 29]), and obtain the minimizing movement scheme via AGS theory of gradient flows [1].

2.1. Energy dissipation inequality. Equation (1.1) defines a dissipative system where the free energy $\mathcal{E}[\boldsymbol{\mu}]$ acts as a Lyapunov functional. Moreover, let $t \mapsto \boldsymbol{\mu}(t)$ be a solution of (1.1), we have the following form of energy dissipation:

$$(2.1) \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}[\boldsymbol{\mu}(t)] = -A\left(\boldsymbol{\mu}(t), -\frac{\delta\mathcal{E}[\boldsymbol{\mu}(t)]}{\delta\boldsymbol{\mu}}\right) \leq 0,$$

where A is the entropy dissipation functional (for $\boldsymbol{\varphi} : \Omega \to \mathbb{R}^{n \times 1}$) given by

$$A(\boldsymbol{\mu}, \boldsymbol{\varphi}) = \int_\Omega \nabla\boldsymbol{\varphi} : M(\boldsymbol{\mu})\nabla\boldsymbol{\varphi} \, \mathrm{d}x.$$

The form of dissipation in (2.1) gives us the following EDI form of the cross diffusion gradient flow (1.1).

LEMMA 2.1 (EDI form of gradient flow). *Let $t \mapsto \boldsymbol{\mu}(t)$ be the solution of (1.1) with initial value $\boldsymbol{\mu}_0$, it can be equivalently characterized by*

$$(2.2) \qquad \mathcal{E}[\boldsymbol{\mu}(T)] + \int_0^T \frac{1}{2}A^*\big(\boldsymbol{\mu}(t), \partial_t\boldsymbol{\mu}(t)\big) + \frac{1}{2}A\left(\boldsymbol{\mu}(t), -\frac{\delta\mathcal{E}[\boldsymbol{\mu}(t)]}{\delta\boldsymbol{\mu}}\right)\mathrm{d}t \leq \mathcal{E}[\boldsymbol{\mu}_0],$$

where A^* is the Legendre transform of A in the second variable, given by

$$A^*\big(\boldsymbol{\mu}, \partial_t\boldsymbol{\mu}\big) = \int_\Omega \nabla\boldsymbol{\varphi}^* : M(\boldsymbol{\mu})\nabla\boldsymbol{\varphi}^* \, \mathrm{d}x,$$

where $\boldsymbol{\varphi}^*$ is the solution to the following Poisson problem:

$$(2.3) \qquad \partial_t\boldsymbol{\mu} - \nabla \cdot \big[M(\boldsymbol{\mu})\nabla\boldsymbol{\varphi}^*\big] = 0 \text{ in } \Omega, \quad \nabla\boldsymbol{\varphi}^* \cdot \mathbf{n} = 0 \text{ on } \partial\Omega.$$

Proof of Lemma 2.1. The proof relies on the following Young's inequality coming from the definition of Legendre transform:

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}[\boldsymbol{\mu}] \, \mathrm{d}t = \left\langle \partial_t\boldsymbol{\mu}, -\frac{\delta\mathcal{E}[\boldsymbol{\mu}]}{\delta\boldsymbol{\mu}} \right\rangle \leq \frac{1}{2}A^*\big(\boldsymbol{\mu}, \partial_t\boldsymbol{\mu}\big) + \frac{1}{2}A\left(\boldsymbol{\mu}, -\frac{\delta\mathcal{E}[\boldsymbol{\mu}]}{\delta\boldsymbol{\mu}}\right),$$

where the equality holds if and only if $\boldsymbol{\mu}$ satisfy (1.1). Therefore, one can obtain an equivalent characterization of (1.1) by integrating the above inequality over $(0, T)$, which gives us (2.2). $\square$

2.2. Transport metric. In this part, we construct the transport metric (1.5) based on the form of energy dissipation in (2.2). In the context of the metric space with a distance $D_*(\boldsymbol{\mu}, \boldsymbol{\nu})$, one has the following energy inequality characterization of curves of maximal slope, i.e., gradient flows (see [1, Remark 1.3.3]):

$$(2.4) \qquad \mathcal{E}[\boldsymbol{\mu}(T)] + \int_0^T \frac{1}{2}|\dot{\boldsymbol{\mu}}(t)|^2 + \frac{1}{2}\big|\partial\mathcal{E}[\boldsymbol{\mu}(t)]\big|^2 \, \mathrm{d}t \leq \mathcal{E}[\boldsymbol{\mu}_0],$$

where $|\dot{\boldsymbol{\mu}}(t)|$ and $\big|\partial\mathcal{E}[\boldsymbol{\mu}(t)]\big|$ are metric derivative and metric slope, respectively, defined by

$$|\dot{\boldsymbol{\mu}}(t)| = \lim_{\varepsilon\to 0} \frac{D_*(\boldsymbol{\mu}(t+\varepsilon), \boldsymbol{\mu}(t))}{\varepsilon}, \quad \big|\partial\mathcal{E}[\boldsymbol{\mu}]\big| = \limsup_{\boldsymbol{\nu}\to\boldsymbol{\mu}} \frac{(\mathcal{E}[\boldsymbol{\mu}] - \mathcal{E}[\boldsymbol{\nu}])_+}{D_*(\boldsymbol{\mu}, \boldsymbol{\nu})},$$

where $(\alpha)_+ = \max\{0, \alpha\}$. For discussion on these concepts, please refers to [1, Chapter 1].

Now, by comparing (2.4) to the EDI form (2.2), we can see that A and A^* in (2.2) correspond to the form of metric slope and metric derivative, respectively. Hence, one can expect that

$$A^*\big(\boldsymbol{\mu}(t), \partial_t \boldsymbol{\mu}(t)\big)^{\frac{1}{2}} = \lim_{\varepsilon \to 0} \frac{D_*(\boldsymbol{\mu}(t + \varepsilon), \boldsymbol{\mu}(t))}{\varepsilon}.$$

By integrating the metric derivative A^* over curves $\boldsymbol{\omega}$, we can formulate the transport metric by using the following least action principal:

$$D_*(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\boldsymbol{\omega}} \left\{ \int_0^1 A^*\big(\boldsymbol{\omega}(t), \partial_t \boldsymbol{\omega}(t)\big) \, \mathrm{d}t, \quad \boldsymbol{\omega}(0) = \boldsymbol{\mu}, \ \boldsymbol{\omega}(1) = \boldsymbol{\nu} \right\}^{\frac{1}{2}}.$$

Let us prove that $D_*(\boldsymbol{\mu}, \boldsymbol{\nu})$ is exactly the transport metric in (1.5).

PROPOSITION 2.2. *Given $\boldsymbol{\mu}, \boldsymbol{\nu}$, we have $D(\boldsymbol{\mu}, \boldsymbol{\nu}) = D_*(\boldsymbol{\mu}, \boldsymbol{\nu})$.*

Proof. The following proof is based on Otto and Westdickenberg [51]. By writing down the form of metric derivative, we have

$$(2.5) \qquad \int_0^1 A^*\big(\boldsymbol{\omega}(t), \partial_t \boldsymbol{\omega}(t)\big) \, \mathrm{d}t = \int_0^1 \int_\Omega \nabla \boldsymbol{\varphi} : M(\boldsymbol{\omega}) \nabla \boldsymbol{\varphi} \, \mathrm{d}x \mathrm{d}t$$

$$\text{where } \partial_t \boldsymbol{\omega} - \nabla \cdot [M(\boldsymbol{\omega}) \nabla \boldsymbol{\varphi}] = 0 \text{ in } \Omega \times (0, 1).$$

i) Because the minimal action $D_*(\boldsymbol{\mu}, \boldsymbol{\nu})$ is the special case of the Benamou-Brenier dynamical formulation $D(\boldsymbol{\mu}, \boldsymbol{\nu})$ when $\mathbf{u} = -\nabla \boldsymbol{\varphi}$, we can show that the minimal action provide a upper bound for the Benamou-Brenier dynamical formulation, i.e., $D(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq D_*(\boldsymbol{\mu}, \boldsymbol{\nu})$.

ii) The converse can be shown as follows. For any given $\boldsymbol{\omega}$ and any admissible pair $(\boldsymbol{\omega}, \mathbf{u})$ that satisfies the continuity equation constraint, we can always find a potential $\boldsymbol{\varphi}$ that solves the following elliptic equation for any time $t \in (0, 1)$:

$$-\nabla \cdot \big[M(\boldsymbol{\omega}(t)) \nabla \boldsymbol{\varphi}(t)\big] = \nabla \cdot \big[M(\boldsymbol{\omega}(t)) \mathbf{u}(t)\big] \text{ in } \Omega.$$

By using an integration by parts and the Neumann boundary condition (2.3) on $\boldsymbol{\varphi}$, we have

$$\int_0^1 \int_\Omega \big[M(\boldsymbol{\omega}) \mathbf{u} + M(\boldsymbol{\omega}) \nabla \boldsymbol{\varphi}\big] : \nabla \boldsymbol{\varphi} \, \mathrm{d}x \mathrm{d}t = 0.$$

By the Cauchy–Schwarz inequality (see, for example, [8, Lemma 3.3.1]), we have

$$\int_\Omega \nabla \boldsymbol{\varphi} : M(\boldsymbol{\omega}) \nabla \boldsymbol{\varphi} \, \mathrm{d}x = - \int_\Omega \nabla \boldsymbol{\varphi} : M(\boldsymbol{\omega}) \mathbf{u} \, \mathrm{d}x$$

$$\leq \left(\int_\Omega \mathbf{u} : M(\boldsymbol{\omega}) \mathbf{u} \, \mathrm{d}x \right)^{\frac{1}{2}} \cdot \left(\int_\Omega \nabla \boldsymbol{\varphi} : M(\boldsymbol{\omega}) \nabla \boldsymbol{\varphi} \, \mathrm{d}x \right)^{\frac{1}{2}},$$

which implies that

$$\int_0^1 \int_\Omega \nabla \boldsymbol{\varphi} : M(\boldsymbol{\omega}) \nabla \boldsymbol{\varphi} \, \mathrm{d}x \mathrm{d}t \leq \int_0^1 \int_\Omega \mathbf{u} : M(\boldsymbol{\omega}) \mathbf{u} \, \mathrm{d}x \mathrm{d}t,$$

and therefore $D_*(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq D(\boldsymbol{\mu}, \boldsymbol{\nu})$. □

2.3. Minimizing movements. By considering the EDI form (2.2) and the construction of distances (Theorem 2.2), we can interpret (1.1) as the gradient flow with respect to the distance $D(\boldsymbol{\mu}, \boldsymbol{\nu})$. Hence, the minimizing movement scheme (1.4):

$$\boldsymbol{\mu}^{k+1} = \arg\min_{\boldsymbol{\mu}} \tau \mathcal{E}[\boldsymbol{\mu}] + \frac{1}{2} D(\boldsymbol{\mu}, \boldsymbol{\mu}^k)^2$$

is a consistent variational formulation to compute (1.1).

In the rest of this part, we introduce a change-of-variable technique in the transport distance [5, 21, 43] to obtain an equivalent convex optimization problem. Consider the so-called "momentum":

$$\mathbf{m} = M(\boldsymbol{\omega})\mathbf{u}.$$

By taking this expression of $\mathbf{m}$ into (1.5), we have the following reformulation of the distance.

LEMMA 2.3 ([43]). *Let $D(\boldsymbol{\mu}, \boldsymbol{\nu})$ be the distance defined in* (1.5), *we have*

$$\frac{1}{2} D(\boldsymbol{\mu}, \boldsymbol{\nu})^2 = \min_{\boldsymbol{\omega}, \mathbf{m}} \left\{ \int_0^1 \int_\Omega f\big(M(\boldsymbol{\omega}), \mathbf{m}\big) \, \mathrm{d}x \mathrm{d}t : \boldsymbol{\omega}(0) = \boldsymbol{\mu} \text{ and } \boldsymbol{\omega}(1) = \boldsymbol{\nu}, \right.$$

(2.6)
$$\left. \partial_t \boldsymbol{\omega} + \nabla \cdot \mathbf{m} = 0 \text{ in } \Omega \times (0,1) \right\}.$$

where $\mathbf{m} = [m_{\alpha,\beta}]_{\alpha,\beta=1}^{n,d}$ is minimized among matrix-valued fields with vanishing normal fluxes on the boundary such that $[m_{\alpha,\beta}]_{\beta=1}^d \cdot \mathbf{n} = 0$ on $\partial\Omega$ for all α, and f is an action function defined by

$$f(M, \mathbf{m}) = \begin{cases} \frac{1}{2}\mathbf{m} : M^\dagger \mathbf{m} & \text{if } \mathbf{m} = M\mathbf{u} \text{ for some } \mathbf{u} \\ & \text{and } 0 \preceq_{\mathcal{S}_+} M, \\ \infty & \text{otherwise.} \end{cases}$$

and $M^\dagger$ denotes the pseudo inverse of M.

Finally, by incorporating (2.6) into the minimizing movement (1.4), it suffices to solve the following convex optimization problem

(2.7)
$$\min_{\boldsymbol{\mu}, \boldsymbol{\omega}, \mathbf{m}} \tau \mathcal{E}[\boldsymbol{\mu}] + \int_0^1 \int_\Omega f\big(M(\boldsymbol{\omega}), \mathbf{m}\big) \, \mathrm{d}x \mathrm{d}t$$
$$\text{s.t. } \partial_t \boldsymbol{\omega} + \nabla \cdot \mathbf{m} = 0 \text{ in } \Omega \times (0,1), \quad \boldsymbol{\omega}(0) = \boldsymbol{\mu}^k, \ \boldsymbol{\omega}(1) = \boldsymbol{\mu} \text{ in } \Omega,$$

where $\mathbf{m} = [m_{\alpha,\beta}]_{\alpha,\beta=1}^{n,d}$ subjects to the vanishing normal fluxes on the boundary.

3. Conforming approximation on staggered grids. In this section, we construct the staggered grid approximation of scheme (2.7), and then discuss its structure-preserving properties. The discrete saddle point form, which is the final form of the numerical scheme, is summarized in the Section 3.4.

3.1. Single-step approximation. We first introduce a single-step approximation to reduce the computation cost which has been considered in [41, 16, 21]. Assume

that $\mathbf{m}$ is constant over inner-time $(0,1)$, we approximate $\boldsymbol{\mu}^{k+1}$ using the optimal $\boldsymbol{\mu}$ that minimizes,

(3.1)
$$\min_{\boldsymbol{\mu},\mathbf{m}} \tau\mathcal{E}[\boldsymbol{\mu}] + \int_\Omega f\left(M\left(\tfrac{1}{2}(\boldsymbol{\mu}+\boldsymbol{\mu}^k)\right),\mathbf{m}\right)\mathrm{d}x$$
$$\text{s.t. } \boldsymbol{\mu}-\boldsymbol{\mu}^k+\nabla\cdot\mathbf{m}=0 \text{ in } \Omega, \quad \boldsymbol{b}_0 \leq \mathcal{M}\boldsymbol{\mu} \leq \boldsymbol{b}_1 \text{ in } \Omega,$$

where $\mathcal{M}$ is a matrix of the size $\mathbb{R}^{p\times n}$ for given p, and $\boldsymbol{b}_0 = [b_{0,\beta}]_{\beta=1}^p, \boldsymbol{b}_1 = [b_{1,\beta}]_{\beta=1}^p$ are global bounds, which are set by specific form of the model to keep $M(\boldsymbol{\mu})$ positive definite.

Remark 3.1. This box constraint is necessary to ensure bounded preservation at the discrete level (see [21, Equation (11)]).

Remark 3.2. It has been shown that single-step approximations in the dynamical transport distance $D(\boldsymbol{\mu},\boldsymbol{\nu})$ does not influence the resolution order of JKO schemes [41, 16, 28].

Remark 3.3. Compared to the scalar case in [21], the bound of the solution may be coupled for (1.1). For example, consider the following mobility matrix from [2] (where σ is a positive constant which is referred to as the saturation level):

$$M(\boldsymbol{\mu}) = \begin{bmatrix} \mu_1(\sigma-(\mu_1+\mu_2)) & 0 \\ 0 & \mu_2(\sigma-(\mu_1+\mu_2)) \end{bmatrix}$$

where the box constraint is given by $\mu_0 \geq 0$, $\mu_1 \geq 0$, $\mu_1+\mu_2 \leq \sigma$. Hence, in the numerical scheme (3.1) the corresponding matrix $\mathcal{M}$ and bounds are given by

$$\mathcal{M} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{b}_0 = \begin{bmatrix} 0 \\ 0 \\ -\infty \end{bmatrix}, \quad \boldsymbol{b}_1 = \begin{bmatrix} \infty \\ \infty \\ \sigma \end{bmatrix}.$$

3.2. Saddle point form. The main difficulty in solving the variational problem (3.1) is the non-smoothness of the action function f. To remedy this, we consider proximal gradient methods for non-smooth optimization problems (see, for example, [17, 53, 18, 45]). In this subsection, we derive a saddle point form of (3.1) which is critical in non-smooth optimization methods.

Closely following the discussion on the scalar action functions [5, 52, 21], we show the following theorem, which states that the action function f is a support function of a convex set, or the Legendre transform of f is the indicator function of a convex set.

THEOREM 3.4. *Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix, and $\mathbf{m} \in \mathbb{R}^{n\times d}$, one has*

$$f(M,\mathbf{m}) = \max_{(Q,\boldsymbol{q})\in K} M : Q + \mathbf{m} : \boldsymbol{q},$$

where K is a closed convex set defined by

(3.2)
$$K = \left\{(Q,\boldsymbol{q}) \in \mathbb{R}_{\mathrm{sym}}^{n\times n} \times \mathbb{R}^{n\times d} : Q + \frac{\boldsymbol{q}\boldsymbol{q}^T}{2} \preceq_{\mathcal{S}_+} 0\right\},$$

where $\mathbb{R}_{\mathrm{sym}}^{n\times n}$ is the space of symmetrical $n\times n$ matrices, and $\preceq_{\mathcal{S}_+}$ is the order relation with respect to the cone of positive semi-definite matrix $\mathcal{S}_+$.

We postpone the proof of this theorem to Section 6. By using the support function form in Theorem 3.4, we have

$$\int_\Omega f\Big(M\big(\tfrac{1}{2}(\boldsymbol{\mu}+\boldsymbol{\mu}^k)\big),\mathbf{m}\Big)\,\mathrm{d}x = \sup\Big\{\int_\Omega M\big(\tfrac{1}{2}(\boldsymbol{\mu}+\boldsymbol{\mu}^k)\big): Q+\mathbf{m}:\boldsymbol{q}\,\mathrm{d}x:$$

$$(Q(x),\boldsymbol{q}(x))\in K \quad \forall x\in\Omega\Big\}.$$

Hence, by using this support function characterization, we can write the variational problem (3.1) in the following saddle point form,

$$\min_{\boldsymbol{\mu},\mathbf{m}}\max_{Q,\boldsymbol{q}} \tau\mathcal{E}[\boldsymbol{\mu}] + \int_\Omega M\big(\tfrac{1}{2}(\boldsymbol{\mu}+\boldsymbol{\mu}^k)\big): Q+\mathbf{m}:\boldsymbol{q}\,\mathrm{d}x,$$

(3.3)

$$\text{s.t. } \boldsymbol{\mu}-\boldsymbol{\mu}^k+\nabla\cdot\mathbf{m}=0, \quad \boldsymbol{b}_0\le\mathcal{M}\boldsymbol{\mu}\le\boldsymbol{b}_1, \quad Q+\frac{\boldsymbol{q}\boldsymbol{q}^T}{2}\preceq_{\mathcal{S}_+} 0.$$

where $\mathbf{m}=[m_{\alpha,\beta}]_{\alpha,\beta=1}^{n,d}$ subjects to the vanishing normal fluxes on the boundary.

3.3. Staggered grid approximation. In this subsection, we construct a staggered grid approximation of (3.1) and proceed to derive the discrete min-max saddle point form corresponding to (3.3). This construction is a multi-species version of the staggered grid approximation in our previous work [21].

We cover the domain Ω with non-overlapping uniform boxes $\Omega_{\mathbf{i}}$ defined by

$$\Omega_{\mathbf{i}} = \Big\{x: \big|x-x_{\mathbf{i}}\big|_\infty \le \tfrac{1}{2}h\Big\},$$

where h is the grid spacing and $x_{\mathbf{i}}$ is the center of the box $\Omega_{\mathbf{i}}$. To write the scheme in a finite-difference stencil, we adopt a global indices to location of the cube $\Omega_{\mathbf{i}}$:

$$\mathfrak{T} = \Big\{\mathbf{i}=\big(i_1,...,i_d\big)\Big\}.$$

Also, the boundary of each box $\Omega_{\mathbf{i}}$ can be partitioned into $2d$-rectangles $\bigcup_{q=1}^d \Gamma_{\mathbf{i}\pm\frac{1}{2}e_q}^q$ of the dimension $d-1$, where $\Gamma_{\mathbf{i}\pm\frac{1}{2}e_q}^q$ corresponds to two bounding interfaces orthogonal to the standard basis vector e_q pointing in the q-th direction. We use the following indices to locate $\Gamma_{\mathbf{i}\pm\frac{1}{2}e_q}^q$:

$$\mathfrak{J} = (\mathfrak{J}^q)_{q=1}^d \text{ where } \mathfrak{J}^q = \Big\{\mathbf{i}=\big(i_1,...,i_q\pm\tfrac{1}{2},...i_d\big)\Big\}.$$

Now, we let $\mathfrak{T}(\mathbb{R})$ and $\mathfrak{J}(\mathbb{R})$ be the discrete scalar field and flux field such that

$$\mathfrak{T}(\mathbb{R}) = \Big\{\sigma:\mathfrak{T}\to\mathbb{R}\Big\}, \quad \mathfrak{J}(\mathbb{R}) = \Big\{\boldsymbol{\sigma}=(\sigma^1,...,\sigma^d): \quad \sigma^q:\mathfrak{J}^q\to\mathbb{R}, \quad \forall 1\le q\le d\Big\}.$$

where we use $\sigma_{\mathbf{i}}$ or $\sigma_{\mathbf{j}}^q$ to denote the value of the filed on the $\mathbf{i}$-th box or $\mathbf{j}$-th interfaces, respectively, for scalar filed $\sigma\in\mathfrak{T}(\mathbb{R})$ or vector field $\boldsymbol{\sigma}=(\sigma^1,...,\sigma^d)\in\mathfrak{J}(\mathbb{R})$. We introduce the space of discrete momentum with vanishing normal fluxes of $\mathbf{m}$ on the boundary:

$$\mathfrak{J}_0(\mathbb{R}) = \Big\{\boldsymbol{\sigma}=(\sigma^1,...,\sigma^d)\in\mathfrak{J}(\mathbb{R}): \quad \sigma_{\mathbf{j}}^q=0 \text{ when } \Gamma_{\mathbf{j}}^q\subseteq\partial\Omega \quad \forall 1\le q\le d\Big\},$$

which encodes the non-flux boundary condition into $\mathbf{m}$ in the optimization.

3.3.1. Construction. In this part, we construct the approximation of the variational problem (3.1) on the staggered grid $\mathfrak{T}(\mathbb{R})$ and $\mathfrak{J}(\mathbb{R})$ constructed above.

1. Constraint: We consider the following discrete admissible set in (3.1):

$$
H = \Big\{ (\boldsymbol{\mu}, \mathbf{m}) \text{ such that } (\mu_\alpha, \mathbf{m}_\alpha) \in \mathfrak{T}(\mathbb{R}) \times \mathfrak{J}_0(\mathbb{R}) :
$$

(3.4)
$$
\mu_\alpha - \mu_\alpha^k + \mathcal{A}\mathbf{m}_\alpha = 0, \quad \forall 1 \le \alpha \le n,
$$

$$
b_{0,\beta} \le \sum_{\alpha=1}^n \mathcal{M}_{\beta\alpha}\mu_{\alpha,\mathbf{i}} \le b_{1,\beta}, \quad \forall \mathbf{i} \in \mathfrak{T} \text{ and } 1 \le \beta \le p \Big\}.
$$

where $\mathcal{A}$ is the discrete divergence operator

$$
\mathcal{A} : \mathfrak{J}(\mathbb{R}) \to \mathfrak{T}(\mathbb{R}), \quad (\mathcal{A}\boldsymbol{\sigma})_{\mathbf{i}} = \frac{1}{h}\sum_{q=1}^d \Big(\sigma^q_{\mathbf{i}+\frac{1}{2}e_q} - \sigma^q_{\mathbf{i}-\frac{1}{2}e_q} \Big) \text{ for } \boldsymbol{\sigma} = (\sigma^1, ..., \sigma^d).
$$

2. Action functional: We consider a centered quadrature rule for the action functional

$$
\int_\Omega f\Big(M\big(\tfrac{1}{2}(\boldsymbol{\mu} + \boldsymbol{\mu}^k)\big), \mathbf{m} \Big) \mathrm{d}x \approx \sum_{\mathbf{i} \in \mathfrak{T}} f\Big(M\big(\tfrac{1}{2}(\boldsymbol{\mu}_{\mathbf{i}} + \boldsymbol{\mu}_{\mathbf{i}}^k)\big), (\mathcal{I}\mathbf{m})_{\mathbf{i}} \Big) h^d
$$

where $\mathcal{I}$ is the interpolation operator

$$
\mathcal{I} : \mathfrak{J}(\mathbb{R}) \to \mathfrak{T}(\mathbb{R})^d, \quad (\mathcal{I}\boldsymbol{\sigma})_{\mathbf{i}} = \left(\frac{\sigma^1_{\mathbf{i}+\frac{1}{2}e_1} + \sigma^1_{\mathbf{i}-\frac{1}{2}e_1}}{2}, \cdots, \frac{\sigma^d_{\mathbf{i}+\frac{1}{2}e_d} + \sigma^d_{\mathbf{i}-\frac{1}{2}e_d}}{2} \right).
$$

Here, by a sight abuse of notation, interpolation $\mathcal{I}\mathbf{m}$ is a row-wise (specie-wise) interpolation when operated on a discrete matrix field in the space $\mathfrak{J}(\mathbb{R})^n$.

3. Free energy functional: For the free energy functional (1.2), we consider a centered quadrature rule for the internal energy and the confinement potential, and a trapezoidal rule for the Dirichlet energy, which has been discussed in [16]:

$$
\mathcal{E}[\boldsymbol{\mu}] \approx \mathcal{E}_h[\boldsymbol{\mu}] = \sum_{\mathbf{i} \in \mathfrak{T}} \Big(U(\boldsymbol{\mu}_{\mathbf{i}}) + V(x_{\mathbf{i}}) \cdot \boldsymbol{\mu}_{\mathbf{i}} \Big) h^d + \sum_{q=1}^d \sum_{\mathbf{j} \in \mathfrak{J}^q} \frac{\varepsilon^2}{2} |\partial_{q,h}\boldsymbol{\mu}|_{\mathbf{j}}^2 h^d,
$$

where $\partial_{q,h}$ is the discrete partial derivative in the q-th direction defined by

$$
\big(\partial_{q,h}\boldsymbol{\mu} \big)_{\mathbf{j}} = \frac{\boldsymbol{\mu}_{\mathbf{j}+\frac{1}{2}e_q} - \boldsymbol{\mu}_{\mathbf{j}-\frac{1}{2}e_q}}{h}.
$$

For the discrete Dirichlet energy $\sum_{q=1}^d \sum_{\mathbf{j} \in \mathfrak{J}^q} \frac{\varepsilon^2}{2} |\partial_{q,h}\boldsymbol{\mu}|_{\mathbf{j}}^2$, its gradient corresponds to a centered approximation of the Laplacian $-\varepsilon^2 \Delta_h \boldsymbol{\mu}$ with a Neumann boundary condition (see [16]).

3.4. Summary. In summary, we solve the following optimization in each time step

(3.5)
$$
\min_{\boldsymbol{\mu},\mathbf{m}} \tau \mathcal{E}_h[\boldsymbol{\mu}] + \sum_{\mathbf{i} \in \mathfrak{T}} f\Big(M\big(\tfrac{1}{2}(\boldsymbol{\mu}_{\mathbf{i}} + \boldsymbol{\mu}_{\mathbf{i}}^k)\big), (\mathcal{I}\mathbf{m})_{\mathbf{i}} \Big) h^d + \iota_H(\boldsymbol{\mu}, \mathbf{m}),
$$

where $\iota_A(\cdot)$ is the indicator function of a set A such that

$$\iota_A(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in A \\ \infty & \text{otherwise.} \end{cases}$$

One can show that optimizing the discrete objective functional in (3.5) preserves the structure of (1.1) at a fully discrete level.

THEOREM 3.5 (Structure-preserving property). *Let* $\boldsymbol{\mu}^{k+1} = [\mu_\alpha^{k+1}]_{\alpha=1}^n$ *be the solution to* (3.5), *one has*
  1. *Energy dissipation:* $\mathcal{E}_h[\boldsymbol{\mu}^{k+1}] \leq \mathcal{E}_h[\boldsymbol{\mu}^k]$.
  2. *Mass conservation:* $\sum_{\mathbf{i}\in\mathfrak{T}} \mu_{\alpha,\mathbf{i}}^{k+1} = \sum_{\mathbf{i}\in\mathfrak{T}} \mu_{\alpha,\mathbf{i}}^k \quad \forall \alpha$.
  3. *Bound preservation:* $0 \preceq_{\mathcal{S}_+} M(\boldsymbol{\mu}_{\mathbf{i}}^{k+1}) \quad \forall \mathbf{i} \in \mathfrak{T}$.

The proof is identical to that of [21, Theorem 1] with minor modifications.

By applying the support function form (Theorem 3.4) to the action function on each grid, we have the following discrete saddle point form corresponding to (3.3), which is our numerical scheme.

SCHEME 1. Let τ be the time step size and h be the spatial grid size. Given $\boldsymbol{\mu}^k$, we approximate the solution $\boldsymbol{\mu}^{k+1}$ of (1.4), and therefore the solution $\boldsymbol{\mu}(t)$ of (1.1) at $t = (k+1)\tau$, by finding the optimal $\boldsymbol{\mu}$ that solves

(3.6)
$$\min_{\boldsymbol{\mu},\mathbf{m}} \max_{Q,\boldsymbol{q}} \tau\widehat{\mathcal{E}}_h[\boldsymbol{\mu}] + \sum_{\mathbf{i}\in\mathfrak{T}} M\big(\tfrac{1}{2}(\boldsymbol{\mu}_{\mathbf{i}} + \boldsymbol{\mu}_{\mathbf{i}}^k)\big) : Q_{\mathbf{i}} + (\mathcal{I}\mathbf{m})_{\mathbf{i}} : \boldsymbol{q}_{\mathbf{i}}$$
$$- \sum_{\mathbf{i}\in\mathfrak{T}} \iota_K\big(Q_{\mathbf{i}}, \boldsymbol{q}_{\mathbf{i}}\big) + \iota_H(\boldsymbol{\mu}, \mathbf{m}),$$

where $\widehat{\mathcal{E}}_h = \mathcal{E}_h/h^d$ denotes the normalized energy functional.

To improve the practical application of the numerical scheme (3.6), we introduce a nonlinear preconditioning proposed in [21]. Let us consider a convex splitting of the energy, and denote,

$$\widehat{\mathcal{E}}_h[\boldsymbol{\mu}] = \mathcal{U}[\mathcal{K}\boldsymbol{\mu}] + \big(\widehat{\mathcal{E}}_h[\boldsymbol{\mu}] - \mathcal{U}[\mathcal{K}\boldsymbol{\mu}]\big) = \mathcal{U}[\mathcal{K}\boldsymbol{\mu}] + \mathcal{V}[\boldsymbol{\mu}],$$

where $\mathcal{U}$ is the preconditioner which requires to be convex, and $\mathcal{K}$ is a linear operator. By applying the Legendre transform to $\mathcal{U}$, we have the following preconditioned numerical scheme.

SCHEME 2. Let τ be the time step size and h be the spatial grid size. Given $\boldsymbol{\mu}^k$, we approximate the solution $\boldsymbol{\mu}^{k+1}$ of (1.4), and therefore the solution $\boldsymbol{\mu}(t)$ of (1.1) at $t = (k+1)\tau$, by finding the optimal $\boldsymbol{\mu}$ that solves

(3.7)
$$\min_{\boldsymbol{\mu},\mathbf{m}} \max_{Q,\boldsymbol{q},\boldsymbol{\nu}} \tau\mathcal{V}[\boldsymbol{\mu}] + \sum_{\mathbf{i}\in\mathfrak{T}} M\big(\tfrac{1}{2}(\boldsymbol{\mu}_{\mathbf{i}} + \boldsymbol{\mu}_{\mathbf{i}}^k)\big) : Q_{\mathbf{i}} + (\mathcal{I}\mathbf{m})_{\mathbf{i}} : \boldsymbol{q}_{\mathbf{i}}$$
$$+ (\mathcal{K}\boldsymbol{\mu})_{\mathbf{i}} \cdot \boldsymbol{\nu}_{\mathbf{i}} - \tau\mathcal{U}^*[\boldsymbol{\nu}/\tau] - \sum_{\mathbf{i}\in\mathfrak{T}} \iota_K\big(Q_{\mathbf{i}}, \boldsymbol{q}_{\mathbf{i}}\big) + \iota_H(\boldsymbol{\mu}, \mathbf{m}),$$

where $\mathcal{U}^*$ is the convex conjugate of $\mathcal{U}$.

4. PDFB splitting optimization. In this section, we present our strategy to solve the saddle point problem (3.7). Here we denote the primal variable and the dual

variable as $\mathbf{x}$ and $\mathbf{y}$ respectively. With this notation, the optimization problem (3.7) can be then compactly written as follows,

$$(4.1) \qquad \min_{\mathbf{x}=(\boldsymbol{\mu},\mathbf{m})} \max_{\mathbf{y}=(Q,\boldsymbol{q},\boldsymbol{\nu})} \Phi(\mathbf{x},\mathbf{y}) - F(\mathbf{y}) + G(\mathbf{x}).$$

where

$$(4.2) \qquad \Phi(\mathbf{x},\mathbf{y}) = \tau\mathcal{V}[\boldsymbol{\mu}] + \sum_{\mathbf{i}\in\mathfrak{T}} M\big(\tfrac{1}{2}(\boldsymbol{\mu_i}+\boldsymbol{\mu_i^k})\big) : Q_{\mathbf{i}} + (\mathcal{I}\mathbf{m})_{\mathbf{i}} : \boldsymbol{q_i} + (\mathcal{K}\boldsymbol{\mu})_{\mathbf{i}} \cdot \boldsymbol{\nu_i}$$

$$G(\mathbf{x}) = \iota_H(\boldsymbol{\mu},\mathbf{m}) \text{ and } F(\mathbf{y}) = \tau\mathcal{U}^*[\boldsymbol{\nu}/\tau] + \sum_{\mathbf{i}\in\mathfrak{T}} \iota_K(Q_{\mathbf{i}},\boldsymbol{q_i}).$$

The optimization problem (4.1) belongs to the class of general convex-concave saddle point problems, where Φ is continuously differentiable while F and G are convex but non-smooth.

4.1. Algorithm. In recent years, there are several optimization algorithms (see, for example, [45, 37, 60]) proposed to cope with general convex-concave saddle point problems (4.1). One of the key ideas behind these methods is the forward-backward splitting approach, in which one combines the gradient descents/ascents of $\Phi(\mathbf{x},\mathbf{y})$ with the proximal gradient updates of $G(\mathbf{x})$ and $F(\mathbf{y})$ in each step of the iteration. Reflection techniques in operator splitting methods can also be considered to improve the optimization algorithms [17, 59, 45].

In the following, we apply the PDFB splitting method developed in our previous work [21, Section 3] to solve the saddle point problem (4.1).

We introduce the Jacobi matrix $\mathcal{J}(\mathbf{x})$ defined by

$$\mathcal{J}(\mathbf{x})\mathbf{w} = \lim_{\varepsilon\to 0} \frac{\nabla_{\mathbf{y}}\Phi(\mathbf{x}+\varepsilon\mathbf{w},\mathbf{y}) - \nabla_{\mathbf{y}}\Phi(\mathbf{x},\mathbf{y})}{\varepsilon}, \quad \forall\mathbf{w}\in\mathbb{R}^{\dim(\mathbf{x})}$$

At each iterate $(\mathbf{x}^\ell,\mathbf{y}^\ell)$, we update ($\gamma$ and $\overline{\gamma}$ are step size parameters of the algorithm)

$$\mathbf{y}^{\ell+1} = \text{prox}_{\gamma F}\big(\mathbf{y}^\ell + \gamma\mathcal{J}(\mathbf{x}^\ell)(\overline{\mathbf{x}}^\ell - \mathbf{x}^\ell) + \gamma\nabla_{\mathbf{y}}\Phi(\mathbf{x}^\ell,\mathbf{y}^\ell)\big)$$
$$\mathbf{x}^{\ell+1} = \text{prox}_{\overline{\gamma}G}\big(\mathbf{x}^\ell - \overline{\gamma}\nabla_{\mathbf{x}}\Phi(\mathbf{x}^\ell,\mathbf{y}^{\ell+1})\big)$$
$$\overline{\mathbf{x}}^{\ell+1} = 2\mathbf{x}^{\ell+1} - \mathbf{x}^\ell - \overline{\gamma}\big(\nabla_{\mathbf{x}}\Phi(\mathbf{x}^{\ell+1},\mathbf{y}^{\ell+1}) - \nabla_{\mathbf{x}}\Phi(\mathbf{x}^\ell,\mathbf{y}^{\ell+1})\big)$$

where prox_F dentoes the proximal operator of a function defined by:

$$\text{prox}_F(\mathbf{x}_0) = \arg\min_v F(\mathbf{x}) + \frac{1}{2}\|\mathbf{x}-\mathbf{x}_0\|^2.$$

For details concerning the motivation and analysis of the algorithm, please refers to [21].

4.2. Implementation details. In each step of the PDFB splitting method, one needs to the compute derivatives of $\Phi(\mathbf{x},\mathbf{y})$. For (4.2), the gradients are trivial to evaluate by chain rules. In the following, we mainly discuss how to compute the proximal operators $G(\mathbf{x})$ and $F(\mathbf{y})$. It is important to note that the dual variables $(Q,\boldsymbol{q})$ and $\boldsymbol{\nu}$ are separable when computing the proximal operator of $F(\mathbf{y})$. We will discuss them as two individual subproblems, where the proximal problem associated with $\boldsymbol{\nu}$ is referred to as the proximal problem of the nonlinear preconditioner.

4.2.1. Proximal operator of primal variable. The proximal problem associated with $\mathrm{prox}_{\gamma G}$ amounts to solve a box constrained quadratic optimization problem for given $(\boldsymbol{\mu}_0, \mathbf{m}_0) \in \big(\mathfrak{T}(\mathbb{R}) \times \mathfrak{J}_0(\mathbb{R}) \big)^n$:

$$\min_{\boldsymbol{\mu}, \mathbf{m}} \frac{1}{2} \|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|^2 + \frac{1}{2} \|\mathbf{m} - \mathbf{m}_0\|^2$$

$$\text{s.t. } \mu_\alpha - \mu_\alpha^k + \mathcal{A}\mathbf{m}_\alpha = 0, \quad \forall 1 \le \alpha \le n,$$

$$b_{0,\beta} \le \sum_{\alpha=1}^n \mathcal{M}_{\beta\alpha} \mu_{\alpha, \mathbf{i}} \le b_{1,\beta}, \quad \forall \mathbf{i} \in \mathfrak{T} \text{ and } 1 \le \beta \le p.$$

This problem can be solved efficiently by the primal-dual active set method [46, page 319], which is also considered in our previous work [21]. Let us consider the Karush-Kuhn-Tucker (KKT) optimality conditions:

(4.3)
$$\begin{cases} \mu_\alpha - \mu_{0,\alpha} + \phi_\alpha + \sum_{\beta=1}^p \mathcal{M}_{\beta\alpha} \lambda_\beta = 0 \\ \mathbf{m}_\alpha - \mathbf{m}_{0,\alpha} + \mathcal{A}^T \phi_\alpha = 0 \qquad \forall 1 \le \alpha \le n, \\ \mu_\alpha - \mu_\alpha^k + \mathcal{A}\mathbf{m}_\alpha = 0 \end{cases}$$

$$\text{and } C_\beta(\boldsymbol{\mu}, \lambda_\beta) = 0, \quad \forall 1 \le \beta \le p.$$

where (ϕ_α) and (λ_β) are the Lagrangian multipliers associated with the linear constraints and box constraints, respectively, and $C_\beta(\boldsymbol{\mu}, \lambda_\beta)$ is the β-th complementary function corresponding to the box constraint:

$$C_\beta(\boldsymbol{\mu}, \lambda_\beta) = \lambda_\beta - \left[\lambda_\beta + \varepsilon_0 \Big(\sum_{\alpha=1}^n \mathcal{M}_{\beta\alpha} \mu_\alpha - b_{0,\beta} \Big) \right]_- - \left[\lambda_\beta + \varepsilon_0 \Big(\sum_{\alpha=1}^n \mathcal{M}_{\beta\alpha} \mu_\alpha - b_{1,\beta} \Big) \right]_+$$

with $\varepsilon_0 > 0$ where we fix $\varepsilon_0 = 1$ in our computation, and $[v]_\pm$ stands for the positive part and negative part of the vector v respectively.

We now begin to describe the primal-dual active set. In this method, we iteratively update the variables using the following three steps in each iteration.

1. Given the ℓ-th update $(\boldsymbol{\mu}^\ell, \mathbf{m}^\ell)$, we compute the active sets (for all $1 \le \beta \le p$):

$$A_\ell^{\beta,-} = \Big\{ \mathbf{i} : \lambda_{\beta,\mathbf{i}}^\ell + (\mathcal{M}\boldsymbol{\mu}_\mathbf{i}^\ell)_\beta < b_{0,\beta} \Big\} \text{ and } A_\ell^{\beta,+} = \Big\{ \mathbf{i} : \lambda_{\beta,\mathbf{i}}^\ell + (\mathcal{M}\boldsymbol{\mu}_\mathbf{i}^\ell)_\beta > b_{1,\beta} \Big\},$$

and inactive set $\mathfrak{T}_\ell^\beta = \mathfrak{T} - (A_\ell^{\beta,-} \cup A_\ell^{\beta,+})$ for all β from 1 to p. Here $A_\ell^{\beta,\pm}$ denotes the failures to satisfy the β-th box constraint at ℓ-th iterate. In order to formulate the optimality condition (4.3) in a sparse linear system form, we concatenate all the variables $\{(\mu_\alpha, \mathbf{m}_\alpha)\}$ and auxiliary variables such that

$$\mathbf{u}^{\ell+1} = \begin{bmatrix} \mu_1^{\ell+1} \\ \vdots \\ \mu_n^{\ell+1} \\ \mathbf{m}_1^{\ell+1} \\ \vdots \\ \mathbf{m}_n^{\ell+1} \end{bmatrix}, \mathbf{u}_0 = \begin{bmatrix} \mu_{0,1} \\ \vdots \\ \mu_{0,n} \\ \mathbf{m}_{0,1} \\ \vdots \\ \mathbf{m}_{0,n} \end{bmatrix}, \mathbf{v}^{\ell+1} = \begin{bmatrix} \phi^{\ell+1} \\ \lambda_{1,A_\ell^{1,-}}^{\ell+1} \\ \lambda_{1,A_\ell^{1,+}}^{\ell+1} \\ \vdots \\ \lambda_{p,A_\ell^{p,-}}^{\ell+1} \\ \lambda_{p,A_\ell^{p,+}}^{\ell+1} \end{bmatrix}, \mathbf{v}_0 = \begin{bmatrix} \mu_1^k \\ \vdots \\ \mu_n^k \\ b_{0,1} e_{A_\ell^{1,-}} \\ b_{1,1} e_{A_\ell^{1,+}} \\ \vdots \\ b_{0,p} e_{A_\ell^{p,-}} \\ b_{1,p} e_{A_\ell^{p,+}} \end{bmatrix}.$$

where $e_A = (1, \cdots, 1)^T$ is a vector with a length equals to the cardinality number $|A|$. Herein ϕ is a vector of the dimension $n \times |\mathfrak{T}|$, corresponding to the Lagrangian multiplier of the discrete continuity constraints, and $\lambda_{\beta, A^{\beta}, \pm}$ are the multipliers associated with the active sets.

2. We update $\mathbf{u}^{\ell+1}$ and $\mathbf{v}^{\ell+1}$ by solving the following saddle point form:

$$
\begin{bmatrix} I & \mathcal{J}_\ell^T \\ \mathcal{J}_\ell & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}^{\ell+1} \\ \mathbf{v}^{\ell+1} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{v}_0 \end{bmatrix}.
$$

where $\mathcal{J}_\ell$ is given by

$$
\mathcal{J}_\ell = \begin{bmatrix} I_{n\times n} \otimes I & I_{n\times n} \otimes \mathcal{A} \\ \mathcal{N}_\ell \end{bmatrix}
$$

where I is the identity matrix of the size $\dim(\mathbf{u}^{\ell+1})$, $\otimes$ is the Kronecker product, and

$$
\mathcal{N}_\ell = \left[\mathcal{M}_{\beta, \alpha} P_{A_\ell^{\beta}, -} \right]_{\beta, \alpha=1}^{p, n} = \begin{bmatrix} \mathcal{M}_{1,1} P_{A_\ell^{1}, -} & & \mathcal{M}_{1,n} P_{A_\ell^{1}, -} \\ \mathcal{M}_{1,1} P_{A_\ell^{1}, +} & & \mathcal{M}_{1,n} P_{A_\ell^{1}, +} \\ & \ddots & \\ & & \ddots \\ \mathcal{M}_{1,p} P_{A_\ell^{p}, -} & & \mathcal{M}_{n,p} P_{A_\ell^{p}, -} \\ \mathcal{M}_{1,p} P_{A_\ell^{p}, -} & & \mathcal{M}_{n,p} P_{A_\ell^{p}, -} \end{bmatrix}
$$

where P_A the selection matrix such that $P_A \lambda$ is the vector of components of λ on the index set A.

3. After obtaining $\mathbf{u}^{\ell+1}$ and $\mathbf{v}^{\ell+1}$ using step 2, we can conversely recover the multiplier (λ_β) by

(4.4) $$ \lambda_\beta^{\ell+1} = P_{A_\ell^{\beta}, -}^T \lambda_{\beta, A_\ell^{\beta}, -}^{\ell+1} + P_{A_\ell^{\beta}, +}^T \lambda_{\beta, A_\ell^{\beta}, +}^{\ell+1}. $$

In the active set method, we repeatedly solve the above system to obtain $\mathbf{u}^{\ell+1}$ and $\mathbf{v}^{\ell+1}$, and update the active sets $A_{\ell+1}^{\beta, \pm}$ and the matrix $\mathcal{J}_{\ell+1}$ by recovering the multiplier $\lambda_\beta^{\ell+1}$ by (4.4) until the active sets remain unchanged, i.e., $A_{\ell+1}^{\beta, \pm} = A_\ell^{\beta, \pm}$ for some ℓ.

Remark 4.1. We note that, if the bound of the solution is not coupled, i.e., $\mathcal{M} = I_{n\times n}$, the proximal operator $\mathrm{prox}_{\gamma G}$ can be evaluated row-wise, and this amounts to solving the following problem for each row:

$$
\min_{\mu, \mathbf{m}} \frac{1}{2} \|\mu - \mu_{\alpha,0}\|^2 + \frac{1}{2} \|\mathbf{m} - \mathbf{m}_{\alpha,0}\|^2,
$$
$$
\text{s.t. } \mu - \mu_\alpha^k + \mathcal{A}\mathbf{m} = 0, \quad b_{0,\alpha} \le \mu_\mathbf{i} \le b_{1,\alpha}, \quad \forall \mathbf{i} \in \mathfrak{T}.
$$

for given $(\mu_{\alpha,0}, \mathbf{m}_{\alpha,0}) \in \mathfrak{T}(\mathbb{R}) \times \mathfrak{J}(\mathbb{R})$. Closed form solutions to this problem has been given in [21].

4.2.2. Proximal operator of nonlinear preconditioner. The proximal problem associated with $\mathrm{prox}_{\tau \mathcal{U}^*[\boldsymbol{\nu}/\tau]}$ amounts to solve the following smooth convex optimization problem

$$
\min_{\boldsymbol{\nu}} \tau \mathcal{U}^*[\boldsymbol{\nu}/\tau] + \frac{1}{2\overline{\gamma}} \|\boldsymbol{\nu} - \boldsymbol{\nu}_0\|^2
$$

for given $\boldsymbol{\nu}_0$. However, $\mathcal{U}^*$ is not always easy to compute explicitly for given $\mathcal{U}$. In this case, one can turn to Moreau's identity (see, for example, [17]):

$$\mathrm{prox}_{\overline{\gamma}\tau\mathcal{U}^*[\boldsymbol{\nu}/\tau]}(\boldsymbol{\nu}_0) = \boldsymbol{\nu}_0 - \overline{\gamma}\mathrm{prox}_{\overline{\gamma}^{-1}\tau\mathcal{U}[\boldsymbol{\nu}]}[\boldsymbol{\nu}_0/\overline{\gamma}].$$

Alternatively, the proximal problem $\mathrm{prox}_{\overline{\gamma}^{-1}\tau\mathcal{U}[\boldsymbol{\nu}]}[\boldsymbol{\nu}_0/\overline{\gamma}]$ amounts to solve the following smooth convex optimization problem

$$\min_{\boldsymbol{\nu}} \tau\mathcal{U}[\boldsymbol{\nu}] + \frac{\overline{\gamma}}{2}\|\boldsymbol{\nu} - \boldsymbol{\nu}_0/\overline{\gamma}\|^2.$$

In either cases, the proximal problem can be solved by Newton's method for smooth optimization

$$\min_{\boldsymbol{\nu}} F_0(\boldsymbol{\nu}) \Longrightarrow \boldsymbol{\nu}^{\ell+1} = (\nabla^2 F_0(\boldsymbol{\nu}^\ell))^{-1}(\boldsymbol{\nu}^\ell - \nabla F_0(\boldsymbol{\nu}^\ell)).$$

4.2.3. Proximal operator of dual variable. The proximal problem associated with $\iota_K(Q_0, \boldsymbol{q}_0)$ amounts to solve the following projection problem for given $(Q_0, \boldsymbol{q}_0) \in \mathbb{R}^{n\times n}_{\mathrm{sym}} \times \mathbb{R}^{n\times d}$

$$(4.5) \qquad \min_{Q,\boldsymbol{q}} \frac{1}{2}\|Q - Q_0\|_2^2 + \frac{1}{2}\|\boldsymbol{q} - \boldsymbol{q}_0\|_2^2 \quad \text{s.t.} \ Q + \frac{\boldsymbol{q}\boldsymbol{q}^T}{2} \preceq_{\mathcal{S}_+} 0,$$

where $\|\boldsymbol{q}\|_2 = (\boldsymbol{q} : \boldsymbol{q})^{\frac{1}{2}}$ denotes the Frobenuous norm. This matrix projection problem involves the control Q of eigenvalue of matrices under a quadratic nonlinearity of $\boldsymbol{q}$. We introduce two possible methods for computing (4.5).

1. A formalize approach is the alternative direction method of multipliers (ADMM). The Lagrangian of (4.5) is given as follows,

$$\min_{Q,\boldsymbol{q}} \max_{0\preceq_{\mathcal{S}_+} Z} \frac{1}{2}\|Q - Q_0\|_2^2 + \frac{1}{2}\|\boldsymbol{q} - \boldsymbol{q}_0\|_2^2 + Z : \left(Q + \frac{\boldsymbol{q}\boldsymbol{q}^T}{2}\right).$$

where Z is the Lagrangian multiplier. We refer readers to [9, Example 2.24] for the introduction of this Lagrangian multiplier. In each iterate $(Q^\ell, \boldsymbol{q}^\ell, Z^\ell)$, we solve

$$Q^{\ell+1} = Q_0 - Z^\ell \ \text{and} \ \boldsymbol{q}^{\ell+1} = (I + Z^\ell)^{-1}\boldsymbol{q}_0$$

$$Z^{\ell+1} = \mathrm{proj}_{\mathcal{S}_+}(Z^\ell + \gamma R^{\ell+1}) \ \text{where} \ R^{\ell+1} = Q^{\ell+1} + \frac{\boldsymbol{q}^{\ell+1}(\boldsymbol{q}^{\ell+1})^T}{2}$$

where $\gamma > 0$ is the stepsize parameter, and $\mathrm{proj}_{\mathcal{S}_+}$ is the projection onto the space of positive semi-definite matrices. Specifically, to conduct the projection $\mathrm{proj}_{\mathcal{S}_+}$ for given Z_0, we first implement an eigenvalue decomposition, and then truncate its eigenvalues to non-negative real numbers:

$$Z_0 = U_Z \mathrm{diag}[\lambda_\alpha]_{\alpha=1}^n U_Z^T, \quad \mathrm{proj}_{\mathcal{S}_+}(Z_0) = U_Z \mathrm{diag}[\max(0, \lambda_\alpha)]_{\alpha=1}^n U_Z^T.$$

In our experiments, we find the parameters γ within $[\frac{1}{2}, 1]$ perform near the optimal. For problems of the size $(n, d) = (2, 1)$, the violation of constrains in the ADMM iterations converges to error 10^{-12} usually within 20 steps when $\gamma = 1$. However, it performs depending on the property of $(Q_0, \boldsymbol{q}_0)$ and the choice of γ, and would be inefficient for problems with large size, for example $(n, d) = (10, 3)$.

2. We also introduce a Newton method. Let $Y = Q + \frac{1}{2}\boldsymbol{q}\boldsymbol{q}^T$. For fixed $\boldsymbol{q}$, one has

$$\min_{Q}\left\{\|Q - Q_0\|_2^2 : Q \preceq_{\mathcal{S}_+} -\tfrac{1}{2}\boldsymbol{q}\boldsymbol{q}^T\right\} = \min_{Y}\left\{\|Y - (Q_0 + \tfrac{1}{2}\boldsymbol{q}\boldsymbol{q}^T)\|_2^2 : Y \preceq_{\mathcal{S}_+} 0\right\}.$$

The RHS of the above equation has minimal expressed by

$$\|Y_+(\boldsymbol{q})\|_2^2 \text{ where } Y(\boldsymbol{q}) := (Q_0 + \tfrac{1}{2}\boldsymbol{q}\boldsymbol{q}^T)$$

where $Y_+(\boldsymbol{q}) = \text{proj}_{\mathcal{S}_+}(Y(\boldsymbol{q}))$ is the positive part of the matrix. Hence, problem (4.5) is equivalent to

$$\min_{\boldsymbol{q}} F_1(\boldsymbol{q}) = \frac{1}{2}\|\boldsymbol{q} - \boldsymbol{q}_0\|_2^2 + \frac{1}{2}\|Y_+(\boldsymbol{q})\|_2^2.$$

Once we get the optimal $\boldsymbol{q}_*$, we can recover

$$Q_* = Y_-(\boldsymbol{q}_*) - \frac{1}{2}\boldsymbol{q}_*\boldsymbol{q}_*^T$$

where $Y_- = \text{proj}_{\mathcal{S}_-}(Y)$ is the negative part of the matrix. To optimize $F_1(\boldsymbol{q})$, we note that $F_1(\boldsymbol{q})$ is convex and has a subgradient given by

$$\partial F_1(\boldsymbol{q}) = \boldsymbol{q} - \boldsymbol{q}_0 + Y_+(\boldsymbol{q})\boldsymbol{q},$$

Hence, it suffice to find the critical point $\partial F_1(\boldsymbol{q}) = 0$. To solve this equation, one can consider the Newton method, where in each iterate we update

(4.6) $$\boldsymbol{q}^{\ell+1} = \boldsymbol{q}^\ell - H(\boldsymbol{q}^\ell)^{-1}[\partial F_1(\boldsymbol{q}^\ell)].$$

Here the Hessian operator $H(\boldsymbol{q})$ is given by product rule as follows,

$$H(\boldsymbol{q})[\boldsymbol{\xi}] = \lim_{\varepsilon \to 0} \frac{\partial F_1(\boldsymbol{q} + \varepsilon\boldsymbol{\xi}) - \partial F_1(\boldsymbol{q})}{\varepsilon} = \boldsymbol{\xi} + Y_+(\boldsymbol{q})\boldsymbol{\xi} + D(Y_+(\boldsymbol{q}))[\tfrac{1}{2}(\boldsymbol{\xi}\boldsymbol{q}^T + \boldsymbol{q}\boldsymbol{\xi}^T)]\boldsymbol{q},$$

with the gradient $D(Y_+(\boldsymbol{q}))$ defined by

$$D(Y_+(\boldsymbol{q}))[Z] = V[\mathcal{A} \circ (V^T Z V)]V^T,$$

where $Y_+(\boldsymbol{q}) = V\text{diag}(\lambda_1, \cdots, \lambda_n)V^T$ and $\circ$ is the Hadamard product, i.e., component wise product, and $\mathcal{A}$ is a matrix depending on the eigenvalue of $Y_+(\boldsymbol{q})$ defined by

$$\mathcal{A}_{\alpha\beta} = \begin{cases} \dfrac{(\lambda_\alpha)_+ - (\lambda_\beta)_+}{\lambda_\alpha - \lambda_\beta} & \text{if } \lambda_\alpha \neq \lambda_\beta \\ \chi_{\{\lambda_\alpha > 0\}} & \text{if } \lambda_\alpha = \lambda_\beta \end{cases}$$

where $(\lambda_\alpha)_+ = \max(0, \lambda_\alpha)$, and χ_A is the characteristic function of set A. Once we have the information of the derivational derivative $H(\boldsymbol{q})[e_{\alpha\beta}]$ for each basis $e_{\alpha\beta}$ of $\mathbb{R}^{n\times d}$, we can expand $\partial F_1(\boldsymbol{q}^\ell)$ as a linear combination of $e_{\alpha\beta}$, and formulate $H(\boldsymbol{q})$ as a matrix on the coefficients. In this case, the Newton iterate (4.6) can be evaluated as a linear system of the coefficients of expansion.

In our experiments, error in the Newton iterate converges to error 10^{-12} usually within 10 steps.

Remark 4.2. For $n = 2$, the eigenvalue decomposition and the matrix inverse can be evaluated analytically, which greatly reduce the computation cost. In addition, a computer code for eigenvalue decomposition of matrix fields is available in [36, Figure 7].

5. Numerical results. In this section, we apply the developed PDFB splitting method to computing several cross-diffusion gradient flows (1.1).

EXAMPLE 1 ([55, 56, 2]). Consider the following SKT population model for $\boldsymbol{\mu} = (\mu_1, \mu_2)$:

$$(5.1) \qquad M(\boldsymbol{\mu}) = \left[\begin{array}{cc} \mu_1(2\mu_1 + \mu_2) & \mu_1\mu_2 \\ \mu_2\mu_1 & \mu_2(\mu_1 + 2\mu_2) \end{array} \right] \text{ and } \frac{\delta\mathcal{E}[\boldsymbol{\mu}]}{\delta\boldsymbol{\mu}} = \left[\begin{array}{c} \log(\mu_1) \\ \log(\mu_2) \end{array} \right].$$

The free energy functional associated with this model is given by

$$\mathcal{E}[\boldsymbol{\mu}] = \int_\Omega \mu_1(\log(\mu_1) - 1) + \mu_2(\log(\mu_2) - 1)\, \mathrm{d}x.$$

In this example, we consider both one dimensional and two dimensional models. For $d = 1$, we consider the following initial value

$$\mu_1(x,0) = \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(\frac{-(x - \frac{1}{2})^2}{2} \right), \quad \mu_2(x,0) = \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(\frac{-(x + \frac{1}{2})^2}{2} \right).$$

In this case, we solve (5.1) on $\Omega = [-5, 5]$ with a grid spacing $h = 0.1$ and a time step size $\tau = 0.1$. For $d = 2$, we consider the following initial value

$$\mu_1(x,y,0) = \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(\frac{-x^2 - (y + \frac{1}{2})^2}{2} \right),$$

$$\mu_2(x,y,0) = \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(\frac{-x^2 - (y - \frac{1}{2})^2}{2} \right).$$

In this case, we solve (5.1) on $\Omega = [-5, 5] \times [-5, 5]$ with a grid spacing $h = 10/256$ and a time step size $\tau = 0.1$. In both cases, to ensure bound preservation $\boldsymbol{\mu} \geq 0$, we set $\mathcal{M} = I_{2\times2}$, $\boldsymbol{b}_0 = [0; 0]$ and $\boldsymbol{b}_1 = [\infty; \infty]$ in the box constraint.

The discrete free energy of the model is

$$\mathcal{E}_h[\boldsymbol{\mu}] = \sum_{\mathbf{i} \in \mathfrak{T}} \mu_{1,\mathbf{i}}(\log(\mu_{1,\mathbf{i}}) - 1)h^d + \mu_{2,\mathbf{i}}(\log(\mu_{2,\mathbf{i}}) - 1)h^d.$$

As the Gibbs-Boltzmann entropy is ill-conditioned, one can consider the convex splitting approach by taking $\mathcal{U}[\boldsymbol{\mu}] = \widehat{\mathcal{E}}_h[\boldsymbol{\mu}]$. The convex conjugate to $\mathcal{U}$ can be explicitly computed as follows,

$$\mathcal{U}^*[\boldsymbol{\nu}] = \sum_{\mathbf{i} \in \mathfrak{T}} e^{-\nu_{1,\mathbf{i}}} + e^{-\nu_{2,\mathbf{i}}} \text{ where } \boldsymbol{\nu} = (\nu_1, \nu_2).$$

The numerical results are given in Figure 1 ($d = 1$) and Figure 2 ($d = 2$). We use reference solutions to justify our PDFB splitting method. These reference solutions are obtained by classical backward Euler method with a time step size $\tau = 10^{-3}$ on the same spatial grid with central difference approximation. In Figure 1.1 (a) and Figure 2 (d, e), we can see that the numerical solutions of the PDFB splitting method closely matches the reference solutions. In addition, to show the accuracy of the time-discrete scheme, in Figure 1.1 (b), we plot the rate of convergence for the relative error with respect to a reference solution obtained by classical backward Euler method with $\tau = 10^{-4}$, and we adopt a grid spacing $h = 10/1024$ for both of the schemes in the comparison.
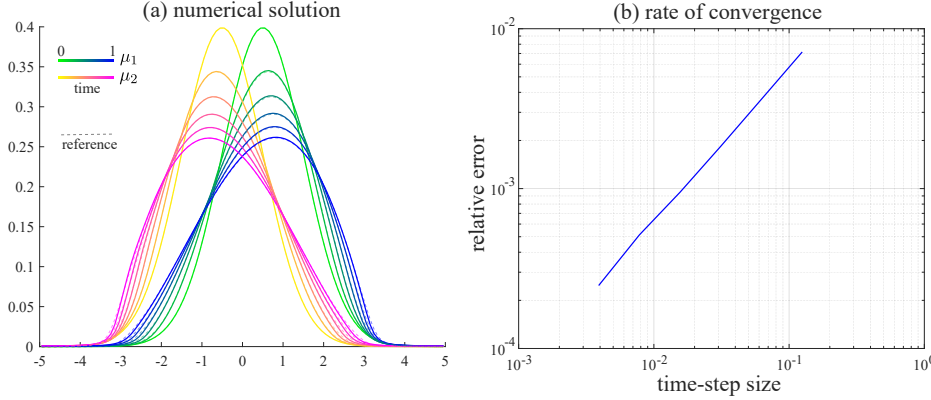
FIG. 1. Numerical results in *Example 1* in one dimension. (a) The numerical solution at several time steps, where the dash-lines represent the reference solution. (b) Plot of the relative errors at $t = 1$ with respect to different time step sizes.

EXAMPLE 2 ([3, 4, 57]). Consider the following model describing a surfactant spreading on a thin viscous film for $\boldsymbol{\mu} = (\mu_1, \mu_2)$:

$$(5.2) \qquad M(\boldsymbol{\mu}) = \begin{bmatrix} \mu_1^3/3 & \mu_1^2\mu_2/2 \\ \mu_1^2\mu_2/2 & \mu_1\mu_2^2 + \varepsilon\mu_2 \end{bmatrix} \text{ and } \frac{\delta\mathcal{E}[\boldsymbol{\mu}]}{\delta\boldsymbol{\mu}} = \begin{bmatrix} -\Delta\mu_1 \\ \log(\mu_2) \end{bmatrix},$$

where μ_1 represents the thickness of the film, μ_2 is the concentration of the surfactant. The free energy functional associated with this example is

$$\mathcal{E}[\boldsymbol{\mu}] = \int_\Omega \frac{1}{2}|\nabla\mu_1|^2 + \mu_2\big(\log(\mu_2) - 1\big)\,\mathrm{d}x.$$

In this example, we use the following initial value considered [3],

$$\mu_1(x,0) = 1, \quad \mu_2(x,0) = \frac{1}{2}(1 - \tanh(10|x| - 5)).$$

Here we solve (5.2) on $\Omega = [-4, 4]$ with two different grid spacings, a coarser grid where $h = 1/128$ and a finer grid where $h = 1/512$, and with a time step size $\tau = 0.1$. To ensure bound preservation $\mu_2 \geq 0$, we set $\mathcal{M} = I_{2\times 2}$, $\boldsymbol{b}_0 = [-\infty; 0]$ and $\boldsymbol{b}_1 = [\infty; \infty]$ in the box constraint.

The discrete free energy of the model in this example is

$$\mathcal{E}_h[\boldsymbol{\mu}] = \sum_{\mathbf{j}\in\mathfrak{J}^1} \frac{1}{2}|\partial_{1,h}\mu_1|_{\mathbf{j}}^2 h + \sum_{\mathbf{i}\in\mathfrak{T}} \mu_{2,\mathbf{i}}(\log(\mu_{2,\mathbf{i}}) - 1)h,$$

We consider the convex splitting though taking $\mathcal{U}[\boldsymbol{\mu}] = \widehat{\mathcal{E}}_h[\boldsymbol{\mu}]$. As the proximal problem associated with $\mathcal{U}[\boldsymbol{\mu}]$ is separable in μ_1 and μ_2, we handle it as two individual subproblems. Herein, the proximal problem of the Dirichlet energy can be evaluated using Moreau's identity via solving the following quadratic optimization problem (for given $\nu_{1,0}$):

$$\min_\nu \tau \sum_{\mathbf{j}\in\mathfrak{J}^1} \frac{1}{2}|\partial_{1,h}\nu|_{\mathbf{j}}^2 + \frac{\overline{\gamma}}{2}\|\nu - \nu_{1,0}/\overline{\gamma}\|^2 \implies \nu_1 = (\overline{\gamma}I - \tau\Delta_h)^{-1}\nu_{1,0}$$
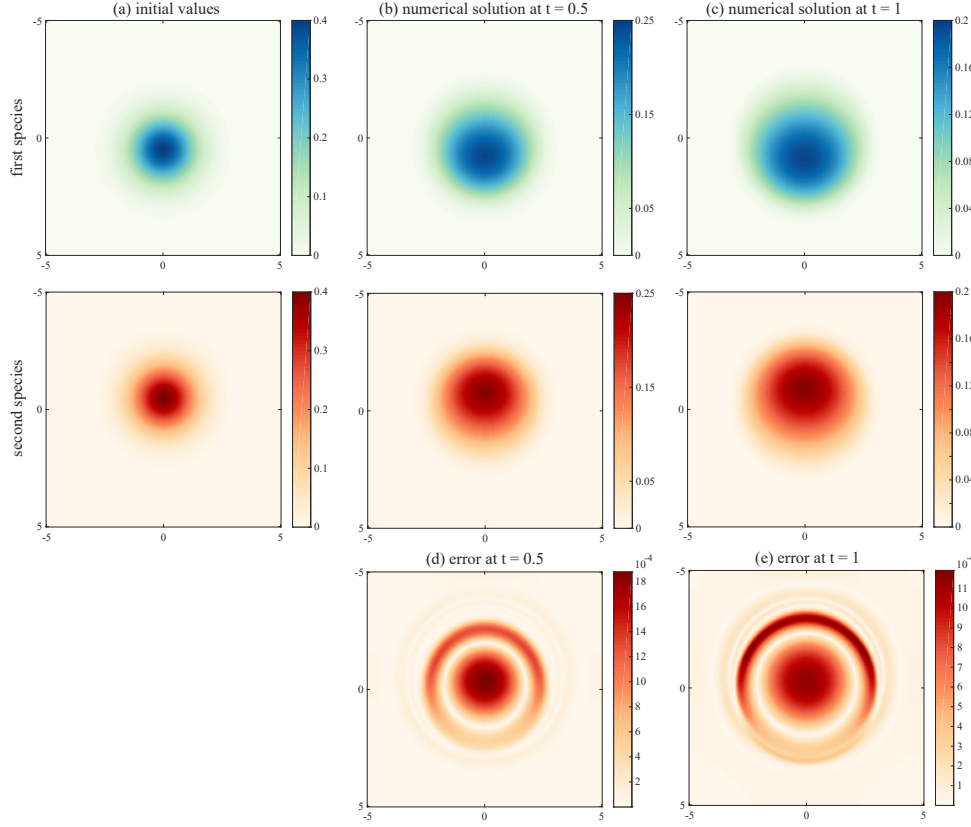
FIG. 2. Numerical results in *Example 1* in two dimension. (a) The initial values of the two species, respectively. (b) Numerical solution of the two species at $t = 0.5$, respectively. (c) Numerical solution of the two species at $t = 1$, respectively. (d) The error of the numerical solution of the second species at $t = 0.5$. (e) The error of the numerical solution of the second species at $t = 1$. In (d) and (e), the error is evaluated via $|\mu_2 - \overline{\mu}_2|$ where $\overline{\mu}_2$ is the reference solution.

where Δ_h is the centered approximation of Laplacian on uniform grids. Furthermore, the proximal operator corresponding to $\sum_{\mathbf{i}\in\mathfrak{T}} \mu_{2,\mathbf{i}}\big(\log(\mu_{2,\mathbf{i}}) - 1\big)$ can be solved the same way as *Example 1.*

The numerical results are given in Figure 3. We observe a similar behavior of the numerical solution (Figure 3 (a)) on the top of the shock as that in [56, Figure 4.3] when the spatial resolution is low, which can be addressed via increasing the spatial resolution. A reference to Figure 3 (b) is given in [3, Figure 4].

EXAMPLE 3 ([38]). Consider the liquid two-layer thin film model for $\boldsymbol{\mu} = (\mu_1, \mu_2)$

(5.3)
$$M(\boldsymbol{\mu}) = \begin{bmatrix} \mu_1^3/3 & \mu_1^3/3 + \mu_1^2(\mu_2 - \mu_1)/2 \\ \mu_1^3/3 + \mu_1^2(\mu_2 - \mu_1)/2 & (\mu_2 - \mu_1)^3/3 + \mu_1\mu_2(\mu_2 - \mu_1) + \mu_1^3/3 \end{bmatrix}$$
$$\text{and } \frac{\delta\mathcal{E}[\boldsymbol{\mu}]}{\delta\boldsymbol{\mu}} = \begin{bmatrix} -\nu\Delta\mu_1 - P'_\varepsilon(\mu_2 - \mu_1) \\ -\Delta\mu_2 + P'_\varepsilon(\mu_2 - \mu_1) \end{bmatrix},$$

where μ_1, μ_2 represents the thickness of different layers, ν is the ratio of viscosities, and P_ε represents a intermolecular potential between the two liquids. In this example,
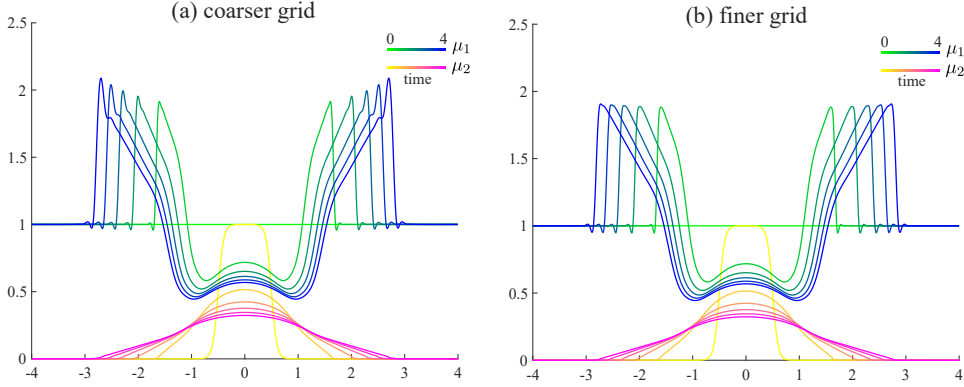
FIG. 3. Numerical results in *Example 2*. (a) Numerical solution at several time steps obtained with a coarser grid spacing $(h = 1/128)$. (b) Numerical solution at several time steps obtained with a finer grid spacing $(h = 1/512)$.

we consider the Lennard-Jones potential [49] given by

$$P_\varepsilon(r) = \frac{\varepsilon^8}{8r^8} - \frac{\varepsilon^2}{2r^2}.$$

The free energy functional associated with this model is given by

$$\mathcal{E}(\boldsymbol{\mu}) = \int_\Omega \frac{\sigma}{2}|\nabla\mu_1|^2 + \frac{1}{2}|\nabla\mu_2|^2 + P_\varepsilon(\mu_2 - \mu_1)\,\mathrm{d}x.$$

In this example, we set $\varepsilon = 0.01$, and use the following initial value,

$$\mu_1(x,0) = \frac{3}{4} - \frac{1}{4}\cos\left(\frac{\pi x}{2}\right) \text{ and } \mu_2(x,0) = 1 + \varepsilon.$$

Here we solve (5.3) on $\Omega = [-1, 1]$ with a grid spacing $h = 0.02$ and a time step size $\tau = 0.0001$. The box constraint is not considered in this example.

The discrete free energy of the model in this example is

$$\mathcal{E}_h[\boldsymbol{\mu}] = \sum_{\mathbf{j}\in\mathfrak{J}^1} \frac{\sigma}{2}|\partial_{1,h}\mu_1|_{\mathbf{j}}^2 h + \frac{1}{2}|\partial_{1,h}\mu_2|_{\mathbf{j}}^2 h + \sum_{\mathbf{i}\in\mathfrak{T}} P_\varepsilon(\mu_{2,\mathbf{i}} - \mu_{1,\mathbf{i}})h,$$

Since the Lennard-Jones potential $P_\varepsilon(\nu)$ is singular around $\nu \approx 0$, we consider the convex splitting via taking

$$\mathcal{U}[\mathcal{K}\boldsymbol{\mu}] = \sum_{\mathbf{i}\in\mathfrak{T}} P_\varepsilon(\mu_{2,\mathbf{i}} - \mu_{1,\mathbf{i}}) \text{ and } \frac{\delta\mathcal{V}[\boldsymbol{\mu}]}{\delta\boldsymbol{\mu}} = \begin{bmatrix} -\sigma\Delta_h\mu_1 \\ -\Delta_h\mu_2 \end{bmatrix},$$

where $\mathcal{K}\boldsymbol{\mu} = \mu_2 - \mu_1$. Herein, the proximal problem of the Lennard-Jones potential can be evaluated using the Moreau's identity via solving the following optimization

$$\min_\nu \tau \sum_{\mathbf{i}\in\mathfrak{T}} P_\varepsilon(\nu_{\mathbf{i}}) + \frac{\overline{\gamma}}{2}\|\nu - \nu_0/\overline{\gamma}\|^2,$$
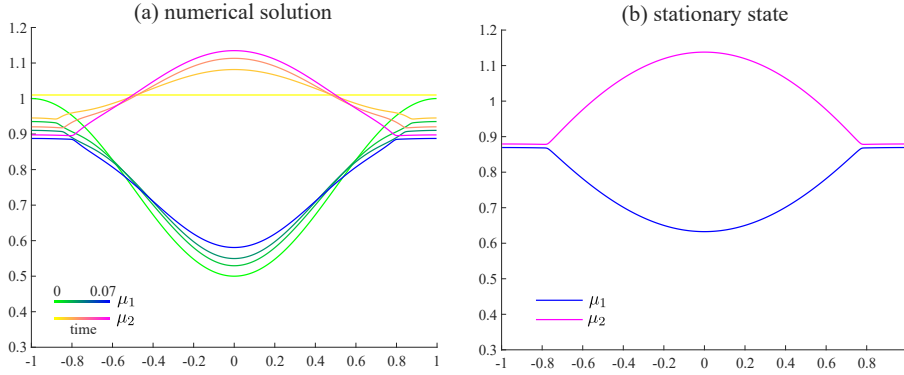
FIG. 4. Numerical results in *Example 3*. (a) The numerical solution at several time steps. (b) The numerical solution computed at $t = 0.5$ which can be seen as an approximation the stationary state of the problem.

which can be solved by several Newton iterates.

The numerical results are given in Figure 4. For large-time, we can see that the numerical solution converges to a stationary state Figure 4 (b) that resembles the theoretical stationary state [38, Fig 4.1].

EXAMPLE 4 ([2]). Consider the following system of Fokker-Planck equations with a saturation effect for $\boldsymbol{\mu} = (\mu_1, \mu_2)$

(5.4)
$$M(\boldsymbol{\mu}) = \left[\begin{array}{cc} \mu_1(1 - (\mu_1 + \mu_2)) & 0 \\ 0 & \mu_2(1 - (\mu_1 + \mu_2)) \end{array} \right]$$
$$\text{and } \frac{\delta \mathcal{E}[\boldsymbol{\mu}]}{\delta \boldsymbol{\mu}} = \left[\begin{array}{c} a(\mu_1 + \mu_2) + \sigma_1 |x|^2/2 \\ a(\mu_1 + \mu_2) + \sigma_2 |x|^2/2 \end{array} \right].$$

The free energy functional associated with this model is given by

$$\mathcal{E}[\boldsymbol{\mu}] = \int_\Omega \frac{a}{2}(\mu_1 + \mu_2)^2 + \frac{\sigma_1}{2}|x|^2\mu_1 + \frac{\sigma_2}{2}|x|^2\mu_2 \, \mathrm{d}x.$$

In this example, we set $a = 0.2$, $\sigma_1 = 4$ and $\sigma_2 = 2$, and use the following initial value,

$$\mu_1(x, 0) = \left[f(x)\left(1 - \tfrac{1}{2}\cos(\omega x)\right) \right]_+ \text{ and } \mu_2(x, 0) = \left[f(x)\left(1 + \tfrac{1}{2}\cos(\omega x)\right) \right]_+.$$

where $f(x) = \frac{2}{5}\left(1 - (\frac{4}{3}x)^2\right)$, and $\omega = 8\pi$. Here we solve (5.4) on $\Omega = [-1, 1]$ with a grid spacing $h = 0.005$ and a time step size $\tau = 0.1$. To ensure bound preservation such that $M(\boldsymbol{\mu}) \geq 0$, we set the parameters in the box constraint as that in *Remark 3.3*.

The discrete free energy of the model in this example is

$$\mathcal{E}_h[\boldsymbol{\mu}] = \sum_{\mathbf{i} \in \mathfrak{I}} \frac{a}{2}(\mu_{1,\mathbf{i}} + \mu_{2,\mathbf{i}})^2 h + \frac{\sigma_1}{2}|x_\mathbf{i}|^2\mu_{1,\mathbf{i}}h + \frac{\sigma_2}{2}|x_\mathbf{i}|^2\mu_{2,\mathbf{i}}h.$$

The convex splitting approach is not considered in this example.

The numerical results are given in Figure 5. We can see that the saturation $\mu_1 + \mu_2$ are confined in the region $[0, 1]$ at the discrete level.
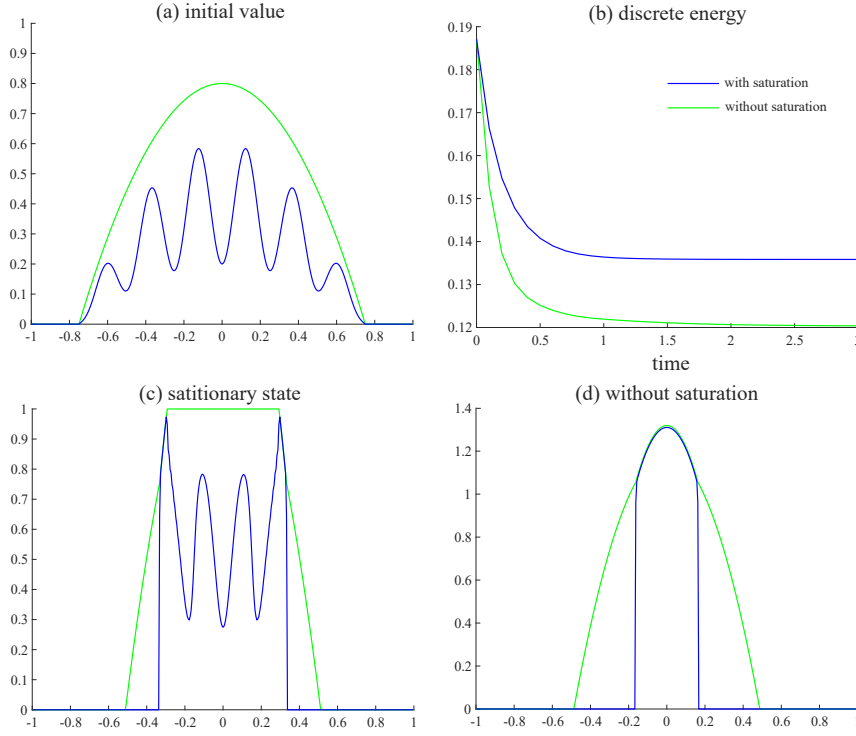
FIG. 5. Numerical results in *Example 4*. (a) The plot of the initial value. (b) The evolution of the discrete free energies. (c) Stationary state of the problem (with saturation effect). (d) Stationary state of the problem without saturation, i.e., the mobility $M(\boldsymbol{\mu}) = \mathrm{diag}(\boldsymbol{\mu})$. In (a), (c), and (d), the blue line represents the numerical solution of μ_1, and the green color represents the line represents the saturation level, i.e., $\mu_1 + \mu_2$.

6. Proof of the support function form. In this part, we prove the support function characterization in Theoerm 3.4. We let $f_{\mathrm{supp}}(M, \mathbf{m})$ be the support function form defined by

$$f_{\mathrm{supp}}(M, \mathbf{m}) = \max_{(Q, \boldsymbol{q}) \in K} M : Q + \mathbf{m} : \boldsymbol{q}.$$

In order to prove Theorem 3.4, we first show the fact that

(6.1) $$f_{\mathrm{supp}}(M, \mathbf{m}) \leq f(M, \mathbf{m}).$$

We then find $(Q_*, \boldsymbol{q}_*) \in K$ such that

(6.2) $$M : Q_* + \mathbf{m} : \boldsymbol{q}_* = f(M, \mathbf{m}),$$

for admissible pairs $(M, \mathbf{m})$, and $f_{\mathrm{supp}}(M, \mathbf{m}) = \infty$ for non-admissible pairs $(M, \mathbf{m})$.

Proof of Inequality (6.1). Consider the eigenvalue decomposition

$$M = U^T \Lambda U$$

where U is an orthonormal matrix, and $\Lambda = \mathrm{diag}[\lambda_\alpha]_{\alpha=1}^n$ is a diagonal matrix. Ap-

plying this eigenvalue decomposition to M in $f(M, \mathbf{m})$, we can show that

$$(6.3) \qquad f(M, \mathbf{m}) = \sum_{\alpha=1}^{n} f_0(\lambda_\alpha, [U\mathbf{m}]_\alpha),$$

where $[U\mathbf{m}]_\alpha$ is the α-th row of the matrix $U\mathbf{m}$, and the function f_0 is defined by,

$$f_0(\lambda_\alpha, \mathbf{m}_\alpha) = \begin{cases} \dfrac{|\mathbf{m}_\alpha|^2}{2\lambda_\alpha} & \lambda_\alpha > 0, \\ 0 & (\lambda_\alpha, \mathbf{m}_\alpha) = (0,0), \\ \infty & \text{otherwise.} \end{cases}$$

The structure of f_0 is well explored for scalar gradient flows [14, 16, 21]. In particular, we have the following support function form [5]:

$$f_0(\lambda_\alpha, [U\mathbf{m}]_\alpha) = \max_{\overline{Q}_\alpha, \overline{\mathbf{q}}_\alpha} \lambda_\alpha \overline{Q}_\alpha + [U\mathbf{m}]_\alpha \cdot \overline{\mathbf{q}}_\alpha \quad \text{s.t. } \overline{Q}_\alpha + \frac{|\overline{\mathbf{q}}_\alpha|^2}{2} \le 0.$$

Then using this formula, we can see that

$$(6.4) \qquad f(M, \mathbf{m}) = \sum_{\alpha=1}^{n} \max_{\overline{Q}_\alpha, \overline{\mathbf{q}}_\alpha} \lambda_\alpha \overline{Q}_\alpha + [U\mathbf{m}]_\alpha \cdot \overline{\mathbf{q}}_\alpha \quad \text{s.t. } \overline{Q}_\alpha + \frac{|\overline{\mathbf{q}}_\alpha|^2}{2} \le 0 \quad \forall \alpha.$$

However, such a support function characterization depends on the form of eigenvalue decomposition of the mobility M. To develop a support function as that in Theorem 3.4, we introduce the following matrix version of (6.4),

$$f(M, \mathbf{m}) = \max_{\overline{Q}, \overline{\mathbf{q}}} \Lambda : \overline{Q} + U\mathbf{m} : \overline{\mathbf{q}} \quad \text{s.t. } (\overline{Q}, \overline{\mathbf{q}}) \in \mathbb{R}_{\text{sym}}^{n \times n} \times \mathbb{R}^{n \times d} \text{ and } \overline{Q}_\alpha + \frac{|\overline{\mathbf{q}}_\alpha|^2}{2} \le 0 \quad \forall \alpha.$$

where $\overline{Q}_\alpha$ is the α-th diagonal element of the symmetrical matrix $\overline{Q}$, and $\overline{\mathbf{q}}_\alpha$ is the α-th row of the matrix $\overline{\mathbf{q}}$. This support function form holds true because the matrix Λ is zero outside the diagonal. Besides, the constraint $\overline{Q}_\alpha + \frac{1}{2}|\overline{\mathbf{q}}_\alpha|^2 \le 0$ in (6.4) is equivalent to the diagonal elements of the matrix $\overline{Q} + \frac{1}{2}\overline{\mathbf{q}}\overline{\mathbf{q}}^T$ being non-positive.

We then consider the change of variable $Q = U^T \overline{Q} U$ and $\mathbf{q} = U^T \overline{\mathbf{q}}$. By using the fact that U is an orthonormal matrix, we have

$$(6.5) \qquad \begin{aligned} \Lambda : \overline{Q} + U\mathbf{m} : \overline{\mathbf{q}} &= U^T \Lambda U : U^T \overline{Q} U + \mathbf{m} : U^T \overline{\mathbf{q}} \\ &= M : Q + \mathbf{m} : \mathbf{q}. \end{aligned}$$

We also note that

$$U\left(Q + \frac{\mathbf{q}\mathbf{q}^T}{2}\right)U^T = \overline{Q} + \frac{\overline{\mathbf{q}}\overline{\mathbf{q}}^T}{2}.$$

Hence, from (6.5) and by using the fact that

$$U_\alpha\left(Q + \frac{\mathbf{q}\mathbf{q}^T}{2}\right)U_\alpha^T = e_\alpha^T\left(\overline{Q} + \frac{\overline{\mathbf{q}}\overline{\mathbf{q}}^T}{2}\right)e_\alpha = \overline{Q}_\alpha + \frac{|\overline{\mathbf{q}}_\alpha|^2}{2} \le 0,$$

where e_α is the α-th vector of the standard basis of $\mathbb{R}^n$ and $U_\alpha = e_\alpha^T U$, one can show that the action function f can be characterized using the following support function

form.

$$f(M, \mathbf{m}) = \max_{Q, \boldsymbol{q}} M : Q + \mathbf{m} : \boldsymbol{q}$$

(6.6)

$$\text{s.t. } (Q, \boldsymbol{q}) \in \mathbb{R}^{n \times n}_{\text{sym}} \times \mathbb{R}^{n \times d} \text{ and } U_\alpha \left(Q + \frac{\boldsymbol{q}\boldsymbol{q}^T}{2} \right) U_\alpha^T \leq 0 \quad \forall \alpha.$$

Finally, we note that the admissible dual variables $(Q, \boldsymbol{q})$ are characterized by $Q + \frac{1}{2}\boldsymbol{q}\boldsymbol{q}^T \preceq_{\mathcal{S}_+} 0$ in $f_{\text{supp}}(M, \mathbf{m})$. By using the variational characterization of eigenvalues, we have $U_*^T(Q + \frac{1}{2}\boldsymbol{q}\boldsymbol{q}^T)U_* \leq 0$ for all $U_* \in \mathbb{R}^n$. Hence, the condition $Q + \frac{1}{2}\boldsymbol{q}\boldsymbol{q}^T \preceq_{\mathcal{S}_+} 0$ is sufficient for the constraint in (6.6), which gives us

$$f(M, \mathbf{m}) \geq f_{\text{supp}}(M, \mathbf{m}) = \max_{Q, \boldsymbol{q}} M : Q + \mathbf{m} : \boldsymbol{q} \quad \text{s.t. } Q + \frac{\boldsymbol{q}\boldsymbol{q}^T}{2} \preceq_{\mathcal{S}_+} 0,$$

as maximization of the same objective function over a smaller set results in a smaller maximum. □

Construction of the recovery sequence (6.2).
1. In the *admissible* case where M is positive semi-definite and invertible, i.e., $0 \preceq_{\mathcal{S}_+} M$ and there exists $\mathbf{u}$ such that $\mathbf{m} = M\mathbf{u}$, we let

$$Q_* = -\frac{1}{2}M^\dagger \mathbf{m}\mathbf{m}^T M^\dagger \quad \boldsymbol{q}_* = M^\dagger \mathbf{m}.$$

It is direct to show that $Q_* + \frac{1}{2}\boldsymbol{q}_*\boldsymbol{q}_*^T = 0$, and therefore $(Q_*, \boldsymbol{q}_*) \in K$. In addition, we have

$$M : Q_* + \mathbf{m} : \boldsymbol{q}_* = \frac{1}{2}\mathbf{m} : M^\dagger \mathbf{m} = f(M, \mathbf{m}).$$

which means that $f_{\text{supp}}(M, \mathbf{m}) = f(M, \mathbf{m})$.

2. In the *non-admissible* case where the mobility is not positive semi-definite $0 \nprec_{\mathcal{S}_+} M$, we consider the eigenvalue decomposition of $M = U^T \Lambda U$, where $\Lambda = \text{diag}[\lambda_\alpha]_{\alpha=1}^n$. Since $0 \nprec_{\mathcal{S}_+} M$, we can always find an negative eigenvalue λ_α of M, and we then consider

$$Q_n = U^T \text{diag}\Big(\underbrace{[0, \cdots, -n, \cdots, 0]}_{\alpha\text{-th is non-zero}} \Big)U, \quad \boldsymbol{q}_n = 0.$$

Hence, we have $M : Q_n + \mathbf{m} : \boldsymbol{q}_n = -n\lambda_\alpha$, which converges to ∞ as $n \to \infty$.

3. In the *non-admissible* case where $\mathbf{m}$ is non-vanishing in the kernel of the mobility M, i.e., $\mathbf{m} \neq M\mathbf{u}$ for all $\mathbf{u}$, we consider the eigenvalue decomposition $M = U^T \Lambda U$, where $\Lambda = \text{diag}[\lambda_\alpha]_{\alpha=1}^n$. Since $\mathbf{m} \neq M\mathbf{u}$ for all $\mathbf{u}$, there exists a row $(U\mathbf{m})_\alpha \neq 0$ such that $\lambda_\alpha = 0$, we then let

$$Q_n = U^T \text{diag}\Big(\underbrace{[0, \cdots, -\tfrac{n^2}{2}|(U\mathbf{m})_\alpha|^2, \cdots, 0]}_{\alpha\text{-th is non-zero}} \Big)U, \quad \boldsymbol{q}_n = nU^T \begin{bmatrix} 0 \\ \vdots \\ (U\mathbf{m})_\alpha \\ \vdots \\ 0 \end{bmatrix}.$$

Hence, we have $M : Q_n + \mathbf{m} : \boldsymbol{q}_n = n|(U\mathbf{m})_\alpha|^2$, which converges to ∞ as $n \to \infty$. □

7. Conclusion. In this work, we constructed a primal-dual forward-backward (PDFB) splitting method for computing cross-diffusion gradient flows via solving their minimizing movements at the fully discrete level. By using the PDFB splitting method developed in [21], we are able to propose a computational framework that can efficiently reduce the minimizing movements to a general convex-concave saddle point problem, where the action functionals involving mobility matrices are decoupled via auxiliary dual variables.

There are several future directions associated with this work. One is to develop numerical solvers for general reaction-diffusion systems [47, 42] based on the current PDFB splitting method. For the numerical analysis, it would be crucial to prove the convergence of the minimizing movements to the cross diffusion gradient flows for general dimensions, where one dimensional case have been addressed in [61]. Furthermore, it can be important to prove the discrete-to-continuous convergence of the staggered grid approximated distance to the "continuous" transport distance (1.5), where these results have been shown for the Wasserstein distance and distances induced by scalar mobilities [32, 34, 30, 35]. Besides, the authors believe that the current PDFB splitting methods for cross-diffusion systems can be potentially extended to computing optimal transport of matrix valued measures [11] which finds its application in quantum mechanics [19], and dynamics of incompressible fluid [10].

REFERENCES

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Springer, 2005.

[2] R. Bailo, J. A. Carrillo, and J. Hu, *Bound-preserving finite-volume schemes for systems of continuity equations with saturation*, SIAM J. Appl. Math., 83 (2023), pp. 1315–1339.

[3] J. W. Barrett, H. Garcke, and R. Nürnberg, *Finite element approximation of surfactant spreading on a thin film*, SIAM J. Numer. Anal., 41 (2003), pp. 1427–1464.

[4] J. W. Barrett and R. Nürnberg, *Convergence of a finite-element approximation of surfactant spreading on a thin film in the presence of van der Waals forces*, IMA J. Numer. Anal., 24 (2004), pp. 323–363.

[5] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[6] J.-D. Benamou, G. Carlier, and M. Laborde, *An augmented lagrangian approach to Wasserstein gradient flows and applications*, ESAIM Proc., 54 (2016), pp. 1–17.

[7] A. Blanchet, V. Calvez, and J. A. Carrillo, *Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model*, SIAM J. Numer. Anal., 46 (2008), pp. 691–721.

[8] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Springer, 2013.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[10] Y. Brenier, *The initial value problem for the Euler equations of incompressible fluids viewed as a concave maximization problem*, Commun. Math. Phys., 364 (2018), pp. 579–605.

[11] Y. Brenier and D. Vorotnikov, *On optimal transport of matrix-valued measures*, SIAM J. Math. Anal., 52 (2020), pp. 2849–2873.

[12] C. Cancès, T. Gallouët, M. Laborde, and L. Monsaingeon, *Simulation of multiphase porous media flows with minimising movement and finite volume schemes*, Eur. J. Math., 30 (2019), pp. 1123–1152.

[13] C. Cancès and A. Zurek, *A convergent finite volume scheme for dissipation driven models with volume filling constraint*, Numer. Math., 151 (2022), pp. 279–328.

[14] J. A. Carrillo, K. Craig, L. Wang, and C. Wei, *Primal dual methods for Wasserstein gradient flows*, Foundations of Computational Mathematics, 22 (2021), pp. 389 – 443.

[15] J. A. Carrillo, R. J. McCann, and C. Villani, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Rev. Mat. Iberoam., 19 (2003), pp. 971–1018.

[16] J. A. Carrillo, L. Wang, and C. Wei, *Structure preserving primal dual methods for gradient flows with nonlinear mobility transport distances*, SIAM J. Numer. Anal., 62 (2024), pp. 376–399.

[17] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[18] A. Chambolle and T. Pock, *On the ergodic convergence rates of a first-order primal–dual algorithm*, Math. Program., 159 (2016), pp. 253–287.

[19] Y. Chen, T. T. Georgiou, and A. Tannenbaum, *Interpolation of matrices and matrix-valued densities: The unbalanced case*, Eur. J. Math., 30 (2019), pp. 458–480.

[20] N. David, T. Dębiec, M. Mandal, and M. Schmidtchen, *A degenerate cross-diffusion system as the inviscid limit of a nonlocal tissue growth model*, SIAM J. Math. Anal., 56 (2024), pp. 2090–2114.

[21] Y. Deng, L. Wang, and C. Wei, *Efficient primal-dual forward-backward splitting method for Wasserstein-like gradient flows with general nonlinear mobilities*, arXiv preprint arXiv:2504.12713, (2025).

[22] L. Desvillettes, K. Fellner, M. Pierre, and J. Vovelle, *Global existence for quadratic systems of reaction-diffusion*, Adv. Nonlinear Stud., 7 (2007), pp. 491–511.

[23] R. Ducasse, F. Santambrogio, and H. Yoldaş, *A cross-diffusion system obtained via (convex) relaxation in the JKO scheme*, Calc. Var. Partial Differential Equations, 62 (2023), p. 29.

[24] C. Elliott and H. Garcke, *Diffusional phase transitions in multicomponent systems with a concentration dependent mobility matrix*, Physica D: Nonlinear Phenomena, 109 (2000), pp. 242–256.

[25] S. Fagioli and O. Tse, *On gradient flow and entropy solutions for nonlocal transport equations with nonlinear mobility*, Nonlinear Anal., 221 (2022), p. 112904.

[26] D. Forkert, J. Maas, and L. Portinale, *Evolutionary Γ-convergence of entropic gradient flow structures for Fokker-Planck equations in multiple dimensions*, SIAM J. Math. Anal., 54 (2022), pp. 4297–4333.

[27] G. Fu, S. Osher, W. Pazner, and W. Li, *Generalized optimal transport and mean field control problems for reaction-diffusion systems with high-order finite element computation*, J. Comput. Phys., 508 (2024), p. 112994.

[28] G. Fu, S. J. Osher, and W. Li, *High order spatial discretization for variational time implicit schemes: Wasserstein gradient flows and reaction-diffusion systems*, J. Comput. Phys., 491 (2023), p. 112375.

[29] Y. Gao and N. K. Yip, *Homogenization of Wasserstein gradient flows*, arXiv preprint arXiv:2312.01584, (2023).

[30] N. García Trillos, *Gromov-Hausdorff limit of Wasserstein spaces on point clouds*, Calc. Var. Partial Differential Equations, 59 (2020), p. 73.

[31] H. Garcke and S. Wieland, *Surfactant spreading on thin viscous films: Nonnegative solutions of a coupled degenerate system*, SIAM J. Math. Anal., 37 (2006), p. 2025–2048.

[32] N. Gigli and J. Maas, *Gromov–Hausdorff convergence of discrete transportation metrics*, SIAM J. Math. Anal., 45 (2013), pp. 879–899.

[33] V. Giovangigli and M. Massot, *Entropic structure of multicomponent reactive flows with partial equilibrium reduced chemistry*, Math. Methods Appl. Sci., 27 (2004), pp. 739–768.

[34] P. Gladbach, E. Kopfer, and J. Maas, *Scaling limits of discrete optimal transport*, SIAM J. Math. Anal., 52 (2020), pp. 2759–2802.

[35] P. Gladbach, E. Kopfer, J. Maas, and L. Portinale, *Homogenisation of one-dimensional discrete optimal transport*, Journal de Mathématiques Pures et Appliquées, 139 (2020), pp. 204–234.

[36] B. Goldluecke, E. Strekalovskiy, and D. Cremers, *The natural vectorial total variation which arises from geometric measure theory*, SIAM J. Imaging Sci., 5 (2012), pp. 537–563.

[37] E. Y. Hamedani and N. S. Aybat, *A primal-dual algorithm for general convex-concave saddle point problems*, arXiv preprint arXiv:1803.01401, 2 (2018), p. 71.

[38] S. Jachalski, R. Huth, G. Kitavtsev, D. Peschka, and B. Wagner, *Stationary solutions of liquid two-layer thin-film models*, SIAM J. Appl. Math., 73 (2012), pp. 1183–1202.

[39] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the fokker–planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17.

[40] I. Kim and A. R. Mészáros, *On nonlinear cross-diffusion systems: an optimal transport approach*, Calc. Var. Partial Differential Equations, 57 (2018), p. 79.

[41] W. Li, J. Lu, and L. Wang, *Fisher information regularization schemes for wasserstein gradient flows*, J. Comput. Phys., 416 (2020), p. 109449.

[42] M. Liero and A. Mielke, *Gradient structures and geodesic convexity for reaction-diffusion systems*, Philos. Trans. A, 371 (2013), pp. 20120346–20120346.

[43] M. Liero, A. Mielke, and G. Savaré, *Optimal transport in competition with reaction: The Hellinger–Kantorovich distance and geodesic curves*, SIAM J. Math. Anal., 48 (2016), pp. 2869–2911.

[44] S. Lisini, D. Matthes, and G. Savar'e, *Cahn–Hilliard and thin film equations with nonlinear mobility as gradient flows in weighted-Wasserstein metrics*, J. Differential Equations, 253 (2012), pp. 814–850.

[45] Y. Malitsky and M. K. Tam, *A forward-backward splitting method for monotone inclusions without cocoercivity*, SIAM J. Optim., 30 (2018), pp. 1451–1472.

[46] A. Manzoni, A. M. Quarteroni, and S. Salsa, *Optimal Control of Partial Differential Equations: Analysis, Approximation, and Applications*, Springer, 2021.

[47] A. Mielke, *A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems*, Nonlinearity, 24 (2011), p. 1329.

[48] T. J. Murphy and N. J. Walkington, *Control volume approximation of degenerate two-phase porous flows*, SIAM J. Numer. Anal., 57 (2019), pp. 527–546.

[49] A. Oron, S. H. Davis, and S. G. Bankoff, *Long-scale evolution of thin liquid films*, Rev. Modern Phys., 69 (1997), pp. 931–980.

[50] F. Otto, *The geometry of dissipative evolution equations: The porous medium equation*, Commun. Partial Differ. Equ., 26 (2001), pp. 101–174.

[51] F. Otto and M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal., 37 (2005), pp. 1227–1255.

[52] N. Papadakis, G. Peyré, and E. Oudet, *Optimal transport with proximal splitting*, SIAM J. Imaging Sci., 7 (2014), pp. 212–238.

[53] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, 2011 International Conference on Computer Vision, (2011), pp. 1762–1769.

[54] B. Seguin and N. Walkington, *Multi-component multiphase porous flow*, Arch. Ration. Mech., 235 (2020), p. 2171–2196.

[55] N. Shigesada, K. Kawasaki, and E. Teramoto, *Spatial segregation of interacting species*, J. Theoret. Biol., 79 (1979), pp. 83–99.

[56] Z. Sun, J. A. Carrillo, and C.-W. Shu, *An entropy stable high-order discontinuous Galerkin method for cross-diffusion gradient flow systems*, Kinet. Relat. Models, (2018).

[57] U. Thiele, A. J. Archer, and M. Plapp, *Thermodynamically consistent description of the hydrodynamics of free surfaces covered by insoluble surfactants of high concentration*, Physics of Fluids, 24 (2012).

[58] C. Villani, *Optimal Transport: Old and New*, Springer Berlin, 2009.

[59] M. Yan, *A new primal–dual algorithm for minimizing the sum of three functions with a linear operator*, J. Sci. Comput., 76 (2016), pp. 1698–1717.

[60] Y. Zhu, D. Liu, and Q. Tran-Dinh, *New primal-dual algorithms for a class of nonsmooth and nonlinear convex-concave minimax problems*, SIAM J. Optim., 32 (2022), pp. 2580–2611.

[61] J. Zinsl and D. Matthes, *Transport distances and geodesic convexity for systems of degenerate diffusion equations*, Calc. Var. Partial Differential Equations, 54 (2015), pp. 3397–3438.