Journal of Computational Physics 491 (2023) 112375

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

High order spatial discretization for variational time implicit schemes: Wasserstein gradient flows and reaction-diffusion systems

Guosheng Fu [a,1], Stanley Osher [b,2], Wuchen Li [c,*,3]

[a] Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556, USA
[b] Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095, USA
[c] Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA

ARTICLE INFO

Article history:
Received 14 March 2023
Received in revised form 10 July 2023
Accepted 11 July 2023
Available online 1 August 2023

Keywords:
High order computation
Entropy dissipation
Metric spaces
Generalized Jordan–Kinderlehrer–Otto schemes
Wasserstein gradient flows
Reversible reaction-diffusion systems

ABSTRACT

We design and compute first-order implicit-in-time variational schemes with high-order spatial discretization for initial value gradient flows in generalized optimal transport metric spaces. We first review some examples of gradient flows in generalized optimal transport spaces from the Onsager principle. We then use a one-step time relaxation optimization problem for time-implicit schemes, namely generalized Jordan-Kinderlehrer-Otto schemes. Their minimizing systems satisfy implicit-in-time schemes for initial value gradient flows with first-order time accuracy. We adopt the first-order optimization scheme ALG2 (Augmented Lagrangian method) and high-order finite element methods in spatial discretization to compute the one-step optimization problem. This allows us to derive the implicit-in-time update of initial value gradient flows iteratively. We remark that the iteration in ALG2 has a simple-to-implement point-wise update based on optimal transport and Onsager's activation functions. The proposed method is unconditionally stable for convex cases. Numerical examples are presented to demonstrate the effectiveness of the methods in two-dimensional PDEs, including Wasserstein gradient flows, Fisher–Kolmogorov-Petrovskii-Piskunov equation, and two and four species reversible reaction-diffusion systems.

1. Introduction

Dissipative dynamics (gradient flows) are essential models in thermodynamics, chemistry, materials science, biological swarming, robotics path panning, and social sciences [19,60]. Nowadays, they also find vast applications in designing machine learning optimization algorithms and Markov-Chain-Monte-Carlo sampling algorithms [2,15,20,27,28,43,46,70]. In physics, dissipative dynamics describe that the systems have maximum efficiency, in which dynamics follow from the direction in which the (negative) entropy/Lyapunov functional dissipates most rapidly. It turns out that the dissipative dynamics are gradient flows in suitable metric spaces. Fast, efficient, and accurate dissipative dynamics simulations are one of the central problems in computational fluid dynamics.

* Corresponding author.
E-mail addresses: gfu@nd.edu (G. Fu), sjo@math.ucla.edu (S. Osher), wuchen@mailbox.sc.edu (W. Li).

[1] G. Fu's work is supported in part by NSF DMS-2134168.
[2] S. Osher's work is supported in part by AFOSR MURI FP 9550-18-1-502, and ONR grants: N00014-20-1-2093, and N00014-20-1-2787.
[3] W. Li's work is supported by AFOSR MURI FP 9550-18-1-502, AFOSR YIP award No. FA9550-23-1-008, NSF DMS-2245097, and NSF RTG: 2038080.

https://doi.org/10.1016/j.jcp.2023.112375
0021-9991/

A particular type of gradient flow has been widely studied in optimal transport, where the metric is known as the Wasserstein-2 metric [3,69]. Typical examples include gradient drift Fokker-Planck equations, porous media equations, aggregation-diffusion equations, etc. One property of simulating gradient flows is that one can design a proximal method for computing a variational implicit time algorithm. This algorithm was first proposed by Jordan-Kinderlehrer-Otto (JKO scheme) to compute Wasserstein gradient flows [33]. Moreover, general gradient flows have been widely studied. They follow the Onsager principle to design optimal transport-type metric spaces [27,39,55]. Similarly, one can develop variational proximal methods to compute and simulate gradient flow dynamics.

This paper designs high-order spatial discretization in simulating gradient flow dynamics using variational proximal schemes in generalized optimal transport metric spaces. We formally illustrate the main computational framework. Consider a reaction-diffusion type equation:

$$\partial_t \rho = \nabla \cdot (V_1(\rho)\nabla \frac{\delta}{\delta \rho}\mathcal{E}(\rho)) - V_2(\rho)\frac{\delta}{\delta \rho}\mathcal{E}(\rho), \tag{1.1}$$

where $\rho\colon \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$ is a scalar density function, $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, is a spatial domain with periodic or Neumann boundary conditions, V_1, $V_2\colon \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$ are positive mobility functions (Onsager activation functions), and $\mathcal{E}(\rho) \in \mathbb{R}$ is a Lyapunov functional (energy). We design a variational implicit time scheme, the linearized JKO scheme [9,42], to update equation (1.1) as below:

$$\rho^n = \arg\min_{\rho} \inf_{(\rho,m,s)} \quad \frac{1}{2\Delta t}\int_\Omega \Big[\frac{\|m\|^2}{V_1(\rho)} + \frac{|s|^2}{V_2(\rho)}\Big]dx + \mathcal{E}(\rho), \tag{1.2a}$$

where $\Delta t \geq 0$ is a stepsize and the minimization is over all functions $\rho\colon \Omega \to \mathbb{R}_+$, $m\colon \Omega \to \mathbb{R}^d$, and $s\colon \Omega \to \mathbb{R}$, subject to the constraint

$$\rho - \rho^{n-1} + \nabla \cdot m = s, \text{ on } \Omega. \tag{1.2b}$$

We use time rescaling of (m, s) in the constraint (1.2b). We then compute variational problem (1.2) iteratively to find the sequence ρ^n, $n = 1, 2, \cdots$. This sequence forms an implicit update for gradient flow dynamic (1.1), which is first-order in time:

$$\frac{\rho^n - \rho^{n-1}}{\Delta t} = \nabla \cdot (V_1(\rho^n)\nabla \frac{\delta}{\delta \rho}\mathcal{E}(\rho^n)) - V_2(\rho^n)\frac{\delta}{\delta \rho}\mathcal{E}(\rho^n) + \mathcal{O}(\Delta t).$$

When V_1, V_2 is concave in terms of ρ, and $\mathcal{E}$ is a convex functional, then the proposed method is unconditionally stable, meaning that we can take large time steps.

Our framework also works for reversible reaction-diffusion systems with detailed balance [55,30,48]. We illustrate the main idea for a simple 2-component reversible reaction-diffusion system: Let X_1, X_2 be two species with a single reversible reaction $X_1 \underset{k_-}{\overset{k_+}{\rightleftharpoons}} X_2$, with $k_-, k_+ > 0$. Let ρ_1 and ρ_2 be the respective densities of X_1 and X_2. This leads to the following PDE system [55,61]:

$$\begin{aligned}
\partial_t \rho_1 - \gamma_1 \Delta \rho_1 &= -(k_+\rho_1 - k_-\rho_2),\\
\partial_t \rho_2 - \gamma_2 \Delta \rho_2 &= (k_+\rho_1 - k_-\rho_2),
\end{aligned}$$

with positive diffusion rates $\gamma_1, \gamma_2 > 0$. By introducing the following mobility functions,

$$V_{1,1}(\rho_1) = \gamma_1\rho_1, \quad V_{1,2}(\rho_2) = \gamma_2\rho_2, \quad V_2(\rho_1, \rho_2) = \frac{k_+\rho_1 - k_-\rho_2}{\log(k_+\rho_1) - \log(k_-\rho_2)}$$

and the energies

$$\mathcal{E}_1(\rho_1) = \int_\Omega \rho_1(\log(k_+\rho_1) - 1)\,dx, \quad \mathcal{E}_2(\rho_2) = \int_\Omega \rho_2(\log(k_-\rho_2) - 1)\,dx,$$

the above PDE system can be recast into the following system version of the form (1.1):

$$\partial_t \rho_1 = \nabla \cdot \left(V_{1,1}(\rho_1)\nabla \frac{\delta \mathcal{E}_1}{\delta \rho}(\rho_1)\right) - V_2(\rho_1, \rho_2)\left(\frac{\delta \mathcal{E}_1}{\delta \rho}(\rho_1) - \frac{\delta \mathcal{E}_2}{\delta \rho}(\rho_2)\right), \tag{1.3a}$$

$$\partial_t \rho_2 = \nabla \cdot \left(V_{1,2}(\rho_2)\nabla \frac{\delta \mathcal{E}_2}{\delta \rho}(\rho_2)\right) + V_2(\rho_1, \rho_2)\left(\frac{\delta \mathcal{E}_1}{\delta \rho}(\rho_1) - \frac{\delta \mathcal{E}_2}{\delta \rho}(\rho_2)\right), \tag{1.3b}$$

which can then be discretized using a similar variational time implicit scheme as (1.2). Here the system is called a *strongly reversible reaction-diffusion system* when $k_+ = k_- > 0$, is called a *reversible reaction-diffusion system with detailed balance* when we allow the two positive reaction rates to be different $k_+ \neq k_- > 0$, and is called an *irreversible reaction-diffusion system* when the backward reaction rate is zero $k_- = 0$; see more detailed in [30,48,55]. Our framework does not directly work for irreversible reaction-diffusion systems as they do not satisfy an energy dissipation law and can not be formulated back to the form (1.3). However, we can approximate an irreversible reaction-diffusion system using a reversible one by using a very small backward reaction rate (see, e.g., [44]) and then solve the reversible system using our formulation.

This paper adopts the augmented Lagrangian (ALG2) optimization method with high-order spatial finite element discretizations to solve the variational problem (1.2). Using finite element spatial discretization, we also develop a point-wise update in the optimization step of computing variational problem (1.2). In this sense, we obtain a high-order spatial discretization scheme in finding the ground state, which is the minimizer of functional $\mathcal{E}$. In this iterative procedure, assuming that the optimization step finds a global minimizer, the Lyapunov functional $\mathcal{E}$ is guaranteed to decay for any large time step sizes.

Computational optimal transport and mean field control/games have been widely investigated in [1,4,6,52,58,63,66]. For example, generalized JKO schemes of Wasserstein gradient flows with first-order time accuracy have been studied in [9,12,26,42,48]. Semi-discretizations of JKO-type schemes have been used in [16]. The Lagrangian type JKO schemes have been investigated in [13,14,51]. It is also worth mentioning that there are methods for high-order time discretizations of gradient flows [31]. Meanwhile, generalized optimal transport metric spaces have recently been introduced in [11,18,22,54,55]. Study of conservative and dissipative operators in non-equilibrium thermodynamics [56,57,72] is an active research area. However, there are limited JKO-type computational results for reaction-diffusion systems. We specifically mention the recently introduced variational operator splitting schemes [48–50] for reversible reaction diffusion systems using the energetic variational framework [29,47]. The studies [48–50] compute implicit schemes for reaction-diffusion equations. They split Wasserstein-type gradient flows and reaction terms and compute them separately. They first compute the reaction term in a constructed algebraic equation and then use the implicit scheme to approximate the Wasserstein-type gradient flows. In our approach, we adopt the generalized Wasserstein-type metric directly, which contains both the Wasserstein metric and the reaction metric with mobility functions V_1 and V_2. Using them, we design a scheme to approximate the proximal operator in generalized Wasserstein-type metric space. This forms the generalized JKO scheme. Thus, both algorithms [48–50] and the proposed method maintain the entropy dissipation properties. The major difference is the implicit time treatment of reaction terms. We emphasize that the method in this paper leverages the optimization structures in generalized JKO schemes.

We note that generalized JKO schemes are examples of mean field control (MFC) problems [6,35], which design optimal control/optimization problems for general initial value evolutionary equations not limited to gradient flows. Computation and modeling studies of MFCs have been conducted in controlling reaction-diffusion equations [39] and conservation laws [40,41] with applications in pandemics modeling [37,38]. Compared to the above approaches, we apply high-order spatial schemes in computing generalized JKO schemes towards initial value gradient flows. We adopt the first-order optimization method, the augmented Lagrangian method (ALG2), to implement the variational time implicit schemes for two and four species reversible reaction-diffusion systems.

This paper is organized as follows. We review some concepts of gradient flows, time implicit schemes, and their first-order optimization methods ALG2 in section 2. Several examples of dynamics, including Wasserstein gradient flows, Fisher–Kolmogorov-Petrovskii-Piskunov (KPP) equation, and reversible reaction-diffusion systems, are presented in section 3. We then present a high-order finite element method and derive all implementation details of the optimization algorithm ALG2 in section 4. Numerical examples are presented for two-dimensional Wasserstein gradient flows of linear, interaction, and potential energies, Fisher-KPP equation, and reversible two and four-species reaction-diffusion systems in section 5.

2. Optimal transport type gradient flows, generalized time implicit schemes, and first-order optimization methods

This section reviews generalized gradient flows and their variational implicit schemes in metric spaces. We also discuss a one-step time discretization relaxation of variational implicit schemes for generalized gradient flows. Entropy dissipation properties of variational implicit schemes are introduced. We then present the augmented Lagrangian method (ALG2) as the optimization solver to compute the variational implicit schemes.

2.1. Optimal transport type gradient flows

In this subsection, we formally review generalized optimal transport gradient flows [14,22,55]. This is known as the Onsager gradient flow [19]. We next discuss a class of variational schemes to compute implicit-in-time solutions of gradient flows.

2.1.1. Gradient flows and entropy dissipations

Consider an initial value equation

$$\begin{aligned}&\partial_t \rho(x,t) = \nabla\cdot(V_1(\rho(x,t))\nabla\frac{\delta}{\delta\rho}\mathcal{E}(\rho)(x,t)) - V_2(\rho(x,t))\frac{\delta}{\delta\rho}\mathcal{E}(\rho)(x,t), \quad t\in[0,\infty)\\ &\rho(x,0)=\rho^0(x).\end{aligned} \tag{2.1}$$

Here $x\in\Omega\subset\mathbb{R}^d$, Ω is a spatial domain with periodic boundary condition or Neumann boundary condition (detailed in later sections), $\rho\colon \Omega\times\mathbb{R}_+\to\mathbb{R}$ is a scalar non-negative density function satisfying

$$\rho(\cdot,t)\in\mathcal{M}=\Big\{\rho\colon \Omega\to\mathbb{R}\colon \rho(x,t)\geq 0\Big\},$$

for any time t, $\mathcal{E}\colon \mathcal{M}\to\mathbb{R}$ is an energy functional, $V_1, V_2\colon \mathbb{R}\to\mathbb{R}_+$ are positive mobility functions, $\frac{\delta}{\delta\rho}$ is the first variation operator in L^2 space, and $\rho^0\in\mathcal{M}$ is an initial condition. Equation (2.1) forms a class of equations, including Wasserstein gradient flows and the Fisher–KPP equation [64,23,34]. Detailed examples of V_1, V_2, and $\mathcal{E}$ are provided in the next section, where we also discuss the extension of (2.1) to reaction-diffusion systems.

Equation (2.1) is purely dissipative. Denote $\rho(\cdot,t)$ as the solution of the PDE (2.1), then the energy functional $\mathcal{E}$ is a Lyapunov functional. In other words, the first-time derivative of the energy functional $\mathcal{E}$ is nonpositive, satisfying

$$\begin{aligned}&\frac{d}{dt}\mathcal{E}(\rho(\cdot,t))\\ &=-\int_\Omega\Big[\|\nabla\frac{\delta}{\delta\rho}\mathcal{E}(\rho)(x,t)\|^2V_1(\rho(x,t))+|\frac{\delta}{\delta\rho}\mathcal{E}(\rho)(x,t)|^2V_2(\rho(x,t))\Big]dx\leq 0,\end{aligned} \tag{2.2}$$

where we use the fact that $V_1(\rho)\geq 0$, and $V_2(\rho)\geq 0$ in the above inequality.

2.1.2. Metric operators and distances

The dissipation of the energy functional also induces a metric function in space $\mathcal{M}$, which further defines distances between two densities $\rho^0,\rho^1\in\mathcal{M}$. This distance designs an implicit time variational problem for computing the gradient flow in metric spaces. See details among optimal transport type gradient flows, distances, and mean-field control problems in [3,39,55].

We directly present generalized optimal transport type distances and the time implicit schemes below for simplicity of discussion.

Definition. *Distance functional.* Define a distance functional $\mathrm{Dist}_{V_1,V_2}\colon \mathcal{M}\times\mathcal{M}\to\mathbb{R}_+$ as below. Consider the following optimal control problem:

$$\begin{aligned}&\mathrm{Dist}_{V_1,V_2}(\rho^0,\rho^1)^2\\ &:=\inf_{\rho,v_1,v_2}\int_0^1\int_\Omega\Big[\|v_1(x,\tau)\|^2V_1(\rho(x,\tau))+|v_2(x,\tau)|^2V_2(\rho(x,\tau))\Big]dxd\tau,\end{aligned} \tag{2.3a}$$

where the infimum is taken among $\rho\colon \Omega\times[0,1]\to\mathbb{R}_+$, $v_1, v_2\colon \Omega\times[0,1]\to\mathbb{R}^d$, such that ρ satisfies a reaction-diffusion type equation with drift vector field v_1, drift mobility V_1, reaction rate v_2, reaction mobility V_2, connecting initial and terminal densities ρ^0, ρ^1:

$$\begin{cases}\partial_\tau\rho(x,\tau)+\nabla\cdot(V_1(\rho(x,\tau))v_1(x,\tau))=V_2(\rho(x,\tau))v_2(x,\tau), \quad \tau\in[0,1],\\ \rho(x,0)=\rho^0(x), \quad \rho(x,1)=\rho^1(x).\end{cases} \tag{2.3b}$$

Variational problem (2.3) is a generalized Benamou-Brenier formula [5], where they consider $V_1(\rho)=\rho$, $V_2(\rho)=0$. One common practice is the following change of variable formula, which leads to a linear constraint optimization problem. Denote a moment vector function $m\colon \Omega\times[0,1]\to\mathbb{R}^d$ and a source function $s\colon \Omega\times[0,1]\to\mathbb{R}$, such that

$$m(x,\tau)=V_1(\rho(x,\tau))v_1(x,\tau), \quad s(x,\tau)=V_2(\rho(x,\tau))v_2(x,\tau).$$

Using variables m, s, variational problem (2.3) satisfies

$$\mathrm{Dist}_{V_1,V_2}(\rho^0,\rho^1)^2:=\inf_{\rho,m,s}\int_0^1\int_\Omega\Big[\frac{\|m(x,\tau)\|^2}{V_1(\rho(x,\tau))}+\frac{|s(x,\tau)|^2}{V_2(\rho(x,\tau))}\Big]dxd\tau,$$

such that

$$\partial_\tau \rho(x,\tau) + \nabla \cdot m(x,\tau) = s(x,\tau), \quad \rho(x,0) = \rho^0(x), \quad \rho(x,1) = \rho^1(x).$$

2.1.3. Variational time implicit schemes and properties

We next design a variational implicit-in-time scheme to update gradient flow (2.1) iteratively.

Definition. *Variational time implicit scheme.* Denote $\Delta t > 0$ as a time step size. Consider the scheme below:

$$\rho^n = \arg\min_{\rho \in \mathcal{M}} \quad \frac{1}{2\Delta t} \mathrm{Dist}_{V_1,V_2}(\rho^{n-1}, \rho)^2 + \mathcal{E}(\rho), \tag{2.4}$$

where $\mathrm{Dist}_{V_1,V_2}(\rho^{n-1}, \rho)^2$ is the distance functional defined in (2.3) between current density ρ and previous step density ρ^{n-1}. After suitable time rescaling, one can show that the minimization scheme (2.4) requires solving the following optimal control problem:

$$\inf_{\rho_{\Delta t},\rho,m,s} \quad \underbrace{\frac{1}{2}\int_0^{\Delta t}\Big[\int_\Omega \frac{\|m(x,\tau)\|^2}{V_1(\rho(x,\tau))} + \frac{|s(x,\tau)|^2}{V_2(\rho(x,\tau))}\Big]dxd\tau}_{=\frac{1}{2\Delta t}\mathrm{Dist}_{V_1,V_2}(\rho^{n-1},\rho)^2} + \mathcal{E}(\rho_{\Delta t}), \tag{2.5a}$$

such that

$$\partial_\tau \rho(x,\tau) + \nabla \cdot m(x,\tau) = s(x,\tau), \quad \tau \in [0,\Delta t], \tag{2.5b}$$

$$\rho(x,0) = \rho^{n-1}(x), \quad \rho(x,\Delta t) = \rho_{\Delta t}(x). \tag{2.5c}$$

The next step solution ρ^n is the density minimizer of (2.5):

$$\rho^n(x) = \rho_{\Delta t}(x).$$

We demonstrate that the variational scheme (2.4) is a first-order accurate implicit in time scheme, i.e.,

$$\frac{\rho^n - \rho^{n-1}}{\Delta t} = \nabla \cdot (V_1(\rho^n)\frac{\delta}{\delta \rho}\mathcal{E}(\rho^n)) - V_2(\rho^n)\frac{\delta}{\delta\rho}\mathcal{E}(\rho^n) + \mathcal{O}(\Delta t).$$

Proof. We write the minimization system of variational problem (2.5). Denote $\Phi(x,\tau) \in \mathbb{R}$, $\tau \in [0,\Delta t]$, as the Lagrange multiplier. The optimal condition of variational problem (2.5) satisfies the following saddle point problem:

$$\inf_{\rho_{\Delta t},\rho,m,s}\sup_{\Phi} \mathcal{L}(\rho_{\Delta t},\rho,m,s,\Phi), \tag{2.6}$$

where

$$\begin{aligned}\mathcal{L}(\rho_{\Delta t},\rho,m,s,\Phi) :=& \frac{1}{2}\int_0^{\Delta t}\int_\Omega \Big[\frac{\|m(x,\tau)\|^2}{V_1(\rho(x,\tau))} + \frac{|s(x,\tau)|^2}{V_2(\rho(x,\tau))}\Big]dxd\tau + \mathcal{E}(\rho_{\Delta t}) \\ &+ \int_0^{\Delta t}\int_\Omega \Phi(x,\tau)\Big(\partial_\tau\rho(x,\tau) + \nabla\cdot m(x,\tau) - s(x,\tau)\Big)dxd\tau\end{aligned}$$

We note that from integration by parts,

$$\begin{aligned}\int_0^{\Delta t}\int_\Omega \Phi(x,\tau)\partial_\tau\rho(x,\tau)dxd\tau =& -\int_0^{\Delta t}\int_\Omega \partial_\tau\Phi(x,\tau)\rho(x,\tau)dxd\tau \\ &+ \int_\Omega \Phi(x,\Delta t)\rho_{\Delta t}(x)dx - \int_\Omega \Phi(x,0)\rho^{n-1}(x)dx.\end{aligned}$$

By computing the saddle point of (2.6), we derive

$$\begin{cases} \frac{\delta}{\delta\rho}\mathcal{L}=0, & \text{if } \rho>0,\\ \frac{\delta}{\delta m}\mathcal{L}=0,\\ \frac{\delta}{\delta s}\mathcal{L}=0,\\ \frac{\delta}{\delta\Phi}\mathcal{L}=0,\\ \frac{\delta}{\delta\rho_{\Delta t}}\mathcal{L}=0, \end{cases} \Rightarrow \begin{cases} -\frac{\|m\|^2}{2V_1(\rho)^2}V_1'(\rho)-\frac{|s|^2}{2V_2(\rho)^2}V_2'(\rho)-\partial_\tau\Phi=0, & \text{if } \rho>0,\\ \frac{m}{V_1(\rho)}-\nabla\Phi=0,\\ \frac{s}{V_2(\rho)}-\Phi=0,\\ \partial_\tau\rho+\nabla\cdot m=s,\\ \Phi(x,\Delta t)+\frac{\delta}{\delta\rho_{\Delta t}}\mathcal{E}(\rho_{\Delta t})=0. \end{cases}$$

Thus we obtain a minimization system:

$$\begin{cases} \partial_\tau\rho(x,\tau)+\nabla\cdot(V_1(\rho(x,\tau))\nabla\Phi(x,\tau))=V_2(\rho(x,\tau))\Phi(x,\tau),\\ \rho(0,x)=\rho^{n-1}(x), \quad \Phi(x,\Delta t)=-\frac{\delta}{\delta\rho}\mathcal{E}(\rho)(x), \end{cases}$$

where Φ satisfies the Hamilton-Jacobi-type equation when $\rho(x,\tau)>0$, such that

$$\partial_\tau\Phi(x,\tau)+\frac{1}{2}\|\nabla\Phi(x,\tau)\|^2V_1'(\rho(x,\tau))+|\Phi(x,\tau)|^2V_2'(\rho(x,\tau))=0.$$

We approximate the equation of $\rho(x,\tau)$ at $\tau=\Delta t$:

$$\begin{aligned}\rho(x,\Delta t)=&\rho(x,0)-\Delta t\Big[\nabla\cdot(V_1(\rho(x,\tau))\nabla\Phi(x,\tau))-V_2(\rho(x,\tau))\Phi(x,\tau)\Big]|_{\tau=\Delta t}+\mathcal{O}(\Delta t)\\ =&\rho(x,0)+\Delta t\Big[\nabla\cdot(V_1(\rho^n(x))\nabla\frac{\delta}{\delta\rho}\mathcal{E}(\rho^n)(x))-V_2(\rho^n(x))\frac{\delta}{\delta\rho}\mathcal{E}(\rho^n)(x)\Big]+\mathcal{O}(\Delta t),\end{aligned}$$

where we denote $\rho^n(x)=\rho(x,\Delta t)$. This finishes the proof. □

In fact, for first-order implicit time accuracy, one can use the one-step approximated minimization scheme. In other words, we only use a local time approximation of distance functional to compute the implicit time scheme.

Definition. *One-step relaxation of variational time implicit scheme.* Consider

$$\inf_{\rho,m,s}\ \underbrace{\frac{1}{2\Delta t}\int_\Omega\Big[\frac{\|m(x)\|^2}{V_1(\rho(x))}+\frac{|s(x)|^2}{V_2(\rho(x))}\Big]dx}_{\approx\frac{1}{2\Delta t}\mathrm{Dist}_{V_1,V_2}(\rho,\rho^{n-1})^2}+\mathcal{E}(\rho), \tag{2.7a}$$

where the minimization is over all functions $m\colon\Omega\to\mathbb{R}^d$, $s\colon\Omega\to\mathbb{R}$, and $\rho\colon\Omega\to\mathbb{R}_+$, such that

$$\rho(x)-\rho^{n-1}(x)+\nabla\cdot m(x)=s(x). \tag{2.7b}$$

Denote the next step solution ρ^n as the density minimizer of (2.7).

We also demonstrate that the variational scheme (2.7) forms a first-order implicit time scheme for the PDE (2.1).

Proof. The proof is similar to the one in (2.5). Denote $\Phi(x)$ as the Lagrange multiplier. The optimal condition of the variational problem (2.7) satisfies the following saddle point problem:

$$\inf_{\rho,m,s}\sup_{\Phi}\ \mathcal{L}(\rho,m,s,\Phi),$$

where

$$\begin{aligned}\mathcal{L}(\rho,m,s,\Phi):=&\frac{1}{2}\int_\Omega\Big[\frac{\|m(x)\|^2}{V_1(\rho(x))}+\frac{|s(x)|^2}{V_2(\rho(x))}\Big]dx+\Delta t\mathcal{E}(\rho)\\ &+\int_\Omega\Phi(x)\Big(\rho(x)-\rho^{n-1}(x)+\nabla\cdot m(x)-s(x)\Big)dx.\end{aligned}$$

By computing saddle point of the above system, we derive

$$\begin{cases} \frac{\delta}{\delta\rho}\mathcal{L}=0, \\ \frac{\delta}{\delta m}\mathcal{L}=0, \\ \frac{\delta}{\delta s}\mathcal{L}=0, \\ \frac{\delta}{\delta\Phi}\mathcal{L}=0, \end{cases} \Rightarrow \begin{cases} -\Big[\frac{\|m\|^2}{2V_1(\rho)^2}V_1'(\rho)+\frac{|s|^2}{2V_2(\rho)^2}V_2'(\rho)\Big]+\Delta t\frac{\delta}{\delta\rho}\mathcal{E}(\rho)+\Phi=0, \\ \frac{m}{V_1(\rho)}-\nabla\Phi=0, \\ \frac{s}{V_2(\rho)}-\Phi=0, \\ \rho-\rho^{n-1}+\nabla\cdot m=s. \end{cases}$$

One can check that $\Phi=-\Delta t\frac{\delta}{\delta\rho}\mathcal{E}(\rho)+o(\Delta t)$. Thus

$$\frac{\rho^n-\rho^{n-1}}{\Delta t}=\nabla\cdot(V_1(\rho^n)\nabla\frac{\delta}{\delta\rho}\mathcal{E}(\rho^n))-V_2(\rho^n)\frac{\delta}{\delta\rho}\mathcal{E}(\rho^n)+\mathcal{O}(\Delta t).$$

Here $\rho^n=\rho$ is the density minimizer. This finishes the proof. □

We remark that solving the variational problem (2.7) is simpler than optimizing (2.5), since (2.7) only involves a local time distance approximation; see [42,9]. We also present some properties of the implicit variational scheme (2.7). The algorithm satisfies the entropy dissipation property for any step size $\Delta t\geq 0$.

Proposition 2.1 *(Time implicit scheme entropy dissipation). Denote the solution $\{\rho^n\}_{n\in\mathbb{N}}$ solving the variational implicit scheme (2.7). For any stepsize $\Delta t\geq 0$, we have*

$$\mathcal{E}(\rho^n)\leq\mathcal{E}(\rho^{n-1}),\qquad \text{for } n\in\mathbb{N}_+.$$

Proof. Denote the objective functional (2.7a) as

$$\mathcal{F}(\rho,m,s)=\frac{1}{2\Delta t}\int_\Omega\Big[\frac{\|m(x)\|^2}{V_1(\rho(x))}+\frac{|s(x)|^2}{V_2(\rho(x))}\Big]dx+\mathcal{E}(\rho). \tag{2.8}$$

Since $(\rho^{n-1},m=0,s=0)$ is a feasible point satisfying the constraint (2.7b), and (ρ^n,m^*,s^*) is an optimal solution of (2.7), we have

$$\mathcal{E}(\rho^n)\leq\mathcal{F}(\rho^n,m^*,s^*)\leq\mathcal{F}(\rho^{n-1},0,0)=\mathcal{E}(\rho^{n-1}),$$

where we use the fact that

$$\mathcal{F}(\rho^n,m^*,s^*)=\mathcal{E}(\rho^n)+\frac{1}{2\Delta t}\int_\Omega\Big[\frac{\|m^*(x)\|^2}{V_1(\rho^n(x))}+\frac{|s^*(x)|^2}{V_2(\rho^n(x))}\Big]dx\geq\mathcal{E}(\rho^n).$$

We finish the proof. □

We also remark that there are issues of convexity in computing minimizers of the variational problem (2.7). If V_1 and V_2 are concave w.r.t. ρ, then the minimization problem (2.7) is always convex for any positive step size Δt. In general, this fact may be lost for general mobility functions V_1 and V_2. In computations, we still apply the first-order optimization algorithm to compute the variational problem (2.7), where we suggest a small stepsize Δt in the iterative update.

2.2. The abstract ALG2 algorithm

In this subsection, we formulate saddle point problems to calculate the variational time implicit schemes (2.7); see also [24,5].

We present the general form of the augmented Lagrangian (ALG2) algorithm [24] for the following saddle point system:

$$\inf_{\boldsymbol{u}}\sup_{\Phi}\quad F(\boldsymbol{u})-G(\Phi)-(\boldsymbol{u},\mathcal{D}\Phi)_\Omega, \tag{2.9}$$

where $\mathcal{D}(\Phi)$ is a linear differential operator for Φ, and $(\cdot,\cdot)_\Omega$ stands for the L^2-inner product on the domain Ω. For the problem (2.7), we choose

$$\boldsymbol{u}=(\rho,m,s),$$

with

$$F(\boldsymbol{u}) = \frac{1}{2}\int_{\Omega}\Big[\frac{\|m\|^2}{V_1(\rho)} + \frac{|s|^2}{V_2(\rho)}\Big]dx + \Delta t\mathcal{E}(\rho), \quad G(\Phi) = \int_{\Omega}\rho^{n-1}\Phi\, dx,$$

and

$$\mathcal{D}\Phi = (-\Phi, \nabla\Phi, \Phi).$$

The algorithm starts with the dual formulation of the saddle-point problem (2.9):

$$\sup_{\boldsymbol{u}}\inf_{\Phi,\boldsymbol{u}^*} \quad F^*(\boldsymbol{u}^*) + G(\Phi) + (\boldsymbol{u}, \mathcal{D}\Phi - \boldsymbol{u}^*)_{\Omega}, \tag{2.10}$$

where $F^*(\boldsymbol{u}^*) = \sup_{\boldsymbol{u}}(\boldsymbol{u}, \boldsymbol{u}^*)_{\Omega} - F(\boldsymbol{u})$ is the Legendre transform. The saddle point of the above system is equivalent to the saddle point of the following augmented Lagrangian form:

$$\sup_{\boldsymbol{u}}\inf_{\Phi,\boldsymbol{u}^*} \quad L_r(\Phi, \boldsymbol{u}, \boldsymbol{u}^*), \tag{2.11}$$

where the augmented Lagrangian

$$L_r(\Phi, \boldsymbol{u}, \boldsymbol{u}^*) := F^*(\boldsymbol{u}^*) + G(\Phi) + (\boldsymbol{u}, \mathcal{D}\Phi - \boldsymbol{u}^*)_{\Omega} + \frac{r}{2}(\mathcal{D}\Phi - \boldsymbol{u}^*, \mathcal{D}\Phi - \boldsymbol{u}^*)_{\Omega},$$

in which r is a positive parameter.

The ALG2 solves the optimization problem (2.11) in a splitting fashion. One iteration contains the following three steps.

Algorithm 1 One iteration of ALG2 algorithm for variational implicit scheme (2.11).

- Step A: update Φ. Minimize $L_r(\Phi, \boldsymbol{u}, \boldsymbol{u}^*)$ with respect to the first argument by solving the elliptic problem: Find Φ^{ℓ} such that it solves

  $$\inf_{\Phi} L_r(\Phi, \boldsymbol{u}^{\ell-1}, \boldsymbol{u}^{*,\ell-1}).$$

- Step B: update $\boldsymbol{u}^*$. Minimize $L_r(\Phi, \boldsymbol{u}, \boldsymbol{u}^*)$ with respect to the last argument by solving the nonlinear problem: Find $\boldsymbol{u}^{*,\ell}$ such that it solves

  $$\inf_{\boldsymbol{u}^*} L_r(\Phi^{\ell}, \boldsymbol{u}^{\ell-1}, \boldsymbol{u}^*).$$

- Step C: update $\boldsymbol{u}$. This is a simple pointwise update for the Lagrange multiplier $\boldsymbol{u}$:

  $$\boldsymbol{u}^{\ell} = \boldsymbol{u}^{\ell-1} + r(\mathcal{D}\Phi^{\ell} - \boldsymbol{u}^{*,\ell}). \tag{2.12}$$

We note that the key success of the ALG2 Algorithm 1 is that Step A is a simple linear reaction-diffusion equation solve, while the nonlinear Step B can be efficiently solved in a point-wise fashion, provided a good spatial discretization is used for the discretization variables; see Algorithm 2 below. We note that for the system case, further splitting in Step A/B for each component calculation will be applied to further save the computational cost; see Algorithm 3 below. We will present details of the implementation in Section 4 where the high-order spatial discretization is introduced. The error in the Lagrange multipliers in two consecutive iterations $\boldsymbol{u}^{\ell} - \boldsymbol{u}^{\ell-1}$ can be used to monitor the convergence of the ALG2 algorithm. Typically, a couple of hundred ALG iterations is sufficient for time accuracy. We take 200 ALG iterations in all our numerical results reported in Section 5. We note that after a spatial finite element discretization (see Section 4 below), the computational complexity of one ALG iteration is *linear* with respect to the total number of degrees of freedom. The proposed algorithm is highly parallelizable, as Step A can be solved using an optimal complexity multigrid solver, and Step B is a pointwise update that is embarrassingly parallelizable. The parallel implementation of this algorithm is our ongoing work.

3. Examples: Wasserstein gradient flow, reaction-diffusion equations, and reversible reaction-diffusion systems

This section presents examples of dissipative dynamic systems that fit in the framework of the previous section: Wasserstein gradient flows, scalar reaction-diffusion equations, and reversible reaction-diffusion systems.

3.1. Wasserestein gradient flow

We consider the following L^2-Wasserstein gradient flow for a time-dependent probability density $\rho : \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$ on a domain $\Omega \subset \mathbb{R}^d$,

$$\partial_t \rho = \nabla \cdot \left(\rho \nabla \frac{\delta}{\delta \rho} \mathcal{E}(\rho) \right), \tag{3.1}$$

subject to Neumann boundary conditions. Typically, the energy functional $\mathcal{E}(\rho)$ takes the following form

$$\mathcal{E}(\rho) := \int_\Omega \left[\alpha U_m(\rho(x)) + \rho(x) V(x) + \frac{1}{2}(W * \rho)(x)\rho(x) \right] dx, \tag{3.2}$$

where $\alpha \geq 0$ is the diffusion coefficient, $U_m(\rho)$ is the diffusion term with

$$U_m(\rho) = \begin{cases} \rho \log(\rho) & \text{if } m = 1, \\ \frac{\rho^m}{m-1} & \text{if } m > 1, \end{cases}$$

ρV is the drift term with drift potential V, and $\frac{1}{2}(W * \rho)\rho$ is the aggregation term with the convolution

$$(W * \rho)(x) := \int_\Omega W(x - y)\rho(y)\, dy,$$

in which $W(\cdot)$ is the symmetric interaction kernel. Its variational derivative is

$$\frac{\delta}{\delta \rho} \mathcal{E} = \alpha U_m'(\rho) + V + W * \rho. \tag{3.3}$$

The equation (3.1) is mass conserving, positivity preserving, and satisfies the energy dissipation law (2.2) with $V_1(\rho) = \rho$ and $V_2(\rho) = 0$.

This model is a special case of (2.1) with $V_1(\rho) = \rho$, $V_2(\rho) = 0$, and energy functional $\mathcal{E}$ in (3.2). The corresponding one-step variational time implicit scheme (2.7) is

$$\inf_{\rho, m} \quad \frac{1}{2\Delta t} \int_\Omega \frac{\|m(x)\|^2}{\rho(x)} dx + \mathcal{E}(\rho), \tag{3.4a}$$

where the minimization is over all functions $m \colon \Omega \to \mathbb{R}^d$, and $\rho \colon \Omega \to \mathbb{R}_+$, such that

$$\rho(x) - \rho^{n-1}(x) + \nabla \cdot m(x) = 0. \tag{3.4b}$$

The next step solution ρ^n is the density minimizer of (2.7), i.e., $\rho^n(x) = \rho(x)$. Here the first term in (3.4a) is the one-step relaxation approximation of the classical Wasserstein distance in Benamou-Brenier's dynamic formulation [5], i.e., the distance in (2.3) with $V_1(\rho) = \rho$ and $V_2(\rho) = 0$. We note that such approximation was originally used in [42,9].

This problem is equivalent to finding the saddle point of (2.9) in which $\boldsymbol{u} = (\rho, m)$,

$$F(\boldsymbol{u}) = \int_\Omega \frac{\|m\|^2}{2\rho} dx + \Delta t \mathcal{E}(\rho), \quad G(\Phi) = \int_\Omega \rho^{n-1} \Phi\, dx,$$

and $\mathcal{D}\Phi = (-\Phi, \nabla\Phi)$, which can be solved using ALG2 Algorithm 1 after a spatial discretization is used; see Section 4.

3.2. Dissipative reaction-diffusion equation

Adding a reaction term of form $-V_2(\rho)\frac{\delta}{\delta\rho}\mathcal{E}$ with a non-negative mobility function $V_2(\rho) \geq 0$ to the PDE (3.1), we get the following reaction-diffusion equation:

$$\partial_t \rho = \nabla \cdot \left(\rho \nabla \frac{\delta}{\delta \rho} \mathcal{E} \right) - V_2(\rho) \frac{\delta \mathcal{E}}{\delta \rho}, \tag{3.5}$$

which is again a special case of (2.1), with $V_1(\rho) = \rho$, and a general non-negative function $V_2(\rho)$. Hence, the corresponding one-step variational time implicit scheme (2.7) is

$$\inf_{\rho, m, s} \quad \frac{1}{2\Delta t} \int_\Omega \left[\frac{\|m(x)\|^2}{\rho(x)} + \frac{|s(x)|^2}{V_2(\rho(x))} \right] dx + \mathcal{E}(\rho), \tag{3.6a}$$

where the minimization is over all functions $m\colon \Omega \to \mathbb{R}^d$, $s\colon \Omega \to \mathbb{R}$, and $\rho\colon \Omega \to \mathbb{R}_+$, such that

$$\rho(x) - \rho^{n-1}(x) + \nabla \cdot m(x) = s(x). \tag{3.6b}$$

This is the saddle point of (2.9) in which $\boldsymbol{u} = (\rho, m, s)$,

$$F(\boldsymbol{u}) = \int_\Omega \Big[\frac{\|m\|^2}{2\rho} + \frac{|s|^2}{2V_2(\rho)}\Big]dx + \Delta t \mathcal{E}(\rho), \quad G(\Phi) = \int_\Omega \rho^{n-1}\Phi\, dx,$$

and $\mathcal{D}\Phi = (-\Phi, \nabla\Phi, \Phi)$.

We will postpone the introduction of a model with a more general $V_1(\rho) \neq \rho$ to Section 3.4.1, where a two-component reversible reaction-diffusion system with detailed balance is discussed.

Below we list three choices of $V_2(\rho)$ along with their corresponding energies that will be used in our numerical experiments:

(i) $V_2(\rho) = c\,\rho^\gamma$ where $c \geq 0$ and $\gamma \in \mathbb{R}$, with a general $\mathcal{E}(\rho)$ given in (3.2). Here $\gamma = 1$ corresponds to the Wasserstein-Fisher-Rao metrics used in [17,45], and $\gamma = 0$ is related to unnormalized optimal transport [36]. Both cases lead to a convex optimization problem (3.6) when the energy is convex; see Remark 4.3 below.

(ii) $V_2(\rho) = c\,\frac{\rho-1}{\log(\rho)}$ where $c \geq 0$ with a general $\mathcal{E}(\rho)$ given in (3.2). This choice also leads to a convex optimization problem for a convex energy.

(iii) $V_2(\rho) = \frac{\rho(\rho-1)}{\alpha\log(\rho)}$, with energy $\mathcal{E}(\rho) := \int_\Omega \alpha\rho(x)(\log(\rho) - 1)dx$, where $\alpha > 0$. This model is the following Fisher–KPP equation; see [39, Example 7]:

$$\frac{\partial \rho}{\partial t} - \nabla \cdot (\alpha \nabla \rho) = \rho(1-\rho). \tag{3.7}$$

It, however, does not lead to a convex optimization problem.

3.3. Strongly reversible reaction-diffusion systems

Our next model deals with the system of strongly reversible reaction-diffusion equations [55]. We consider M different chemical species $X_1, \ldots, X_M$ reacting according to R mass-action laws:

$$\alpha_1^p X_1 + \cdots + \alpha_M^p X_M \underset{k_-^p}{\overset{k_+^p}{\rightleftharpoons}} \beta_1^p X_1 + \cdots + \beta_M^p X_M, \tag{3.8}$$

where $p = 1, \cdots, R$ is the number of possible reactions, $\boldsymbol{\alpha}^p = (\alpha_1^p, \cdots, \alpha_M^p)$, $\boldsymbol{\beta}^p = (\beta_1^p, \cdots, \beta_M^p) \in \mathbb{N}_0^M$ are the vectors of the stoichiometric coefficients, and k_+^p, k_-^p are the positive forward and backward reaction rates. For simplicity, we restrict ourselves to the *strongly reversible* case where $k_+^p = k_-^p = k^p > 0$ in this subsection. The next subsection will discuss the more general case of reversible reaction-diffusion systems with the detailed balance that allows $k_+^p \neq k_-^p > 0$.

Combining the mass-action laws (3.8) with (independent) isotropic linear diffusion with energy $\mathcal{E}_i(\rho_i) = \int_\Omega \rho_i(\log(\rho_i) - 1)\,dx$ for each density ρ_i of species X_i, we get the following reaction-diffusion system:

$$\partial_t \rho_i - \nabla \cdot \left(\gamma_i \rho_i \nabla \frac{\delta}{\delta \rho}\mathcal{E}_i(\rho_i)\right) = -\sum_{p=1}^R k^p(\alpha_i^p - \beta_i^p)(\boldsymbol{\rho}^{\boldsymbol{\alpha}^p} - \boldsymbol{\rho}^{\boldsymbol{\beta}^p}), \tag{3.9}$$

for $1 \leq i \leq M$, where $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_M)$ and the multi-index notation $\boldsymbol{\rho}^{\boldsymbol{\alpha}^p} := \prod_{i=1}^M \rho_i^{\alpha_i^p}$ is used. Here the potential $\frac{\delta}{\delta\rho}\mathcal{E}_i(\rho_i) = \log(\rho_i)$ is simply the logarithm.

Next, we recast the above system (3.9) back to a system version of the general dissipative form (2.1) using appropriate mobility functions. We introduce the following function; see [55]:

$$\ell(x, y) = \begin{cases} \frac{x-y}{\log(x)-\log(y)} & \text{for } x \neq y, \\ y & \text{for } x = y, \end{cases} \tag{3.10}$$

and denote the following mobility functions:

$$V_{1,i}(\rho_i) = \gamma_i \rho_i, \quad \forall 1 \leq i \leq M, \tag{3.11a}$$

$$V_{2,p}(\boldsymbol{\rho}) = k^p\, \ell\left(\boldsymbol{\rho}^{\boldsymbol{\alpha}^p}, \boldsymbol{\rho}^{\boldsymbol{\beta}^p}\right), \quad \forall 1 \leq p \leq R. \tag{3.11b}$$

Using these notations, it can be shown that (3.9) is equivalent to

$$\begin{aligned}\partial_t \rho_i = &\nabla \cdot \left(V_{1,i}(\rho_i) \nabla \frac{\delta}{\delta \rho} \mathcal{E}_i(\rho_i) \right) \\ &- \sum_{p=1}^{R} V_{2,p}(\boldsymbol{\rho})(\alpha_i^p - \beta_i^p) \sum_{j=1}^{M} (\alpha_j^p - \beta_j^p) \frac{\delta}{\delta \rho} \mathcal{E}_j(\rho_j).\end{aligned} \tag{3.12}$$

It is now clear that the above system is purely dissipative as for the scalar case (2.1). That is, the first-time derivative of the total energy functional is nonpositive and satisfies

$$\begin{aligned}\frac{d}{dt} \sum_{i=1}^{M} \mathcal{E}_i(\rho_i(\cdot, t)) = &- \sum_{i=1}^{M} \int_{\Omega} \| \nabla \frac{\delta}{\delta \rho} \mathcal{E}_i(\rho_i)(x,t) \|^2 V_{1,i}(\rho_i)\, dx \\ &- \sum_{p=1}^{R} \int_{\Omega} \left| \sum_{j=1}^{M} (\alpha_j^p - \beta_j^p) \frac{\delta}{\delta \rho} \mathcal{E}_j(\rho_j) \right|^2 V_{2,p}(\boldsymbol{\rho})\, dx.\end{aligned} \tag{3.13}$$

As in the scalar case in Definition (2.3), we consider an optimal transport type distance:

$$\begin{aligned}\mathbf{Dist}_{V_1,V_2}(\boldsymbol{\rho}^0, \boldsymbol{\rho}^1)^2 = \inf_{\boldsymbol{\rho}, \boldsymbol{m}, \boldsymbol{s}} \Big\{ &\int_0^1 \int_{\Omega} \left(\sum_{i=1}^{M} \frac{|m_i|^2}{V_{1,i}(\rho_i)} + \sum_{p=1}^{R} \frac{|s_p|^2}{V_{2,p}(\boldsymbol{\rho})} \right) dx d\tau : \\ &\partial_\tau \rho_i + \nabla \cdot m_i = \textstyle\sum_{p=1}^{R} (\alpha_i^p - \beta_i^p) s_p, \forall 1 \le i \le M, \\ &\boldsymbol{\rho}(\cdot, 0) = \boldsymbol{\rho}^0, \quad \boldsymbol{\rho}(\cdot, 1) = \boldsymbol{\rho}^1. \Big\},\end{aligned}$$

where $\boldsymbol{m} = (m_1, \cdots, m_M)$ is the collection of fluxes, and $\boldsymbol{s} = (s_1, \cdots, s_R)$ is the collection of sources. Using this distance, the variational time implicit scheme is defined as follows (compare Definition (2.4) for the scalar case).

Definition. *Variational time implicit scheme for system* (3.12). Denote $\Delta t > 0$ as a time step size. Consider the scheme below:

$$\boldsymbol{\rho}^n = \arg \min_{\boldsymbol{\rho} \in [\mathcal{M}]^M} \quad \frac{1}{2\Delta t} \mathbf{Dist}_{V_1,V_2}(\boldsymbol{\rho}^{n-1}, \boldsymbol{\rho})^2 + \sum_{i=1}^{M} \mathcal{E}_i(\rho_i). \tag{3.14}$$

Its one-step relaxation is given as follows, which is the starting point of our spatial discretization to be discussed in the next section.

Definition. *One-step relaxation of variational time implicit schemes for system* (3.12). Consider

$$\inf_{\boldsymbol{\rho}, \boldsymbol{m}, \boldsymbol{s}} \quad \frac{1}{2\Delta t} \left(\sum_{i=1}^{M} \int_{\Omega} \frac{\|m_i\|^2}{V_{1,i}(\rho_i)} dx + \sum_{p=1}^{R} \int_{\Omega} \frac{\|s_p\|^2}{V_{2,p}(\boldsymbol{\rho})} dx \right) + \sum_{i=1}^{M} \mathcal{E}_i(\rho_i), \tag{3.15a}$$

where the minimization is over all functions $\boldsymbol{m}\colon \Omega \to [\mathbb{R}^d]^M$, $\boldsymbol{s}\colon \Omega \to [\mathbb{R}]^R$, and $\boldsymbol{\rho}\colon \Omega \to [\mathbb{R}_+]^M$, such that

$$\rho_i(x) - \rho_i^{n-1}(x) + \nabla \cdot m_i(x) = \sum_{p=1}^{R} (\alpha_i^p - \beta_i^p) s_p(x), \quad \forall 1 \le i \le M. \tag{3.15b}$$

The next step solution $\boldsymbol{\rho}^n$ is the density minimizer of (3.15). It is the saddle point of (2.9) in which

$$\begin{aligned}&\boldsymbol{u} = (\rho_1, \cdots, \rho_M, m_1, \cdots, m_M, s_1, \cdots, s_R), \Phi = (\Phi_1, \cdots, \Phi_M), \\ &F(\boldsymbol{u}) = \frac{1}{2\Delta t} \left(\sum_{i=1}^{M} \int_{\Omega} \frac{\|m_i\|^2}{V_{1,i}(\rho_i)} dx + \sum_{p=1}^{R} \int_{\Omega} \frac{\|s_p\|^2}{V_{2,p}(\boldsymbol{\rho})} dx \right), \quad G(\Phi) = \sum_{i=1}^{M} \int_{\Omega} \rho_i^{n-1} \Phi_i\, dx,\end{aligned}$$

and

$$\begin{aligned}\mathcal{D}\Phi = (&-\Phi_1, \cdots, -\Phi_M, \nabla\Phi_1, \cdots, \nabla\Phi_M, \\ &\sum_{i=1}^{M} (\alpha_i^1 - \beta_i^1)\Phi_i, \cdots, \sum_{i=1}^{M} (\alpha_i^R - \beta_i^R)\Phi_i).\end{aligned}$$

3.4. Reversible reaction-diffusion system with detailed balance

Note that the strongly reversible reaction-diffusion system (3.12) uses the same energy $\mathcal{E}_i(\rho_i) = \int_\Omega \rho_i(\log(\rho_i) - 1)\,dx$ for all species. By simply relaxing this requirement and rescaling the energy as

$$\mathcal{E}_i(\rho_i) = \int_\Omega \rho_i(\log(\kappa_i\rho_i) - 1)\,dx, \tag{3.16}$$

with $\kappa_i > 0$ being a positive constant to be determined by the reaction rates $k_\pm^p$, we will recover reversible reaction-diffusion systems with detailed balance; see [30,48–50]. For the above choice of energy, there holds

$$\frac{\delta}{\delta\rho}\mathcal{E}_i(\rho_i) = \log(\kappa_i\rho_i).$$

Below we give two specific examples that will be used in the numerical results section.

3.4.1. A two species model

We consider two species X_1, X_2 with a single reversible reaction

$$X_1 + 2X_2 \underset{k_-}{\overset{k_+}{\rightleftharpoons}} 3X_2,$$

with $k_-, k_+ > 0$. Denoting the following coefficients and mobility functions,

$$\kappa_1 = k_+, \quad \kappa_2 = k_-, \tag{3.17a}$$

$$V_{1,1}(\rho_1) = \gamma_1(\rho_1)^m, \quad V_{1,2}(\rho_2) = \gamma_2\rho_2, \tag{3.17b}$$

$$V_2(\rho_1, \rho_2) = \ell(\kappa_1\rho_1\rho_2^2, \kappa_2\rho_2^3), \tag{3.17c}$$

with $\ell(\cdot,\cdot)$ given in (3.10), $\gamma_1, \gamma_2 > 0$, $m \geq 1$, and using the energy (3.16), the system (3.12) written in component-wise notation is given as follows:

$$\partial_t\rho_1 = \nabla\cdot\left(V_{1,1}(\rho_1)\nabla\frac{\delta\mathcal{E}_1}{\delta\rho}(\rho_1)\right) - V_2(\rho_1,\rho_2)\left(\frac{\delta\mathcal{E}_1}{\delta\rho}(\rho_1) - \frac{\delta\mathcal{E}_2}{\delta\rho}(\rho_2)\right), \tag{3.18a}$$

$$\partial_t\rho_2 = \nabla\cdot\left(V_{1,2}(\rho_2)\nabla\frac{\delta\mathcal{E}_2}{\delta\rho}(\rho_2)\right) + V_2(\rho_1,\rho_2)\left(\frac{\delta\mathcal{E}_1}{\delta\rho}(\rho_1) - \frac{\delta\mathcal{E}_2}{\delta\rho}(\rho_2)\right). \tag{3.18b}$$

This is the following two-component reversible reaction-diffusion system studied in [48,49], which has potential applications in modeling tumor growth (see [53,62]):

$$\partial_t\rho_1 - \frac{\gamma_1}{m}\Delta\rho_1^m = -(k_+\rho_1\rho_2^2 - k_-\rho_2^3),$$
$$\partial_t\rho_2 - \gamma_2\Delta\rho_2 = (k_+\rho_1\rho_2^2 - k_-\rho_2^3).$$

3.4.2. A reversible four-component Gray-Scott model

Our final example is the reversible four-component Gray-Scott model originally proposed in [44] and numerically studied in [49]. We consider four species X_1, X_2, X_3, X_4 with three reversible reactions

$$X_1 + 2X_2 \underset{k_-^1}{\overset{k_+^1}{\rightleftharpoons}} 3X_2, \quad X_2 \underset{k_-^2}{\overset{k_+^2}{\rightleftharpoons}} X_3, \quad X_1 \underset{k_-^3}{\overset{k_+^3}{\rightleftharpoons}} X_4.$$

The reaction-diffusion system that combines these reactions with linear diffusion (with $M = 4, R = 3$) can be written into the form (3.12) by the following specific choices of κ-values, and mobility functions $V_{1,i}$ and $V_{2,p}$:

$$\kappa_1 = 1, \quad \kappa_2 = \frac{k_-^1}{k_+^1}, \quad \kappa_3 = \frac{k_-^1}{k_+^1}\frac{k_-^2}{k_+^2}, \quad \kappa_4 = \frac{k_-^3}{k_+^3}, \tag{3.19a}$$

$$V_{1,1}(\rho_1) = \gamma_1\rho_1, \quad V_{1,2}(\rho_2) = \gamma_2\rho_2, \quad V_{1,3}(\rho_3) = V_{1,4}(\rho_4) = 0, \tag{3.19b}$$

$$V_{2,1}(\boldsymbol{\rho}) = \ell(k_+^1\boldsymbol{\rho}^{\alpha^1}, k_-^1\boldsymbol{\rho}^{\beta^1}) = \frac{k_+^1\rho_1\rho_2^2 - k_-^1\rho_2^3}{\log(\kappa_1\rho_1) - \log(\kappa_2\rho_2)}, \tag{3.19c}$$

$$V_{2,2}(\boldsymbol{\rho}) = \ell(k_+^2\boldsymbol{\rho}^{\alpha^2}, k_-^2\boldsymbol{\rho}^{\beta^2}) = \frac{k_+^2\rho_2 - k_-^2\rho_3}{\log(\kappa_2\rho_2) - \log(\kappa_3\rho_3)}, \tag{3.19d}$$

$$V_{2,3}(\boldsymbol{\rho}) = \ell(k_+^2\boldsymbol{\rho}^{\alpha^3}, k_-^3\boldsymbol{\rho}^{\beta^3}) = \frac{k_+^3\rho_1 - k_-^3\rho_4}{\log(\kappa_1\rho_1) - \log(\kappa_4\rho_4)}. \tag{3.19e}$$

For completeness, we write down the PDE system (3.12) with the above choice of parameters using a standard component-wise notation in the following:

$$\partial_t\rho_1 = \gamma_1\Delta\rho_1 - (k_+^1\rho_1\rho_2^2 - k_-^1\rho_2^3) - (k_+^3\rho_1 - k_-^3\rho_4), \tag{3.20a}$$
$$\partial_t\rho_2 = \gamma_2\Delta\rho_2 + (k_+^1\rho_1\rho_2^2 - k_-^1\rho_2^3) - (k_+^2\rho_2 - k_-^2\rho_3), \tag{3.20b}$$
$$\partial_t\rho_3 = (k_+^2\rho_2 - k_-^2\rho_3), \tag{3.20c}$$
$$\partial_t\rho_4 = (k_+^3\rho_1 - k_-^3\rho_4). \tag{3.20d}$$

This is the reversible Gray-Scott model proposed in [44] to approximate the following two-component irreversible Gray-Scott model [32]:

$$\partial_t\rho_1 = \gamma_1\Delta\rho_1 - k_+^1\rho_1\rho_2^2 - k_+^3(\rho_1 - 1), \tag{3.21a}$$
$$\partial_t\rho_2 = \gamma_2\Delta\rho_2 + k_+^1\rho_1\rho_2^2 - k_+^2\rho_2, \tag{3.21b}$$

which can form spatially complex patterns [59], and is widely used to study pattern formations. We comment that by requiring

$$\kappa_-^1\rho_2^3 \approx 0, \quad k_-^3\rho_4 \approx k_+^3, \quad \text{and} \quad \kappa_-^2\rho_3 \approx 0, \tag{3.22}$$

the reversible Gray-Scott model (3.20) formally converges to the irreversible Gray-Scott model (3.21). We refer interested readers to [44] for a theoretical study. Formally, the conditions (3.22) can be achieved by taking very small backward reaction rates $\kappa_-^1, \kappa_-^2, \kappa_-^3 \ll 1$, and using initial value for ρ_4 such that $\rho_4 = \frac{\kappa_+^3}{\kappa_-^3} \gg 1$. As a side note, we mention that spatially complex patterns were not observed in the numerical results [49, Example 4.3], which uses a second-order operator splitting scheme via an energetic variational formulation. We found that the reason for no pattern formation in the test case in [49] was due to inappropriate choices of a too large backward reaction rate k_-^3 and the initial condition. With more careful choices of diffusion coefficients, reaction rates, and initial conditions, we numerically observe complex pattern formations in both 1D and 2D reversible Gray-Scott models; see our simulation results in Section 5.7.

4. High-order spatial discretization for generalized time implicit schemes

In this section, we first apply high-order spatial discretization to the time implicit schemes (3.4), (3.6) and their system version (3.15), and then discuss the practical implementation of each step of the ALG2 Algorithm 1. We restrict ourselves to the two-dimensional setting with a rectangular domain Ω, which is triangulated using a uniform rectangular mesh $\mathcal{T}_h = \{T\}$. While our method can work on general unstructured triangular meshes, see [25], the restriction to uniform rectangular meshes has a huge advantage in computing the convolution term in the energy (3.2), where the Fast Fourier transform can be applied.

4.1. The finite element spaces and notation

The spatial discretization is adopted from our previous work on high-order schemes for optimal transport and mean field games [25]. Specifically, the high-order H^1-conforming finite element space

$$V_h^k := \{v \in H^1(\Omega) : \ v|_T \in \mathcal{Q}^k(T) \ \forall T \in \mathcal{T}_h\}, \tag{4.1}$$

is used to approximate the Φ variable, and the high-order L^2-conforming discontinuous polynomial space,

$$W_h^k := \{w \in L^2(\Omega) : \ w|_T \in \mathcal{Q}^k(T) \ \forall T \in \mathcal{T}_h\}, \tag{4.2}$$

is used to approximate the other variables where derivative information is not needed. Here $\mathcal{Q}^k(T)$ is the space of tensor-product polynomial spaces of degree no greater than $k \geq 1$ in each direction. We equip the space W_h^k with a set of *nodal basis* $\{\varphi_i\}_{i=1}^{N_W} \subset W_h^k$ that satisfies

$$\varphi_i(\xi_j) = \delta_{ij}, \quad \forall 1 \leq j \leq N_W, \tag{4.3}$$

where N_W is the dimension of the space W_h^k, δ_{ij} is the Kronecker delta function, and $\{\xi_i\}_{i=i}^{N_W}$ is the collection of N_W Gauss-Legendre integration points with corresponding weights $\{\omega_i\}_{i=1}^{N_W}$ on the mesh $\mathcal{T}_h$. For the current work, only evaluation on quadrature points for functions in W_h^k is needed in the algorithm, not their derivatives. Hence, given a function $u_h \in W_h^k$

expressed as $u_h = \sum_{i=1}^{N_W} \mathrm{u}_i \varphi_i(x)$, we simply need to store and update its coefficient vector $[\mathrm{u}_i, \cdots, \mathrm{u}_{N_W}]^T$, which makes its practical implementation extremely simple. Moreover, we denote the discrete $L^2(\Omega)$-inner product $(\cdot,\cdot)_h$ as

$$(u, v)_h := \sum_{i=1}^{N_W} u(\xi_i) v(\xi_i) \omega_i, \tag{4.4}$$

we have $(u_h, v_h)_h = \sum_{i=1}^{N_W} \mathrm{u}_i \mathrm{v}_i \omega_i$, for any function $u_h = \sum_{i=1}^{N_W} \mathrm{u}_i \varphi_i(x) \in W_h^k$ and $v_h = \sum_{i=1}^{N_W} \mathrm{v}_i \varphi_i(x) \in W_h^k$.

4.2. High-order FEM for the reaction diffusion equation

Since the variation time implicit scheme for the Wasserstein gradient flow problem (3.4) is a special case of the reaction-diffusion problem (3.6) with no reaction $V_2(\rho) = 0$, we only present the high-order spatial discretization for (3.6). We first write the discrete saddle point problem in its augmented Lagrangian form (2.11): given mesh $\mathcal{T}_h$, polynomial degree $k \geq 1$, time step size $\Delta t > 0$ and density approximation ρ_h^{old} at the previous time step, find $\boldsymbol{u}_h, \boldsymbol{u}_h^* \in [W_h^k]^4$, and $\Phi_h \in V_h^k$, such that

$$\inf_{\boldsymbol{u} \in [W_h^k]^4} \sup_{\Phi_h \in V_h^k, \boldsymbol{u}_h^* \in [W_h^k]^4} L_{r,h}(\Phi_h, \boldsymbol{u}_h, \boldsymbol{u}_h^*), \tag{4.5}$$

where $\boldsymbol{u}_h = (\rho_h, m_h^0, m_h^1, s_h)$ is the collection of density ρ_h, (two-dimensional) flux $m_h = (m_h^0, m_h^1)$, and source term s_h, $\boldsymbol{u}_h^* = (\rho_h^*, m_h^{0,*}, m_h^{1,*}, s_h^*)$ is its dual, and the discrete augmented Lagrangian is

$$\begin{aligned} L_{r,h}(\Phi_h, \boldsymbol{u}_h, \boldsymbol{u}_h^*) := {} & F_h^*(\boldsymbol{u}_h^*) + G_h(\Phi_h) + (\boldsymbol{u}_h, \mathcal{D}\Phi_h - \boldsymbol{u}_h^*)_h \\ & + \frac{r}{2} (\mathcal{D}\Phi_h - \boldsymbol{u}_h^*, \mathcal{D}\Phi_h - \boldsymbol{u}_h^*)_h. \end{aligned} \tag{4.6}$$

Here $(\cdot,\cdot)_h$ is the volume integration rule given in (4.4), the operators

$$\mathcal{D}\Phi_h := (-\Phi_h, \partial_{x_0}\Phi_h, \partial_{x_1}\Phi_h, \Phi_h), \tag{4.7}$$

$$G_h(\Phi_h) := (\rho_h^{\mathrm{old}}, \Phi_h)_h, \tag{4.8}$$

$$F_h^*(\boldsymbol{u}_h^*) := \sup_{\boldsymbol{u}_h \in [W_h^k]^4} (\boldsymbol{u}_h^*, \boldsymbol{u}_h)_h - F_h(\boldsymbol{u}_h), \tag{4.9}$$

where $(\partial_{x_0}, \partial_{x_1}) = \nabla$ is the gradient, and F_h is given as

$$F_h(\boldsymbol{u}_h) := \left(\frac{|m_h^0|^2 + |m_h^1|^2}{2\rho_h} + \frac{s_h^2}{2V_2(\rho_h)}, 1 \right)_h + \Delta t\, \mathcal{E}_h(\rho_h), \tag{4.10}$$

in which the discrete total energy

$$\mathcal{E}_h(\rho_h) := (\alpha U_m(\rho_h) + \rho_h V(x), 1)_h + \frac{1}{2} (W * \rho_h, \rho_h)_h \tag{4.11}$$

for energy of the form (3.2). We note that when the interaction kernel $W(x)$ is smooth, the convolution term $W * \rho_h$ in the above expression can be simply evaluated using the same integration rule (4.4). On the other hand, for singular kernels with $W(0) = \pm\infty$, we shall use alternative integration rules to avoid the evaluation of $W(0)$ when evaluating this convolution term.

Note that a similar formulation can be used for the more general case (2.7) for the equation (2.1) where the denominator in the first term in (4.10) is replaced by a general mobility function $V_1(\rho_h)$.

Remark 4.1 *(On polynomial degree for Φ_h and $\boldsymbol{u}_h$).* We note that in our previous work [25], the polynomial degree for the discontinuous functions $\boldsymbol{u}_h$ associated with the integration rule space W_h^k is taken to be one order lower than that for the continuous function Φ_h. Here our numerical experiments suggest that increasing the integration rule space order to be the same as the continuous space V_h^k leads to a more accurate result. Hence we use equal order approximations for all our numerical results.

We next provide a practical implementation of each step of the ALG2 Algorithm 1 for solving the saddle point problem.

4.2.1. Step A: scalar case

Taking infinium of $L_{r,h}$ with respect to Φ_h, we arrive at a constant coefficient reaction-diffusion equation: find $\Phi_h^\ell \in V_h^k$ such that

$$(\mathcal{D}\Phi_h^\ell, \mathcal{D}\Psi_h)_h = (\boldsymbol{u}_h^{*,\ell-1} - \frac{1}{r}\boldsymbol{u}_h^{\ell-1}, \mathcal{D}\Psi_h)_h - \frac{1}{r}(\rho_h^{\text{old}}, \Psi_h)_h, \quad \forall \Psi_h \in V_h^k. \tag{4.12}$$

Using the definition in (4.7), we write the above equation using physical variables:

$$\begin{aligned} 2(\Phi_h^\ell, \Psi_h)_h + (\nabla\Phi_h^\ell, \nabla\Psi_h)_h = (s_h^{*,\ell-1} - \rho_h^{*,\ell-1} + \frac{\rho_h^{\ell-1} - s_h^{\ell-1} - \rho_h^{\text{old}}}{r}, \Psi_h)_h \\ + (\boldsymbol{m}_h^{*,\ell-1} - \frac{\boldsymbol{m}_h^{\ell-1}}{r}, \nabla\Psi_h)_h. \end{aligned}$$

This symmetric positive definite linear system can be efficiently solved using, e.g., a multigrid algorithm [8,71].

4.2.2. Step B/C: scalar case

The next step is to take infinium of $L_{r,h}$ with respect to $\boldsymbol{u}_h^*$. Find $\boldsymbol{u}_h^{*,\ell} \in [W_h^k]^4$, such that it solves

$$\underset{\boldsymbol{u}_h^* \in [W_h^k]^4}{\text{argmin}} \quad F_h^*(\boldsymbol{u}_h^*) - (\boldsymbol{u}_h^{\ell-1}, \boldsymbol{u}_h^*)_h + \frac{r}{2}(\mathcal{D}\Phi_h^\ell - \boldsymbol{u}_h^*, \mathcal{D}\Phi_h^\ell - \boldsymbol{u}_h^*)_h.$$

Without loss of generality, we abuse the notation and denote $\mathcal{D}\Phi_h^\ell$ as its interpolation onto the space $[W_h^k]^4$. We further denote

$$\overline{\boldsymbol{u}}_h := \mathcal{D}\Phi_h^\ell + \frac{1}{r}\boldsymbol{u}_h^{\ell-1} \in [W_h^k]^4. \tag{4.13}$$

Then the above minimization problem is equivalent to

$$\underset{\boldsymbol{u}_h^* \in [W_h^k]^4}{\text{argmin}} \quad F_h^*(\boldsymbol{u}_h^*) + \frac{r}{2}(\boldsymbol{u}_h^* - \overline{\boldsymbol{u}}_h, \boldsymbol{u}_h^* - \overline{\boldsymbol{u}}_h)_h. \tag{4.14}$$

After this minimizer is computed, the last step is to update the Lagrangian multiplier $\boldsymbol{u}_h^\ell$ according to (2.12):

$$\boldsymbol{u}_h^\ell = \boldsymbol{u}_h^{\ell-1} + r(\mathcal{D}\Phi_h^\ell - \boldsymbol{u}_h^{*,\ell}) = r(\overline{\boldsymbol{u}}_h - \boldsymbol{u}_h^{*,\ell}) \in [W_h^k]^4, \tag{4.15}$$

where we used the definition (4.13) in the last step.

Due to the complicated form of the energy (3.2), it might be challenging to compute an explicit expression of the convex conjugate $F_h^*(\boldsymbol{u}_h^*)$. Here we present a practical way to solve the minimization problem (4.14) without explicitly computing this convex conjugate using duality. The main idea is presented in the next result.

Proposition 4.1. *Let $\boldsymbol{u}_h^{*,\ell} \in [W_h^k]^4$ be the minimizer to the problem (4.14), and let $\boldsymbol{u}_h^\ell$ be given according to (4.15). Then, $\boldsymbol{u}_h^\ell$ is the minimizer to the following problem*

$$\boldsymbol{u}_h^\ell = \underset{\boldsymbol{u}_h \in [W_h^k]^4}{\text{argmin}} \quad F_h(\boldsymbol{u}_h) + \frac{1}{2r}(\boldsymbol{u}_h - r\overline{\boldsymbol{u}}_h, \boldsymbol{u}_h - r\overline{\boldsymbol{u}}_h)_h, \tag{4.16}$$

which we refer to as the dual problem of (4.14). Furthermore, there holds

$$\boldsymbol{u}_h^{*,\ell} = \overline{\boldsymbol{u}}_h - \boldsymbol{u}_h^\ell/r. \tag{4.17}$$

Proof. The equation (4.17) is a simple rewriting of (4.15). Let us now prove (4.16). By definition (4.9), we have $\boldsymbol{u}_h^{*,\ell}$ is part of the saddle point solution

$$\inf_{\boldsymbol{u}_h^* \in [W_h^k]^4} \sup_{\boldsymbol{u}_h \in [W_h^k]^4} (\boldsymbol{u}_h, \boldsymbol{u}_h^*)_h - F_h(\boldsymbol{u}_h) + \frac{r}{2}(\boldsymbol{u}_h^* - \overline{\boldsymbol{u}}_h, \boldsymbol{u}_h^* - \overline{\boldsymbol{u}}_h)_h. \tag{4.18}$$

Taking the derivative with respect to $\boldsymbol{u}_h^*$ in the above expression, we get

$$\boldsymbol{u}_h^* = \overline{\boldsymbol{u}}_h - \boldsymbol{u}_h/r.$$

Plugging this expression back to (4.18), we easily see that the primal variable $\boldsymbol{u}_h$ is the minimizer to the dual problem (4.16). By (4.15), it is clear that this optimizer is nothing but the solution $\boldsymbol{u}_h^\ell$. This completes the proof. □

Proposition 4.1 suggests to first solve for the primal variable $\boldsymbol{u}_h^\ell$ using the minimization problem (4.16), then update $\boldsymbol{u}_h^{*,\ell}$ using (4.17), which is the approach we adopt in our implementation. It is in general more convenient than the (equivalent) original ALG2 algorithm that first solve for the dual variable $\boldsymbol{u}_h^{*,\ell}$ using (4.14) then update $\boldsymbol{u}_h^\ell$ using (4.15), which requires the computation of the dual functional (4.9).

Next, using the particular form of F_h in (4.10), we show that the minimization problem (4.16) can be efficiently solved by first locally expressing flux m_h^0, m_h^1 and source s_h in terms of density ρ_h and then solving a nonlinear optimization problem for ρ_h alone. We record this procedure in the following result.

Proposition 4.2. *Let $\boldsymbol{u}_h^\ell$ be the solution to (4.16). Then there holds*

$$m_h^{0,\ell} = \frac{r\rho_h^\ell}{r+\rho_h^\ell}\overline{m}_h^0, \quad m_h^{1,\ell} = \frac{r\rho_h^\ell}{r+\rho_h^\ell}\overline{m}_h^1, \quad s_h^\ell = \frac{rV_2(\rho_h^\ell)}{r+V_2(\rho_h^\ell)}\overline{s}_h, \tag{4.19}$$

where

$$\overline{\boldsymbol{u}}_h = (\overline{\rho}_h, \overline{m}_h^0, \overline{m}_h^1, \overline{s}_h),$$

and ρ_h^ℓ is the minimizer to the following reduced problem:

$$\begin{aligned}\operatorname*{argmin}_{\rho_h \in W_h^k} \frac{1}{2r}\left(|\rho_h - r\overline{\rho}_h|^2, 1\right)_h + \left(\frac{r^2(|\overline{m}_h^0|^2 + |\overline{m}_h^1|^2)}{2(r+\rho_h)}, 1\right)_h \\ + \left(\frac{r^2\,|\overline{s}_h|^2}{2(r+V_2(\rho_h))}, 1\right)_h + \Delta t\,\mathcal{E}_h(\rho_h).\end{aligned} \tag{4.20}$$

Proof. The derivatives of the functional in (4.16) at the saddle point vanish. Taking derivatives with respect to m_h^0, m_h^1 and s_h, we get the relations (4.19). Plugging these relations back to (4.16) and simplifying, we get the optimization problem (4.20) for ρ_h^ℓ. □

Remark 4.2 *(On pointwise update for (4.20)).* The problem (4.20) can be solved by computing its critical point. Taking the variation of the function in (4.20) with respect to ρ_h, we have

$$\frac{1}{r}(\rho_h - r\overline{\rho}_h) - \frac{r^2(|\overline{m}_h^0|^2 + |\overline{m}_h^1|^2)}{2(r+\rho_h)^2} - \frac{r^2V_2'(\rho_h)\overline{s}_h^2}{2(r+V_2(\rho_h))^2} + \Delta t\,\frac{\delta\mathcal{E}_h}{\delta\rho}(\rho_h) = 0. \tag{4.21}$$

By the choice of the function space (4.2), it is clear that (4.21) is satisfied on all quadrature points ξ_i for $1 \le i \le N_W$. Using definition of the energy (4.11), we have

$$\frac{\delta\mathcal{E}_h}{\delta\rho}(\rho_h) = \alpha U_m'(\rho_h) + V(x) + W * \rho_h.$$

In the absence of interaction kernel where $W(x) = 0$, the equation (4.21) can be solved in a pointwise fashion per quadrature point thanks to the particular choice of the nodal basis (4.3) for the space (4.2), using, e.g., Newton's method.

On the other hand, when aggregation effects are included, the term $W * \rho_h$ prohibits such pointwise update due to the nonlocal effect of this convolution. In this case, we treat the convolution term $W * \rho_h$ explicitly in (4.21) by evaluating it at the previous time step, i.e.,

$$W * \rho_h \approx W * \rho_h^{\text{old}},$$

and then solve the modified pointwise local problem (4.21) using the Newton's method. This is the choice we use in all our simulation results with aggregation effects. Similar treatment was used in, e.g., [10,7].

Remark 4.3 *(On convexity).* Let us briefly comment on convexity of the problem (4.20). When aggregation effects are included, we extrapolate the nonlocal convolution term according to Remark 4.2. The problem (4.20) is a pointwise minimization problem per quadrature point. Taking its second-order variation, we obtain

$$\frac{1}{r} + \frac{r^2(|\overline{m}_h^0|^2 + |\overline{m}_h^1|^2)}{(r+\rho_h)^3} + \frac{r^2|\overline{s}_h|^2\left(2V_2'(\rho_h)^2 - (r+V_2(\rho_h))V_2''(\rho_h)\right)}{2(r+V_2(\rho_h))^3} + \alpha\Delta t U_m''(\rho_h). \tag{4.22}$$

It is clear that the first, second, and last terms of the above expression are always nonnegative as long as $\rho_h \ge 0$. Moreover, if

$$2V_2'(\rho_h)^2 - (r + V_2(\rho_h))V_2''(\rho_h) \geq 0, \tag{4.23}$$

then the third term is also nonnegative. For such a choice of mobility V_2, the minimization problem is convex, and the uniqueness of the solution is guaranteed unconditionally for any time step size Δt. In the absence of aggregation effects, the overall ALG2 algorithm with V_2 satisfying (4.23) can also be shown to be unconditionally convergent; see, e.g., [21].

We note that the convexity condition (4.23) is ensured if we take $V_2(\rho) = c\rho^\gamma$ for $c > 0$ and $0 \leq \gamma \leq 1$, or $V_2(\rho) = \frac{\rho - \bar{\rho}}{\log(\rho) - \log(\bar{\rho})}$ for any $\bar{\rho} > 0$. The latter choice will be used in the system case. On the other hand, the mobility $V_2(\rho) = \frac{\rho(1-\rho)}{\log(\rho)}$ for the Fisher-KPP equation (3.7) does not satisfy the convexity condition (4.23). For this case, we may use a small time step size Δt to get a stable simulation.

We finally note that small time step size Δt may also be needed for the general case with an interaction potential W, where extrapolation is used to approximate the problem (4.20) as mentioned in Remark 4.2.

For completeness, we collect one iteration of this algorithm as follows.

Algorithm 2 One iteration of ALG2 algorithm for (4.5).

- Step A: update Φ_h^ℓ. Find $\Phi_h^\ell \in V_h^k$ such that the equation (4.12) holds.
- Step B/C: update $\boldsymbol{u}_h^\ell, \boldsymbol{u}_h^{*,\ell}$. First, find ρ_h^ℓ such that it is the minimizer to (4.20). Then update $m_h^{0,\ell}, m_h^{1,\ell}, s_h^\ell$ according to (4.19). Finally, update $\boldsymbol{u}_h^{*,\ell}$ according to (4.17).

We note that the positivity of density approximation ρ_h can be easily enforced in the pointwise optimization problem (4.20).

4.3. High-order FEM for strongly reversible reaction diffusion systems

We now present the high-order FEM discretization of the variational time implicit scheme (3.15) and discuss its practical (modified) ALG2 implementation. Given time step size $\Delta t > 0$ and density approximations

$$\boldsymbol{\rho}_h^{\text{old}} = (\rho_{1,h}^{\text{old}}, \cdots, \rho_{M,h}^{\text{old}}) \in [W_h^k]^M$$

at the previous time step, find $\boldsymbol{u}_h, \boldsymbol{u}_h^* \in [W_h^k]^{3M+R}$, and $\boldsymbol{\Phi}_h \in [V_h^k]^M$, such that

$$\inf_{\boldsymbol{u}_h \in [W_h^k]^{3M+R}} \sup_{\boldsymbol{\Phi}_h \in [V_h^k]^M, \boldsymbol{u}_h^* \in [W_h^k]^{3M+R}} \underline{L_{r,h}}(\boldsymbol{\Phi}_h, \boldsymbol{u}_h, \boldsymbol{u}_h^*), \tag{4.24}$$

where

$$\boldsymbol{u}_h = (\rho_{1,h}, m_{1,h}^0, m_{1,h}^1, \cdots, \rho_{M,h}, m_{M,h}^0, m_{M,h}^1, s_{1,h}, \cdots, s_{R,h})$$

is the collection of densities $\boldsymbol{\rho}_h$, fluxes

$$\boldsymbol{m}_h = (m_{1,h}^0, m_{1,h}^1, \cdots, m_{M,h}^0, m_{M,h}^1),$$

and source terms $\boldsymbol{s}_h = (s_{1,h}, \cdots, s_{R,h})$, $\boldsymbol{u}_h^*$ is its dual, $\boldsymbol{\Phi}_h = (\Phi_{1,h}, \ldots, \Phi_{M,h})$, and the discrete augmented Lagrangian is

$$\begin{aligned} \underline{L_{r,h}}(\boldsymbol{\Phi}_h, \boldsymbol{u}_h, \boldsymbol{u}_h^*) := {} & \underline{F_h}^*(\boldsymbol{u}_h^*) + \underline{G_h}(\boldsymbol{\Phi}_h) + (\boldsymbol{u}_h, \underline{\mathcal{D}}\boldsymbol{\Phi}_h - \boldsymbol{u}_h^*)_h \\ & + \frac{r}{2}(\underline{\mathcal{D}}\boldsymbol{\Phi}_h - \boldsymbol{u}_h^*, \underline{\mathcal{D}}\boldsymbol{\Phi}_h - \boldsymbol{u}_h^*)_h. \end{aligned} \tag{4.25}$$

Here the operators

$$\begin{aligned} \underline{\mathcal{D}}\boldsymbol{\Phi}_h := \Big(& -\Phi_{1,h}, \partial_{x_0}\Phi_{1,h}, \partial_{x_1}\Phi_{1,h}, \cdots, -\Phi_{1,h}, \partial_{x_0}\Phi_{1,h}, \partial_{x_1}\Phi_{1,h}, \\ & \sum_{i=1}^M (\alpha_i^1 - \beta_i^1)\Phi_{i,h}, \cdots, \sum_{i=1}^M (\alpha_i^R - \beta_i^R)\Phi_{i,h} \Big), \end{aligned} \tag{4.26}$$

$$\underline{G_h}(\boldsymbol{\Phi}_h) := \sum_{i=1}^M (\rho_{i,h}^{\text{old}}, \Phi_{i,h})_h, \tag{4.27}$$

$$\underline{F_h}^*(\boldsymbol{u}_h^*) := \sup_{\boldsymbol{u}_h \in [W_h^k]^{3M+R}} (\boldsymbol{u}_h^*, \boldsymbol{u}_h)_h - \underline{F_h}(\boldsymbol{u}_h), \tag{4.28}$$

and $\underline{F_h}$ is given as

$$\underline{F_h}(\boldsymbol{u}_h) := \left(\sum_{i=1}^{M} \frac{|m_{i,h}^0|^2 + |m_{i,h}^1|^2}{2V_{1,i}(\rho_{i,h})} + \sum_{p=1}^{R} \frac{|s_{p,h}|^2}{2V_{2,p}(\boldsymbol{\rho}_h)}, 1 \right)_h + \Delta t \sum_{i=1}^{M} \mathcal{E}_{i,h}(\rho_{i,h}), \tag{4.29}$$

where the mobility functions are given in (3.11) and the discrete energy

$$\mathcal{E}_{i,h}(\rho_{i,h}) = (\rho_{i,h}(\log(\rho_{i,h}) - 1), 1)_h.$$

We now discuss a modified implementation of the ALG2 Algorithm 1 for the saddle point system (4.24), where further componentwise splitting is introduced to drive down the overall computational cost.

4.3.1. Step A: system case

Taking infinium of $\underline{L_{r,h}}$ with respect to $\boldsymbol{\Phi}_h$, we arrive at a coupled system of constant coefficient reaction-diffusion equations: find $\boldsymbol{\Phi}_h^\ell \in [V_h^k]^M$ such that

$$(\underline{\mathcal{D}}\boldsymbol{\Phi}_h^\ell, \underline{\mathcal{D}}\boldsymbol{\Psi}_h)_h = (\boldsymbol{u}_h^{*,\ell-1} - \frac{1}{r}\boldsymbol{u}_h^{\ell-1}, \underline{\mathcal{D}}\boldsymbol{\Psi}_h)_h - \frac{1}{r}(\boldsymbol{\rho}_h^{\text{old}}, \boldsymbol{\Psi}_h)_h, \tag{4.30}$$

for all $\boldsymbol{\Psi}_h \in [V_h^k]^M$. Using the definition in (4.26), we write the above system back using the physical variables:

$$\begin{aligned}
&(\Phi_{i,h}^\ell, \Psi_{i,h})_h + (\nabla\Phi_{i,h}^\ell, \nabla\Psi_{i,h})_h + \sum_{p=1}^{R}\sum_{j=1}^{M} \left((\alpha_j^p - \beta_j^p)\Phi_{j,h}^\ell, (\alpha_i^p - \beta_i^p)\Psi_{i,h} \right)_h \\
&= (-\rho_{i,h}^{*,\ell-1} + \frac{\rho_{i,h}^{\ell-1} - \rho_{i,h}^{\text{old}}}{r}, \Psi_{i,h})_h + (\boldsymbol{m}_{i,h}^{*,\ell-1} - \frac{\boldsymbol{m}_{i,h}^{\ell-1}}{r}, \nabla\Psi_{i,h})_h \\
&\quad + \sum_{p=1}^{R} \left(s_{r,h}^{*,\ell-1} - \frac{1}{r}s_{r,h}^{\ell-1}, (\alpha_i^p - \beta_i^p)\Psi_{i,h} \right)_h,
\end{aligned}$$

for all $1 \le i \le M$. This coupled linear system might be expensive to solve. Here we propose to solve these M equations in parallel by treating the coupling term on the left hand side of the above equation explicitly. Specifically, for each $1 \le i \le M$, we compute $\Phi_{i,h} \in V_h^k$ such that it solves the following scalar linear reaction-diffusion equation:

$$\begin{aligned}
&(\Phi_{i,h}^\ell, \Psi_{i,h})_h + (\nabla\Phi_{i,h}^\ell, \nabla\Psi_{i,h})_h + \sum_{p=1}^{R} \left((\alpha_i^p - \beta_i^p)\Phi_{i,h}^\ell, (\alpha_i^p - \beta_i^p)\Psi_{i,h} \right)_h \\
&= (-\rho_{i,h}^{*,\ell-1} + \frac{\rho_{i,h}^{\ell-1} - \rho_{i,h}^{\text{old}}}{r}, \Psi_{i,h})_h + (\boldsymbol{m}_{i,h}^{*,\ell-1} - \frac{\boldsymbol{m}_{i,h}^{\ell-1}}{r}, \nabla\Psi_{i,h})_h \\
&\quad + \sum_{p=1}^{R} \left(s_{r,h}^{*,\ell-1} - \frac{1}{r}s_{r,h}^{\ell-1}, (\alpha_i^p - \beta_i^p)\Psi_{i,h} \right)_h \\
&\quad - \sum_{p=1}^{R}\sum_{\substack{j=1 \\ j\neq i}}^{M} \left((\alpha_j^p - \beta_j^p)\Phi_{j,h}^{\ell-1}, (\alpha_i^p - \beta_i^p)\Psi_{i,h} \right)_h,
\end{aligned} \tag{4.31}$$

for all $\Phi_{i,h} \in V_h^k$. These are M decoupled scalar constant-coefficient linear reaction-diffusion equations, which are easy to solve.

One may also solve the equation (4.31) sequentially (in a Gauss-Seidel manner), which uses the updated $\Phi_{j,h}^\ell$ for $j < i$ when computing the variable $\Phi_{i,h}^\ell$.

4.3.2. Step B/C: system case

Similar to the scalar case in Subsection 4.2.2, we first compute the solutions $\boldsymbol{u}_h^\ell$ according to the following system version of (4.16):

$$\boldsymbol{u}_h^\ell = \underset{\boldsymbol{u}_h \in [W_h^k]^{3M+R}}{\operatorname{argmin}} \quad \underline{F_h}(\boldsymbol{u}_h) + \frac{1}{2r}(\boldsymbol{u}_h - r\overline{\boldsymbol{u}}_h, \boldsymbol{u}_h - r\overline{\boldsymbol{u}}_h)_h, \tag{4.32}$$

where

$$\overline{\boldsymbol{u}}_h := \underline{\mathcal{D}}\boldsymbol{\Phi}_h^\ell + \frac{1}{r}\boldsymbol{u}_h^{\ell-1},$$

with the understanding that $\underline{\mathcal{D}}\boldsymbol{\Phi}_h^\ell$ is its interpolation onto the space $[W_h^k]^{3M+R}$, and then update $\boldsymbol{u}_h^{*,\ell}$ according to

$$\boldsymbol{u}_h^{*,\ell} = \overline{\boldsymbol{u}}_h - \boldsymbol{u}_h^\ell/r. \tag{4.33}$$

Again, we solve the problem (4.32) by first locally expressing all other variables in terms of the densities, and then solving pointwise optimization problems for these densities on each quadrature point.

Proposition 4.3. *Let $\boldsymbol{u}_h^\ell$ be the solution to (4.32). Then there holds*

$$m_{i,h}^{k,\ell} = \frac{rV_{1,i}(\rho_{i,h}^\ell)}{r + V_{1,i}(\rho_{i,h}^\ell)}\overline{m}_{i,h}, \forall k = 0, 1, \text{ and } 1 \le i \le M, \tag{4.34}$$

$$s_{p,h}^\ell = \frac{rV_{2,p}(\boldsymbol{\rho}_h^\ell)}{r + V_{2,p}(\boldsymbol{\rho}_h^\ell)}\overline{s}_{p,h}, \quad \forall 1 \le p \le R, \tag{4.35}$$

and the collection of densities $\boldsymbol{\rho}_h^\ell$ is the minimizer of the following reduced problem:

$$\begin{aligned}\operatorname*{argmin}_{\boldsymbol{\rho}_h \in [W_h^k]^M} & \sum_{i=1}^M \frac{1}{2r}\left(|\rho_{i,h} - r\overline{\rho}_{i,h}|^2, 1\right)_h + \sum_{i=1}^M \left(\frac{r^2(|\overline{m}_{i,h}^0|^2 + \overline{m}_{i,h}^0|^2)}{2(r + V_{1,i}(\rho_{i,h}))}, 1\right)_h \\ & + \sum_{p=1}^R \left(\frac{r^2|\overline{s}_{p,h}|^2}{2(r + V_{2,p}(\boldsymbol{\rho}_h))}, 1\right)_h + \Delta t \sum_{i=1}^M \mathcal{E}_{i,h}(\rho_{i,h}).\end{aligned} \tag{4.36}$$

By the choice of the integration rule space (4.2) and its nodal basis (4.3), it is clear that the minimization problem (4.36) can be solved in a pointwise fashion per quadrature point. On each quadrature point, it is an M-dimensional minimization problem, where the coupling is introduced in the reaction term in the second row of (4.36). Again, we propose to solve M independent single-variable minimization problems in parallel by treating the reaction term semi-implicitly. Specifically, the solution $\rho_{i,h}^\ell$ for each $1 \le i \le M$ is obtained by solving the following problems in parallel:

$$\begin{aligned}\operatorname*{argmin}_{\rho_{i,h} \in W_h^k} & \frac{1}{2r}\left(|\rho_{i,h} - r\overline{\rho}_{i,h}|^2, 1\right)_h + \left(\frac{r^2(|\overline{m}_{i,h}^0|^2 + \overline{m}_{i,h}^0|^2)}{2(r + V_{1,i}(\rho_{i,h}))}, 1\right)_h \\ & + \sum_{p=1}^R \left(\frac{r^2|\overline{s}_{p,h}|^2}{2(r + V_{2,p}(\tilde{\boldsymbol{\rho}}_h))}, 1\right)_h + \Delta t\mathcal{E}_{i,h}(\rho_{i,h}).\end{aligned} \tag{4.37}$$

Here

$$\tilde{\boldsymbol{\rho}}_h^i = (\rho_{1,h}^{\ell-1}, \cdots, \rho_{i-1,h}^{\ell-1}, \rho_{i,h}, \rho_{i+1,h}^{\ell-1}, \cdots, \rho_{M,h}^{\ell-1}),$$

i.e., all other densities are evaluated explicitly at level $\ell - 1$. By the choice of mobility functions in (3.11), it is easy to show that the problem (4.37) is convex and hence has a unique global minimizer. We collect this modified ALG2 implementation in the following algorithm.

Algorithm 3 One iteration of modified ALG2 algorithm for (4.24).

- Step A: update $\boldsymbol{\Phi}_h^\ell$. Find $\Phi_{i,h}^\ell \in V_h^k$ such that the equation (4.31) holds for each $1 \le i \le M$.
- Step B/C: update $\boldsymbol{u}_h^\ell, \boldsymbol{u}_h^{*,\ell}$. First, find $\rho_{i,h}^\ell$ such that it is the minimizer to (4.37) for each $1 \le i \le M$. Then update $m_{i,h}^{k,\ell}$ for $k = 0, 1$ according to (4.34) and update $s_{p,h}^\ell$ for $1 \le p \le R$ according to (4.35). Finally, update $\boldsymbol{u}_h^{*,\ell}$ according to (4.33).

4.4. High-order FEM for reversible reaction-diffusion systems with detailed balance

For a reversible reaction-diffusion system with detailed balance, the spatial discretization and the corresponding practical ALG2 implementation are the same as the one in a strongly reversible case, with the only change that the discrete energy now takes the following form:

$$\mathcal{E}_{i,h}(\rho) = (\rho(\log(\kappa_i \rho) - 1), 1)_h,$$

where $\kappa_i > 0$ depends on the reaction rates.

Here a small modification (with a reduced cost) is needed to simulate the reversible Gray-Scott model in Example 3.4.2 since it does not include diffusion for the last two species. Specifically, we do not need flux approximations for the last two species, and the variables and operators in the fully discrete algorithm (4.24) for the system (3.20) are recorded below for completeness:

$$\boldsymbol{u}_h = (\rho_{1,h}, m^0_{1,h}, m^1_{1,h}, \rho_{2,h}, m^0_{2,h}, m^1_{2,h}, \rho_{3,h}, \rho_{4,h}, s_{1,h}, s_{2,h}, s_{3,h}), \tag{4.38a}$$

$$\Phi_h = (\Phi_{1,h}, \Phi_{2,h}, \Phi_{3,h}, \Phi_{4,h}), \tag{4.38b}$$

$$\begin{aligned}\underline{\mathcal{D}}\boldsymbol{\Phi}_h := \Big(&- \Phi_{1,h}, \partial_{x_0}\Phi_{1,h}, \partial_{x_1}\Phi_{1,h}, -\Phi_{2,h}, \partial_{x_0}\Phi_{2,h}, \partial_{x_1}\Phi_{2,h}, \\ &- \Phi_{3,h}, -\Phi_{4,h}, \Phi_{1,h} - \Phi_{2,h}, \Phi_{2,h} - \Phi_{3,h}, \Phi_{1,h} - \Phi_{4,h}\Big),\end{aligned} \tag{4.38c}$$

$$\underline{G_h}(\boldsymbol{\Phi}_h) := \sum_{i=1}^{4} (\rho^{\text{old}}_{i,h}, \Phi_{i,h})_h, \tag{4.38d}$$

$$\begin{aligned}\underline{F_h}(\boldsymbol{u}_h) := &\left(\sum_{i=1}^{2} \frac{|m^0_{i,h}|^2 + |m^1_{i,h}|^2}{2V_{1,i}(\rho_{i,h})} + \sum_{p=1}^{3} \frac{|s_{p,h}|^2}{2V_{2,p}(\boldsymbol{\rho}_h)}, 1 \right)_h \\ &+ \Delta t \sum_{i=1}^{4} \mathcal{E}_{i,h}(\rho_{i,h}),\end{aligned} \tag{4.38e}$$

where the parameters and mobility functions are given in (3.19). Note that Step A of Algorithm 3 now becomes two scalar linear reaction-diffusion equation updates for $\Phi^{\ell}_{1,h}$ and $\Phi^{\ell}_{2,h}$, and two simple mass matrix updates for $\Phi^{\ell}_{3,h}$ and $\Phi^{\ell}_{4,h}$.

5. Numerical experiments

In this section, we first present a 3D numerical example for the heat equation to demonstrate the first order accuracy in time of Algorithm 2. We then conduct comprehensive 2D experiments to show the efficiency and effectiveness of the proposed numerical algorithms. Throughout, we take the augmented Lagrangian parameter to be $r = 1$, and perform 200 ALG iterations in each time step for all test cases. We use the previous step solution as the initial guess for the ALG2 algorithm. Our numerical simulations are performed using the open-source finite-element software NGSolve [65], https://ngsolve.org/.

5.1. Temporal convergence rate

We first consider the heat equation in three dimensions

$$\partial_t \rho - \triangle \rho = 0,$$

on the domain $\Omega = [0, 1]^3$ with homogeneous Neumann boundary conditions. It is a Wasserstein gradient flow of the form (3.1) with energy $\mathcal{E}(\rho) := \int_\Omega \rho(\log \rho - 1)dx$. We take the initial condition to be

$$\rho_0(x_0, x_1, x_2) = 1 + \frac{1}{2}\cos(\pi x_0)\cos(\pi x_1)\cos(\pi x_2).$$

The exact solution is

$$\rho_{ex}(x_0, x_1, x_2, t) = 1 + \frac{1}{2}\cos(\pi x_0)\cos(\pi x_1)\cos(\pi x_2)\exp(-3\pi^2 t).$$

We perform a temporal mesh convergence study for the scheme (4.5) using Algorithm 2 on a fixed uniform mesh of size $16 \times 16 \times 16$ with polynomial degree $k = 3$. The final time of simulation is $T = 0.1$. We use uniform temporal meshes of sizes $N = 4 \times 2^{\ell}$, where $\ell = 0, 1, 2, 3$ is the temporal mesh level. Here the temporal error dominates the spatial error. The L^2-convergence in the density ρ at final time $T = 0.1$ is recorded in Table 1. We clearly observe the first order convergence in time.

Table 1
Temporal convergence rate for the L^2-error in density at $T = 0.1$ of scheme (4.5) applied to a 3D heat equation.

mesh level ℓ	error	rate
0	1.012e-02	–
1	5.090e-03	0.99
2	2.551e-03	1.00
3	1.290e-03	0.98

Table 2
Convergence rates of scheme (4.5) with different polynomial degrees k applied to a 2D steady Fokker Planck equation.

$\dim(V_h^k)$	$k=1$		$k=2$		$k=4$	
81	2.362e-03	–	2.409e-04	–	2.628e-05	–
289	5.923e-04	2.00	3.298e-05	2.87	1.424e-06	4.21
1089	1.482e-04	2.00	4.232e-06	2.96	5.589e-08	4.67
4225	3.705e-05	2.00	5.326e-07	2.99	1.884e-09	4.89

5.2. Spatial convergence rates

We now consider the nonlinear Fokker-Plank equation

$$\partial_t \rho - \triangle \rho^3 = \nabla \cdot (\rho x),$$

on the domain $\Omega = [-1, 1] \times [-1, 1]$ with homogeneous Neumann boundary conditions. It is a Wasserstein gradient flow of the form (3.1) with energy

$$\mathcal{E}(\rho) := \int_\Omega \left(\frac{1}{2}\rho(x)^3 + \frac{1}{2}(x_0^2 + x_1^2)\rho(x) \right) dx,$$

where $x = (x_0, x_1)$. This problem reaches a steady state solution:

$$\rho_{\text{steady}}(x_0, x_1) = \sqrt{\frac{(2C - (x_0^2 + x_1^2))_+}{3}},$$

that satisfies either

$$\frac{\delta \mathcal{E}}{\delta \rho} = \frac{3}{2}\rho^2 + \frac{1}{2}(x_0^2 + x_1^2) = C,$$

or $\rho = 0$. Here the constant C depends on the total mass of the initial condition, which we set to be $C = 2$ so that the solution on Ω is positive and smooth.

We perform a mesh convergence study for the scheme (4.5) using Algorithm 2 with polynomial degree $k = 1, 2, 4$ on a sequence of uniformly refined meshes. The coarse mesh is of size 8×8 for $k = 1$, 4×4 for $k = 2$, and 2×2 for $k = 4$, so that the total number of degrees of freedom for Φ is the same on each mesh level for different polynomial degrees. As accuracy in time does not play a role here in the steady state solution, we take large time step size with $\Delta t = 1$, and perform 10 time steps of simulation where the numerical solution reaches the steady state. The L^2-convergence in the density ρ is recorded in Table 2. We clearly observe the $k+1$-th order of convergence for each case. In particular, the high-order method leads to a smaller error when a same number of total degrees of freedom is used.

5.3. Aggregation-drift-diffusion equations

We consider Wasserstein gradient flow (3.1) with five choices of energies (3.2) that include aggregation effects. The specific form of the energy, along with the domain size L where the computational domain $\Omega = [-L, L] \times [-L, L]$, and the initial conditions are given in Table 3. Here $\chi_{[-3,3]\times[-3,3]}$ is the characteristic function on $[-3, 3] \times [-3, 3]$ for Case 5. All cases were considered in [12], except Case 4 which adds an additional diffusion to the energy in Case 3.

Note that the interaction kernel $W(x)$ for Cases 2/3/4 is singular at zero. Here we use a higher-order numerical integration rule, which avoids the evaluation of $W(x)$ at zero to compute the convolution

$$W * \rho(\xi_i), \quad \forall 1 \le i \le N_W,$$

Table 3
Example 5.3. Five choices of energies, domain size, and initial condition.

Case	$\alpha U_m(\rho)$	$V(x)$	$W(x)$	L	I.C.
1	0	0	$\frac{\lvert x\rvert^4}{4} - \frac{\lvert x\rvert^2}{2}$	1	$\frac{25}{2\pi}\exp(-\frac{25}{2}\lvert x\rvert^2)$
2	0	0	$\frac{\lvert x\rvert^2}{2} - \log(\lvert x\rvert)$	1.5	$\frac{25}{8\pi}\exp(-\frac{25}{8}\lvert x\rvert^2)$
3	0	$-\frac{1}{4}\log(\lvert x\rvert)$	$\frac{\lvert x\rvert^2}{2} - \log(\lvert x\rvert)$	1.5	$\frac{25}{8\pi}\exp(-\frac{25}{8}\lvert x\rvert^2)$
4	$0.1\rho^2$	$-\frac{1}{4}\log(\lvert x\rvert)$	$\frac{\lvert x\rvert^2}{2} - \log(\lvert x\rvert)$	1.5	$\frac{25}{8\pi}\exp(-\frac{25}{8}\lvert x\rvert^2)$
5	$0.1\rho^3$	0	$-\exp(-\lvert x\rvert^2)/\pi$	4	$0.25\chi_{[-3,3]\times[-3,3]}$

at the quadrature points $\{\xi_i\}_{i=1}^{N_W}$. Fast Fourier transform is used to evaluate these convolutions all together.

For all cases, we take the computational mesh to be a 32×32 uniform square mesh, and use polynomial degree $k = 4$ in the scheme (4.5). We take time step size $\Delta t = 0.05$ for the first four cases, and $\Delta t = 0.5$ for the last case. The final time of simulation is $T = 10$ for Case 1, $T = 3$ for Cases 2/3/4, and $T = 15$ for Case 5. Snapshots of the density contours at different times are shown in Fig. 1. We find the results for Cases 1/2/3 and 5 are qualitatively similar to the results reported in [12]. In particular, Case 1 converges to a steady Dirac ring solution; Case 2 converges to a steady constant solution with a circular shape; Case 3 converges to a characteristic function for the torus due to the drift effects that pushes away the density from the origin; and the competition between median range aggregation with short/long range diffusion is observed for Case 5. Moreover, the diffusion effects of Case 4 comparing with Case 3 are also clearly seen.

5.4. Scalar reaction-diffusion equation

We take the Case 4 energy in Table 3, but consider the reaction-diffusion equation (3.5). Three choices of mobility coefficient $V_2(\rho)$ are used in this example, namely,

$$\begin{cases} \text{Type 1: } V_2(\rho) = 0.1, \\ \text{Type 2: } V_2(\rho) = 0.1\rho, \\ \text{Type 3: } V_2(\rho) = 0.1\frac{\rho-1}{\log(\rho)}. \end{cases} \tag{5.1}$$

The same discretization setup as in the previous example is used, i.e., using polynomial degree $k = 4$ on a 32×32 uniform mesh with time step size $\Delta t = 0.05$, and final time $T = 3$.

Snapshots of the density contours for each case at different times are shown in Fig. 2. It is clear from the color range of these plots that reaction effects leads to mass loss, with the Type 1 reaction has the most mass loss, followed by Type 3 reaction.

5.5. Fisher-KPP equation

Our next example deals with the Fisher-KPP equation (3.7). Here we slightly modify the PDE (3.7) to allow for anisotropic diffusion:

$$\partial_t \rho - \lambda_1 \partial_{x_0 x_0}\rho - \lambda_2 \partial_{x_1 x_1}\rho = \mu\rho(1-\rho).$$

We use a similar setup as in [68, Section 3.1], where the diffusion parameters are taken to be $\lambda_1 = 0.1, \lambda_2 = 0.01$, and $\mu > 0$ is the reaction coefficient to be specified. Initial condition is a flat top Gaussian:

$$\rho_0(x_0, x_1) = \begin{cases} 1, & \text{if } x_0^2 + 4x_1^2 \le 0.25 \\ \exp(-10(x_0^2 + 4x_1^2 - 0.25)), & \text{otherwise} \end{cases}$$

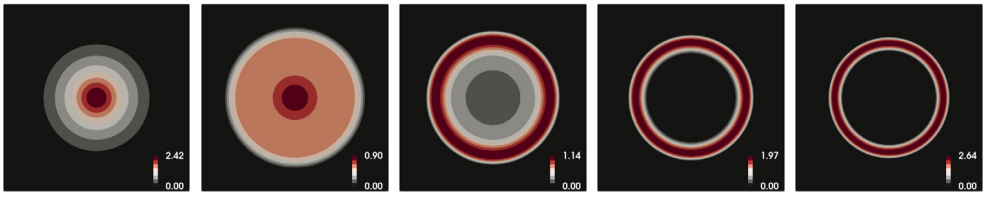
The computational domain is a rectangle $\Omega = [-2, 2] \times [-1, 1]$, which is discretized with a 32×16 square mesh. We use polynomial degree $k = 4$ for the scheme (4.5), in which the functional F_h in (4.10) is adjusted as follows to allow for anisotropic diffusion:

$$F_h(\boldsymbol{u}_h) := \left(\frac{|m_h^0|^2}{2V_{1,0}(\rho_h)} + \frac{|m_h^1|^2}{2V_{1,1}(\rho_h)} + \frac{|s_h|^2}{2V_2(\rho_h)}, 1 \right)_h + \Delta t\, \mathcal{E}_h(u_{0,h}),$$

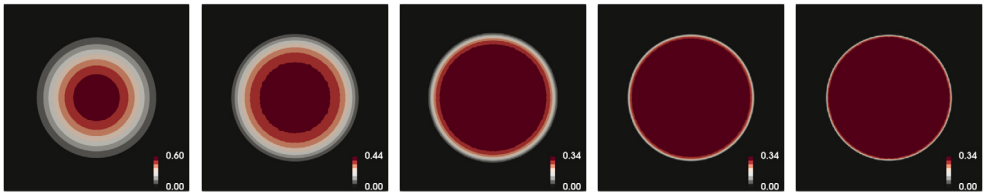
where $V_{1,0}(\rho) := \lambda_1\rho$, $V_{1,1}(\rho) := \lambda_2\rho$, $V_2(\rho) := \mu\frac{\rho(\rho-1)}{\log(\rho)}$, and the energy satisfies

$$\mathcal{E}_h(\rho) = (\rho(\log(\rho) - 1), 1)_h.$$

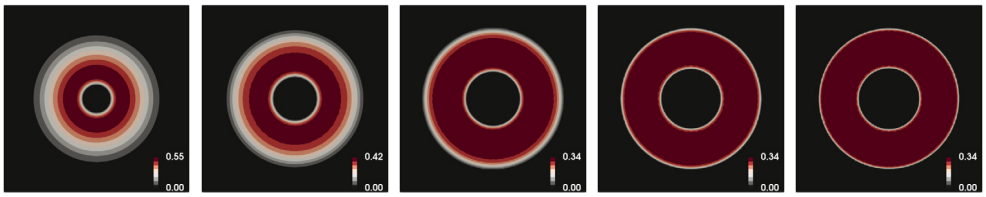
We take time step size $\Delta t = 0.1$, and the final time is $T = 4$.



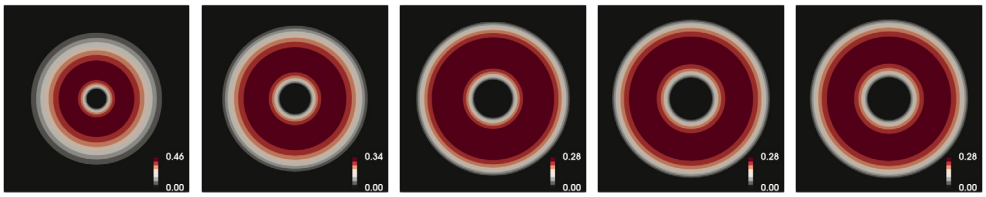
(a) Case 1. Left to right time: $t = 0.5, 1.5, 3.0, 6.0, 10$



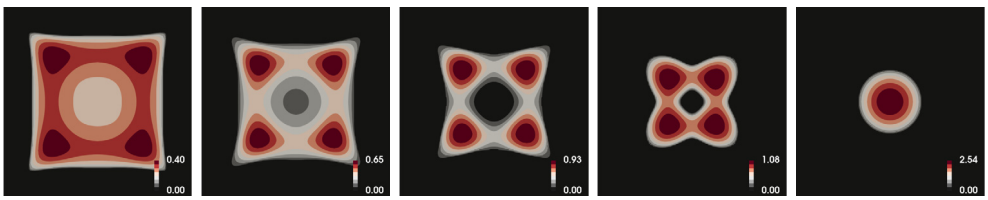
(b) Case 2. Left to right time: $t = 0.2, 0.5, 1.5, 2.0, 3.0$



(c) Case 3. Left to right time: $t = 0.2, 0.5, 1.5, 2.0, 3.0$



(d) Case 4. Left to right time: $t = 0.2, 0.5, 1.5, 2.0, 3.0$



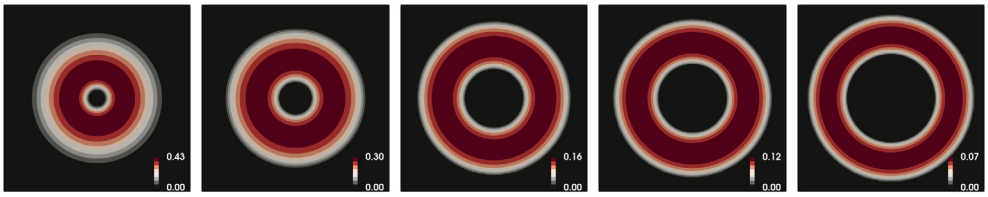
(e) Case 5. Left to right time: $t = 2, 4, 6, 10, 15$

Fig. 1. Example 5.3. Snapshots of density contours at different times for different test cases. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

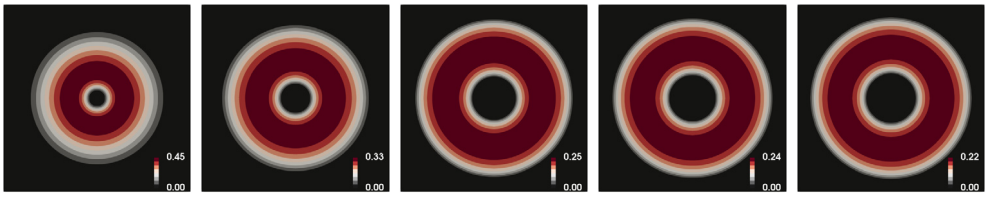
Snapshots of the density contours for $\mu = 0.1$ (weak reaction) $\mu = 0.5$ (medium reaction), and $\mu = 1.0$ (strong reaction) at different times are shown in Fig. 3. We further plot the evolution of energy $\mathcal{E}_h(\rho_h)$ and total mass $\int_\Omega \rho_h \, dx$ over time for the three cases in Fig. 4. It is clear that the energy is monotonically decreasing for all three cases and the total mass is monotonically increasing, where a faster decay of energy is observed when the reaction coefficient μ is larger.

5.6. Two-component reversible reaction-diffusion system with detailed balance

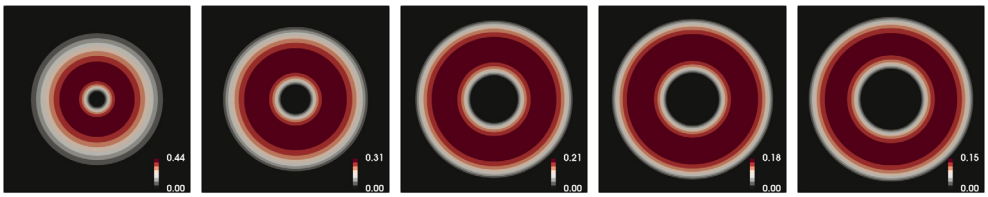
We consider the two-species model discussed in Section 3.4. In particular, we consider the system (3.18) with parameters $k_+ = 1$ and $k_- = 0.1$, $\gamma_1 = 0.2$, $\gamma_2 = 0.1$, and $V_{1,1}(\rho) = \gamma_1 \rho^m$ and $V_{1,2}(\rho) = \gamma_2 \rho$ with four choices of $m \in \{1, 2, 3, 4\}$. Here porous medium type diffusion is used for the first species with density ρ_1 and linear diffusion is used for the second species



(a) Case 4 energy, Type 1 reaction. Left to right time: $t = 0.2, 0.5, 1.5, 2.0, 3.0$

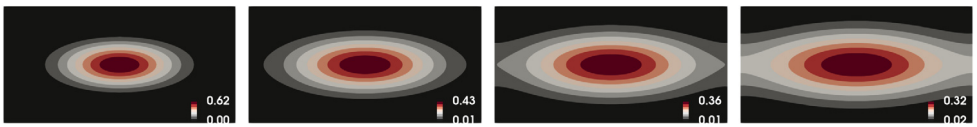


(b) Case 4 energy, Type 2 reaction. Left to right time: $t = 0.2, 0.5, 1.5, 2.0, 3.0$

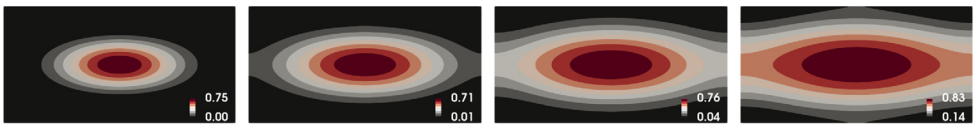


(c) Case 4 energy, Type 3 reaction. Left to right time: $t = 0.2, 0.5, 1.5, 2.0, 3.0$

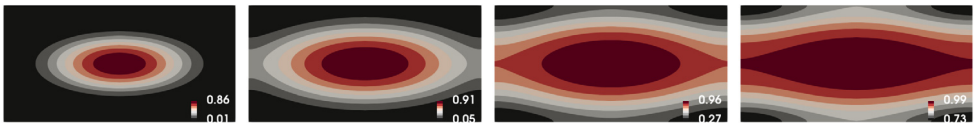
Fig. 2. Example 5.4. Snapshots of density contours at different times for different reaction mobility functions.



(a) Reaction coefficient $\mu = 0.1$. Left to right time: $t = 1, 2, 3, 4$



(b) Reaction coefficient $\mu = 0.5$. Left to right time: $t = 1, 2, 3, 4$



(c) Reaction coefficient $\mu = 1.0$. Left to right time: $t = 1, 2, 3, 4$

Fig. 3. Example 5.4. Snapshots of density contours at different times for different reaction coefficients.

with density ρ_2. A similar model was used in [48,49]. The problems are solved on the domain $\Omega = [-1, 1] \times [-1, 1]$ with the following initial data

$$\rho_1(x, 0) = \frac{1}{2}\left(1 - \tanh(10(\sqrt{x_0^2 + x_1^2} - 0.2))\right),$$
$$\rho_2(x, 0) = \frac{1}{2}\left(1 + \tanh(10(\sqrt{x_0^2 + x_1^2} - 0.2)).\right).$$

Final time is taken to be $T = 2$.

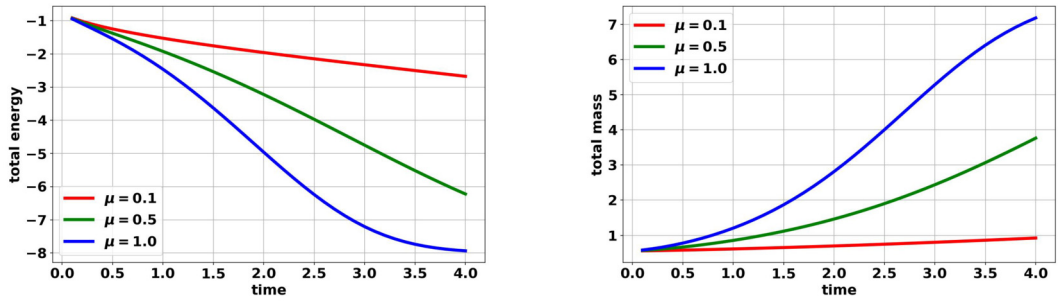
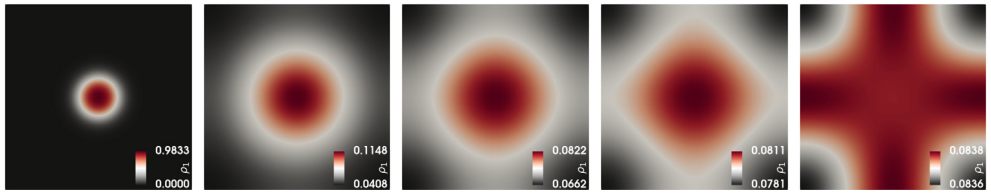
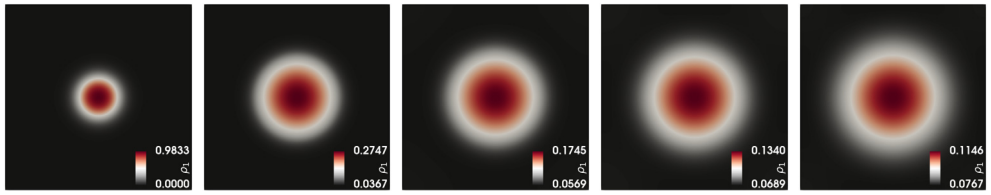


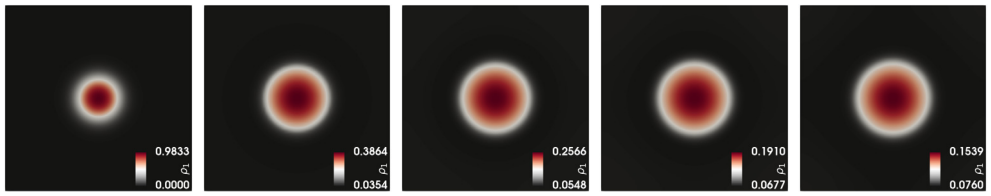
Fig. 4. Example 5.4. Evolution of total energy (left) and total mass (right) over time.



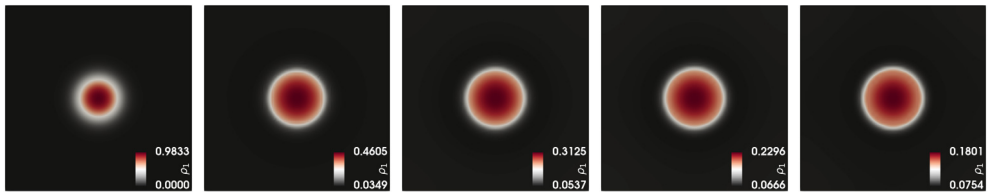
(a) $V_{1,1}(\rho) = \gamma_1 \rho$. Left to right time: $t = 0, 0.5, 1, 1.5, 2$



(b) $V_{1,1}(\rho) = \gamma_1 \rho^2$. Left to right time: $t = 0, 0.5, 1, 1.5, 2$



(c) $V_{1,1}(\rho) = \gamma_1 \rho^3$. Left to right time: $t = 0, 0.5, 1, 1.5, 2$

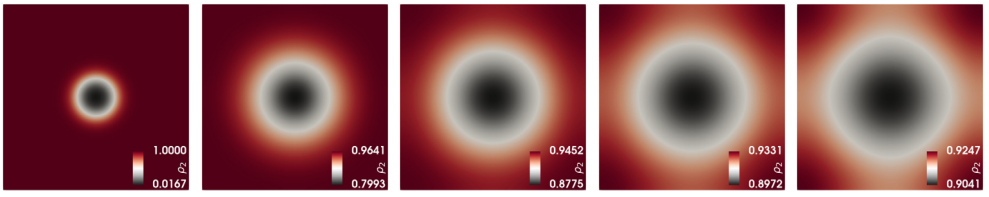


(d) $V_{1,1}(\rho) = \gamma_1 \rho^4$. Left to right time: $t = 0, 0.5, 1, 1.5, 2$

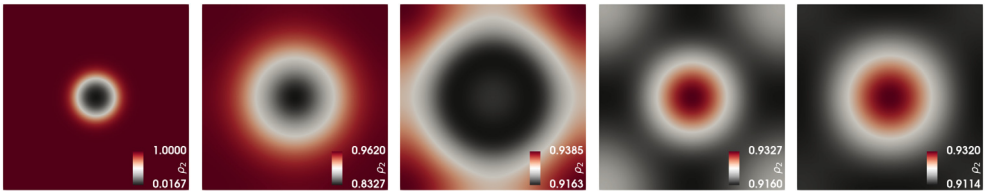
Fig. 5. Example 5.6. Snapshots of first-component density contours at different times for different $V_{1,1}(\rho)$.

We use the scheme (4.24) with polynomial degree $k = 4$ on a 16×16 mesh with time step size $\Delta t = 0.05$. We apply Algorithm 3 to solve the resulting saddle point problem. Snapshots of the density contours at different times are shown in Fig. 5 for the first component and in Fig. 6 for the second component. It is clear that increasing the power m leads to a slower diffusion for the first species.

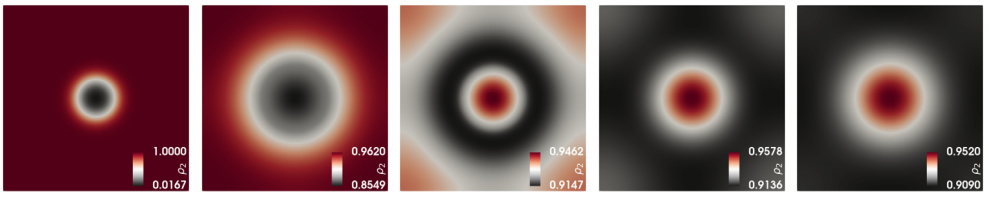
We further plot the time evolution of the total energy $\mathcal{E}_{\text{total}} = \mathcal{E}_{1,h}(\rho_{1,h}) + \mathcal{E}_{2,h}(\rho_{2,h})$ and total mass $\int_\Omega (\rho_{1,h} + \rho_{2,h})\, dx$ for the four cases in Fig. 7. Moreover, the total mass conservation is kept well within an error of 10^{-4} for all cases.



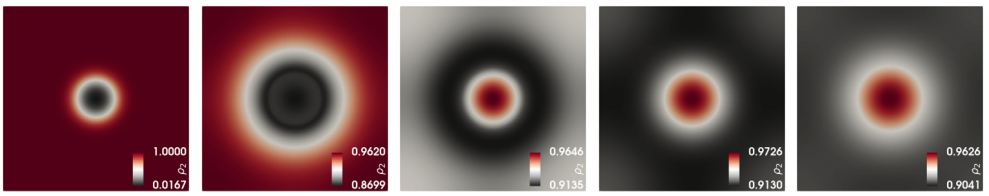
(a) $V_{1,1}(\rho) = \rho$. Left to right time: $t = 0, 0.5, 1, 1.5, 2$



(b) $V_{1,1}(\rho) = \rho^2$. Left to right time: $t = 0, 0.5, 1, 1.5, 2$



(c) $V_{1,1}(\rho) = \rho^3$. Left to right time: $t = 0, 0.5, 1, 1.5, 2$



(d) $V_{1,1}(\rho) = \rho^4$. Left to right time: $t = 0, 0.5, 1, 1.5, 2$

Fig. 6. Example 5.6. Snapshots of second-component density contours at different times for different $V_{1,1}(\rho)$.

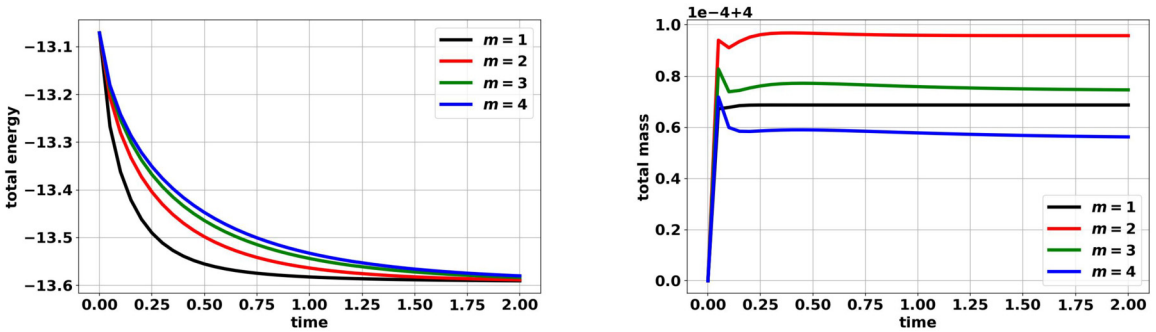
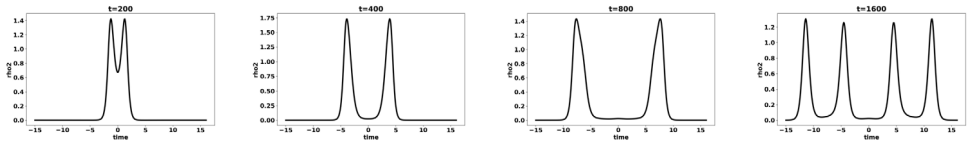


Fig. 7. Example 5.6. Evolution of total energy (left) and total mass (right) over time with $V_{1,1}(\rho) = \gamma_1 \rho^m$.

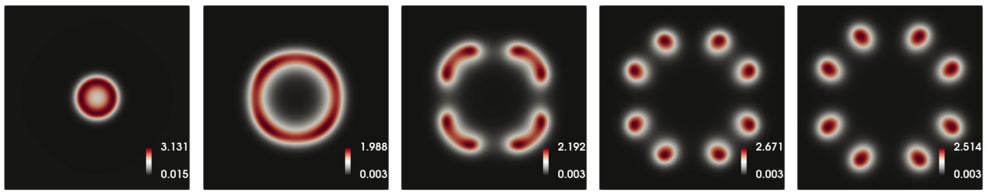
5.7. Reversible Gray-Scott model

In our last example, we simulate the 4-component reversible Gray-Scott model (3.20) using the Algorithm 3 for the fully discrete scheme (4.24) with variables/operators (4.38). The physical parameters are chosen to be the following:

$$\gamma_1 = 1, \quad \gamma_2 = 0.01,$$



(a) 1D results. Left to right time: $t = 200, 400, 800, 1600$.



(b) 2D results. Left to right time: $t = 100, 200, 300, 400, 500$

Fig. 8. Example 5.7. Snapshots of second-component density contours at different times for 1D (top) and 2D (bottom) simulations.

$$k_+^1 = 1, \quad k_-^1 = 10^{-3},$$
$$k_+^2 = 8.4 \times 10^{-2}, \quad k_-^2 = 8.4 \times 10^{-5},$$
$$k_+^3 = 2.4 \times 10^{-2}, \quad k_-^3 = 2.4 \times 10^{-5},$$

where the backward reaction rates are taken to be 1000 times smaller than the forward reaction rates. This provides a good approximation to the irreversible Gray-Scott model (3.21). We consider both 1D and 2D simulations for this problem. The initial conditions for the second component density ρ_2 are taken to be

$$\rho_2(x,0) = \begin{cases} 0.15 + \frac{1}{4}x^2(x+1)^2 & \text{if } -1 \le x \le 0, \\ 0.15 + \frac{1}{4}x^2(1-x)^2 & \text{if } 0 \le x \le 1, \\ 0.15 & \text{elsewhere}, \end{cases}$$

in one dimension, and

$$\rho_2(x,0) = \begin{cases} 0.15 + 4x^2(x+1)^2y^2(y+1)^2 & \text{if } -1 \le x \le 0 \text{ and } -1 \le y \le 0, \\ 0.15 + 4x^2(x+1)^2y^2(1-y)^2 & \text{if } -1 \le x \le 0 \text{ and } 0 \le y \le 1, \\ 0.15 + 4x^2(1-x)^2y^2(y+1)^2 & \text{if } 0 \le x \le 1 \text{ and } -1 \le y \le 0, \\ 0.15 + 4x^2(1-x)^2y^2(1-y)^2 & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1, \\ 0.15 & \text{elsewhere}, \end{cases}$$

in two dimensions. The initial conditions for the other densities are taken to be

$$\rho_1(x,0) = 1 - 2\rho_2(x,0), \quad \rho_3(x,0) = 1, \quad \rho_4(x,0) = k_+^3/k_-^3 = 1000.$$

For the 1D simulation, we take the computation domain to be $\Omega_{1D} = [-16, 16]$ and set the final time of simulation to be $T = 1600$. For the 2D simulation, we take a smaller computational domain with $\Omega_{2D} = [-8, 8] \times [-8, 8]$ and set the final time of simulation to be $T = 500$.

We apply the scheme (4.24) with $k = 4$ on a uniform mesh with mesh size $h = 1$ (32 elements in 1D, and 16×16 elements in 2D) for both problems. Here we gradually increase the time step size from $\Delta t = 0.01$ to $\Delta t = 0.1$ as initially taking $\Delta t = 0.1$ leads to numerical instability. This may be caused by our splitting version of the ALG2 implementation in Algorithm 3. A theoretical investigation on the stability of the algorithm with respect to the time step size Δt is the subject of our on-going work.

We record the snapshots of the second-component density at various times in Fig. 8. For both cases, we observe pattern formations and the solution reaches a nontrivial steady state at large time. Finally, we plot the evolution of total energy for both cases in Fig. 9, where we observe the expected monotone energy decay.

6. Conclusion

This paper applies high-order accurate finite element methods in space to compute first-order accuracy implicit-in-time gradient flows. Our formulation applies a one-step time discretization of the generalized JKO scheme and then uses the ALG2 to calculate optimization problems in each generalized JKO time step. The method is unconditionally stable when

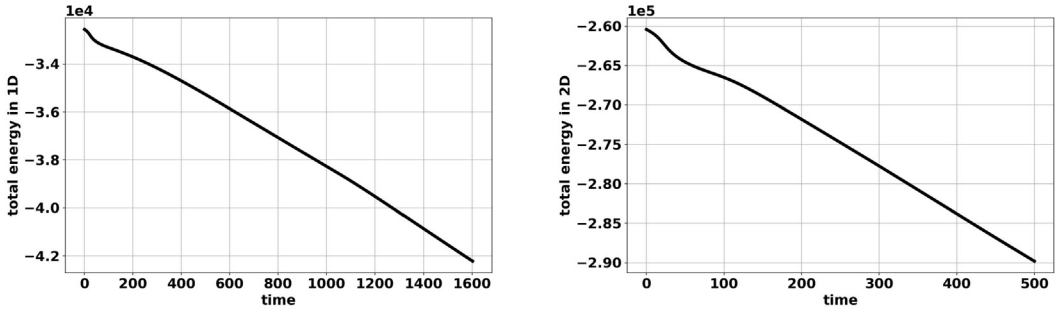


Fig. 9. Example 5.7. Evolution of total energy in 1D (left) and 2D (right).

the optimization problem is convex. Numerical experiments in two-dimensional gradient flow dynamics, such as Wasserstein gradient flows, Fisher–KPP dynamics, and reversible reaction-diffusion systems, demonstrate the effectiveness of the proposed method with high-order spatial accuracy.

We note that for dissipative dynamics, such as strongly reversible reaction-diffusion systems, different entropies $\mathcal{E}$, and optimal transport-type metrics V_1, V_2, could produce the same evolutionary equation. In simulations, we suggest selecting a suitable class of entropies and metrics to develop simple and efficient optimization procedures. Some limitations exist for computing implicit-in-time gradient flows in generalized optimal transport metric spaces. The constructed functions V_1 and V_2 should be nonnegative for entropy dissipation schemes. Our generalized JKO scheme is unstable for many reaction-diffusion equations, e.g., the Allen-Cahn-type equations [67]. We also remark that the current computations are limited to the first–order time accuracy variational-implicit schemes of gradient flows. In future work, we shall design and compute generalized optimal transport and mean field control problems for implicit-in-time fluid dynamics with general conservative-dissipative formulations. Typical examples include regularized conservation laws [40,41]. The other important question is the high-order implicit time variational schemes for initial value PDEs. This requires careful design of optimization problems related to energies, metrics, and stepsizes.

CRediT authorship contribution statement

Guosheng Fu: Conceptualization, Methodology, Software, Writing, Reviewing and Editing. **Stanley Osher**: Methodology, Reviewing and Editing. **Wuchen Li**: Conceptualization, Methodology, Writing, Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Yves Achdou, Victor Perez, Iterative strategies for solving linearized discrete mean field games systems, Netw. Heterog. Media 7 (2) (2012) 197.
[2] Shun-ichi Amari, Natural gradient works efficiently in learning, Neural Comput. 10 (2) (1998) 251–276.
[3] Luigi Ambrosio, Nicola Gigli, Giuseppe Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, 2ed edition, Lectures in Mathematics ETH Zürich, Birkhäuser, Basel, 2008.
[4] Jean-David Benamou, Yann Brenier, A numerical method for the optimal time-continuous mass transport problem and related problems, Contemp. Math. 226 (1999) 1–12.
[5] Jean-David Benamou, Yann Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math. 84 (3) (2000) 375–393.
[6] Jean-David Benamou, Guillaume Carlier, Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations, J. Optim. Theory Appl. 167 (1) (2015) 1–26.
[7] Jean-David Benamou, Guillaume Carlier, Maxime Laborde, An augmented Lagrangian approach to Wasserstein gradient flows and applications, ESAIM Proc. Surv. 54 (2016) 1–17.
[8] James H. Bramble, Multigrid Methods, Pitman Research Notes in Mathematics Series, vol. 294, Longman Scientific & Technical, Harlow, 1993, copublished in the United States with John Wiley & Sons, Inc., New York.
[9] Clément Cancès, Thomas O. Gallouët, Gabriele Todeschi, A variational finite volume scheme for Wasserstein gradient flows, Numer. Math. 146 (3) (2020) 437–480.
[10] Guillaume Carlier, Maxime Laborde, Remarks on continuity equations with nonlinear diffusion and nonlocal drifts, J. Math. Anal. Appl. 444 (2) (2016) 1690–1702.
[11] J.A. Carrillo, S. Lisini, G. Savaré, D. Slepčev, Nonlinear mobility continuity equations and generalized displacement convexity, J. Funct. Anal. 258 (4) (2010) 1273–1309.

[12] José A. Carrillo, Katy Craig, Li Wang, Chaozhen Wei, Primal dual methods for Wasserstein gradient flows, Found. Comput. Math. 22 (2) (2022) 389–443.
[13] José A. Carrillo, Bertram Düring, Daniel Matthes, David S. McCormick, A Lagrangian scheme for the solution of nonlinear diffusion equations using moving simplex meshes, J. Sci. Comput. 75 (3) (2018) 1463–1499.
[14] José A. Carrillo, Helene Ranetbauer, Marie-Therese Wolfram, Numerical simulation of nonlinear continuity equations by evolving diffeomorphisms, J. Comput. Phys. 327 (2016) 186–202.
[15] Yifan Chen, Wuchen Li, Optimal transport natural gradient for statistical manifolds with continuous sample space, Inf. Geom. 3 (1) (2020) 1–32.
[16] Qing Cheng, Jie Shen, Global constraints preserving scalar auxiliary variable schemes for gradient flows, SIAM J. Sci. Comput. 42 (4) (2020) A2489–A2513.
[17] Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, François-Xavier Vialard, Unbalanced optimal transport: dynamic and Kantorovich formulations, J. Funct. Anal. 274 (11) (2018) 3090–3123.
[18] Shui-Nee Chow, Wen Huang, Yao Li, Haomin Zhou, Fokker–Planck equations for a free energy functional or Markov process on a graph, Arch. Ration. Mech. Anal. 203 (3) (2012) 969–1008.
[19] Masao Doi, Onsager's variational principle in soft matter, J. Phys. Condens. Matter 23 (28) (2011) 284118.
[20] E. Weinan, Chao Ma, Lei Wu, Machine learning from a continuous viewpoint, i, Sci. China Math. (2020) 1–34.
[21] Jonathan Eckstein, Dimitri P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, Math. Program., Ser. A 55 (3) (1992) 293–318.
[22] Matthias Erbar, A gradient flow approach to the Boltzmann equation, arXiv:1603.00540 [math.AP], 2016.
[23] R.A. Fisher, The wave of advance of advantageous genes, Ann. Eugen. 7 (1937) 353–369.
[24] Michel Fortin, Roland Glowinski, Augmented Lagrangian Methods. Applications to the Numerical Solution of Boundary Value Problems, Studies in Mathematics and Its Applications, vol. 15, North-Holland Publishing Co., Amsterdam, 1983. Translated from the French by B. Hunt and D.C. Spicer.
[25] Guosheng Fu, Siting Liu, Stanley Osher, Wuchen Li, High order computation of optimal transport, mean field planning, and mean field games, arXiv: 2302.02308 [math.NA], 2023.
[26] Thomas O. Gallouët, Léonard Monsaingeon, A JKO splitting scheme for Kantorovich–Fisher–Rao gradient flows, SIAM J. Math. Anal. 49 (2) (2017) 1100–1130.
[27] Yuan Gao, Wuchen Li, Jian-Guo Liu, Master equations for finite state mean field games with nonlinear activations, arXiv:2212.05675 [math.OC], 2022.
[28] Alfredo Garbuno-Inigo, Franca Hoffmann, Wuchen Li, Andrew M. Stuart, Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler, SIAM J. Appl. Dyn. Syst. 19 (1) (2020) 412–441.
[29] Mi-Ho Giga, Arkadz Kirshtein, Chun Liu, Variational modeling and complex fluids, in: Yoshikazu Giga, Antonin Novotny (Eds.), Handbook of Mathematical Analysis in Mechanics of Viscous Fluids, Springer International Publishing, 2017, pp. 1–41.
[30] Annegret Glitzky, Alexander Mielke, A gradient structure for systems coupling reaction-diffusion effects in bulk and interfaces, Z. Angew. Math. Phys. 64 (1) (2013) 29–52.
[31] Yuezheng Gong, Jia Zhao, Qi Wang, Arbitrarily high-order unconditionally energy stable schemes for thermodynamically consistent gradient flow models, SIAM J. Sci. Comput. 42 (1) (2020) B135–B156.
[32] P. Gray, S.K. Scott, Sustained oscillations and other exotic patterns of behavior in isothermal reaction, J. Phys. Chem. 59 (1985) 22–32.
[33] Richard Jordan, David Kinderlehrer, Felix Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. 29 (1) (1998) 1–17.
[34] A. Kolmogorov, I. Petrovskii, N. Piskunov, A study of the diffusion equation with increase in the amount of substance, in: V.M. Tikhomirov (Ed.), Selected Works of A. N. Kolmogorov I, Kluwer, 1991, pp. 248–270.
[35] Jean-Michel Lasry, Pierre-Louis Lions, Mean field games, Jpn. J. Math. 2 (1) (2007) 229–260.
[36] Wonjun Lee, Rongjie Lai, Wuchen Li, Stanley Osher, Generalized unnormalized optimal transport and its fast algorithms, J. Comput. Phys. 436 (2021) 110041.
[37] Wonjun Lee, Siting Liu, Wuchen Li, Stanley Osher, Mean field control problems for vaccine distribution, Res. Math. Sci. 9 (3) (2022) 51.
[38] Wonjun Lee, Siting Liu, Hamidou Tembine, Wuchen Li, Stanley Osher, Controlling propagation of epidemics via mean-field control, SIAM J. Appl. Math. 81 (1) (2021) 190–207.
[39] Wuchen Li, Wonjun Lee, Stanley Osher, Computational mean-field information dynamics associated with reaction-diffusion equations, J. Comput. Phys. 466 (2022) 111409.
[40] Wuchen Li, Siting Liu, Stanley Osher, Controlling conservation laws i: entropy-entropy flux, arXiv:2111.05473, 2021.
[41] Wuchen Li, Siting Liu, Stanley Osher, Controlling conservation laws ii: compressible Navier–Stokes equations, J. Comput. Phys. 463 (2022) 111264.
[42] Wuchen Li, Jianfeng Lu, Li Wang, Fisher information regularization schemes for Wasserstein gradient flows, J. Comput. Phys. 416 (2020) 109449.
[43] Wuchen Li, Guido Montúfar, Natural gradient via optimal transport, Inf. Geom. 1 (2) (2018) 181–214.
[44] Jiangyan Liang, Ning Jiang, Chun Liu, Yiwei Wang, Teng-Fei Zhang, On a reversible Gray-Scott type system from energetic variational approach and its irreversible limit, J. Differ. Equ. 309 (2022) 427–454.
[45] Matthias Liero, Alexander Mielke, Giuseppe Savaré, Optimal transport in competition with reaction: the Hellinger–Kantorovich distance and geodesic curves, SIAM J. Math. Anal. 48 (4) (2016) 2869–2911.
[46] Alex Tong Lin, Wuchen Li, Stanley Osher, Guido Montúfar, Wasserstein proximal of GANs, in: Frank Nielsen, Frédéric Barbaresco (Eds.), in: Geometric Science of Information, vol. 12829, Springer International Publishing, Cham, 2021, pp. 524–533.
[47] Chun Liu, An introduction of elastic complex fluids: an energetic variational approach, in: Multi-Scale Phenomena in Complex Fluids: Modeling, Analysis and Numerical Simulation, World Scientific, 2009, pp. 286–337.
[48] Chun Liu, Cheng Wang, Yiwei Wang, A structure-preserving, operator splitting scheme for reaction-diffusion equations with detailed balance, J. Comput. Phys. 436 (2021) 110253.
[49] Chun Liu, Cheng Wang, Yiwei Wang, A second-order accurate, operator splitting scheme for reaction-diffusion systems in an energetic variational formulation, SIAM J. Sci. Comput. 44 (4) (2022) A2276–A2301.
[50] Chun Liu, Cheng Wang, Yiwei Wang, Steven M. Wise, Convergence analysis of the variational operator splitting scheme for a reaction-diffusion system with detailed balance, SIAM J. Numer. Anal. 60 (2) (2022) 781–803.
[51] Chun Liu, Yiwei Wang, On Lagrangian schemes for porous medium type generalized diffusion equations: a discrete energetic variational approach, J. Comput. Phys. (2020) 109566.
[52] Hailiang Liu, Wumaier Maimaitiyiming, A dynamic mass transport method for Poisson-Nernst-Planck equations, J. Comput. Phys. 473 (2023) 111699.
[53] Jian-Guo Liu, Min Tang, Li Wang, Zhennan Zhou, An accurate front capturing scheme for tumor growth models with a free boundary limit, J. Comput. Phys. 364 (2018) 73–94.
[54] Jan Maas, Gradient flows of the entropy for finite Markov chains, J. Funct. Anal. 261 (8) (2011) 2250–2292.
[55] Alexander Mielke, A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems, Nonlinearity 24 (4) (2011) 1329–1346.
[56] L. Onsager, S. Machlup, Fluctuations and irreversible processes, Phys. Rev. 91 (6) (1953) 1505–1512.
[57] Hans Christian Öttinger, Miroslav Grmela, Dynamics and thermodynamics of complex fluids. II. Illustrations of a general formalism, Phys. Rev. E 56 (6) (1997) 6633–6655.
[58] Nicolas Papadakis, Gabriel Peyré, Edouard Oudet, Optimal transport with proximal splitting, SIAM J. Imaging Sci. 7 (1) (2014) 212–238.

[59] J.E. Pearson, Complex patterns in a simple system, Science 261 (1993) 189–192.
[60] Mark A. Peletier, Variational modelling: energies, gradient flows, and large deviations, arXiv:1402.1990 [math-ph], 2014.
[61] Mark A. Peletier, Giuseppe Savaré, Marco Veneroni, From diffusion to reaction via Γ-convergence, SIAM J. Math. Anal. 42 (4) (2010) 1805–1825.
[62] Benoît Perthame, Fernando Quirós, Juan Luis Vázquez, The Hele-Shaw asymptotics for mechanical models of tumor growth, Arch. Ration. Mech. Anal. 212 (1) (2014) 93–127.
[63] Gabriel Peyré, Marco Cuturi, Computational optimal transport, Found. Trends Mach. Learn. 11 (5–6) (2019) 355–607.
[64] Filippo Santambrogio, Lectures on optimal transport [book review of 4294651]; *an invitation to optimal transport, Wasserstein distances, and gradient flows* [book review of 4331435], Eur. Math. Soc. Mag. 124 (2022) 60–63.
[65] J. Schöberl, C++11 Implementation of Finite Elements in NGSolve, Institute for Analysis and Scientific Computing, Vienna University of Technology, 2014. ASC Report 30/2014.
[66] Jie Shen, Jie Xu, Unconditionally positivity preserving and energy dissipative schemes for Poisson–Nernst–Planck equations, Numer. Math. 148 (3) (2021) 671–697.
[67] Jie Shen, Xiaofeng Yang, Numerical approximations of Allen-Cahn and Cahn-Hilliard equations, Discrete Contin. Dyn. Syst., Ser. A 28 (4) (2010) 1669–1691.
[68] S. Tang, S. Qin, R.O. Weber, Numerical studies on 2-dimensional reaction-diffusion equations, J. Aust. Math. Soc. Ser. B 35 (2) (1993) 223–243.
[69] Cédric Villani, Optimal Transport: Old and New, vol. 338, Springer Science & Business Media, 2008.
[70] Yifei Wang, Wuchen Li, Accelerated information gradient flow, J. Sci. Comput. 90 (1) (2022) 11.
[71] J. Xu, The auxiliary space method and optimal multigrid preconditioning techniques for unstructured grids, in: International GAMM-Workshop on Multi-Level Methods, Meisdorf, 1994, Computing 56 (3) (1996) 215–235.
[72] Yi Zhu, Liu Hong, Zaibao Yang, Wen-An Yong, Conservation-dissipation formalism of irreversible thermodynamics, J. Non-Equilib. Thermodyn. 40 (2) (2015).