

COMPUTATIONAL MEAN-FIELD INFORMATION DYNAMICS ASSOCIATED WITH REACTION-DIFFUSION EQUATIONS

WUCHEN LI, WONJUN LEE, AND STANLEY OSHER

ABSTRACT. We formulate and compute a class of mean-field information dynamics for reaction-diffusion equations. Given a class of nonlinear reaction-diffusion equations and entropy type Lyapunov functionals, we study their gradient flows formulations with generalized optimal transport metrics and mean-field control problems. We apply the primal-dual hybrid gradient algorithm to compute the mean-field control problems with potential energies. A byproduct of the proposed method contains a new and efficient variational scheme for solving implicit in time schemes of mean-field control problems. Several numerical examples demonstrate the solutions of mean-field control problems.

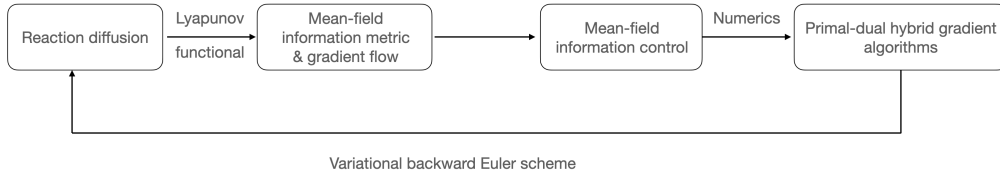
1. INTRODUCTION

Metrics [2] are essential in mathematical physics equations with applications in scientific computing and Bayesian sampling problems. One popular example in this field is the optimal transport metric, a.k.a. Wasserstein metric [2, 40], defined in probability density space. The study of gradient flow in optimal transport metric space has found applications in Markov-chain-Monte-Carlo (MCMC) methods. An example is that the heat flow is the gradient descent flow of negative Boltzman-Shannon entropy in Wasserstein space [38]. In addition, the optimal control problems and differential games in Wasserstein space are known as mean-field control problems and mean-field games, respectively [4, 24, 37]. Moreover, all the above variational formulations are useful in modeling a large number of particles simultaneously, which are essential in modeling inverse problems, AI inference, and optimization problems [15, 16, 17, 29, 31].

Recently, generalizations of optimal transport metrics and gradient flows have been studied by [6, 11, 29, 37]; see many references therein. They are helpful in studying nonlinear diffusion equations. Meanwhile, information geometry has been using the Fisher-Rao information metric, which can be applied to study the pure reaction equations, arisen in population games, AI inference, and modeling [1]. A “linear” combination of information geometry and optimal transport metrics has been applied to the study of reaction-diffusion equations [35, 36]. One typical example is the unbalanced optimal transport metric [10, 22, 30]. In this direction, [13, 25] also propose generalized optimal transport distances in unnormalized density space.

Key words and phrases. Reaction-diffusion equations; Onsager principle; Optimal transport; Information geometry; Mean-field games; Primal-dual hybrid gradient algorithms; Implicit schemes.

This research is supported by AFOSR MURI FA9550-18-1-0502, ONR grants N000142012093 and N000141812527.



This paper introduces numerical schemes for general metric spaces and mean-field control problems for reaction-diffusion equations. Given a Lyapunov functional (entropy) and a nonlinear reaction-diffusion equation, we study a metric space in which the given reaction-diffusion equation forms a gradient flow. In the designed metric space, we derive mean-field Hamiltonian flows and Hamilton-Jacobi equations in positive density space. They are derived from mean-field optimal control problems of nonlinear reaction-diffusion equations. We call these equations *mean-field information dynamics*. We then design primal-dual algorithms to compute the proposed dynamics. Numerically, an additional potential energy is also proposed to improve the convexity of the problem. We apply Newton's method to compute the sub-optimization problems in mean-field control problems. Our mean-field control problem provides a variational formulation for implicit schemes of mean-field information dynamics. And the primal-dual hybrid gradient algorithm solves the implicit scheme with a simple component by component update. The flowchart is summarized in the above figure.

Various gradient flows have been studied in [18, 35, 36]; see many references therein. These formulations are motivated and derived from the Onsager principle. The principle is used to derive many evolution equations from soft matter physics and chemistry [39]. This work focuses on both modeling and computational formalism for gradient flows, such as reaction-diffusion equations. We study mean-field control problems and generalized optimal transport metrics for reaction-diffusion equations. By using primal-dual hybrid gradient algorithms, we can efficiently compute the mean-field control problems implicitly in time. It also has a potential to compute reaction-diffusion equations implicit in time; see similar studies in [28].

The paper is organized as follows. In section 2, we briefly review both gradient flows and Hamiltonian flows in a finite-dimensional Euclidean space. In section 3, we introduce the mean-field information distance in positive density space, in which we formulate both gradient flows and Hamiltonian flows in the positive density space. In section 4, we demonstrate several concrete examples of proposed mean-field information dynamics. In section 5, we design the primal-dual hybrid-gradient methods to compute the mean-field information dynamics. It also provides a scheme to compute the mean-field control problem implicitly in time. Several numerical examples are presented.

2. REVIEW

In this section, we review some facts on gradient systems and optimal control in a d -dimensional Euclidean space. We next apply these facts into the infinite-dimensional space.

2.1. Gradient flows. Consider an optimization problem in \mathbb{R}^d

$$\min_{x \in \mathbb{R}^d} f(x),$$

where the function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ is a given smooth objective function. To find the minimizer of function f , consider an initial value dynamical system

$$\frac{dx(t)}{dt} = -g(x(t))^{-1} \nabla f(x(t)), \quad x(0) = x_0, \quad (1)$$

where $g: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ is a given matrix function. In practice, there are several natural choices of matrix functions g .

- (i) If $g(x) = \mathbb{I}$, where \mathbb{I} is an identity matrix. Dynamic (1) forms the gradient flow in Euclidean space.
- (ii) If $g(x) = \nabla^2 f(x)$, where ∇^2 is the Euclidean Hessian operator. Dynamic (1) satisfies Newtonian flow in Euclidean space.

Assume that matrix function g is positive definite. We observe that the objective function f decays along the dynamic (1). In other words,

$$\frac{d}{dt} f(x(t)) = \nabla f(x(t))^\top \frac{dx(t)}{dt} = -\nabla f(x(t))^\top g(x(t))^{-1} \nabla f(x(t)) \leq 0. \quad (2)$$

The above decaying behavior is known as a Lyapunov method, in which the objective function is a “natural” Lyapunov function for equation (1). Dynamic (1) can be viewed as a gradient flow in the metric space (\mathbb{R}^d, g) . The matrix function g is often named the metric tensor. It is also called the matrix operator or the preconditioner matrix.

2.2. Optimal control. In this metric space (\mathbb{R}^d, g) , one often considers the following variational problem. Denote $k: \mathbb{R}^d \rightarrow \mathbb{R}$ as a given smooth potential function and formulate $L: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ as a Lagrangian function:

$$L(x, v) = \frac{1}{2} v^\top g(x) v - k(x).$$

Consider

$$\frac{1}{2} D(x_0, x_1)^2 := \inf_{x: [0,1] \rightarrow \mathbb{R}^d} \int_0^1 L(x(t), \frac{dx(t)}{dt}) dt, \quad (3)$$

where the infimum is taken among all smooth paths $x(t) \in \mathbb{R}^d$, $t \in [0, 1]$ with fixed initial and terminal functions x_0, x_1 . By direct calculations, the Euler-Lagrange equation of problem (3) is formulated below. Denote a Hamiltonian $H: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ as the convex conjugate of L .

$$H(x, p) = \sup_{v \in \mathbb{R}^d} p^\top v - L(x, v) = \frac{1}{2} p^\top g(x)^{-1} p + k(x). \quad (4)$$

where $x \in \mathbb{R}^d$ represents the state variable and $p \in \mathbb{R}^d$ is the momentum variable. The minimizer of variational problem (3) satisfies $x(0) = x_0$, $x(1) = x_1$, with

$$\begin{cases} \frac{dx(t)}{dt} = \nabla_p H(x(t), p(t)), \\ \frac{dp(t)}{dt} = -\nabla_x H(x(t), p(t)). \end{cases} \quad (5)$$

If $k = 0$, equation (5) is called the geodesic equation in metric space (\mathbb{R}^d, g) , and $D(x_0, x_1)$ is the distance function. The above flow defines the characteristics of Hamilton-Jacobi equation. Consider a value function $U: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$, such that

$$\partial_t U(t, x) + H(x, \nabla_x U(t, x)) = 0.$$

And

$$p(t) = \nabla_x U(t, x).$$

2.3. Gradient flows and variational time discretizations. We remark that both gradient flow (1) and Hamiltonian flow (5) are different but connected with each other. One can design a variational implicit scheme for gradient flow (1).

Formally speaking, the gradient flow (1) can be written below:

$$\begin{cases} \frac{dx(t)}{dt} = \nabla_p H(x(t), p(t)), \\ p(t) = -\nabla_x f(x(t)), \end{cases}$$

where H is the quadratic Hamiltonian function defined in (4) with $k = 0$. We observe that the equation $p(t) = -\nabla_x f(x(t))$ does not satisfy the second equation in (5) directly. However, one can construct a time approximation variational scheme, which enforces $p(t) = -\nabla f(x(t))$ at the terminal time.

Denote a time stepsize $\Delta t > 0$. Construct a sequence $\{x_k\}_{k=1}^\infty$ below. Consider an iterative variational sequence

$$\inf_{x: [t_k, t_{k+1}] \rightarrow \mathbb{R}^d} \left\{ \int_{t_k}^{t_{k+1}} L(x(t), \frac{dx(t)}{dt}) dt + f(x(t_{k+1})) : x(t_k) = x_k \right\}. \quad (6)$$

Write

$$x_{k+1} = x(t_{k+1}),$$

where $x(t_{k+1})$ is the minimizer of variational problem (6).

Proposition 1. *The minimizer of variational problem (6) satisfies*

$$\begin{cases} \frac{dx(t)}{dt} = \nabla_p H(x(t), p(t)), & t \in [t_k, t_{k+1}) \\ \frac{dp(t)}{dt} = -\nabla_x H(x(t), p(t)), & t \in [t_k, t_{k+1}), \\ p(t_{k+1}) = -\nabla_x f(x(t_{k+1})). \end{cases} \quad (7)$$

Then $\{x_k\}_{k=1}^\infty$ is a first-order time discretization of gradient flow (1).

Proof of Proposition 1. To see this fact, one can solve the minimization problem (6) and observe that

$$\begin{cases} \frac{dx}{ds} = \nabla_p H(x, p), & \frac{dp}{ds} = -\nabla_x H(x, p), \\ x(t_k) = x_k, & p(t_{k+1}) = -\nabla_x f(x(t_{k+1})). \end{cases}$$

The update forms

$$\begin{aligned}
x_{k+1} &= x(t_{k+1}) = x_k + \int_{t_k}^{t_{k+1}} \frac{dx(t)}{dt} dt \\
&= x_k + \int_{t_k}^{t_{k+1}} \nabla_p H(x(t), p(t)) dt \\
&= x_k + (t_{k+1} - t_k) \nabla_p H(x(t), p(t))|_{t=h} + o(h) \\
&= x_k - hg(x_{k+1})^{-1} \nabla_x f(x_{k+1}) + o(h).
\end{aligned}$$

Hence $\{x_k\}_{k=1}^\infty$ is a backward Euler time discretization of $x(t)$ up to a small order perturbation. We notice that the implicit scheme is a first order time discretization in term of stepsize h . \square

Remark 1. We remark that equation (5), (6) are consequences of Pontryagin's maximum principle. A modification of this connects with gradient flows. These modifications induce implicit schemes to approximate gradient flows.

3. MEAN-FIELD INFORMATION METRIC SPACES AND THEIR DYNAMICS

In this section, we first develop motivations and examples, including reaction diffusion equations and Lyapunov functionals. We next review the mean-field information metric space in positive density space; see [2, 35]. Finally, we formulate both gradient flows and Hamiltonian flows in positive density metric space. We define a class of mean-field control problems.

3.1. Motivation and examples. In this subsection, we review some known facts about reaction-diffusion equations in term of optimal transport type gradient flows. See related studies of diffusion equations in [38], and reaction-diffusion equations [35, 36]. In a word, we can construct metrics and mean-field control problems for some nonlinear reaction-diffusion equations from Lyapunov functionals.

Consider a scalar nonlinear reaction-diffusion equation

$$\partial_t u(t, x) = \Delta F(u(t, x)) + R(u(t, x)), \quad (8)$$

where $x \in \Omega$, $\Omega \subset \mathbb{R}^d$ is compact convex set, $u \in \mathcal{M}(\Omega) = \{u \in C^\infty(\Omega) : u \geq 0\}$, and Δ is the Euclidean Laplacian operator. We assume periodic boundary conditions on the boundary of the spatial domain Ω , and $F, R: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ are smooth functions.

We next construct a Lyapunov functional $\mathcal{G}: \mathcal{M}(\Omega) \rightarrow \mathbb{R}$ to study equation (8). Consider

$$\mathcal{G}(u) = \int G(u(x)) dx,$$

where $G: \mathbb{R} \rightarrow \mathbb{R}$ is a convex function with $G''(u) > 0$. In this case, along the reaction-diffusion equation (8), we observe that

$$\begin{aligned}
\frac{d}{dt}\mathcal{G}(u(t, \cdot)) &= \int G'(u(t, x)) \cdot \partial_t u(t, x) dx \\
&= \int G'(u(t, x))(\Delta F(u(t, x)) + R(u(t, x))) dx \\
&= - \int (\nabla G'(u(t, x)), \nabla F(u(t, x))) dx + \int G'(u(t, x)) R(u(t, x)) dx \\
&= - \int (\nabla G'(u(t, x)), \nabla u(t, x)) F'(u(t, x)) dx + \int G'(u(t, x)) R(u(t, x)) dx \\
&= - \int (\nabla G'(u(t, x)), \nabla u(t, x)) G''(u(t, x)) \frac{F'(u(t, x))}{G''(u(t, x))} dx + \int G'(u(t, x))^2 \frac{R(u(t, x))}{G'(u(t, x))} dx \\
&= - \int (\nabla G'(u(t, x)), \nabla G'(u(t, x))) \frac{F'(u(t, x))}{G''(u(t, x))} dx + \int G'(u(t, x))^2 \frac{R(u(t, x))}{G'(u(t, x))} dx,
\end{aligned}$$

where we apply integration by parts in the third equality and $\nabla G'(u) = G''(u) \nabla u$ in the last equality.

We assume that $R \in C^1(\Omega)$ is a given function with $-\frac{R}{G'} > 0$, and $F'(u) > 0$ for $u > 0$. Under these assumptions, it is clear that

$$\frac{d}{dt}\mathcal{G}(u) \leq 0.$$

This indicates that functional $\mathcal{G}(u)$ is not increasing along flow (8).

In fact, the above decay behavior indicates a gradient flow formulation for dynamics (8). We introduce the following notations. Denote an inverse of the weighted elliptic operator

$$g(u) := \left(-\nabla \cdot \left(\frac{F'(u)}{G''(u)} \nabla \right) - \frac{R(u)}{G'(u)} \right)^{-1}.$$

We have

$$\begin{aligned}
\partial_t u &= -g(u)^{-1} \frac{\delta}{\delta u} \mathcal{G}(u) \\
&= - \left(-\nabla \cdot \left(\frac{F'(u)}{G''(u)} \nabla \right) - \frac{R(u)}{G'(u)} \right) \frac{\delta}{\delta u} \mathcal{G}(u) \\
&= \nabla \cdot \left(\frac{F'(u)}{G''(u)} \nabla G'(u) \right) + \frac{R(u)}{G'(u)} G'(u) \\
&= \Delta F(u) + R(u),
\end{aligned} \tag{9}$$

where $\frac{\delta}{\delta u}$ represents the L^2 first variation w.r.t. $u \in \mathcal{M}(\Omega)$. In the above notation, the dissipation of Lyapunov functional \mathcal{G} along equation (8) satisfies

$$\frac{d}{dt}\mathcal{G}(u) = - \int \left(\frac{\delta}{\delta u} \mathcal{G}(u), g(u)^{-1} \frac{\delta}{\delta u} \mathcal{G}(u) \right) dx \leq 0.$$

Clearly, our assumptions on F , R are sufficient conditions to guarantee that $g(u)$ is a “positive definite” operator.

3.2. Mean-field information metrics and their gradient flows. In this subsection, we illustrate a formal definition of metric space and gradient flows. See details in [2, 34, 36].

Denote a smooth positive density space as

$$\mathcal{M} = \left\{ u \in C^\infty(\Omega) : u > 0 \right\}.$$

Given $F, G: \mathbb{R} \rightarrow \mathbb{R}$ satisfying $F'(u) > 0$ if $u > 0$, and $G''(u) > 0$. Denote

$$V_1(u) = \frac{F'(u)}{G''(u)}, \quad V_2(u) = -\frac{R(u)}{G'(u)}.$$

Denote the tangent space of \mathcal{M} at $u \in \mathcal{M}$ as

$$T_u\mathcal{M} = \left\{ \sigma \in C^\infty(\Omega) \right\}.$$

We define the F, G, R induced metric in the positive density space.

Definition 2 (Mean-field information metric). *The inner product $g(u): T_u\mathcal{M} \times T_u\mathcal{M} \rightarrow \mathbb{R}$ is given below. For any $\sigma_1, \sigma_2 \in T_u\mathcal{M}$, define*

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} \sigma_1 \left(-\nabla \cdot (V_1(u)\nabla) + V_2(u) \right)^{-1} \sigma_2 dx,$$

where

$$\left(-\nabla \cdot (V_1(u)\nabla) + V_2(u) \right)^{-1} : T_u\mathcal{M} \rightarrow T_u\mathcal{M},$$

denotes the inverse operator of weighted elliptic operator $-\nabla \cdot (V_1(u)\nabla) + V_2(u)$. The other formulation of metric is given below. Denote $\Phi_i \in C^\infty(\Omega)$, such that

$$\sigma_i = -\nabla \cdot (V_1(u)\nabla\Phi_i) + V_2(u)\Phi_i, \quad i = 1, 2,$$

Hence the metric satisfies

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} (\nabla\Phi_1, \nabla\Phi_2)V_1(u)dx + \int_{\Omega} \Phi_1\Phi_2V_2(u)dx.$$

Remark 2. We remark that the special case of the above metric is the L^2 -Wasserstein metric, which is well-studied in optimal transport. It corresponds to $V_1 = u, V_2 = 0$. It also contains the Fisher-Rao metric, which is important in information geometry. It corresponds to $V_1 = 0, V_2 = u$. There are several interactive studies of them in unbalanced optimal transport $V_1 = V_2 = u$ [10, 30] and unnormalized optimal transport $V_1 = u, V_2 = 1$ [13, 25]. They are different choices of metric operators $g(u)$, depending on the Lyapunov functional. We call the above metrics *mean-field information metrics*.

We are now ready to formulate gradient flows in (\mathcal{M}, g) .

Proposition 3 (Mean-field information Gradient flow). *Given an energy functional $\mathcal{E}: \mathcal{M} \rightarrow \mathbb{R}$, the gradient flow of \mathcal{E} in $(\mathcal{M}(\Omega), g)$ satisfies*

$$\partial_t u(t, x) = \nabla \cdot (V_1(u)\nabla \frac{\delta}{\delta u} \mathcal{E}(u))(t, x) - V_2(u) \frac{\delta}{\delta u} \mathcal{E}(u)(t, x). \quad (10)$$

If

$$\mathcal{E}(u) = \mathcal{G}(u) = \int G(u)dx,$$

then equation (10) forms the reaction-diffusion equation (8).

Proof. The proof follows the definition. The gradient operator in $(\mathcal{M}(\Omega), g)$ is defined by

$$g(\sigma, \text{grad}\mathcal{E}(u)) = \int \frac{\delta}{\delta u(x)} \mathcal{E}(u) \cdot \sigma(x) dx, \quad \text{for any } \sigma(x) \in T_u \mathcal{M}. \quad (11)$$

In other words,

$$\begin{aligned} \text{grad}\mathcal{E}(u) &= g(u)^{-1} \frac{\delta}{\delta u} \mathcal{F}(u) \\ &= -\nabla \cdot (V_1(u) \nabla \frac{\delta}{\delta u} \mathcal{E}(u)) + V_2(u) \frac{\delta}{\delta u} \mathcal{E}(u), \end{aligned}$$

which finishes the proof. Thus the gradient flow in $(\mathcal{M}(\Omega), g)$ satisfies

$$\partial_t u(t, x) = -\text{grad}\mathcal{E}(u)(t, x) = \nabla \cdot (V_1(u) \nabla \frac{\delta}{\delta u} \mathcal{E}(u)) - V_2(u) \frac{\delta}{\delta u} \mathcal{E}(u).$$

If $\mathcal{E}(u) = \int_{\Omega} G(u) dx$, then

$$\begin{aligned} \partial_t u &= \nabla \cdot (V_1(u) \nabla \frac{\delta}{\delta u} \mathcal{E}(u)) - V_2(u) \frac{\delta}{\delta u} \mathcal{E}(u) \\ &= \nabla \cdot \left(\frac{F'}{G''} \nabla G' \right) + G' \cdot \frac{R}{G'} \\ &= \nabla \cdot \left(\frac{F'}{G''} G'' \nabla u \right) + G' \cdot \frac{R}{G'} \\ &= \Delta F(u) + R(u). \end{aligned}$$

□

We next present the decay of the Lyapunov functional along gradient flow equation (8).

Proposition 4 (Mean-field information De-Brujin identity). *Suppose $u(t, x)$ satisfies (8), then*

$$\frac{d}{dt} \mathcal{G}(u) = -\mathcal{I}(u),$$

where $\mathcal{I}: \mathcal{M}(\Omega) \rightarrow \mathbb{R}$ is a functional:

$$\mathcal{I}(u) = \int_{\Omega} \|\nabla G'(u)\|^2 V_1(u) dx + \int_{\Omega} |G'(u)|^2 V_2(u) dx. \quad (12)$$

Proof. The proof follows from the definition of gradient flow. Note that along the gradient flow (8),

$$\begin{aligned} \frac{d}{dt} \mathcal{G}(u) &= \int_{\Omega} \left(\frac{\delta}{\delta u} \mathcal{G}(u), \partial_t u \right) dx \\ &= \int_{\Omega} (G'(u), \nabla \cdot (V_1(u) \nabla G'(u)) - V_2(u) G'(u)) dx \\ &= - \int_{\Omega} \|\nabla G'(u)\|^2 V_1(u) dx - \int_{\Omega} |G'(u)|^2 V_2(u) dx \\ &= -\mathcal{I}(u), \end{aligned}$$

where the third equality holds following the integration by parts formula. □

Remark 3. We remark that if $G = u \log u$, $V_1 = u$, $V_2 = 0$, then the gradient flow satisfies the heat equation. And the decay of Lyapunov functional along the heat flow satisfies

$$\mathcal{I}(u) = \int_{\Omega} \|\nabla \log u\|^2 u dx.$$

In literature, the relation $\frac{d}{dt}\mathcal{G}(u) = -\mathcal{I}(u)$ is often named the De-Brujin identity. And $\mathcal{I}(u)$ is called the Fisher information functional. Following this spirit, we name the generalized dissipation property “mean-field information De-Brujin identity”. And we call \mathcal{I} the “mean-field information functional”; see examples in [29].

Remark 4. The gradient flow not only works for a scalar function u . One can define a similar metric operator for a vector valued function u ; see examples in [36]. It is worth mentioning that there are more general choices of V_1 , which includes kernel functions; see examples in [6, 11, 27].

Remark 5. In information geometry [1] and its applications in machine learning, the Fisher-Rao gradient flow is known as the natural gradient flow. The Fisher-Rao metric refers to $V_1(u) = 0$, $V_2(u) = u$. This metric and gradient flow has been widely used in machine learning. In addition, the mean-field information gradient flow is the generalization of the “natural gradient” flow. The terminology “natural” corresponds to the “projection” operation. In other words, one projects the infinite dimensional metric space into finite dimensional parameterized models, e.g. neural networks. In this paper, we focus on the infinite dimensional gradient flows, and design classical finite volume methods to solve the related dynamics. We postpone the related AI scientific computing methods in future work. See an initial approach in [31].

3.3. Mean-field information control problems. In this subsection, we state the main variational problem studied in this paper, for which we will design fast numerical methods. We first study the critical point of a variational problem in positive density space. We call the derived system *mean-field information dynamics*.

Definition 5 (Mean-field information control problems). *Denote an energy functional $\mathcal{F}: \mathcal{M}(\Omega) \rightarrow \mathbb{R}$, and write*

$$V_1(u) = \frac{F'(u)G'(u)}{G''(u)}, \quad V_2(u) = -\frac{R(u)}{G'(u)}.$$

Consider a variational problem

$$\inf_{v_1, v_2, u} \int_0^1 \left[\int_{\Omega} \frac{1}{2} \|v_1(t, x)\|^2 V_1(u(t, x)) + \frac{1}{2} |v_2(t, x)|^2 V_2(u(t, x)) dx - \mathcal{F}(u) \right] dt, \quad (13a)$$

where the infimum is taken among all density functions $u: [0, 1] \times \Omega \rightarrow \mathbb{R}$, vector fields $v_1: [0, 1] \times \Omega \rightarrow \mathbb{R}^d$, and reaction rate functions $v_2: [0, 1] \times \Omega \rightarrow \mathbb{R}$, such that

$$\partial_t u(t, x) + \nabla \cdot (V_1(u(t, x))v_1(t, x)) = v_2(t, x)V_2(u(t, x)), \quad (13b)$$

with fixed initial and terminal density functions $u_0, u_1 \in \mathcal{M}(\Omega)$.

We briefly explain variational problem (13) with a modeling perspective. It is a generalized optimal control problem in optimal transport [3, 40] and mean-field control [24, 37]. Suppose an infinite number of identical particles/agents evolve under both transportation

and reaction. The transportation mobility is selected as V_1 , and the reaction mobility is chosen as V_2 . Suppose that the mean-field limit of these particles exists, which satisfies an unnormalized density function. And the evolution of density function satisfies equation (13b). Given two sets of densities u_0, u_1 , what is the optimal way to move or control density u_0 to density u_1 ? The “optimal” is in the sense of the objective functional, which combines transportation and reaction kinetic energies with a potential energy. We notice that the consideration of general reaction mobility functions V_2 has not been considered in mean-field control/game communities [24, 37]. We expect that the proposed variational problems will be useful in controlling reaction-diffusion models, which arise in biology, chemistry, and, recently, social dynamics and pandemic evolution.

We next obtain the critical point for the variational problem (13). Assume that a minimizer for variational problem (13) exists. We formally present the derivation of the critical point.

Proposition 6 (Mean-field information Hamiltonian flows). *Assume $u(t, x) > 0$ for $t \in [0, 1]$. Then there exists a function $\Phi: [0, 1] \times \Omega \rightarrow \mathbb{R}$, such that the critical points of variational problem (13) satisfy*

$$v_1(t, x) = \nabla \Phi(t, x), \quad v_2(t, x) = \Phi(t, x),$$

with

$$\begin{cases} \partial_t u(t, x) + \nabla \cdot (V_1(u(t, x)) \nabla \Phi(t, x)) = V_2(u(t, x)) \Phi(t, x), \\ \partial_t \Phi(t, x) + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 V_1'(u(t, x)) + \frac{1}{2} |\Phi(t, x)|^2 V_2'(u(t, x)) + \frac{\delta}{\delta u} \mathcal{F}(u)(t, x) = 0, \end{cases} \quad (14)$$

and

$$u(0, x) = u_0(x), \quad u(1, x) = u_1(x).$$

Proof. We first rewrite the variables in variational formula (13) as

$$m_1(t, x) = V_1(u)v(t, x), \quad m_2(t, x) = V_2(u)v_2(t, x),$$

Then variational problem (13) forms

$$\begin{aligned} \inf_{m_1, m_2, u} \left\{ \int_0^1 \int_{\Omega} \frac{\|m_1(t, x)\|^2}{2V_1(u(t, x))} + \frac{|m_2(t, x)|^2}{2V_2(u(t, x))} - \mathcal{F}(u) dx dt : \right. \\ \left. \partial_t u(t, x) + \nabla \cdot m_1(t, x) = m_2(t, x), \quad \text{fixed } u_0, u_1 \right\}. \end{aligned} \quad (15)$$

Denote the Lagrange multiplier of problem (15) by Φ . We consider the following saddle point problem

$$\inf_{m_1, m_2, u} \sup_{\Phi} \mathcal{L}(m_1, m_2, u, \Phi),$$

with

$$\begin{aligned} \mathcal{L}(m_1, m_2, u, \Phi) = \int_0^1 \int_{\Omega} \left\{ \frac{\|m_1(t, x)\|^2}{2V_1(u(t, x))} + \frac{|m_2(t, x)|^2}{2V_2(u(t, x))} \right. \\ \left. + \Phi(t, x) \left(\partial_t u(t, x) + \nabla \cdot m_1(t, x) - m_2(t, x) \right) \right\} dx dt. \end{aligned}$$

By finding the saddle point of \mathcal{L} , we have

$$\begin{cases} \frac{\delta}{\delta m_1} \mathcal{L} = 0, \\ \frac{\delta}{\delta m_2} \mathcal{L} = 0, \\ \frac{\delta}{\delta u} \mathcal{L} = 0, \\ \frac{\delta}{\delta \Phi} \mathcal{L} = 0, \end{cases} \Rightarrow \begin{cases} \frac{m_1}{V_1} = \nabla \Phi, \\ \frac{m_2}{V_2} = \Phi, \\ -\frac{1}{2} \frac{\|m_1\|^2}{V_1^2} V_1' - \frac{1}{2} \frac{|m_2|^2}{V_2^2} V_2' - \frac{\delta}{\delta u} \mathcal{F} - \partial_t \Phi = 0, \\ \partial_t u + \nabla \cdot m_1 - m_2 = 0, \end{cases}$$

where $\frac{\delta}{\delta m_1}$, $\frac{\delta}{\delta m_2}$, $\frac{\delta}{\delta u}$, $\frac{\delta}{\delta \Phi}$ are L^2 first variations w.r.t. functions m_1 , m_2 , u , Φ , respectively. Substituting the above two row equations into the last two row equations, we derive the PDE pair (14) in $\mathcal{M}(\Omega)$. \square

Remark 6. If $V_1 = u$, $V_2 = 0$, the above formulation corresponds to the well-known Benamou-Brenier formula [3] in optimal transport.

Remark 7. If V_1, V_2 are positive functions and are convex w.r.t. u , and the functional \mathcal{F} is convex w.r.t. u , then the objective functional of problem (15) is convex. In this case, the derived flow is a minimizer of variational problem (15).

Proposition 7 (Functional Hamilton-Jacobi equations in positive density space). *The Hamilton-Jacobi equation in positive density space satisfies*

$$\partial_t \mathcal{U}(t, u) + \frac{1}{2} \int_{\Omega} \left\| \nabla \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right\|^2 V_1(u) dx + \frac{1}{2} \int_{\Omega} \left| \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right|^2 V_2(u) dx + \mathcal{F}(u) = 0,$$

where $\mathcal{U}: [0, 1] \times L^2(\Omega) \rightarrow \mathbb{R}$ is a value functional.

Proof. Denote the Hamiltonian functional $\mathcal{H}: L^2(\Omega) \times L^2(\Omega) \rightarrow \mathbb{R}$ as

$$\mathcal{H}(u, \Phi) = \int_{\Omega} \left(\frac{1}{2} \|\nabla \Phi\|^2 V_1(u) + \frac{1}{2} |\Phi|^2 V_2(u) \right) dx + \mathcal{F}(u). \quad (16)$$

Then the minimizer system (14) satisfies

$$\partial_t u = \frac{\delta}{\delta \Phi} \mathcal{H}(u, \Phi), \quad \partial_t \Phi = -\frac{\delta}{\delta u} \mathcal{H}(u, \Phi), \quad \text{fixed } u_0, u_1.$$

Here the density function u is the state variable, while potential function Φ is the momentum variable in positive density space. The above flow forms the characteristic equation of Hamilton-Jacobi equation in positive density space. This is true by the fact that

$$\frac{\delta}{\delta u(x)} \mathcal{U}(t, u) = \Phi(t, x).$$

\square

3.4. Variational time discretization. In this section, we present a mean-field control problem, which gives a first order accuracy in small time interval limit for the reaction-diffusion equation (8).

From now on, we consider

$$\mathcal{H}(u, \Phi) = \frac{1}{2} \int_{\Omega} \left[\|\nabla \Phi\|^2 V_1(u) + |\Phi|^2 V_2(u) \right] dx. \quad (17)$$

Proposition 8. *The reaction-diffusion equation (8) can be formulated as*

$$\begin{cases} \partial_t u(t, x) = \frac{\delta}{\delta \Phi(x)} \mathcal{H}(u, \Phi) = -\nabla \cdot (V_1(u(t, x)) \nabla \Phi(t, x)) + V_2(u(t, x)) \Phi(t, x), \\ \Phi(t, x) = -\frac{\delta}{\delta u(x)} \mathcal{G}(u) = -G'(u(t, x)). \end{cases}$$

Proof. The proof is based on a direct calculation. Notice

$$\frac{\delta}{\delta \Phi} \mathcal{H}(u, \Phi) = -\nabla \cdot (V_1 \nabla \Phi) + V_2 \Phi.$$

Hence

$$\begin{aligned} \frac{\delta}{\delta \Phi} \mathcal{H}(u, \Phi)|_{\Phi=-G'} &= \nabla \cdot (V_1 \nabla G') - V_2 G' \\ &= \nabla \cdot \left(V_1 \frac{F' G''}{V_1} \nabla u \right) + \frac{R}{G'} G' \\ &= \Delta F + R. \end{aligned}$$

The second equality follows from the definition of V_1 , V_2 . \square

We notice that proposition 8 is useful in designing a variational implicit time discretization. It is the mean-field control generalization of Jordan-Kinderlehrer-Otto (JKO) scheme [5, 21], where they select $V_1(u) = u$, $V_2(u) = 0$, and $\mathcal{F}(u) = 0$. Similarly, we consider an iterative sequence of variational problems, which approximates equation (8) sequentially, in each time interval $[t_k, t_{k+1}]$.

Definition 9 (Iterative variational formulations for reaction-diffusion equations). *Denote a time stepsize as $\Delta t > 0$, and $t_k = k\Delta t$, $k = 0, 1, 2, \dots$, and $u_0(x) = u(0, x)$. Consider the following iterative variational problem*

$$\inf_{v_1, v_2, u(\cdot, \cdot), u_{k+1}} \int_{t_k}^{t_{k+1}} \left\{ \int_{\Omega} \frac{1}{2} \|v_1(t, x)\|^2 V_1(u(t, x)) + \frac{1}{2} |v_2(t, x)|^2 V_2(u(t, x)) dx \right\} dt + \mathcal{G}(u(t_{k+1}, \cdot)), \quad (18)$$

where the infimum is taken among all density functions $u: [t_k, t_{k+1}] \times \Omega \rightarrow \mathbb{R}$, vector fields $v_1: [t_k, t_{k+1}] \times \Omega \rightarrow \mathbb{R}^d$, and reaction functions $v_2: [t_k, t_{k+1}] \times \Omega \rightarrow \mathbb{R}$, such that

$$\partial_t u(t, x) + \nabla \cdot (V_1(u(t, x)) v_1(t, x)) = v_2(t, x) V_2(u(t, x)),$$

with a fixed initial value function $u(t_k, x) = u_k(x)$ and a terminal energy functional $\mathcal{G}(u_{k+1})$. Denote the update as

$$u_{k+1}(x) = u(t_{k+1}, x), \quad k = 1, 2, 3, \dots$$

where $u(t_{k+1}, x)$ is the minimizer for variational problem (18).

Proposition 10. *Consider the minimizer of variational problem (18):*

$$\begin{cases} \partial_t u + \nabla \cdot (V_1 \nabla \Phi) = V_2 \Phi, & t \in [t_k, t_{k+1}), \\ \partial_t \Phi + \frac{1}{2} \|\nabla \Phi\|^2 V_1' + |\Phi|^2 V_2' = 0, & t \in [t_k, t_{k+1}), \\ u(t_k, x) = u_k(x), \quad \Phi(t_{k+1}, x) = -\frac{\delta}{\delta u(x)} \mathcal{G}(u)|_{t=t_{k+1}} = -G'(u_{k+1}). \end{cases} \quad (19)$$

Then $\{u_k\}_{k=1}^\infty$ in (19) approximates the reaction-diffusion equation (8) with the first order accuracy in time.

Proof. As the proof in Proposition 17, we derive the minimizer system for variational problem (18). We note that the minimizer system follows the Pontryagin maximum principle. Again, denote the Lagrangian multiplier of problem (18) as Φ . We consider the following saddle point problem

$$\inf_{m_1, m_2, u(t, \cdot), u_{k+1}} \sup_{\Phi} \mathcal{L}_1(m_1, m_2, u, \Phi),$$

where

$$\begin{aligned} \mathcal{L}_1(m_1, m_2, u, \Phi, u_{k+1}) = & \int_{t_k}^{t_{k+1}} \int_{\Omega} \left\{ \frac{\|m_1(t, x)\|^2}{2V_1(u(t, x))} + \frac{|m_2(t, x)|^2}{2V_2(u(t, x))} \right. \\ & \left. + \Phi(t, x) \left(\partial_t u(t, x) + \nabla \cdot m_1(t, x) - m_2(t, x) \right) \right\} dx dt + \mathcal{G}(u_{k+1}). \end{aligned}$$

Similarly, by finding the saddle point of \mathcal{L}_1 , we have

$$\begin{cases} \frac{\delta}{\delta m_1} \mathcal{L}_1 = 0, & \frac{\delta}{\delta m_2} \mathcal{L}_1 = 0, \\ \frac{\delta}{\delta u} \mathcal{L}_1 = 0, \\ \frac{\delta}{\delta \Phi} \mathcal{L}_1 = 0, \\ \frac{\delta}{\delta u_{k+1}} \mathcal{G}(u_{k+1}) = 0, \end{cases} \Rightarrow \begin{cases} \frac{m_1}{V_1} = \nabla \Phi, & \frac{m_2}{V_2} = \Phi, \\ -\frac{1}{2} \frac{\|m_1\|^2}{V_1^2} V_1' - \frac{1}{2} \frac{|m_2|^2}{V_2^2} V_2' - \partial_t \Phi = 0, \\ \partial_t u + \nabla \cdot m_1 - m_2 = 0, \\ \Phi(t_{k+1}, x) + \frac{\delta}{\delta u_{k+1}(x)} \mathcal{G}(u_{k+1}) = 0. \end{cases}$$

In other words, we have

$$\begin{cases} \partial_t u = \frac{\delta}{\delta \Phi} \mathcal{H}(u, \Phi), & \partial_t \Phi = -\frac{\delta}{\delta u} \mathcal{H}(u, \Phi), \\ u_0(x) = u_k(x), & \Phi(t_{k+1}, x) = -\frac{\delta}{\delta u_{k+1}(x)} \mathcal{G}(u_{k+1}), \end{cases}$$

where

$$\mathcal{H}(u, \Phi) = \frac{1}{2} \int_{\Omega} \left[\|\nabla \Phi\|^2 V_1(u) + |\Phi|^2 V_2(u) \right] dx.$$

We notice that the sequence $\{u_k\}_{k=1}^\infty$ forms an approximation for reaction-diffusion (8):

$$\begin{aligned}
u_{k+1}(x) &= u_0(x) + \int_{t_k}^{t_{k+1}} \partial_t u(t, x) dt \\
&= u_k(x) + \int_{t_k}^{t_{k+1}} \frac{\delta}{\delta \Phi(x)} \mathcal{H}(u(t, x), \Phi(t, x)) dt \\
&= u_k(x) + (t_{k+1} - t_k) \cdot \frac{\delta}{\delta \Phi(x)} \mathcal{H}(u, \Phi)|_{\Phi=\Phi(t_{k+1}, x)} + o(\Delta t) \\
&= u_k(x) + \Delta t \cdot \frac{\delta}{\delta \Phi(x)} \mathcal{H}(u, \Phi)|_{\Phi=-\frac{\delta}{\delta u_{k+1}} \mathcal{G}(u_{k+1})} + o(\Delta t).
\end{aligned}$$

The above update is a time discretization for equation (8), which is true for a small order time increment Δt . We finish the derivation. \square

Remark 8. We remark that equation (14), (19) are again consequences of Pontryagin maximum principles. They are generalizations of equations in optimal transport and mean field control/game problems.

Remark 9. We note that equation (19) is different from equation (8). In gradient flow (8), Φ is chosen as the L^2 gradient of Lyapunov functional \mathcal{G} , while in (19), Φ satisfies a dual equation. However, if we intentionally “ignore” the equation (19) of Φ and keep solving the equation (19) of u , we obtain a time approximation scheme for reaction-diffusion equations. In this way, we let the terminal condition $\Phi(t_{k+1}) = -G'(u(t_{k+1}))$ enter the system. And the first equation of system (19) does approximate the original reaction-diffusion equation.

Remark 10. If there is a Lyapunov functional \mathcal{G} and functional \mathcal{H} , such that variational problem (18) becomes a convex optimization. We can develop a convex optimization method to approximate reaction-diffusion equation implicitly in time. We leave these careful studies and computations for future work. In the numerical section of this paper, we develop a new and efficient algorithm for solving problem (13).

Remark 11. Variational problem (18) can be viewed as a generalized Moreau envelope problem in (\mathcal{M}, g) . Consider

$$u_{k+1} = \arg \inf_{u \in \mathcal{M}} \frac{1}{2h} \text{Dist}(u_k, u)^2 + \mathcal{G}(u).$$

Here the distance functional $\text{Dist}(u_k, u)^2$ is the value function in variational problem (13), where we select $\mathcal{F} = 0$. In detail,

$$\text{Dist}(u_k, u)^2 := \inf_{v_1, v_2, u} \int_0^1 \left[\int_{\Omega} \|v_1(t, x)\|^2 V_1(u(t, x)) + |v_2(t, x)|^2 V_2(u(t, x)) \right] dx dt,$$

where the infimum is taken among u, v_1, v_2 , such that

$$\partial_t u(t, x) + \nabla \cdot (V_1(u(t, x))v_1(t, x)) = v_2(t, x)V_2(u(t, x)),$$

with fixed initial and terminal density functions u_0, u_1 .

Remark 12. We remark that the implicit time discretizations of gradient flows are not unique. There are many other semi-implicit variational time discretization methods, e.g. Crank–Nicolson algorithm [5, 18].

Remark 13. It is worth mentioning that the variational scheme could be useful in computing “nonlinear” reaction-diffusion equations, when there exists a non-quadratic functional \mathcal{H} . See examples in [2, 37].

4. EXAMPLES

In this section, we list several examples. They are designed by using both Lyapunov functionals and reaction-diffusion equations. From now on, we also study an additional energy functional \mathcal{F} for the mean-field control problem (15). We shall design numerical schemes for them using primal-dual hybrid gradient methods.

Example 1 (Wasserstein metric and heat flow). *Let*

$$G(u) = u \log u - 1, \quad F(u) = u, \quad R(u) = 0,$$

thus

$$V_1(u) = \frac{F'(u)}{G''(u)} = u, \quad V_2(u) = -\frac{R(u)}{G'(u)} = 0.$$

The metric forms

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} (\nabla \Phi_1(x), \nabla \Phi_2(x)) u(x) dx,$$

with $\sigma_i = -\nabla \cdot (u \nabla \Phi_i)$, $i = 1, 2$. In this case, the mean-field information metric coincides with the Wasserstein-2 metric [2, 14, 38, 40]. The gradient flow of $\mathcal{G}(u)$, named negative Boltzmann-Shannon entropy, in $(\mathcal{M}(\Omega), g)$ forms the heat equation, i.e.

$$\partial_t u = \nabla \cdot (u \nabla G'(u)) = \nabla \cdot (u G''(u) \nabla u) = \Delta u.$$

The dissipation of $\mathcal{G}(u)$ forms

$$\mathcal{I}(u) = \int_{\Omega} \|\nabla \log u(x)\|^2 u(x) dx.$$

And the Hamilton-Jacobi equation in $(\mathcal{M}(\Omega), g)$ follows

$$\partial_t \mathcal{U}(t, u) + \frac{1}{2} \int_{\Omega} \left\| \nabla \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right\|^2 u(x) dx + \mathcal{F}(u) = 0.$$

Its “characteristics” in $(\mathcal{M}(\Omega), g)$ satisfy

$$\begin{cases} \partial_t u + \nabla \cdot (u \nabla \Phi) = 0, \\ \partial_t \Phi + \frac{1}{2} \|\nabla \Phi\|^2 + \frac{\delta}{\delta u} \mathcal{F}(u) = 0. \end{cases}$$

Example 2 (Generalized Wasserstein metric and nonlinear heat flow). *Choose functions F , G , R , such that*

$$\frac{F'(u)}{G''(u)} = u^\alpha, \quad R(u) = 0,$$

where $\alpha \in \mathbb{R}$. The metric forms

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} (\nabla \Phi_1(x), \nabla \Phi_2(x)) u^\alpha(x) dx,$$

with $\sigma_i = -\nabla \cdot (u^\alpha \nabla \Phi_i)$, $i = 1, 2$. The gradient flow of \mathcal{G} in $(\mathcal{M}(\Omega), g)$ forms

$$\partial_t u = \nabla \cdot (V_1(u) \nabla G'(u)) = \nabla \cdot \left(\frac{F'(u)}{G''(u)} G''(u) \nabla u \right) = \nabla \cdot (F'(u) \nabla u) = \Delta F(u),$$

and the dissipation of $\mathcal{G}(u)$ satisfies

$$\mathcal{I}(u) = \int_{\Omega} \|\nabla G'(u)\|^2 u^\alpha dx.$$

And the Hamilton-Jacobi equation in $(\mathcal{M}(\Omega), g)$ follows

$$\partial_t \mathcal{U}(t, u) + \frac{1}{2} \int_{\Omega} \left\| \nabla \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right\|^2 u^\alpha(x) dx + \mathcal{F}(u) = 0.$$

Its “characteristics” in $(\mathcal{M}(\Omega), g)$ satisfy

$$\begin{cases} \partial_t u + \nabla \cdot (u^\alpha \nabla \Phi) = 0, \\ \partial_t \Phi + \frac{\alpha}{2} \|\nabla \Phi\|^2 u^{\alpha-1} + \frac{\delta}{\delta u} \mathcal{F}(u) = 0. \end{cases}$$

Example 3 (H^{-1} metric and nonlinear heat flow). Consider $\alpha = 0$ in the above example. We choose functions F , G , R , such that

$$\frac{F'(u)}{G''(u)} = 1, \quad R(u) = 0.$$

In this case,

$$V_1(u) = 1, \quad V_2(u) = 0.$$

The metric forms

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} (\nabla \Phi_1(x), \nabla \Phi_2(x)) dx,$$

with $\sigma_i = -\nabla \cdot (\nabla \Phi_i)$, $i = 1, 2$. The gradient flow of \mathcal{G} in $(\mathcal{M}(\Omega), g)$ satisfies

$$\partial_t u = \Delta F(u),$$

and the dissipation of $\mathcal{G}(u)$ satisfies

$$\mathcal{I}(u) = \int_{\Omega} \|\nabla G'(u)\|^2 dx.$$

And the Hamilton-Jacobi equation in $(\mathcal{M}(\Omega), g)$ follows

$$\partial_t \mathcal{U}(t, u) + \frac{1}{2} \int_{\Omega} \left\| \nabla \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right\|^2 dx + \mathcal{F}(u) = 0.$$

Its “characteristics” in $(\mathcal{M}(\Omega), g)$ satisfy

$$\begin{cases} \partial_t u + \nabla \cdot (\nabla \Phi) = 0, \\ \partial_t \Phi + \frac{\delta}{\delta u} \mathcal{F}(u) = 0. \end{cases}$$

Example 4 (Fisher-Rao metric and birth-death equation). Consider

$$F(u) = 0, \quad G(u) = u \log u - u, \quad R(u) = -u \log u,$$

then

$$V_1(u) = \frac{F'(u)}{G''(u)} = 0, \quad V_2(u) = -\frac{R(u)}{G'(u)} = u.$$

The metric satisfies

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} \Phi_1(x) \Phi_2(x) u(x) dx,$$

with $\sigma_i(x) = \Phi_i u$, $i = 1, 2$. In this case, the mean-field information metric forms the Fisher-Rao metric in positive density space; see information geometry [1]. The gradient flow of \mathcal{G} in $(\mathcal{M}(\Omega), g)$ satisfies the birth-death dynamics

$$\partial_t u = -V_2(u) G'(u) = -V_2(u) \log u = -u \log u.$$

And the dissipation of $\mathcal{G}(u)$ forms

$$\mathcal{I}(u) = \int_{\Omega} |\log u(x)|^2 u(x) dx.$$

And the Hamilton-Jacobi equation in $(\mathcal{M}(\Omega), g)$ follows

$$\partial_t \mathcal{U}(t, u) + \frac{1}{2} \int_{\Omega} \left| \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right|^2 u(x) dx + \mathcal{F}(u) = 0.$$

Its “characteristics” in $(\mathcal{M}(\Omega), g)$ satisfy

$$\begin{cases} \partial_t u - u \Phi = 0, \\ \partial_t \Phi + \frac{1}{2} |\Phi|^2 + \frac{\delta}{\delta u} \mathcal{F}(u) = 0. \end{cases}$$

Example 5. Consider

$$F(u) = u, \quad G(u) = u \log u - u, \quad R(u) = -u^\alpha \log u,$$

where $\alpha \in \mathbb{R}$ is a given value. In this case,

$$V_1(u) = \frac{F'(u)}{G''(u)} = u, \quad V_2(u) = -\frac{R(u)}{G'(u)} = u^\alpha.$$

The metric forms

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} (\nabla \Phi_1(x), \nabla \Phi_2(x)) u(x) dx + \int_{\Omega} \Phi_1(x) \Phi_2(x) u(x)^\alpha dx,$$

with $\sigma_i = -\nabla \cdot (u \nabla \Phi_i) + \Phi_i u^\alpha$, $i = 1, 2$. The gradient flow of \mathcal{G} in $(\mathcal{M}(\Omega), g)$ forms

$$\partial_t u = \Delta u + u^\alpha \log u,$$

and the dissipation of $\mathcal{G}(u)$ satisfies

$$\mathcal{I}(u) = \int_{\Omega} \|\nabla \log u\|^2 u dx + \int_{\Omega} |\log u|^2 u^\alpha dx.$$

And the Hamilton-Jacobi equation in $(\mathcal{M}(\Omega), g)$ follows

$$\partial_t \mathcal{U}(t, u) + \frac{1}{2} \int_{\Omega} \left\| \nabla \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right\|^2 u(x) dx + \frac{1}{2} \int_{\Omega} \left| \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right|^2 u(x)^\alpha dx + \mathcal{F}(u) = 0.$$

Its “characteristics” in $(\mathcal{M}(\Omega), g)$ satisfy

$$\begin{cases} \partial_t u + \nabla \cdot (u \nabla \Phi) = \Phi u^\alpha, \\ \partial_t \Phi + \frac{1}{2} \|\nabla \Phi\|^2 + \alpha \Phi u^{\alpha-1} + \frac{\delta}{\delta u} \mathcal{F}(u) = 0. \end{cases}$$

Example 6 (Constant regularized optimal transport metric). *Consider*

$$F(u) = u, \quad G(u) = (u+1)\log(u+1), \quad R(u) = 0.$$

Thus

$$V_1(u) = \frac{F'(u)}{G''(u)} = u+1, \quad V_2(u) = 0.$$

The metric forms

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} (\nabla \Phi_1, \nabla \Phi_2)(u+1)dx,$$

with $\sigma_i = -\nabla \cdot ((u+1)\nabla \Phi_i)$, $i = 1, 2$. The gradient flow of \mathcal{G} in $(\mathcal{M}(\Omega), g)$ satisfies

$$\partial_t u(t, x) = \Delta u(t, x).$$

And the dissipation of $\mathcal{G}(u)$ satisfies

$$\mathcal{I}(u) = \int_{\Omega} \|\nabla \log(u+1)\|^2 (u+1)dx.$$

And the Hamilton-Jacobi equation in $(\mathcal{M}(\Omega), g)$ follows

$$\partial_t \mathcal{U}(t, u) + \frac{1}{2} \int_{\Omega} \left\| \nabla \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right\|^2 (u+1)dx + \mathcal{F}(u) = 0.$$

Its “characteristics” in $(\mathcal{M}(\Omega), g)$ satisfy

$$\begin{cases} \partial_t u + \nabla \cdot ((u+1)\nabla \Phi) = 0, \\ \partial_t \Phi + \frac{1}{2} \|\nabla \Phi\|^2 + \frac{\delta}{\delta u} \mathcal{F}(u) = 0. \end{cases}$$

Example 7 (Fisher-KPP metric and Fisher-KPP equation). *Consider the Fisher-KPP equation*

$$\partial_t u = u(1-u) + \Delta u.$$

Consider

$$F(u) = u, \quad G(u) = u \log u - u, \quad R(u) = u(1-u).$$

Thus

$$V_1(u) = \frac{F'(u)}{G''(u)} = u, \quad V_2(u) = -\frac{R(u)}{G'(u)} = \frac{u(u-1)}{\log u}.$$

The metric forms

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} (\nabla \Phi_1, \nabla \Phi_2) u dx + \int_{\Omega} \Phi_1 \Phi_2 \frac{u(u-1)}{\log u} dx,$$

with $\sigma_i = -\nabla \cdot (u \nabla \Phi_i) + \frac{u(u-1)}{\log u} \Phi_i$, $i = 1, 2$. The gradient flow of \mathcal{G} in $(\mathcal{M}(\Omega), g)$ satisfies the Fisher-KPP equation

$$\partial_t u(t, x) = u(1-u) + \Delta u.$$

And the dissipation of $\mathcal{G}(u)$ satisfies

$$\mathcal{I}(u) = \int_{\Omega} \|\nabla \log u\|^2 u dx + \int_{\Omega} |\log u|^2 \frac{u(u-1)}{\log u} dx.$$

And the Hamilton-Jacobi equation in $(\mathcal{M}(\Omega), g)$ follows

$$\partial_t \mathcal{U}(t, u) + \frac{1}{2} \int_{\Omega} \left\| \nabla \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right\|^2 u dx + \frac{1}{2} \int_{\Omega} \left| \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right|^2 \frac{u(u-1)}{\log u} dx + \mathcal{F}(u) = 0.$$

Its “characteristics” in $(\mathcal{M}(\Omega), g)$ satisfy

$$\begin{cases} \partial_t u + \nabla \cdot (u \nabla \Phi) - \frac{u(u-1)}{\log u} \Phi = 0, \\ \partial_t \Phi + \frac{1}{2} \|\nabla \Phi\|^2 + |\Phi|^2 \frac{(2u-1) \log u + 1 - u}{(\log u)^2} + \frac{\delta}{\delta u} \mathcal{F}(u) = 0. \end{cases}$$

Example 8 (Allen-Cahn metric and Allen-Cahn equation). *Let $f \in C^2(\mathbb{R})$ be a given function. Consider*

$$F(u) = u, \quad G(u) = f(u), \quad R(u) = -f'(u).$$

Thus

$$V_1(u) = \frac{F'(u)}{G''(u)} = f''(u)^{-1}, \quad V_2(u) = -\frac{R(u)}{G'(u)} = 1.$$

The metric forms

$$g(u)(\sigma_1, \sigma_2) = \int_{\Omega} (\nabla \Phi_1, \nabla \Phi_2) f''(u)^{-1} dx + \int_{\Omega} \Phi_1 \Phi_2 dx,$$

with $\sigma_i = -\nabla \cdot (f''(u)^{-1} \nabla \Phi_i) + \Phi_i$, $i = 1, 2$. The gradient flow of \mathcal{G} in $(\mathcal{M}(\Omega), g)$ satisfies

$$\partial_t u(t, x) = \Delta u(t, x) - f'(u(t, x)).$$

And the dissipation of $\mathcal{G}(u)$ satisfies

$$\mathcal{I}(u) = \int_{\Omega} \|\nabla f'(u)\|^2 f''(u)^{-1} dx + \int_{\Omega} |f'(u)|^2 dx.$$

And the Hamilton-Jacobi equation in $(\mathcal{M}(\Omega), g)$ follows

$$\partial_t \mathcal{U}(t, u) + \frac{1}{2} \int_{\Omega} \left\| \nabla \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right\|^2 f''(u(x))^{-1} dx + \frac{1}{2} \int_{\Omega} \left| \frac{\delta}{\delta u(x)} \mathcal{U}(t, u) \right|^2 dx + \mathcal{F}(u) = 0.$$

Its “characteristics” in $(\mathcal{M}(\Omega), g)$ satisfy

$$\begin{cases} \partial_t u + \nabla \cdot (f''(u)^{-1} \nabla \Phi) - \Phi = 0, \\ \partial_t \Phi - \frac{1}{2} \|\nabla \Phi\|^2 \frac{f'''(u)}{f''(u)^2} + \frac{\delta}{\delta u} \mathcal{F}(u) = 0. \end{cases}$$

5. ALGORITHMS

In this section, we propose an algorithm to solve the mean-field information variational problem (15) in two dimensions. Section 5.1 presents the main optimization tool we use to solve the variational problem. We use the primal-dual hybrid gradient (PDHG) algorithm [7, 8], which is a popular first-order optimization method to solve saddle point problems. More specifically, we use the general proximal primal dual hybrid gradient (G-prox PDHG) method from [20], which is a variation of the PDHG algorithm with a precondition matrix. Section 5.2 shows the implementation of G-Prox PDHG algorithm to solve the variational problem. Section 5.3 provides additional algorithm when V_1 and V_2 in (15) are affine functions with specific forms (see (33)). In Section 5.4, we give details of the discretization of the algorithms to solve the variational problem on a compact set in 2-dimensional space. This section also shows the solution of the algorithm is equivalent

to the solution of an implicit finite difference scheme that is stable and convergent for all ratios of Δt and Δx .

5.1. PDHG for mean-field control problems. We first review the PDHG algorithm. Consider the following convex optimization problem.

$$\min_z f(Az) + g(z),$$

where z is a variable to be minimized, f and g are convex functions and A is a linear operator. Recall the Legendre transform f^* of f is

$$f^*(p) = \sup_z \langle z, p \rangle - f(z).$$

It is well-known that if f is convex then $f^{**} = f$. Thus, we have

$$f(w) = f^{**}(w) = \sup_p \langle w, p \rangle - f^*(p).$$

Using this property of convex functions, the minimization problem can be converted to a saddle point problem

$$\min_z \max_p g(z) + \langle Az, p \rangle - f^*(p) =: \mathcal{L}(z, p) \quad (20)$$

where \mathcal{L} is a Lagrangian functional. The PDHG algorithm solves the problem by iterating

$$\begin{aligned} p^{(k+1)} &= \operatorname{argmax}_p \mathcal{L}(z^{(k+1)}, p) - \frac{1}{2\sigma} \|p - p^{(k)}\|_{L^2}^2 \\ z^{(k+1)} &= \operatorname{argmin}_z \mathcal{L}(z, 2p^{(k+1)} - p^{(k)}) + \frac{1}{2\tau} \|z - z^{(k)}\|_{L^2}^2. \end{aligned} \quad (21)$$

The scheme converges if the step sizes τ and σ satisfy

$$\tau\sigma \|A^T A\|_{L^2} < 1, \quad (22)$$

where $\|\cdot\|_{L^2}$ is the operator norm in L^2 . G-Prox PDHG provides an appropriate choice of norms for the algorithm and the authors prove that choosing the proper norms allows the algorithm to have larger step sizes and faster convergence than the original PDHG algorithm. The G-prox PDHG iterates

$$\begin{aligned} p^{(k+1)} &= \operatorname{argmax}_p \mathcal{L}(z^{(k+1)}, p) - \frac{1}{2\sigma} \|p - p^{(k)}\|_H^2, \\ z^{(k+1)} &= \operatorname{argmin}_z \mathcal{L}(z, 2p^{(k+1)} - p^{(k)}) + \frac{1}{2\tau} \|z - z^{(k)}\|_{L^2}^2. \end{aligned} \quad (23)$$

Note that the norm in the first line is changed to H from L^2 . The norm $\|\cdot\|_H$ is defined as

$$\|p\|_H^2 = \|A^\top p\|_{L^2}^2.$$

For example, in our problem, we define $z = (m_1, m_2, u)$ as a vector of functions, $p = \Phi$, and the linear operator A as

$$A(m_1, m_2, u)(t, x) = \partial_t u(t, x) + \nabla \cdot m_1(t, x) - m_2(t, x). \quad (24)$$

Note that the differential operator ∂_t and the divergence operator $\nabla \cdot$ are linear operators, which make A a linear operator. We define the inner product as

$$\langle z_1, z_2 \rangle = \int_0^1 \int_{\Omega} (m_1)_1(t, x) \cdot (m_1)_2(t, x) + (m_2)_1(t, x)(m_2)_2(t, x) + u_1(t, x)u_2(t, x) dx dt$$

where $z_i = ((m_1)_i, (m_2)_i, u_i)$ for $i = 1, 2$, and

$$\langle p_1, p_2 \rangle = \int_0^1 \int_{\Omega} \Phi_1(t, x)\Phi_2(t, x) dx dt.$$

where $p_i = \Phi_i$ for $i = 1, 2$. Thus, we have

$$\langle Az, p \rangle = \int_0^1 \int_{\Omega} \Phi(t, x) \left(\partial_t u(t, x) + \nabla \cdot m_1(t, x) - m_2(t, x) \right) dx dt.$$

These inner products induce the L^2 norm of z and p , such that

$$\begin{aligned} \|z\|_{L^2}^2 &= \int_0^1 \int_{\Omega} |m_1(t, x)|^2 + m_2(t, x)^2 + u(t, x)^2 dx dt \\ \|p\|_{L^2}^2 &= \int_0^1 \int_{\Omega} \Phi^2(t, x) dx dt \end{aligned} \tag{25}$$

and the H norm of p can be written as

$$\|p\|_H^2 = \int_0^1 \int_{\Omega} (\partial_t \Phi(t, x))^2 + |\nabla \Phi(t, x)|^2 + (\Phi(t, x))^2 dx dt \tag{26}$$

where the A^\top is computed using integration by parts. With abuse of notation, we also define the L^2 norm of m_1 , m_2 , and u as follows

$$\begin{aligned} \|m_1\|_{L^2}^2 &= \int_0^1 \int_{\Omega} |m_1(t, x)|^2 dx dt \\ \|m_2\|_{L^2}^2 &= \int_0^1 \int_{\Omega} m_2(t, x)^2 dx dt \\ \|u\|_{L^2}^2 &= \int_0^1 \int_{\Omega} u(t, x)^2 dx dt \end{aligned} \tag{27}$$

By choosing such norm based on [20], the step sizes of the algorithm only need to satisfy

$$\sigma\tau < 1,$$

which is independent of the operator A . From the definition of A in (24), the operator involves an unbounded operator ∇ . Thus, this step size condition in PDHG allows us to run the algorithm with larger step sizes independent of the grid sizes.

5.2. Implementations of the algorithm. To implement the algorithm in the variational problem (15), we define z , p , and the linear operator A as above. Furthermore, we set convex functionals g as

$$g(m_1, m_2, u) = \int_0^1 \int_{\Omega} \frac{|m_1(t, x)|^2}{2V_1(u(t, x))} + \frac{m_2(t, x)^2}{2V_2(u(t, x))} dx - \mathcal{F}(u(t, \cdot)) dt + \mathcal{G}(u(1, \cdot)) \tag{28}$$

where the functional \mathcal{F} is of the form

$$\mathcal{F}(u(t, \cdot)) = - \int_{\Omega} s(u(t, x)) dx$$

and the terminal functional \mathcal{G} is of the form

$$\mathcal{G}(u(1, \cdot)) = \int_{\Omega} G(u(1, x)) dx$$

where $s, G : \mathbb{R} \rightarrow \mathbb{R}$ are convex functions. Set a convex functional f as

$$\begin{aligned} f(w) &= \int_0^1 \int_{\Omega} i_{\infty}(w(t, x)) dx dt \\ i_{\infty}(t) &= \begin{cases} 0 & \text{if } t = 0 \\ \infty & \text{otherwise.} \end{cases} \end{aligned} \tag{29}$$

From the definition of function f , we can compute the Legendre transform of f

$$f^*(\Phi) = \sup_w \int_0^1 \int_{\Omega} \Phi(t, x) w(t, x) - i_{\infty}(w(t, x)) dx dt = 0.$$

Thus, from the definition of a Lagrangian functional in (30), we have

$$\begin{aligned} \mathcal{L}(m_1, m_2, u, \Phi) &= g(m_1, m_2, u) + \langle A(m_1, m_2, u), \Phi \rangle - f^*(\Phi) \\ &= \int_0^1 \int_{\Omega} \left\{ \frac{|m_1(t, x)|^2}{2V_1(u(t, x))} + \frac{m_2(t, x)^2}{2V_2(u(t, x))} \right. \\ &\quad \left. + \Phi(t, x) \left(\partial_t u(t, x) + \nabla \cdot m_1(t, x) - m_2(t, x) \right) \right\} dx - \mathcal{F}(u(t, \cdot)) dt + \mathcal{G}(u(1, \cdot)). \end{aligned} \tag{30}$$

G-Prox PDHG computes the saddle point $(m_1^*, m_2^*, u^*, \Phi^*)$ by iterating

$$\begin{aligned} \Phi^{(k+1)} &= \operatorname{argmax}_{\Phi: [0,1] \times \Omega \rightarrow \mathbb{R}} \mathcal{L}(m_1^{(k)}, m_2^{(k)}, u^{(k)}, \Phi) - \frac{1}{2\sigma} \|\Phi - \Phi^{(k)}\|_H^2 \\ m_1^{(k+1)} &= \operatorname{argmin}_{m_1: [0,1] \times \Omega \rightarrow \mathbb{R}} \mathcal{L}(m_1, m_2^{(k)}, u^{(k)}, 2\Phi^{(k+1)} - \Phi^{(k)}) + \frac{1}{2\tau} \|m_1 - m_1^{(k)}\|_{L^2}^2 \\ m_2^{(k+1)} &= \operatorname{argmin}_{m_2: [0,1] \times \Omega \rightarrow \mathbb{R}^d} \mathcal{L}(m_1^{(k)}, m_2, u^{(k)}, 2\Phi^{(k+1)} - \Phi^{(k)}) + \frac{1}{2\tau} \|m_2 - m_2^{(k)}\|_{L^2}^2 \\ u^{(k+1)} &= \operatorname{argmin}_{u: [0,1] \times \Omega \rightarrow \mathbb{R}} \mathcal{L}(m_1^{(k+1)}, m_2^{(k+1)}, u, 2\Phi^{(k+1)} - \Phi^{(k)}) + \frac{1}{2\tau} \|u - u^{(k)}\|_{L^2}^2. \end{aligned} \tag{31}$$

Here, L^2 and H norms are defined in (26) and (27).

From the optimality conditions, we can find the explicit formula for each variable $\Phi^{(k+1)}, m_1^{(k+1)}, m_2^{(k+1)}, u^{(k+1)}(t, \cdot)$.

Proposition 11. *The variables $m_1^{(k+1)}, m_2^{(k+1)}, \Phi^{(k+1)}, u^{(k+1)}(1, \cdot)$ from (31) satisfy the following explicit formulas:*

$$\begin{aligned}\Phi^{(k+1)}(t, x) &= \Phi^{(k)}(t, x) + \sigma(AA^T)^{-1} \left(\partial_t u^{(k)}(t, x) + \nabla \cdot m_1^{(k)}(t, x) - m_2^{(k)}(t, x) \right), \\ m_1^{(k+1)}(t, x) &= \frac{V_1(u^{(k)}(t, x))}{\tau + V_1(u^{(k)}(t, x))} \left(m_1^{(k)}(t, x) + \tau \nabla (2\Phi^{(k+1)}(t, x) - \Phi^{(k)}(t, x)) \right), \\ m_2^{(k+1)}(t, x) &= \frac{V_2(u^{(k)}(t, x))}{\tau + V_2(u^{(k)}(t, x))} \left(m_2^{(k)}(t, x) + \tau (2\Phi^{(k+1)}(t, x) - \Phi^{(k)}(t, x)) \right),\end{aligned}$$

for $(t, x) \in [0, 1] \times \Omega$ and

$$u^{(k+1)}(1, x) = (G^*)'(-2\Phi^{(k+1)}(1, x) + \Phi^{(k)}(1, x)), \quad x \in \Omega.$$

The variable $u^{(k+1)}$ also satisfies the following optimality condition:

$$\begin{aligned}-\frac{\|m_1^{(k+1)}\|^2}{2V_1^2(u^{(k+1)})}V_1'(u^{(k+1)}) - \frac{|m_2^{(k+1)}|^2}{2V_2^2(u^{(k+1)})}V_2'(u^{(k+1)}) \\ - \partial_t(2\Phi^{(k+1)} - \Phi^{(k)}) - s'(u^{(k+1)}) + \frac{1}{\tau}(u^{(k+1)} - u^{(k)}) = 0\end{aligned}\tag{32}$$

for $(t, x) \in [0, 1] \times \Omega$. As one can see, the solution $u^{(k+1)}$ of equation depends on V_1 , V_2 , and \mathcal{F} . We first present the algorithm for general V_1 , V_2 and \mathcal{F} . Assuming V_1 , V_2 and \mathcal{F} are smooth, we can compute $u^{(k+1)}$ using Newton's method.

Algorithm 1: Newton's method to compute $u^{(k+1)}(t, x)$

Input: $u^{(k)}(t, x), m_1(t, x), m_2(t, x), \partial_t \Phi(t, x)$.

Output: $u^{(k+1)}(t, x)$.

Initialize $u^0 = u^{(k)}(t, x)$ and set $\epsilon > 0$.

For $j \in \mathbb{N}$

$$\begin{aligned}a &= -\frac{\|m_1\|^2}{2V_1^2(u^j)}V_1'(u^j) - \frac{|m_2|^2}{2V_2^2(u^j)}V_2'(u^j) - \partial_t \Phi - s'(u^j) + \frac{1}{\tau}(u^j - u^{(k)}); \\ b &= \frac{\|m_1\|^2}{V_1^3(u^j)}(V_1'(u^j))^2 - \frac{\|m_1\|^2}{2V_1^2(u^j)}V_1''(u^j) + \frac{|m_2|^2}{V_2^3(u^j)}(V_2'(u^j))^2 - \frac{|m_2|^2}{2V_2^2(u^j)}V_2''(u^j) + \frac{c}{u^j} + \frac{1}{\tau}; \\ u^{j+1} &= \max(0, u^j - \frac{a}{b});\end{aligned}$$

Stop the iteration when $|\frac{a}{b}| < \epsilon$.

In the algorithm 1, we omit (t, x) in $u^j, u^{j+1}, u^{(k)}, m_1, m_2, \partial_t \Phi$ to simplify the notations. Using Proposition 11, we can rewrite the algorithm (31).

Algorithm 2: G-prox PDHG for mean-field information variational problem
Input: Initial density u_0 .

Output: $u, \Phi, m_2 : [0, T] \times \Omega \rightarrow \mathbb{R}, m_1 : [0, T] \times \Omega \rightarrow \mathbb{R}^d$.

For $k \in \mathbb{N}$
For $(t, x) \in [0, 1] \times \Omega$

$$\Phi^{(k+1)}(t, x) = \Phi^{(k)}(t, x) + \sigma(AA^T)^{-1} \left(\partial_t u^{(k)}(t, x) + \nabla \cdot m_1^{(k)}(t, x) - m_2^{(k)}(t, x) \right);$$

$$m_1^{(k+1)}(t, x) = \frac{V_1(u^{(k)}(t, x))}{\tau + V_1(u^{(k)}(t, x))} \left(m_1^{(k)}(t, x) + \tau \nabla (2\Phi^{(k+1)}(t, x) - \Phi^{(k)}(t, x)) \right);$$

$$m_2^{(k+1)}(t, x) = \frac{V_2(u^{(k)}(t, x))}{\tau + V_2(u^{(k)}(t, x))} \left(m_2^{(k)}(t, x) + \tau (2\Phi^{(k+1)}(t, x) - \Phi^{(k)}(t, x)) \right);$$

$$u^{(k+1)}(t, x) \leftarrow \text{Compute using Algorithm 1} \\ \text{with } m_1 = m_1^{(k+1)}, m_2 = m_2^{(k+1)}, \Phi = 2\Phi^{(k+1)} - \Phi^{(k)};$$

$$u^{(k+1)}(1, x) = (G^*)'(-2\Phi^{(k+1)}(1, x) + \Phi^{(k)}(1, x)).$$

The convergence of Newton's method from Algorithm 1 depends on functions V_1, V_2 and a functional \mathcal{F} . In the numerical experiments, we use

$$\mathcal{F}(u(t, \cdot))(x) = -c \int_{\Omega} \left(u(t, x) \log u(t, x) - u(t, x) \right) dx,$$

with a given positive constant c . This functional regularizes equation (32). Thus the algorithm converges faster. In the actual numerical experiments, Newton's method only requires less than 10 iterations to reach the residual error $|a/b|$ less than 10^{-10} . We use FFTW library to compute $(AA^T)^{-1}$ by Fast Fourier Transform (FFT). It takes $O(n \log n)$ operations per iteration, where n is the number of points in both time and spatial grids. Overall, the algorithm takes just $O(n \log n)$ operations per iteration.

In the next subsection, we present an alternative algorithm when V_1 and V_2 are linear functions with specific forms. The optimality condition for $u^{(k+1)}$ leads to a third-order polynomial equation which has an analytical solution and can easily be computed without using the second-order optimization method just as in Algorithm 1.

5.3. Affine Case. Suppose V_1 and V_2 are affine functions, such that

$$V_1(u) = c_1(u + c_3), \quad V_2(u) = c_2(u + c_3) \tag{33}$$

where c_1, c_2, c_3 are constants, and the functional \mathcal{F} is of the form

$$\mathcal{F}(u(t, \cdot)) = - \int_{\Omega} s(u(t, x)) dx$$

where $s : \mathbb{R} \rightarrow \mathbb{R}$ is a smooth convex function. Note that from the algorithm (31), the optimality condition for $u^{(k+1)}$ involves a functional \mathcal{F} . We can get rid of a functional

\mathcal{F} from the optimality condition for $u^{(k+1)}$ by introducing an extra Lagrangian multiplier $\Psi : [0, 1] \times \Omega \rightarrow \mathbb{R}$ to the saddle point problem (20).

$$\begin{aligned} & \inf_{m_1, m_2, u} \sup_{\Phi, \Psi} \mathcal{L}(m_1, m_2, u, \Phi, \Psi) \\ &= \inf_{m_1, m_2, u} \sup_{\Phi, \Psi} \int_0^1 \int_{\Omega} \left\{ \frac{|m_1(t, x)|^2}{2V_1(u(t, x))} + \frac{m_2(t, x)^2}{2V_2(u(t, x))} \right. \\ & \quad \left. + \Phi(t, x) \left(\partial_t u(t, x) + \nabla \cdot m_1(t, x) - m_2(t, x) \right) \right\} dx dt \\ & \quad + \int_0^1 \int_{\Omega} \Psi(t, x) u(t, x) - s^*(\Psi(t, x)) dx dt + \mathcal{G}(u(1, \cdot)). \end{aligned}$$

Using G-Prox PDHG algorithm, the saddle point can be computed by iterating

$$\begin{aligned} \Phi^{(k+1)} &= \operatorname{argmax}_{\Phi} \mathcal{L}(m_1^{(k)}, m_2^{(k)}, u^{(k)}, \Phi, \Psi^{(k)}) - \frac{1}{2\sigma} \|\Phi - \Phi^{(k)}\|_H^2 \\ \Psi^{(k+1)} &= \operatorname{argmax}_{\Psi} \mathcal{L}(m_1^{(k)}, m_2^{(k)}, u^{(k)}, \Phi^{(k)}, \Psi) - \frac{1}{2\sigma} \|\Psi - \Psi^{(k)}\|_{L^2}^2 \\ m_1^{(k+1)} &= \operatorname{argmin}_m \mathcal{L}(m, m_2^{(k)}, u^{(k)}, 2\Phi^{(k+1)} - \Phi^{(k)}, 2\Psi^{(k+1)} - \Psi^{(k)}) + \frac{1}{2\tau} \|m - m_1^{(k)}\|_{L^2}^2 \\ m_2^{(k+1)} &= \operatorname{argmin}_m \mathcal{L}(m_1^{(k)}, m, u^{(k)}, 2\Phi^{(k+1)} - \Phi^{(k)}, 2\Psi^{(k+1)} - \Psi^{(k)}) + \frac{1}{2\tau} \|m - m_2^{(k)}\|_{L^2}^2 \\ u^{(k+1)} &= \operatorname{argmin}_u \mathcal{L}(m_1^{(k+1)}, m_2^{(k+1)}, u, 2\Phi^{(k+1)} - \Phi^{(k)}, 2\Psi^{(k+1)} - \Psi^{(k)}) + \frac{1}{2\tau} \|u - u^{(k)}\|_{L^2}^2. \end{aligned} \tag{34}$$

From the optimality conditions, we can find the explicit formula for each variable $\Phi^{(k+1)}, \Psi^{(k+1)}, m_1^{(k+1)}, m_2^{(k+1)}$.

Proposition 12. *The variables $m_1^{(k+1)}, m_2^{(k+1)}, \Phi^{(k+1)}, \Psi^{(k+1)}, u^{(k+1)}(1, \cdot)$ from (34) satisfy the following explicit formulas:*

$$\begin{aligned} \Phi^{(k+1)}(t, x) &= \Phi^{(k)}(t, x) + \sigma(AA^T)^{-1} \left(\partial_t u^{(k+1)}(t, x) + \nabla \cdot m_1^{(k+1)}(t, x) - m_2^{(k+1)}(t, x) \right), \\ \Psi^{(k+1)}(t, x) &= (Id + \sigma(s^*)')^{-1} (\Psi^{(k)}(t, x) + \sigma u^{(k)}(t, x)), \\ m_1^{(k+1)}(t, x) &= \frac{V_1(u^{(k)}(t, x))}{\tau + V_1(u^{(k)}(t, x))} \left(m_1^{(k)}(t, x) + \tau \nabla (2\Phi^{(k+1)}(t, x) - \Phi^{(k)}(t, x)) \right), \\ m_2^{(k+1)}(t, x) &= \frac{V_2(u^{(k)}(t, x))}{\tau + V_2(u^{(k)}(t, x))} \left(m_2^{(k)}(t, x) + \tau (2\Phi^{(k+1)}(t, x) - \Phi^{(k)}(t, x)) \right), \end{aligned}$$

for $(t, x) \in [0, 1] \times \Omega$,

$$u^{(k+1)}(1, x) = (G^*)'(-2\Phi^{(k+1)}(1, x) + \Phi^{(k)}(1, x)), \quad x \in \Omega,$$

and Id is an identity operator.

For the explicit formula of $\Psi^{(k+1)}$, if $s(t) = \frac{c}{2}t^2$ with $c > 0$, then the formula can be simplified to

$$\Psi^{(k+1)}(t, x) = \frac{1}{1 + \sigma/c} (\Psi^{(k)}(t, x) + \sigma u^{(k)}(t, x)).$$

The following proposition shows the equation from the optimality condition of $u^{(k+1)}$ from (31).

Proposition 13. *If V_1 and V_2 are affine functions in (33), $u^{(k+1)}(t, x)$ satisfies the following equation for $(t, x) \in [0, 1] \times \Omega$:*

$$(u^{(k+1)}(t, x))^3 + (u^{(k+1)}(t, x))^2(2c_3 + k_2) + u^{(k+1)}(t, x)(c_3^2 + 2c_3k_2) + (k_1 + k_2c_3^2) = 0 \quad (35)$$

where

$$k_1(t, x) = -\frac{\tau|m_1^{(k+1)}(t, x)|^2}{2c_1} - \frac{\tau m_2^{(k+1)}(t, x)^2}{2c_2},$$

$$k_2(t, x) = -\tau \partial_t(2\Phi^{(k+1)}(t, x) - \Phi^{(k)}(t, x)) + \tau(2\Psi^{(k+1)}(t, x) - \Psi^{(k)}(t, x)) - u^{(k)}(t, x).$$

Proof. The optimality condition for $u^{(k+1)}$ from (31) gives

$$-\frac{|m_1^{(k+1)}|^2}{2c_1(u^{(k+1)} + c_3)^2} - \frac{(m_2^{(k+1)})^2}{2c_2(u^{(k+1)} + c_3)^2} - \partial_t(2\Phi^{(k+1)} - \Phi^{(k)}) + 2\Psi^{(k+1)} - \Psi^{(k)} + \frac{1}{\tau}(u^{(k+1)} - u^{(k)}) = 0.$$

Simple algebra leads to (35). □

From Proposition 13, the optimality condition for $u^{(k+1)}$ leads to the third order polynomial which has an analytical solution. Thus, $u^{(k+1)}$ can be computed through

$$u^{(k+1)}(t, x) = \text{root}_+(2c_3 + k_2, c_3^2 + 2c_3k_2, k_1 + k_2c_3^2), \quad (t, x) \in [0, 1] \times \Omega$$

where $\text{root}_+(q_1, q_2, q_3)$ is a positive root of a third-order polynomial $x^3 + q_1x^2 + q_2x + q_3 = 0$. Together with Proposition 12 and 13, we present the algorithm for the linear case.

Algorithm 3: G-prox PDHG for mean-field information variational problem with V_1, V_2 defined in (33)

Input: Initial density u_0 .

Output: $u, \Phi, m_2 : [0, T] \times \Omega \rightarrow \mathbb{R}, m_1 : [0, T] \times \Omega \rightarrow \mathbb{R}^d$.

For $k \in \mathbb{N}$

For $(t, x) \in [0, 1] \times \Omega$

$$\Phi^{(k+1)}(t, x) = \Phi^{(k)}(t, x) + \sigma(AA^T)^{-1} \left(\partial_t u^{(k)}(t, x) + \nabla \cdot m_1^{(k)}(t, x) - m_2^{(k)}(t, x) \right);$$

$$\Psi^{(k+1)}(t, x) = (\text{Id} + \sigma(s^*)')^{-1} (\Psi^{(k)}(t, x) + \sigma u^{(k)}(t, x));$$

$$m_1^{(k+1)}(t, x) = \frac{V_1(u^{(k)}(t, x))}{\tau + V_1(u^{(k)}(t, x))} \left(m_1^{(k)}(t, x) + \tau \nabla (2\Phi^{(k+1)}(t, x) - \Phi^{(k)}(t, x)) \right);$$

$$m_2^{(k+1)}(t, x) = \frac{V_2(u^{(k)}(t, x))}{\tau + V_2(u^{(k)}(t, x))} \left(m_2^{(k)}(t, x) + \tau (2\Phi^{(k+1)}(t, x) - \Phi^{(k)}(t, x)) \right);$$

$$u^{(k+1)}(t, x) = \text{root}_+(2b + k_2, b^2 + 2bk_2, k_1 + k_2b^2), \quad k_1, k_2 \text{ are defined in Proposition 13}$$

$$u^{(k+1)}(1, x) = (G^*)'(-2\Phi^{(k+1)}(1, x) + \Phi^{(k)}(1, x)).$$

Note that by introducing an extra dual variable Ψ , the optimality condition for $u^{(k+1)}$ becomes a third-order polynomial which has an analytical solution. Thus, the linear case algorithm does not require a second-order optimization method to compute $u^{(k+1)}$ as in Algorithm 2. The computational complexity of Algorithm 3 is similar to Algorithm 2. We use FFTW library to compute $(AA^T)^{-1}$ by Fast Fourier Transform (FFT), which takes $O(n \log n)$ operations per iteration, where n is the number of points in time and spatial grids. The other variables, Ψ , m_1 , m_2 , and u , takes $O(n)$ operations to compute over $(t, x) \in [0, 1] \times \Omega$. Overall, the algorithm takes $O(n \log n)$ operations per iteration.

5.4. Discretization in 2D. Let $\Omega = [0, 1]^2$ be a unit square in \mathbb{R}^2 and the terminal time $T = 1$. Since we simulate the formulation on a compact set, we set the following boundary conditions on m_1 for the variational problem (15).

$$m_1(t, x) \cdot \vec{n}(t, x) = 0, \quad (t, x) \in [0, 1] \times \partial\Omega \quad (36)$$

where \vec{n} is an outward normal vector and $\partial\Omega$ is a boundary of Ω . This boundary condition is equivalent to no-flux condition that means no mass flows through a boundary. For the numerical experiments, the domain $[0, 1] \times \Omega$ is discretized with the regular Cartesian grid. Denote

$$\Delta x_1 = \frac{1}{N_{x_1}}, \quad \Delta x_2 = \frac{1}{N_{x_2}}, \quad \Delta t = \frac{1}{N_t - 1}$$

where N_{x_1} , N_{x_2} are the number of discretized points in x_1 -axis and x_2 -axis, and N_t is the number of discretized points in time. Denote grid points in space and time as

$$\begin{aligned} x_{jl} &= ((j + 0.5)\Delta x_1, (l + 0.5)\Delta x_2), \quad j = 0, \dots, N_{x_1} - 1 \\ &\quad l = 0, \dots, N_{x_2} - 1 \\ t_n &= n\Delta t, \quad n = 0, \dots, N_t - 1. \end{aligned}$$

Using notations, we have the following approximations:

$$\begin{aligned} u(t_n, x_{jl}) &\approx u_{[n,jl]}, \\ \Phi(t_n, x_{jl}) &\approx \Phi_{[n,jl]}, \\ m_1(t_n, x_{jl}) &\approx [m_1]_{[n,jl]} = ([m_{1,x_1}]_{[n,jl]}, [m_{1,x_2}]_{[n,jl]}), \\ m_2(t_n, x_{jl}) &\approx [m_2]_{[n,jl]}. \end{aligned}$$

The subscript n represents the approximation at $t_n \in [0, 1]$ and the subscript jl represents the approximation at $x_{jl} \in \Omega$. Here, m_{1,x_1} and m_{1,x_2} are x_1 and x_2 components of m_1 , respectively.

The gradient operators are calculated using the first order approximation.

$$\begin{aligned} \partial_{x_1} \Phi_{[n,jl]} &= \frac{\Phi_{[n,jl]} - \Phi_{[n,j-1,l]}}{\Delta x_1}, & \partial_{x_2} \Phi_{[n,jl]} &= \frac{\Phi_{[n,jl]} - \Phi_{[n,j,l-1]}}{\Delta x_2} \\ \partial_t \Phi_{[n,jl]} &= \frac{\Phi_{[n+1,jl]} - \Phi_{[n,jl]}}{\Delta t}, & \partial_t u_{[n,jl]} &= \frac{u_{[n,jl]} - u_{[n-1,jl]}}{\Delta t} \end{aligned}$$

where ∂_{x_i} is a partial derivative with respect to x_i ($i = 1, 2$) axis and ∂_t is a partial derivative with respect to time t . The divergence operator for m_1 is calculated using the first order approximation as well.

$$\nabla \cdot [m_1]_{[n,jl]} = \partial_{x_1}[m_{1,x_1}]_{[n,jl]} + \partial_{x_2}[m_{1,x_2}]_{[n,jl]}$$

$$\partial_{x_1}[m_{1,x_1}]_{[n,jl]} = \frac{[m_{1,x_1}]_{[n,j+1,l]} - [m_{1,x_1}]_{[n,j,l]}}{\Delta x_1}, \quad \partial_{x_2}[m_{1,x_2}]_{[n,jl]} = \frac{[m_{1,x_2}]_{[n,j,l+1]} - [m_{1,x_2}]_{[n,j,l]}}{\Delta x_2}$$

Note that the partial derivative of Φ with respect to t uses forward difference scheme and that of u uses backward difference scheme. The partial derivatives of Φ with respect to x_1 and x_2 use backward difference scheme and those of m_1 use forward difference scheme. The following proposition justifies our choices of finite difference schemes for each variable.

Proposition 14. *Define the discrete partial differential operators of u and m_1 as follows*

$$\begin{aligned} \partial_t u_{[n,jl]} &= \frac{u_{[n,jl]} - u_{[n-1,jl]}}{\Delta t} \\ \nabla \cdot [m_1]_{[n,jl]} &= \partial_{x_1}[m_{1,x_1}]_{[n,jl]} + \partial_{x_2}[m_{1,x_2}]_{[n,jl]} \\ \partial_{x_1}[m_{1,x_1}]_{[n,jl]} &= \frac{[m_{1,x_1}]_{[n,j+1,l]} - [m_{1,x_1}]_{[n,j,l]}}{\Delta x_1} \\ \partial_{x_2}[m_{1,x_2}]_{[n,jl]} &= \frac{[m_{1,x_2}]_{[n,j,l+1]} - [m_{1,x_2}]_{[n,j,l]}}{\Delta x_2}. \end{aligned} \tag{37}$$

Suppose u satisfies the Dirichlet boundary condition in time

$$\begin{aligned} u_{0,jl} &= u_0(x_{jl}), \quad j = 0, \dots, N_{x_1} - 1 \\ l &= 0, \dots, N_{x_2} - 1 \end{aligned}$$

and m_1 satisfies the no-flux boundary condition in space

$$\begin{aligned} [m_{1,x_1}]_{[n,0,l]} &= [m_{1,x_1}]_{[n,N_{x_1},l]} = 0, \quad n = 0, \dots, N_t - 1, \\ l &= 0, \dots, N_{x_2} - 1 \end{aligned}$$

and

$$\begin{aligned} [m_{1,x_2}]_{[n,j,0]} &= [m_{1,x_2}]_{[n,j,N_{x_2}]} = 0, \quad n = 0, \dots, N_t - 1, \\ j &= 0, \dots, N_{x_1} - 1. \end{aligned}$$

Then, $\Phi_{[n,jl]}$, $[m_1]_{[n,jl]}$, $[m_2]_{[n,jl]}$ satisfy the following optimality conditions:

$$\partial_t u_{[n,jl]} + \nabla \cdot [m_1]_{[n,jl]} - [m_2]_{[n,jl]} = 0 \tag{38}$$

$$\frac{[m_{1,x_i}]_{[n,jl]}}{V_1(u_{[n,jl]})} - \partial_{x_i} \Phi_{[n,jl]} = 0, \quad (i = 1, 2) \tag{39}$$

$$\frac{[m_2]_{[n,jl]}}{V_2(u_{[n,jl]})} - \Phi_{[n,jl]} = 0, \tag{40}$$

for $n = 0, \dots, N_t - 1$, $j = 0, \dots, N_{x_1} - 1$, $l = 0, \dots, N_{x_2} - 1$. The optimality condition for $u_{[n,jl]}$ is

$$-\frac{([m_{1,x_1}]_{[n,jl]})^2 + ([m_{1,x_2}]_{[n,jl]})^2}{2V_1(u_{[n,jl]})^2} V_1'(u_{[n,jl]}) - \frac{([m_2]_{[n,jl]})^2}{2V_2(u_{[n,jl]})^2} V_2'(u_{[n,jl]}) - \partial_t \Phi_{[n,jl]} + s'(u_{[n,jl]}) = 0 \tag{41}$$

for $n = 1, \dots, N_t - 2$, $j = 0, \dots, N_{x_1} - 1$, $l = 0, \dots, N_{x_2} - 1$ and

$$\Phi_{[N_t-1,j,l]} + G'(u_{[N_t-1,j,l]}) = 0 \quad (42)$$

for $j = 0, \dots, N_{x_1} - 1$, $l = 0, \dots, N_{x_2} - 1$. The differential operators of Φ with respect to time and space are defined as

$$\begin{aligned} \partial_t \Phi_{[n,j,l]} &= \frac{\Phi_{[n+1,j,l]} - \Phi_{[n,j,l]}}{\Delta t}, \\ \partial_{x_1} \Phi_{[n,j,l]} &= \frac{\Phi_{[n,j,l]} - \Phi_{[n,j-1,l]}}{\Delta x_1}, \\ \partial_{x_2} \Phi_{[n,j,l]} &= \frac{\Phi_{[n,j,l]} - \Phi_{[n,j,l-1]}}{\Delta x_2}, \end{aligned}$$

with Neumann boundary conditions in time and space

$$\begin{aligned} \Phi_{[1,j,l]} &= \Phi_{[0,j,l]}, \quad j = 0, \dots, N_{x_1} - 1, \\ l &= 0, \dots, N_{x_2} - 1. \end{aligned}$$

$$\begin{aligned} \Phi_{[n,j,0]} &= \Phi_{[n,j,-1]}, \quad n = 0, \dots, N_t - 1, \\ j &= 0, \dots, N_{x_1} - 1. \end{aligned}$$

$$\begin{aligned} \Phi_{[n,0,l]} &= \Phi_{[n,-1,l]}, \quad n = 0, \dots, N_t - 1, \\ l &= 0, \dots, N_{x_2} - 1 \end{aligned}$$

and a boundary condition in time at $t = 1$ from (42).

Proof. Using the discretization notations that are introduced in this subsection, we discretize the Lagrangian functional in (30).

$$\begin{aligned} \mathcal{L}(m_1, m_2, u, \Phi) &\approx \Delta t \Delta x_1 \Delta x_2 \sum_{n=1}^{N_t-1} \sum_{j=0}^{N_{x_1}-1} \sum_{l=0}^{N_{x_2}-1} \left(\frac{([m_{1,x_1}]_{[n,j,l]})^2 + ([m_{1,x_1}]_{[n,j,l]})^2}{2V_1(u_{[n,j,l]})} + \frac{([m_2]_{[n,j,l]})^2}{2V_2(u_{[n,j,l]})} \right. \\ &\quad \left. + \Phi_{[n,j,l]} \left(\partial_t u_{[n,j,l]} + \nabla \cdot [m_1]_{[n,j,l]} - [m_2]_{[n,j,l]} \right) + s(u_{[n,j,l]}) \right) \\ &\quad + \Delta x_1 \Delta x_2 \sum_{j=1}^{N_{x_1}-1} \sum_{l=1}^{N_{x_2}-1} G(u_{[N_t-1,j,l]}). \end{aligned}$$

The optimality condition for $\Phi_{[n,j,l]}$ can be obtained by differentiating with respect to $\Phi_{[n,j,l]}$

$$\partial_t u_{[n,j,l]} + \nabla \cdot [m_1]_{[n,j,l]} - [m_2]_{[n,j,l]} = 0$$

for $n = 0, \dots, N_t - 1$, $j = 0, \dots, N_{x_1} - 1$, $l = 0, \dots, N_{x_2} - 1$.

Let us take a closer look at the terms with gradients and divergence operators. Using the difference schemes defined in (37), it can be rewritten as

$$\begin{aligned} & \sum_{n=1}^{N_t-1} \sum_{j=0}^{N_{x_1}-1} \sum_{l=0}^{N_{x_2}-1} \Phi_{[n,jl]} \left(\partial_t u_{[n,jl]} + \nabla \cdot [m_1]_{[n,jl]} \right) \\ &= \sum_{n=1}^{N_t-1} \sum_{j=0}^{N_{x_1}-1} \sum_{l=0}^{N_{x_2}-1} \Phi_{[n,jl]} \left(\frac{u_{[n,jl]} - u_{[n-1,jl]}}{\Delta t} + \frac{[m_{1,x_1}]_{n,j+1l} - [m_{1,x_1}]_{[n,jl]}}{\Delta x_1} + \frac{[m_{1,x_2}]_{[n,j,l+1]} - [m_{1,x_2}]_{[n,jl]}}{\Delta x_2} \right). \end{aligned}$$

By rearranging the indices,

$$\begin{aligned} &= - \sum_{n=1}^{N_t-2} \sum_{j=0}^{N_{x_1}-1} \sum_{l=0}^{N_{x_2}-1} \left(\frac{\Phi_{[n+1,jl]} - \Phi_{[n,jl]}}{\Delta t} \right) u_{[n,jl]} \\ &\quad - \sum_{n=1}^{N_t-1} \sum_{j=0}^{N_{x_1}-1} \sum_{l=0}^{N_{x_2}-1} \left(\frac{\Phi_{[n,jl]} - \Phi_{[n,j-1,l]}}{\Delta x_1} \right) [m_{1,x_1}]_{[n,jl]} \\ &\quad - \sum_{n=1}^{N_t-1} \sum_{j=0}^{N_{x_1}-1} \sum_{l=0}^{N_{x_2}-1} \left(\frac{\Phi_{[n,jl]} - \Phi_{[n,j,l-1]}}{\Delta x_2} \right) [m_{1,x_2}]_{[n,jl]} \\ &\quad + \sum_{j=1}^{N_{x_1}-1} \sum_{l=1}^{N_{x_2}-1} \frac{1}{\Delta t} (\Phi_{[N_t-1,jl]} u_{[N_t-1,jl]} - \Phi_{[0,jl]} u_{[0,jl]}). \end{aligned}$$

From above, we define partial differential operators of Φ as

$$\begin{aligned} \partial_t \Phi_{[n,jl]} &= (\Phi_{[n+1,jl]} - \Phi_{[n,jl]}) / \Delta t, \\ \partial_{x_1} \Phi_{[n,jl]} &= (\Phi_{[n,jl]} - \Phi_{[n,j-1,l]}) / \Delta x_1, \\ \partial_{x_2} \Phi_{[n,jl]} &= (\Phi_{[n,jl]} - \Phi_{[n,j,l-1]}) / \Delta x_2. \end{aligned}$$

Putting back to the Lagrangian functional, we get

$$\begin{aligned} & \mathcal{L}(m_1, m_2, u, \Phi) \\ & \approx \Delta t \Delta x_1 \Delta x_2 \sum_{n=1}^{N_t-1} \sum_{j=0}^{N_{x_1}-1} \sum_{l=0}^{N_{x_2}-1} \left(\frac{([m_{1,x_1}]_{[n,jl]})^2 + ([m_{1,x_2}]_{[n,jl]})^2}{2V_1(u_{[n,jl]})} + \frac{([m_2]_{[n,jl]})^2}{2V_2(u_{[n,jl]})} - \Phi_{[n,jl]} [m_2]_{[n,jl]} \right. \\ & \quad \left. - \partial_{x_1} \Phi_{[n,jl]} [m_{1,x_1}]_{[n,jl]} - \partial_{x_2} \Phi_{[n,jl]} [m_{1,x_2}]_{[n,jl]} + s(u_{[n,jl]}) \right) \\ & \quad - \Delta t \Delta x_1 \Delta x_2 \sum_{n=1}^{N_t-2} \sum_{j=0}^{N_{x_1}-1} \sum_{l=0}^{N_{x_2}-1} \partial_t \Phi_{[n,jl]} u_{[n,jl]} \\ & \quad + \Delta x_1 \Delta x_2 \sum_{j=1}^{N_{x_1}-1} \sum_{l=1}^{N_{x_2}-1} \Phi_{[N_t-1,jl]} u_{[N_t-1,jl]} - \Phi_{[0,jl]} u_{[0,jl]} + G(u_{[N_t-1,jl]}). \end{aligned}$$

By differentiating with respect to $[m_{1,x_i}]_{[n,jl]}$ ($i = 1, 2$), $[m_2]_{[n,jl]}$, $u_{[n,jl]}$, we achieve the optimality conditions (39), (40), (41).

□

From Proposition 14, the variational problem solves the following PDE:

$$\begin{aligned} \partial_t u + \nabla \cdot m_1 &= m_2 \\ \partial_t \Phi + \frac{1}{2} V_1'(u) |\nabla \Phi|^2 + \frac{1}{2} V_2'(u) \Phi^2 &= s'(u) \\ u(0, x) &= u_0(x), \quad \Phi(1, x) = -G'(u(1, x)), \quad x \in \Omega. \end{aligned} \tag{43}$$

Note that solving this system of PDEs is computationally challenging. The PDE of u evolves forward in time with an initial condition u_0 and that of Φ evolves backward in time with a terminal condition $-G'(u(1, x))$. Solving the PDEs using the finite difference methods is difficult due to the strict CFL condition (the ratio of Δt , Δx_1 , Δx_2). Furthermore, using finite difference schemes that are implicit in time is also challenging because there is no simple formula of $u_{[n,jl]}$ in terms of $u_{[n-1,jl]}$, $[m_1]_{[n,jl]}$, $[m_2]_{[n,jl]}$ that satisfy the PDE of Φ (43). With nonlinear V_1 and V_2 , it becomes even more difficult to use any finite difference schemes.

Instead of solving PDEs directly using finite difference schemes, the algorithm computes the saddle point of a variational problem (15) that satisfies the optimality conditions in Proposition 14. Thus, the saddle points are in fact the solutions to the PDE (43). Furthermore, by (38) and (41), the algorithm solves (43) with implicit finite difference schemes (forward in time for u and backward in time for Φ). Thus, the algorithm circumvents the numerical difficulties coming from the strict CFL conditions.

6. NUMERICAL EXAMPLES

In this section, we present two sets of numerical experiments using the Algorithm 2 with different V_1 , V_2 functions. We wrote C++ codes to run the numerical experiments. For all the experiments we used the discretization sizes $N_{x_1} = N_{x_2} = 128$, $N_t = 30$ and $c = 0.1$. Furthermore, in the numerical experiments, we use

$$\mathcal{F}(u(t, \cdot))(x) = -c \int_{\Omega} \left(u(t, x) \log u(t, x) - u(t, x) \right) dx,$$

with a given positive constant c .

6.0.1. *Numerical Example 1.* The numerical simulations are based on the choice of

$$V_1(u) = u \text{ and } V_2(u) = u^\alpha$$

where $\alpha > 0$. The choices are from Example 5 in Section 4. We show how varying the exponent α affects the evolution of the densities. Below are the initial and terminal densities used in the simulations.

$$\begin{aligned} u_0(0, x_1, x_2) &= 10 \exp(-60 ((x_1 - 0.3)^2 + (x_2 - 0.7)^2)) + 1 \\ u_1(0, x_1, x_2) &= 20 \exp(-60 ((x_1 - 0.7)^2 + (x_2 - 0.3)^2)) + 1. \end{aligned}$$

Furthermore, we use the terminal functional

$$\mathcal{G}(u(1, \cdot)) = \int_{\Omega} i_{u_1(x)}(u(1, x)) dx \quad (44)$$

where

$$i_{u_1(x)}(u(1, x)) = \begin{cases} 0 & \text{if } u(1, x) = u_0(x) \\ \infty & \text{otherwise.} \end{cases}$$

This function is equivalent to the following discrete form:

$$i_{[u_1]_{[jl]}}(u_{[0,jl]}) = \begin{cases} 0 & \text{if } u_{[N_t-1,jl]} = [u_1]_{[jl]} \\ \infty & \text{otherwise} \end{cases}$$

for $j = 0, \dots, N_{x_1} - 1$ and $l = 0, \dots, N_{x_2} - 1$. Recall that, from Proposition 14, the boundary condition of Φ at $t = 1$ is given by

$$\Phi_{[N_t-1,jl]} = -(i_{[u_1]_{[jl]}})'(u_{[N_t-1,jl]}).$$

However, dealing with the derivative of $i_{[u_1]_{[jl]}}$ is computationally challenging because it is non-differentiable and discontinuous. Since the function is convex, we may use a property Legendre transform, i.e.,

$$(f^*)' = (f')^{-1}$$

for any convex function f . Using the property, we get the boundary condition of u at time $t = 1$,

$$u_{[N_t-1,jl]} = (i_{[u_1]_{[jl]}}^*)'(-\Phi_{[N_t-1,jl]})$$

where

$$i_{[u_1]_{[jl]}}^*(\Phi) = \sup_{u \in \mathbb{R}} u \cdot \Phi - i_{[u_1]_{[jl]}}(u) = [u_1]_{[jl]} \Phi.$$

Thus, we have

$$u_{[N_t-1,jl]} = [u_1]_{[jl]}.$$

In other words, the terminal functional (44) is equivalent to the boundary condition of u in time at $t = 1$.

The variational problem (15) is solved with these initial conditions using Algorithm 2. The solution of the variational problem represents the evolution of the density u from initial density u_0 at time $t = 0$ to terminal density u_1 at $t = 1$ that satisfies the system of PDEs (43). The results are shown in Figure 1, Figure 2 and Figure 3. In Figure 1, the plots show the evolution of densities from time $t = 0$ to $t = 1$. The first row is from $V_2(u) = u$, the second row is from $V_2(u) = u^2$, the third row is from $V_2(u) = u^3$ and the last row is from $V_2(u) = u^4$. The Figure 2 shows the cross sections of the 2d plots from Figure 1 along a line $x + y = 1$.

Note that the function $(u, m) \mapsto \frac{\|m\|_2^2}{2V(u)}$ is nonconvex when $V(u) = u^\alpha$ and $\alpha > 1$. Thus, the solutions for $V_2(u) = u^2, u^3, u^4$ of the variational problems are not unique and the numerical results may not represent the global minimizer of the formulation (15). However, the algorithm converges to a local minimum and the convergence plot can be seen in Figure 3. In Figure 3, the vertical axis represents the value of the energy functional given in (13) and the horizontal axis represents iterations.

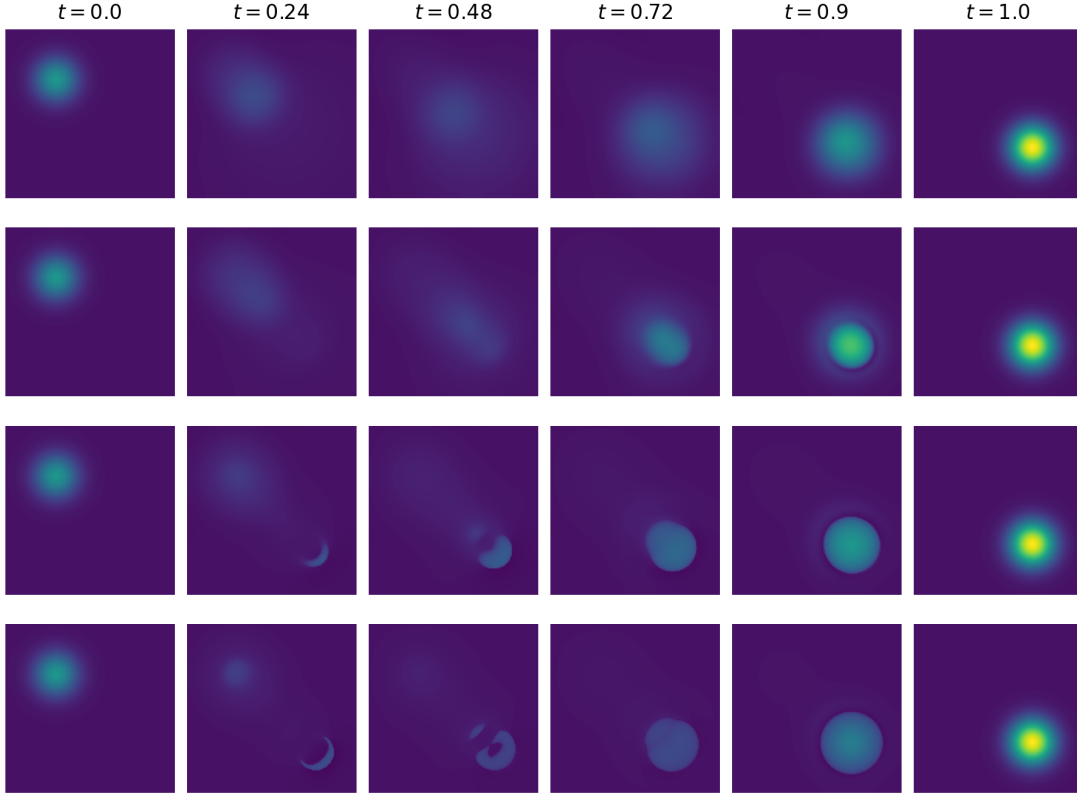


FIGURE 1. Example 1: Snapshots of the densities at different times t . Each column shows the results at $t = 0$ (column 1), $t = 0.24$ (column 2), $t = 0.48$ (column 3), $t = 0.72$ (column 4), $t = 0.9$ (column 5), $t = 1$ (column 6). The first, the second, the third and the fourth rows represent $V_2(u) = u$, $V_2(u) = u^2$, $V_2(u) = u^3$, $V_2(u) = u^4$, respectively. $V_1(u) = u$ for all simulations.

6.0.2. *Numerical Example 2.* The numerical simulations are based on the choice of $V_2(u) = \frac{u(u-1)}{\log u}$ which is from the Fisher-KPP equation in Example 7. In particular, this result compares the following three sets of functions.

	$V_1(u)$	$V_2(u)$
1	u	$u(u-1)/\log u$
2	\sqrt{u}	$u(u-1)/\log u$
3	\sqrt{u}	u

Below are the initial and terminal densities used in the simulations.

$$u_0(0, x) = 15 \exp(-80((x - 0.5)^2 + (y - 0.5)^2)) + 1$$

$$u_1(0, x) = 15 \exp(-80((x - 0.3)^2 + (y - 0.3)^2)) + 15 \exp(-80((x - 0.7)^2 + (y - 0.7)^2)) + 1.$$

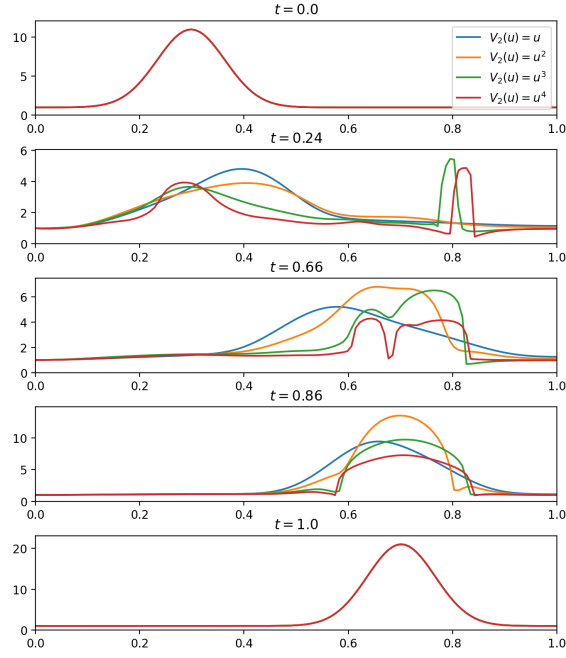


FIGURE 2. Example 1: Cross sections along a line $x + y = 1$ for different time t . Each color represents $V_2(u) = u$, $V_2(u) = u^2$, $V_2(u) = u^3$, $V_2(u) = u^4$.

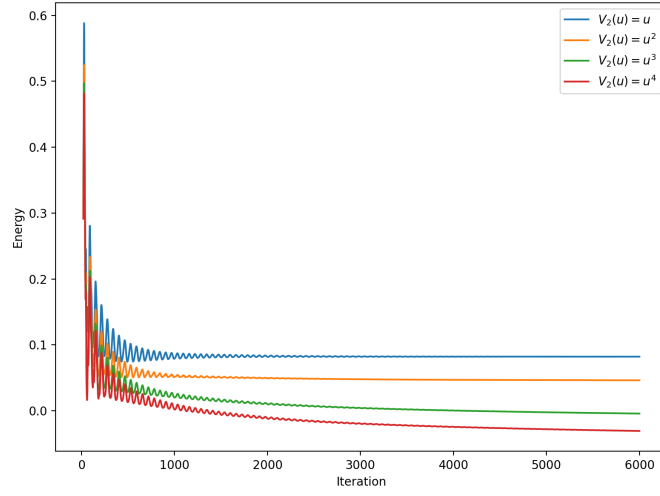


FIGURE 3. Example 1: The convergence plot from the Algorithm. The x-axis represents iterations and y-axis represents the energy defined in (13).

As in example 1, the terminal functional is defined as

$$\mathcal{G}(u(1, \cdot)) = \int_{\Omega} i_{u_1(x)}(u(1, x)) dx.$$

The variational problem (15) with these initial conditions are solved using Algorithm 2. The results are shown in Figure 4 and Figure 5. This example is also computed from our Algorithm 2. Figure 4 demonstrates the evolution of densities for different V_1 and V_2 for the time interval $[0, 1]$. Figure 5 shows the optimal control energy functional value for different choices of V_1, V_2 . The evolution of density functions are similar, but the energy functionals are different.

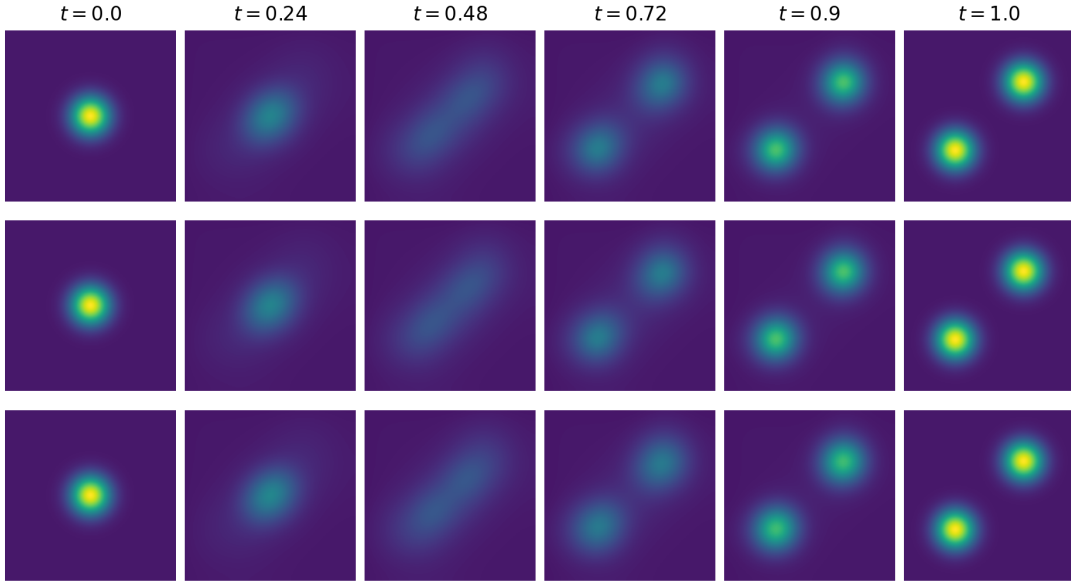


FIGURE 4. Example 2: Snapshots of the densities at different times t . Each column shows the results at $t = 0$ (column 1), $t = 0.24$ (column 2), $t = 0.48$ (column 3), $t = 0.72$ (column 4), $t = 0.9$ (column 5), $t = 1$ (column 6). The first row is from $(V_1(u) = u, V_2(u) = \frac{u(u-1)}{\log u})$, the second row is from $(V_1(u) = \sqrt{u}, V_2(u) = \frac{u(u-1)}{\log u})$ and the last row is from $(V_1(u) = \sqrt{u}, V_2(u) = u)$.

7. DISCUSSION

In this paper, we applied a novel generalized mean field control for nonlinear reaction-diffusion equations. Several mean-field information variational problems and dynamics in unnormalized density space are presented. In computations, we focus on mean field control problems. And we design primal-dual hybrid gradient methods to compute these generalized optimal transport and mean-field control problems. In future work, we shall design and compute unconditional stable implicit time schemes for reaction-diffusion equations using mean field control problems.

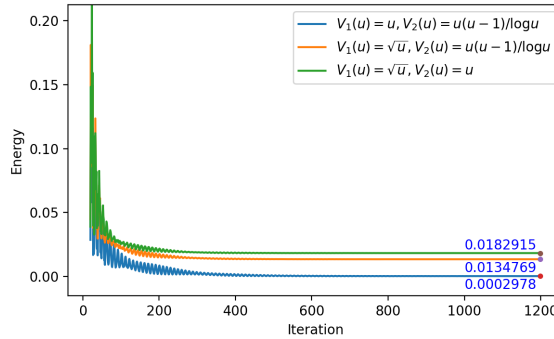


FIGURE 5. Example 2: The convergence plot from the Algorithm. The x-axis represents iterations and y-axis represents the energy defined in (13). The blue line represents the solution of $V_1(u) = u$ and $V_2(u) = \frac{u(u-1)}{\log u}$ and converges to 0.00029. The orange line represents the solution of $V_1(u) = \sqrt{u}$ and $V_2(u) = \frac{u(u-1)}{\log u}$ and converges to 0.0135. The green line represents the solution of $V_1(u) = \sqrt{u}$ and $V_2(u) = u$ and converges to 0.0183.

REFERENCES

- [1] S. Amari. *Information Geometry and Its Applications*. Springer Publishing Company, Incorporated, 1st edition, 2016.
- [2] L. Ambrosio, N. Gigli and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zurich, Birkhauser Verlag, Basel, 2nd ed. 2008.
- [3] J.D. Benamou and Y. Brenier. A Computational Fluid Mechanics Solution to the Monge-Kantorovich Mass Transfer Problem. *Numerische Mathematik*, 84(3):375–393, 2000
- [4] P. Cardaliaguet, F. Delarue, J. Lasry, and P. Lions. The master equation and the convergence problem in mean-field games. *arXiv:1509.02505*, 2015.
- [5] J.A. Carrillo, K. Craig, L. Wang and C.Z. Wei. Primal Dual Methods for Wasserstein Gradient Flows. *Found Comput Math*, 2021.
- [6] J.A. Carrillo, S. Lisini, G. Savaré and D. Slepcev. Nonlinear mobility continuity equations and generalized displacement convexity. *Journal of Functional Analysis* 258(4):1273-1309, 2009.
- [7] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision.*, 40(1):120-145, 2011.
- [8] A. Chambolle and T. Pock. On the ergodic convergence rates of a first-order primal-dual algorithm. *Math. Program.*, 159(1-2, Ser. A):253-287, 2016.
- [9] J. Chayes, S. Osher and J. Ralston. On singular diffusion equations with applications to self-organized criticality. *Communications on Pure and Applied Mathematics*, 1993.
- [10] L. Chizat, G. Peyre, B. Schmitzer, and F.X. Vialard. An Interpolating Distance Between Optimal Transport and Fisher-Rao Metrics. *Found Comput Math* 18, 1-44, 2018.
- [11] J. Dolbeault, B. Nazaret, and G. Savaré. A new class of transport distances. *Calculus of Variations and Partial Differential Equations*, (2):193-231, 2009.
- [12] T.O. Gallouet, and L. Monsaingeon. A JKO Splitting Scheme for Kantorovich-Fisher-Rao Gradient Flows. *SIAM J. Math. Anal.*, 49(2), 1100-1130, 2017.
- [13] W. Gangbo, W. Li, S. Osher, and M. Puthawala. Unnormalized Optimal Transport. *Journal of Computational Physics*, 2019.
- [14] W. Gangbo, T. Nguyen, and A. Tudorascu. Hamilton-Jacobi equations in the Wasserstein space. *Meth. Appl. Anal.* Vol. 15 no 2, 155-184, 2008.
- [15] Y. Gao, and J.G. Liu. A note on parametric Bayesian inference via gradient flows. *Annals of Mathematical Sciences and Applications*, 2020.

- [16] Y. Gao, and J.G. Liu. Revisit of macroscopic dynamics for non-equilibrium chemical reactions from a Hamiltonian viewpoint. *arXiv:2108.05396*, 2021.
- [17] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A.M. Stuart. Interacting Langevin Diffusions: Gradient Structure and Ensemble Kalman Sampler. *SIAM J. Appl. Dyn. Syst.*, (1), 412–441, 2019.
- [18] Y. Gong, J. Zhao, and Q. Wang. Arbitrarily high-order unconditionally energy stable schemes for thermodynamically consistent gradient flow models. *SIAM Journal on Scientific Computing*, 42 (1), B135-B156, 2020.
- [19] M. Huang, R.P. Malhame, and P. Caines. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221-252, 2006.
- [20] M. Jacobs, F. Léger, W. Li, and S. Osher. Solving Large-Scale Optimization Problems with a Convergence Rate Independent of Grid Size. *SIAM Journal on Numerical Analysis.*, 57(3), 1100-1123, 2019.
- [21] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1-17, 1998.
- [22] S. Kondratyev, and D. Vorotnikov. Spherical Hellinger-Kantorovich Gradient Flows. *SIAM Journal on Mathematical Analysis* 51:3, 2053-2084, 2019.
- [23] J. D. Lafferty. The density manifold and configuration space quantization. *Transactions of the American Mathematical Society*, 305(2):699-741, 1988.
- [24] J. M. Lasry, and P. L. Lions, mean-field games. *Japanese Journal of Mathematics*, 2, 229-260, 2007.
- [25] W. Lee, R. Lai, W. Li, and S. Osher. Generalized Unnormalized Optimal Transport and its fast algorithms. *Journal of Computational Physics*, Volume 436, 1, 2021.
- [26] W. Li. Transport information geometry: Riemannian calculus on probability simplex. *Information Geometry*, 2020.
- [27] W. Li. Hessian metric via transport information geometry. *Journal of Mathematical Physics*, 62, 033301, 2021.
- [28] W. Li, S. Liu, and S. Osher. Controlling conservation laws I: entropy-entropy flux. *arXiv:2111.05473*, 2021.
- [29] W. Li, and L. Ying. Hessian transport gradient flows. *Res Math Sci*, 6, 34, 2019.
- [30] M. Liero, A. Mielke, G. Savaré. *Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. Inventiones mathematicae*, 211(3) 969-1117, 2018
- [31] A. Lin, W. Li, S. Osher and G. Montufar. Wasserstein Proximal of GANs. Geometry information of science, 2021.
- [32] S. Lisini, D. Matthes, and G. Savaré. Cahn-Hilliard and thin film equations with nonlinear mobility as gradient flows in weighted-Wasserstein metrics. *Journal of differential equations*, (253): 814-850, 2012.
- [33] Q. Liu, and D. Wang. Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm. *NeurIPS*, 2016.
- [34] J. Maas, and A. Mielke. Modeling of Chemical Reaction Systems with Detailed Balance Using Gradient Structures. *Journal of statistical physics*, 181, 2257-2303, 2020.
- [35] A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 4(4):1329, 2011.
- [36] A. Mielke. Free energy, free entropy, and a gradient structure for thermoplasticity. In “K. Weinberg and A. Pandolfi (eds): *Innovative Numerical Approaches for Multi-Field and Multi-Scale Problems*”. Lecture Notes in Appl. Comp. Mechanics Vol. 81, pages 135-160, Springer 2016.
- [37] A. Mielke, D. R. M. Renger, and M. A. Peletier. A generalization of Onsagers reciprocity relations to gradient flows with nonlinear mobility. *J. Non-Equil. Thermodyn.*, 41(2), 141-149, 2016.
- [38] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101-174, 2001.
- [39] L. Onsager. Reciprocal relations in irreversible processes, *I+II. Physical Review*, 37, 405D426, 1931.
- [40] C. Villani. *Optimal Transport: Old and New*. Number 338 in Grundlehren Der Mathematischen Wissenschaften. Springer, Berlin, 2009.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTH CAROLINA, COLUMBIA

Email address: `wlee@math.ucla.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES

Email address: `sjo@math.ucla.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES