

Algorithm for Hamilton-Jacobi equations in density space via a generalized Hopf formula

Yat Tin Chow* Wuchen Li[†] Stanley Osher* Wotao Yin*

Abstract

We design fast numerical methods for Hamilton-Jacobi equations in density space (HJD), which arises in optimal transport and mean field games. We propose an algorithm using a generalized Hopf formula¹ in density space. The formula helps transforming a problem from an optimal control problem in density space, which are constrained minimizations supported on both spatial and time variables, to an optimization problem over only one spatial variable. This transformation allows us to compute HJD efficiently via multi-level approaches and coordinate descent methods. Rigorous derivation of the Hopf formula is provided under restricted assumptions and for a relatively narrow case; meanwhile our practical investigation allows us to conjecture that the actual range of applicability should be wider, and therefore we conjecture the formula can be applied to a wider class of practical examples.

Keywords: Hamilton-Jacobi equation in density space; Generalized Hopf formula; Mean field games; Optimal transport.

1 Introduction

In recent years, optimal control problems in density space have started to play vital roles in physics [30], fluid dynamics [5] and probability [9]. Two typical examples are mean field games (MFGs) [25, 27] and optimal transportation [32]. For these optimal control problems, Hamilton-Jacobi equation in density space (HJD) determines the global information of the system [22, 23], which describes the time evolution of the optimal value in density space. More precisely, HJD refers to the functional differential equation as follows: Let $x \in X$ where X be a compact Riemannian manifold, and $\rho(\cdot) \in \mathcal{P}(X)$ represent the probability density space supported on X . Let $U: [0, \infty) \times \mathcal{P}(X) \rightarrow \mathbb{R}$ be the value function. Consider

$$\begin{cases} \partial_s U(s, \rho) + \mathcal{H}(\rho, \delta_\rho U) = 0 \\ U(0, \rho) = G(\rho), \end{cases}$$

where δ_ρ is the L^2 first variation w.r.t. ρ and \mathcal{H} represents the total Hamiltonian function in $\mathcal{P}(X)$:

$$\mathcal{H}(\rho, \delta_{\rho(x)} U) := \int_X H(x, \nabla_x \delta_{\rho(x)} U) \rho(x) dx + F(\rho), \quad (1.1)$$

*Department of Mathematics, UCR, Riverside, CA 92521. (yattinc@ucr.edu) Research supported by Research supported by AFOSR MURI proposal number 18RT0073, ONR N000141712162 and NSF DMS-1720237.

[†]Department of Mathematics, UCLA, Los Angeles, CA 90095-1555 (wcli@math.ucla.edu, sjo@math.ucla.edu, wotaoyin@math.ucla.edu). Research supported by AFOSR MURI proposal number 18RT0073, ONR grant: N00014-1-0444, N00014-16-1-2119, N00014-16-215-1, NSF grant ECCS-1462398 and DOE grant DE-SC00183838.

¹We drop the word “generalized” in what follows.

with the given Hamiltonian function H on X . Here, $F, G: \mathcal{P}(X) \rightarrow \mathbb{R}$ are given interaction potential and initial cost functional in density space, respectively.

In applications, HJD has been shown very effective at modeling population differential games, also known as MFGs, which study strategic dynamical interactions in large populations by extending finite players' differential games. This setting provides powerful tools for modeling macro-economics, stock markets, and wealth distribution [24]. In this setting, a Nash equilibrium (NE) describes a status in which no individual player in the population is willing to change his/her strategy unilaterally. A widely-studied special class of MFG is the potential game [29], where all players face the same cost function or potential, and every player minimizes this potential. This amounts to solving an optimal control problem in density space. In this case, a NE refers to the characteristics of HJD, which form a PDE system consisting of continuity equation and Hamilton-Jacobi equation in X . These two equations represent the dynamical evolutions of the population density and the cost value, respectively.

Despite the importance of HJD, solving it numerically is not a simple task. It is known that computing Hamilton-Jacobi equations using a grid in a dimension greater than or equal to three is difficult. The cost increases exponentially with the dimension, which is known as the curse of dimensionality [17]. HJD is even harder to compute since it involves an infinite-dimensional functional PDE. In this paper, expanding the ideas in [16, 15, 14, 17], we overcome the curse of infinite dimensionality in HJD by exploiting a Hopf formula in density space. This approach considers a particular primal-dual formulation associated with the optimal control problem in density space. Specifically, the Hopf formula is given as

$$U(t, \rho) := \sup_{\Phi_t} \left\{ \int_X \rho_t \Phi_t dx - \int_0^t \left(F(\rho_s) - \int_X \rho_s \delta_{\rho_s} F(\rho_s) dx \right) ds - G^*(\Phi_0) : \right. \\ \left. \begin{aligned} \partial_s \rho_s &= \delta_{\Phi_s} \mathcal{H}(\rho_s, \Phi_s), & \partial_s \Phi_s &= -\delta_{\rho_s} \mathcal{H}(\rho_s, \Phi_s) \\ \rho(x, t) &= \rho_t(x), & \Phi(x, t) &= \Phi_t(x) \end{aligned} \right\},$$

where $\Phi_0(x) = \Phi(0, x)$ and

$$G^*(\Phi_0) := \sup_{\rho_0 \in \mathcal{P}(X)} \int_X \rho_0 \Phi_0 dx - G(\rho_0).$$

In here, we would like to compare similarity and difference between the formula in [16, 15, 14, 17] and that is stated above. In [16, 15, 14, 17], we only have Hamiltonian H but there is no potential/interaction energy F . In this work, however, we have both Hamiltonian H and interaction F . Nonetheless, one may combine all the terms to obtain a total Hamiltonian \mathcal{H} as in (1.1). Now we may regard \mathcal{H} as a full Hamiltonian and interpret that as an analog of the Hamiltonian stated in [16, 15, 14, 17].

The rigorous derivation of the Hopf formula suggested is provided for a relatively narrow case; however, the practical investigation allows to conjecture that the actual range of applicability should be wider. We further discretize the above variational problem following the same discretization as in optimal transport on graphs [11, 12, 13, 20, 28]. We then apply a multi-level stochastic gradient descent method to optimize the discretized problem.

In the literature of numerical methods for potential MFGs are seminal works of Achdou, Camilli, and Dolcetta [1, 2, 3]. Their approaches utilize the primal-dual structure of the optimal control formulation, simplifying it by a Legendre transform and applying Newton's method to the resulting saddle point system. Different from their approaches, we focus on solving the dual problem, in which the optimal control problem is an optimization problem over the terminal

adjoint state $\Phi(x) := \Phi(x, t)$, satisfying the MFG system. Since this is a functional of a single variable, many optimization techniques for high-dimensional problems can be applied, for example, coordinate gradient descent methods. Also, numerical methods for special cases of potential games were introduced in [7]. They transform the optimal control problem into a regularized linear program. Unlike these methods, our methods can be applied to general Lagrangians for optimal control problems in density space. Yet another well-known line of research focuses on stationary MFG systems [6, 4], for which proximal splitting methods have been used. They are different from our focus on time-dependent MFGs. We would like to emphasize that rigorous derivation of the Hopf formula suggested is performed only for a relatively narrow case; however, the practical investigation allows to conjecture that the actual range of applicability should be wider.

The Hopf maximization principle gives us an optimal balance between the indirect method (Pontryagin's maximum principle), e.g. the well-known MFG system (49)-(51) below in [27], and the direct method (optimization over the spaces of curves), e.g. the primal-dual formulation in [1, 2, 3] and Hopf formula (57)-(59) in [27]. This balance leads to computational efficiency. There are several existing formulations for solving HJD numerically: (i) the original formulation in (2.2a)-(2.2b) or its resulting (primal) Lagrangian formulation, (ii) the intermediate primal-dual formulation in [1, 2, 3], (iii) the dual formulation (Hopf formula) (2.3) in (57)-(59) in [27], (iv) the resulting KKT optimality condition (2.1) (the MFG system (49)-(51) in [27]), and (v) the proposed Hopf formulation (3.4) in this paper. Under suitable conditions, the five formulations are equivalent, but their effects on computation are different. Formulations (i), (ii), and (iii) involve a larger number of variables, which lead to high complexities on problems with high dimensions and long time intervals. The reason of that is that in these three formulations, both $\{\rho(x, s)\}_{x \in X, s \in [t, T]}$ and $\{\Phi(x, s)\}_{x \in X, s \in [t, T]}$ needs to be stored for each iteration; (iv) is a forward-backward system that needs different numerical methods. Our approach (v) is a balance between the indirect and direct methods and reduces the number of variables to a single terminal adjoint state $\Phi(x, t)$. With this, since the computation of the PDE is done by moving forward in time, at each time step s , only one time-step of $\{\rho(x, s)\}_{x \in X}$ and $\{\Phi(x, s)\}_{x \in X}$ is needed, and also only the initial $\Phi(x, t)$ is needed to be saved for each iteration after the whole descent step. The memory requirement is thus reduced in our approach. On the other hand, we keep a variable and a functional such that our algorithm produces a descending sequence that converges to a local minimum.

We utilize the coordinate descent method, which avoids the difficulties coming from a non-smooth functional. We remark that the proposed approach can handle Hamiltonians of homogeneous degree 1, which can be used as a mean-field level set approach for the reachability problem. Moreover, we choose the Hopf formulation to handle the case where the Hamiltonian H is non-convex. We propose to check the computed limit (i.e., whether it is a global minimum) via the condition $\Phi(x, 0) \in \partial G(\rho(x, 0))$.

The rest of this paper is organized as follows. In Section 2, we briefly review potential MFGs and related HJD and formally derive the Hopf formula in density space. We also rigorously established some of the solutions to HJD under mild assumptions and convexity assumptions, and described a solution in a special case with non-convex Hamiltonian. We then propose a rigorous approach on discrete grid approximations of optimal control problems and show the validity of the Hopf formula under proper assumptions in Section 3. In Section 4, we design a fast multi-level random coordinate descent method for solving the discrete Hopf formula that we obtained in Section 3. Several numerical examples are presented in Section 5 to illustrate the effectiveness of the proposed algorithm.

2 Hopf formula in mean field games

In this section, we briefly review potential MFGs. They are related to optimal control problems in density space, which induce Hamilton-Jacobi equations in density space. We propose the Hopf formula in density space for subsequent numerical computation.

2.1 Potential mean field games

Consider a differential game played by one population, which contains countably infinitely many agents. The state of each agent is described by a point on a compact manifold X , e.g. d -dimensional torus. The choice of a compact manifold goes in line with classical literature, e.g. in the seminal work of [9] on proving the convergence of N -player game to MFG master equation and the important works of [1, 2, 3] on numerical methods on MFG. A main reason of this choice is that compactness in many cases simplifies the proofs. The choice as a d -dimensional torus is for convinient sake since the space is flat. We follow this convension to consider X being compact. The aggregated state of the population can now be described by the population state $\rho(x) \in \mathcal{P}(X) = \{\rho(\cdot): \int_X \rho(x)dx = 1, \rho(x) \geq 0\}$, where $\rho(x)$ represents the population density of players choosing strategy $x \in X$. The game assumes that each player's cost is independent of his/her identity (autonomous game). In a differential game, each agent plays the game dynamically facing the same Lagrangian $L: T(X) \rightarrow \mathbb{R}$, where $T(X)$ represents the tangent space of X . The term "mean field" makes sense when each player's potential energy f and terminal cost g rely on mean-field quantities of all other players' choices, mathematically written as $f, g: X \times \mathcal{P}(X) \rightarrow \mathbb{R}$. The controls of the game is now describes by a tangent vector field, or a smooth section of $T(X)$.

Under quite general and mild assumptions, it is proven rigorously e.g. in [8, 9], that a Nash equilibrium (NE) exists for a class of mean field games, and is also the limit of N -player game as $N \rightarrow \infty$ of a N -player ϵ -Nash equilibrium in the setting of closed loop strategy (possibly being associated with an open-loop control obtained by solving an ODE or SDE) when the cost functions are permutation invariant. To recall, the NE now describes a status in which no player in population is willing to change his/her strategy unilaterally. In a MFG, it is represented as a primal-dual dynamical system:

$$\begin{cases} \partial_s \rho(x, s) + \nabla_x \cdot (\rho(x, s) D_p H(x, \nabla_x \Phi(x, s))) = 0 \\ \partial_s \Phi(x, s) + H(x, \nabla_x \Phi(x, s)) + f(x, \rho(\cdot, s)) = 0 \\ \rho(x, t) = \rho(x), \quad \Phi(x, 0) = g(x, \rho(\cdot, 0)), \end{cases} \quad (2.1)$$

where the Hamiltonian H is defined as

$$H(x, p) := \sup_{v \in T_x X} \langle v, p \rangle - L(x, v).$$

Here H relates to the Lagrangian L through a Legendre transform in v . And $\rho(s, \cdot)$ represents the population state at time s satisfying the continuity equation while $\Phi(s, \cdot)$ governs the velocity of population according to the Hamilton-Jacobi equation.

A game is called a potential game when there exists a differentiable potential energy $F: \mathcal{P}(X) \rightarrow \mathbb{R}$ and terminal cost $G: \mathcal{P}(X) \rightarrow \mathbb{R}$ such that

$$\delta_{\rho(x)} F(\rho) = f(x, \rho), \quad \delta_{\rho(x)} G(\rho) = g(x, \rho),$$

where $\delta_{\rho(x)}$ is the L^2 first variation operator. The above definition represents that the incentives of all the players can be globally modeled by a functional called the potential [9]. In this case,

the game is modeled as the following optimal control problem in density space:

$$\inf_{\rho, v} \left\{ \int_0^t \left[\int_X L(x, v(x, s)) \rho(x, s) dx - F(\rho(\cdot, s)) \right] ds + G(\rho(\cdot, 0)) \right\}, \quad (2.2a)$$

where the infimum is taken among all vector fields $v(x, s)$ and densities $\rho(x, s)$ subject to the continuity equation

$$\begin{cases} \frac{\partial}{\partial s} \rho(x, s) + \nabla \cdot (\rho(x, s) v(x, s)) = 0, & 0 \leq s \leq t, \\ \rho(x, t) = \rho(x). \end{cases} \quad (2.2b)$$

It can be shown that, under suitable conditions of L , F , G , NEs are minimizers of potential games. In other words, every NE (2.1) satisfies the Euler-Lagrange equation (Karush-Kuhn-Tucker conditions) of the optimal control problem (2.2). Let $\mathcal{H}(\rho, \Phi)$ denote the total Hamiltonian defined over the primal-dual pair (ρ, Φ) :

$$\mathcal{H}(\rho, \Phi) := \int_X \rho(x) H(x, \nabla_x \Phi(x)) dx + F(\rho(\cdot)),$$

where δ_ρ , δ_Φ are L_2 first variations w.r.t. ρ and Φ . Then, NE (2.1) is given as

$$\begin{cases} \partial_s \rho_s = \delta_{\Phi_s} \mathcal{H}(\rho_s, \Phi_s), & \partial_s \Phi_s = -\delta_{\rho_s} \mathcal{H}(\rho_s, \Phi_s) \\ \rho_t = \rho(x), & \Phi_0 = \delta_{\rho_0} G(\rho_0). \end{cases}$$

The time evolution of the minimal value in optimal control satisfies the Hamilton-Jacobi equation. In the case of density space, the optimal value function in (2.2a) is denoted by $U: [0, +\infty) \times \mathcal{P}(X) \rightarrow \mathbb{R}$. As shown in [22, 23], U satisfies the Hamilton-Jacobi equation in density space

$$\begin{cases} \partial_s U(s, \rho(\cdot)) + \mathcal{H}(\rho(\cdot), \delta_\rho U) = 0 \\ U(0, \rho) = G(\rho). \end{cases}$$

Here, HJD is a functional partial differential equation. If U is solved, then its characteristics in density space, i.e. (ρ, Φ) , are known. In particular, $\Phi(t, x) = \delta_{\rho(x)} U(t, \rho)$. Thus, NE (2.1) is found. Next, we shall design a fast numerical algorithm for HJD.

2.2 Hopf formula in density space

Our approach is based on a primal-dual reformulation of the optimal control problem (2.2), which we call the Hopf formula, assuming the duality gap is zero and the saddle point problem admits a saddle point.

Proposition 2.1 (Hopf formula in density space). *Assume the duality gap between the primal problem (2.2) and its dual problem is zero and the resulting primal dual problem admits a saddle point, then*

$$\begin{aligned} U(t, \rho) &:= \sup_{\Phi} \left\{ \int_X \rho(x) \Phi(x) dx - \int_0^t \left(F(\rho(\cdot, s)) - \int_X \rho(s, x) \delta_{\rho_s} F(\rho(\cdot, s)) dx \right) ds - G^*(\Phi(\cdot, 0)) : \right. \\ &\quad \left. \begin{aligned} \partial_s \rho(x, s) + \nabla \cdot (\rho(x, s) D_p H(x, \nabla \Phi(x, s))) &= 0 \\ \partial_s \Phi(x, s) + H(x, \nabla \Phi(x, s)) + \delta_{\rho_s(x)} F(\rho(\cdot, s)) &= 0 \\ \rho(x, t) &= \rho(x), \quad \Phi(x, t) = \Phi(x) \end{aligned} \right\} \end{aligned} \quad (2.3)$$

where

$$G^*(\Phi(\cdot, 0)) := \sup_{\rho(\cdot, 0) \in \mathcal{P}(X)} \left\{ G(\rho(\cdot, 0)) - \int_X \rho(x, 0) \Phi(x, 0) dx \right\}.$$

Formal derivation. We first define the flux function $m(s, x) := \rho(s, x)v(s, x)$ in (2.2). Thus problem (2.2) takes the form

$$U(t, \rho) := \inf_{m, \rho} \left\{ \int_0^t \left[\int_X L \left(x, \frac{m(x, s)}{\rho(x, s)} \right) \rho(x, s) dx - F(\rho(\cdot, s)) \right] ds + G(\rho(\cdot, 0)) \right\},$$

where the infimum is taken among all flux functions $m(x, s)$ and densities $\rho(x, s)$ subject to

$$\begin{cases} \frac{\partial}{\partial s} \rho(x, s) + \nabla \cdot m(x, s) = 0, & 0 \leq s \leq t, \\ \rho(x, t) = \rho(x). \end{cases}$$

Next, we compute the dual of the optimal control problem (2.2). Assume that, under suitable assumptions of F , G , L , the duality gap of optimal control problem (2.2) is zero. Hence we can switch “inf” and “sup” signs in our derivations. Let the Lagrange multiplier of continuity equation (2.2b) be denoted by $\Phi(x, s)$. The optimal control problem (2.2) becomes

$$\begin{aligned} U(t, \rho) &= \inf_{m(\cdot, s), \rho(\cdot, s), \rho(\cdot, t) = \rho} \sup_{\Phi(\cdot, t)} \left\{ \int_0^t \int_X L \left(x, \frac{m(x, s)}{\rho(x, s)} \right) \rho(x, s) dx ds - \int_0^t F(\rho(\cdot, s)) ds + G(\rho(\cdot, 0)) \right. \\ &\quad \left. + \int_0^t \int_X (\partial_s \rho(x, s) + \nabla \cdot m(x, s)) \Phi(x, s) dx ds \right\} \\ &= \sup_{\Phi(\cdot, s)} \inf_{m(\cdot, s), \rho(\cdot, s), \rho(\cdot, t) = \rho} \left\{ \int_0^t \int_X L \left(x, \frac{m(x, s)}{\rho(x, s)} \right) \rho(x, s) dx ds - \int_0^t F(\rho(\cdot, s)) ds + G(\rho(\cdot, 0)) \right. \\ &\quad \left. + \int_0^t \int_X (\partial_s \rho(x, s) + \nabla \cdot m(x, s)) \Phi(x, s) dx ds \right\} \\ &= \sup_{\Phi(\cdot, s)} \inf_{m(\cdot, s), \rho(\cdot, s), \rho(\cdot, t) = \rho} \left\{ \int_0^t \int_X \left[L \left(x, \frac{m(x, s)}{\rho(x, s)} \right) - \frac{m(x, s)}{\rho(x, s)} \cdot \nabla \Phi(x, s) \right] \rho(x, s) dx ds \right. \\ &\quad \left. - \int_0^t F(\rho(\cdot, s)) ds + G(\rho(\cdot, 0)) + \int_0^t \int_X \partial_s \rho(x, s) \Phi(x, s) dx ds \right\} \\ &= \sup_{\Phi(\cdot, s)} \inf_{\rho(\cdot, s), \rho(\cdot, t) = \rho} \left\{ - \int_0^t \int_X \rho(x, s) H(x, \nabla \Phi(x, s)) dx ds - \int_0^t F(\rho(\cdot, s)) ds \right. \\ &\quad \left. + G(\rho(\cdot, 0)) + \int_0^t \int_X \partial_s \rho(x, s) \Phi(x, s) dx ds \right\}, \end{aligned}$$

where the third equality is given by integration by parts, and the fourth equality follows by the Legendre transform in the third equality, i.e., with $v(x, s) := \frac{m(x, s)}{\rho(x, s)}$,

$$H(x, \nabla \Phi) = \sup_{v \in TX} \nabla \Phi \cdot v - L(x, v).$$

By integration by parts w.r.t. s for the functional $\int_0^t \int_X \partial_s \rho(x, s) \Phi(x, s) dx ds$, we obtain

$$\begin{aligned} U(t, \rho) = \sup_{\Phi(\cdot, s)} \inf_{\rho(\cdot, s), \rho(\cdot, t) = \rho} & \left\{ - \int_0^t \int_X \rho(x, s) H(x, \nabla \Phi(x, s)) dx ds - \int_0^t F(\rho(\cdot, s)) ds \right. \\ & + G(\rho(\cdot, 0)) - \int_X \rho(x, 0) \Phi(x, 0) dx \\ & \left. + \int_X \rho(x, t) \Phi(x, t) dx - \int_0^t \int_X \rho(x, s) \partial_s \Phi(x, s) dx ds \right\}. \end{aligned}$$

Then, we see that, separating the supremum over the functions $\Phi(\cdot, s)$ into first taking supremum over $\Phi(\cdot, s)$ given $\Phi(\cdot, t) = \Phi$ then taking supremum over Φ , we obtain

$$\begin{aligned} U(t, \rho) = \sup_{\Phi} \sup_{\Phi(\cdot, s), \Phi(\cdot, t) = \Phi} \inf_{\rho(\cdot, s), \rho(\cdot, t) = \rho} & \left\{ - \int_0^t \int_X \rho(x, s) H(x, \nabla \Phi(x, s)) dx ds - \int_0^t F(\rho(\cdot, s)) ds \right. \\ & \left. - G^*(\Phi(\cdot, 0)) + \int_X \rho(x, t) \Phi(x, t) dx - \int_0^t \int_X \rho(x, s) \partial_s \Phi(x, s) dx ds \right\}. \end{aligned} \quad (2.4)$$

We optimize the above formula w.r.t. $\rho(x, s)$ and $\phi(x, s)$. Suppose for a fixed $\Phi = \Phi(t, \cdot)$, the saddle point problem

$$\begin{aligned} \sup_{\Phi(\cdot, s), \Phi(\cdot, t) = \Phi} \inf_{\rho(\cdot, s), \rho(\cdot, t) = \rho} & \left\{ - \int_0^t \int_X \rho(x, s) H(x, \nabla \Phi(x, s)) dx ds - \int_0^t \int_X \rho(x, s) \partial_s \Phi(x, s) dx ds \right. \\ & \left. - \int_0^t F(\rho(\cdot, s)) ds - G^*(\Phi(\cdot, 0)) + \int_X \rho(x, t) \Phi(x, t) dx \right\} \end{aligned}$$

has a solution. It is simple to check that this saddle point satisfies (2.1). Substituting (2.1) into (2.4), we derive the Hopf formula (2.3). \square

Equation (2.3) can be viewed as the Hopf formula of the optimal control problem (2.2). This goes in line with [16, 15, 14]. That means that (2.3) contains an optimization problem and uses a minimal number of unknown variables. We develop fast algorithms based on this formula.

Remark 2.2. *We would like to emphasize that our derivation heavily relies on suitable assumptions of F , G , L that the duality gap of optimal control problem (2.2) is zero, and therefore we may switch “inf” and “sup” signs in our derivations. This may happen when there exists a saddle point for the saddle point problem, which can be imposed under convexity/monotonicity, coercivity and smoothness of $(-F)$, G and L . We would like to remark, however, that although our derivation of the generalized Hopf formula is provided only for a relatively narrow case, practical investigation allows us to conjecture the validity of the formula shall be wider.*

Remark 2.3. *When $(-F)$, G and L are convex and smooth, the discrete formulation of the primal dual formulation of (2.2) has been used for numerical computation in [1, 2, 3] along with Newton’s method. We, on the other hand, prefer sticking to the formulation (2.3) since we hope to solve for non-convex $(-F)$, G and L with nonsmooth $H(x, p)$, while keeping a minimal number of variables. In addition, the Hopf formula (2.3) can be further simplified into*

$$U(t, \rho) = \sup_{\Phi(\cdot, s)} \left\{ \int_X \rho(x) \Phi(x) dx - \int_0^t F^*(\Phi(\cdot, s)) ds - G^*(\Phi(\cdot, 0)) \right\}, \quad (2.5)$$

which coincides with (57)-(59) in [27]. However, the formulation (2.5), similar to the Lagrangian formulation (2.2), has more independent variables after discretization of $\Phi(x, s)$. Hence, it is not ideal for numerical computation.

Remark 2.4. The Hopf formula (2.3) is also related to the dual formulation of an optimal transport problem. When $F(\rho) = 0$, the primal equation in (2.3) can be dropped. Let $\rho = \rho_1$ in

$$U(t, \rho_1) = \sup_{\Phi} \left\{ \int_X \rho(x) \Phi(x) dx - G^*(\Phi(\cdot, 0)) : \partial_s \Phi(x, s) + H(x, \nabla \Phi(x, s)) = 0, \Phi(x, t) = \Phi(x) \right\}.$$

This is precisely the Kantorovich dual of the optimal transport problem from ρ_0 to ρ_1 when we choose $G(\rho) = \iota_{\rho_0}(\rho)$ and let $t = 1$. Here, for a set A and a subset $B \subset A$, the indicator function $\iota_B : A \rightarrow \{0, \infty\}$ is defined as

$$\iota_B(x) = \begin{cases} 0 & \text{if } x \in B \\ \infty & \text{if } x \notin B \end{cases}.$$

If $B = \{x_0\}$ is a singleton, we write $\iota_{x_0}(x) := \iota_{\{x_0\}}(x)$, abusing the notation.

Remark 2.5. As in remark 2.4, our Hopf formula (2.3) reduces to Monge-Kantorovich duality of the optimal transport with a specific choices of F , G and t . Moreover, the simplified formula can be used to compute the proximal map of p -Wasserstein distance in the L^2 sense. Let us recall the connection between optimal transport and (2.2). The optimal transport problem can be formulated in an optimal control problem in density space, known as the Benamou-Brenier formula [32]. Consider $L(x, q) = \frac{1}{2}|q|_2^2$. Then,

$$\begin{aligned} U(1, \rho_1) & \quad (\text{Definition}) \\ & = \inf_{v(\cdot, s), \rho(\cdot, s)} \left\{ \int_0^1 \int_X L(v(x, s)) \rho(x, s) dx ds + G(\rho(x, 0)) : \right. \\ & \quad \left. \partial_s \rho + \nabla \cdot (\rho v) = 0, \rho(1) = \rho_1 \right\} \\ & \quad (\text{Benamou-Brenier}) \\ & = \inf_{\rho_0} \left\{ (W_p(\rho_0, \rho_1))^p + G(\rho_0) \right\} \\ & \quad (\text{Kantorovich duality}) \\ & = \inf_{\rho_0} \sup_{\Phi_1} \left\{ \int_Y \Phi(y, 1) \rho_1(y) dy - \int_X \Phi(x, 0) \rho_0(x) dx + G(\rho_0) : \right. \\ & \quad \left. \partial_s \Phi(x, s) + H(\nabla \Phi(x, s)) \leq 0, \Phi(x, 1) = \Phi_1(x) \right\} \\ & \quad (\text{Convexity of } G, H) \\ & = \sup_{\Phi_1} \left\{ \int_Y \Phi(y, 1) \rho_1(y) dy - G^*(\Phi(\cdot, 0)) : \right. \\ & \quad \left. \partial_s \Phi(x, s) + H(\nabla \Phi(x, s)) = 0, \Phi(x, 1) = \Phi_1(x) \right\}, \end{aligned}$$

where $W_p(\rho_0, \rho_1)$ is the p -Wasserstein metric which can be defined via the Benamou-Brenier formulation as follows:

$$(W_p(\rho_0, \rho_1))^p := \inf_{v(\cdot, s), \rho(\cdot, s)} \left\{ \int_0^1 \int_X L(v(x, s)) \rho(x, s) dx ds : \partial_s \rho + \nabla \cdot (\rho v) = 0, \rho(0) = \rho_0, \rho(1) = \rho_1 \right\}.$$

If one aims to consider a general optimization problem over G regularized by W_p^p as in

$$\min_{\rho_0} \{\beta W_p^p(\rho_0, \rho) + G(\rho_0)\},$$

we can either apply the above formulation directly or apply a splitting method, in which we need the proximal maps of W_p^p (in L^2 sense) as

$$\begin{aligned} \text{Prox}_{\beta W_p^p(\cdot, \rho)}(\rho_1) &= \underset{\rho_0}{\operatorname{argmin}} \left\{ \beta W_p^p(\rho_0, \rho) + \frac{1}{2} \|\rho_0 - \rho_1\|^2 \right\} \\ &= \rho_1 - \beta \tilde{\Phi}, \end{aligned}$$

where

$$\begin{aligned} \tilde{\Phi} &:= \underset{\Phi}{\operatorname{argmax}} \left\{ \int_X \rho(x) \tilde{\Phi}(x) dx - \int_X \rho_1(x) \Phi(\tilde{\Phi}, 0, \cdot) dx + \frac{\beta}{2} \|\Phi(\tilde{\Phi}, 0, \cdot)\|^2 : \right. \\ &\quad \left. \partial_s \Phi + H(x, \nabla_x \Phi) = 0, \Phi(\tilde{\Phi}, t, \cdot) = \tilde{\Phi} \right\}. \end{aligned}$$

2.3 Some rigorously known solutions to HJD for specific Hamiltonians

In this subsection, we would like to describe some rigorously known solutions to the HJD which provides us a platform to compare them with the Hopf solution and understand the solutions that we computed in Section 5.

2.3.1 Optimal transport problem with some types of general cost

In this subsection we will discuss the viscosity solution to a HJ PDE coming from a general optimal transport problem. Some of the following results are standard textbook materials and shown in e.g. [5]. We are still providing here for completeness.

In what follows let us denote \mathcal{A} to be either $\mathcal{A} = C^{0,1}[0, T]$ the Lipschitz space, or $\mathcal{A} = W^{1,q}[0, T]$ a Sobelov space for some $1 < q < \infty$.

Lemma 2.6. *Let $L : T(X) \rightarrow [0, +\infty]$ be such $L(x(s), \dot{x}(s))$ is Lebesgue measurable on $[0, T]$ for all $x(s) \in \mathcal{A}$. Define, for $0 < T \leq \infty$,*

$$d_{L,T}(x, y) := \inf_{\gamma \in \mathcal{A}} \left\{ \int_0^T L(\gamma(s), v(s)) ds : \dot{\gamma} = v; \gamma(0) = x, \gamma(T) = y \right\} \geq 0$$

Then for all $x, y, z \in X$, the followings hold:

1. $d_{L,T}(x, y) = d_{L,T}(y, x)$
2. $d_{L,T_1+T_2}(x, z) \leq d_{L,T_1}(x, y) + d_{L,T_2}(y, z)$
3. $d_{L,T}(x, y) \leq d_{L,S}(x, y)$ if $0 < S < T \leq \infty$.
4. If we have L satisfying $L(x, v) \leq r^{-1} L(x, rv)$ for $0 < r \leq 1$, then for $0 < S < T \leq \infty$

$$d_{L,T}(x, y) = d_{L,T}(x, y) \quad \text{and hence} \quad d_{L,T}(x, z) \leq d_{L,T}(x, y) + d_{L,T}(y, z).$$

Proof. The integral is well-defined since $\dot{\gamma}$ exists a.e. w.r.t. Lebegue measure, and $L(x(s), \dot{\gamma}(s))$ is Lebegue on $[0, T]$ if $\gamma \in \mathcal{A}$. The first three statements are direct consequences of the definition of infimum via change of variable and set inclusions. The last statement comes from a change of variable which gives

$$d_{L,S}(x, y) \leq d_{\frac{T}{S}L(\cdot, \frac{S}{T}\cdot), S}(x, y) = d_{L,T}(x, y) \leq d_{L,S}(x, y).$$

□

Notice the above did not prevent $d_{L,T}(x, y) = \infty$ nor does this necessarily give $d_{L,T}(x, y) = 0 \Rightarrow x = y$, and hence this is not a distance function even when $L(x, v) \leq r^{-1}L(x, rv)$ for $0 < r \leq 1$. Another remark is that we are not claiming a minimizer exists in any sense in this very general setting. However, the followings are true.

Now let us define the following function for any compact subset $C \subset X$.

Definition 2.7. Let $(x, v) \mapsto L(x, v)$ be lower semi-continuous and C be compact. Let us define

$$Proj_{C, d_{L,T}}(x) := \left\{ y \in C : \exists \{y_n\}_{n \in \mathbb{N}} \in C \text{ s.t. } d_{L,T}(x, y_n) \rightarrow \inf_{z \in C} d_{L,T}(x, z) \text{ and } y_n \rightarrow y \text{ as } n \rightarrow \infty \right\}$$

to be a set-valued function.

Definition 2.8. Given $(x, v) \mapsto L(x, v)$ be lower semi-continuous and C be compact. The projection operator $Proj_{C, d_{L,T}}$ is said to satisfies property (Proj) for all $x \in X$, $y_x \in Proj_{C, d_{L,T}}$, there exists a minimizing curve (geodesic) $\gamma_{x, y_x} \in \mathcal{A}$ such that

$$\int_0^T L(\gamma_{x, y_x}(s), \dot{\gamma}_{x, y_x}(s)) ds = d_{L,T}(x, y_x),$$

and hence

$$Proj_{C, d_{L,T}}(x) = \operatorname{argmin}_{z \in C} d_{L,T}(x, z)$$

as a set-valued function.

Lemma 2.9. Assume L is lower semi-continuous. Then

1. $Proj_{C, d_{L,T}}(x)$ is a non-empty set-valued function.
2. For all $y_x \in Proj_{C, d_{L,T}}(x)$, there exists a family of curves $\{\gamma_{x, y_x, n}(t)\}_{n \in \mathbb{N}} \in \mathcal{A}$ satisfies $\gamma_{x, n}(0) = x$, $C \ni \gamma_{x, n}(T) \rightarrow y_x$ as $n \rightarrow \infty$ and

$$\int_0^T L(\gamma_{x, y_x, n}(s), \dot{\gamma}_{x, y_x, n}(s)) ds \rightarrow \inf_{z \in C} d_{L,T}(x, z) \text{ as } n \rightarrow \infty.$$

3. Let $\mathcal{A} = C^{0,1}[0, T]$. If there exists $R > 0$ such that $L(x, v) = \infty$ for $\|v\| > R$, then for all $y_x \in Proj_{C, d_{L,T}}(x)$, there exists R -Lipschitz geodesic $\gamma_{x, y_x}(t) \in \mathcal{A}$ with $\gamma_{x, y_x}(0) = x$, $C \ni \gamma_{x, y_x}(T) = y_x$ and

$$\int_0^T L(\gamma_{x, y_x}(s), \dot{\gamma}_{x, y_x}(s)) ds = d_{L,T}(x, y_x) = \inf_{z \in C} d_{L,T}(x, z),$$

Hence $Proj_{C, d_{L,T}}$ satisfies property (Proj).

4. Let $\mathcal{A} = W^{1,q}[0, 1]$. Assume $\frac{1}{K}\|v\|^q - K \leq L(x, v) \leq K\|v\|^q + K$ for some K where $1 < q < \infty$; and $(x, v) \mapsto L(x, v)$ is differentiable w.r.t. x and convex w.r.t. v . Write $p := q/(q-1)$, there exists an $C^{0,\alpha}$ geodesic $\gamma_{x,y_x}(t) \in \mathcal{A}$ for $\alpha < 1/p$ with $\gamma_{x,y_x}(0) = x$, $C \ni \gamma_{x,y_x}(T) = y_x$ and

$$\int_0^T L(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)) ds = d_{L,T}(x, y_x) = \inf_{z \in C} d_{L,T}(x, z),$$

Hence $\text{Proj}_{C,d_{L,T}}$ satisfies property (Proj).

Proof. 1. First, notice that $L(x(s), v(s))$ is a non-negative Lebesgue measurable function, and thus $d_{L,T}$ is well-defined. Now fix $x \in X$, consider the problem $\inf_{z \in C} d_{L,T}(x, z) \geq 0$, by definition, there exists a sequence $\{y_{x,n} \in C\}_{n \in \mathbb{N}}$ such that $d_{L,T}(x, y_{x,n}) \rightarrow \inf_{y \in C} d_{L,T}(x, y)$. By compactness of C and renaming a subsequence, $y_{x,n} \rightarrow y_x$ for some $y_x \in C$ as $n \rightarrow \infty$. Hence $\text{Proj}_{C,d_{L,T}}(x)$ is non-empty.

2. Let $y_x \in \text{Proj}_{C,d_{L,T}}(x)$ be given. Then there exists $\{y_{x,n}\}_{n \in \mathbb{N}} \in C$ such that $d_{L,T}(x, y_{x,n}) \rightarrow \inf_{z \in C} d_{L,T}(x, z)$ and $y_{x,n} \rightarrow y_x$ as $n \rightarrow \infty$. Now from the definition of $d_{L,T}(x, y_{x,n})$, for each n , there exists minimizing sequence $\{\gamma_{x,y_{x,n}}^m(t)\}_{0 \leq t \leq T} \in C^{0,1}[0, T]$ such that $\gamma_{x,y_{x,n}}^m(0) = x$, $\gamma_{x,y_{x,n}}^m(T) = y_{x,n}$ and

$$\int_0^T L(\gamma_{x,y_{x,n}}^m(s), \dot{\gamma}_{x,y_{x,n}}^m(s)) ds \rightarrow d_{L,T}(x, y_{x,n}) \text{ as } m \rightarrow \infty.$$

A standard diagonal argument now provides an increasing sequence $\{m_n\}$ such that $\{\gamma_{x,y_{x,n}}(t) := \gamma_{x,y_{x,n}}^{m_n}(t)\}_{0 \leq t \leq T}$ satisfying the desired properties.

3. Since there exists $R > 0$ such that $L(x, v) = \infty$ for $\|v\| > R$, we can always choose the minimizing sequences $\{\gamma_{x,y_{x,n}}^m(t)\}_{0 \leq t \leq T} \in C^{0,1}[0, T]$ with $\|\gamma_{x,y_{x,n}}^m\|_{C^{0,1}} \leq R$ in Part 2. Hence the resulting diagonal sequence $\{\gamma_{x,y_{x,n}}(t)\}_{0 \leq t \leq T}$ is R -Lipschitz and equicontinuous. By Arzela-Ascoli theorem and again renaming the subsequence, we have $\gamma_{x,y_{x,n}} \rightarrow \gamma_{x,y_x}$ uniformly as $n \rightarrow \infty$ for some R -Lipschitz $\gamma_{x,y_x}(s)$ with $\gamma_{x,y_x}(0) = x$ and $\gamma_{x,y_x}(T) = y_x$. Now by Fatou's lemma and lower semi-continuity of L ,

$$\begin{aligned} d_{L,T}(x, y_x) &\leq \int_0^T L(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)) ds \\ &\leq \int_0^T \liminf_n L(\gamma_{x,y_{x,n}}(s), \dot{\gamma}_{x,y_{x,n}}(s)) ds \\ &\leq \liminf_n \int_0^T L(\gamma_{x,y_{x,n}}(s), \dot{\gamma}_{x,y_{x,n}}(s)) ds \\ &= \inf_{z \in C} d_{L,T}(x, z) \leq d_{L,T}(x, y_x). \end{aligned}$$

4. Let $\mathcal{A} = W^{1,q}[0, 1]$. If $K\|v\|^q - K \leq L(x, v) \leq K\|v\|^q + K$ for some K where $1 < q < \infty$, we can always choose the minimizing sequences $\{\gamma_{x,y_{x,n}}^m(t)\}_{0 \leq t \leq T} \in W^{1,q}[0, 1]$ in Part 2 with

$$\int_0^T \|\dot{\gamma}_{x,y_{x,n}}^m(s)\|^q ds < C$$

for some C . By Morrey's inequality, the resulting diagonal sequence $\{\gamma_{x,y_{x,n}}(t)\}_{0 \leq t \leq T}$ is then uniformly bounded in $C^{0,1/p}$. By Arzela-Ascoli theorem, for all $\alpha < 1/p$, with

a renaming of a subsequence, we have $\gamma_{x,y_x,n} \rightarrow \gamma_{x,y_x}$ in $C^{0,\alpha}$ norm as $n \rightarrow \infty$. Moreover, by reflexivity of strongly convex space and Mazur lemma, we have weak compactness of $W^{1,p}[0, T]$ and hence with a renaming of a subsequence, we may further obtain $\gamma_{x,y_x} \in W^{1,p}[0, T]$ and $\gamma_{x,y_x,n} \rightharpoonup \gamma_{x,y_x}$ in $L^q[0, T]$ and $\dot{\gamma}_{x,y_x,n} \rightharpoonup \dot{\gamma}_{x,y_x}$ in $L^q[0, T]$. Now by differentiability of L w.r.t. x , convexity of L w.r.t. v , we have $r(x, v) := \partial_x L(x, v)$ and a choice of $s(x, v) \in \partial_v L(x, v)$ subdifferential of L w.r.t. v at (x, v) with

$$\begin{aligned} & \langle r(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)), \gamma_{x,y_x}(s) - \gamma_{x,y_x,n}(s) \rangle + \langle s(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)), \dot{\gamma}_{x,y_x}(s) - \dot{\gamma}_{x,y_x,n}(s) \rangle \\ & \leq L(\gamma_{x,y_x,n}(s), \dot{\gamma}_{x,y_x,n}(s)) - L(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)) + o(\|\gamma_{x,y_x} - \gamma_{x,y_x,n}\|_{C^0[0,1]}) \end{aligned}$$

Now since $|r(x, v)| \leq \tilde{K}(|v|^q + 1)$ and $|s(x, v)| \leq \tilde{K}(|v|^{q-1} + 1)$ for some \tilde{K} , hence $r(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)) \in L^1[0, T]$ and $s(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)) \in L^p[0, T]$, then by pointwise convergence and weak convergence, we have as $n \rightarrow \infty$,

$$\begin{aligned} \left| \int_0^T \langle r(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)), \gamma_{x,y_x}(s) - \gamma_{x,y_x,n}(s) \rangle ds \right| & \leq \|\gamma_{x,y_x} - \gamma_{x,y_x,n}\|_{C^0[0,T]} \rightarrow 0 \\ \int_0^T \langle s(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)), \dot{\gamma}_{x,y_x}(s) - \dot{\gamma}_{x,y_x,n}(s) \rangle ds & \rightarrow 0. \end{aligned}$$

Therefore

$$\begin{aligned} 0 &= \lim_n \int_0^T \langle r(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)), \gamma_{x,y_x}(s) - \gamma_{x,y_x,n}(s) \rangle ds \\ &\quad + \lim_n \int_0^T \langle s(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)), \dot{\gamma}_{x,y_x}(s) - \dot{\gamma}_{x,y_x,n}(s) \rangle ds \\ &\leq \liminf_n \int_0^T L(\gamma_{x,y_x,n}(s), \dot{\gamma}_{x,y_x,n}(s)) ds - \int_0^T L(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)) ds \end{aligned}$$

Therefore $\gamma_{x,y_x}(s) \in \mathcal{A} \cap C^{0,\alpha}[0, T]$ and the definition of $\gamma_{x,y_x,n}(s)$ gives

$$\begin{aligned} d_{L,T}(x, y_x) &\leq \int_0^T L(\gamma_{x,y_x}(s), \dot{\gamma}_{x,y_x}(s)) ds \\ &\leq \liminf_n \int_0^T L(\gamma_{x,y_x,n}(s), \dot{\gamma}_{x,y_x,n}(s)) ds \\ &\leq \inf_{z \in C} d_{L,T}(x, z) \\ &\leq d_{L,T}(x, y_x). \end{aligned}$$

□

Note that with our general definition, if $d_{L,T}(x, y) = \infty$ for all $y \in C$, $\text{Proj}_{C,d_{L,T}}(x) = C$. Also, if $C = \{y\}$, then $\text{Proj}_{C,d_{L,T}}(x) = \{y\}$.

Lemma 2.10. *Assume L is lower semi-continuous and $C \subset X$ be a compact subset. Assume further that $\text{Proj}_{C,d_{L,T}}$ satisfies property (Proj). Then there exists (a possibly non-unique) $Y : X \rightarrow C, x \mapsto Y(x)$ and a (possibly non-unique) family of curves $\{\gamma_{x,Y(x)}(t) \in \mathcal{A}\}_{x \in X, 0 \leq t \leq T}$ with*

1. $Y(x) \in \text{Proj}_{C,d_{L,T}}(x)$;
2. $\{\gamma_{x,Y(x)}(t)\}$ satisfies conclusions in Lemma 2.9, part 3.

3. If $\gamma_{x_1, Y(x_1)}(t) = \gamma_{x_2, Y(x_2)}(t)$ for some $x_1, x_2 \in X$ and $0 \leq t \leq T$, then $\gamma_{x_1, Y(x_1)}(s) = \gamma_{x_2, Y(x_2)}(s)$ for all $t < s \leq T$.

Proof. Now for all $x \in X$, applying Lemma 2.9, $\text{Proj}_{C, d_{L, T}}(x)$ is non-empty and for any $y_x \in \text{Proj}_{C, d_{L, T}}(x)$, by property (Proj), there exists a minimizing $\gamma_{x, y_x}(t)$ joining x and y_x . What remains to be proved is that there exists a choice of y_x such the third property holds.

Now let us consider the collection of subsets $L(X) \subset 2^X := \{\tilde{X} \subset X\}$ (partially ordered by set inclusions) such that there is a choice $\tilde{Y} : \tilde{X} \rightarrow C, x \mapsto \tilde{Y}(x)$ if $x_1, x_2 \in \tilde{X}$ and $0 \leq t \leq T$, $\gamma_{x_1, \tilde{Y}(x_1)}(t) = \gamma_{x_2, \tilde{Y}(x_2)}(t)$, then $\gamma_{x_1, \tilde{Y}(x_1)}(s) = \gamma_{x_2, \tilde{Y}(x_2)}(s)$. We want to show that $L(X) = 2^X$.

Now take any subset $\tilde{X} \in L(X)$. Consider a point $p \notin \tilde{X}$, let us obtain a curve $\gamma_{p, y_p}(t)$ satisfying Lemma 2.9, part 3. Now let $\tilde{t} = \inf\{t \leq T : \gamma_{p, y_p}(t) = \gamma_{x, \tilde{Y}(x)}(t) \text{ for some } x \in \tilde{X}\}$.

If $\tilde{t} = T$, then the image $\{\gamma_{x, \tilde{Y}(x)}(t)\}_{x \in \tilde{X}, 0 \leq t < T}$ and $\{\gamma_{p, y_p}(t)\}_{0 \leq t < T}$ are disjoint. Therefore $\tilde{Z} : \tilde{X} \cup \{p\} \rightarrow C$ as $\tilde{Z}(x) = \tilde{Y}(x)$ if $x \neq p$ and $\tilde{Z}(p) = y_p$. Then $\{\tilde{Z}\}$ and $\{\gamma_{x, \tilde{Y}(x)}(t)\}_{x \in \tilde{X}} \cup \{\tilde{\gamma}_p^n(t)\}$ satisfy the desired property.

If $\tilde{t} < T$ by definition there exists a point $a \in \tilde{X}$ such that $\gamma_{p, y_p}(\tilde{t}) = \gamma_{a, \tilde{Y}(a)}(\tilde{t})$. Now let us define

$$\tilde{\gamma}_p(t) = \begin{cases} \gamma_{p, y_p}(t) & \text{for } t < \tilde{t} \\ \gamma_{a, \tilde{Y}(a)}(t) & \text{for } t \geq \tilde{t} \end{cases}$$

Now write $\tilde{Z} : \tilde{X} \cup \{p\} \rightarrow C$ as $\tilde{Z}(x) = \tilde{Y}(x)$ if $x \neq p$ and $\tilde{Z}(p) = \tilde{Y}(a)$. Then notice the new family of curve $\{\gamma_{x, \tilde{Y}(x)}(t)\}_{x \in \tilde{X}} \cup \{\tilde{\gamma}_p(t)\}$ satisfies the third property. We now hope to argue

$$\int_{\tilde{t}}^T L(\gamma_{a, \tilde{Y}(a)}(s), \dot{\gamma}_{a, \tilde{Y}(a)}(s)) ds \leq \int_{\tilde{t}}^T L(\gamma_{p, y_p}(s), \dot{\gamma}_{p, y_p}(s)) ds.$$

Assume otherwise. Then there is a curve

$$\tilde{\gamma}_a(t) = \begin{cases} \gamma_{a, \tilde{Y}(a)}(t) & \text{for } t < \tilde{t} \\ \gamma_{p, y_p}(t) & \text{for } t \geq \tilde{t} \end{cases}$$

such that $\tilde{Y}(a) \in C$ and

$$\inf_{z \in C} d_{L, T}(a, z) = \int_0^T L(\gamma_{a, \tilde{Y}(a)}(s), \dot{\gamma}_{a, \tilde{Y}(a)}(s)) ds > \int_0^T L(\tilde{\gamma}_a(s), \dot{\tilde{\gamma}}_a(s)) ds \geq \inf_{z \in C} d_{L, T}(a, z)$$

and hence it yields a contradiction. Therefore we must have instead

$$\inf_{z \in C} d_{L, T}(p, z) \leq \int_{\tilde{t}}^T L(\tilde{\gamma}_p(s), \dot{\tilde{\gamma}}_p(s)) ds \leq \int_0^T L(\gamma_{p, y_p}(s), \dot{\gamma}_{p, y_p}(s)) ds = \inf_{z \in C} d_{L, T}(p, z).$$

Thus we have $\tilde{Z}(p) = \tilde{Y}(a) = \tilde{\gamma}_p(T) \in \text{Proj}_{C, d_{L, T}}(p)$ and

$$\tilde{\gamma}_p \in \text{argmin}_{\gamma \in \mathcal{A}} \left\{ \int_0^T L(\gamma(s), \dot{\gamma}(s)) ds : \dot{\gamma} = v; \gamma(0) = a, \gamma(T) = \tilde{Z}(p) \right\}$$

Either case, we have $\tilde{X} \cup \{p\} \in L(X)$ if $\tilde{X} \in L(X)$. Since $L(X) \subset 2^X$ is partially ordered by inclusion. By axiom of choice/Zorn's lemma, $L(X) = 2^X$. Therefore $X \in L(X)$. \square

Lemma 2.11. Assume L is lower semi-continuous and $C \subset X$ be a compact subset. Assume further that $\text{Proj}_{C,d_{L,T}}$ satisfies property (Proj). Assume further that fixing x , $v \mapsto L(x, v)$ be furthermore convex, and $G(\rho) = \iota_{\mathcal{P}(C)}$. Then

$$U(\rho, t) = \int_0^t \int_X \inf_{z \in C} d_{L,t}(z, x) \rho(x) dx ds$$

where Y is given by Lemma 2.10, and furthermore the optimizer of $\rho(x, t)$ in the problem $U(\rho, t)$ is given by $\rho(x, s) dx = \gamma(\cdot, s) \# [\rho(x) dx]$ where $\gamma(x, s) := \gamma_{x, Y(x)}(s)$ given by Lemma 2.10

Proof. Let us consider

$$\begin{aligned} U(\rho, t) &= \inf_{\rho(\cdot, s), v(\cdot, s)} \left\{ G(\rho(0)) + \int_0^t \int_X L(x, v(x, s)) \rho(x, s) dx ds : \partial_s \rho + \nabla \cdot (v \rho) = 0, \rho_t = \rho \right\} \\ &= \inf_{m(\cdot, s), v(\cdot, s)} \left\{ G(\rho(0)) + \int_0^t \int_X f(x, \rho(x, s), m(x, s)) dx ds : \partial_s \rho + \nabla \cdot m = 0, \rho_t = \rho \right\} \end{aligned}$$

where $f(x, \rho, m)$ is given by

$$f(x, \rho, m) = \sup_{(a, b)} \{ a\rho + bm - a^+ - H(x, b) \} = \begin{cases} L(x, \frac{m}{\rho})\rho & \text{if } \rho > 0, m > 0 \\ 0 & \text{if } \rho = 0, m = 0 \\ \infty & \text{otherwise} \end{cases}.$$

Note now that fixing x since the function $(a, b) \mapsto a^+ + H(x, b)$ is convex, then $(\rho, m) \mapsto f(x, \rho, m)$ is also convex, and hence $(\rho(x, s), m(x, s)) \mapsto \int_0^t \int_X L(x, v(x, s)) \rho(x, s) dx ds$ is now convex with lower bound. Moreover the constraint $\partial_t \rho + \nabla \cdot m = 0$ is linear. Hence an minimizer exists if G is convex and there exists a bounded set κ such that $G|_{\mathcal{P}(X) \setminus \kappa} > \inf G > -\infty$.

Now consider $G(\rho) = \iota_{\mathcal{P}(C)}$ and we try to manipulate the expression. For $t > 0$ we have

$$\begin{aligned} U(\rho, t) &= \inf_{\rho(\cdot, s), v(\cdot, s)} \left\{ \int_0^t \int_X L(x, v(x, s)) \rho(x, s) dx ds : \partial_s \rho + \nabla \cdot (v \rho) = 0, \rho_t = \rho, \rho_0 \in \mathcal{P}(C) \right\} \\ &= \inf_{\gamma(\cdot, s)} \left\{ \int_0^t \int_X L(\gamma(x, s), \partial_s \gamma(x, s)) \rho(x) dx ds : \rho(\gamma(x, t)) \in \mathcal{P}(C) \right\} \end{aligned}$$

Now we have

$$\begin{aligned} U(\rho, t) &\geq \inf_{Z(\cdot)} \left\{ \int_0^t \int_X \inf_{\gamma_x \in C^{1,0}[0, t]} \{ L(\gamma_x(s), \partial_s \gamma_x(s)) : x_t = Z(x), x_0 = x \} \rho(x) dx ds : \rho(Z(x)) \in \mathcal{P}(C) \right\} \\ &\geq \inf_{Z(\cdot)} \left\{ \int_0^t \int_X d_{L,t}(Z(x), x) \rho(x) dx ds : \rho(Z(x)) \in \mathcal{P}(C) \right\} \\ &\geq \int_0^t \int_X \inf_{z \in C} d_{L,t}(z, x) \rho(x) dx ds. \end{aligned}$$

Now by Lemma 2.10, running t backward, there exists a $Y : X \rightarrow C$ such that $Y(x) \in \text{Proj}_{C,d_{L,t}}(x)$ and $\{\gamma_{x, Y(x)}(s)\}$ satisfies conclusions in Lemma 2.9, part 3 (except now that $\gamma_{x, Y(x)}(0) = Y(x)$ and $\gamma_{x, Y(x)}(t) = x$), and moreover if $\gamma_{x_1, Y(x_1)}(l) = \gamma_{x_2, Y(x_2)}(l)$ for some $x_1, x_2 \in X$ and $0 \leq l \leq t$, then $\gamma_{x_1, Y(x_1)}(l) = \gamma_{x_2, Y(x_2)}(l)$ for all $0 < s \leq l$. Now if $\mathcal{A} = C^{0,1}[0, t]$, $v(x, t) := \partial_t^- \gamma_{x, Y(x)}(t)$ is well-defined a.e. w.r.t. Lebesgue measure where ∂_t^- is the left derivative w.r.t. t and is absolute continuous. Otherwise if $\mathcal{A} = W^{1,q}[0, t]$, $\gamma_{x, Y(x)}(t)$ is absolute continuous

and $v(x, t) := \partial_t \gamma_{x, Y(x)}(t)$ is in $L^q[0, t]$. In both cases, by fundamental theorem of calculus for absolute continuous function, we have $\gamma(x, t) := x - \int_t^T v(x, s) ds$. Now we may see that for $\partial_s \rho + \nabla \cdot (v \rho) = 0$, $\rho_t = \rho$, then $\rho(x, s) dx = \gamma(\cdot, s) \# [\rho(x) dx]$ and $\rho(x, 0) \in \mathcal{P}(C)$. Moreover, by Lemma 2.10, we have

$$\int_0^t \int_X L(x, v(x, s)) \rho(x, s) dx ds = \int_0^t \int_X d_{L,t}(Y(x), x) \rho(x) dx ds = \int_0^t \int_X \inf_{z \in C} d_{L,t}(z, x) \rho(x) dx ds$$

Hence the minimum is attained. \square

Combining Lemma 2.9 and Lemma 2.11, we directly have the following two corollaries.

Corollary 2.12. *Assume L is lower semi-continuous and convex w.r.t. v and $C \subset X$ is compact. Moreover if there exists $R > 0$ such that $L(x, v) = \infty$ for $\|v\| > R$, then conclusion of Lemma 2.11 holds.*

Corollary 2.13. *Assume L is differentiable w.r.t. x , lower semi-continuous and convex w.r.t. v , and $C \subset X$ is compact. If $K\|v\|^q - K \leq L(x, v) \leq K\|v\|^q + K$ for some K where $1 < q < \infty$, then conclusion of Lemma 2.11 holds.*

Note that [10] considered a similar problem, but $\rho_t \in \mathcal{P}(X \setminus C)$ for all t instead, hence C represents an obstacle. The problem considered here is related but is different in a sense that in our problem, we hope to move the density ρ to C .

Note that also from [21], if we considered the case when $L(x, v) = h(v)$ where h is strongly convex, then for any μ_0 vanishing on any rectifiable set of dimension $d - 1$, we have that the existence and uniqueness of optimal transport map $Y = Id - \nabla h^{-1} \circ \nabla \phi$, where ϕ is the Kantorovich potential, which equals to $\phi(x) = \Phi(x, 1)$ in the maximizer of the dual formulation; and the product of support of the two measure satisfies $d_{L,1}$ -cyclical monotonicity. Our previous result can be generalized to that case, with where $v(x, t) = \nabla_p(H \nabla \phi(x, t))$.

2.3.2 Differential game on probability space with two players driving the mass

We give a brief derivation of the link of a special form of non-convex HJPDE and a (2 player) differential game problems on probability space. We follow discussions in [15, 14, 16, 17, 19], see also [18], about optimal control and also for differential games, and their links with HJ PDE.

We consider two compact convex sets $A \subset C^{0,1}(X; \mathbb{R}^d)$ and $B \subset C^{0,1}(X; \mathbb{R}^d)$ (e.g. bounded sets in $C^{0,1}(X; \mathbb{R}^d)$), in which control parameters lie. Then let us denote $\mathcal{A} = \{m : [t, T] \rightarrow A : m \text{ is measurable}\}$, which is referred to as the admissible set of Player I; and $\mathcal{B} = \{n : [t, T] \rightarrow B : n \text{ is measurable}\}$, which is referred to as the admissible set of Player II. We call the measurable functions $m : [t, T] \rightarrow A$ in the set \mathcal{A} and the function $n : [t, T] \rightarrow B$ in the set \mathcal{B} as *controls* performed by Players I and II respectively. Now consider the following system with $0 \leq t < T$, $\rho \in (\mathcal{P}_2(X), W_2)$,

$$\begin{cases} \frac{d}{ds} \rho(x, s) = \nabla \cdot (m(x, s) - n(x, s)) & t \leq s \leq T, \\ \rho(t) = \rho, \end{cases} \quad (2.6)$$

Now notice that $(m, n) \mapsto \nabla \cdot (m - n)$ is well-defined a.e. w.r.t. Lebegue measure and

$$\|\nabla \cdot (m - n)\|_{L^\infty} \leq C_1(\|m\|_{C^{0,1}} + \|n\|_{C^{0,1}}) \leq C_2,$$

for some constant C_1 , $m \in A$, $n \in B$. The unique solution to (2.6) is called the response of the controls $m(\cdot), n(\cdot)$. Then we introduce the *payoff* functional for a given pair of (t, ρ) :

$$P(m, n) := P_{t, \rho}(m(\cdot), n(\cdot)) := \int_t^T \int_X f(x, \rho(s), m(s), n(s)) dx ds + G(\rho(T)),$$

where $G : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies

$$\begin{cases} |G(\rho)| \leq C_2 \\ |G(\rho_1) - G(\rho_2)| \leq C_2 d_{W_2}(\rho_1, \rho_2), \end{cases}$$

and f satisfies

$$f(x, \rho, m, n) = f_2(x, \rho, m) - f_1(x, \rho, n)$$

where

$$f_i(x, \rho, m) = \sup_{(a,b)} \{a\rho + bm - a^+ - H_i(x, b)\} = \begin{cases} L_i(x, \frac{m}{\rho})\rho & \text{if } \rho > 0, m > 0 \\ 0 & \text{if } \rho = 0, m = 0 \\ \infty & \text{otherwise} \end{cases}.$$

which are both convex and satisfying

$$\begin{cases} |\int_X f(x, \rho, m, n) dx| \leq C_3 \\ |\int_X f(x, \rho_1, m, n) dx - \int_X f(x, \rho_2, m, n) dx| \leq C_3 d_{W_2}(\rho_1, \rho_2), \end{cases}$$

for some constants C_2, C_3 , $m \in A$, $n \in B$. In a differential game of the mass, the goal of player I is to maximize the functional P by choosing his control m whereas that of player II is to minimize P by choosing his control n .

We now define the lower and upper values of the differential game, based on the notation introduced above. We first define the two sets containing the respective controls of players I and II:

$$\begin{aligned} M(t) &:= \{m : [t, T] \rightarrow A : m \text{ is measurable.}\}, \\ N(t) &:= \{n : [t, T] \rightarrow B : n \text{ is measurable.}\}. \end{aligned}$$

Define a strategy for player I as the map

$$\alpha : N(t) \rightarrow M(t)$$

for each $t \leq s \leq T$ and $n, \hat{n} \in N(t)$ such that

$$n(\tau) = \hat{n}(\tau) \text{ for a.e. } t \leq \tau \leq s \quad \Rightarrow \quad \alpha[n](\tau) = \alpha[\hat{n}](\tau) \text{ for a.e. } t \leq \tau \leq s.$$

Therefore a strategy for player I $\alpha[n]$ is the control of player I given that of player II as n . Similarly, let us define a strategy for player II as

$$\beta : M(t) \rightarrow N(t)$$

for each $t \leq s \leq T$ and $m, \hat{m} \in M(t)$ such that

$$m(\tau) = \hat{m}(\tau) \text{ for a.e. } t \leq \tau \leq s \quad \Rightarrow \quad \beta[m](\tau) = \beta[\hat{m}](\tau) \text{ for a.e. } t \leq \tau \leq s.$$

Again a strategy for player II $\beta[m]$ is the control of player II given that of player I as m .

Let $\Gamma(t)$ denote the set of all strategies for I and $\Delta(t)$ for II beginning at time t . We define the upper and lower values of the differential game. The lower value $V(t, \rho)$ is defined as

$$V(t, \rho) := \inf_{\beta \in \Delta(t)} \sup_{m \in M(t)} P_{t, \rho}(m, \beta[m]).$$

where $x(\cdot)$ solves (2.6) for a given pair of (t, ρ) . Likewise, the upper value $U(t, \rho)$ is defined as

$$U(t, \rho) := \sup_{\alpha \in \Gamma(t)} \inf_{n \in N(t)} P_{t, \rho}(\alpha[n], n).$$

Similar to [19], one may derive from dynamic programming optimality conditions in the lower and upper values V and U are the viscosity solutions of a certain possibly nonconvex HJ PDE.

Hence we obtain Hamiltonians

$$\begin{aligned} \tilde{\mathcal{H}}^+(\rho, \Xi) &:= \max_{n \in B} \min_{m \in A} \left\{ \int_X -\langle m(x) - n(x), \nabla \Xi(x) \rangle - f(x, \rho(x), m(x), n(x)) dx \right\} \\ &= \int_X H_1(x, \nabla \Xi(x)) \rho(x) - H_2(x, \nabla \Xi(x)) \rho(x) dx \\ \tilde{\mathcal{H}}^-(\rho, \Xi) &:= \min_{m \in A} \max_{n \in B} \left\{ \int_X -\langle m(x) - n(x), \nabla \Xi(x) \rangle - f(x, \rho(x), m(x), n(x)) dx \right\} \\ &= \int_X H_1(x, \nabla \Xi(x)) \rho(x) - H_2(x, \nabla \Xi(x)) \rho(x) dx \end{aligned}$$

Hence it happens that $\tilde{\mathcal{H}}^+$ and $\tilde{\mathcal{H}}^-$ coincide. Writing

$$\tilde{\mathcal{H}}(\rho, \Xi) = \tilde{\mathcal{H}}_1(\rho, \Xi) - \tilde{\mathcal{H}}_2(\rho, \Xi)$$

where

$$\tilde{\mathcal{H}}_i(\rho, \Xi) = \int_X H_i(x, \nabla \Xi(x)) \rho(x) dx$$

Following a same argument [19], it is routine to show the dynamical programming principle and also the following well-known theorem:

Lemma 2.14. [19] *The function U and V are equal, and U is the viscosity solution to the HJ PDE :*

$$\begin{cases} \frac{\partial}{\partial t} U - \tilde{\mathcal{H}}(x, \delta_\rho U) = 0 & \text{on } \mathcal{P}(X) \times (-\infty, T), \\ U(T, \rho) = G(\rho) & \text{on } \mathcal{P}(X). \end{cases}$$

Now by a change of variable from t or to variable $T - t$, we have a variational representation of solution to when the Hamiltonian is in the specific form $\tilde{\mathcal{H}}(\rho, \Xi) = \tilde{\mathcal{H}}_1(\rho, \Xi) - \tilde{\mathcal{H}}_2(\rho, \Xi)$.

3 Rigorous treatment after discretization

In this section, we aim to give a rigorous treatment to the discrete spatial states in potential MFGs. Our spatial discretization follows the same work on optimal transport on graphs as in [20, 28] and our proof follows the ideas in [16]. This discretization follows also the works in [11, 12, 13]. Although we do not claim full generality of the following discrete framework with the aforementioned continuous framework, we expect a close connection under mild assumptions, with numerical evidence that closed form solutions in previous solutions are compared with numerically computed solution via the Hopf formula shown in Section 5.

For illustrative purposes, we focus on the following special form of the Lagrangian:

$$L(x, v) := \sum_{i=1}^n L(v_i),$$

where $L : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ is a proper function, define the Hamiltonian $H : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ as

$$H(p) := \sup_{v \in \mathbb{R}^1} \{pv - L(v)\}.$$

Consider (V, E) as a uniform toroidal graph with equal spacing $\Delta x = \frac{1}{M}$ in each dimension. Here, V is a vertex set with $|V| = (M+1)^d$ nodes, and each node, $i = (i_k)_{k=1}^d \in V$, $1 \leq k \leq d$, $0 \leq i_k \leq n$, represents a cube with length Δx :

$$C_i = \{(x_1, \dots, x_d) \in [0, 1]^d : |x_1 - i_1 \Delta x| \leq \Delta x/2, \dots, |x_d - i_d \Delta x| \leq \Delta x/2\}.$$

Here E is an edge set, where $i + \frac{e_v}{2} := \text{edge}(i, i + e_v)$, and e_v is a unit vector at v th column.

Define

$$\rho_i := \int_{C_i} \rho(x) dx \in [0, 1]$$

on each $i \in V$. Let the discrete flux function be $m := (m_{i+\frac{e_v}{2}})_{i+\frac{e_v}{2} \in E}$, where $m_{i+\frac{e_v}{2}}$ represents the discrete flux on the edge $i + \frac{e_v}{2}$, i.e.,

$$m_{i+\frac{e_v}{2}} \approx \int_{C_{i+\frac{e_v}{2}}(x)} m_v(x) dx,$$

where $m(x) = (m_v(x))_{v=1}^d$ is the flux function in continuous space.

Thus the discrete divergence operator is:

$$\text{div}(m)|_i = \frac{1}{\Delta x} \sum_{v=1}^d (m_{i+\frac{1}{2}e_v} - m_{i-\frac{1}{2}e_v}).$$

The discretized cost functional forms

$$\mathcal{L}(m, \rho) := \sum_{i \in V, i+\frac{e_v}{2} \in E} \tilde{L}\left(m_{i+\frac{1}{2}e_v}, \theta_{i+\frac{1}{2}e_v}\right)$$

where

$$\tilde{L}\left(m_{i+\frac{1}{2}e_v}, \theta_{i+\frac{1}{2}e_v}\right) := \begin{cases} L\left(\frac{m_{i+\frac{1}{2}e_v}}{\theta_{i+\frac{1}{2}e_v}}\right) \theta_{i+\frac{1}{2}e_v} & \text{if } \theta_{i+\frac{1}{2}e_v} > 0; \\ 0 & \text{if } \theta_{i+\frac{1}{2}e_v} = 0 \text{ and } m_{i+\frac{e_v}{2}} = 0; \\ +\infty & \text{Otherwise.} \end{cases}$$

and $\theta_{i+\frac{1}{2}e_v} := \frac{1}{2}(\rho_i + \rho_{i+e_v})$ is the discrete probability on the edge $i + \frac{e_v}{2} \in E$.

We further introduce a time discretization. The time interval $[0, 1]$ is divided into N intervals with endpoints $t_n = n\Delta t$, $\Delta t = \frac{1}{N}$, $l = 0, 1, \dots, N$. Combining the above spatial discretization and a forward finite difference scheme on the time variable, we arrive at the following discrete optimal control problem:

$$\tilde{U}(t, \rho) := \inf_{m, \rho} \left\{ \sum_{n=1}^N \Delta t \mathcal{L}(m^n, \rho^n) - \sum_{n=1}^N \Delta t F(\rho^n) + G(\rho^0) \right\} \quad (3.1a)$$

where the minimizer is taken among $\{\rho\}_i^n$, $\{m\}_{i+\frac{e_v}{2}}^n$, such that for $n = 0, \dots, N-1$

$$\begin{cases} \rho_i^{n+1} - \rho_i^n + \Delta t \cdot \text{div}(m^{n+1})|_i = 0, \\ \rho_i^N = \rho_i. \end{cases} \quad (3.1b)$$

We next derive the discrete Hopf formula for minimization (3.1). Denote $\rho_i^N := \rho_i$ and

$$(\mathbf{m}, \rho, \Phi) := \left(\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1}, \{\rho_i^n\}_{n=0}^{N-1}, \{\Phi_i^n\}_{n=1}^N \right) \in \mathbb{R}^{|E|N} \times [0, 1]^{|V|N} \times \mathbb{R}^{|V|N}.$$

Hence by an application of Lagrange multiplier at (3.1), then we have

$$\tilde{U}(t, \{\rho_i\}) = \inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in \mathbb{R}^{|E|N}, \{\rho_i^n\}_{n=0}^{N-1} \in [0, 1]^{|V|N}} \sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \mathcal{F}(\mathbf{m}, \rho, \Phi)$$

where

$$\begin{aligned} \mathcal{F}(\mathbf{m}, \rho, \Phi) &:= \sum_{n=0}^{N-1} \sum_{i \in V, i+\frac{e_v}{2} \in E} \Delta t \mathcal{L} \left(m_{i+\frac{1}{2}e_v}^{n+1}, \theta_{i+\frac{1}{2}e_v}^{n+1} \right) - \sum_{n=0}^{N-1} \Delta t F(\{\rho\}_i^{n+1}) + G(\{\rho\}_i^0) \\ &\quad + \sum_{n=0}^{N-1} \sum_{i \in V} \Phi_i^{n+1} (\rho_i^{n+1} - \rho_i^n + \Delta t \cdot \operatorname{div}(m^{n+1})|_i). \end{aligned} \quad (3.3)$$

For a rigorous treatment, we assume:

(A1) The Lagrangian $L : \mathbb{R} \rightarrow \mathbb{R}$ is a proper, lower semi-continuous, convex functional.

(A2) The Lagrangian $L : \mathbb{R} \rightarrow \mathbb{R}$ has the following properties:

- for any fixed $x \neq 0$, $\lim_{y \rightarrow 0^+} L\left(\frac{x}{y}\right) y = \infty$;
- for any fixed y , the function $(x, y) \rightarrow L\left(\frac{x}{y}\right) y$ is equi-coercive (under parameter y) w.r.t x in the following sense: for all $N > 0$, there exists K (independent of y)

$$\left| L\left(\frac{x}{y}\right) y \right| \geq K$$

whenever $|x| \geq N$.

(A3) The functional $F : [0, 1]^{|V|} \rightarrow \mathbb{R}$ is a proper, upper semi-continuous, strongly concave functional.

(A4) The functional $G : [0, 1]^{|V|} \rightarrow \mathbb{R}$ is proper, lower semi-continuous, and convex in $\{\rho_i\}_{i=1}^{|V|}$.

(A5) $H : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ is in C^2 , and $F : [0, 1]^{|V|} \rightarrow \mathbb{R}$ and $G : [0, 1]^{|V|} \rightarrow \mathbb{R}$ are in $C^2([0, 1]^{|V|})$.

(A6) Denote the Legendre transform of the function $F : [0, 1]^{|V|} \rightarrow \mathbb{R}$ by F^* . Suppose F^* is coercive, i.e.

$$|F^*(x)| \rightarrow \infty$$

as $x \rightarrow \infty$.

(A7) The derivative of the function $F : [0, 1]^{|V|} \rightarrow \mathbb{R}^1$ satisfies, for any $\bar{\rho} \in \{0, 1\}^{|V|}$

$$|\partial_\rho F(\{\rho_i\})|_2^2 \rightarrow \infty$$

whenever $\{\rho_i\} \rightarrow \bar{\rho}$.

Under the above assumptions, we introduce the discrete Hopf formula by the following theorem

Theorem 3.1. *If (A1)-(A7) holds, then the value function $\tilde{U}(t, \{\rho_i\})$ in (3.1) equals*

$$\tilde{U}(t, \{\rho_i\}) = \sup_{\{\Phi_i\} \in \mathbb{R}^{|V|}} \left\{ \sum_{i \in V} \Phi_i^N \rho_i - \sum_{n=1}^N \Delta t (F(\{\rho\}_i^n) - \sum_i [\nabla_\rho F(\{\rho\}_i^n)]_i \rho_i^n) - G^*(\{\Phi\}_i^0) : \right. \\ \left. \begin{aligned} &\rho_i^n - \rho_i^{n-1} + \Delta t \sum_{v=1}^d D_p H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) \theta_{i+\frac{1}{2}e_v}^n = 0 \\ &\Phi_i^{n+1} - \Phi_i^n + \frac{\Delta t}{4} \sum_{v=1}^d H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) + \Delta t [\nabla_\rho F(\{\rho\}_i^n)]_i = 0 \\ &\rho_i^N = \rho_i, \Phi_i^N = \Phi_i \end{aligned} \right\} \quad (3.4)$$

Remark 3.2. *We remark that if $(\{\rho_i^n\}, \{\Phi_i^n\})$ are computed according to the constraints given in (3.4) for all $n = 0, \dots, N-1$, then for each n , the numerical Hamiltonian*

$$\mathbb{H}(\rho^n, \Phi^n) := \sum_{i \in V, i+\frac{e_v}{2} \in E} H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) \theta_{i+\frac{1}{2}e_v}^n + F(\{\rho\}_i^n) \quad (3.5)$$

is conserved, where we write $\rho^n := \{\rho_i^n\}$ and $\Phi^n := \{\Phi_i^n\}$.

Remark 3.3. *If $F(\rho) = 0$, (3.4) is an unstable scheme for initial value Hamilton-Jacobi equations. In computations, we handle it using a monotone scheme; see section 4 (Remark 4.1) below.*

Remark 3.4. *We note in numerical examples in Section 5 that our formula appears to be valid beyond the assumptions (A1)-(A7), e.g., in the case when H is a nonsmooth, nonconvex Hamiltonian. The continuous analog of (3.4) is discussed and proposed in Section 2.2. The minimal assumptions of validity for (3.4) to hold may be an interesting direction to explore, and some possibilities are discussed in, e.g., [16, 27, 33].*

We prove Theorem 3.1 by showing the following three lemmas.

Lemma 3.5. *Assume (A1). Then, the functional $\mathcal{L}(m, \rho)$ is convex.*

Proof. We shall show that \mathcal{L} is convex. Since

$$\mathcal{L}(m, \rho) := \sum_{i \in V, i+\frac{e_v}{2} \in E} \tilde{L} \left(m_{i+\frac{1}{2}e_v}, \theta_{i+\frac{1}{2}e_v} \right)$$

We only need to show that \tilde{L} is convex. In other words, for $y > 0$, $L(\frac{x}{y})y$ is convex for (x, y) . In fact, that is true since

$$\text{Hess} \left(L \left(\frac{x}{y} \right) y \right) = L'' \left(\frac{x}{y} \right) \begin{pmatrix} \frac{1}{y} & -\frac{x}{y^2} \\ -\frac{x}{y^2} & \frac{x^2}{y^3} \end{pmatrix} \succeq 0.$$

□

We now proceed as in [16] and obtain the following lemma. This lemma is similar to the primal-dual formulation in [1, 2, 3]. The argument to obtain this lemma is also similar to [16], but we provide it here for the sake of completeness.

Lemma 3.6. *Write*

$$(\mathbf{m}, \rho, \Phi) := \left(\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1}, \{\rho_i^n\}_{n=0}^{N-1}, \{\Phi_i^n\}_{n=1}^N \right) \in \mathbb{R}^{|E|N} \times [0, 1]^{|V|N} \times \mathbb{R}^{|V|N}.$$

and

$$(\tilde{\rho}, \Phi) := \left(\{\rho_i^n\}_{n=0}^{N-1}, \{\Phi_i^n\}_{n=1}^N \right) \in [0, 1]^{|V|N} \times \mathbb{R}^{|V|(N-1)}.$$

and let $\mathcal{F}(\mathbf{m}, \rho, \Phi)$ given in (3.3), and

$$\begin{aligned} \tilde{\mathcal{F}}(\tilde{\rho}, \Phi) &:= - \sum_{n=1}^N \sum_{i \in V, i+\frac{e_v}{2} \in E} H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) \theta_{i+\frac{1}{2}e_v}^n \Delta t - \sum_{n=1}^N \Delta t F(\{\rho_i^n\}) - G^*(\{\Phi_i^n\}) \\ &\quad + \sum_{n=1}^{N-1} \sum_{i \in V} (\Phi_i^n - \Phi_i^{n+1}) \rho_i^n + \sum_{i \in V} \Phi_i^N \rho_i \end{aligned}$$

If (A1), (A2), (A3), (A4) are satisfied, then

$$\begin{aligned} &\inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in \mathbb{R}^{|E|N}, \{\rho_i^n\}_{n=0}^{N-1} \in [0, 1]^{|V|N}} \sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \mathcal{F}(\mathbf{m}, \rho, \Phi) \\ &= \sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \inf_{\{\rho_i^n\}_{n=1}^{N-1} \in [0, 1]^{|V|(N-1)}} \tilde{\mathcal{F}}(\tilde{\rho}, \Phi). \end{aligned} \tag{3.6}$$

Proof. In fact, from the equi-coercivity (A2), we find that there exists a closed and bounded interval $C \subset \mathbb{R}$ s.t.

$$\begin{aligned} &\inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in \mathbb{R}^{|E|N}, \{\rho_i^n\}_{n=0}^{N-1} \in [0, 1]^{|V|N}} \sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \mathcal{F}(\mathbf{m}, \rho, \Phi) \\ &= \inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in C^{|E|N}, \{\rho_i^n\}_{n=0}^{N-1} \in [0, 1]^{|V|N}} \sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \mathcal{F}(\mathbf{m}, \rho, \Phi) \end{aligned}$$

Now that $\mathcal{F}(\mathbf{m}, \rho, \Phi)$ is lower-semicontinuous and quasi-convex w.r.t. (\mathbf{m}, ρ) (from (A1), (A3) and (A4)) and upper-semicontinuous and quasi-concave w.r.t. Φ (from linearity), as well as $C^{|E|N} \times [0, 1]^{|V|N}$, we have by an application of Sion's minimax theorem [26, 31] that

$$\begin{aligned} &\inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in C^{|E|N}, \{\rho_i^n\}_{n=0}^{N-1} \in [0, 1]^{|V|N}} \sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \mathcal{F}(\mathbf{m}, \rho, \Phi) \\ &= \sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in C^{|E|N}, \{\rho_i^n\}_{n=0}^{N-1} \in [0, 1]^{|V|N}} \mathcal{F}(\mathbf{m}, \rho, \Phi) \\ &= \sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in \mathbb{R}^{|E|N}, \{\rho_i^n\}_{n=0}^{N-1} \in [0, 1]^{|V|N}} \mathcal{F}(\mathbf{m}, \rho, \Phi) \end{aligned}$$

where the last equality is again obtained by equi-coercivity in (A2).

Now let us fix $(\tilde{\rho}, \Phi) = (\{\rho_i^n\}_{n=1}^{N-1}, \{\Phi_i^n\}_{n=1}^N)$, and consider the optimization

$$\inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in \mathbb{R}^{|E|N}, \{\rho_i^0\} \in [0, 1]^{|V|}} \mathcal{F}(\mathbf{m}, \rho, \Phi).$$

We next derive its duality formula. Following the summation by part, then

$$\begin{aligned}
& \inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in \mathbb{R}^{|E|^N}, \{\rho_i^0\} \in [0,1]^{|V|}} \mathcal{F}(\mathbf{m}, \rho, \Phi) \\
&= \inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in \mathbb{R}^{|E|^N}, \{\rho_i^0\} \in [0,1]^{|V|}} \left\{ \sum_{n=0}^{N-1} \sum_{i \in V, i+\frac{e_v}{2} \in E} \Delta t \mathcal{L} \left(m_{i+\frac{1}{2}e_v}^{n+1}, \theta_{i+\frac{1}{2}e_v}^{n+1} \right) - \sum_{n=0}^{N-1} \Delta t F(\{\rho\}_i^{n+1}) + G(\{\rho\}_i^0) \right. \\
&\quad \left. + \sum_{n=0}^{N-1} \sum_{i \in V} \Phi_i^{n+1} (\rho_i^{n+1} - \rho_i^n + \Delta t \cdot \text{div}(m^{n+1})|_i) \right\} \\
&= \inf_{\{\rho_i^0\} \in [0,1]^{|V|}} \left\{ \sum_{n=0}^{N-1} \sum_{i \in V, i+\frac{e_v}{2} \in E} \Delta t \inf_{m_{i+\frac{1}{2}e_v}} \left\{ L \left(\frac{m_{i+\frac{1}{2}e_v}^{n+1}}{\theta_{i+\frac{1}{2}e_v}^{n+1}} \right) \theta_{i+\frac{1}{2}e_v}^{n+1} + \frac{1}{\Delta x} (\Phi_i^{n+1} - \Phi_{i+e_v}^{n+1}) m_{i+\frac{1}{2}e_v}^{n+1} \right\} \right. \\
&\quad \left. - \sum_{n=0}^{N-1} \Delta t F(\{\rho\}_i^{n+1}) + G(\{\rho\}_i^0) + \sum_{n=0}^{N-1} \sum_{i \in V} \Phi_i^{n+1} (\rho_i^{n+1} - \rho_i^n) \right\}
\end{aligned}$$

where the last equality is from the spatial integration by parts for $\sum_{n=0}^{N-1} \sum_{i \in V} \Phi_i^{n+1} \text{div}(m^{n+1})|_i$. From the Legendre transform

$$H(p) = \sup_{v \in \mathbb{R}^1} pv - L(v)$$

with $p = \frac{1}{\Delta x} (\Phi_{i+e_v}^n - \Phi_i^n)$ and $v = \frac{m_{i+\frac{1}{2}e_v}^{n+1}}{\theta_{i+\frac{1}{2}e_v}^{n+1}}$, we have

$$\begin{aligned}
& \inf_{\{m_{i+\frac{1}{2}e_v}^n\}_{n=0}^{N-1} \in \mathbb{R}^{|E|^N}, \{\rho_i^0\} \in [0,1]^{|V|}} \mathcal{F}(\mathbf{m}, \rho, \Phi) \\
&= \inf_{\{\rho_i^0\} \in [0,1]^{|V|}} \left\{ - \sum_{n=1}^N \sum_{i \in V, i+\frac{e_v}{2} \in E} H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) \theta_{i+\frac{1}{2}e_v}^n \Delta t - \sum_{n=1}^N F(\{\rho\}_i^{n+1}) \Delta t \right. \\
&\quad \left. + G(\{\rho\}_i^0) + \sum_{n=0}^{N-1} \sum_{i \in V} \Phi_i^{n+1} (\rho_i^{n+1} - \rho_i^n) \right\} \\
&= \inf_{\{\rho_i^0\} \in [0,1]^{|V|}} \left\{ - \sum_{n=1}^N \sum_{i \in V, i+\frac{e_v}{2} \in E} H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) \theta_{i+\frac{1}{2}e_v}^n \Delta t - \sum_{n=1}^N F(\{\rho\}_i^{n+1}) \Delta t + G(\{\rho\}_i^0) \right. \\
&\quad \left. + \sum_{n=1}^{N-1} \sum_{i \in V} (\Phi_i^n - \Phi_i^{n+1}) \rho_i^n + \sum_{i \in V} \Phi_i^N \rho_i - \sum_{i \in V} \Phi_i^0 \rho_i^0 \right\} \\
&= \tilde{\mathcal{F}}(\tilde{\rho}, \Phi),
\end{aligned}$$

where the last line follows from the definition of Legendre transform for $\{\rho^0\}_i^V$. \square

Remark 3.7. We remark that [1, 2, 3] utilized a similar version of the above lemma and computed the saddle point using Newton's method. However, since we aim to reduce the number of dimensions in our numerical scheme and also aim to handle nonsmooth cases, we do not stop at this formulation.

Lemma 3.8. *If (A1), (A3), (A4), (A5), (A6), (A7) are satisfied, then*

$$\begin{aligned} & \sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \inf_{\{\rho_i^n\}_{n=1}^{N-1} \in [0,1]^{|V|(N-1)}} \tilde{\mathcal{F}}(\tilde{\rho}, \Phi) \\ &= \sup_{\{\Phi_i\} \in \mathbb{R}^{|V|}} \left\{ \sum_{i \in V} \Phi_i^N \rho_i - \sum_{n=1}^N \Delta t \left(F(\{\rho\}_i^n) - \sum_i [\nabla_\rho F(\{\rho\}_i^n)]_i \rho_i^n \right) - G^*(\{\Phi\}_i^0) : \right. \\ & \quad \left. \begin{aligned} & \rho_i^n - \rho_i^{n-1} + \Delta t \sum_{v=1}^d D_p H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) \theta_{i+\frac{1}{2}e_v}^n = 0 \\ & \Phi_i^{n+1} - \Phi_i^n + \frac{\Delta t}{4} \sum_{v=1}^d H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) + \Delta t [\nabla_\rho F(\{\rho\}_i^n)]_i = 0 \\ & \rho_i^N = \rho_i, \Phi_i^N = \Phi_i \end{aligned} \right\}. \end{aligned}$$

Proof. Writing $(\tilde{\rho}, \tilde{\Phi}) = (\{\Phi_i^n\}_{n=1}^{N-1}, \{\rho_i^n\}_{n=1}^{N-1})$, then we have

$$\sup_{\{\Phi_i^n\}_{n=1}^N \in \mathbb{R}^{|V|N}} \inf_{\{\rho_i^n\}_{n=1}^{N-1} \in [0,1]^{|V|(N-1)}} \tilde{\mathcal{F}}(\tilde{\rho}, \Phi) = \sup_{\{\Phi_i^N\} \in \mathbb{R}^N} \sup_{\{\Phi_i^n\}_{n=1}^{N-1} \in \mathbb{R}^{|V|(N-1)}} \inf_{\{\rho_i^n\}_{n=1}^{N-1} \in [0,1]^{|V|(N-1)}} \tilde{\mathcal{F}}(\tilde{\rho}, \tilde{\Phi}, \Phi^N).$$

Given $(\tilde{\Phi}, \Phi_i^N)$, from (A3) and (A7), we have the resulting functional $\tilde{\mathcal{F}}(\tilde{\rho}, \tilde{\Phi}, \Phi^N)$ to be C^2 w.r.t. $\tilde{\rho}$. Together with compactness of $[0,1]^{|V|(N-1)}$, we therefore have that the infimum

$$\inf_{\{\rho_i^n\}_{n=1}^{N-1} \in [0,1]^{|V|(N-1)}} \tilde{\mathcal{F}}(\tilde{\rho}, \tilde{\Phi}, \Phi^N)$$

is attained, and the minimum is attained in the interior of the domain $[0,1]^{|V|(N-1)}$ thanks to (A7) and convexity assumption of F , that F is growing near the boundary. From the smoothness given by (A5), the minimum $\tilde{\rho}^*(\tilde{\Phi}, \Phi^N)$ shall satisfy the first order optimality condition

$$\frac{\Phi_i^{n+1} - \Phi_i^n}{\Delta t} + \frac{1}{4} \sum_{v=1}^d H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) + [\nabla_\rho F(\{\rho^*(\tilde{\Phi}, \Phi^N)\}_i^n)]_i = 0$$

and by smoothness given by (A5) that the above equation is in C^1 , and strong convexity that $\nabla_\rho^2 F$ is non-singular, then the implicit function theorem applies and $\tilde{\rho}^*(\tilde{\Phi}, \Phi^N)$ smoothly depends on $(\tilde{\Phi}, \Phi_i^N)$. Now let us fix Φ^N . From the definition, we have

$$\begin{aligned} & \inf_{\{\rho_i^n\}_{n=1}^{N-1} \in [0,1]^{|V|(N-1)}} \tilde{\mathcal{F}}(\tilde{\rho}, \tilde{\Phi}, \Phi^N) \\ &= - \sum_{n=1}^N \Delta t F^* \left(\left\{ \frac{\Phi_i^{n+1} - \Phi_i^n}{\Delta t} + \frac{1}{4} \sum_{v=1}^d H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) \right\}_i^n \right) - G^*(\{\Phi\}_i^0) + \sum_{i \in V} \Phi_i^N \rho_i. \end{aligned}$$

Now for any given $\{V_i^n\}_{n=1}^{N-1}$, by iteratively solving the difference equation, there exists $\{\Phi_i^n\}_{n=1}^{N-1}$ such that

$$\frac{\Phi_i^{n+1} - \Phi_i^n}{\Delta t} + \frac{1}{4} \sum_{v=1}^d H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) = V_i^{n-1}.$$

Therefore we have bijectivity between $\{V_i^n\}_{n=1}^{N-1}$ and $\{\Phi_i^n\}_{n=1}^{N-1}$, and therefore

$$\begin{aligned} & \sup_{\{\Phi_i^n\}_{n=1}^{N-1} \in \mathbb{R}^{|V|(N-1)}} \inf_{\{\rho_i^n\}_{n=1}^{N-1} \in [0,1]^{|V|(N-1)}} \tilde{\mathcal{F}}(\tilde{\rho}, \tilde{\Phi}, \Phi^N) \\ &= \sup_{\{V_i^n\}_{n=1}^{N-1} \in \mathbb{R}^{|V|(N-1)}} \left\{ - \sum_{n=0}^{N-1} \Delta t F^*(V_i^n) - G^*(\{\Phi\}_i^0) + \sum_{i \in V} \Phi_i^N \rho_i \right\}. \end{aligned} \quad (3.7)$$

Note that by (A3), strong convexity of F implies strongly smoothness (i.e. at least differentiability) and convexity of F^* . Note that by (A6), the above function is also coercive, and therefore the supremum in the above is attained, and hence there exists a maximum point $\{V_i^{*n}\}_{n=0}^{N-1}$ for the functional $-\sum_{n=0}^{N-1} F^*(V_i^n)$. Now go back to find $\{\Phi_i^{*n}\}_{n=1}^{N-1}$ such that

$$\frac{\Phi_i^{*n+1} - \Phi_i^{*n}}{\Delta t} + \frac{1}{4} \sum_{v=1}^d H \left(\frac{1}{\Delta x} (\Phi_i^{*n} - \Phi_{i+e_v}^{*n}) \right) = V_i^{*n-1}.$$

Now, fixing Φ^N and noticing $\{\Phi_i^{*n}\}_{n=1}^{N-1}$ is a maximum value of the function

$$\tilde{\Phi} = \{\Phi_i^n\}_{n=1}^{N-1} \mapsto \inf_{\{\rho_i^n\}_{n=1}^{N-1} \in [0,1]^{|V|(N-1)}} \tilde{\mathcal{F}}(\tilde{\rho}, \tilde{\Phi}, \Phi^N) = \tilde{\mathcal{F}}(\tilde{\rho}^*(\tilde{\Phi}, \Phi^N), \tilde{\Phi}, \Phi^N),$$

we see that the maximum is attained and by (A5), we have the smooth first order optimality condition for $\{\Phi_i^{*n}\}_{n=1}^{N-1}$ and it can be characterized by

$$\rho^*(\tilde{\Phi}, \Phi^N)_i^n - \rho^*(\tilde{\Phi}, \Phi^N)_i^{n-1} + \Delta t \sum_{v=1}^d D_p H \left(\frac{1}{\Delta x} (\Phi_i^{*n} - \Phi_{i+e_v}^{*n}) \right) \theta_{i+\frac{1}{2}e_v}^n(\tilde{\Phi}, \Phi^N) = 0.$$

Concluding the above argument, for each Φ^N , we have

$$\sup_{\{\Phi_i^n\}_{n=1}^{N-1} \in \mathbb{R}^{|V|(N-1)}} \inf_{\{\rho_i^n\}_{n=1}^{N-1} \in [0,1]^{|V|(N-1)}} \tilde{\mathcal{F}}(\tilde{\rho}, \tilde{\Phi}, \Phi^N) = \tilde{\mathcal{F}}(\tilde{\rho}^*(\Phi^N), \tilde{\Phi}^*(\Phi^N), \Phi^N),$$

which satisfies the following pair of equations

$$\begin{cases} \rho_i^n - \rho_i^{n-1} + \Delta t \sum_{v=1}^d D_p H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) \theta_{i+\frac{1}{2}e_v}^n = 0 \\ \Phi_i^{n+1} - \Phi_i^n + \frac{\Delta t}{4} \sum_{v=1}^d H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) + \Delta t [\nabla_p F(\{\rho\}_i^n)]_i = 0 \\ \rho_i^N = \rho_i, \Phi_i^N = \Phi_i \end{cases} \quad (3.8)$$

for all $n = 0, 1, \dots, N-1$. The conclusion of the lemma thus follows. \square

Remark 3.9. We note that taking supremum of (3.7) over Φ^N yields the discrete version of the Hopf formula given in (57)-(59) of [27]. However, we are not trying to get back to (57)-(59) in [27] since the formulation, though very elegant mathematically, contains too many variables for numerical optimization of a low memory requirement, and is thus not our first choice.

Combining all the three lemmas above, we have proved Theorem 3.1.

Remark 3.10. The full set of (A1) to (A7) are indeed a bit too strong; however, as is illustrated above, they make the argument as simple as possible. Some of the assumptions may be weakened to obtain weaker but similar forms, e.g. Lemma 3.6 has an assumption weaker than Lemma 3.8, in that in many cases smoothness may be dropped by replacing derivatives with a version of subdifferentials. We expect a proof of existence of saddle point shall be viable without much modification. However the full generality of our conjecture involves the hope of dropping the convexity assumption of H with respect p . Similar to [16], our hope is that the full generality involves only a pseudoconvexity assumption (Definition 3.11 in [16]) and a similar form of Conjecture 3.12 shall hold.

4 Algorithm

In next section, we apply the discrete Hopf formula (3.4) to design numerical methods for HJD. In this section, we compute the optimizer in the Hopf formula in (3.4). We shall perform the following multi-level stochastic gradient descent method.

We first consider a sequence of step-size $h_i = 2^{-i}$ where $i = 0, \dots, N$ and a nested sequence of finite dimensional subspaces $V_{h_0} \subset V_{h_2} \subset \dots \subset V_{h_{N-1}} \subset V_{h_N}$ of a function space over tensor product grid in d -dimensions over $[-1, 1]^d \subset X$. Now, we also define a family of restriction and extension operators:

$$R_{ij} : V_{h_i} \rightarrow V_{h_j} \quad \text{and} \quad E_{ji} : V_{h_j} \rightarrow V_{h_i}$$

Now, let us define the following approximation of the functional $G(\cdot)$ as

$$\begin{aligned} G_j : V_{h_j} &\rightarrow \mathbb{R} \\ G_j(\tilde{\Phi}) &:= \int_X R_{Nj}[\rho](x) \tilde{\Phi}(x) dx - G^*(\Phi(E_{jN}[\tilde{\Phi}], 0, \cdot)) \\ &\quad - \int_0^t \left(F(\rho(E_{jN}[\tilde{\Phi}], s, \cdot)) - \int_X [\nabla_\rho F(\rho(E_{jN}[\tilde{\Phi}], s, \cdot))] \rho(E_{jN}[\tilde{\Phi}], s, x) dx \right) dt \end{aligned} \quad (4.1)$$

where $(\rho(E_{jN}[\tilde{\Phi}], s, x), \Phi(E_{jN}[\tilde{\Phi}], s, x))$ numerically solves the following terminal value problem:

$$\begin{cases} \partial_s \rho(E_{jN}[\tilde{\Phi}], s, x) + \nabla \cdot (\rho(E_{jN}[\tilde{\Phi}], s, x) D_p H(x, \nabla \Phi(E_{jN}[\tilde{\Phi}], s, x))) = 0 \\ \partial_s \Phi(E_{jN}[\tilde{\Phi}], s, \cdot) + H(x, \nabla \Phi(E_{jN}[\tilde{\Phi}], s, x)) + \nabla_\rho F(\rho(E_{jN}[\tilde{\Phi}], s, \cdot)) = 0 \\ \rho(E_{jN}[\tilde{\Phi}], t, x) = \rho(x), \quad \Phi(E_{jN}[\tilde{\Phi}], t, x) = E_{jN}[\tilde{\Phi}](x). \end{cases}$$

The numerical method to compute this Cauchy problem will be discussed after we present the main algorithm in Remark 4.1.

With the above notation, we are ready to present our variant of stochastic gradient descent to optimize $G_N(\cdot)$. We utilize the following coordinate descent algorithm:

Algorithm 1. Take an initial guess $[\Phi_0]^1 \in V_{h_0}$.

For $i = 0, \dots, N$, do:

- Take an initial guess of the Lipschitz constant L_i , set $\text{count} := 0$, $v_i := 1/L_i$, and an error tolerance ε_i .

- For $k = 1, \dots, M$, do:

- 1: Randomly select $I_k \in 2^{\{1, \dots, 2^{d_i}\}}$,
- 2: Compute the following unit vector V_{I_k} where

$$[V_{I_k}]_l = \begin{cases} 1/\sqrt{|I_k|} & \text{if } l \in I_k, \\ 0 & \text{otherwise.} \end{cases}$$

- 3: Compute

$$\begin{cases} [\Phi_i]_l^{k+1} = [\Phi_i]_l^k - v_i [V_{I_k}]_l \partial_{V_{I_k}} G_i([\Phi_i]_l^k) & \text{if } l \in I \\ [\Phi_i]_l^{k+1} = [\Phi_i]_l^{k+1} & \text{otherwise.} \end{cases}$$

4: If $Er_{i,k+1} \left(\{[\Phi_i]_l^r\}_{r=1}^{k+1} \right) \geq \varepsilon_i$, then set $\text{count} := 0$. If $k = M$, then reset $k := 0$ and set $v_i := v_i/2$, (i.e. let $L_i := 2L_i$.)

5: If $Er_{i,k+1} \left(\{[\Phi_i]_l^r\}_{r=1}^{k+1} \right) < \varepsilon_i$, set $\text{count} := \text{count} + 1$.

5: If $\text{count} = 2^{di}$, define $[\Phi_i]_{\text{final}} := [\Phi_i]_I^{k+1}$, stop.

- If $i < N$, set $[\Phi_{i+1}]^1 = E_{i(i+1)}([\Phi_i]_{\text{final}})$.

Output $[\Phi_N]_{\text{final}}$.

where we now denote, for a given history of iteration $\{[\Phi_i]_I^r\}_{r=1}^k$, the consecutive ergodic error as the following

$$\text{Er}_{i,k} \left(\{[\Phi_i]_I^r\}_{r=1}^k \right) := \left| \frac{\sum_{r=1}^k G_i([\Phi_i]_I^r)}{k} - \frac{\sum_{r=1}^{k-1} G_i([\Phi_i]_I^r)}{k-1} \right| \quad (4.2)$$

The choice of such is in response to the fact that random stochastic gradient descent has an averaged (ergodic) convergence rate of $O(1/k)$. We would also like to mention that the above choice of I_k and v_{I_k} give us a random coordinate descent along a random direction of the form $v_{I_k} = (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)/|I_k|$. This choice of descent direction is to enhance speed of convergence via an average effect of grid points that are far away from each other to allow information passing in a greater speed. Note that the size of the above random choice of I_k does not depend on k . The coefficients, e.g. L_i , M and N , are suitably chosen (and is currently only done later in examples.)

Remark 4.1. For computation of $(\rho(E_{jN}[\tilde{\Phi}], s, x), \Phi(E_{jN}[\tilde{\Phi}], s, x))$ with numerical PDE techniques, we notice that the primal-dual system in (3.4), i.e. the conservation law and the HJD, may not provide a stable PDE algorithm. One way to address this pathology is to modify the numerical Hamiltonian that we have implicitly chosen when we derive our algorithm. We notice indeed that the system (3.4) is a symplectic scheme that conserves the following numerical Hamiltonian (writing $\rho^n := \{\rho_i^n\}$ and $\Phi^n := \{\Phi_i^n\}$):

$$\mathbb{H}(\rho^n, \Phi^n) = \sum_{i+\frac{e_v}{2} \in E} H \left(\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right) \theta_{i+\frac{1}{2}e_v}^n + F(\{\rho_i^n\}).$$

On the other hand, we notice that the choice of such Hamiltonian is not unique: we can choose another numerical Hamiltonian that corresponds to an upwind (monotone) scheme for the primal system and monotone Hamiltonian for the dual system as follows (see also [1, 2, 3]):

$$\mathbb{H}(\rho^n, \Phi^n) := \sum_i \sum_{v=1}^d H \left(\left[\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right]^+ \right) \rho_i^n + F(\{\rho_i^n\}),$$

where $[\cdot]^+ := \max(0, \cdot)$ and $[\cdot]^- := \min(0, \cdot)$. With this, the primal dual system will instead be as follows:

$$\begin{aligned} \frac{\rho_i^n - \rho_i^{n-1}}{\Delta t} + \sum_{v=1}^d D_p H \left(\left[\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right]^+ \right) \rho_i^n + \sum_{v=1}^d D_p H \left(\left[\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right]^- \right) \rho_{i+e_v}^n &= 0 \\ \frac{\Phi_i^{n+1} - \Phi_i^n}{\Delta t} + \frac{1}{2} \sum_{v=1}^d H \left(\left[\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right]^+ \right) + [\nabla_\rho F(\{\rho_i^n\})]_i &= 0. \end{aligned}$$

To further enhance stability, we can add, given a regularization parameter β , a Lax-Friedrichs scheme numerical diffusion term:

$$\begin{aligned} & \frac{\rho_i^n - \rho_i^{n-1}}{\Delta t} + \left\{ \sum_{v=1}^d D_p H \left(\left[\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right]^+ \right) \rho_i^n \right. \\ & \quad \left. + \sum_{v=1}^d D_p H \left(\left[\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right]^- \right) \rho_{i+e_v}^n \right\} + \frac{\beta \Delta x}{2(\Delta x)^2} \sum_{v=1}^d (\rho_i^n - \rho_{i+e_v}^n) = 0 \\ & \frac{\Phi_i^{n+1} - \Phi_i^n}{\Delta t} + \frac{1}{2} \sum_{v=1}^d H \left(\left[\frac{1}{\Delta x} (\Phi_i^n - \Phi_{i+e_v}^n) \right]^+ \right) + [\nabla_\rho F(\{\rho\}_i^n)]_i + \frac{\beta \Delta x}{2(\Delta x)^2} \sum_{v=1}^d (\Phi_i^n - \Phi_{i+e_v}^n) = 0. \end{aligned}$$

This adds a magnitude of $\beta \Delta x$ numerical diffusion in the system. We notice in our numerical examples that stability improves after imposing $\beta > 0$ and considering an upwind monotone scheme.

5 Numerical results

In this section, we present numerical results for solving HJD by Algorithm 1. We tested several cases with the different Hamiltonians, including the convex

$$H_1(x, p) = \frac{1}{2}(|p_1|^2 + |p_2|^2),$$

the non-convex

$$H_2(x, p) = \frac{1}{2}(|p_1|^2 - |p_2|^2),$$

and the convex 1-homogeneous Hamiltonian

$$H_3(x, p) = |p_1| + |p_2|.$$

For a given center x_0 and radius R , we consider

$$G(\rho) = \inf_{\tilde{\rho} \in \mathcal{P}(X)} \left\{ \iota_{\mathcal{P}(B_R(x_0))}(\tilde{\rho}) + \frac{1}{2\eta} \int_X [\tilde{\rho} - \rho(x)]^2 dx \right\},$$

where η is a regularization parameter, and we recall that, for a given convex subset $\mathcal{B} \subset \mathcal{P}(X)$, the indicator function $\iota_{\mathcal{B}}(\rho) = 0$ if $\rho \in \mathcal{B}$ and $\iota_{\mathcal{B}}(\rho) = \infty$ otherwise. A direct computation shows

$$\iota_{\mathcal{P}(B_R(x_0))}^*(\Phi) = \sup_{\rho \in \mathcal{P}(B_R(x_0))} \int_X \rho(x) \Phi(x) dx = \sup_{x \in B_R(x_0)} \Phi(x).$$

With the correspondence of summation and infimum convolution via Legendre transform, we arrive at

$$G^*(\Phi) = \sup_{x \in B_R(x_0)} \Phi(x) + \frac{\zeta}{2} \int_X [\Phi(x)]^2 dx.$$

In numerical examples, we set $\zeta = 10^{-3}$. This helps us compute a regularized projection of a given ρ to the set of all the measures supported at an unit ball. For simplicity, we set $F(\rho) = 0$ in all our examples.

We utilize Algorithm 1 for numerical computations. The number of levels $N = 5$ is always chosen. The Lipschitz constant is always chosen as $L_i = L = 2$. For numerical approximation of PDE, we choose the upwind numerical Hamiltonian, together with an addition of Lax-Friedrichs numerical diffusion where $\beta = 2$ is chosen. The discretization parameters are chosen as $\Delta x = 0.08$ and $\Delta t = 0.008$. In all experiments, we consider $X = \mathbb{T}^2$.

Example 5.1. *In this example, we consider the Hamiltonian H_1 and the input distribution $\rho(x)$ as follows:*

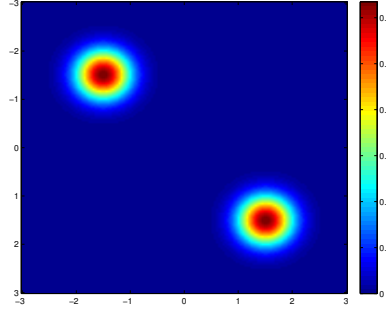


Figure 1: The distribution of the input ρ .

We choose the center and radius (x_0, R) which helps to define $G(\rho)$ as $x_0 = (0, 0), R = 1$.

In order to provide illustration for the multi-level effect using Algorithm 1 to compute (3.4), we have plot an optimizer obtained for (4.1) in each level as in Figure 7, for $i = 1, \dots, 4$ in one trial of randomization. As one may see, the minimizers are successively refined. The randomness and asymmetry comes from the random nature of stochastic gradient descent.

Figure 3 gives the optimizer $\tilde{\Phi}$ (left) in (3.4) in the finest level and its gradient $\nabla_x \tilde{\Phi}$ (right) computed using Algorithm 1 when $t = 1$ in the Hamiltonian. The gradient $\nabla_x \tilde{\Phi}$ generates the final kick of the drift for the masses to be flown accordingly.

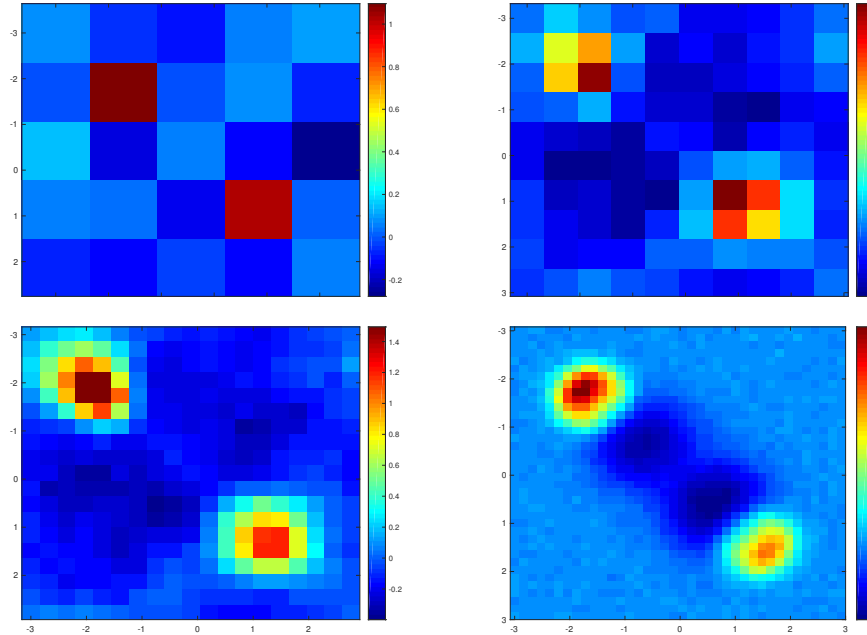


Figure 2: Left: optimizer $\tilde{\Phi}$ in $G_i([\Phi_i]_l^k)$ in (4.1) on coarse grids, $i = 1, \dots, 4$.

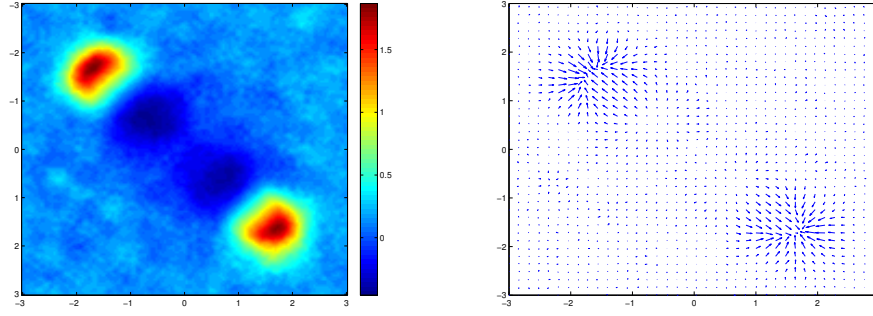


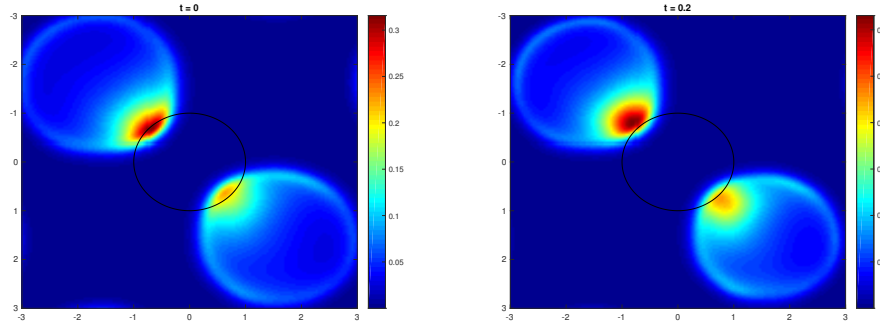
Figure 3: Left: optimizer $\tilde{\Phi}$ in $G_i([\Phi_i]_I^k)$ in (4.1), $i = 5$, i.e. the finest approximation of $U(t, \rho)$ where $t = 1$ in (2.3), right: vector field $\nabla_x \tilde{\Phi}$.

To illustrate the multi-level effect using Algorithm 1, Table 1 shows also, for each i, k , the respective number of epochs is defined as $\text{ceil}(k/\text{DoF})$, the minimum value function $G_i([\Phi_i]_I^k)$ obtained, where k is given by $[\Phi_i]_{\text{final}} = [\Phi_i]_I^{k+1}$ and G_i is given as (4.1), and the consecutive ergodic error $Er_{i,k}$ defined as in (4.2). This illustrates the successive refinement nature of the method.

Level i	Degree of Freedom (DoF)	Number of Epoch	Value function $G_i([\Phi_i]_I^k)$	Error $Er_{i,k}$
1	25	4	0.5029	7.6999×10^{-3}
2	100	10	0.7729	5.2091×10^{-4}
3	361	13	0.9741	3.0454×10^{-5}
4	1444	99	1.0156	7.3882×10^{-7}
5	5625	125	1.0506	5.1721×10^{-7}

Table 1: Number of epochs, minimum value function $G_i([\Phi_i]_I^k)$ obtained and consecutive ergodic error $Er_{i,k}$ defined as in (4.2).

In Figure 4, we plot the distributions $\rho(t, x)$ for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. It describes the transportation of the masses according to the flow generated by the gradient of $\Phi(t, x)$ at different times t .



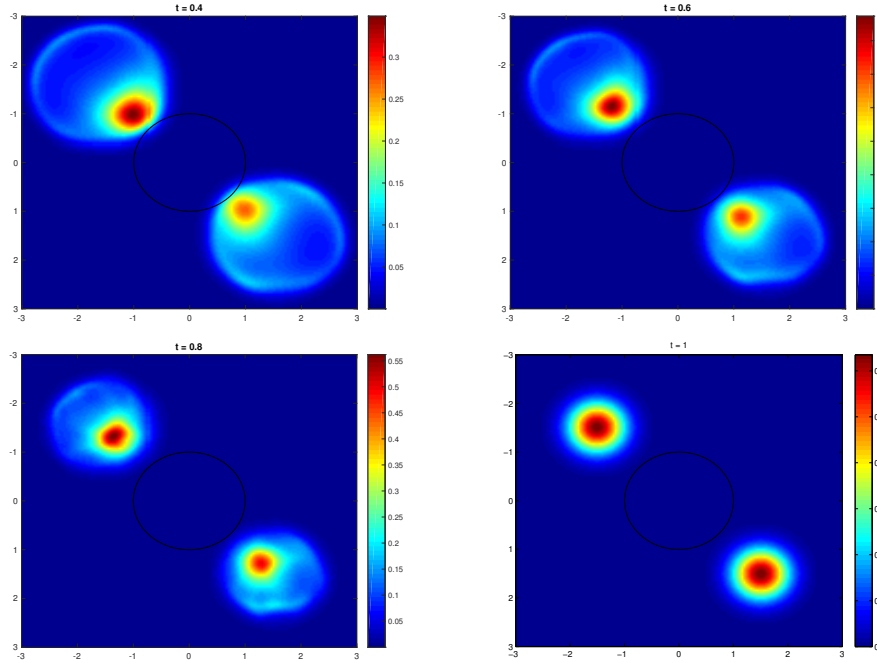
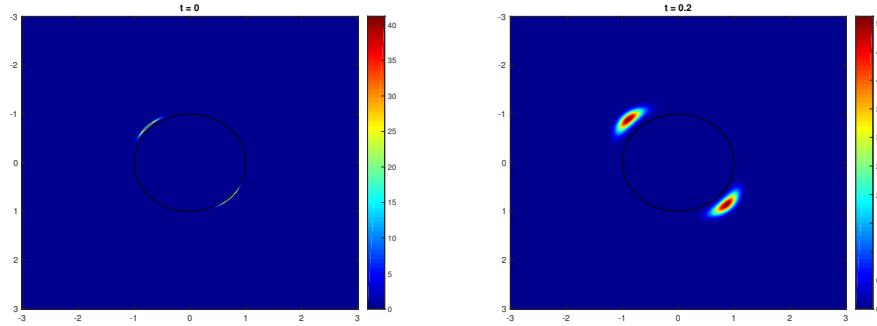


Figure 4: The distribution $\rho(t, x)$ generating the convection on the mass for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. The black circle is the boundary of $B_1(0)$.

From the figures, we can see that our proposed method identifies simultaneously the two non-unique points closest from two mass lumps at antipodal positions to the ball in the center. In particular, the projected measure is the average of the two Dirac masses at the boundary of the circle, where each of them is the closest point of the mass lumps. The algorithm uses reversed time, and the reconstruction moves from the points on the balls to the two respective masses. Also, randomness of stochastic gradient descent and exponential error growth with respect to initial condition of the initial value in PDE also contribute to some error and assymetry.

As for comparison, we have also computed the closed form solution given by Lemma 2.13 when $\zeta = 0$. In Figure 5, we plot the distributions $\rho(t, x)$ for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. As is shown in the figures, the movements of the major lumps/bulks matches, except with some minor deflects. The deflects are caused by both numerical diffusion of the numerical scheme, and as well as regularization parameters ζ .



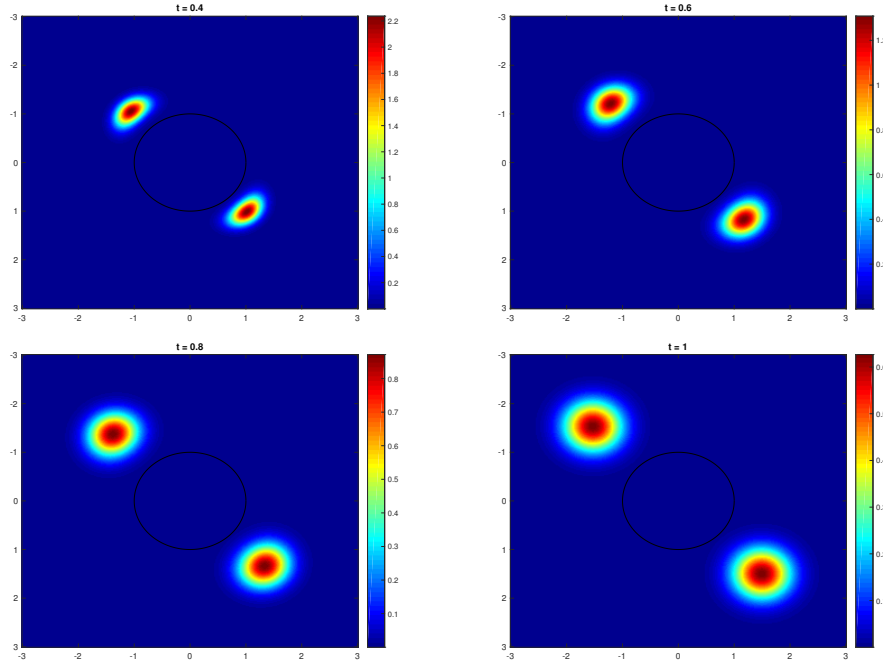


Figure 5: The distribution $\rho(t, x)$ generating the convection on the mass for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$ given by Lemma 2.11 The black circle is the boundary of $B_1(0)$.

Example 5.2. In this example, we consider the Hamiltonian H_1 above and the input distribution $\rho(x)$ as follows:

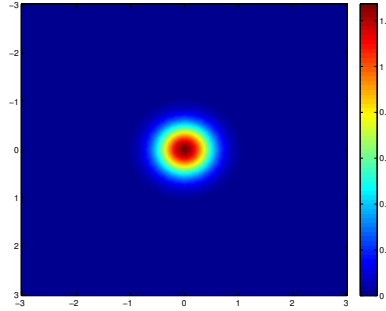


Figure 6: The distribution of the input ρ .

We choose the center and radius (x_0, R) , which helps to define $G(\rho)$ as $x_0 = (3, 0), R = 2$. Since this is a torus, the mass sees a “non-convex” object from both sides from afar. We fix $L_i = 2$ in Algorithm 1.

So as to provide illustration for the multi-level effect using Algorithm 1 to compute (3.4), an optimizer obtained for (4.1) in each level as in Figure 7, for $i = 1, \dots, 4$ in one trial of randomization is plot. As one may see, the minimizers are successively refined. The randomness and asymmetry comes from the random nature of stochastic gradient descent.

Figure 8 gives the optimizer $\tilde{\Phi}$ (left) in (2.3) in the finest level and its gradient $\nabla_x \tilde{\Phi}$ (right) computed using Algorithm 1 when $t = 1$ in the Hamiltonian.

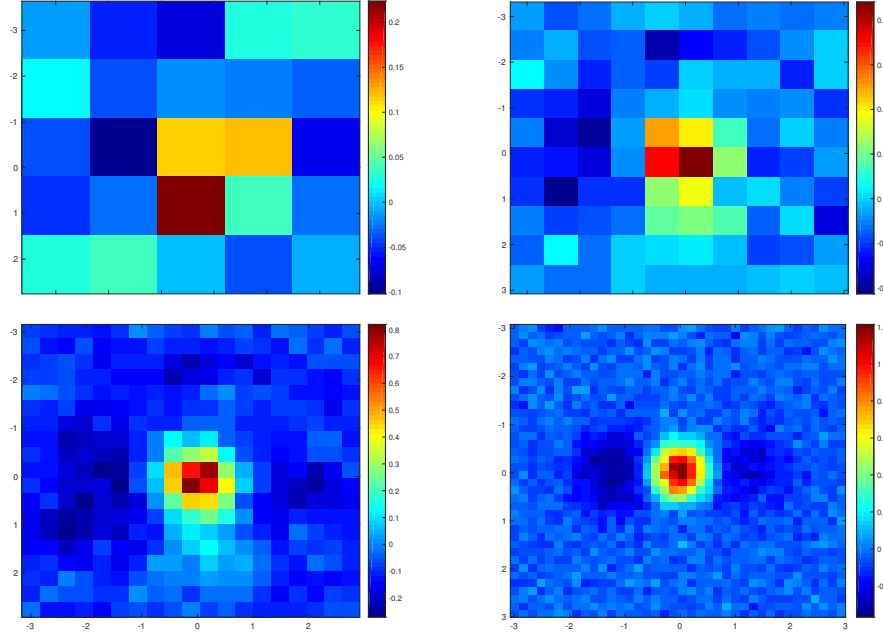


Figure 7: Left: optimizer $\tilde{\Phi}$ in $G_i([\Phi_i]_l^k)$ in (4.1) on coarse grids, $i = 1, \dots, 4$.

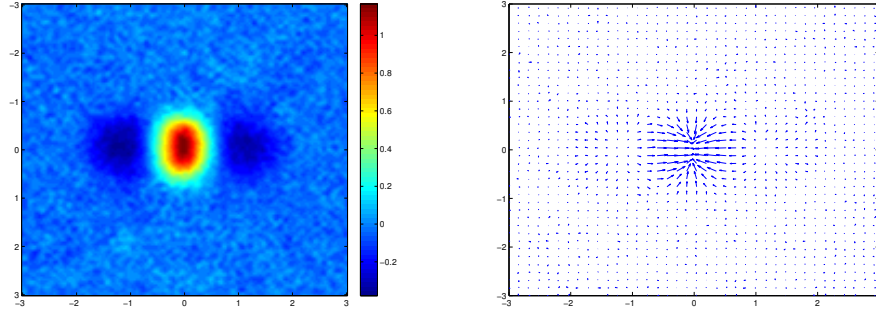


Figure 8: Left: optimizer $\tilde{\Phi}$ in $G_i([\Phi_i]_l^k)$ in (4.1), $i = 5$, i.e. the finest approximation of $U(t, \rho)$ where $t = 1$ in (2.3), right: vector field $\nabla_x \tilde{\Phi}$.

Table 2 now shows also, for each i, k , the respective number of epochs is defined as $\text{ceil}(k/\text{DoF})$, the minimum value function $G_i([\Phi_i]_l^k)$ obtained, where k is given by $[\Phi_i]_{\text{final}} = [\Phi_i]_l^{k+1}$ and G_i is given as (4.1), and the consecutive ergodic error $Er_{i,k}$ defined as in (4.2). This illustrates the successive refinement nature of the method, and also the effectiveness of the method, that a small number of epochs

Level i	Degree of Freedom (DoF)	Number of Epochs	Value function $G_i([\Phi_i]_l^k)$	Error $Er_{i,k}$
1	25	2	0.1724	1.7999×10^{-3}
2	100	2	0.2348	4.5542×10^{-4}
3	361	4	0.4236	8.3531×10^{-5}
4	1444	17	0.5373	9.4764×10^{-6}
5	5625	1	0.5560	1.0753×10^{-6}

Table 2: Number of epochs, minimum value function $G_i([\Phi_i]_l^k)$ obtained and consecutive ergodic error $Er_{i,k}$ defined as in (4.2).

In Figure 9, we plot the distributions $\rho(t, x)$ for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$.

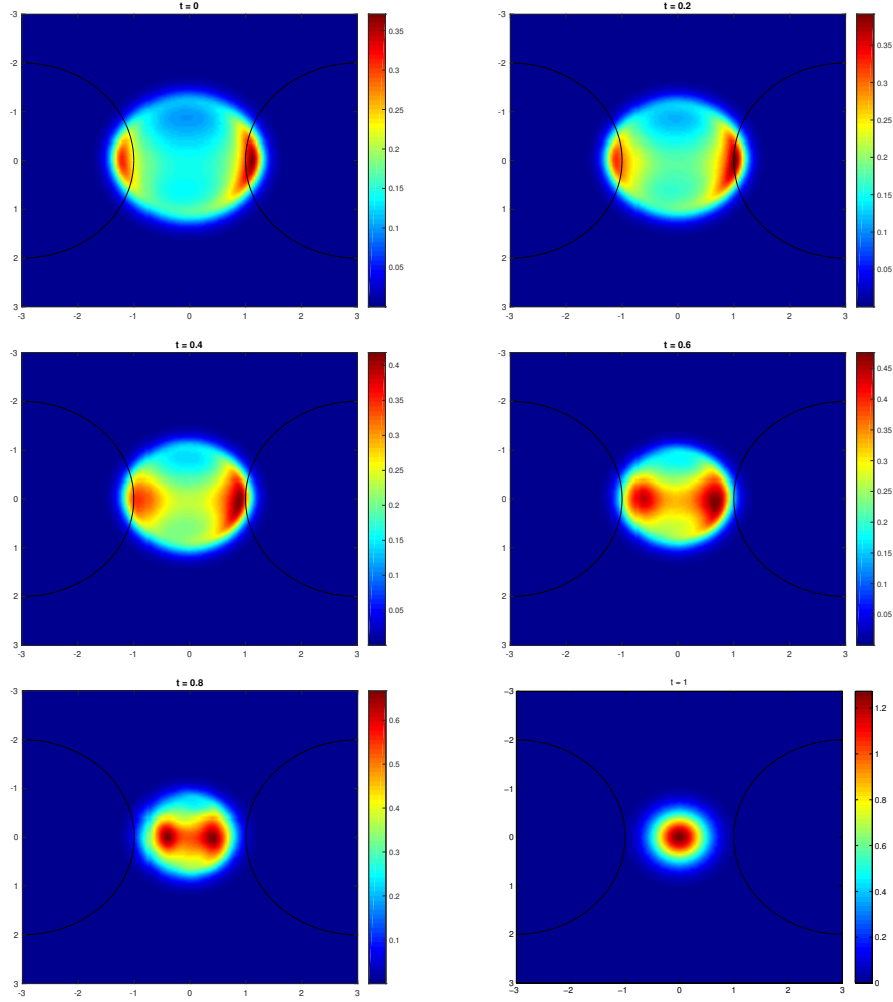


Figure 9: The distribution $\rho(t, x)$ generating the convection on the mass for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. The black circle is the boundary of $B_1(0)$.

In this example, our method accurately finds the projections of a mass to the two non-unique closest points on the “non-convex” body (which is in fact a ball “split” in two). In particular, the flow of the mass splits into two opposite directions; each brings half of the densities to the boundary of the target body.

We also computed the closed form solution given by Lemma 2.13 when $\zeta = 0$. In Figure 10, we plot the distributions $\rho(t, x)$ for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. With non-unique closest points on the “non-convex” body, the movement of the ball splits as in that from Hopf formula. The movements of the split matches, again except with some minor deflects. The deflects are caused by both numerical diffusion of the numerical scheme, and as well as regularization parameters ζ . Moreover, randomness of stochastic gradient descent and exponential error growth with respect to initial condition of the initial value in PDE also contribute to some error and asymmetry.

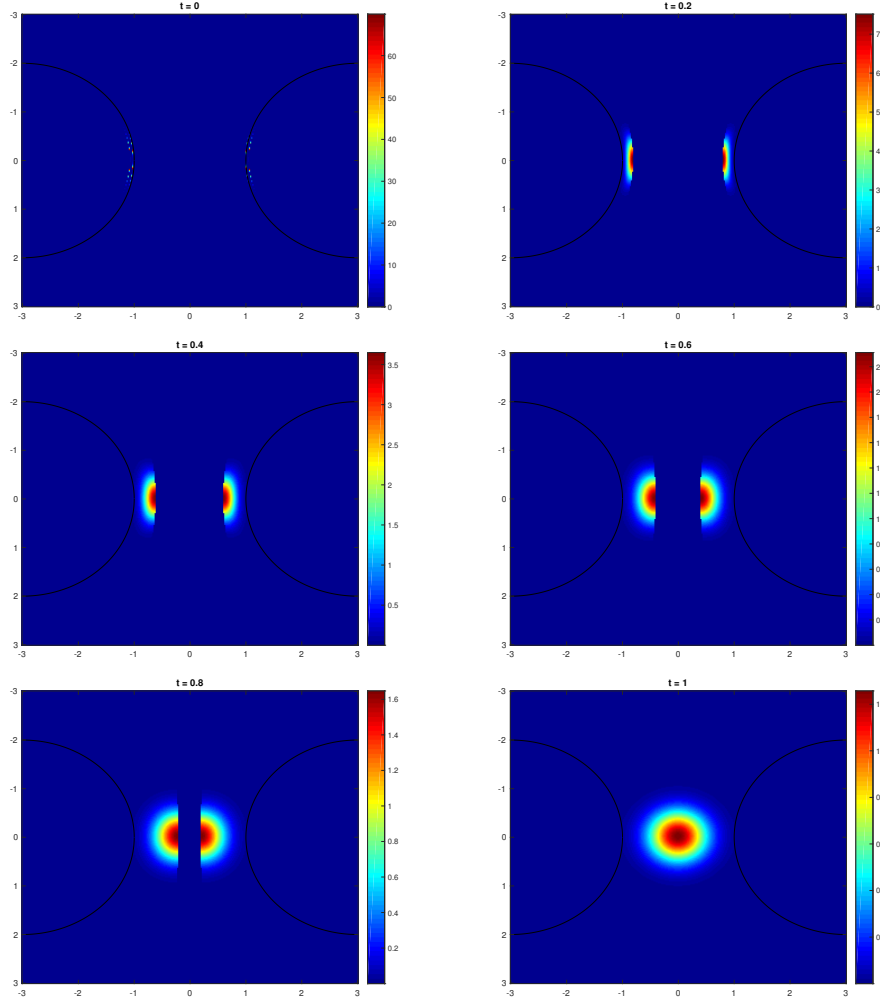


Figure 10: The distribution $\rho(t, x)$ generating the convection on the mass for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$ given by Lemma 2.11 The black circle is the boundary of $B_1(0)$.

Example 5.3. In this example, we consider the non-smooth but convex Hamiltonian H_3 above and the input distribution $\rho(x)$ in Figure 11.

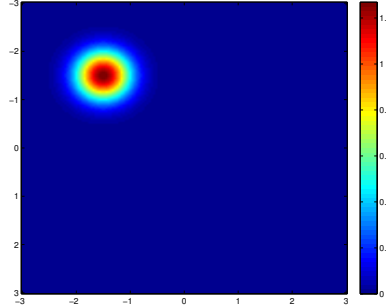


Figure 11: The distribution of the input ρ .

In order to compute the absolute value in the conservation law in a stable way, we replace the sign function with a soft sign function as follows

$$\text{soft-sign}(x) := \frac{2}{\pi} \arctan(cx),$$

where we choose $c = 20$. With this regularization, the Hamiltonian under consideration is actually

$$H_{3,c}(p) := H_c(p_1) + H_c(p_2)$$

where

$$H_c(s) = \frac{2}{\pi} s \arctan(s) - \frac{1}{c\pi} \log(1 + c^2 s^2),$$

and thus the Lagrangian cost functional is

$$L_{3,c}(v) := L_c(v_1) + L_c(v_2),$$

where

$$L_c(t) = \begin{cases} \frac{1}{c\pi} \log(\sec^2(\frac{\pi t}{2})) & \text{if } |t| < 1, \\ \infty & \text{if } |t| \geq 1. \end{cases}$$

As in Example 1, we choose the center and radius (x_0, R) which helps to define $G(\rho)$ as $x_0 = (0, 0)$, $R = 1$.

To illustrate for the multi-level effect using Algorithm 1 when we compute compute (3.4), an optimizer obtained for (4.1) in each level as in Figure 12, for $i = 1, \dots, 4$ in one trial of randomization is plot. As one may see, the minimizers are successively refined. The randomness and assymetry comes from comes the random nature of stochastic gradient descent, but you see the optimizer looks similar to the finer level in a very early stage.

Figure 13 gives the optimizer $\tilde{\Phi}$ (left) in (2.3) in the finest level and its gradient $\nabla_x \tilde{\Phi}$ (right) computed using Algorithm 1 when $t = 1$ in the Hamiltonian.

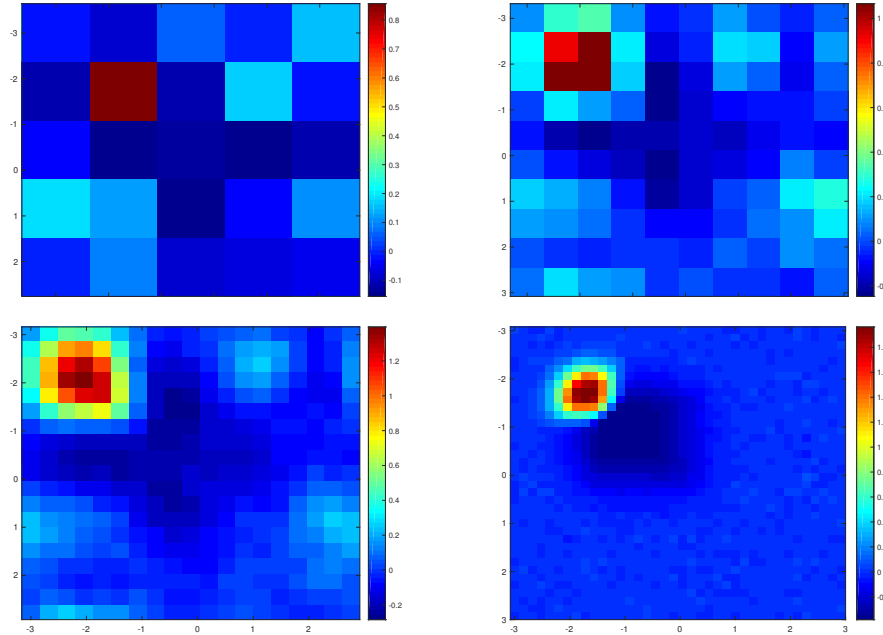


Figure 12: Left: optimizer $\tilde{\Phi}$ in $G_i([\Phi_i]_l^k)$ in (4.1) on coarse grids, $i = 1, \dots, 4$.

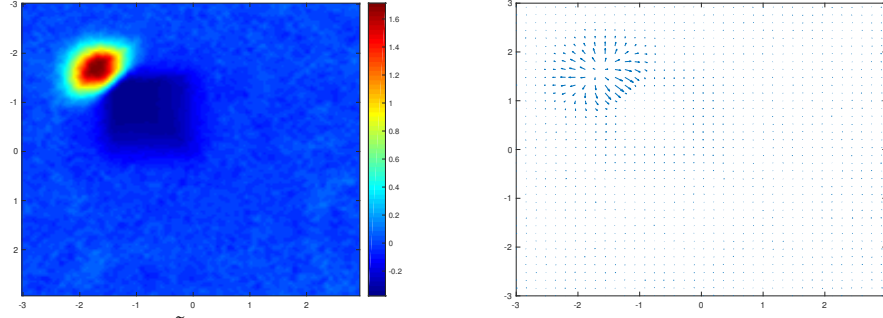


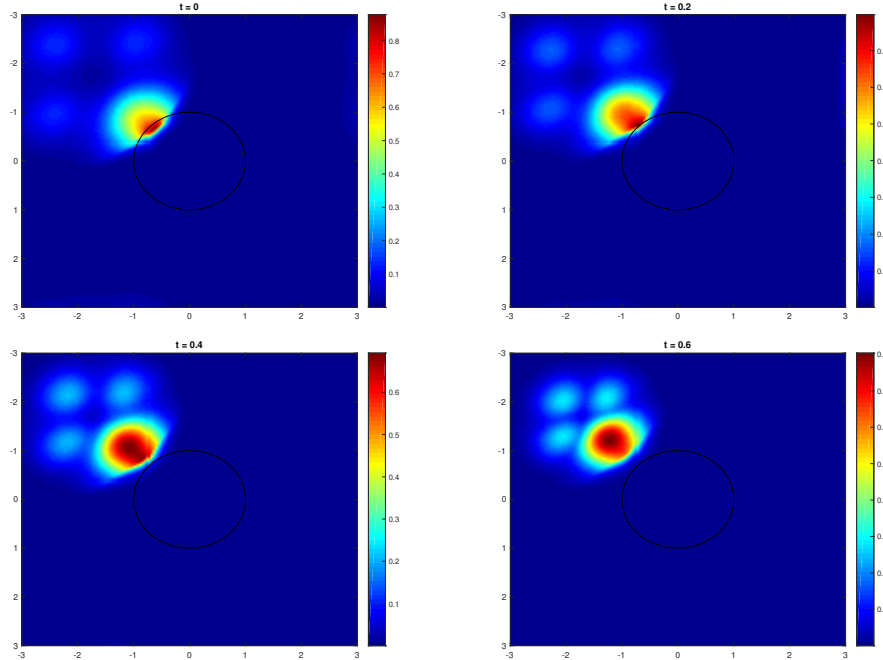
Figure 13: Left: optimizer $\tilde{\Phi}$ in $G_i([\Phi_i]_l^k)$ in (4.1), $i = 5$, i.e. the finest approximation of $U(t, \rho)$ where $t = 1$ in (2.3), right: vector field $\nabla_x \tilde{\Phi}$.

Table 3 now shows also, for each i, k , the respective number of epochs is defined as $\text{ceil}(k/\text{DoF})$, the minimum value function $G_i([\Phi_i]_l^k)$ obtained, where k is given by $[\Phi_i]_{\text{final}} = [\Phi_i]_l^{k+1}$ and G_i is given as (4.1), and the consecutive ergodic error $Er_{i,k}$ defined as in (4.2).

Level i	Degree of Freedom (DoF)	Number of Epoch	Value function $G_i([\Phi_i]_l^k)$	Error $Er_{i,k}$
1	25	4	0.6898	3.5999×10^{-3}
2	100	10	0.9309	2.2014×10^{-4}
3	361	27	0.9741	1.9953×10^{-5}
4	1444	99	1.3015	1.6259×10^{-6}
5	5625	76	1.3526	1.5398×10^{-6}

Table 3: Number of epochs, minimum value function $G_i([\Phi_i]_l^k)$ obtained and consecutive ergodic error $Er_{i,k}$ defined as in (4.2).

Figure 18 plots the distributions $\rho(t, x)$ for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$.



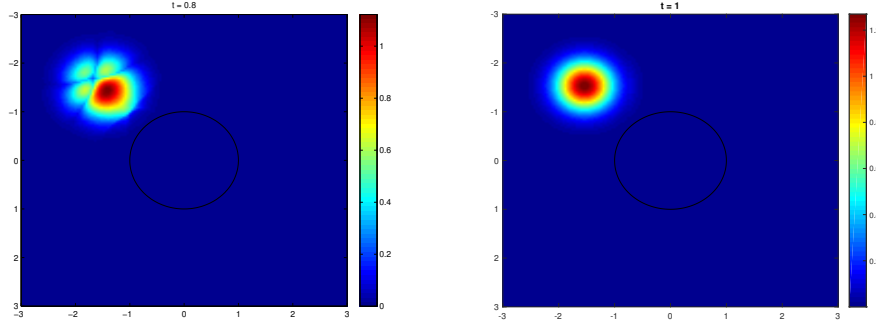
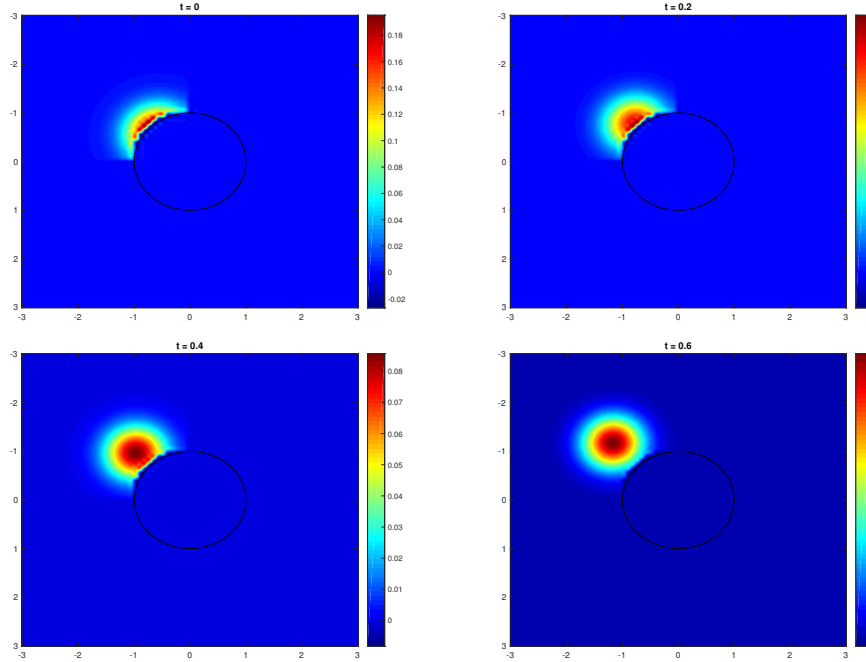


Figure 14: The distribution $\rho(t, x)$ generating the convection on the mass for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. The black circle is the boundary of $B_1(0)$.

We identify reachability of the measure to the boundary of the ball w.r.t. to the l_1 Hamiltonian. With our choice of regularization, we, however, see a defect in our numerical computation: there are three small tails that are left behind in the conservation law as the mass is moving since the exact cutoff of the absolute value is regularized. Nonetheless, the solution makes perfect sense in term of identifying reachability.

To compare the behavior, we also computed the closed form solution given by Lemma 2.13 when $\zeta = 0$. In Figure 15, we plot the distributions $\rho(t, x)$ for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$.

Since $L(x, v) = \iota_{\{\|\cdot\|_\infty \leq 1\}}(v)$, we have non-uniqueness of even the projection as some points will have infinity, rendering the value function to be infinite if ζ is zero. However, we may still obtain a sensible solution that we shall expect by projecting the points as close possible to the unique circle as possible using the infinity-norm ball. Comparing that with the solution computed from Hopf formula, the movement of the mass still matches with some defects caused by non-uniqueness, regularization (now both by ζ and c), and numerical diffusion. However as we can see, the major part of the mass still matches and move with constant speed as we expect.



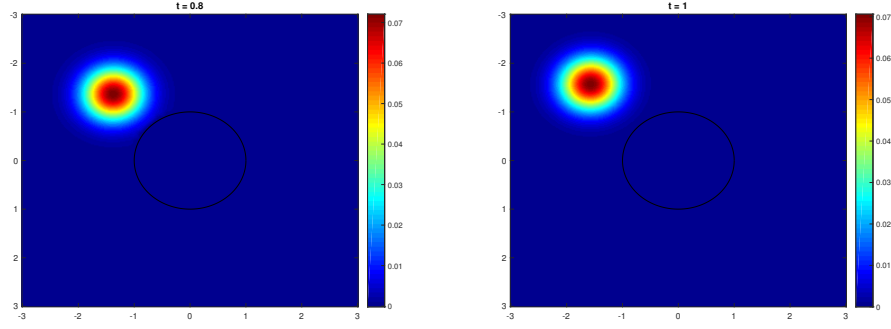


Figure 15: The distribution $\rho(t, x)$ generating the convection on the mass for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$ given by Lemma 2.11 The black circle is the boundary of $B_1(0)$.

Example 5.4. In this example, we consider the non-convex Hamiltonian H_2 above and the input distribution $\rho(x)$ the same as in Example 1 in Figure 1. Again, we choose the center and radius (x_0, R) , which helps to define $G(\rho)$ as $x_0 = (0, 0), R = 1$.

For the illustration of the multi-level effect using Algorithm 1 when we compute compute (3.4), an optimizer obtained for (4.1) in each level as in Figure 16, for $i = 1, \dots, 4$ in one trial of randomization is plot. The successive refinement property is still observed, that on a coarse level the solution looks similar to the fine level. The stochastic nature again contributes to the randomness and assymetry.

Figure 17 gives the optimizer $\tilde{\Phi}$ (left) in (2.3) in the finest level and its gradient $\nabla_x \tilde{\Phi}$ (right) computed using Algorithm 1 when $t = 1$ in the Hamiltonian.

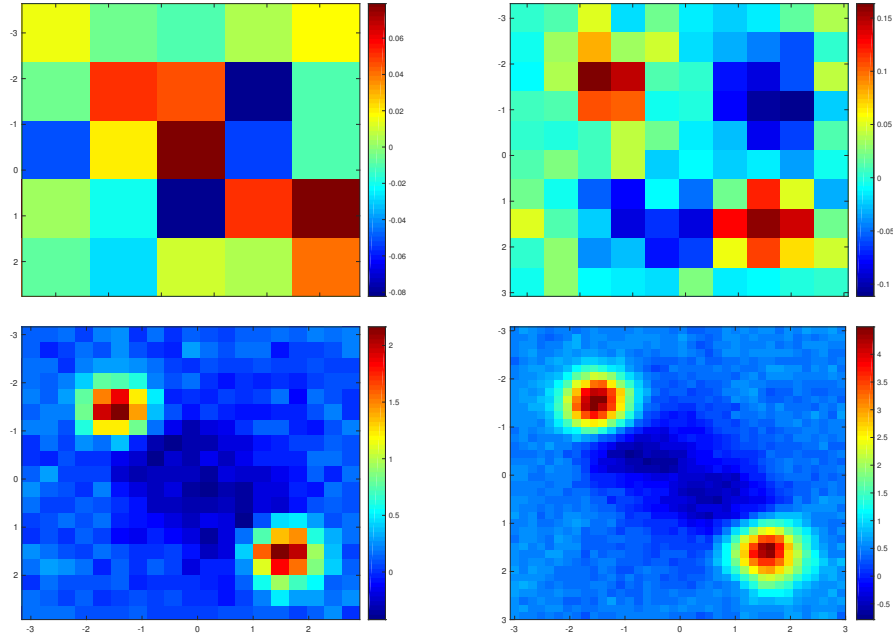


Figure 16: Left: optimizer $\tilde{\Phi}$ in $G_i([\Phi_i]_l^k)$ in (4.1) on coarse grids, $i = 1, \dots, 4$.

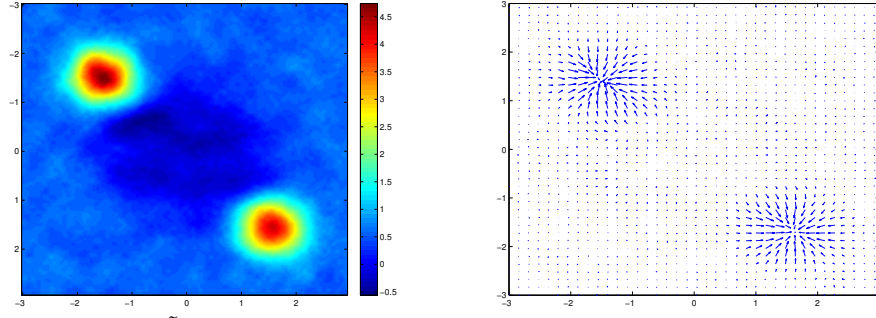


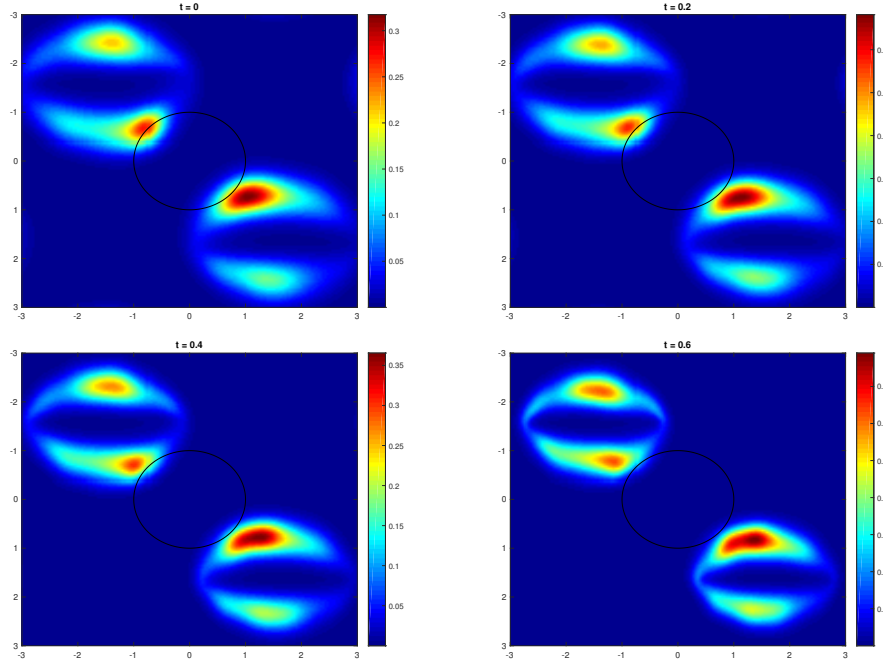
Figure 17: Left: optimizer $\tilde{\Phi}$ in $G_i([\Phi_i]_l^k)$ in (4.1), $i = 5$, i.e. the finest approximation of $U(t, \rho)$ where $t = 1$ in (2.3), right: vector field $\nabla_x \tilde{\Phi}$.

We now have Table 4 showing, for each i, k , the respective number of epochs is defined as $\text{ceil}(k/\text{DoF})$, the minimum value function $G_i([\Phi_i]_l^k)$ obtained, where k is given by $[\Phi_i]_{\text{final}} = [\Phi_i]_l^{k+1}$ and G_i is given as (4.1), and the consecutive ergodic error $Er_{i,k}$ defined as in (4.2).

Level i	Degree of Freedom (DoF)	Number of Epoch	Value function $G_i([\Phi_i]_l^k)$	Error $Er_{i,k}$
1	25	1	0.0246	5.3129×10^{-4}
2	100	4	0.0975	2.1833×10^{-4}
3	361	27	1.1421	5.4065×10^{-5}
4	1444	42	2.5169	1.0120×10^{-5}
5	5625	77	3.4000	9.9348×10^{-6}

Table 4: Number of epochs, minimum value function $G_i([\Phi_i]_l^k)$ obtained and consecutive ergodic error $Er_{i,k}$ defined as in (4.2).

Figure 18 plots the distributions $\rho(t, x)$ for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$.



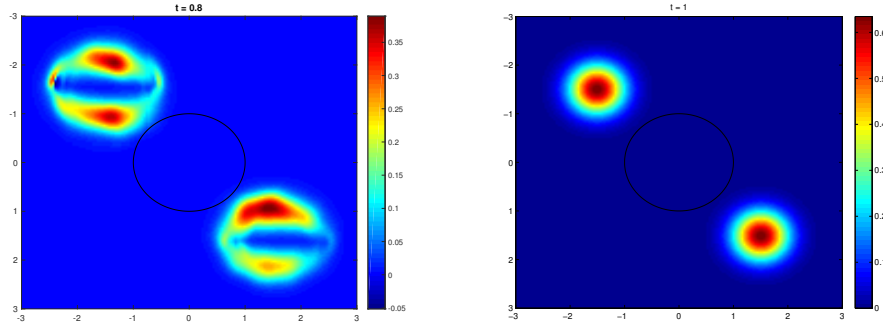


Figure 18: The distribution $\rho(t, x)$ generating the convection on the mass for different $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. The black circle is the boundary of $B_1(0)$.

We can see the competing nature of the Hamiltonian, where one part of the Hamiltonian tries to drift the mass to the ball inward along one direction, while the other part of the Hamiltonian tries to drift the mass away along another direction. In fact one may see clearly that they are drifting along the characteristics lines of H_2 , although half in forward direction and half in backward directions. This tears each mass lump apart into two lumps. The numerical behavior of the solution shows the competing nature of a differential game problem in the mean field setting.

To understand our computed solution more thoroughly, we may refer our readers to section 2.3.2. Although it is not known that Lemma 2.14 holds when $\zeta = 0$ which gives G not satisfying the bounded and Lipschitz condition, Lemma 2.14 does hold after regularization when $\zeta > 0$, which gives G satisfying the bounded and Lipschitz condition. Therefore we may understand our numerical solution as the solution both as the upper value and lower value according to Lemma 2.14. Indeed, as we can see, with our Hamiltonian H_2 , control I is trying to have the vector component m stays in $|m| < 1$ and pushing the support of $\rho(0)$ out from C and control II is trying to have the vector component n stays in $|n| < 1$ and pushing the support of $\rho(0)$ out into C .

6 Discussions

To summarize, we propose a generalized Hopf formula for potential mean field games. Our algorithm inherits main ideas in optimal transport on graphs and the Hopf formula for state-dependent optimal control problems.

Compared to the existing methods, the advantage of the proposed algorithm is three fold. First, the Hopf formula in density space introduces a minimization with variables depending on solely spatial grids. It has a lower complexity than the original optimal control problem. Second, the Hopf formula gives a simple parameterization for boundary problems in NE. This parameterization helps us design a simple first-order gradient descent method. This property allows us to compute the case of nonconvex Hamiltonians efficiently. Finally, our spatial discretization follows the dual of optimal transport on graphs. Hence, it is approximately discrete time reversible. This property conserves the primal-dual structure of potential mean field games.

We would like to emphasize that our rigorous derivation of the generalized Hopf formula is provided only for a relatively narrow case. We notice that the Hopf formula in density space appears to go beyond monotonicity conditions and give legitimate numerical results, as shown in Section 5. Although it is beyond the scope of this paper, it is interesting to search for the precise conditions for the validity of the Hopf formula. Also, our current study only considers potential games without noise perturbations in players' decision processes. We will extend it to compute NEs for general non-potential games in future work.

References

- [1] Y. Achdou, F. Camilli, and I. Capuzzo-Dolcetta. Mean Field Games: Numerical Methods for the Planning Problem. *SIAM Journal on Control and Optimization*, 50(1):77–109, 2012.
- [2] Y. Achdou, F. Camilli, and I. Capuzzo-Dolcetta. Mean Field Games: Convergence of a Finite Difference Method. *SIAM Journal on Numerical Analysis*, 51(5):2585–2612, 2013.
- [3] Y. Achdou and I. Capuzzo-Dolcetta. Mean Field Games: Numerical Methods. *SIAM Journal on Numerical Analysis*, 48(3):1136–1162, 2010.
- [4] N. Almula, R. Ferreira, and D. Gomes. Two Numerical Approaches to Stationary Mean-Field Games. *Dynamic Games and Applications*, 7(4):657–682, 2017.
- [5] 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.
- [6] J.-D. Benamou and G. Carlier. Augmented Lagrangian Methods for Transport Optimization, Mean Field Games and Degenerate Elliptic Equations. *Journal of Optimization Theory and Applications*, 167(1):1–26, 2015.
- [7] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman Projections for Regularized Transportation Problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.
- [8] P. Cardaliaguet. Notes on mean field games. *preprint*, 2013.
- [9] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions. The master equation and the convergence problem in mean field games. *arXiv:1509.02505 [math]*, 2015.
- [10] P. Cardaliaguet and C. Jimenez. Optimal transport with convex obstacle. *Journal of Mathematical Analysis and Applications*, 381(1):43–63, 2011.
- [11] S.-N. Chow, L. Dieci, W. Li, and H. Zhou. Entropy dissipation semi-discretization schemes for Fokker-Planck equations. *arXiv:1608.02628 [math]*, 2016.
- [12] S.-N. Chow, W. Li, and H. Zhou. A discrete Schrodinger equation via optimal transport on graphs. *arXiv:1705.07583 [math]*, 2017.
- [13] S.-N. Chow, W. Li, and H. Zhou. Entropy dissipation of Fokker-Planck equations on graphs. *arXiv:1701.04841 [math]*, 2017.
- [14] Y. T. Chow, J. Darbon, S. Osher, and W. Yin. Algorithm for Overcoming the Curse of Dimensionality For Time-Dependent Non-convex Hamilton–Jacobi Equations Arising From Optimal Control and Differential Games Problems. *Journal of Scientific Computing*, 73(2-3):617–643, 2017.
- [15] Y. T. Chow, J. Darbon, S. Osher, and W. Yin. Algorithm for overcoming the curse of dimensionality for certain non-convex hamilton-jacobi equations, projections and differential games. *Annals of Mathematical Sciences and Applications*, 3:369–403, 2018.
- [16] Y. T. Chow, J. Darbon, S. Osher, and W. Yin. Algorithm for Overcoming the Curse of Dimensionality for State-dependent Hamilton–Jacobi equations. *Journal of Computational Physics*, 387:376–409, 2019.

- [17] J. Darbon and S. Osher. Algorithms for overcoming the curse of dimensionality for certain Hamilton–Jacobi equations arising in control theory and elsewhere. *Research in the Mathematical Sciences*, 3(1), 2016.
- [18] L. C. Evans. *Partial Differential Equations*. Number v. 19 in Graduate studies in mathematics. American Mathematical Society, Providence, R.I, 2nd ed edition, 2010.
- [19] L. C. Evans and P. E. Souganidis. Differential games and representation formulas for solutions of hamilton-jacobi-isaacs equations. Technical report, WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER, 1983.
- [20] W. Gangbo, W. Li, and C. Mou. Geodesic of minimal length in the set of probability measures on graphs. *arXiv:1712.09266 [math]*, 2017.
- [21] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.
- [22] W. Gangbo, T. Nguyen, and A. Tudorascu. *Hamilton-Jacobi Equations in the Wasserstein Space*. 2008.
- [23] W. Gangbo and A. Swiech. Existence of a solution to an equation arising from the theory of Mean Field Games. *Journal of Differential Equations*, 259(11):6573–6643, 2015.
- [24] O. Guéant, J.-M. Lasry, and P.-L. Lions. Mean Field Games and Applications. In J.-M. Morel, F. Takens, and B. Teissier, editors, *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003, pages 205–266. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [25] M. Huang, R. P. Malhamé, and P. E. 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.
- [26] H. Komiya. Elementary proof for Sion’s minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7, 1988.
- [27] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.
- [28] W. Li, P. Yin, and S. Osher. Computations of Optimal Transport Distance with Fisher Information Regularization. *Journal of Scientific Computing*, 2017.
- [29] D. Monderer and L. S. Shapley. Potential Games. *Games and Economic Behavior*, 14(1):124–143, 1996.
- [30] E. Nelson. Derivation of the Schrödinger Equation from Newtonian Mechanics. *Physical Review*, 150(4):1079–1085, 1966.
- [31] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.
- [32] C. Villani. *Optimal Transport: Old and New*. Number 338 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.

- [33] I. Yegorov and P. Dower. Perspectives on characteristics based curse-of-dimensionality-free numerical approaches for solving Hamilton-Jacobi equations. *arXiv:1711.03314 [math]*, 2017.