MATHEMATICS OF COMPUTATION
Volume 91, Number 335, May 2022, Pages 1019–1075
https://doi.org/10.1090/mcom/3726
Article electronically published on March 14, 2022

TIME DISCRETIZATIONS OF WASSERSTEIN–HAMILTONIAN FLOWS

JIANBO CUI, LUCA DIECI, AND HAOMIN ZHOU

ABSTRACT. We study discretizations of Hamiltonian systems on the probability density manifold equipped with the L^2-Wasserstein metric. Based on discrete optimal transport theory, several Hamiltonian systems on a graph (lattice) with different weights are derived, which can be viewed as spatial discretizations of the original Hamiltonian systems. We prove consistency of these discretizations. Furthermore, by regularizing the system using the Fisher information, we deduce an explicit lower bound for the density function, which guarantees that symplectic schemes can be used to discretize in time. Moreover, we show desirable long time behavior of these symplectic schemes, and demonstrate their performance on several numerical examples. Finally, we compare the present approach with the standard viscosity methodology.

1. Introduction

In recent years, there has been a lot of interest in studying Hamiltonian systems defined on the probability space endowed with the L^2-Wasserstein metric, also known as Wasserstein manifold, and several authors have been concerned with their connections to some well-known partial differential equations (PDEs); e.g., see [1, 21, 35, 45].

Our present study is influenced by the point of view in [13], where the authors showed that the push-forward density of a classical Hamiltonian vector field in phase space is a Hamiltonian flow on the Wasserstein manifold. To be more precise, consider a Hamiltonian system subject to initial condition $(q_0, v(0, q_0))$:

$$\begin{aligned} \frac{d}{dt} q(t, q(t, q_0)) &= \frac{\partial H}{\partial v}\big(v(t, q(t, q_0)), q(t, q_0)\big), \\ \frac{d}{dt} v(t, q(t, q_0)) &= -\frac{\partial H}{\partial q}\big(v(t, q(t, q_0)), q(t, q_0)\big), \end{aligned} \tag{1.1}$$

where the position $q(t, q_0) \in \mathbb{R}^d$, the conjugate momenta $v(t, q(t, q_0)) \in \mathbb{R}^d$, and H is the real-valued Hamiltonian $H(v, q) = \frac{1}{2}|v|^2 + F(q)$ with a smooth potential $F(q)$. Here, $|v|^2 = v \cdot v$. If the initial position q_0 is a random vector having probability

Received by the editor June 11, 2020, and, in revised form, September 9, 2021.
2020 *Mathematics Subject Classification.* Primary 65P10; Secondary 35R02, 58B20, 65M12.
Key words and phrases. Wasserstein–Hamiltonian flow, symplectic schemes, optimal transport, Fisher information.
The research was partially supported by Georgia Tech Mathematics Application Portal (GT-MAP) and by research grants NSF DMS-1620345. DMS-1830225, and ONR N00014-18-1-2852. The research of the first author was partially supported by start-up funds (P0039016) from Hong Kong Polytechnic University and the CAS AMSS-PolyU Joint Laboratory of Applied Mathematics.
The first author is the corresponding author.

density μ^0, it can be shown that the position $q(t, q_0)$, following (1.1), is also a random vector whose density $\rho(t, q)$ satisfies

$$\begin{aligned} \frac{\partial}{\partial t}\rho(t,q) + \nabla \cdot \big(\frac{\partial H}{\partial v}\big(v(t,q),q\big)\rho(t,q)\big) = 0, \\ \frac{\partial}{\partial t}v(t,q) + \nabla v(t,q)\cdot v(t,q) + \nabla \cdot \frac{\partial H}{\partial q}\big(v(t,q),q\big) = 0, \end{aligned} \tag{1.2}$$

where ∇ is the gradient operator with respect to q. In other words, viewing (1.1) as a push-forward map, it transforms μ^0 to $\rho(t, q)$ according to (1.2). When $v(t, q)$ has a primitive function $S(t, q)$, meaning $v(t, q) = \nabla S(t, q)$, the second equation in (1.2) can be written as a Hamilton–Jacobi equation

$$\frac{\partial}{\partial t}S(t,q) + H(\nabla S(t,q),q) = C(t),$$

where $C(t)$ is an arbitrary function independent of q. Taking $C(t) = 0$, (1.2) becomes

$$\begin{aligned} \frac{\partial}{\partial t}\rho(t,q) + \nabla \cdot \big(\nabla S(t,q)\rho(t,q)\big) = 0, \\ \frac{\partial}{\partial t}S(t,q) + H(\nabla S(t,q),q) = 0, \end{aligned} \tag{1.3}$$

which can be viewed as a Hamiltonian flow on the Wasserstein manifold [13].

Inspired by the relationship between (1.1) and (1.3), in this paper we focus on the following coupled system of PDEs,

$$\begin{aligned} \frac{\partial}{\partial t}\rho(t,q) + \nabla \cdot \big(\frac{\partial H}{\partial v}\big(\nabla S(t,q),q\big)\rho(t,q)\big) = 0, \\ \frac{\partial}{\partial t}S(t,q) + H(\nabla S(t,q),q) = 0. \end{aligned} \tag{1.4}$$

This can be rewritten as

$$\begin{aligned} \frac{\partial}{\partial t}\rho(t,q) &= \frac{\delta}{\delta S(t,q)}\mathcal{H}(\rho(t,\cdot),S(t,\cdot)), \\ \frac{\partial}{\partial t}S(t,q) &= -\frac{\delta}{\delta \rho(t,q)}\mathcal{H}(\rho(t,\cdot),S(t,\cdot)), \end{aligned}$$

where $\frac{\delta}{\delta\rho}$, $\frac{\delta}{\delta S}$ are functional derivatives, and

$$\mathcal{H}(\rho,S) = \int_{\mathbb{R}^d}\frac{1}{2}|\nabla S(q)|^2\rho(q)dq + \int_{\mathbb{R}^d}F(q)\rho(q)dq$$

with $|\nabla S|^2 = \nabla S\cdot\nabla S$. This viewpoint explains that (1.4) is a *Wasserstein–Hamiltonian* system as defined in [13].

The formulation (1.4) is remarkably powerful and general. Indeed, with different choices of the Hamiltonian H, the Wasserstein–Hamiltonian system (1.4) leads to differential equations arising in many different applications. For example, by taking $H(v, q) = \frac{1}{2}|v|^2$, one obtains the well-known geodesic equations between two densities μ^0 and μ^1 on the Wasserstein manifold:

$$\begin{aligned} \frac{\partial}{\partial t}\rho(t,q) + \nabla \cdot \big(\rho(t,q)\nabla S(t,q)\big) = 0, \\ \frac{\partial}{\partial t}S(t,q) + \frac{1}{2}|\nabla S(t,q)|^2 = 0, \end{aligned} \tag{1.5}$$

subject to boundary conditions $\rho(0) = \mu^0, \rho(1) = \mu^1$. In the seminal paper [2], it has been proven that the solution of (1.5) is a minimizer of the following variational problem, commonly known as the *Benamou-Brenier formula*:

$$\begin{aligned} g_W^2(\mu^0,\mu^1) = \inf_{v(t)} \Big\{ \int_0^1 \langle v(t), v(t)\rangle_{\rho(t)} dt : \\ \frac{\partial}{\partial t}\rho(t,q) + \nabla\cdot\big(\rho(t,q)v(t,q)\big) = 0, \rho(0) = \mu^0, \rho(1) = \mu^1 \Big\}, \end{aligned} \tag{1.6}$$

where $\langle v(t), v(t)\rangle_{\rho(t)} = \int_{\mathbb{R}^d} |v(t,q)|^2 \rho(t,q)dq$. As shown in [2], the optimal value $g_W(\mu^0,\mu^1)$ is the L^2-Wasserstein distance between μ^0 and μ^1.

Similarly, a problem known as the *Schrödinger Bridge Problem* can be stated as

$$\begin{aligned} \inf_{v(t)} \Big\{ \int_0^1 \frac{1}{2}\langle v(t), v(t)\rangle_{\rho(t)} + \frac{\hbar^2}{8} I(\rho(t)) dt \ : \\ \frac{\partial}{\partial t}\rho(t,q) + \nabla\cdot\big(\rho(t,q)v(t,q)\big) = 0, \rho(0) = \mu^0, \rho(1) = \mu^1 \Big\}, \end{aligned} \tag{1.7}$$

where $\hbar > 0$ is the Planck constant and $I(\rho) = \langle \nabla \log(\rho), \nabla\log(\rho)\rangle_\rho$ is the *Fisher information*. The minimizer of (1.7) satisfies the Wasserstein–Hamiltonian system (1.4) with the energy $\mathcal{H}(v,\rho) = \frac{1}{2}\int_{\mathbb{R}^d} |v(q)|^2\rho(q)dq - \frac{\hbar^2}{8} I(\rho)$ in density space. Although the Schrödinger Bridge problem is nearly 100 years old, it has recently received attention in control theory and machine learning, see [32, 40, 44].

If we change the sign of the Fisher information in (1.7), we get

$$\begin{aligned} \inf_{v(t)} \Big\{ \int_0^1 \frac{1}{2}\langle v(t), v(t)\rangle_{\rho} - \frac{\hbar^2}{8} I(\rho(t)) dt \ : \\ \frac{\partial}{\partial t}\rho(t,x) + \nabla\cdot\big(\rho(t,x)v(t,x)\big) = 0, \rho(0) = \mu^0, \rho(1) = \mu^1 \Big\}, \end{aligned} \tag{1.8}$$

which is the variational formula that Nelson used to derive the Schrödinger equation, [38]. Its reformulation as Wasserstein–Hamiltonian system becomes the well-known Madelung system [37].

Remark 1.1. The Benamou-Brenier formula (1.6) has been extensively used to study Wasserstein gradient flows; e.g., see [30, 39, 45, 46]. However, much less is known for Wasserstein–Hamiltonian flows, hence for solutions of (1.4) for given initial values. The problem is subtle, mainly because—depending on the initial condition—the solution of (1.4) may develop singularities. Moreover, there are several important properties of the Wasserstein–Hamiltonian flow, such as preservation of symplectic structure and other quantities, which make the numerical approximation of Wasserstein–Hamiltonian flows quite challenging. These considerations have motivated us to carry out the present numerical study.

The Wasserstein–Hamiltonian system we investigate in this work may find potential connections to two other active research fields, namely mean-field control and mean-field games, in which various Hamiltonian structures on the Wasserstein manifold are of interest. For instance, the first-order necessary optimality conditions for optimal control problems in multi-agent systems analysis [6–8, 42], as well as in machine learning problems [19, 27], often lead to Hamiltonian flows on the density space. Besides, the Hamilton–Jacobi equation in Wasserstein space, the second equation in (1.4), plays a central role in mean-field control [5, 11, 28], mean-field games [9, 10, 31], and also in the study of the so-called "master equation" [4, 23] and in problems arising from fluid mechanics [22].

To the best of our knowledge, prior to our present work, there are no numerical analysis results on the full (i.e., space and time) discretization of Wasserstein–Hamiltonian systems. The way we approach this problem is by first using discrete optimal transport techniques to obtain Wasserstein–Hamiltonian systems defined on general graphs, and view these as spatial discretizations of the original Wasserstein–Hamiltonian system. We explicitly show the consistency of the semi-discretizations, and derive lower bounds for the probability density function on different graphs. Then, we combine ideas from discrete optimal transport and symplectic integration to construct fully discrete numerical schemes for the solution of Wasserstein–Hamiltonian systems.

We would like to emphasize the crucial role of the Fisher information in our study. The Fisher information is widely used in many areas in statistics, physics and biology (see e.g. [20]). It appears naturally in some Wasserstein–Hamiltonian systems, such as (1.7), and it has recently been used as a regularization term in computations of optimal transport and Wasserstein gradient flows (see [33, 34] and references therein). In this paper, we advocate the use of Fisher information as a regularization technique to construct symplectic schemes for Wasserstein–Hamiltonian systems. Our analysis, as presented in Theorem 4.1, indicates its clear benefits in numerical computations. Not only the Fisher information helps to maintain the positivity of the density function, but it is also conducive to time reversible and gauge invariant schemes. Moreover, the resulting regularized schemes preserve mass and symplectic structure, and they almost preserve energy for very long times (of $\mathcal{O}(\tau^{-r})$, where r is the order of the numerical scheme and τ is the time step size). In addition, the proposed schemes can be used to solve the two-point boundary value optimal transport problem; see (1.6). Indeed, by combining a multiple shooting method with the proposed structure-preserving numerical methods for (1.4), in [17] we construct a continuation multiple shooting method to solve the optimal transport problem. We want to mention that other regularization techniques, like entropic regularization, have been used in numerical methods for optimal transport problems. For more information, we refer to [3, 41, 43] and references therein.

This paper is organized as follows. In Section 2, we introduce the Wasserstein–Hamiltonian vector field on graphs and study its properties. In Section 3, we give an explicit lower bound of the probability density for the discrete Wasserstein–Hamiltonian flow on different graphs; the proofs of the technical results in this section are in the Appendix. Section 4 is devoted to constructing and analyzing time discretizations, and in particular we develop and study symplectic schemes. To compare with the results we obtain using Fisher information as regularization device, in this Section 4 we also investigate regularized schemes obtained by adding a viscosity term; here (see e.g. [16]), this means adding a small multiple of the discrete Laplacian operator to the Hamilton–Jacobi equation (1.4). Several numerical examples are given in Section 5, and in Section 6 we give a limited discussion of the the limiting behavior as the Fisher regularization parameter goes to 0.

2. Wasserstein–Hamiltonian flow on a finite graph

Our goal in this section is three-fold: to introduce the Wasserstein–Hamiltonian flow on a graph, to recognize it as a consistent spatial discretization of the PDE (1.4), and to show relevant properties of the associated flow. The latter effort is a prelude to Section 4 where also the time discretization is examined.

2.1. **Wasserstein–Hamiltonian flows via discrete optimal transport.** Throughout this paper, we will consider an undirected, connected, graph G, with no self loops or multiple edges (see e.g. [25] for more details on the graph theory). That is, $G = (V, E, \Omega)$, where the node set is $V = \{a_i, 1 \leq i \leq N\}$, the edge set is E, with the corresponding weights $\Omega = \{\omega_{jl}\}_{j,l=1}^N$ satisfying $\omega_{jl} = \omega_{lj} > 0$ if there is an edge between a_j and a_l, and $\omega_{jl} = \omega_{lj} = 0$ otherwise. Below, we will write $(i, j) \in E$ to denote the edge in E between the vertices a_i and a_j, and finally we let $N(i) = \{a_j \in V : (i, j) \in E\}$ be the set of neighbors of the node a_i.

Let us denote the set of discrete probabilities on the graph by $\mathcal{P}(G)$:

$$\mathcal{P}(G) = \Big\{(\rho_i)_{i=1}^N \ : \sum_{i=1}^N \rho_i = 1, \rho_i \geq 0, \text{ for } a_i \in V\Big\},$$

and $\mathcal{P}_o(G)$ is its interior (i.e., all $\rho_i > 0$ for $a_i \in V$).

To describe the discrete Lagrange functional and Hamiltonian flow on the graph, we introduce the following notation. A vector field v on E is a $N \times N$ skew-symmetric matrix, and a weighted inner product of two vector fields u, v is defined by

$$\langle u, v\rangle_{\theta(\rho)} = \frac{1}{2} \sum_{(j,l)\in E} u_{jl} v_{jl} \theta_{jl}(\rho),$$

where the coefficient $1/2$ is taken because every edge in G is counted twice. The density dependent weight $\theta_{ij}(\rho)$ on the edge $(i, j) \in E$ was introduced in [12, 14, 36] to capture the long time dynamics for Fokker-Planck equations on graphs and the dispersion relationship for the Schrödinger equation on graphs. By comparison, we denote the standard l^2-inner product of vectors u, v by $\langle u, v\rangle = \sum_{i=1}^N u_i v_i$. We introduce the discrete divergence of the flux function ρv as

$$div_G^\theta(\rho v) = (\sum_{j\in N(i)} \sqrt{\omega_{ij}} v_{ij}\theta_{ij})_{i=1}^N.$$

Accordingly, the gradient operator ∇_G on the graph is given through the following dual formulation

$$-\langle div_G^\theta(\rho v), \widetilde{\Phi}\rangle = \langle v, \nabla_G \widetilde{\Phi}\rangle_{\theta(\rho)},$$

where $\widetilde{\Phi}$ is a function defined on every node of the graph. Clearly, ∇_G maps a node defined function to a vector field on the edges. For convenience, on graphs we denote

$$\rho(t) = (\rho_i(t))_{i=1}^N,\ v(t) = (v_{ij}(t))_{i,j=1}^N,\ \text{for } t \geq 0.$$

Given real-valued functions $\mathbb{V}_i$ on each node a_i and $\mathbb{W}_{ij}$, with $\mathbb{W}_{ij} = \mathbb{W}_{ji}$, on each edge $(i, j) \in E$, we define a linear potential functional $\mathcal{V}$ and an interaction potential functional $\mathcal{W}$ on G by

$$\mathcal{V}(\rho) = \sum_{i=1}^N \mathbb{V}_i\rho_i,\ \mathcal{W}(\rho) = \frac{1}{2}\sum_{i,j=1}^N \mathbb{W}_{ij}\rho_i\rho_j.$$

The *discrete Fisher information* is defined by

$$I(\rho) = \frac{1}{2}\sum_{i=1}^N \sum_{j\in N(i)} \widetilde{\omega}_{ij}|\log(\rho_i) - \log(\rho_j)|^2\widetilde{\theta}_{ij}(\rho). \tag{2.1}$$

Remark 2.1. Note that in (2.1), we are allowing use of edge weights $\widetilde{\omega}$ and probability weights $\widetilde{\theta}$, different from ω and θ; this added flexibility can be exploited to obtain more robust space discretizations than those obtained when choosing $\widetilde{\omega} = \omega$ and $\widetilde{\theta} = \theta$, as it was done in [12].

For a given parameter $\beta \geq 0$, we define the discrete Lagrangian by

$$\mathcal{L}(\rho, v) = \int_0^T \Big[\frac{1}{2}\langle v(t), v(t)\rangle_{\theta(\rho(t))} - \mathcal{V}(\rho(t)) - \mathcal{W}(\rho(t)) - \beta I(\rho(t))\Big]dt, \tag{2.2}$$

where $T > 0$, $\rho(t) \in \mathcal{P}_o(G), t \in [0, T]$, is subject to the constraints $\rho(0) = \mu^0, \rho(1) = \mu^1$, and

$$\frac{d\rho(t)}{dt} + div_G^\theta\big(\rho(t)v(t)\big) = 0. \tag{2.3}$$

The goal is to study the critical point of $\mathcal{L}(\rho, v)$ subject to the constraint (2.3). Here ρ, v in $\mathcal{L}$ means $\rho(t), v(t), t \in [0, T]$.

As shown in [12], the critical point (ρ, v) of $\mathcal{L}$ can be written as $v = \nabla_G S = \big(\sqrt{\omega_{jl}}(S_j - S_l)\big)_{(j,l)\in E}$ for some function S on V. Consequently, the minimization problem leads to the following Wasserstein–Hamiltonian flow on the cotangent bundle of $\mathcal{P}(G)$:

$$\begin{aligned} &\frac{d\rho_i(t)}{dt} + \sum_{j\in N(i)} \omega_{ij}(S_j(t)-S_i(t))\theta_{ij}(\rho(t)) = 0, \\ &\frac{dS_i(t)}{dt} + \frac{1}{2}\sum_{j\in N(i)} \omega_{ij}(S_i(t)-S_j(t))^2 \frac{\partial\theta_{ij}(\rho(t))}{\partial\rho_i} + \beta\frac{\partial I(\rho(t))}{\partial\rho_i} + \mathbb{V}_i \\ &\quad + \sum_{j=1}^N \mathbb{W}_{ij}\rho_j(t) = 0. \end{aligned} \tag{2.4}$$

With respect to the variables ρ and S, we can rewrite (2.4) as a Hamiltonian system

$$\begin{aligned} \frac{d\rho(t)}{dt} &= \frac{\partial}{\partial S}\mathcal{H}(\rho(t),S(t)), \\ \frac{dS(t)}{dt} &= -\frac{\partial}{\partial\rho}\mathcal{H}(\rho(t),S(t)), \end{aligned} \tag{2.5}$$

where the Hamiltonian $\mathcal{H}(\rho,S) = K(\rho,S) + \mathcal{F}(\rho)$, with

$$K(\rho,S) = \frac{1}{2}\langle\nabla_G S, \nabla_G S\rangle_{\theta(\rho)} \text{ and } \mathcal{F}(\rho) = \beta I(\rho) + \mathcal{V}(\rho) + \mathcal{W}(\rho). \tag{2.6}$$

Particularly, if $T = 1$, $\beta = 0$, $\mathcal{V} = 0$, and $\mathcal{W} = 0$, the infimum of $\mathcal{L}(\rho, v)$ induces the Wasserstein metric on the graph, which is a discrete version of the Benamou-Brenier formula:

$$\begin{aligned} W(\rho^0,\rho^1) = \inf_{v(t)}\Big\{\Big(\int_0^1 \langle v(t),v(t)\rangle_{\theta(\rho(t))}dt\Big)^{\frac{1}{2}} : \\ \frac{d\rho(t)}{dt} + div_G^\theta\big(\rho(t)v(t)\big) = 0,\ \rho(0)=\mu^0,\ \rho(1)=\mu^1\Big\}. \end{aligned}$$

Example 2.1 illustrates the importance of adding the Fisher information in order to regularize the discrete Hamiltonian, so as to avoid singularities when solving the initial value problem (2.4). (Here, and later, the word singularity refers to either the density becoming negative, or S blowing up, at a finite time.)

Example 2.1. Consider a 2-point graph G. Let $\rho_1(0), \rho_2(0) > 0$, let $S_1(0), S_2(0)$ be the corresponding initial values on the two nodes, and take the weights to be constant (e.g., take them to be 1). By choosing $\theta_{12}(\rho) = \theta_{21}(\rho) = \frac{\rho_1+\rho_2}{2} = \frac{1}{2}$, (2.4) becomes

(2.7)

$$\begin{aligned} \frac{d\rho_1(t)}{dt} &= -\frac{1}{2}(S_2(t)-S_1(t)), \quad \frac{d\rho_2(t)}{dt} = -\frac{1}{2}(S_1(t)-S_2(t)), \\ \frac{dS_1(t)}{dt} &= -\frac{1}{4}|S_2(t)-S_1(t)|^2 - \rho_1(t), \quad \frac{dS_2(t)}{dt} = -\frac{1}{4}|S_1(t)-S_2(t)|^2 - \rho_2(t). \end{aligned}$$

Combining the above equations and using $\rho_1(t) + \rho_2(t) = 1$, we get

$$\frac{d(\rho_1(t) - \rho_2(t))}{dt} = -(S_2(t) - S_1(t)),$$
$$\frac{d(S_1(t) - S_2(t))}{dt} = \rho_2(t) - \rho_1(t).$$

The solution of this system may may give negative values for ρ_1 and ρ_2 at a finite time. Indeed, we have that $\rho_1(t) - \rho_2(t) = (\rho_1(0) - \rho_2(0))\cos(t) + (S_1(0) - S_2(0))\sin(t)$. Then we obtain explicit expressions for ρ_1 and ρ_2 as

$$\rho_1(t) = \frac{1}{2} + \frac{1}{2}\cos(t)(\rho_1(0) - \rho_2(0)) + \frac{1}{2}\sin(t)(S_1(0) - S_2(0)),$$
$$\rho_2(t) = \frac{1}{2} + \frac{1}{2}\cos(t)(\rho_2(0) - \rho_1(0)) + \frac{1}{2}\sin(t)(S_2(0) - S_1(0)).$$

It is clear that either ρ_1 or ρ_2 can exit the boundary of $\mathcal{P}(G)$, i,e, $\rho_1 < 0$ or $\rho_2 < 0$, at a finite time if the initial condition satisfies $|S_1(0) - S_2(0)| > 1$. When taking $S_1(0) = S_2(0)$, the solution can be given in the following cases,

$$\rho_1(t) = \rho_2(t) = \frac{1}{2}, \text{ if } \rho_1(0) = \rho_2(0),$$
$$\rho_1(t) > 0, \rho_2(t) > 0, \text{ if } |\rho_1(0) - \rho_2(0)| < 1,$$
$$\rho_1(n\pi) = 0, \text{ or } \rho_2(n\pi) = 0, \text{ if } |\rho_1(0) - \rho_2(0)| = 1.$$

From this example, we see that a smooth potential in (2.4) on $\mathcal{P}(G)$ can lead to the development of a singularity of (ρ, S) at a finite time. Let T^* be the first time when a singularity occurs, if any. That is, if it is finite, $T^* = T^*_{\rho^0,S^0} > 0$ is the first time such that

$$\lim_{t\to T^*} \rho_i(t) \le 0, \text{ or } \lim_{t\to T^*} S_i(t) = \infty$$

for some index i. In the rest of this article, we mainly consider the initial value problem of (2.4) before T^*, since several properties of (2.4), including the local existence of a unique solution, can be established.

Proposition 2.1. *For any $\mu^0 \in \mathcal{P}_o(G)$, S^0 on V and $\beta \ge 0$, there exists $T^* > 0$ such that* (2.4) *has a unique solution and it satisfies properties* (i)–(vi) *below.*

(i) *Mass is conserved, i.e.: before time T^*,*

$$\sum_{i=1}^N \rho_i(t) = \sum_{i=1}^N \mu_i^0 = 1.$$

(ii) *Energy is conserved, i.e.: before time T^*,*

$$\mathcal{H}(\rho(t), S(t)) = \mathcal{H}(\mu^0, S^0).$$

(iii) *The solution is time reversible in the following sense: if $(\rho(t), S(t))$ is the solution of* (2.4)*, then $(\rho(-t), -S(-t))$ also solves* (2.4)*.*

(iv) *The solution is time transverse invariant with respect to the linear potential: if* $\mathbb{V}^\alpha = \mathbb{V} + \alpha$*, then* $S^\alpha = S + \alpha t$ *is the solution of* (2.4) *with potential* $\mathbb{V}^\alpha$*.*

(v) *If a time invariant* $\rho^* \in \mathcal{P}_o(G)$ *and* $S^*(t) = -\nu t$ *form an interior stationary solution of* (2.4)*, then* ρ^* *is the critical point of*

$$\rho^* \in \arg\min_{\rho \in \mathcal{P}_o(G)} \mathcal{H}(\rho, S) \quad \textit{and} \quad \nu = \mathcal{H}(\rho^*) + \frac{1}{2}\sum_{i,j=1}^N \mathbb{W}_{ij}\rho_i\rho_j,$$

where $\mathcal{H}(\rho^*) = \mathcal{H}(\rho^*, -\nu t)$ *is independent of the time* t*.*

(vi) *Assuming that* $\beta > 0$ *and* $\widetilde{\theta}_{ij}(\rho) = 0$ *only if* $\rho_i = \rho_j = 0$*, then there exists a compact set* $B \subset \mathcal{P}_o(G)$ *such that* $\rho(t) \in B$ *for all* $t \in [0, +\infty)$*.*

Proof. The proof of properties (i)–(v) is the same (except for the use of θ_{ij} instead of $\widetilde{\theta}_{ij}$) as that of [12, Theorem 6], thus we omit it. We prove (vi). Since the coefficient of (2.4) is locally Lipschitz and $\rho(0) = \mu^0 \in \mathcal{P}_o(G)$, it is not difficult to obtain the local existence of a unique solution $(\rho(t), S(t))$ in $[0, T^*)$, where $T^* > 0$ is the largest time for which $(\rho(t), S(t))$ exists and $\rho(t) \in \mathcal{P}_o(G)$. Thus, it suffices to show that the local solution can be extended to $T^* = \infty$, i.e., to show that the boundary is a repeller for $\rho(t)$. More precisely, we show that for any given $\mu^0 \in \mathcal{P}_o(G)$, there exists a compact set $B \subset \mathcal{P}_o(G)$ such that $T^* = +\infty$ and $\rho(t) \in B$.

We define B as follows:

$$B = \Big\{\rho \in \mathcal{P}_o(G) \mid \beta I(\rho) \le \mathcal{H}(\mu^0, S^0) - \min_{\rho \in \mathcal{P}(G)} [\mathcal{V}(\rho) + \mathcal{W}(\rho)]\Big\}. \tag{2.8}$$

Since $\mathcal{H}(\mu^0, S^0) < \infty$, the set B is nonempty. Due to the fact that $\mathcal{P}(G)$ is a compact subset of $\mathbb{R}^N$, to show that B is a compact set, it suffices to prove that $I(\rho)$ is positive infinity on the boundary of $\mathcal{P}(G)$. Assume that this is not true. It means that there exists a constant $M > 0$, such that $\beta I(\rho) \le M$ if $\min_{i=1,\dots,N} \rho_i = 0$. This implies that

$$M \ge \frac{\beta}{2}\sum_{i=1}^N \sum_{j \in N(i)} \widetilde{\omega}_{ij}(\log(\rho_i) - \log(\rho_j))^2 \widetilde{\theta}_{ij}(\rho).$$

Consequently, if we have that $\rho_i = 0$ for some i and that for $j \in N(i)$,

$$\beta\widetilde{\omega}_{ij}(\log(\rho_i) - \log(\rho_j))^2 \widetilde{\theta}_{ij}(\rho) \le 2M,$$

which yields $\widetilde{\theta}_{ij}(\rho) = 0$ for any $j \in N(i)$. Since G is connected and V is a finite set, we get that $\min_{i=1,\dots,N} \rho_i = 0$, which leads to a contradiction. Therefore, B is a compact set.

Finally, we show that $T^* = +\infty$ and $\rho(t) \in B$ for $t \in [0, +\infty)$. The energy conservation (ii) yields $\mathcal{H}(\rho(t), S(t)) = \mathcal{H}(\mu^0, S^0)$. Denoting $M =$

$\mathcal{H}(\mu^0, S^0) - \min_{\rho\in\mathcal{P}(G)}[\mathcal{V}(\rho) + \mathcal{W}(\rho)]$, we have

$$\begin{aligned}\beta I(\rho(t)) &= \mathcal{H}(\rho(t), S(t)) - \frac{1}{2}\langle \nabla_G S(t), \nabla_G S(t)\rangle_{\theta(\rho(t))} - (\mathcal{V}(\rho(t)) + \mathcal{W}(\rho(t)))\\ &\le \mathcal{H}(\mu^0, S^0) - \min_{\rho\in\mathcal{P}(G)}[\mathcal{V}(\rho) + \mathcal{W}(\rho)] = M.\end{aligned}$$

Thus $\rho(t) \in B \subset \mathcal{P}_o(G)$ for $t \in [0, +\infty)$. □

From Property (vi) in Proposition 2.1, it is clear that the Fisher information term helps maintain positivity of the density function in the Wasserstein–Hamiltonian flow. This fact motivated us to regularize the discretized Wasserstein–Hamiltonian system (2.4) by adding the Fisher information, and the details are discussed in Section 4.2.

There are many possible choices for θ_{ij} and $\widetilde{\theta}_{ij}$, as long as we require that $\widetilde{\theta}_{ij}(\rho) = 0$ only if $\rho_i = \rho_j = 0$. Indeed, this property is needed in order to get the lower bound estimate on the density in Section 3. For instance, the probability weight can be chosen as one of the following three options

$$\theta_{ij}(\rho) = \begin{cases} \theta^U_{ij}(\rho) = \rho_i, \text{ if } S_j > S_i, & (2.9)\\ \theta^A_{ij}(\rho) = \dfrac{\rho_i+\rho_j}{2} & (2.10)\\ \theta^L_{ij}(\rho) = \dfrac{\rho_i-\rho_j}{\log(\rho_i)-\log(\rho_j)}. & (2.11)\end{cases}$$

Here θ^U_{ij} is called the upwind weight (see e.g. [14] for more details). When the graph is a lattice, θ^U_{ij} corresponds to an upwind numerical scheme for hyperbolic PDEs. Recently, it has also been applied to study the stochastic Hamiltonian process on a finite graph, [18]. θ^A_{ij} is simply the average weight (see e.g. [12]), and θ^L_{ij} is called the logarithmic weight (see e.g. [36]), both of which have been used in the study of the continuous time Markov chain as gradient flow of the discrete entropy.

Remark 2.2. The above results hold even when G is not connected, in the following sense. Consider the decomposition of G into disjoint connected components, i.e. $G = \cup_{j=1}^l G_j$. Then on each subgraph $G_j = (V^j, E^j, \Omega^j)$, the properties (i)–(vi) in Proposition 2.1 also hold, with

$$\sum_{a_i\in V^j}\rho_i(t) = \sum_{a_i\in V^j}\mu_i^0.$$

2.2. **Spatial consistency for Wasserstein–Hamiltonian flows.** In this subsection, we show that (2.4) is a spatially consistent discretization for the following system of PDEs on a domain $\mathcal{M} \subset \mathbb{R}^d$:

$$
\begin{aligned}
&\frac{\partial}{\partial t}\rho(t,x) + \nabla\cdot(\rho(t,x)\nabla S(t,x)) = 0,\\
&\frac{\partial}{\partial t}S(t,x) + \frac{1}{2}|\nabla S(t,x)|^2 + \beta\frac{\delta}{\delta\rho(t,x)}\mathcal{I}(\rho(t,\cdot)) + \mathbb{V}(x)\\
&\qquad + \int_{\mathcal{M}}\mathbb{W}(x,y)\rho(t,y)dy = 0.
\end{aligned}
\tag{2.12}
$$

Next, introduce the Hamiltonian

$$\widetilde{H}(\rho,S) = \int_{\mathcal{M}}\frac{1}{2}|\nabla S(x)|^2\rho(x)dx + \mathcal{F}(\rho),$$

where

$$\mathcal{F}(\rho) = \int_{\mathcal{M}}\mathbb{V}(x)\rho(x)dx + \frac{1}{2}\int_{\mathcal{M}}\int_{\mathcal{M}}\mathbb{W}(x,y)\rho(x)\rho(y)dxdy + \beta\mathcal{I}(\rho),$$

and

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

Then, Equation (2.12) can be rewritten as a Wasserstein–Hamiltonian flow, that is,

$$
\begin{aligned}
\frac{\partial}{\partial t}\rho(t,x) &= \frac{\delta}{\delta S(t,x)}\widetilde{H}\big(\rho(t,\cdot),S(t,\cdot)\big),\\
\frac{\partial}{\partial t}S(t,x) &= -\frac{\delta}{\delta\rho(t,x)}\widetilde{H}\big(\rho(t,\cdot),S(t,\cdot)\big).
\end{aligned}
$$

Here $\frac{\delta}{\delta\rho}$, $\frac{\delta}{\delta S}$ are functional derivatives (see e.g. [24]). We would like to mention that in some cases there is a time $T^* > 0$ such that a unique smooth solution (ρ, S) of (2.12) exists for $t \in [0, T^*)$. For example, for the geodesic equation ($\beta = 0$), the local existence of a unique smooth solution can be obtained as follows. By the characteristic line method (see e.g. [29, section 2.2]), one can get the classical solution $S \in C^\infty([0,T^*);\mathbb{R}^d)$ for the Hamilton–Jacobi equation; then, using the fact that the continuity equation is linear and smoothness of S, one can also get $\rho \in C^\infty([0,T^*);\mathbb{R}^d)$. For the local existence result of a smooth solution for the Schrödinger equation, $\beta > 0$, we refer to [47] and references therein. Finally, for more properties of Wasserstein–Hamiltonian flow, we refer to [1, 6, 7, 12]. The Hamiltonian structure is also helpful for studying time periodic solutions and the long time behavior (e.g., see [15] in potential mean field games).

To illustrate the spatial consistency of (2.4), we consider a lattice graph G which is a Cartesian product of d 1-dimensional lattices: $G = G_1 \times \cdots \times G_d$ with $G_k = (V_k, E_k)$, $k = 1, \dots, d$. For simplicity, let us assume that $\mathcal{M} = \mathbb{R}^d$, and there is no interaction potential ($\mathbb{W} = 0$) in both (2.4) and (2.12).

We denote h the mesh size of the lattice, $\omega_{ij} = \frac{1}{h^2}$ for $(i,j) \in E$, $T > 0$ is a constant representing the length of the time interval. We also let $i = (i_1, i_2, \cdots, i_d)$ represents a point $x(i)$ in $\mathbb{R}^d$. It can be seen that its neighboring set $N(i) = \cup_{k=1}^d N_k(i)$ with

$$N_k(i) = \{(i_1, \cdots, i_{k-1}, j_k, i_{k+1}, \cdots, i_d) \ : \ (i_k, j_k) \in E_k\}.$$

We call (2.4) a consistent numerical scheme for (2.12) (equivalently, a consistent discretization of (2.12)) if for any smooth functions

$$(\rho, S) \in \mathcal{C}^\infty([0,T]; (0,1)^d) \times \mathcal{C}^\infty([0,T]; \mathbb{R}^d),$$

it holds that for any $t \in [0,T]$ and $x \in \mathbb{R}^d$,

$$\begin{aligned}
&\frac{\partial}{\partial t}\rho(t,x) + \nabla \cdot (\rho(t,x)\nabla S(t,x)) - \frac{d\rho_i(t)}{dt} \\
&\qquad - \sum_{j \in N(i)} \omega_{ij}(S_j(t) - S_i(t))\theta_{ij}(\rho(t)) \to 0, \\
&\frac{\partial}{\partial t}S(t,x) + \frac{1}{2}|\nabla S(t,x)|^2 + \beta\frac{\delta}{\delta\rho(t,x)}\mathcal{I}(\rho(t,\cdot)) + \mathbb{V}(x) \\
&- \frac{dS_i(t)}{dt} - \frac{1}{2}\sum_{j \in N(i)} \omega_{ij}(S_i(t) - S_j(t))^2 \frac{\partial\theta_{ij}(\rho(t))}{\partial\rho_i} - \beta\frac{\partial I(\rho(t))}{\partial\rho_i} - \mathbb{V}_i \to 0,
\end{aligned}$$

as $h \to 0$ and $x(i) \to x$. Here we use the notation $\rho_j(t) = \rho(t, x(j))$, $S_j(t) = S(t, x(j))$, $\rho(t) = (\rho_j(t))_{j \in V}$, $S(t) = (S_j(t))_{j \in V}$, and $\mathbb{V}_j = \mathbb{V}(x(j))$.

For the probability weights $\theta_{ij}(\rho)$ and $\widetilde{\theta}_{ij}(\rho)$ in (2.4), we assume that

$$\theta_{ij}(\rho) = \Theta(\rho_i, \rho_j), \quad \widetilde{\theta}_{ij}(\rho) = \widetilde{\Theta}(\rho_i, \rho_j),$$

where Θ and $\widetilde{\Theta}$ are symmetric and belong to $\mathcal{C}^2((0,1) \times (0,1)) \cap \mathcal{C}([0,1] \times [0,1])$. In order to show the spatial consistency of (2.4), we further assume that

$$\frac{\partial\Theta(x,x)}{\partial x} = \frac{1}{2}, \ \Theta(x,x) = x. \tag{2.13}$$

As customary, to show consistency of the scheme in Proposition 2.2, we will tacitly assume that ρ and S are sufficiently smooth.

Proposition 2.2. *The system* (2.4) *with* θ *and* $\widetilde{\theta}$ *being selected to satisfy* (2.13) *is a consistent finite difference scheme for* (2.12).

Proof. Let $e_1, \ldots, e_d$, be the standard unit vectors. The lattice graph in the e_k direction contains two points near i, i.e., $x(i) + e_k h$ and $x(i) - e_k h$, which we label i^+ and i^- for short. By Taylor expansion at i in every e_k direction, we obtain

$$\begin{aligned}
&-\sum_{j\in N(i)} \omega_{ij}(S_j(t) - S_i(t))\theta_{ij}(\rho(t)) \\
&= \sum_k \frac{1}{h^2}\big(S_i(t) - S_{i^+}(t)\big)\theta_{ii^+}(\rho(t)) + \sum_k \frac{1}{h^2}\big(S_i(t) - S_{i^-}(t)\big)\theta_{ii^-}(\rho(t)) \\
&= \sum_k \frac{1}{h^2}\Big(-\frac{\partial S}{\partial x_k}(t,x(i))h - \frac{1}{2}\frac{\partial^2 S}{\partial x_k^2}(t,x(i))h^2 + \mathcal{O}(h^3)\Big) \\
&\qquad \times \Big(\theta_{ii}(\rho(t)) + \frac{\partial \theta_{ii}(\rho(t))}{\partial \rho_i}\frac{\partial \rho_i(t)}{\partial x_k}h + \mathcal{O}(h^2)\Big) \\
&+ \sum_k \frac{1}{h^2}\Big(\frac{\partial S}{\partial x_k}(t,x(i))h - \frac{1}{2}\frac{\partial^2 S}{\partial x_k^2}(t,x(i))h^2 + \mathcal{O}(h^3)\Big) \\
&\qquad \times \Big(\theta_{ii}(\rho(t)) - \frac{\partial \theta_{ii}(\rho(t))}{\partial \rho_i(t)}\frac{\partial \rho_i(t)}{\partial x_k}h + \mathcal{O}(h^2)\Big) \\
&= -\sum_k \Big(\frac{\partial^2 S}{\partial x_k^2}(t,x(i))\theta_{ii}(\rho) + 2\frac{\partial S}{\partial x_k}(t,x(i))\frac{\partial \theta_{ii}(\rho(t))}{\partial \rho_i}\frac{\partial \rho_i(t)}{\partial x_k}\Big) + \mathcal{O}(h),
\end{aligned}$$

where the summation is over $k = 1, \cdots, d$. Similarly,

$$\begin{aligned}
&-\frac{1}{2}\sum_{j\in N(i)} \omega_{ij}(S_i(t) - S_j(t))^2 \frac{\partial \theta_{ij}(\rho(t))}{\partial \rho_i} \\
&= -\frac{1}{2}\sum_k \frac{1}{h^2}\big(S_{i^+}(t) - S_i(t)\big)^2 \frac{\partial \theta_{ii^+}(\rho(t))}{\partial \rho_i} \\
&- \frac{1}{2}\sum_k \frac{1}{h^2}\big(S_{i^-}(t) - S_i(t)\big)^2 \frac{\partial \theta_{ii^-}(\rho(t))}{\partial \rho_i} \\
&= -\frac{1}{h^2}\sum_k \Big(\frac{\partial S}{\partial x_k}(t,x(i))h + \mathcal{O}(h^2)\Big)^2 \Big(\frac{\partial \theta_{ii}(\rho(t))}{\partial \rho_i} + \mathcal{O}(h)\Big)
\end{aligned}$$

and

$$
\begin{aligned}
-\beta\frac{\partial I(\rho(t))}{\partial \rho_i} &= -\beta\sum_k \frac{1}{h^2}|\log(\rho_{i^+}(t)) - \log(\rho_i(t))|^2\frac{\partial\widetilde{\theta}_{ii^+}(\rho(t))}{\partial\rho_i}\\
&\quad -\beta\sum_k \frac{1}{h^2}|\log(\rho_{i^-}(t)) - \log(\rho_i(t))|^2\frac{\partial\widetilde{\theta}_{ii^-}(\rho(t))}{\partial\rho_i}\\
&\quad -2\beta\sum_k \frac{1}{h^2}\Big(\log(\rho_i(t)) - \log(\rho_{i^+}(t))\Big)\frac{1}{\rho_i(t)}\theta_{ii^+}(\rho(t))\\
&\quad -2\beta\sum_k \frac{1}{h^2}\Big(\log(\rho_i(t)) - \log(\rho_{i^-}(t))\Big)\frac{1}{\rho_i(t)}\theta_{ii^-}(\rho(t))\\
&= -\sum_k\Big(2\beta\sum_k|\frac{\partial\log(\rho_i(t))}{\partial x_k}|^2\frac{\partial\widetilde{\theta}_{ii}(\rho(t))}{\partial\rho_i}\Big)\\
&\quad +2\beta\sum_k\frac{1}{\rho_i(t)}\frac{\partial^2}{\partial x_k^2}\rho_i(t) + \mathcal{O}(h).
\end{aligned}
$$

Thus, if $\frac{\partial\theta_{ii}(\rho)}{\partial\rho_i} = \frac{\partial\widetilde{\theta}_{ii}(\rho)}{\partial\rho_i} = \frac{1}{2}, \widetilde{\theta}_{ii}(\rho) = \theta_{ii}(\rho) = \rho_i$, we have

$$
\begin{aligned}
&\frac{d\rho(t,x(i))}{dt} + \sum_{j\in N(i)}\omega_{ij}(S_j(t) - S_i(t))\theta_{ij}(\rho(t))\\
&= \frac{\partial\rho(t,x(i))}{\partial t} + \nabla\cdot\big(\nabla S(t,x(i))\rho(t,x(i))\big) + \mathcal{O}(h),\\
&\frac{dS(t,x(i))}{dt} + \frac{1}{2}\sum_{j\in N(i)}\omega_{ij}(S_i(t) - S_j(t))^2\frac{\partial\theta_{ij}(\rho(t))}{\partial\rho_i} + \beta\frac{\partial I(\rho(t))}{\partial\rho_i} + \mathbb{V}(x(i))\\
&= \frac{\partial S(t,x(i))}{\partial t} + \frac{1}{2}|\nabla S(t,x(i))|^2 + \beta\frac{\delta I}{\delta\rho_i}(\rho(t,\cdot)) + \mathbb{V}(x(i)) + \mathcal{O}(h).
\end{aligned}
$$

Taking the Taylor expansion of $\rho(t,x)$ and $S(t,x)$ at $x(i)$ completes the proof. ☐

As we show next, even if Θ and $\widetilde{\Theta}$ are not sufficiently regular, spatial consistency still holds as long as (2.13) holds. For example, one can take θ as the upwind weight, $\theta^U_{ij}(\rho) = \rho_i$, if $S_j > S_i$, $\widetilde{\theta}$ satisfies (2.13) and $\widetilde{\Theta} \in \mathcal{C}^2((0,1)\times(0,1))\cap\mathcal{C}([0,1]\times[0,1])$ is symmetric.

Proposition 2.3. *Assume that $\theta = \theta^U$, and that $\widetilde{\theta}$ satisfies* (2.13). *Then* (2.4) *with θ^U and $\widetilde{\theta}$ is a consistent spatial discretization of* (2.12).

Proof. We use the same notation as in the proof of Proposition 2.2. For simplicity, we assume that $S(t,x(i)+e_kh) \le S(t,x(i)) \le S(t,x(i)-e_kh)$. Similarly, we can show the same results for other possible configurations.

By Taylor expansion, we obtain

$$\begin{aligned}
&-\sum_{j\in N(i)}\omega_{ij}(S_j(t)-S_i(t))\theta_{ij}(\rho(t))\\
&\sum_k \frac{1}{h^2}(S_i(t)-S_{i^+}(t))\theta_{ii^+(t)}(\rho)+\sum_k \frac{1}{h^2}(S_i(t)-S_{i^-}(t))\theta_{ii^-(t)}(\rho)\\
&=\sum_k \frac{1}{h^2}(S(t,x(i))-S(t,x(i)+e_kh))\rho_{i^+}(t)\\
&+\sum_k \frac{1}{h^2}(S(t,x(i))-S(t,x(i)-e_kh))\rho_i(t)\\
&=\sum_k \frac{1}{h^2}(-\frac{\partial S}{\partial x_k}(t,x(i))h-\frac{1}{2}\frac{\partial^2 S}{\partial x_k^2}(t,x(i))h^2+\mathcal{O}(h^3))\rho_{i^+}(t)\\
&+\sum_k \frac{1}{h^2}(\frac{\partial S}{\partial x_k}(t,x(i))h-\frac{1}{2}\frac{\partial^2 S}{\partial x_k^2}(t,x(i))h^2+\mathcal{O}(h^3))\rho_i(t)\\
&=-\nabla\cdot(\rho(t,x(i))\nabla S(t,x(i)))+\mathcal{O}(h)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{2}\sum_{j\in N(i)}\omega_{ij}(S_i(t)-S_j(t))^2\frac{\partial\theta_{ij}(\rho(t))}{\partial\rho_i}\\
&=\frac{1}{2}\sum_k\frac{1}{h^2}(S_{i^+}(t)-S_i(t))^2\frac{\partial\theta_{ii^+}(\rho(t))}{\partial\rho_i}+\frac{1}{2}\sum_k\frac{1}{h^2}(S_{i^-}(t)\\
&\qquad -S_i(t))^2\frac{\partial\theta_{ii^-}(\rho(t))}{\partial\rho_i}\\
&=\frac{1}{2}\sum_k\Big|\frac{\partial S}{\partial x_k}(t,x(i))\Big|^2+\mathcal{O}(h).
\end{aligned}$$

Based on the above estimates and the calculation of $-\beta\frac{\partial I(\rho(t))}{\partial\rho_i}$ in the proof of Proposition 2.2, we have that

$$\begin{aligned}
&\frac{d\rho(t,x(i))}{dt}+\sum_{j\in N(i)}\omega_{ij}(S_j(t)-S_i(t))\theta_{ij}(\rho(t))\\
&=\frac{\partial\rho(t,x(i))}{\partial t}+\nabla\cdot(\nabla S(t,x(i))\rho(t,x(i)))+\mathcal{O}(h),\\
&\frac{dS(t,x(i))}{dt}+\frac{1}{2}\sum_{j\in N(i)}\omega_{ij}(S_i(t)-S_j(t))^2\frac{\partial\theta_{ij}(\rho(t))}{\partial\rho_i}+\beta\frac{\partial I(\rho(t))}{\partial\rho_i}+\mathbb{V}(x(i))\\
&=\frac{\partial S(t,x(i))}{\partial t}+\frac{1}{2}\big|\nabla S(t,x(i))\big|^2+\beta\frac{\partial I}{\partial\rho(t,x(i))}(\rho(t,x(i)))+\mathbb{V}(x(i))+\mathcal{O}(h).
\end{aligned}$$

Taking the Taylor expansion of $\rho(t,x)$ and $S(t,x)$ at $x(i)$ completes the proof. □

Remark 2.3. Let us take $\beta = \frac{\hbar^2}{8} > 0$ in (2.4), where $\hbar$ is the Planck constant (see e.g. [37, 38]) . By introducing the discrete Madelung transformation

$$u(t) = (u_j(t))_{j=1}^N = \left(\sqrt{\rho_j(t)} e^{\mathbf{i}\frac{S_j(t)}{\hbar}}\right)_{j=1}^N, \tag{2.14}$$

(2.4) can be rewritten as

$$\hbar\mathbf{i}\frac{du_j(t)}{dt} = -\frac{\hbar^2}{2}(\Delta_G u(t))_j + u_j(t)\mathbb{V}_j + u_j(t)\sum_{l=1}^N \mathbb{W}_{jl}|u_l(t)|^2, \tag{2.15}$$

where the Laplacian on the graph (see e.g. [12]) is defined by

$$\begin{aligned}(\Delta_G u)_j = -u_j\Big(\frac{1}{|u_j|^2}\big[&\sum_{l\in N(j)} \omega_{jl}(Im\big(\log(u_j)\big) - Im\big(\log(u_l)\big))\theta_{jl} \\ &+ \sum_{l\in N(j)} \widetilde{\omega}_{jl}(Re\big(\log(u_j)\big) - Re\big(\log(u_l)\big))\widetilde{\theta}_{jl}\big] \\ &+ \sum_{l\in N(j)} \omega_{jl}\big|Im(\log(u_j) - \log(u_l))\big|^2\frac{\partial\theta_{jl}}{\partial\rho_j} \\ &+ \sum_{l\in N(j)} \widetilde{\omega}_{jl}\big|Re(\log(u_j) - \log(u_l))\big|^2\frac{\partial\widetilde{\theta}_{jl}}{\partial\rho_j}\Big).\end{aligned} \tag{2.16}$$

When the graph is a lattice graph of a bounded domain, (2.15) becomes a nonlinear spatial approximation of the nonlinear Schrödinger equation

$$\hbar\mathbf{i}\frac{\partial}{\partial t}u(t,x) = -\frac{\hbar^2}{2}u(t,x) + u(t,x)\mathbb{V}(x) + u(t,x)\int_{\mathbb{R}^d}\mathbb{W}(x,y)|u(t,y)|^2 dy.$$

This viewpoint may give advantages insofar as the preservation of desirable properties, like dispersion relationship, Hamiltonian structure and time reversibility.

Remark 2.4. Taking $\beta = 0$, $\mathbb{W} = 0$, the Wasserstein–Hamiltonian flow (2.4) also has a close relationship to the linear Vlasov equation

$$\frac{\partial}{\partial t}f(t,x,v) + v\cdot\nabla_x f(t,x,v) - \nabla\mathbb{V}(x)\cdot\nabla_v f(t,x,v) = 0, \tag{2.17}$$

which is the governing equation for the joint density of the particle system

$$\frac{d}{dt}X(t,x) = v(t,X(t,x)),\ \frac{d}{dt}v(t,X(t,x)) = -\nabla\mathbb{V}(X(t,x))$$

with random initial positions and vector field. More precisely, if the initial distribution $f(0,x,v)$ is $\rho(0,x)\delta_{v(0,x)}(v)$, the joint density $f(t,x,v)$ enjoys the form $\rho(t,x)\delta_{v(t,x)}(v)$ before T^* when the characteristic lines first intersect (see e.g. [29]). By [12, section 3], the joint distribution $f(t,x,v)$ $(t < T^*)$

can be characterized by the Wasserstein–Hamiltonian flow,

$$\begin{aligned}&\frac{\partial}{\partial t}\rho(t,x)+\nabla\cdot(\rho(t,x)\nabla S(t,x))=0,\\&\frac{\partial}{\partial t}S(t,x)+\frac{1}{2}|\nabla S(t,x)|^2=-\mathbb{V}(x)+C(t),\end{aligned}\tag{2.18}$$

where $C(t)$ is a constant function independent of x. Therefore, (2.4) can be viewed as a special discretization of (2.17) up to a constant function $C(t)$.

3. Lower bound estimate of the density

In this section, we give an explicit lower bound for the density function in (2.4) with the logarithmic weight $\widetilde{\theta}_{ij}=\theta^L_{ij}$ defined in (2.11). We take two basic finite graph structures to illustrate the derivation of the lower bound. We believe that similar techniques can also be used, with appropriate modifications, to obtain the lower bounds for more general graphs and different probability weights.

To introduce the type of graphs we consider we need Definition 3.1.

Definition 3.1. Given a graph $G=(V,E,\Omega)$, we call $a_i\in V$ a boundary node if it has only one edge connecting it to other nodes in V. We call boundary set the set of all boundary nodes, and denote it as ∂V.

3.1. Lower bound for graph with empty boundary set.

Here, our goal is to analyze the properties of the extreme point of the Fisher information (2.1) over $\mathcal{P}_o(G)$ on a graph $G=(V,E,\Omega)$ whose edge set satisfies

$$\begin{cases}(i,j)\in E, & \text{if} \quad j=i+1 \quad \text{or} \quad j=i-1,\\(1,j)\in E, & \text{if} \quad j=2 \quad \text{or} \quad j=N,\\(N,j)\in E, & \text{if} \quad j=N-1 \quad \text{or} \quad j=1.\end{cases}$$

To illustrate, an instance of the graph we consider is that resulting from discretizing a circle of circumference 1. The corresponding finite graph G is defined by the set of N equidistant points on the circle $V=\{a_i, i\le N\}$, $(i,j)\in E$ if and only if $j=i-1$ or $i+1$ (mod N), and $\omega_{ij}=\widetilde{\omega}_{ij}=\frac{1}{h^2}$, $h=\frac{1}{N}$. We take θ_{ij} in Section 2.2 satisfying (2.13). When $\widetilde{\theta}=\theta^L$, the Fisher information (2.1) in this case can be rewritten as

$$I(\rho)=\sum_{i=1}^N\widetilde{\omega}_{i,i+1}(\log(\rho_i)-\log(\rho_{i+1}))(\rho_i-\rho_{i+1})\ ,\ \ \rho_{N+1}=\rho_1\ .\tag{3.1}$$

Denote the tangent space at $\rho\in\mathcal{P}_o(G)$ by

$$T_\rho\mathcal{P}_o(G)=\Big\{(\sigma_i)_{i=1}^N\in\mathbb{R}^N\ \Big|\ \sum_{i=1}^N\sigma_i=0\Big\}.$$

Lemma 3.1. *The Fisher information in* (2.1) *is strictly convex on* $\mathcal{P}_o(G)$ *and achieves its unique minimum at the uniform distribution.*

Proof. The convexity of I can be obtained by directly calculating its Hessian matrix, $\mathrm{Hess}(I(\rho))$, and proving that

$$\min_{\sigma\in T_\rho\mathcal{P}_o(G)}\{\sigma^\top \mathrm{Hess}(I(\rho))\sigma|\sigma^\top\sigma=1\}>0.$$

Direct calculation yields that

$$\frac{\partial}{\partial\rho_i\partial\rho_j}I(\rho)=\begin{cases}\widetilde{\omega}_{i,i+1}\dfrac{1}{\rho_i^2}(\rho_i+\rho_{i+1})+\widetilde{\omega}_{i,i-1}\dfrac{1}{\rho_i^2}(\rho_i+\rho_{i-1}) & \text{for} \quad j=i;\\ -\widetilde{\omega}_{i,i+1}\dfrac{1}{\rho_i\rho_{i+1}}(\rho_i+\rho_{i+1}) & \text{for} \quad j=i+1;\\ -\widetilde{\omega}_{i,i-1}\dfrac{1}{\rho_i\rho_{i-1}}(\rho_i+\rho_{i-1}) & \text{for} \quad j=i-1;\\ 0 & \text{otherwise.}\end{cases}$$

Thus, we obtain

$$\begin{aligned}&\sigma^T\mathrm{Hess}I(\rho)\sigma\\&=\sum_{i=1}^N\Big(\widetilde{\omega}_{i,i+1}\frac{1}{\rho_i^2}\big(\rho_i+\rho_{i+1}\big)+\widetilde{\omega}_{i,i-1}\frac{1}{\rho_i^2}\big(\rho_i+\rho_{i-1}\big)\Big)\sigma_i^2\\&\quad+\sum_{i=1}^N\Big(\widetilde{\omega}_{i,i+1}\frac{1}{\rho_i\rho_{i+1}}(\rho_i+\rho_{i+1})\sigma_i\sigma_{i+1}+\widetilde{\omega}_{i,i-1}\frac{1}{\rho_i\rho_{i-1}}\big(\rho_i+\rho_{i-1}\big)\sigma_i\sigma_{i-1}\Big)\\&=\sum_{i=1}^N\widetilde{\omega}_{i,i+1}\big(\rho_i+\rho_{i+1}\big)\big(\frac{\sigma_i}{\rho_i}-\frac{\sigma_{i+1}}{\rho_{i+1}}\big)^2\ge 0,\end{aligned}$$

which implies that $\mathrm{Hess}(I(\rho))$ is positive semi-definite. To show strict convexity, assume that there exists a unit vector σ^* such that ${\sigma^*}^T\mathrm{Hess}I(\rho)\sigma^*=0$. Then we have $\frac{\sigma_1}{\rho_1}=\frac{\sigma_i}{\rho_i}$ for $i=2,\cdots,N$. Since $\sigma\in T_\rho\mathcal{P}_o(G)$, then

$$\sum_{i=1}^N\sigma_i=\sigma_1(1+\sum_{i=2}^N\frac{\rho_i}{\rho_1})=0.$$

As $\rho\in\mathcal{P}_o(G)$, we conclude that $\sigma_i=0$ for all i, which contradicts that ${\sigma^*}^\top\sigma^*=1$.

The strict convexity implies that there is a unique minimum point for the Fisher information on $\mathcal{P}_o(G)$.

We use the Lagrange multiplier technique to find the minimum of $I(\rho)$ under the constraint $\sum_{i=1}^N\rho_i=1$. We consider the critical point of

$$I(\rho)-\lambda\Big(\sum_{i=1}^N\rho_i-1\Big),$$

where $\lambda\in\mathbb{R}$.

Taking the first derivative with respect to ρ, we obtain that the extreme point satisfies

$$\widetilde{\omega}_{i,i+1}\phi(\frac{\rho_{i+1}}{\rho_i}) + \widetilde{\omega}_{i-1,i}\phi(\frac{\rho_{i-1}}{\rho_i}) = \lambda, \quad \text{for} \quad i \le N,$$

where $\phi(t) = 1 - t - \log(t), t \in (0,\infty)$. It can be verified that ϕ is strictly decreasing, convex, and $\phi(1) = 0$. Then when $\lambda = 0$, $\rho_i = \frac{1}{N}$, and the extreme point $\rho_i = \frac{1}{N}$ is the unique minimum point and gives $I(\rho) = 0$. ☐

Below we present the lower bound estimate of the density $\rho(t), t \in [0,+\infty)$ of (2.4) with $\beta > 0$ by studying the lower bound of the Fisher information. Our analysis is based on the following two facts. On one hand, by Proposition (2.1)-(vi) , $\rho(t) \in B, t \in [0,+\infty)$, where B is defined in (2.8). Denote

$$M = \mathcal{H}(\mu^0, S^0) - \min_{\rho\in\mathcal{P}(G)} [\mathcal{V}(\rho) + \mathcal{W}(\rho)].$$

It follows that $\sup_{t\ge0} I(\rho(t)) \ge \frac{M}{\beta}$. On the other hand, the continuity and convexity of I yields that for any $C > 0$, there exists $c < \frac{1}{N}$ such that

$$\inf_{0<\min_i(\rho_i)\le c} I(\rho(t)) \ge C.$$

Thus, if we could find c such that $I(\rho) \ge \frac{M}{\beta}$ over the subset

$$\Big\{\rho \in \mathcal{P}_o(G) \;\Big|\; \min_{i\le N} \rho_i \le c\Big\},$$

the explicit lower bound can be obtained.

The following result gives the anticipated lower bound, and its proof is given in the Appendix, where we assume that $\widetilde{\omega}_{i,i+1} = 1$ for simplicity. We denote with $[y]$ the integer part of $y \in \mathbb{R}$.

Proposition 3.1. *Let* $\min_{i\le N} \mu_i^0 < \frac{1}{N}$. *Then it holds that*

$$\sup_{t\ge0} \min_{i\le N} \rho_i(t) \ge \min\big(\kappa_0, \kappa_1\big),$$

where

$$\kappa_0 = \frac{1}{2}\min_{i\le N} \mu_i^0 \quad \text{and} \quad \kappa_1 = \Big(1 + N\exp\big(\frac{M(N-1)([\frac{N-1}{2}]+1)}{\beta}\big)\Big)^{-1}.$$

Proof. See the Appendix. ☐

3.2. **Lower bound for graph with nonempty boundary set.** Again, our goal is to analyze the extreme point(s) of the Fisher information (2.1), but now the graph $G = (V, E, \Omega)$ has $\partial V \ne \emptyset$.

To illustrate the graph structure, consider that resulting from discretization of $[0, 1]$. V has nodes $a_i = (i-1)h$, $i = 1,\dots,N$, $h = \frac{1}{N}$, $\omega_{ij} = \widetilde{\omega}_{ij} = \frac{1}{h^2}$ and $(i,j) \in E$ if and only if $j = i-1$ or $i+1$, for $i = 2,\dots,N-1$. In this example, the boundary nodes of G are a_1 and a_N. We take θ_{ij} satisfying

(2.13), and when $\widetilde{\theta} = \theta^L$, the Fisher information (2.1) in the present case is rewritten as

$$I(\rho) = \sum_{i=1}^{N-1} \widetilde{\omega}_{i,i+1}(\log(\rho_i) - \log(\rho_{i+1}))(\rho_i - \rho_{i+1}), \tag{3.2}$$

which corresponds to Neumann boundary conditions ($\frac{\partial \rho}{\partial x}(0) = \frac{\partial \rho}{\partial x}(1) = 0$).

Similarly to Lemma 3.1, and with much the same proof, we have strict convexity of $I(\rho)$.

Lemma 3.2. *The Fisher information in* (2.1) *is strictly convex on* $\mathcal{P}_o(G)$ *and achieves its unique minimum at the uniform distribution.*

To obtain a lower bound on the density, we let $\kappa \leq N-1$ be the number of nodes in ∂V, and d_{max} be the largest distance[1] between two nodes in ∂V.

The proof of the following lower bound estimate is given in the Appendix, where for simplicity we assume that $\widetilde{\omega}_{ii+1} = 1$.

Proposition 3.2. *Let* $\min_{i\leq N}(\mu_i^0) < \frac{1}{N}$. *Then it holds that*

$$\sup_{t\geq 0} \min_{i\leq N} \rho_i(t) \geq \min\big(\kappa_0, \kappa_1\big),$$

where

$$\kappa_0 = \frac{1}{2}\min_{i\leq N} \mu_i^0 \quad \text{and} \quad \kappa_1 = \Big(1 + \kappa(d_{max}-1)\exp\big(2\frac{M(d_{max}-1)(N-1)}{\beta}\big)\Big)^{-1}.$$

Proof. See the Appendix. ☐

4. Time discretization of Wasserstein–Hamiltonian systems on graph

In this section, we construct and study the full discretization of Wasserstein–Hamiltonian systems. In particular, we discuss the time discretization of the (regularized) spatial discretizations (2.4) and (2.12). Our main goal is to devise a symplectic discretization of the Wasserstein–Hamiltonian flow (2.4) with $\beta > 0$. We also discuss general regularization strategies for (2.12).

To illustrate our strategy, we need to introduce an equivalent formulation of the Lagrangian functional (2.2), the symplectic structure, as well as some preliminary estimates for the Wasserstein–Hamiltonian flows with $\beta > 0$ on graph.

First, we introduce an equivalent formulation of (2.2) to derive (2.4) on graph, that is, we seek the critical point of

$$\mathcal{L}(\rho, \nabla_G S) = \int_0^T L(\rho(t), \nabla_G S(t))dt$$

[1] The distance d_{ij} between two nodes a_i and a_j is the smallest number of edges connecting a_i and a_j.

subject to the constraint

$$\frac{d}{dt}\rho(t) = -div_G^\theta(\rho(t)\nabla_G S(t)),$$

with the Langrangian functional defined by

$$L(\rho(t), \nabla_G S(t)) = \frac{1}{2}\langle \nabla S(t), \nabla S(t)\rangle_{\theta(\rho(t))} - \mathcal{V}(\rho(t)) - \mathcal{W}(\rho(t)) - \beta I(\rho(t)),$$

where $\rho(t) \in \mathcal{P}_o(G), t \in [0,T]$, $S(t) \in L^2([0,T], \mathbb{R}^N)$. Let us denote $(-\Delta_\rho^\theta)^\dagger$ as the pseudo inverse of the operator $\Delta_\rho^\theta \cdot = div_G^\theta(\rho\nabla_G \cdot)$ as that defined in [12]. This pseudo inverse operator, together with the constraint, induces the Legendre transformation $S(t) = (-\Delta^\theta_{\rho(t)})^\dagger \dot\rho(t)$ between the primal coordinate $(\rho(t), \dot\rho(t))$ with $\dot\rho(t) = \frac{d}{dt}\rho(t)$ and the dual coordinate $(\rho(t), S(t))$. By using the integration by part formula and $S(t) = (-\Delta^\theta_{\rho(t)})^\dagger \dot\rho(t)$, the functional $L(\rho, \nabla_G S)$ can also be written as $L(\rho, \dot\rho)$ in the primal coordinates as follows,

$$\begin{aligned}
L(\rho(t), \dot\rho(t)) &= \frac{1}{2}\langle \nabla_G S(t), \nabla_G S(t)\rangle_{\theta_{\rho(t)}} - \mathcal{F}(\rho(t)) \\
&= \frac{1}{2}\langle S(t), \Delta^\theta_{\rho(t)} S\rangle - \mathcal{F}(\rho(t)) \\
&= \frac{1}{2}\langle \nabla_G((-\Delta^\theta_{\rho(t)})^\dagger \dot\rho(t)), \nabla_G((-\Delta^\theta_{\rho(t)})^\dagger \dot\rho(t))\rangle_{\theta_{\rho(t)}} - \mathcal{F}(\rho(t)) \\
&= \frac{1}{2}\langle (-\Delta^\theta_{\rho(t)})^\dagger \dot\rho(t)), (-\Delta^\theta_{\rho(t)})(-\Delta^\theta_{\rho(t)})^\dagger \dot\rho(t)\rangle - \mathcal{F}(\rho(t)),
\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the l^2-inner product and $\mathcal{F}(\rho(t)) = \beta I(\rho(t)) + \mathcal{V}(\rho(t)) + \mathcal{W}(\rho(t)), \beta > 0$.

Recall that the critical point (2.4) is a Hamiltonian system on graph since it has the form (2.5). Consequently, its phase flow preserves the symplectic structure

$$d\rho(t) \wedge dS(t) = d\rho(0) \wedge dS(0), \tag{4.1}$$

i.e., the sum of oriented areas of projections onto the coordinate planes $(\rho_1, S_1), \cdots, (\rho_N, S_N)$ is an integral invariant (see e.g. [26]).

At last, we present some estimates of the coefficients and their derivatives of (2.4). We assume that

$$\begin{aligned}
&c_0 \le \omega_{ij} \le C_0, c_0 \le \widetilde{\omega}_{ij} \le C_0, \text{ for some } c_0, C_0 > 0, \\
&M_0 = \max_{i \le N} \mathbb{V}_i + \max_{(i,j)\in E} \mathbb{W}_{ij}, \ \mathcal{H}_0 = \mathcal{H}(\mu^0, S^0).
\end{aligned}$$

For simplicity, we also restrict our consideration to $\theta_{ij} = \theta^A_{ij}$ in (2.10), and $\widetilde{\theta}_{ij} = \theta^L_{ij}$ in (2.11). Denote the maximum numbers of edges connecting to a node by E_{max}. Let c be the uniform lower bound of ρ derived in Section 3.

We first give a uniform upper bound estimate of $|S_i(t)-S_j(t)|$ and $|\log(\rho_i(t))-\log(\rho_j(t))|, t\in[0,T]$ as follows. Recall that $\mathcal{H}(\rho,S)=K(\rho,S)+\mathcal{F}(\rho)$ in (2.6). Due to the conservation of $\mathcal{H}$ in (2.4), we have that

$$\begin{aligned}
K(S(\rho(t)),\rho(t))+\beta I(\rho(t)) &= \frac{1}{4}\sum_{i=1}^N\sum_{j\in N(i)}\omega_{ij}|S_i(t)-S_j(t)|^2\theta_{ij}(\rho(t))\\
&\quad+\beta\frac{1}{2}\sum_{i,j=1}^N\widetilde{\omega}_{ij}(\log(\rho_i(t))-\log(\rho_j(t)))^2\widetilde{\theta}_{ij}(\rho(t))\\
&\le \mathcal{H}_0-\min_{\rho\in\mathcal{P}(G)}\Big[\mathcal{V}(\rho)+\mathcal{W}(\rho)\Big]=M.
\end{aligned}$$

Then we get

$$\begin{aligned}
\max_{i\le N}|S_i-S_j|^2 &\le \frac{2M}{c_0\min_{(i,j)\in E}\theta_{ij}(\rho(t))}\le\frac{2M}{c_0c},\\
\max_{i\le N}|\log(\rho_i)-\log(\rho_j)|^2 &\le \frac{M}{c_0\min_{(i,j)\in E}\theta_{ij}(\rho(t))}\le\frac{M}{c_0c},
\end{aligned}$$

where c is the lower bound of the denisty function of (2.4) in Proposition 3.1 for the graph without boundary or Proposition 3.2 for the graph with boundary. Since $x-y\le\log(x)-\log(y)$ for $0<y\le x<1$, we also obtain

$$\max_{i\le N}|\rho_i-\rho_j|^2\le\frac{M}{c_0\min_{(i,j)\in E}\theta_{ij}(\rho(t))}\le\frac{M}{c_0c}.$$

By Proposition 2.1, we have that $(\rho(t),S(t))$ always stay in a subset $\widetilde{B}=\{\rho\in\mathcal{P}_o(G),S\in\mathbb{R}^N|K(S,\rho)+I(\rho)\le M\}$. Hence, we can get the estimates of the coefficients and their derivatives of (2.4) on $\widetilde{B}$. Taking the first derivatives of $\mathcal{H}$, we obtain

$$\begin{aligned}
\Big\|\frac{\partial\mathcal{H}}{\partial S}\Big\|_{l^\infty} &= \max_{i\le N}\Big|\sum_{j\in N(i)}\frac{1}{2}(S_i-S_j)\omega_{ij}\theta_{ij}\Big|\\
&\le E_{max}C_0\sqrt{\frac{2M}{cc_0}},\\
\Big\|\frac{\partial\mathcal{H}}{\partial\rho}\Big\|_{l^\infty} &\le \Big|\frac{1}{2}\sum_{j\in N(i)}\omega_{ij}(S_i-S_j)^2\frac{\partial\theta_{ij}}{\partial\rho_i}+\beta\frac{\partial I}{\partial\rho_i}(\rho)\Big|+M_0\\
&\le E_{max}C_0\Big(\frac{1}{2}\frac{M}{cc_0}+\beta\sqrt{\frac{M}{cc_0}}+\beta\frac{1}{c}\Big)+M_0.
\end{aligned}$$

Similarly, we obtain the estimates for higher derivatives of $\mathcal{H}$,

(4.2)

$$\Big\|\frac{\partial^2\mathcal{H}}{\partial S^2}\Big\|_{l^\infty} \le \max_{i\le N}\Big|\sum_{j\in N(i)}\omega_{ij}\theta_{ij}(\rho)\Big| \le C_0 E_{\max},$$

$$\Big\|\frac{\partial^2\mathcal{H}}{\partial S\partial\rho}\Big\|_{l^\infty} \le \max_{i\le N}\Big|\frac{1}{2}w\sum_{j\in N(i)}\omega_{ij}\big(S_i-S_j\big)\Big| \le \frac{1}{2}C_0E_{\max}\sqrt{\frac{2M}{cc_0}},$$

$$\Big\|\frac{\partial^2\mathcal{H}}{\partial\rho^2}\Big\|_{l^\infty} \le \max_{i\le N}\Big|\beta\sum_{j\in N(i)}\widetilde{\omega}_{ij}\big(1+\frac{\rho_j}{\rho_i^2}\big)\Big| + M_0 \le \beta C_0E_{\max}\Big(1+\frac{1}{c^2}\Big)+M_0.$$

By recursive calculations, we verify that

$$\Big\|\frac{\partial^n\mathcal{H}}{\partial\rho^n}\Big\|_{l^\infty} \le \beta C_0E_{max}((n-1)!\frac{1}{c^n}, \text{ for } n\ge 3, \tag{4.3}$$
$$\Big\|\frac{\partial^3\mathcal{H}}{\partial S^2\partial\rho}\Big\|_{l^\infty} \le \frac{1}{2}C_0E_{max}.$$

and all other partial derivatives are 0 for $n \ge 3$.

4.1. **Symplectic methods.** Let $T>0$ and τ be the time step size such that $t_m = m\tau$, $m \le \widetilde{M}$ and $T = t_{\widetilde{M}}$. In this part, we use the discrete Hamiltonian principle in [26, Chapter VI] to obtain the symplectic schemes whose numerical solution preserves the discrete symplectic structure (4.1)

$$d\rho^m \wedge dS^m = d\rho^0\wedge dS^0. \tag{4.4}$$

This is a standard procedure to derive symplectic schemes. We present it here for the completeness of this paper. Given the initial density ρ^0 and target density $\rho^{\widetilde{M}}$, the discrete Hamiltonian principle is looking for the critical point $(\rho^m)_{m=0}^{\widetilde{M}}$ of the functional

$$\mathbb{L}_\tau((\rho^m)_{m=0}^{\widetilde{M}}) = \sum_{m=0}^{\widetilde{M}-1} L_\tau(\rho^m,\rho^{m+1}).$$

Here the discrete Lagrangian $L_\tau(\rho^m,\rho^{m+1})$ is an approximation of the continuous Lagrangian $\int_{t_m}^{t_{m+1}} L(\rho(t),\dot\rho(t))dt$, where $\rho(t)$ is the solution of (2.4) with boundary values $\rho(t_m)=\rho^m$ and $\rho(t_{m+1})=\rho^{m+1}$.

Then, letting $\frac{\partial\mathcal{S}_\tau}{\partial\rho^m}=0$, for $m=1,\cdots,\widetilde{M}-1$, we get the discrete Euler-Lagrange equation

$$\frac{\partial L_\tau}{\partial x}(\rho^m,\rho^{m+1}) + \frac{\partial L_\tau}{\partial y}(\rho^{m-1},\rho^m) = 0,$$

where $\frac{\partial L_\tau}{\partial x}$ and $\frac{\partial L_\tau}{\partial y}$ refer to the partial derivatives with respect to the first and second argument.

Thus it gives a three-term difference scheme determining $\rho^1, \cdots, \rho^{\widetilde{M}-1}$. By introducing the discrete momenta via the discrete Legendre transformation $S^m = -\frac{\partial L_\tau}{\partial x}(\rho^m, \rho^{m+1})$, we can get

$$d\mathbb{L}_\tau = S^{\widetilde{M}} d\rho^{\widetilde{M}} - S^0 d\rho^0,$$

where $\mathbb{L}_\tau$ is also called symplecticity generating function. This implies the symplecticity of the map $(S^0, \rho^0) \to (S^{\widetilde{M}}, \rho^{\widetilde{M}})$ (see e.g. [26, Chapter VI]). Indeed, we get

$$S^m = -\frac{\partial L_\tau}{\partial x}(\rho^m, \rho^{m+1}),\ S^{m+1} = \frac{\partial L_\tau}{\partial y}(\rho^m, \rho^{m+1}).$$

Next, we give the derivation of symplectic Runge–Kutta methods as an example of symplectic integrators. We consider the numerical integration of $L(\rho, \dot{\rho})$ at $[t_0, t_1]$,

$$L_\tau(\rho^0, \rho^1) = \tau \sum_{l=1}^{s} b_l L(u(c_l\tau), \dot{u}(c_l\tau)),$$

where $0 \le c_1 < \cdots < c_s \le 1$ and $u(t)$ is the polynomial of degree s with $u(0) = \rho^0$ and $u(\tau) = \rho^1$ which extremizes the right-hand side (see [26] for more details on the collocation polynomial). Thus, our goal is to find the critical point of

$$L_\tau(\rho^0, \rho^1) = \tau \sum_{l=1}^{s} b_l L(\Phi^l, \dot{\Phi}^l),$$

$$\Phi^l = \rho^0 + \tau \sum_{n=1}^{s} a_{ln} \dot{\Phi}^n,$$

where $\Phi^l \in \mathcal{P}(G)$ and $\dot{\Phi}^l \in \mathbb{R}^N$ are chosen to extremize the above sum subject to the constraint

$$\rho^1 = \rho^0 + \tau \sum_{l=1}^{s} b_l \dot{\Phi}^l.$$

We assume that all the b_l are nonzero and that their sum equals 1. By applying the Lagrange multiplier method to the above problem (see e.g. [26, Chapter VI] for a standard derivation), the extreme point satisfies

$$S^1 = S^0 + \tau \sum_{l=1}^{s} b_l \dot{\Xi}^l,\ \rho^1 = \rho^0 + \tau \sum_{l=1}^{s} b_l \dot{\Phi}^l,$$

$$\Xi^l = p_0 + \tau \sum_{n=1}^{s} \widehat{a}_{ln} \dot{\Xi}^n,\ \Phi^l = q_0 + \tau \sum_{n=1}^{s} a_{ln} \dot{\Phi}^n,$$

where

$$\dot{\Xi}^l = \frac{\partial L}{\partial x}(\Phi^l, \dot{\Phi}^l),\ \ \Xi^l = \frac{\partial L}{\partial y}(\Phi^l, \dot{\Phi}^l)$$

with $\widehat{a}_{ln} = b_n - \frac{b_n a_{nl}}{b_l}$. By noticing that $\mathcal{H}(\rho, S) = S^\top \dot{\rho} - L(\rho, \dot{\rho})$, one may also rewrite the above system into its dual coordinates, that is,

$$
\begin{aligned}
S^1 &= S^0 - \tau \sum_{l=1}^{s} b_l \frac{\partial \mathcal{H}(\Phi^l, \Xi^l)}{\partial \rho}, \\
\rho^1 &= \rho^0 + \tau \sum_{l=1}^{s} b_l \frac{\partial \mathcal{H}(\Phi^l, \Xi^l)}{\partial S}, \\
\Xi^l &= S^0 - \tau \sum_{n=1}^{s} \widehat{a}_{ln} \frac{\partial \mathcal{H}(\Phi^n, \Xi^n)}{\partial \rho}, \\
\Phi^l &= \rho^0 + \tau \sum_{n=1}^{s} a_{ln} \frac{\partial \mathcal{H}(\Phi^n, \Xi^n)}{\partial S},
\end{aligned}
\tag{4.5}
$$

where the coefficients satisfy the condition $\widehat{a}_{ln} b_l + a_{nl} b_n = b_l b_n$, of partitioned Runge Kutta symplectic methods for the Wasserstein–Hamiltonian system (2.4).

Example 4.1. By taking $s = 1$, $a_{11} = 0, \widehat{a}_{11} = 0$, $b_1 = 1$ in (4.5), we obtain the symplectic Euler method, i.e.,

$$
\begin{aligned}
\rho_i^{m+1} &= \rho_i^m + \frac{\partial \mathcal{H}(S^{m+1}, \rho^m)}{\partial S} \tau, \\
&= \rho_i^m - \sum_{j \in N(i)} \omega_{ij} (S_j^{m+1} - S_i^{m+1}) \theta_{ij}(\rho^m) \tau \\
S_i^{m+1} &= S_i^m - \frac{\partial \mathcal{H}(S^{m+1}, \rho^m)}{\partial \rho} \tau, \\
&= S_i^m - \frac{1}{2} \sum_{j \in N(i)} \omega_{ij} (S_i^{m+1} - S_j^{m+1})^2 \frac{\partial \theta_{ij}(\rho^m)}{\partial \rho_i} \tau - \frac{\partial \mathcal{F}(\rho^m)}{\partial \rho_i} \tau,
\end{aligned}
\tag{4.6}
$$

where $\mathcal{F}(\rho) = \beta I(\rho) + \mathcal{V}(\rho) + \mathcal{W}(\rho)$. □

In the following, we focus on the case of symplectic Runge–Kutta methods, i.e., $\widehat{a}_{ln} = a_{ln}$. With minor modifications, similar results hold for the partitioned symplectic Runge–Kutta methods (see [26] for more details), i.e., $\widehat{a}_{ln} b_l + a_{nl} b_n = b_l b_n$.

Theorem 4.1. *Assume that $G = (V, E, \Omega)$ is a connected weighted graph and that $\min_{1 \le i \le N} \rho_i^0 > 0, S^0 \in \mathbb{R}^N$. Let τ be the time step size and $T > 0$. Then the symplectic Runge–Kutta scheme* (4.5) *satisfies the following properties:*

(i) *It preserves mass:*

$$\sum_{i=1}^{N} \rho_i^m = \sum_{i=1}^{N} \rho_i^0.$$

(ii) *It preserves symplectic structure:*

$$d\rho^m \wedge dS^m = d\rho^0 \wedge dS^0.$$

(iii) *If* (4.5) *is symmetric, then it is time reversible: if* (ρ^m, S^m) *is the solution of the full discretization, then* $(\rho^{-m}, -S^{-m})$ *is also the solution of* (4.5).

(iv) *It is time transverse (gauge) invariant: if* $\mathbb{V}^\alpha = \mathbb{V} + \alpha$, *then* $(S^\alpha)^m = S^m + \alpha m\tau$ *is the solution of* (4.5) *with linear potential* $\mathbb{V}^\alpha$.

(v) *If a time invariant* $\rho^* \in \mathcal{P}_o(G)$ *and* $(S^*)^m = -\nu m\tau$ *form an interior stationary solution of* (4.5), *then* ρ^* *is the critical point of* $\mathcal{H}(\rho, S)$ *and*

$$\nu = \mathcal{H}(\rho^*, S^*) + \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N \mathbb{W}_{ij}\rho_i^*\rho_j^*.$$

(vi) *When* $\frac{M}{\beta}$ *is small enough, the scheme almost preserves the Hamiltonian up to time* $\widetilde{T} = \mathcal{O}(\tau^{-r})$:

$$\mathcal{H}(S^m, \rho^m) = \mathcal{H}(S^0, \rho^0) + \mathcal{O}(\tau^r),$$

where r *is the order of the symplectic numerical scheme.*

Proof. Property (i) holds since $\frac{\partial \mathcal{H}}{\partial S}$ in (4.5) is linear with respect to ρ. Property (ii) can be verified by using the symplecticity condition $a_{ij}b_i + a_{ji}b_j = b_i b_j$.

Denote the exact flow of the original system (2.4) by

$$\Gamma_t(\rho^0, S^0) = (\rho(t, \rho^0), S(t, S^0)).$$

It follows from Proposition 2.1 that Γ_t is g-reversible, i.e., $g \circ \Gamma_t = \Gamma_t^{-1} \circ g$, with $g(S, \rho) = (-S, \rho)$. Since the one-step method (4.5) is symmetric, i.e., $\widetilde{\Gamma}_\tau(\rho^0, S^0) = (\rho^1, S^1)$ satisfies $\widetilde{\Gamma}_\tau \circ \widetilde{\Gamma}_{-\tau} = I$, then $\widetilde{\Gamma}_\tau$ is g-reversible, i.e., $g \circ \widetilde{\Gamma}_\tau = \widetilde{\Gamma}_\tau^{-1} \circ g$, by [26, Theorem 1.5 Chapter V], and (iii) holds.

Property (iv) can be directly verified because $K(\rho, S)$ is an even function of $S_i - S_j, (i, j) \in E$ and the potential $\mathcal{V}$ is linear with respect to ρ. To show Property (v), we only need to show that ρ^* satisfies the Karush-Kuhn-Tucker conditions of the optimality for the minimization of

$$\min_{\rho \in \mathcal{P}_o(G)} \mathcal{H}(\rho, S^*) = \min_{\rho \in \mathcal{P}_o(G)} (\beta I(\rho) + \mathcal{V}(\rho) + \mathcal{W}(\rho)),$$

which can be done using the Lagrange multiplier method as in the continuous case.

We next focus on the proof of (vi). Denote $(y^m)^\top = (\rho^m, S^m)^\top$, $(\widetilde{y}^{m,l})^\top = (\Phi^{m,l}, \Xi^{m,l})^\top, l \leq s$, and $f(y)^\top = (\frac{\partial \mathcal{H}}{\partial S}(\rho, S), -\frac{\partial \mathcal{H}}{\partial \rho}(\rho, S))^\top$ for $y^\top = (\rho, S)^\top$.

Then r-th order Runge–Kutta scheme can be rewritten as

$$y^1 = y^0 + \tau \sum_{i=1}^{s} b_i f(\tilde{y}^i),$$
$$\tilde{y}^i = y^0 + \tau \sum_{j=1}^{s} a_{ij} f(\tilde{y}^j).$$

Notice that

$$(y_i^0)_{i=1}^N \in B = \Big\{\rho \in \mathcal{P}_o(G) \mid \beta I(\rho) \le \mathcal{H}_0 - \min_{\rho\in\mathcal{P}(G)} (\mathcal{V}(\rho) + \mathcal{W}(\rho))\Big\}.$$

Let K be the smallest number such that $(y_i^{K+1})_{i=1}^N \notin B$. Thus, there exists some index $j \le N$, $y_j^{K+1} = \min_{i=1,\dots,N} |y_i^{K+1}| = \alpha c, 0 < \alpha < 1$, where c is the lower bound of the density of (2.4). By applying the Taylor expansion repeatedly and using (4.2) and (4.3), there exists a constant $C_{r,M,c_0,C_0} > 0$ such that

$$\begin{aligned} |y_i(t_{K+1}) - y_i^{K+1}| &\le |y_i(t_K) - y_i^K| + C_{r,M,c_0,C_0}(1 + \frac{\beta}{c^{r+1}})\tau^{r+1} \\ &\le K\tau C_{r,M,c_0,C_0}(1 + \frac{\beta}{c^{r+1}})\tau^r, \end{aligned}$$

which implies that for $i = 1, \cdots, N$,

$$y_i^{K+1} \ge y_i(t_{K+1}) - K\tau C_{r,M,c_0,C_0} \frac{\beta + c^{r+1}}{c^{r+1}} \tau^r.$$

Thus we have a lower bound of K, that is,

$$K\tau \ge \frac{c^{r+1}}{\tau^r C_{r,M,c_0,C_0}(c^{r+1} + \beta)}(1 - \alpha).$$

As a consequence, before $\widetilde{T} = K\tau$, the lower bound of the density is uniformly controlled by c since $\frac{\mathcal{H}^0}{\beta}$ is small by the lower bound of (2.4) in Section 3. We complete the proof of (vi) by using the Taylor expansion on the energy before $\widetilde{T} = K\tau$. □

By (4.2), the solvability of the scheme requires the classical condition,

$$\max\Big(C_0 E_{max}, \frac{1}{2} C_0 E_{max}\sqrt{\frac{2M}{cc_0}}, \beta C_0 E_{max}(1 + \frac{1}{c^2}) + M_0\Big) \le \frac{1}{\tau}\text{constant},$$

where the constant only depends on the numerical method.

4.1.1. *Backward error analysis.* In spite of point (vi) in Theorem 4.1, symplectic methods nearly preserve the Hamiltonian for times much longer than $\mathcal{O}(\tau^{-r})$, since the backward error analysis allows for an exponentially small error between the symplectic scheme and its modified equation. To apply the backward error analysis, we need to verify that the coefficients of the equation admit an analytic extension on the complex domain, which we will do next.

By choosing the principle value of the logarithm of z in $\mathbb{C}/\{0\}$, denoted by $\mathrm{Log}(z) := \log|z| + i\mathrm{Arg}(z)$, it is known that $\mathrm{Log}(z)$ is analytic except along the negative real axis. Since $\frac{1}{\rho_i}$ and $\log(\rho_i)$ can be extended to analytic complex functions for $\rho_i \in \mathbb{C}/\{0\}$, we can extend

$$\begin{aligned} f(y)^\top &= \Big(\frac{\partial \mathcal{H}}{\partial S}, -\frac{\partial \mathcal{H}}{\partial \rho}\Big)^\top \\ &= \Big(\frac{1}{2}\sum_{j\in N(i)} \omega_{ij}(S_i - S_j)(\rho_i+\rho_j), \\ &\qquad -\frac{1}{4}\sum_{j\in N(i)} \omega_{ij}(S_i-S_j)^2 - \sum_{j\in N(i)} \widetilde{\omega}_{ij}\Big(1-\frac{\rho_j}{\rho_i} - \log\big(\frac{\rho_j}{\rho_i}\big)\Big)\Big)^\top \end{aligned}$$

to a complex function in $\mathbb{C}^{2N}$ such that for any interior point y^0 of B, $f(y)$ is analytic in the neighborhood of y^0 and that there exists $R>0$, $M_c>0$ such that

$$\|f(y)\| \le M_c, \text{ for } \|y-y^0\| \le 2R.$$

Here M_c is constant depending on c. This is applicable since we can choose $R \le \frac{1}{4}\mathrm{dist}(y_0, \partial B)$, where ∂B is the boundary of B in (2.8), such that

$$\min_{i=1,\dots,N} |y_i| = \min_{i=1,\dots,N} |\rho_i| \ge c,$$

and

$$\|f(y)\|_{l^\infty} \le \max\Big(C_0 E_{max}, \frac{1}{2}C_0 E_{max}\sqrt{\frac{2M}{cc_0}}, \beta C_0 E_{max}\big(1+\frac{1}{c^2}\big) + M_0\big)\Big)$$

by (4.2). Thus, we can apply the backward error analysis in our case. We first introduce the truncated modified differential equation of (2.4) with respect to an r-th order numerical scheme,

$$\dot{\widetilde{y}} = F_{\mathcal{N}}(\widetilde{y}),\ F_{\mathcal{N}}(\widetilde{y}) = f(\widetilde{y}) + \tau^r f_{r+1}(\widetilde{y}) + \dots + \tau^{N-1} f_{\mathcal{N}}(\widetilde{y}) \tag{4.7}$$

with $\widetilde{y}(0) = y(0)$. It is well-known that the above modified equation is also a Hamiltonian system with the modified Hamiltonian $\widetilde{\mathcal{H}}(y) = \mathcal{H}(y) + \tau^r \mathcal{H}_{r+1}(y) + \dots + \tau^{N-1}\mathcal{H}_{\mathcal{N}}(y)$. According to [26, Theorem 7.2 and Theorem 7.6], we have that for (4.5), if $f(y)$ is analytic and $\|f(y)\| \le M_c$ in the complex ball $B_{2R}(y_0)$, then the coefficients d_j in the Taylor expansion of the numerical method

$$\widetilde{\Gamma}_\tau(y) = y + \tau f(y) + \tau^2 d_2(y) + \dots + \tau^j d_j(y) + \dots,$$

are analytic and satisfy $\|d_j(y)\| \le C\big(\frac{M_c}{R}\big)^{j-1}$ in $B_R(y_0)$ for a constant $C>0$. If $\tau \le \tau_0$ with $\tau_0 \le \frac{R}{CM_c}$, then there exists $\mathcal{N} = \mathcal{N}(\tau)$ satisfying $\tau\mathcal{N} \le h_0$ such that

$$\|\widetilde{\Gamma}_\tau(y^0) - \widetilde{\phi_{N,\tau}}(y^0)\| \le C\tau M_c e^{-\frac{\tau_0}{\tau}},$$

where $y^1 = \widetilde{\Gamma}_\tau(y^0)$ is the numerical solution and $\widetilde{\phi_{\mathcal{N},\tau}}(y^0)$ is the exact solution of (4.7) at $t=\tau$.

As a consequence of the above, by [26, Theorem 8.1] the long-time energy conservation is obtained as follows. Assume that the numerical solution of the symplectic method (4.5) $\widetilde{\Gamma}_\tau(y)$ stays in the compact set B, then there exists R, τ_0 and $N(\tau_0)$ such that

$$|\widetilde{\mathcal{H}}(y^m) - \widetilde{\mathcal{H}}(y^0)| \le m\tau M_c e^{-\frac{\tau_0}{\tau}},$$

$$|\mathcal{H}(y^m) - \mathcal{H}(y^0)| \le C\frac{M_c^{p+1}}{R^p}\tau^p.$$

We summarize the above discussion in Corollary 4.1.

Corollary 4.1. *Under the same condition of Theorem* 4.1*, when* $\frac{M}{\beta}$ *is small enough, there exist* τ_0 *small enough,* $C_M > 0$*, and a modified energy* $\widetilde{\mathcal{H}}$*,* $\mathcal{O}(\tau^r)$*-close to* $\mathcal{H}$*, such that for any* $\tau < \tau_0$*,* $m\tau < T$*,*

$$|\widetilde{\mathcal{H}}(S^m, \rho^m) - \widetilde{\mathcal{H}}(S^0, \rho^0)| \le m\tau C_M e^{-\frac{\tau_0}{\tau}}.$$

4.2. **Regularizations.** Here we look at two instances of regularization for the geodesic equation (2.12) with $\mathbb{V} = 0, \mathbb{W} = 0$ and $\beta = 0$. The first is based on the Fisher information, and the other is based on the standard viscosity solution. We assume that $\mathcal{M} = [0, 1] \subset \mathbb{R}$ and for simplicity restrict to (2.12) subject to periodic boundary conditions without the term $\mathcal{F}(\rho)$. The initial condition $\rho(0) > 0$, and $S(0)$, are smooth and bounded functions on $\mathcal{M}$.

4.2.1. *Fisher information regularization symplectic scheme.* For the geodesic equation, its Lagrangian formalism is equivalent to its Hamiltonian formalism. We can directly apply the Fisher information regularization symplectic scheme (4.5) to the semi-discretization of the Hamiltonian system. We use the mid-point scheme applied to the graph generated by the central difference scheme under the periodic condition as an example of a fully discrete scheme,

(4.8)

$$\begin{aligned}
\rho_i^{m+1} &= \rho_i^m + \frac{\partial \mathcal{H}(S^{m+\frac{1}{2}}, \rho^{m+\frac{1}{2}})}{\partial S_i}\tau, \\
&= \rho_i^m - \sum_{j\in N(i)} \frac{\tau}{h^2}(S_j^{m+\frac{1}{2}} - S_i^{m+\frac{1}{2}})\theta_{ij}(\rho^{m+\frac{1}{2}}) \\
S_i^{m+1} &= S_i^m - \frac{\partial \mathcal{H}(S^{m+\frac{1}{2}}, \rho^{m+\frac{1}{2}})}{\partial \rho_i}\tau, \\
&= S_i^m - \frac{1}{2}\sum_{j\in N(i)} \frac{\tau}{h^2}\big(S_i^{m+\frac{1}{2}} - S_j^{m+\frac{1}{2}}\big)^2 \frac{\partial \theta_{ij}}{\partial \rho_i}(\rho^{m+\frac{1}{2}}) - \beta\frac{\partial I(\rho^{m+\frac{1}{2}})}{\partial \rho_i}\tau,
\end{aligned}$$

where h is the space step size, τ be the time step size such that $h = \frac{1}{N}$, $T = \widetilde{M}\tau$ and $i \le N-1, m \le \widetilde{M} - 1$. Then all the properties in Theorem

4.1 hold. According to the lower bound estimate of ρ in Section 3 and (4.2), we have the following space-time step size restriction,

$$\tau \max\Big(C_0 E_{max}, \frac{1}{2} C_0 E_{max}\sqrt{\frac{2M}{cc_0}}, \beta C_0 E_{max}(1+\frac{1}{c^2}) + M_0\Big) = \mathcal{O}(1),$$

where

$$c \geq \min\big(\kappa_0, \kappa_1\big) \text{ and } c_0 = C_0 = \frac{1}{h^2}$$

with

$$\kappa_0 = \frac{1}{2}\min_{i=1}^{N} \rho_i(0),\ \kappa_1 = \Big(1 + N\exp(\frac{M(N-1)([\frac{N-1}{2}]+1)h^2}{\beta})\Big)^{-1}.$$

Theorem 4.2. *Let $h > 0$ be the spatial step size, τ the temporal step size and $T = \widetilde{M}\tau > 0$. Assume that G is the lattice graph of $\mathcal{M}$ with the periodic boundary condition and that $\min_{1\leq i\leq N} \rho_i^0 > 0, S^0 \in \mathbb{R}^N$. Then the symplectic Runge–Kutta scheme* (4.5) *has the following properties:*

(i) *It preserves mass:*

$$\sum_{i=1}^{N} \rho_i^m h = \sum_{i=1}^{N} \rho_i^0 h.$$

(ii) *It preserves symplectic structure:*

$$d\rho^m \wedge dS^m = d\rho^0 \wedge dS^0.$$

(iii) *If* (4.5) *is symmetric, then it is time reversible: if (ρ^m, S^m) is the solution of the full discretization, then $(\rho^{-m}, -S^{-m})$ is also the solution of* (4.5).

(iv) *It is time transverse (gauge) invariant: if $\mathbb{V}^\alpha = \mathbb{V} + \alpha$, then $(S^\alpha)^m = S^m + \alpha m\tau$ is the solution of* (4.5) *with linear potential $\mathbb{V}^\alpha$.*

(v) *When $\frac{M}{\beta}$ is small enough, the scheme almost preserves the Hamiltonian up to time $\widetilde{T} = \mathcal{O}(\tau^{-r})$:*

$$\mathcal{H}(S^m, \rho^m) = \mathcal{H}(S^0, \rho^0) + \mathcal{O}(\tau^r),$$

where r is the order of the symplectic numerical scheme.

If we do not add a regularization term, like the Fisher information, to the numerical scheme of (2.4), then the numerical scheme may develop singularities and produce unstable behavior. Example 4.2 indicates that even the structure-preserving numerical scheme which uses the upwind weight θ^U without regularization may fail—at a finite step m—to maintain positivity for ρ_i^m, or it may lead to blow up in S_i^m.

Example 4.2. Assume that the graph has only two points. Assume that $\rho_1(0), \rho_2(0) > 0$ and $S_1(0), S_2(0)$ are the corresponding initial densities and potentials of the two points:

$$\theta_{ij}^U(\rho) = \rho_j, \text{ if } S_i > S_j,$$
$$\theta_{ij}^U(\rho) = \rho_i, \text{ if } S_i < S_j.$$

For simplicity, take $S_1(0) > S_2(0)$, $\mathcal{F}(\rho) = 0$. Then the finite dimensional system becomes

$$\dot\rho_1(t) = (S_1(t) - S_2(t))\rho_2,\ \dot\rho_2(t) = (S_2(t) - S_1(t))\rho_2,$$
$$\dot S_1(t) = 0,\ \dot S_2(t) = -\frac{1}{2}|S_1(t) - S_2(t)|^2.$$

Then $S_1(t) - S_2(t) = \frac{S_1(0)-S_2(0)}{1-\frac{1}{2}(S_1(0)-S_2(0))t)}$. Until $t < \frac{2}{S_1(0)-S_2(0)}$, ρ_1 and ρ_2 are strictly positive, but when $t = \frac{2}{S_1(0)-S_2(0)}$, $\rho_1 = 1$, $\rho_2 = 0$.

4.2.2. *Regularization by adding viscosity.* As alternative to adding the Fisher information as regularization term, a classical regularization procedure is adding small viscosity in order to obtain monotone schemes for S (see e.g. [16]). For example, by introducing the viscosity regularization term

$$\alpha_i(S^m) := \alpha(S_{i+1}^m - 2S_i^m - S_{i-1}^m), \tag{4.9}$$

where $\alpha \in \mathbb{R}$ is used to guarantee the monotonicity of S_i^{m+1}. This is a standard procedure (*elliptic regularization*), which we detail and further use it in our numerical tests for comparison purposes. As we shall see, although adding viscosity does lead to a well defined discretization (see (4.10)), unlike the regularization scheme (4.8), the resulting numerical scheme does not preserve time reversibility, symplectic structure nor nearly preserves the energy. This will be highlighted in the numerical tests in Section 5.

Let h be the space step size, τ be the time step size such that $h = \frac{1}{N}$, $T = \widetilde{M}\tau$ and $i \le N, m \le \widetilde{M}$. Assume that $\max_{i\le N-1}\max_{m\le \widetilde{M}} |\frac{S_{i+1}^m - S_i^m}{h}| \le R$ for some $R > 0$. Then, we can choose α $(0 < \alpha < \frac{1}{2}, \alpha \ge R\frac{\tau}{h})$ such that

$$1 - \frac{\tau}{h}\Big(\frac{\big(S_{i+1}^m - S_i^m\big)^+}{h} + \frac{(S_{i-1}^m - S_i^m)^+}{h}\Big) - 2\alpha \ge 0,$$
$$-\frac{\tau}{h}\frac{\big(S_{i+1}^m - S_i^m\big)^+}{h} + \alpha \ge 0,$$
$$-\frac{\tau}{h}\frac{\big(S_{i-1}^m - S_i^m\big)^+}{h} + \alpha \ge 0.$$

Doing so, we get the following scheme (*viscosity regularization*):

$$\begin{aligned}\rho_i^{m+1} &= \rho_i^m + \tau\Big(\frac{S_i^m - S_{i+1}^m}{h^2}\Big)^+\rho_{i+1}^m + \tau\Big(\frac{S_i^m - S_{i-1}^m}{h^2}\Big)^+\rho_{i-1}^m \\ &\quad + \tau\Big(\frac{S_i^m - S_{i+1}^m}{h^2}\Big)^-\rho_i^m + \tau\Big(\frac{S_i^m - S_{i-1}^m}{h^2}\Big)^-\rho_i^m \\ S_i^{m+1} &= S_i^m - \frac{1}{2}\tau\Big|\frac{(S_i^m - S_{i+1}^m)^-}{h}\Big|^2 - \frac{1}{2}\tau\Big|\frac{(S_i^m - S_{i-1}^m)^-}{h}\Big|^2 + \alpha_i(S^n).\end{aligned} \tag{4.10}$$

Let ρ^0 and S^0 be the grid function of $\rho(0)$ and $S(0)$ on the grid G. Then (4.10) enjoys the following properties, which implies that the numerical viscosity term leads to positivity of the density function and uniform boundedness of S.

Theorem 4.3. *Assume that* $\max_{i\le N-1}\max_{m\le \widetilde{M}}|\frac{S^m_{i+1}-S^m_i}{h}|\le R,\ \alpha\ge R\frac{\tau}{h}$. *Then there exists a unique solution* $(\rho^m,S^m)_{m=1}^{\widetilde{M}}$ *of* (4.10) *and it satisfies the following properties.*

(i) *Mass is preserved: for* $m\le \widetilde{M}$,

$$\sum_{i=1}^N \rho_i^m=\sum_{i=1}^N \rho_i^0.$$

(ii) *It preserves strict positivity: if* $\min_{i=1,\dots,N}\rho_i^0>0$, *then* $\min_{i=1,\dots,N}\rho_i^m>0$ *for* $m\le \widetilde{M}$.

(iii) *If* $\frac{\tau}{h}$ *is sufficiently small, and* $\tau,h\to 0$, *then* S_i^m *converges to the viscosity solution of the Hamilton–Jacobi equation.*

(iv) *Let* $\tau>0$ *be fixed. Then it holds that* $\lim_{m\to\infty}S^m=S^\infty$ *and* $\lim_{m\to\infty}\rho^m=\rho^\infty$, *where* $\rho^\infty\in\mathcal{P}(G)$.

(v) *It holds that*

$$\|S^m\|_{l^\infty}\le\|S^0\|_{l^\infty},\ \|\rho^m\|_{l^\infty}\le\max((1+R\frac{\tau}{h})^m\|\rho^0\|_{l^\infty},1/h).$$

Proof. For Properties (i), (iii) and (v), we refer to [16] for their proofs relative to the numerical approximation

$$S_i^{m+1}=S_i^m-\frac{1}{2}\tau\Big|\frac{S^m_{i+1}-S^m_i}{h}\Big|^2+\alpha_i(S^m).$$

We proceed to prove (ii) and (iv).

Due to the expression of ρ_i^{m+1}, we get

$$\rho_i^{m+1}\ge\rho_i^m+\frac{\tau}{h}\Big((\frac{S_i^m-S_{i+1}^m}{h})^-+(\frac{S_i^m-S_{i-1}^m}{h})^-\Big)\rho_i^m\ge(1-2R\frac{\tau}{h})\rho_i^m,$$

which leads to

$$\rho_i^m\ge(1-2R\frac{\tau}{h})^m\rho_i^0.$$

Thus we have that

$$\rho_i^m\ge e^{-c_1\frac{\tau m}{h}}\min_{i=1,\dots,N}\rho_i^0$$

for some $c_1>0$ and thus (ii) holds.

Now we are in a position to show (iv). Since S^m is uniformly bounded with respect to m, there exists a sub-sequence $(S^{m_k})_{k=1}^\infty$ converging to a constant S^∞. By using the comparison principle, we get that for any $k,l,n\in\mathbb{N}^+$,

$$\|S^{m_k+n}-S^{m_l+n}\|_{l^\infty}\le\|S^{m_k}-S^{m_l}\|_{l^\infty}.$$

Thus $(S^{m_k})_{k=1}^{\infty}$ is a Cauchy sequence in $l^{\infty}(V \times \mathbb{N}^+)$ and converges to the same limit S^{∞}. On the other hand, one can also check that the solution of the following relation

$$\frac{1}{2}\Big|\frac{(S_i^{\infty} - S_{i+1}^{\infty})^-}{h}\Big|^2 + \frac{1}{2}\Big|\frac{(S_i^{\infty} - S_{i-1}^{\infty})^-}{h}\Big|^2 + \alpha_i(S^{\infty}) = 0 \tag{4.11}$$

must be 0. Indeed, let us assume that there is a nonzero solution for (4.11). From the fact that $\alpha_i(S^{\infty}) > 0$ if $S_i^{\infty} - S_{i+1}^{\infty} < 0$, $S_i^{\infty} - S_{i+1}^{\infty} < 0$ and $\alpha_i(S^{\infty}) < 0$, if $S_i^{\infty} - S_{i+1}^{\infty} > 0$, $S_i^{\infty} - S_{i+1}^{\infty} > 0$, the nonzero solution of (4.11) should have different signs for $S_i^{\infty} - S_{i+1}^{\infty}$ and $S_i^{\infty} - S_{i-1}^{\infty}$ at each node a_i. For simplicity assume that $S_i^{\infty} - S_{i+1}^{\infty} < 0$ and $S_i^{\infty} - S_{i-1}^{\infty} > 0$. Now, adding all the equations together, we obtain that

$$\sum_{i=1}^{N} \frac{1}{2}\Big|\frac{(S_i^{\infty} - S_{i+1}^{\infty})^-}{h}\Big|^2 = 0,$$

which contradicts the fact that $S_i^{\infty} - S_{i+1}^{\infty} < 0$ for $i = 1, \cdots, N$. Repeating this argument, it follows that for any $1 \le j \le N$, the solution of the following relation

$$\sum_{i=1}^{j} \left[\frac{1}{2}\Big|\frac{(S_i^{\infty} - S_{i+1}^{\infty})^-}{h}\Big|^2 + \frac{1}{2}\Big|\frac{(S_i^{\infty} - S_{i-1}^{\infty})^-}{h}\Big|^2 + \alpha_i(S^{\infty})\right] = 0$$

must be 0. As a consequence, for any subsequence $(S^{m_k})_k^{\infty}$, we have that the quantity

$$\frac{S_i^{m_k+1} - S_i^{m_k}}{\tau} = -\frac{1}{2}\Big|\frac{(S_i^{m_k} - S_{i+1}^{m_k})^-}{h}\Big|^2 - \frac{1}{2}\Big|\frac{(S_i^{m_k} - S_{i-1}^{m_k})^-}{h}\Big|^2 + \alpha_i(S^{m_k})$$

converges to

$$\frac{1}{2}\Big|\frac{(S_i^{\infty} - S_{i+1}^{\infty})^-}{h}\Big|^2 + \frac{1}{2}\Big|\frac{(S_i^{\infty} - S_{i-1}^{\infty})^-}{h}\Big|^2 + \alpha_i(S^{\infty}) = 0,$$

which only possesses the unique zero solution. Since $\|\rho\|_{l^1} = 1$, there exists a subsequence $(\rho^{m_k})_{k=1}^{\infty}$ which converges to a density probability ρ^{∞}. From (4.10) and the convergence of S, we are in a position to show that all the subsequence of $(\rho^m)_{m=1}^{\infty}$ converges to the same limit ρ^{∞}.

Next, we show that for given k sufficient large, then $\{\rho^{m_k+n}\}_{n=1}^{\infty}$ is a Cauchy sequence. Indeed, we have

$$\begin{aligned}
\|\rho_i^{m_k+1} - \rho_i^{m_k}\|_{l^\infty} \le{}& \tau\|\rho_{i+1}^{m_k}\|_{l^\infty}\Big\|\Big((\frac{S_i^{m_k} - S_{i+1}^{m_k}}{h^2})^+ - (\frac{S_i^{\infty} - S_{i+1}^{\infty}}{h^2})^+\Big)\Big\|_{l^\infty} \\
&+ \tau\|\rho_{i+1}^{m_k}\|_{l^\infty}\Big\|\Big(\frac{S_i^{m_k} - S_{i-1}^{m_k}}{h^2})^+ - \frac{S_i^{\infty} - S_{i-1}^{\infty}}{h^2})^+\Big)\Big\|_{l^\infty} \\
&+ \tau\|\rho_i^{m_k}\|_{l^\infty}\Big\|\Big(\frac{S_i^{m_k} - S_{i+1}^{m_k}}{h^2})^- - (\frac{S_i^{\infty} - S_{i+1}^{\infty}}{h^2})^-\Big)\Big\|_{l^\infty} \\
&+ \tau\|\rho_i^{m_k}\|_{l^\infty}\Big\|\Big(\frac{S_i^{m_k} - S_{i-1}^{m_k}}{h^2})^- - (\frac{S_i^{\infty} - S_{i-1}^{\infty}}{h^2})^-\Big)\Big\|_{l^\infty},
\end{aligned}$$

which, together with the uniform convergence of S, implies that ρ^{m_k+n} is a Cauchy sequence and possesses the same limit ρ^∞. □

5. Numerical examples

Here we show performance of the numerical schemes on several examples. All the numerical tests are performed under periodic boundary conditions in space, for given initial conditions $\rho(0,x) = \mu^0(x)$ and $S(0,x) = S^0(x)$, as specified below.

Example 5.1. Geodesic equation. This is (1.5), but we replace q with x following the convention in PDEs:

$$\frac{\partial}{\partial t}\rho(t,x) + \nabla \cdot (\rho(t,x)\nabla S(t,x)) = 0,$$
$$\frac{\partial}{\partial t}S(t,x) + \frac{1}{2}|\nabla S(t,x)|^2 = 0.$$

We report on the results of two different strategies: the upwind scheme (4.10) with numerical viscosity, and the Fisher information regularization symplectic scheme (4.8). We look at three different sets of initial values, cases (I)-(II)-(III), to compare the evolution of the density function and energy. (The different behaviors of S and ∇S for (4.10) and (4.8) are not of interest, since for (4.10) S will always converge to a constant; see Theorem 4.3.)

(I) In Figure 5.1, we show the behavior of (4.10) and (4.8) with initial value $\mu^0(x) = \frac{\exp(-10(x-0.5)^2)}{K}$ and $S^0(x) = -\frac{1}{5}\log(\cosh(5(x-0.5)))$, for $0 \le x \le 1$. Here K is a normalization constant so that $\int_0^1 \mu^0(x)dx = 1$. We observe that for $T < 0.15$ the two scheme behave quite closely to each other and the density concentrates at the point 0.5. But, after $T = 0.15$, the density of (4.8) begins to oscillate. Here, we choose spatial step size $h = 5\times 10^{-3}$, temporal step size $\tau = 10^{-4}$, viscosity coefficient $\alpha = 1/12$ for (4.10), and $\theta_{ij}(\rho) = \theta_{ij}^U(\rho)$, $\widetilde{\theta}_{ij}(\rho) = \theta_{ij}^L(\rho)$, $\beta = 10^{-5}$ for (4.8). In Figure 5.2, we also plot the density functions computed for (4.8) with different schemes and different temporal and spatial step sizes, and clearly the oscillations appear to be independent of the choice of schemes and mesh sizes; this indicates that the oscillations exist for the continuous system.

(II) In Figure 5.3, for $0 \le x \le 1$, we take $\mu^0(x) = 1$ and

$$S^0(x) = -\frac{1}{5}\log(\cosh(5(x-0.5))).$$

We choose spatial step size $h = 1.5\times 10^{-3}$, temporal step size $\tau = 1.3863\times 10^{-5}$, viscosity coefficient $\alpha = 8\times 10^{-2}$ for (4.10), and $\theta_{ij}(\rho) = \theta_{ij}^U(\rho)$, $\widetilde{\theta}_{ij}(\rho) = \theta_{ij}^L(\rho)$, $\beta = 5\times 10^{-7}$ for (4.8).

(III) In Figure 5.4, for $0 \le x \le 2$, we choose $\mu^0 = \frac{1}{2}$, $S^0 = \frac{1}{8}\sin(2\pi x)$, the spatial step size $h = 10^{-2}$, temporal step size $\tau = 10^{-4}$, viscosity coefficient

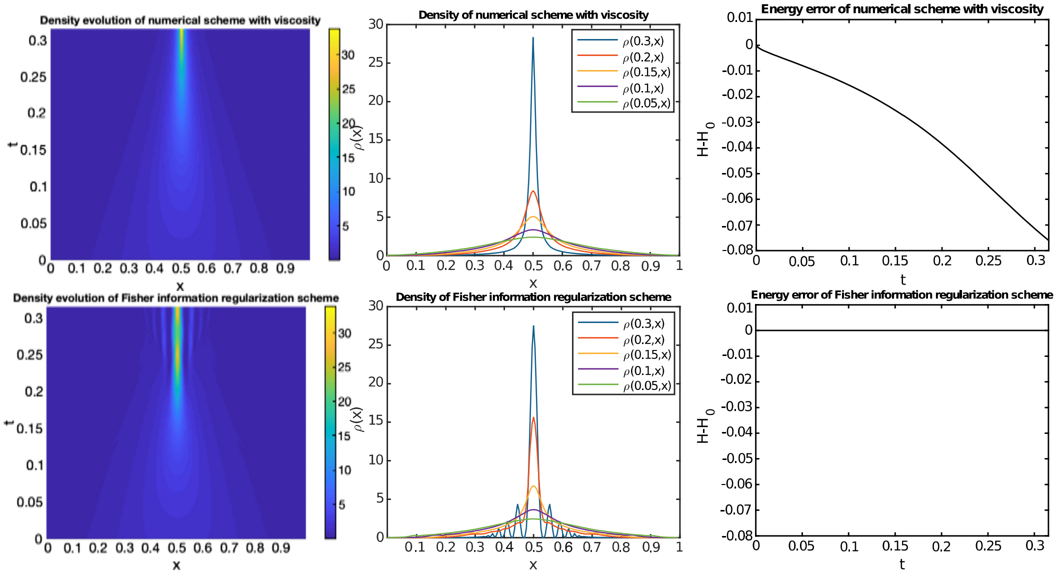


FIGURE 5.1. Example 5.1, case (I). The contour plot of $\rho(t,x)$ (left), snapshots of $\rho(t,x)$ at $t=(0.3,0.2,0.15,0.1,0.05)$ (middle) and energy error before $T=0.315$ (right) for the upwind scheme (4.10) with numerical viscosity (top) and the Fisher information regularization symplectic scheme (4.8) (bottom).

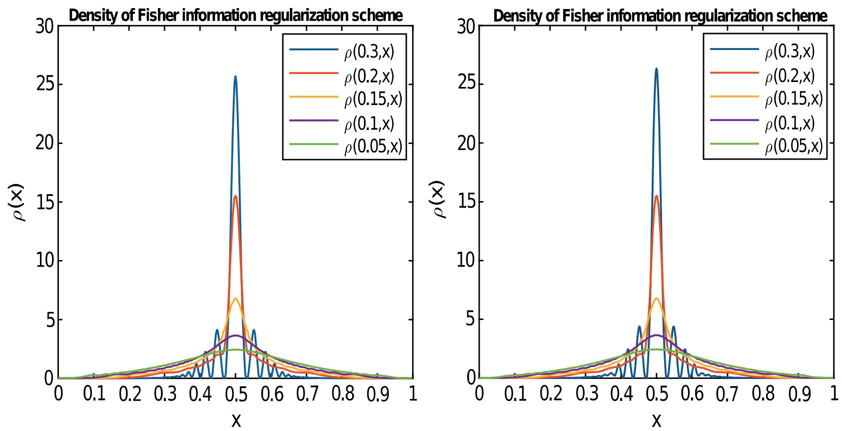
$\alpha = 5\times 10^{-2}$ for (4.10), and $\theta_{ij}(\rho)=\theta^U_{ij}(\rho)$, $\widetilde{\theta}_{ij}(\rho)=\theta^L_{ij}(\rho)$, $\beta=10^{-4}$ for (4.8).

All the numerical tests show that the Fisher information regularization scheme (4.8) preserves more of the relevant structures for (2.12), such as the energy evolution and time transverse invariance, compared to the numerical scheme (4.10). Meanwhile (4.8) causes oscillatory behavior after the singularity of (2.12) is developed.

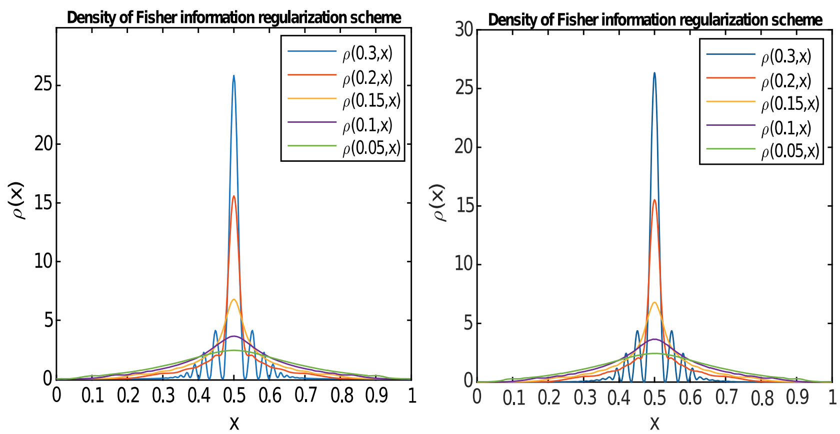
Figure 5.5 shows the relationship between β and the largest time step size τ in (4.8) that still gives the correct approximation to the solution. In this numerical test, we use $h=5\times 10^{-2}, T=4$, $0\le x\le 1$, $S^0(x)=\frac{\sin(\pi x)}{\pi}$, $\mu^0(x)=1$. From Figure 5.5, we can see that the relationship between $\frac{\mathcal{H}_0}{\beta}$ and τ is very sensitive when $\frac{\mathcal{H}_0}{\beta}$ is large.

Example 5.1 shows the dramatic difference between the numerical schemes regularized via viscosity, (4.10), or via the Fisher information, (4.8), when the continuous system (1.5) develops a singularity. Next, we test the impact of β in (4.8) when (1.5) does not develop any singularity in Example 5.2. It turns out that the numerical solution of (4.8) converges to the true solution as β goes to 0, just as the standard Hamilton–Jacobi solver (i.e., (4.10)) does.

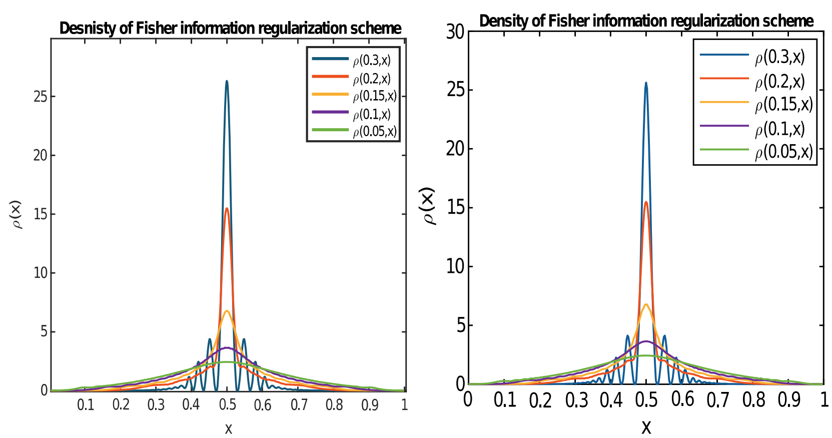
Example 5.2. Push-forward density in optimal transport problem. We use the Wasserstein geodesic equations from optimal transport in a case where the exact solution is known, to check whether the Fisher information



(a) Midpoint scheme



(b) Implicit Euler scheme



(c) Symplectic Euler scheme

FIGURE 5.2. Example 5.1, case (I). In (a) and (c), there are snapshots of $\rho(t,x)$ at $t = (0.3, 0.2, 0.15, 0.1, 0.05)$ for the solution computed wirth (4.8) and (4.6) with $h = 0.25 \times 10^{-2}, \tau = 0.25 \times 10^{-4}$ (left) and $h = 0.125 \times 10^{-2}, \tau = 0.2 \times 10^{-4}$ (right). In (b), we show snapshots of $\rho(t,x)$ at $t = (0.3, 0.2, 0.15, 0.1, 0.05)$ for (4.8) with $h = 0.25 \times 10^{-2}, \tau = 1/3 \times 10^{-4}$ (left) and $h = 1/8 \times 10^{-2}, \tau = 1/2 \times 10^{-5}$ (right).

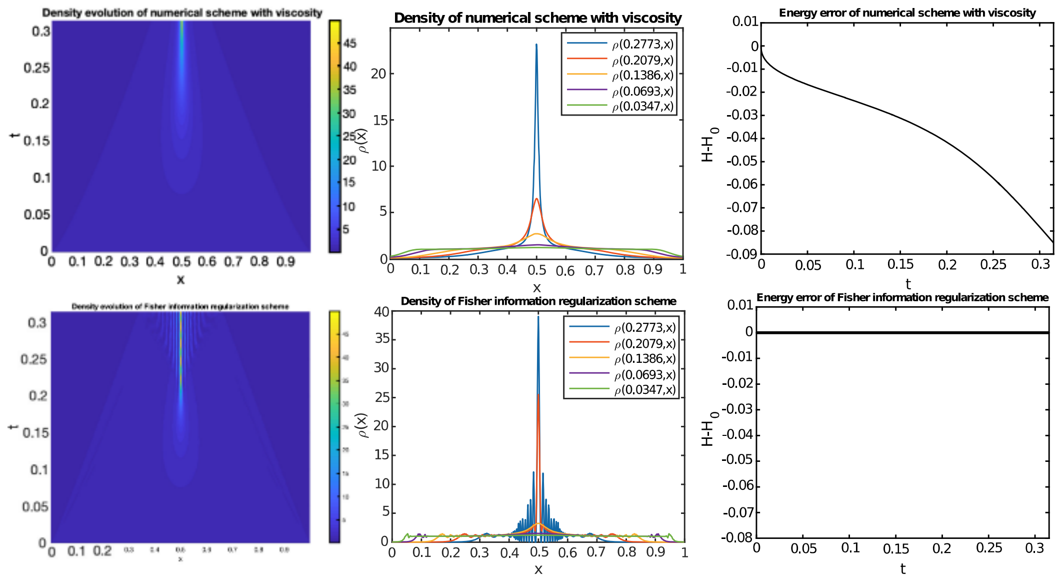


FIGURE 5.3. Example 5.1, case (II). Contour plot of $\rho(t,x)$ (left), snapshots of $\rho(t,x)$ at $t = (0.2773, 0.2079, 0.1386, 0.0693, 0.0347)$ and the energy error before $T = 0.315$ or the upwind scheme (4.10) with numerical viscosity (top) and the Fisher information regularization symplectic scheme (4.8) (bottom).

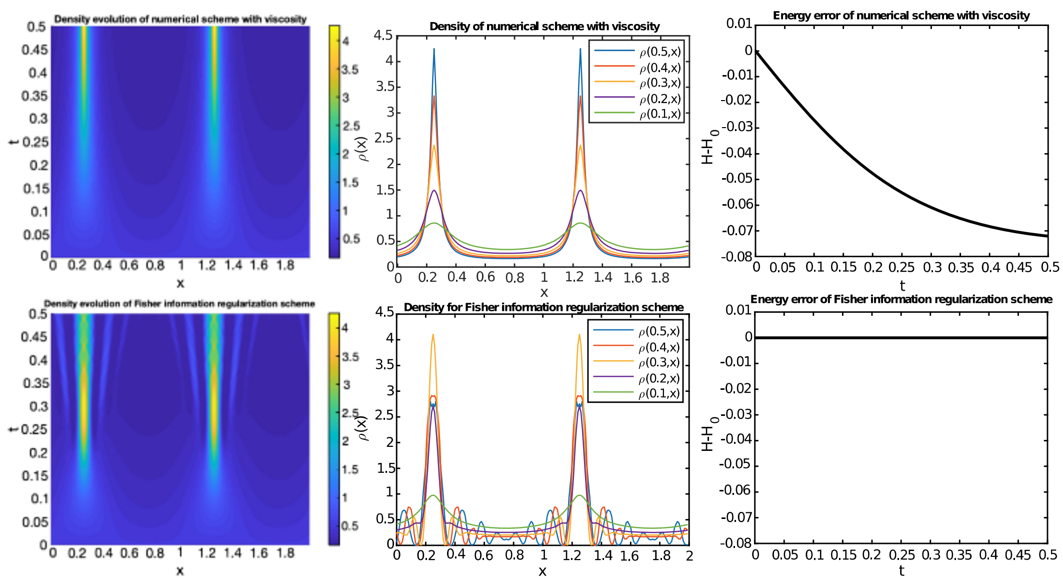


FIGURE 5.4. Example 5.1, case (III). Contour plot of $\rho(t,x)$ (left), snapshots of $\rho(t,x)$ at $t = (0.5, 0.4, 0.3, 0.2, 0.1)$ and the energy error before $T = 0.5$ for the upwind scheme (4.10) with numerical viscosity (top) and the Fisher information regularization symplectic scheme (4.8) (bottom).

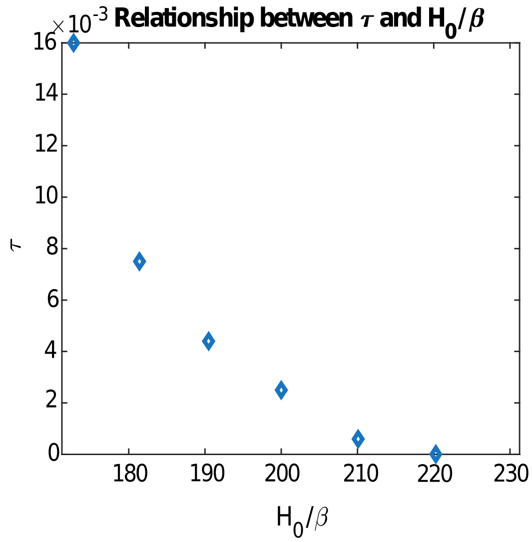


FIGURE 5.5. Example 5.1, $0 \le x \le 1$, $S_0(x) = \frac{\sin(\pi x)}{x}$, $\mu_0(x) = 1$, $T = 4$. Relationship between $\frac{H_0}{\beta}$ and the largest time step size τ for which (4.8) gives the correct behavior, for values of $\beta = 0.005788, 0.005513, 0.00525, 0.005, 0.00476, 0.00454$.

TABLE 1. Example 5.2. The L_∞-error between $\rho(t, x)$ produced by (4.8) and the density produced by the Hamilton–Jacobi solver.

β	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
L_∞-error	0.2312	0.1945	0.0940	0.021	0.0029	3.337×10^{-4}	6.291×10^{-5}

regularization scheme converges to the true solution. Following the approach in [2, 43], we take $\rho(1, x) = 1$ and $\rho(0, x) = 1 - \gamma \sin(2\pi x)(2\pi)^2$ with $\gamma = \frac{1}{8}(2\pi)^{-2}$. It can be verified that the minimizer of (1.6) is concentrated on the characteristic line $X(t, x) = x + \gamma cos(2\pi x)t, t \in [0.1]$. In this case, the potential function is $S(0, x) = \gamma \sin(2\pi x)$ up to a constant.

In Figure 5.7, we present the evolution of probability density function against different scales of the Fisher information, i.e., $\beta = 10^{-1}, \cdots, 10^{-6}$. In Table 1, we compare the numerical solutions of (4.8) to those produced by the Hamilton–Jacobi solver (4.10). In these experiments, all the parameters except β are kept fixed. It can be observed that when $\beta \to 0$, the difference between the two computed solutions vanishes. Still, from Figure 5.6 we observe that the Fisher information solution (obtained with $\beta = 10^{-6}$) maintains energy on the time interval $[0, 1]$ whereas for the viscosity solution there is a linear deterioration in energy preservation.

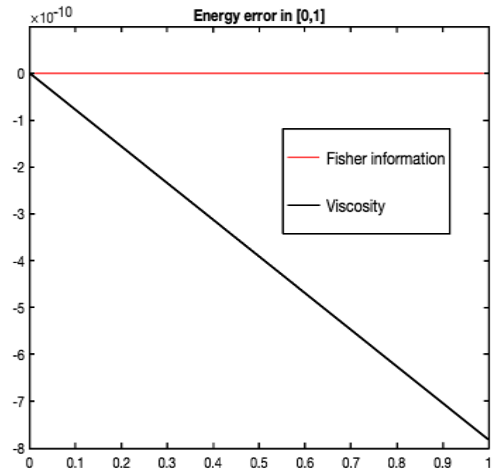


FIGURE 5.6. Example 5.2. Energy error between $\mathcal{H}(t)-\mathcal{H}(0)$ for Fisher information regularization scheme and viscosity regularization scheme.

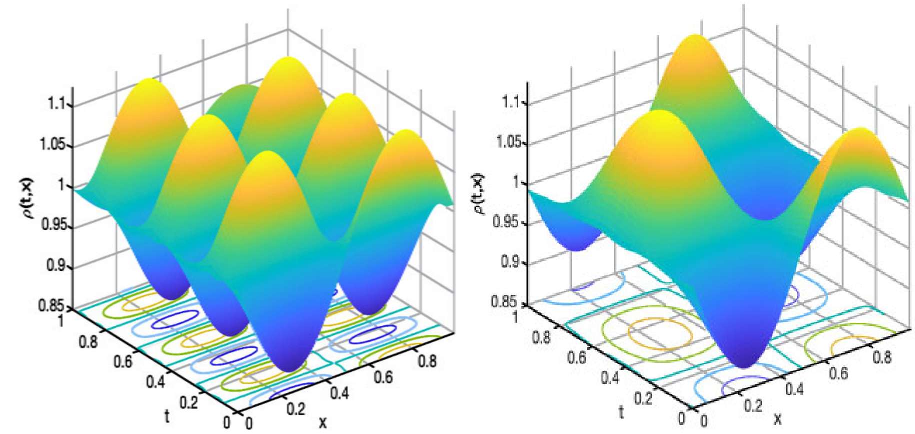
Example 5.3. Madelung system. Consider the Wasserstein–Hamiltonian system (2.12):

$$\frac{\partial}{\partial t}\rho(t,x)+\nabla\cdot(\rho(t,x)\nabla S(t,x))=0,$$
$$\frac{\partial}{\partial t}S(t,x)+\frac{1}{2}|\nabla S(t,x)|^2+\beta\frac{\partial}{\partial\rho(t,x)}I(\rho(t,\cdot))+\mathbb{V}(x)$$
$$+\int_{\mathcal{M}}W(x,y)|\rho(y)|^2dy=0.$$

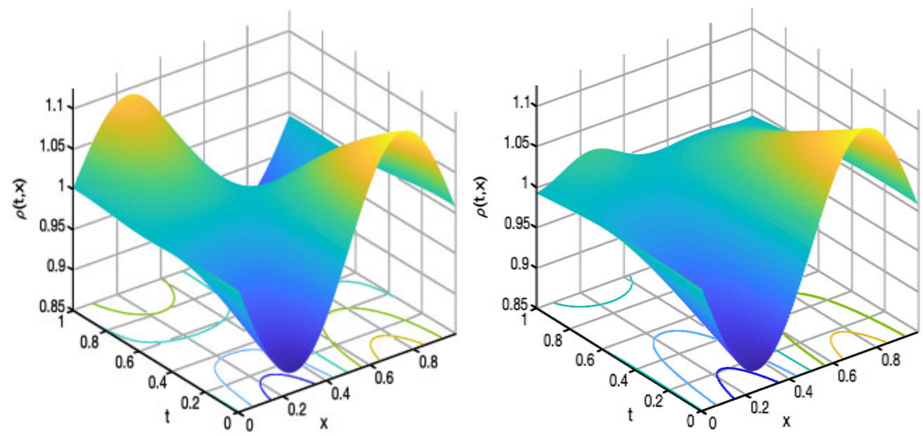
We use the scheme (4.8) for given $\beta>0$. Figure 5.8 shows the behaviors of ρ and S when $\mathbb{V}=0$ and $\mathbb{W}=0$, as well as the mass and energy evolution. Here, for the evolution of ρ and S, we choose $\beta=1$, $T=0.5$, $\tau=10^{-3}$, $h=10^{-2}$, $S^0(x)=\frac{1}{2}\sin(2\pi x)$, $\mu^0(x)=1$. We also plot the evolution of energy error $\mathcal{H}(t)-\mathcal{H}_0$ with $\mathcal{H}(t)=\mathcal{H}(\rho(t),S(t))$ and mass error up to $T=400$, which shows the good long time behavior of the proposed scheme. In Figure 5.9 and Figure 5.10, we present the behavior of ρ and S, as well as the mass and energy evolution, for $\mathbb{V}=1,\mathbb{W}=0$, and $\mathbb{V}=1,\mathbb{W}(x,y)=2\delta(x-y)$, respectively. It can be seen that the contour plot of S i is influenced by the linear and interaction potentials. The step sizes and initial value are chosen to be the same as in the case $\mathbb{V}=0,\mathbb{W}=0$. Note the very small scale in these figures. The preservation of desirable long time behavior for the schemes using the Fisher information as the regularization of choice is observed in all cases.

6. Discussion on the limit of vanishing regularization

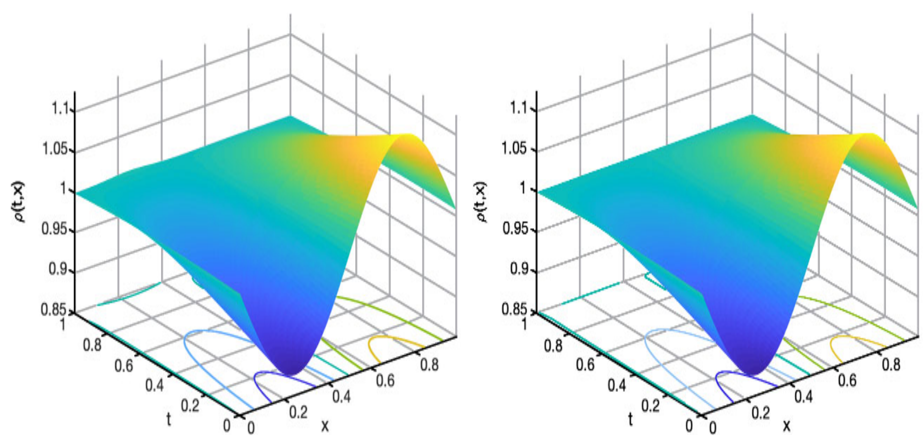
From Example 5.2, we observed that the solution obtained by using the Fisher information regularization scheme converges to the exact solution



(a) $\beta = 10^{-1}$ (left), 10^{-2} (right)

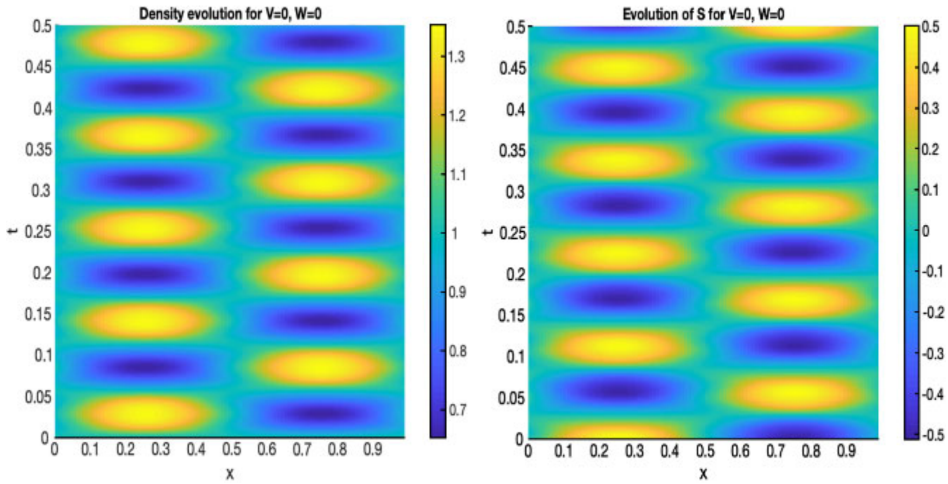


(b) $\beta = 10^{-3}$ (left), 10^{-4} (right)

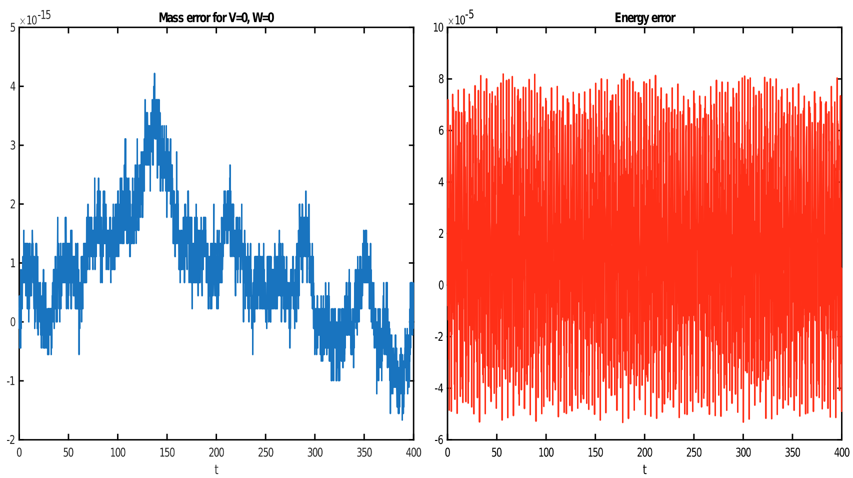


(c) $\beta = 10^{-5}$ (left), 10^{-6} (right)

FIGURE 5.7. Example 5.2. The plot of $\rho(t,x)$ by the Fisher information regularization symplectic scheme (4.8) with different $\beta = 10^{-1}, \cdots, 10^{-6}$.



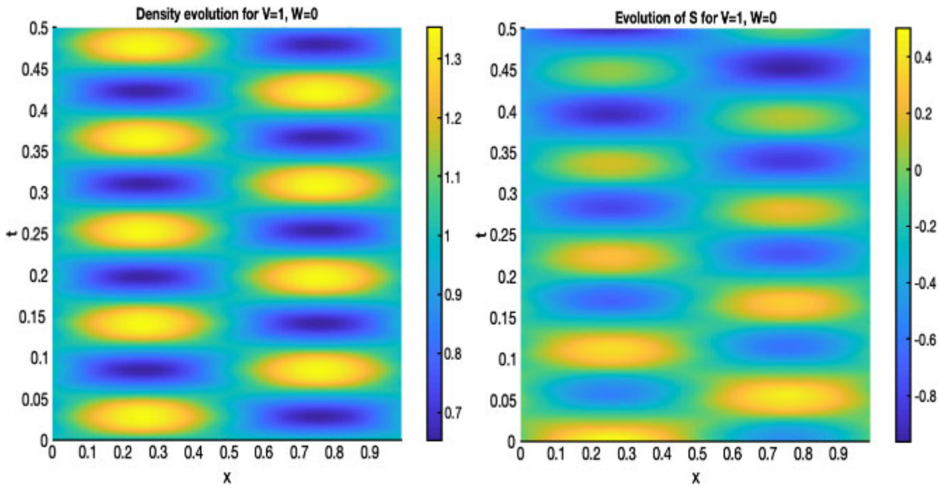
(a)



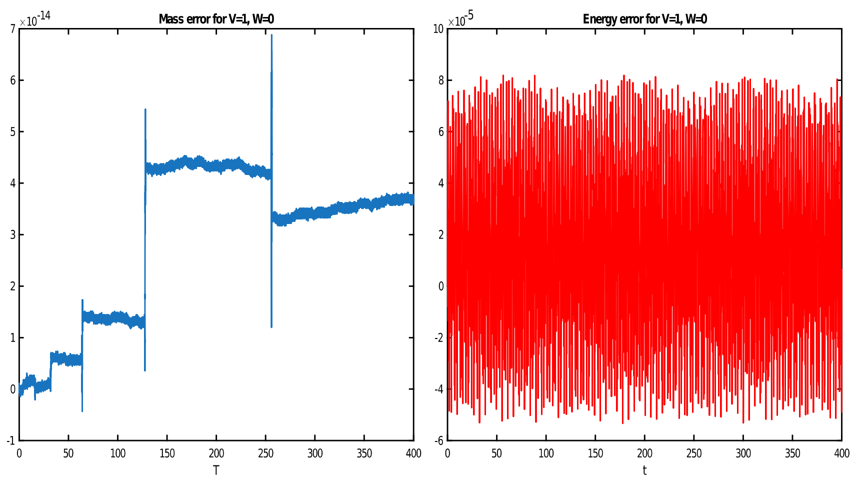
(b)

FIGURE 5.8. The evolutions of ρ and S ($\mathbb{V} = 1$, $\mathbb{W} = 0$) before $T = 0.5$ (a), the mass conservation law and the energy error before $T = 400$ (b). Note the extremely small scales in the plots.

as $\beta \to 0$ when the classical solution of (1.4) exists. On the other hand, Example 5.1 shows that the numerical solution of the Fisher information regularization scheme can be computed beyond the blow-up time of (1.4). It inspires us to theoretically study the behaviors of the density, via vanishing β in front of the Fisher information regularization even when the classical solution does not exists.



(a)

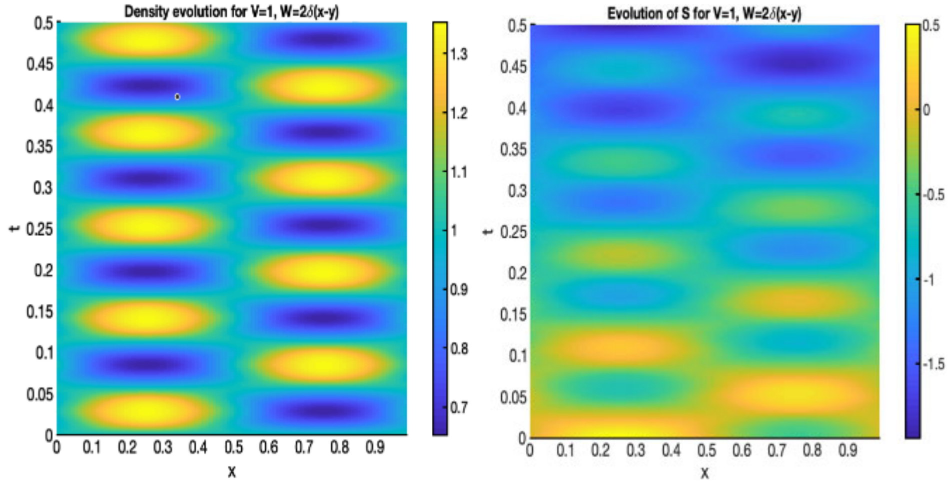


(b)

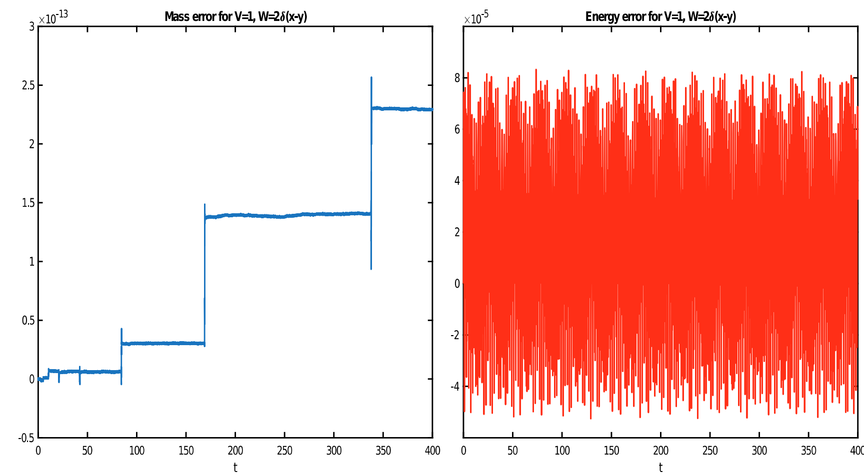
FIGURE 5.9. Example 5.3. Evolution of ρ and S ($\mathbb{V} = 1$, $\mathbb{W} = 2\delta(x-y)$) before $T = 0.5$ (a), the mass conservation law and the energy error before $T = 400$ (b).

Our first consideration is whether the weak limit of semi-discrete scheme (2.4) exists as $\beta \to 0$ or not. To simplify the discussion, we assume that $\mathbb{V} = 0$ and $\mathbb{W} = 0$. The corresponding continuous system is

$$\frac{\partial}{\partial t}\rho(t,x) + \nabla \cdot (\nabla S(t,x)\rho(t,x)) = 0,$$
$$\frac{\partial}{\partial t}S(t,x) + \frac{1}{2}|\nabla S(t,x)|^2 + \beta \frac{\delta}{\delta\rho(t,x)}\mathcal{I}(\rho(t,\cdot)) = 0.$$



(a)



(b)

FIGURE 5.10. Example 5.3. Evolutions of ρ and S ($\mathbb{V} = 1$, $\mathbb{W}(x, y) = 2\delta(x - y)$) before $T = 0.5$ (a), the mass conservation law and the energy error before $T = 400$ (b).

on $\mathcal{M} = [0, 1]$ with periodic boundary condition. Then (2.4) becomes

$$
\begin{aligned}
&\frac{d}{dt}\rho_i(t) + \frac{1}{h^2}(S_{i+1}(t) - S_i(t))\frac{\rho_i(t) + \rho_{i-1}(t)}{2} + \frac{1}{h^2}(S_{i-1}(t) \\
&\qquad - S_i(t))\frac{\rho_i(t) + \rho_{i+1}(t)}{2} = 0, \\
&\frac{d}{dt}S_i(t) + \frac{1}{4h^2}(S_i - S_{i+1})^2 + \frac{1}{4h^2}(S_i - S_{i+1})^2 + \frac{1}{4h^2}(S_i - S_{i+1})^2 \\
&\qquad + \beta\frac{\partial I}{\partial \rho_i}(\rho(t)) = 0,
\end{aligned}
$$

where the corresponding graph G is an one dimensional lattice graph, i.e. $(i,j)\in E$ if and only if $j=i-1$ or $i+1$, $N=\frac{1}{h}$. Here we take $\omega_{ij}=\widetilde{\omega}_{ij}=\frac{1}{h^2}$, $\theta_{ij}(\rho)=\frac{\rho_i+\rho_j}{2}$ and $\widetilde{\theta}_{ij}(\rho)=\frac{\rho_i-\rho_j}{\log(\rho_i)-\log(\rho_j)}$.

Assume that $\rho(0,x)>0$ and $S(0,x)$ are smooth functions,

$$\widetilde{K}(\rho(0,\cdot),S(0,\cdot))=\frac{1}{2}\int_{\mathcal{M}}|\nabla S(0,x)|^2\rho(0,x)dx<\infty.$$

To avoid confusion while emphasizing the β dependence, in this part we denote the density and potential function of (2.4) by $\rho^\beta(t)$ and $S^\beta(t)$. Recall that the discrete kinetic energy functional

$$\mathcal{K}(\rho,S)=\frac{1}{2}\sum_{i=1}^N|\delta_h^+S_i|^2\theta_{ii+1}(\rho)h$$

with $\delta_h^+S_i=\frac{S_{i+1}-S_i}{h}$. Then it follows that $\mathcal{K}(\mu^0,S^0)=\widetilde{K}(\rho(0,\cdot),S(0,\cdot))+\mathcal{O}(h)$ by Taylor expansion. Moreover, we define the density function on $\mathcal{M}$ by $\rho^{\beta,h}(x)=\theta_{ii+1}(\rho^\beta)=\frac{1}{2}(\rho_i^\beta+\rho_{i+1}^\beta)$ for $x\in(x(i),x(i{+}1)]$, and denote $\vartheta^{\beta,h}$ as its corresponding Borel measure, i.e. $\vartheta^{\beta,h}$ has density function $\rho^{\beta,h}$. Similarly, we can define a piecewise function $\widetilde{v}^{\beta,h}(x)=\delta_h^+S_i^\beta$ if $x\in(x(i),x(i{+}1)]$. Hence, the discrete flux function $\rho^{\beta,h}(x)\widetilde{v}^{\beta,h}(x)$ is the density function of a signed measure $E^{\beta,h}$.

Proposition 6.1. *For any $t>0$ and $h>0$, the sequences of measures with density functions $(\rho^{\beta,h}(t,x),\rho^{\beta,h}\widetilde{v}^{\beta,h}(t,x))_{\beta\le 1}$ produced by* (2.4) *have weak limits $(\widetilde{\vartheta}^h,\widetilde{E}^h)$ satisfying*

$$\int_{\mathcal{M}}d\widetilde{\vartheta}^h=1,\quad \int_{\mathcal{M}}\Big|\frac{d\widetilde{E}^h}{d\widetilde{\vartheta}^h}\Big|^2d\widetilde{\vartheta}^h\le 2\mathcal{K}(\mu^0,S^0).$$

Proof. See Appendix A.2. □

Another interesting question to consider the weak limits of $(\widetilde{\vartheta}^h,\widetilde{E}^h)_{h>0}$, as $h\to 0$. We can show that they exist. They are a nonnegative measure denoted by $\widetilde{\vartheta}$ and a signed measure $\widetilde{E}$ respectively. We can further derive the equation governing the time evolution of $\widetilde{\vartheta}$. Those results are stated in Proposition 6.2.

Proposition 6.2. *Under the same conditions of Proposition* (6.1)*, $(\widetilde{\vartheta}^{h_m}$, $\widetilde{E}^{h_m})$ weakly converges to $(\widetilde{\vartheta},\widetilde{E})$, as $h_m\to 0$, satisfying*

$$\frac{\partial\widetilde{\vartheta}}{\partial t}+\nabla\cdot(\widetilde{E})=0. \tag{6.1}$$

Proof. See Appendix A.2. □

However, the evolution equation for $\widetilde{E}$ is not known yet. This question and a few others are of interests theoretically. We feel they are beyond the scope of the current paper, and hope to report more findings in the future.

Appendix

A.1. **Lower bound of the density.**

Proof of Proposition 3.1. It suffices to find a constant $0 < c < \frac{1}{N}$ such that

$$\inf_{0\le \min_{i=1,\dots,N} \rho_i \le c} I(\rho) \ge \frac{M_0}{\beta}.$$

Since the graph is finite, we have that

$$\inf_{0\le \min_{i=1}^N \rho_i \le c} I(\rho) = \min_{i\le N} \inf_{0\le \rho_i\le c} I(\rho).$$

Due to the convexity of $I(\rho)$ on $0 \le \rho_i \le c$ for a fixed $i \le N$, and the fact that $I(\rho)$ approaches ∞ when ρ approaches the boundary of $\mathcal{P}_o(G)$, $I(\rho)$ takes the minimum at the boundary, i.e., $\inf_{0\le\rho_i\le c} I(\rho) = \inf_{\rho_i=c} I(\rho)$ on $\mathcal{P}_o(G)$. Because of the periodic boundary condition, without loss of generality we can assume that $\rho_1 = c$. By calculating the Hessian matrix of $I(\rho)$, we get for any $\sigma \ne 0$,

$$\begin{aligned}
\sigma^T \mathrm{Hess} I(\rho)\sigma &= \sum_{i=3}^{N-1} \Big(\frac{1}{\rho_i^2}\big(\rho_i + \rho_{i+1} + \rho_{i-1}\big)\Big)\sigma_i^2 \\
&\quad + \sum_{i=3}^{N-1} \Big(\frac{1}{\rho_i\rho_{i+1}}\big(\rho_i + \rho_{i+1}\big)\sigma_i\sigma_{i+1} + \frac{1}{\rho_i\rho_{i-1}}\big(\rho_i + \rho_{i-1}\big)\sigma_i\sigma_{i-1}\Big) \\
&\quad + \frac{1}{\rho_2^2}\big(2\rho_i + \rho_3 + c\big) + \frac{1}{\rho_2\rho_3}\big(\rho_2 + \rho_3\big) \\
&\quad + \frac{1}{\rho_N^2}\big(2\rho_N + \rho_{N-1} + c\big) + \frac{1}{\rho_N\rho_{N-1}}\big(\rho_N + \rho_{N-1}\big) \\
&= \sum_{i=2}^{N-1} (\rho_i + \rho_{i+1})\big(\frac{\sigma_i}{\rho_i} - \frac{\sigma_{i+1}}{\rho_{i+1}}\big)^2 + \frac{1}{\rho_2^2}\big(\rho_2 + c\big)\sigma_2^2 + \frac{1}{\rho_N^2}\big(\rho_N + c\big)\sigma_N^2 \\
&> 0,
\end{aligned}$$

which implies the strict convexity of $I(c, \cdot)$ on $\sum_{i=2}^N \rho_i = 1 - c$. Using the Lagrange multiplier technique on $I(c, \rho_2, \cdots, \rho_N) - \lambda\big(\sum_{i=2}^N \rho_i - 1 + c\big)$, we get that the unique minimum point satisfies

$$\begin{aligned}
&\phi\big(\frac{c}{\rho_2}\big) + \phi\big(\frac{\rho_3}{\rho_2}\big) = \lambda, \\
&\varphi\big(\frac{\rho_{i-1}}{\rho_i}\big) + \phi\big(\frac{\rho_{i+1}}{\rho_i}\big) = \lambda, \text{ if } 3 \le i \le N-1, \\
&\phi\big(\frac{\rho_{N-1}}{\rho_N}\big) + \phi\big(\frac{c}{\rho_N}\big) = \lambda,
\end{aligned} \tag{A.2}$$

where $\phi(t) = 1 - t - \log(t)$.

We claim that $\rho_{N-i+1} = \rho_{i+1}$, for $i = 1, \cdots, \frac{N-1}{2}$, if $N-1$ is even number. When $N - 1$ is odd, we have $\rho_{N-i+1} = \rho_{i+1}$, for $i = 1, \cdots, [\frac{N-1}{2}]$, where $[s]$ is the largest integer smaller than or equal to $s \in \mathbb{R}$. To prove this claim, it suffices to show that $\rho_2 = \rho_N$. Assume that $\rho_2 > \rho_N$, Due to the monotonicity of ϕ, we have $\frac{c}{\rho_2} < \frac{c}{\rho_N}$,

$$\frac{\rho_3}{\rho_2} > \frac{\rho_{N-1}}{\rho_N}, \frac{\rho_4}{\rho_3} > \frac{\rho_{N-2}}{\rho_{N-1}}, \cdots, \frac{\rho_{i+2}}{\rho_{i+1}} > \frac{\rho_{N-i}}{\rho_{N-i+1}}, \text{ for } 1 \le i \le [\frac{N-1}{2}]. \tag{A.3}$$

If $N-1$ is even, we obtain that

$$\phi(\frac{\rho_{\frac{N-1}{2}+2}}{\rho_{\frac{N-1}{2}+1}}) < \phi(\frac{\rho_{\frac{N-1}{2}+1}}{\rho_{\frac{N-1}{2}+2}}),$$

which leads to $\frac{\rho_{\frac{N-1}{2}+1}}{\rho_{\frac{N-1}{2}+2}} < \frac{\rho_{\frac{N-1}{2}+2}}{\rho_{\frac{N-1}{2}+1}}$, i.e., $\rho_{\frac{N-1}{2}+2} > \rho_{\frac{N-1}{2}+1}$. Thus, we can conclude from (A.3) that

$$\frac{\rho_N}{\rho_2} > \frac{\rho_{N-1}}{\rho_3} > \cdots > \frac{\rho_{\frac{N-1}{2}+2}}{\rho_{\frac{N-1}{2}+1}} > 1,$$

which contradicts the assumption $\rho_2 > \rho_N$. If $N-1$ is odd, similar arguments yield that

$$\phi(\frac{\rho_{[\frac{N-1}{2}]+2}}{\rho_{[\frac{N-1}{2}]+1}}) < \phi(\frac{\rho_{[\frac{N-1}{2}]+2}}{\rho_{[\frac{N-1}{2}]+3}}),$$

which implies that $\rho_{[\frac{N-1}{2}]+3} > \rho_{[\frac{N-1}{2}]+1}$. Thus from (A.3), we have that

$$\frac{\rho_N}{\rho_2} > \frac{\rho_{N-1}}{\rho_3} > \cdots > \frac{\rho_{[\frac{N-1}{2}]+3}}{\rho_{[\frac{N-1}{2}]+1}} > 1,$$

which contradicts the assumption $\rho_2 > \rho_N$. One can show that $\rho_2 < \rho_N$ is also impossible by the same arguments. As a consequence, $\rho_2 = \rho_N$. By further using (A.2), we immediately get $\rho_{N-i+1} = \rho_{i+1}$, for $i = 1, \cdots, [\frac{N-1}{2}]$.

Now, we are going to show that the extreme point possesses the monotonicity along the path starting from a_1. Indeed, ρ_{i+1} is increasing when $d_{1,i+1}$ is increasing for $i \le [\frac{N-1}{2}]$ if N is odd and for $i \le [\frac{N-1}{2}] + 1$ if N is even. We use Figure A.1 to illustrate these two different cases.

Step 1. The Lagrange multiplier $\lambda > 0$. Since $\lambda = 0$ if and only if $\rho_i = \frac{1}{N}$, then $I(\rho) = 0$ which contradicts the fact that $\inf_{0 \le \min_{i=1}^N \rho_i \le c} I(\rho) > 0$. Assume that $\lambda < 0$. Then (A.2), together with the symmetry $\rho_{i+1} = \rho_{N-i+1}, i = 1, \cdots, [\frac{N-1}{2}]$, implies that when $N-1$ is even, it holds that

$$\begin{aligned} &\phi(\frac{\rho_{i-1}}{\rho_i}) + \phi(\frac{\rho_{i+1}}{\rho_i}) = \lambda, \text{ if } 2 \le i \le \frac{N-1}{2} - 1, \\ &\phi(\frac{\rho_{\frac{N-1}{2}}}{\rho_{\frac{N-1}{2}+1}}) = \lambda. \end{aligned} \tag{A.4}$$

Since $\lambda < 0$, we obtain that

$$\rho_{\frac{N-1}{2}+1} < \rho_{\frac{N-1}{2}} < \cdots < \rho_2 < \rho_1 = c,$$

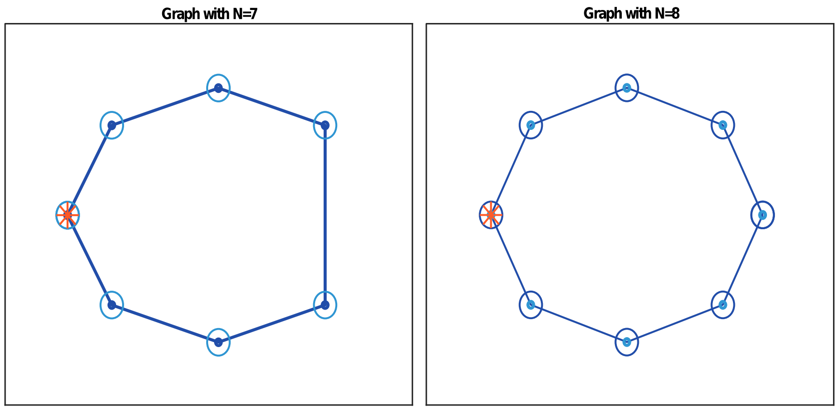


FIGURE A.1. The picture of the graph with $N = 7$ (left) and with $N = 8$ (right), where the red node represents v_1.

which contradicts the fact that $\sum_{i=2}^{N} \rho_i = 1 - c$. When $N-1$ is odd, then (A.2) and symmetry of ρ_i imply that

$$\text{(A.5)}\quad \begin{aligned} \phi(\frac{\rho_{i-1}}{\rho_i}) + \phi(\frac{\rho_{i+1}}{\rho_i}) = \lambda, \text{ if } 2 \le i \le [\frac{N-1}{2}], \\ 2\phi(\frac{\rho_{[\frac{N-1}{2}]+1}}{\rho_{[\frac{N-1}{2}]+2}}) = \lambda. \end{aligned}$$

Then we get $\rho_{[\frac{N-1}{2}]+2} < \rho_{[\frac{N-1}{2}]+1} < \cdots < \rho_2 < \rho_1 = c$, which is also not possible. Thus it holds that $\lambda > 0$. This indicates that

$$\rho_{[\frac{N-1}{2}]+2} > \rho_{[\frac{N-1}{2}]+1} > \cdots > \rho_2 > \rho_1 = c.$$

Step 2. $\frac{\rho_{i+1}}{\rho_i}$ is strictly decreasing. If $N-1$ is even, $\frac{\rho_{i+1}}{\rho_i}$ is strictly decreasing for $1 \le i \le [\frac{N-1}{2}]$. According to (A.4), it holds that

$$\begin{aligned} \phi(\frac{\rho_{\frac{N-1}{2}-1}}{\rho_{\frac{N-1}{2}}}) &= \lambda - \phi(\frac{\rho_{\frac{N-1}{2}+1}}{\rho_{\frac{N-1}{2}}}) = \phi(\frac{\rho_{\frac{N-1}{2}}}{\rho_{\frac{N-1}{2}+1}}) - \phi(\frac{\rho_{\frac{N-1}{2}+1}}{\rho_{\frac{N-1}{2}}}), \\ \phi(\frac{\rho_{\frac{N-1}{2}-i-1}}{\rho_{\frac{N-1}{2}-i}}) &= \lambda - \phi(\frac{\rho_{\frac{N-1}{2}-i+1}}{\rho_{\frac{N-1}{2}-i}}) = \phi(\frac{\rho_{\frac{N-1}{2}-i}}{\rho_{\frac{N-1}{2}-i+1}}) - \phi(\frac{\rho_{\frac{N-1}{2}-i+1}}{\rho_{\frac{N-1}{2}-i}}) \\ &\quad + \phi(\frac{\rho_{\frac{N-1}{2}-i+2}}{\rho_{\frac{N-1}{2}-i+1}}), \end{aligned}$$

where $i = 1, \cdots, \frac{N-5}{2}$. The monotonicity of ρ_i, $i \le \frac{N-1}{2}$, together with $\lambda > 0$, leads to

$$\phi(\frac{\rho_{\frac{N-1}{2}-i-1}}{\rho_{\frac{N-1}{2}-i}}) > \phi(\frac{\rho_{\frac{N-1}{2}-i}}{\rho_{\frac{N-1}{2}-i+1}}), \text{ for } i = 0, \cdots, \frac{N-5}{2}.$$

If $N-1$ is odd, $\frac{\rho_{i+1}}{\rho_i}$ is strictly decreasing for $1\le i\le [\frac{N-1}{2}]+1$. From (A.5), it follows that

$$\begin{aligned}\phi\Big(\frac{\rho_{[\frac{N-1}{2}]}}{\rho_{[\frac{N-1}{2}]+1}}\Big) &= \lambda-\phi\Big(\frac{\rho_{[\frac{N-1}{2}]+2}}{\rho_{[\frac{N-1}{2}]+1}}\Big)\\ &=2\phi\Big(\frac{\rho_{[\frac{N-1}{2}]+1}}{\rho_{[\frac{N-1}{2}]+2}}\Big)-\phi\Big(\frac{\rho_{[\frac{N-1}{2}]+2}}{\rho_{[\frac{N-1}{2}]+1}}\Big),\\ \phi\Big(\frac{\rho_{[\frac{N-1}{2}]-i-1}}{\rho_{[\frac{N-1}{2}]-i}}\Big) &= \lambda-\phi\Big(\frac{\rho_{[\frac{N-1}{2}]-i+1}}{\rho_{[\frac{N-1}{2}]-i}}\Big)\\ &=\phi\Big(\frac{\rho_{[\frac{N-1}{2}]-i}}{\rho_{[\frac{N-1}{2}]-i+1}}\Big)-\phi\Big(\frac{\rho_{[\frac{N-1}{2}]-i+1}}{\rho_{[\frac{N-1}{2}]-i}}\Big)+\phi\Big(\frac{\rho_{[\frac{N-1}{2}]-i+2}}{\rho_{[\frac{N-1}{2}]-i+1}}\Big),\end{aligned}$$

where $i=0,\cdots,[\frac{N-1}{2}]+1$. From the monotonicity of ϕ, it follows that $\frac{\rho_{i+1}}{\rho_i}$ is strictly decreasing for $1\le i\le [\frac{N-1}{2}]$.

Step 3. Lower bound for $\frac{\rho_{i+1}}{\rho_i}, i=1,\cdots,[\frac{N-1}{2}]$. We first deal with the case that $N-1$ is even. Due to monotonicity of $\frac{\rho_{i+1}}{\rho_i}$, its minimum is $k=\frac{\rho_{[\frac{N-1}{2}]+1}}{\rho_{[\frac{N-1}{2}]}}$. Since $\sum_{i=1}^N \rho_i=1$, we have

$$c+2\sum_{i=2}^{[\frac{N-1}{2}]}\rho_i=c\big(1+2\sum_{i=2}^{[\frac{N-1}{2}]}\frac{\rho_i}{c}\big)=1.$$

To find a lower bound of $\frac{\rho_{i+1}}{\rho_i}$, it suffices to find an upper bound of k such that

$$1+\sum_{i=2}^{[\frac{N-1}{2}]}k^{i-1}=\frac{k^{[\frac{N-1}{2}]}-1}{k-1}<\frac{1+c}{2c}.$$

Let $k\le \big(\frac{1-c}{2c[\frac{N-1}{2}]}\big)^{\frac{1}{[\frac{N-1}{2}]}}$. Then it holds that

$$\sum_{i=2}^{[\frac{N-1}{2}]+1}k^{i-1}\le [\frac{N-1}{2}]k^{[\frac{N-1}{2}]}\le \frac{1-c}{2c}.$$

Finally, we get that

$$\begin{aligned}\inf_{\rho_1=c}I(\rho)&=2\sum_{i=1}^{[\frac{N-1}{2}]}\big(\log(\rho_i^*)-\log(\rho_{i+1}^*)\big)(\rho_i^*-\rho_{i+1}^*)\\ &\ge 2\log\Big(\frac{\rho^*_{[\frac{N-1}{2}]+1}}{\rho^*_{[\frac{N-1}{2}]}}\Big)\big(\rho^*_{[\frac{N-1}{2}]+1}-c\big).\end{aligned}$$

Since there exists at least $\rho_j^*, j \le N$ such that $\rho_j^* > \frac{1-c}{N-1}$, thus it holds that

(A.6)

$$\inf_{\rho_1=c} I(\rho) \ge 2\log(\frac{\rho^*_{[\frac{N-1}{2}]+1}}{\rho^*_{[\frac{N-1}{2}]}})\big(\frac{1-c}{N-1}-c\big) \ge 2\log(k)\big(\frac{1-c}{N-1}-c\big) \ge \frac{M}{\beta}.$$

Now, we are able to show the desired lower bound estimate. If there exists

$$\frac{1}{2}\min_{i=1}^{N}\rho_i(0)N \le \alpha < \min_{i=1}^{N}\rho_i(0)N,\ c = \alpha\frac{1}{N},$$

such that $\inf_{\rho_1=c} I(\rho) \ge \frac{M}{\beta}$, then

$$\sup_{t\ge 0}\min_{i\le N}\rho_i(t) \ge \frac{1}{2}\min_{i\le N}\rho_i(0).$$

Otherwise, $c < \frac{1}{N}\alpha$, for $\alpha \le \frac{1}{2}\min_{i\le N}\rho_i(0)N$. From the estimate (A.6), it follows that if

$$c < \frac{1}{1+2[\frac{N-1}{2}]\exp(\frac{M(N-1)([\frac{N-1}{2}])}{2\beta(1-\alpha)})},$$

then $\inf_{\rho_1=c} I(\rho) > \frac{M}{\beta}$. Based on the above estimates, we have the following lower bound for ρ,

$$\sup_{t\ge 0}\min_{i\le N}\rho_i(t) \ge \frac{1}{1+2[\frac{N-1}{2}]\exp(\frac{M(N-1)[\frac{N-1}{2}]}{2\beta(1-\alpha)})} \ge \frac{1}{1+N\exp(\frac{M(N-1)[\frac{N-1}{2}]}{\beta})}.$$

Denote $\kappa_0 = \frac{1}{2}\min_{i=1}^{N}\rho_i(0)$ and

$$\kappa_1 = \frac{1}{1+2[\frac{N-1}{2}]\exp(\frac{M(N-1)[\frac{N-1}{2}]}{\beta})}.$$

Thus, it holds that

$$\sup_{t\ge 0}\min_{i\le N}\rho_i(t) \ge \min\big(\kappa_0,\kappa_1\big).$$

Similar arguments yield the estimate when $N-1$ is odd.

□

Proof of Proposition 3.2. We use an induction argument and similar techniques to those used in the proof of Proposition 3.1. Like the proof of Proposition 3.1, it suffices to find the largest $0<c<\frac{1}{N}$ such that $\inf_{0\le \min_{i=1}^N \rho_i \le c} I(\rho) \ge \frac{M}{\beta}$. Since the graph is finite and $I(\rho)$ is convex, we have that

$$\inf_{0\le \min_{i=1}^N \rho_i \le c} I(\rho) = \min_{i\le N}\inf_{\rho_i=c} I(\rho).$$

When $N=3$, then the graph only has two boundary nodes and we only need to consider the case that $\rho_1 = c$ and $\rho_2 = c$, due to the symmetry on boundary nodes.

When $\rho_1 = c$, the Lagrange multiplier method yields that the extreme point satisfies

$$\phi(\frac{c}{\rho_2}) + \phi(\frac{\rho_3}{\rho_2}) = \lambda, \quad \phi(\frac{\rho_2}{\rho_3}) = \lambda, \quad \phi(t) = 1 - t - \log(t) \ .$$

Then we get that $\lambda > 0$, $\rho_3 > \rho_2 > c$ and $\frac{c}{\rho_2} < \frac{\rho_2}{\rho_3}$. When $\rho_2 = c$, the Lagrange multiplier method yields that the extreme point satisfies

$$\phi(\frac{c}{\rho_1}) = \lambda, \quad \phi(\frac{c}{\rho_3}) = \lambda,$$

and so we obtain that $\lambda > 0$, $\rho_3 > c, \rho_1 > c$. From these, similarly to the proof of Proposition 3.1, we obtain

$$\sup_t \min_{i=1}^{3} \rho_i(t) \geq \min\Big(\frac{1}{2}\min_{i=1}^{3} \rho_i(0), \frac{1}{1 + 2\exp(4\frac{M}{\beta})}\Big).$$

Now we proceed with the induction steps. Assume that for the graph with $N-1$ nodes, if $\inf_{j \leq N-1} \inf_{|\rho_j| \leq c} I(\rho) = \inf_{\rho_i = c} I(\rho)$ for some i, then we get $\lambda > 0$ in the Lagrange multiplier method, and that for any path $a_{l_0}a_{l_1}a_{l_2}\cdots a_{l_m}$, $m \leq N-1$, starting from $a_{l_0} = a_i$ to a boundary point a_{l_m}, the probability density ρ_{l_j}, $0 \leq j \leq m$ is increasing and $\frac{\rho_{l_{j+1}}}{\rho_{l_j}}, 0 \leq j \leq m-1$, is decreasing. We are going to prove that the above statement also holds for the graph with N nodes. Let $\inf_{j \leq N} \inf_{|\rho_j| \leq c} I(\rho) = \inf_{\rho_i = c} I(\rho)$ for some i. Then either a_i is a boundary vertex of the the graph, or a_i is an interior vertex of the graph.

Case 1: a_i is an interior node of the graph. Assume that the number of edges connecting to a_i is n_i. By using the Lagrange multiplier method and taking the partial derivative with respect to ρ_j, $j \neq i$, we obtain $N-1$ equations. Since a_i is an interior node, these $N-1$ equations can be rewritten as n_i systems of equations which are related to n_i subgraphs sharing the same node a_i. Notice that the number of the nodes of each subgraphs is smaller than $N-1$. According to our induction assumption, it holds that $\lambda > 0$, for any path $a_{l_0}a_{l_1}a_{l_2}\cdots a_{l_m}, m \leq N-1$, from $a_{l_0} = a_i$ to a boundary point a_{l_m}, the probability density $\rho_{l_j}, 0 \leq j \leq m$ is increasing and $\frac{\rho_{l_{j+1}}}{\rho_{l_j}}, 0 \leq j \leq m-1$, is decreasing.

Case 2: a_i is a boundary node of the graph. By the Lagrange multiplier method, with $\phi(t) = 1 - t - \log(t)$, we obtain

$$\sum_{l \in N(j)} \phi(\frac{\rho_l}{\rho_j}) = \lambda, \text{ if } j \notin N(i),$$

$$\sum_{l \in N(j), l \neq i} \phi(\frac{\rho_l}{\rho_j}) + \phi(\frac{c}{\rho_j}) = \lambda, \text{ if } j \in N(i).$$

We first show that $\lambda > 0$. Assume that $\lambda \leq 0$. If ∂V has only two nodes, then by the monotonicity of ϕ, it holds that ρ is decreasing along the path $a_{l_0}a_{l_1}a_{l_2}\cdots a_{l_m}$ from a_{l_0} to any other node a_{l_m}. From the connectivity of the graph, we have $c \geq \rho_l, l \leq N$, which leads to the contradiction that

$\sum_{l=1}^{N} \rho_l = 1 \leq Nc < 1$. Similarly, ff ∂V has more than two nodes, we could also show that $\lambda > 0$ and the increasing property of ρ_{l_j} along the path $a_{l_0} a_{l_1} a_{l_2} \cdots a_{l_m}$, $m \leq N-1$ from $a_{l_0} = a_i$ to any boundary node $a_{l_m} \in \partial V$.

Next, we show the decreasing property of $\frac{\rho_{l_{j+1}}}{\rho_{l_j}}$. Notice that

$$\sum_{l \in N(l_1), l \neq i, l_2} \phi(\frac{\rho_l}{\rho_{l_1}}) + \phi(\frac{c}{\rho_{l_1}}) + \phi(\frac{\rho_{l_2}}{\rho_{l_1}}) = \lambda > 0,$$

$$\sum_{l \in N(l_j), l \neq l_{j-1}, l_j} \phi(\frac{\rho_l}{\rho_{l_j}}) + \phi(\frac{\rho_{l_{j-1}}}{\rho_{l_j}}) + \phi(\frac{\rho_{l_{j+1}}}{\rho_{l_j}}) = \lambda > 0, \ 2 \leq j \leq m-1,$$

$$\phi(\frac{\rho_{l_{m-1}}}{\rho_{l_m}}) = \lambda > 0.$$

The increasing property of ρ along any path from a_i to the node in ∂V yields that

$$\phi(\frac{\rho_{l_{m-2}}}{\rho_{l_{m-1}}}) = \lambda - \sum_{l \in N(l_j), l \neq l_m, l_{m-2}} \phi(\frac{\rho_l}{\rho_{l_{m-1}}}) - \phi(\frac{\rho_{l_m}}{\rho_{l_{m-1}}}) > \phi(\frac{\rho_{l_{m-1}}}{\rho_{l_m}}).$$

The monotonicity of ϕ leads to $\frac{\rho_{l_{m-2}}}{\rho_{l_{m-1}}} < \frac{\rho_{l_{m-1}}}{\rho_{l_m}}$. By repeating the above procedures on a_{l_j}, $1 \leq j \leq m-2$, we obtain that

$$\phi(\frac{\rho_{l_{j-1}}}{\rho_{l_j}}) + \phi(\frac{\rho_{l_j}}{\rho_{l_{j+1}}}) + \sum_{l \in N(l_j), l \neq l_{j-1}, l_{j+1}} \phi(\frac{\rho_l}{\rho_{l_j}}) + \phi(\frac{\rho_{l_{j+1}}}{\rho_{l_j}}) = \lambda + \phi(\frac{\rho_{l_j}}{\rho_{l_{j+1}}}).$$

Notice that $\phi(t) + \phi(1/t) \leq 0, t > 0$ and that $\phi(\frac{\rho_l}{\rho_{l_j}}) < 0$ when $l \neq j-1$. As a consequence, we get that

$$\phi(\frac{\rho_{l_{j-1}}}{\rho_{l_j}}) \geq \lambda + \phi(\frac{\rho_{l_j}}{\rho_{l_{j+1}}}),$$

which implies that $\frac{\rho_{l_{j+1}}}{\rho_{l_j}}, 0 \leq j \leq m-1$ is decreasing along the path from a_i to any node in ∂V. Thus the results holds for the graph with N nodes.

Now, we are going to derive the desired lower bound of the ρ_t. Assume that $\kappa \leq N-1$ is the number of nodes in ∂V and that d_{max} is largest distance $d_{i,l_m} \leq N - \kappa + 1$ from a_i to a_{l_m}. Since $\sum_{i=1}^{N} \rho_i = 1$, there exists at least a node a_n such that the density at $a_n > \frac{1-c}{N-1}$. Then for the path $a_{l_0} a_{l_1} \cdots a_{l_j} \cdots a_{l_m}, a_{l_0} = v_i, a_{l_m} \in \partial V, a_{l_j} = a_n$, $m \leq d_{max} - 1$, we have

$$\sum_{r=0}^{m} \rho_{l_r} = c(1 + \sum_{r=1}^{m} \frac{\rho_{l_r}}{c}).$$

Adding all the paths, which have a_i as a common node, together, we obtain

$$c(1 + \sum_{s=1}^{\kappa} \sum_{r=1}^{m_s} \frac{\rho_{l_r^s}}{c}) \geq 1.$$

To find a lower bound of the ratio of $\frac{\rho_{l^s_{r+1}}}{\rho_{l^s_r}}$ for all the paths, we denote $k = \min_{s\le \kappa} \frac{\rho_{l^s_{m_s}}}{\rho_{l^s_{m_s-1}}}$ and let

$$c(1+\sum_{s=1}^{\kappa}\sum_{r=1}^{m_s} k^r) < 1.$$

It suffices to require that

$$1+\kappa(d_{max}-1)k^{d_{max}-1} < \frac{1}{c}, \quad \text{i.e.,} \quad k \le (\frac{1-c}{c\kappa(d_{max}-1)})^{\frac{1}{d_{max}-1}}.$$

Thus it holds that

$$\min_{s\le\kappa} \frac{\rho_{l^s_{m_s}}}{\rho_{l^s_{m_s-1}}} \ge (\frac{1-c}{c\kappa(d_{max}-1)})^{\frac{1}{d_{max}-1}}, \text{ if } c \le \frac{1}{\kappa(d_{max}-1)+1}.$$

When $c \le \frac{1}{\kappa(d_{max}-1)+1}$, we get that for some path which contains the node a_{l_j} whose density is large than $\frac{1-c}{N-1}$,

$$\begin{aligned}\min_{i\le N}\inf_{\rho_i=c} I(\rho) &\ge \min_{i\le N}\inf_{\rho_i=c} \sum_{r=0}^{m^i_s-1} (\log(\rho^*_{l^i_r}) - \log(\rho^*_{l^i_{r+1}}))(\rho^*_{l^i_r} - \rho^*_{l^i_{r+1}})\\ &\ge \min_{i\le N}\inf_{\rho_i=c} \log(\frac{\rho^*_{m^i_s}}{\rho^*_{m^i_s-1}})(\rho_{m^i_s} - c).\\ &\ge \frac{1}{d_{max}-1}\log(\frac{1-c}{c\kappa(d_{max}-1)})(\frac{1-c}{N-1}-c).\end{aligned}$$

If there exists $\frac{1}{2}\min_i \rho_i(0)N \le \alpha < \min_i \rho_i(0)N, c = \alpha\frac{1}{N}$ such that $\inf_{\rho_1=c} I(\rho) \ge \frac{M_0}{\beta}$, then

$$\sup_{t\ge 0}\min_{i\le N} \rho_i(t) \ge \frac{1}{2}\min_i \rho_i(0).$$

Otherwise, taking

$$\begin{aligned}\frac{1}{d_{max}-1}\log(\frac{1-c}{c\kappa(d_{max}-1)})(\frac{1-c}{N-1}-c) \ge \frac{\mathcal{H}_0}{\beta},\\ c < \min(\alpha\frac{1}{N}, \frac{1}{(d_{max}-1)\kappa+1}),\end{aligned}$$

where $\alpha < \frac{1}{2}N\min_i \rho_i(0)$, we obtain the lower bound as

$$\sup_t \min_i \rho_i(t) \ge \frac{1}{1+\kappa(d_{max}-1)\exp(2\frac{M(d_{max}-1)(N-1)}{\beta})}.$$

Denote $\kappa_0 = \frac{1}{2}\min_i(\rho_i(0))$ and

$$\kappa_1 = \Big(1+\kappa(d_{max}-1)\exp(2\frac{M(d_{max}-1)(N-1)}{\beta})\Big)^{-1}.$$

Combining all cases above, we have the following lower bound estimate

$$\sup_t \min_i \rho_i(t) \geq \min\big(\kappa_0, \kappa_1\big).$$

□

A.2. **Proof of Propositions 6.1 and 6.2.**

Proof of Propositions 6.1. By using the energy-preserving property in Proposition 2.1 and the smoothness of $\rho(0,x)$ and $S(0,x)$, we have that for any $t>0$,

$$\begin{aligned}\mathcal{K}(\rho^\beta(t),S^\beta(t))+\beta I(\rho^\beta(t)) &= \mathcal{K}(\mu^0,S^0(0))+\beta I(\mu^0)\\ &\leq \widetilde{K}(\rho(0),S(0))+\beta\mathcal{I}(\rho(0))+C_{h,S(0),\rho(0)},\end{aligned}$$

where $C_{h,S^0,\rho^0}\to 0$ when $h\to 0$. Denote the spatial piecewise density on $\mathcal{M}$ by $\rho^{\beta,h}(t,x)$, $\rho^{\beta,h}(t,x)v^{\beta,h}(t,x)$. Thus when $h>0$ is fixed, by the mass conservation law, it follows that there exists a convergent subsequence $\{\rho^{\beta_n,h}\}_{n\geq 1}$ with a strong limit $\rho^h\in\mathcal{P}(G)$ when $\beta_n\to 0$ if $n\to\infty$. By choosing such subsequence, we have that

$$\begin{aligned}\int_{\mathcal{M}}\Big|\frac{\rho^{\beta_n,h}(x)v^{\beta_n h}(x)}{\rho^{\beta_n,h}(x)}\Big|^2\rho^{\beta_n,h}(x)dx &= \sum_{i=1}^N\frac{((\theta_{ii+1}(\rho^{\beta_n}))\delta_h^+S_i^{\beta_n})^2}{(\theta_{ii+1}(\rho^{\beta_n}))^2}\theta_{ii+1}(\rho^{\beta_n})h\\ &\leq \mathcal{K}(\rho(0),S(0))+\beta_n\mathcal{I}(\rho(0)),\end{aligned}$$

which implies that the signed measure $E^{\beta_n,h}$ with the density $\rho^{\beta_n,h}v^{\beta_n,h}$ is absolute continuous with respect to the measure $\vartheta^{\beta_n,h}$ with the density $\rho^{\beta_n,h}$ and its Radon–Nikodym derivative is mean square integrable with respect to $\vartheta^{\beta_n,h}$.

Now we are going to show the tightness of the signed measures $E^{\beta_n,h}$ with the density $\rho^{\beta_n,h}v^{\beta_n,h}$, i.e, for any $\epsilon>0$, there exists a compact set K_ϵ of $\mathcal{M}$ such that

$$\sup_n\Big|\int_{\mathcal{M}/K_\epsilon}\frac{\rho^{\beta_n,h}(x)v^{\beta_n,h}(x)}{\rho^{\beta_n,h}(x)}\rho^{\beta_n,h}(x)dx\Big|\leq\epsilon.$$

Here we omit the dependence on t since t is fixed. By using the conservation of the energy, Hölder inequality and the boundedness of $\mathcal{M}$, we have that

$$\Big|\int_{\mathcal{M}/K_\epsilon}\frac{\rho^{\beta_n,h}v^{\beta_n,h}}{\rho^{\beta_n,h}}\rho^{\beta_n,h}dx\Big|\leq\sqrt{\int_{\mathcal{M}/K_\epsilon}\Big|\frac{\rho^{\beta_n,h}v^{\beta_n,h}}{\rho^{\beta_n,h}}\Big|^2\rho^{\beta_n,h}dx}\sqrt{\int_{\mathcal{M}/K_\epsilon}\rho^{\beta_n,h}dx}.$$

The tightness of the $\vartheta^{\beta_n,h}$ implies the tightness of the the signed measures $E^{\beta_n,h}$. Thus there exists a weak limit $(\widetilde{\vartheta}^h, \widetilde{E}^h)$ of the subsequence of measures with densities $(\rho^{\beta_n,h}, \rho^{\beta_n,h}v^{\beta_n,h})_n$. Since the density functions of $\widetilde{\vartheta}^h$ and $\widetilde{E}^h$ are the linear combination of step functions, one can choose suitable subsequence such that the density of $\widetilde{\vartheta}^h$ and $\widetilde{E}^h$ is also strong limit of $\{(\rho^{\beta_n,h}, \rho^{\beta_n,h}v^{\beta_n,h})\}_{n\geq 1}$. By Fatou's lemma, we complete the proof. □

Proof of Proposition 6.2. Since $(\widetilde{\vartheta}^h)_h$ in the proof of Proposition 6.1 is a sequence of probability measures on $\mathcal{M}$, there exists a subsequence of $(\widetilde{\vartheta}^{h_m})_{h_m}$ which weakly converges to a probability measure $\widetilde{\vartheta}$ by the Riesz theorem and Alaoglu's theorem. By choosing such a subsequence, it is sufficient to show that $\widetilde{E}^{h_m}$ possesses a weakly convergent subsequence. By the Hölder inequality, for any subset K of $\mathcal{M}$,

$$\int_{\mathcal{M}/K} |\frac{d\widetilde{E}^{h_m}}{d\widetilde{\vartheta}^{h_m}}|\widetilde{\vartheta}^{h_m}(dx) \leq \sqrt{\int_{\mathcal{M}/K} |\frac{d\widetilde{E}^{h_m}}{d\widetilde{\vartheta}^{h_m}}|^2\widetilde{\vartheta}^{h_m}(dx)}\sqrt{\int_{\mathcal{M}/K} \widetilde{\vartheta}^{h_m}(dx)}.$$

The weak convergence of $(\widetilde{\vartheta}^{h_m})_{h_m}$ implies the tightness of $(\widetilde{\vartheta}^{h_m})_{h_m}$, and thus leads to the tightness of $\widetilde{E}^{h_m}$, which yields the existence of a weak limit $\widetilde{E}$.

According to (2.4), we have that

$$\frac{\partial}{\partial t}\Big(\frac{\rho_i^\beta(t)+\rho_{i+1}^\beta(t)}{2}\Big) = -\frac{1}{2h}\Big(\delta_h^+ S_{i+1}^\beta\big(\frac{\rho_{i+1}^\beta(t)+\rho_{i+2}^\beta(t)}{2}\big) - \delta_h^+ S_{i-1}^\beta\big(\frac{\rho_{i-1}^\beta(t)+\rho_i^\beta(t)}{2}\big)\Big). \tag{A.7}$$

By the definition of the piecewise density function $\rho^{\beta,h}$, we get that

$$\begin{aligned}&\frac{\partial}{\partial t}\rho^{\beta,h}(t,x)\chi_{(x(i),x(i+1)]}\\ &= -\frac{1}{2h}\Big(\delta_h^+ S_{i+1}^\beta \frac{\rho_{i+1}^\beta(t)+\rho_{i+2}^\beta(t)}{2} - \delta_h^+ S_{i-1}^\beta \frac{\rho_{i-1}^\beta(t)+\rho_i^\beta(t)}{2}\Big)\chi_{(x(i),x(i+1)]},\end{aligned}$$

where χ is the indicator function. Next, taking the inner product with a smooth test function ϕ on both sides of the above equation and using the

Taylor expansion, we obtain that

$$\begin{aligned}
&\frac{\partial}{\partial t}\int_{\mathcal{M}}\rho^{\beta,h}(t,x)\phi(x)dx\\
&=\frac{\partial}{\partial t}\sum_{i=1}^{N}\int_{x(i)}^{x(i+1)}\rho^{\beta,h}(t,x)\phi(x)dx\\
&=\sum_{i=1}^{N}\int_{x(i)}^{x(i+1)}-\frac{1}{2h}\Big(\delta_h^{+}S_{i+1}^{\beta}\big(\frac{\rho_{i+1}^{\beta}(t)+\rho_{i+2}^{\beta}(t)}{2}\big)\\
&\qquad\qquad -\delta_h^{+}S_{i-1}^{\beta}\big(\frac{\rho_{i-1}^{\beta}(t)+\rho_{i}^{\beta}(t)}{2}\big)\Big)\phi(x)dx\\
&=\sum_{i=1}^{N}-\frac{1}{2h}\delta_h^{+}S_{i+1}^{\beta}(t)\frac{\rho_{i+1}^{\beta}(t)+\rho_{i+2}^{\beta}(t)}{2}\Big(\int_{x(i)}^{x(i+1)}\phi(x)dx-\int_{x(i+2)}^{x(i+3)}\phi(x)dx\Big)\\
&=\sum_{i=1}^{N}\int_{x(i)}^{x(i+1)}\delta_h^{+}S_{i+1}^{\beta}(t)\frac{\rho_{i+1}^{\beta}(t)+\rho_{i+2}^{\beta}(t)}{2}\nabla\phi(x)dx+\mathcal{O}(h)\\
&=\int_{\mathcal{M}}\rho^{\beta,h}(t,x)v^{\beta,h}(t,x)\nabla\phi(x)dx+\mathcal{O}(h).
\end{aligned}$$

Taking the subsequence β_n, h_m to 0 in the proof Proposition 6.1, by the weak convergence of ϑ^{β_n,h_m} and $\widetilde{E}^{h_m}$, we obtain (6.1) in weak sense. □

References

[1] L. Ambrosio and W. Gangbo, *Hamiltonian ODEs in the Wasserstein space of probability measures*, Comm. Pure Appl. Math. **61** (2008), no. 1, 18–53, DOI 10.1002/cpa.20188. MR2361303

[2] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), no. 3, 375–393, DOI 10.1007/s002110050002. MR1738163

[3] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, *Iterative Bregman projections for regularized transportation problems*, SIAM J. Sci. Comput. **37** (2015), no. 2, A1111–A1138, DOI 10.1137/141000439. MR3340204

[4] A. Bensoussan, J. Frehse, and P. Yam, *Mean Field Games and Mean Field Type Control Theory*, SpringerBriefs in Mathematics, Springer, New York, 2013, DOI 10.1007/978-1-4614-8508-7. MR3134900

[5] A. Bensoussan, J. Frehse, and S. C. P. Yam, *The master equation in mean field theory*, J. Math. Pures Appl. (9) **103** (2015), no. 6, 1441–1474, DOI 10.1016/j.matpur.2014.11.005. MR3343705

[6] M. Bongini, M. Fornasier, F. Rossi, and F. Solombrino, *Mean-field Pontryagin maximum principle*, J. Optim. Theory Appl. **175** (2017), no. 1, 1–38, DOI 10.1007/s10957-017-1149-5. MR3707918

[7] B. Bonnet and H. Frankowska, *Necessary Optimality Conditions for Optimal Control Problems in Wasserstein Spaces*, Appl. Math. Optim. **84** (2021), no. suppl. 2, 1281–1330, DOI 10.1007/s00245-021-09772-w. MR4356896

[8] B. Bonnet and F. Rossi, *The Pontryagin maximum principle in the Wasserstein space*, Calc. Var. Partial Differential Equations **58** (2019), no. 1, Paper No. 11, 36, DOI 10.1007/s00526-018-1447-2. MR3881882

[9] P. Cardaliaguet. *Notes on mean-field games*, Available at `https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf`, 2012.

[10] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions, *The Master Equation and the Convergence Problem in Mean Field Games*, Annals of Mathematics Studies, vol. 201, Princeton University Press, Princeton, NJ, 2019, DOI 10.2307/j.ctvckq7qf. MR3967062

[11] G. Cavagnari, A. Marigonda, and B. Piccoli, *Generalized dynamic programming principle and sparse mean-field control problems*, J. Math. Anal. Appl. **481** (2020), no. 1, 123437, 45, DOI 10.1016/j.jmaa.2019.123437. MR4008538

[12] S.-N. Chow, W. Li, and H. Zhou, *A discrete Schrödinger equation via optimal transport on graphs*, J. Funct. Anal. **276** (2019), no. 8, 2440–2469, DOI 10.1016/j.jfa.2019.02.005. MR3926122

[13] S.-N. Chow, W. Li, and H. Zhou, *Wasserstein Hamiltonian flows*, J. Differential Equations **268** (2020), no. 3, 1205–1219, DOI 10.1016/j.jde.2019.08.046. MR4029003

[14] S.-N. Chow, W. Huang, Y. Li, and H. Zhou, *Fokker-Planck equations for a free energy functional or Markov process on a graph*, Arch. Ration. Mech. Anal. **203** (2012), no. 3, 969–1008, DOI 10.1007/s00205-011-0471-6. MR2928139

[15] M. Cirant and L. Nurbekyan, *The variational structure and time-periodic solutions for mean-field games systems*, Minimax Theory Appl. **3** (2018), no. 2, 227–260. MR3882531

[16] M. G. Crandall and P.-L. Lions, *Two approximations of solutions of Hamilton-Jacobi equations*, Math. Comp. **43** (1984), no. 167, 1–19, DOI 10.2307/2007396. MR744921

[17] J. Cui, L. Dieci, and H. Zhou, *A continuation multiple shooting method for Wasserstein geodesic equation*, `arXiv:2105.09502`, 2021.

[18] J. Cui, S. Liu, and H. Zhou, *What is a stochastic Hamiltonian process on finite graph? An optimal transport answer*, J. Differential Equations **305** (2021), 428–457, DOI 10.1016/j.jde.2021.10.009. MR4330908

[19] W. E, J. Han, and Q. Li, *A mean-field optimal control formulation of deep learning*, Res. Math. Sci. **6** (2019), no. 1, Paper No. 10, 41, DOI 10.1007/s40687-018-0172-y. MR3891852

[20] B. R. Frieden, *Physics from Fisher Information*, Cambridge University Press, Cambridge, 1998. A unification, DOI 10.1017/CBO9780511622670. MR1676801

[21] W. Gangbo, H. K. Kim, and T. Pacini, *Differential forms on Wasserstein space and infinite-dimensional Hamiltonian systems*, Mem. Amer. Math. Soc. **211** (2011), no. 993, vi+77, DOI 10.1090/S0065-9266-2010-00610-0. MR2808856

[22] W. Gangbo, T. Nguyen, and A. Tudorascu, *Hamilton-Jacobi equations in the Wasserstein space*, Methods Appl. Anal. **15** (2008), no. 2, 155–183, DOI 10.4310/MAA.2008.v15.n2.a4. MR2481677

[23] W. Gangbo and A. Święch, *Existence of a solution to an equation arising from the theory of mean field games*, J. Differential Equations **259** (2015), no. 11, 6573–6643, DOI 10.1016/j.jde.2015.08.001. MR3397332

[24] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963. Revised English edition translated and edited by Richard A. Silverman. MR0160139

[25] J. L. Gross and J. Yellen, editors. *Handbook of graph theory.* Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, 2004.

[26] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration*, 2nd ed., Springer Series in Computational Mathematics, vol. 31, Springer-Verlag, Berlin, 2006. Structure-preserving algorithms for ordinary differential equations. MR2221614

[27] J.-F. Jabir, D. Šiška, and L. Szpruch, *Mean-field neural odes via relaxed optimal control.* `arXiv:1912.05475`, 2019.

[28] C. Jimenez, A. Marigonda, and M. Quincampoix, *Optimal control of multiagent systems in the Wasserstein space*, Calc. Var. Partial Differential Equations **59** (2020), no. 2, Paper No. 58, 45, DOI 10.1007/s00526-020-1718-6. MR4073204

[29] S. Jin, P. Markowich, and C. Sparber, *Mathematical and computational methods for semiclassical Schrödinger equations*, Acta Numer. **20** (2011), 121–209, DOI 10.1017/S0962492911000031. MR2805153

[30] J. D. Lafferty, *The density manifold and configuration space quantization*, Trans. Amer. Math. Soc. **305** (1988), no. 2, 699–741, DOI 10.2307/2000885. MR924776

[31] J.-M. Lasry and P.-L. Lions, *Mean field games*, Jpn. J. Math. **2** (2007), no. 1, 229–260, DOI 10.1007/s11537-007-0657-8. MR2295621

[32] C. Léonard, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst. **34** (2014), no. 4, 1533–1574, DOI 10.3934/dcds.2014.34.1533. MR3121631

[33] W. Li, J. Lu, and L. Wang, *Fisher information regularization schemes for Wasserstein gradient flows*, J. Comput. Phys. **416** (2020), 109449, 24, DOI 10.1016/j.jcp.2020.109449. MR4107049

[34] W. Li, P. Yin, and S. Osher, *Computations of optimal transport distance with Fisher information regularization*, J. Sci. Comput. **75** (2018), no. 3, 1581–1595, DOI 10.1007/s10915-017-0599-0. MR3798113

[35] J. Lott, *Some geometric calculations on Wasserstein space*, Comm. Math. Phys. **277** (2008), no. 2, 423–437, DOI 10.1007/s00220-007-0367-3. MR2358290

[36] J. Maas, *Gradient flows of the entropy for finite Markov chains*, J. Funct. Anal. **261** (2011), no. 8, 2250–2292, DOI 10.1016/j.jfa.2011.06.009. MR2824578

[37] E. Madelung, *Quanten theorie in hydrodynamischer form*, Z. Phys. **40** (1927), no. 3–4, 322–326.

[38] E. Nelson, *Derivation of the Schrödinger equation from newtonian mechanics*, Phys. Rev. **150** (1966), 1079–1085.

[39] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174, DOI 10.1081/PDE-100002243. MR1842429

[40] M. Pavon, *Quantum Schrödinger Bridges*, Directions in Mathematical Systems Theory and Optimization, Lect. Notes Control Inf. Sci., vol. 286, Springer, Berlin, 2003, pp. 227–238, DOI 10.1007/3-540-36106-5_17. MR2014794

[41] G. Peyré, *Entropic approximation of Wasserstein gradient flows*, SIAM J. Imaging Sci. **8** (2015), no. 4, 2323–2351, DOI 10.1137/15M1010087. MR3413589

[42] N. Pogodaev and M. Staritsyn, *Impulsive control of nonlocal transport equations*, J. Differential Equations **269** (2020), no. 4, 3585–3623, DOI 10.1016/j.jde.2020.03.007. MR4097258

[43] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Progress in Nonlinear Differential Equations and their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling, DOI 10.1007/978-3-319-20828-2. MR3409718

[44] E. Schrödinger, *Uber die Umkehrung der Naturgesetze*, Sitzungsber. Preuss Akad. Wiss. Phys. Math. **144** (1931), 144–153.

[45] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003, DOI 10.1090/gsm/058. MR1964483

[46] C. Villani, *Optimal Transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009. Old and new, DOI 10.1007/978-3-540-71050-9. MR2459454

[47] M.-K. von Renesse, *An optimal transport view of Schrödinger's equation*, Canad. Math. Bull. **55** (2012), no. 4, 858–869, DOI 10.4153/CMB-2011-121-9. MR2994690

School of Mathematics, Georgia Tech, Atlanta, Georgia 30332
Current address: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
Email address: `jianbo.cui@polyu.edu.hk`

School of Mathematics, Georgia Tech, Atlanta, Georgia 30332
Email address: `dieci@math.gatech.edu`

School of Mathematics, Georgia Tech, Atlanta, Georgia 30332
Email address: `hmzhou@math.gatech.edu`