Available online at www.sciencedirect.com

ScienceDirect

Journal of Differential Equations 305 (2021) 428–457

Journal of Differential Equations

www.elsevier.com/locate/jde

What is a stochastic Hamiltonian process on finite graph? An optimal transport answer

Jianbo Cui [a,b], Shu Liu [a,*], Haomin Zhou [a]

[a] *School of Mathematics, Georgia Tech, Atlanta, GA 30332, USA*
[b] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*

Received 8 March 2021; revised 31 July 2021; accepted 5 October 2021
Available online 25 October 2021

Abstract

We present a definition of stochastic Hamiltonian process on finite graph via its corresponding density dynamics in Wasserstein manifold. We demonstrate the existence of stochastic Hamiltonian process in many classical discrete problems, such as the optimal transport problem, Schrödinger equation and Schrödinger bridge problem (SBP). The stationary and periodic properties of Hamiltonian processes are also investigated in the framework of SBP.

MSC: 35R02; 60J20; 53C22; 49Q22

Keywords: Wasserstein-Hamiltonian flow; Schrödinger bridge problem; Optimal transport; Time-inhomogeneous Markov process

1. Introduction

Hamiltonian systems, including both ordinary or partial differential equations (ODEs or PDEs respectively), are ubiquitous in applications. Their mathematical studies have a long and rich history (see e.g., [25,1,21]). Traditionally, the ambient space on which to define a Hamiltonian system is continuous, such as Euclidean space $\mathbb{R}^n$ or smooth manifolds like torus $\mathbb{T}^2$. What is

* Corresponding author.
E-mail addresses: jcui82@gatech.edu, jianbo.cui@polyu.edu.hk (J. Cui), sliu459@math.gatech.edu (S. Liu), hmzhou@math.gatech.edu (H. Zhou).

https://doi.org/10.1016/j.jde.2021.10.009

a Hamiltonian process if the underlying space becomes discrete, such as a finite graph? This is the question that we would like to explore within the framework of optimal transport (OT) in this study.

Our motivation to consider this question is 3-fold. Curiosity is at the first place. Secondly, the notion of gradient flow on graph has been investigated extensively using OT theory (see e.g. [19,7] and references therein). For example, an irreducible and reversible continuous time Markov chain on graph can be viewed as the gradient flow of entropy with respect to the discrete Wasserstein metric [19]. Naturally, we are inspired to ask whether the concept of Hamiltonian process on graph exists or not. To the best of our knowledge, the Hamiltonian mechanics on graph has not been explored yet. Finally and most importantly, recent developments in several practical problems, which can be defined in both continuous and discrete spaces, demonstrate Hamiltonian principles on graph. They are

(i) the OT problem (see e.g. [30]),

$$W_2^2(\rho_0,\rho_1)=\inf_v\Big\{\int_0^1 \mathbb{E}[|\dot X_t|^2]dt\ :\ \dot X_t=v(t,X_t),\, X_0\sim\rho^0,\, X_1\sim\rho^1\Big\}, \tag{1}$$

(ii) the SBP (see e.g. [26]),

$$\inf_v\Big\{\int_0^1 \frac{1}{2}\mathbb{E}[|v(t,X_t)|^2]dt : \dot X_t=v(t,X_t)+\sqrt{\hbar}\dot B_t,\ X_0\sim\rho^0,\ X_1\sim\rho^1\Big\} \tag{2}$$

and (iii) the Schrödinger equation (see e.g. [23,20,9]),

$$\inf_v\Big\{\int_0^T \frac{1}{2}\mathbb{E}[|\dot X_t|^2]dt : \dot X_t=v(t,X_t)+\sqrt{\hbar}\dot B_t,\ X_0\sim\rho^0,\ X_1\sim\rho^1\Big\}. \tag{3}$$

The above formulations are presented in Euclidean space where $v\in\mathbb{R}^d$ can be any smooth vector field, X_t is a stochastic process with prescribed probability densities ρ^0 and ρ^1 at time 0 and 1 respectively, B_t is the standard Brownian motion and $\hbar>0$ is a constant.

A common property shared by these problems is that their critical points obey the Hamiltonian principle. For instance, the minimizer of OT problem (1) satisfies a Hamiltonian PDE with the Hamiltonian $H(x,v,t)=\frac{1}{2}|v|^2$ (see e.g. [2]). The minimizer of SBP (2) is the solution of a Hamiltonian PDE with $H(x,v)=\frac{1}{2}|v|^2-\frac{1}{8}\hbar\frac{\delta}{\delta\rho}I(\rho)(t,x)$ where the Fisher information $I(\rho)=\int_{\mathbb{R}^d}|\nabla\log\rho(x)|^2\rho(x)dx$ (see e.g. [24,17]). Needless to say, the critical point of (3) satisfies the Schrödinger equation, which is a well-known Hamiltonian system. The problems stated in (1), (2) and (3) can be posed, with nominal changes, on a graph, and the density functions of their critical points have been studied on the Wasserstein manifold (see [14], [18,8], [9]) showing that they satisfy Hamiltonian ODEs. Based on those results, we investigate the properties of stochastic process $X(t)$ and provide an answer to the question in the title of the paper within the OT framework.

Defining Hamiltonian process on graph must face several intrinsic difficulties. The most obvious one is that $X(t)$ is a stochastic process jumping from node to node on the graph, while its continuous space counterpart trajectory is a spatial-temporal continuous function. Another challenge is about characteristic line. In fact, it is not clear how to define characteristic on graph. Furthermore, there is no reported result about examining whether a stochastic process, such as discrete OT and SBP, can preserve Hamiltonian along its trajectory, just like a classical Hamiltonian system does in continuous space.

To fill the gaps on finite graph, our idea is lifting the process on graph into a motion on its density manifold. To be more precise, we define the Hamiltonian process by a random process whose density and generators of instantaneous transition rate matrix form a Wasserstein Hamiltonian flow on the cotangent bundle of density manifold. Meanwhile, we show that such defined Hamiltonian processes exist in numerous practical problems, such as the discrete OT problem and SBP. Two important classes of Hamiltonian processes, namely the stationary Hamiltonian process and the periodic Hamiltonian process, are also discussed via the framework of SBP. They correspond to the invariant measure and the periodic solution of the Hamiltonian flow on the density space. We would like to mention that the Wasserstein Hamiltonian flow is firstly studied by Nelson's mechanics (see e.g. [23,4]). It is also pointed out that the Hamiltonian flows in density space are probability transition equations of classical Hamiltonian ODEs (see [30,10] and references therein).

There are several works with titles related to Hamiltonian systems on graphs, like the port-Hamiltonian system on graphs (see e.g. [28,29] and the references therein). Our current work is different from them. The port-Hamiltonian systems are the generalization of classical Hamiltonian system which describes the dynamics in interaction with control units, energy dissipating or energy storing units. The graph structure is used to characterize the interaction of the systems with ports, and their underlying phase variables are still in continuous spaces, like $\mathbb{R}^d$ or smooth manifold.

This paper is organized as follows. In section 2, we use the discrete optimal transport problem as the motivation of studying the Hamiltonian process on finite graph. In section 3, we present the definition and several properties of the Hamiltonian process on graph. In section 4, we study several different Hamiltonian dynamics derived from the discrete SBP from two different perspectives. We also discuss the existence of stationary and periodic Hamiltonian processes of the discrete SBP. We provide more examples of Hamiltonian process on graph in section 5.

2. Preliminary

In this section, we first briefly recall the relationship between the continuous OT problem and Hamiltonian systems. Then we introduce our motivation example on a graph and review some notations for inhomogeneous Markov process, which is used in our definition for Hamiltonian process.

It is known that in a continuous OT problem (1) with given marginal distributions ρ^0 and ρ^1, the optimal transfer $\{X_t\}_{t\in[0,1]}$ induces a trajectory concentrating on the geodesic path whose position and momentum obey the Hamiltonian principle (see e.g. [30]). More precisely, recalling that $H(x,v)=\frac{1}{2}|v|^2$, the critical point of the OT problem (1) in density manifold satisfies the Wasserstein–Hamiltonian flow,

$$\begin{aligned}\partial_t\rho+\nabla\cdot(\frac{\partial H}{\partial v}(x,\nabla S)\rho)&=0,\\ \partial_t S+H(x,\nabla S)&=C(t),\end{aligned}\tag{4}$$

where $C(t)$ is a function depending only on t and $v = \nabla S$ with $|\nabla S|^2 = \nabla S \cdot \nabla S$. Being a Hamiltonian system on its own, (4) can also be connected to the following classic Hamiltonian system closely (see e.g. [9]):

$$\begin{aligned} d_t v &= -\frac{\partial H}{\partial x}(X, v), \\ d_t X &= \frac{\partial H}{\partial v}(X, v), \end{aligned} \tag{5}$$

where $X \in \mathbb{R}^d$, the conjugate momenta $v \in \mathbb{R}^d, d \in \mathbb{N}^+$, and the Hamiltonian H is smooth. If the initial position $X(0)$ is random following a distribution with density ρ^0, the trajectory X_t is random too. Its density function ρ, defined by the pushforward operator induced by the X_t, $\rho_t = X_t^{\#}\rho^0$, satisfies the Wasserstein-Hamiltonian flow (4). However, directly mimicking the relationship between (4) and (5) is impossible if the underlying space becomes a graph. In the next subsection, we illustrate the challenges in detail by an example on graph.

2.1. A motivation example

Consider a graph $G = (V, E, \mathbf{W})$ with a node set $V = \{a_i\}_{i=1}^N$, an edge set E, and $w_{jl} \in \mathbf{W}$ are the weights of the edges: $w_{jl} = w_{lj} > 0$, if there is an edge between a_j and a_l, and 0 otherwise. Below, we write $(i, j) \in E$ to denote the edge in E between the vertices a_i and a_j. We assume that G is an undirected and connected graph with no self loops or multiple edges for simplicity. Let us denote the set of discrete probabilities on the graph by:

$$\mathcal{P}(G) = \Big\{(\rho)_{j=1}^N \ : \sum_j \rho_j = 1, \rho_j \geq 0, \text{ for } j \in V\Big\},$$

and let $\mathcal{P}_o(G)$ be its interior (i.e., all $\rho_j > 0$, for $a_j \in V$). Inspired by [9,12,18], we consider the following discrete OT problem whose minimizer is the so-called geodesic random walk.

Example 2.1. OT on G (geodesic random walk).

The OT problem on a finite graph is related to the Wasserstein distance on $\mathcal{P}(G)$, which can be defined by the discrete Benamou–Brenier formula:

$$W(\rho^0, \rho^1) := \inf_{v,\rho}\Bigg\{\sqrt{\int_0^1 \langle v, v\rangle_{\theta(\rho)} dt} \ : \ \frac{d\rho}{dt} + div_G^\theta(\rho v) = 0, \ \rho(0) = \rho^0, \ \rho(1) = \rho^1\Bigg\},$$

where $\rho^0, \rho^1 \in \mathcal{P}(G)$, $\rho \in H^1([0, 1], \mathbb{R}^N)$ and v is a skew matrix valued function. The inner product of two vector fields u, v is defined by

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

with the weight θ_{ij} depending on ρ_i and ρ_j. The divergence of the flux function ρv is defined as

$$(div_G^\theta(\rho v))_j := -(\sum_{l\in N(j)} w_{jl}v_{jl}\theta_{jl}),$$

where $N(i) = \{a_j \in V : (i, j) \in E\}$ is the adjacency set of node a_i. Then its critical point (ρ, v), with $v = \nabla_G S := (S_j - S_l)_{(j,l)\in E}$ for some function S on V, satisfies the following discrete Wasserstein-Hamiltonian flow on the graph G:

$$\begin{aligned} &\frac{d\rho_i}{dt} + \sum_{j\in N(i)} w_{ij}(S_j - S_i)\theta_{ij}(\rho) = 0, \\ &\frac{dS_i}{dt} + \frac{1}{2}\sum_{j\in N(i)} w_{ij}(S_i - S_j)^2\frac{\partial\theta_{ij}(\rho)}{\partial\rho_i} = 0. \end{aligned} \tag{6}$$

We may view this equation as a discrete analog of (4). Consequently, its Hamiltonian only consists of the kinetic energy

$$\mathcal{H}(\rho, S) = \frac{1}{4}\sum_{ij}(S_i - S_j)^2\theta_{ij}(\rho)w_{ij}.$$

As discussed in [18], the goal of the discrete OT problem is to find an optimal transport of the informal minimization problem

$$\inf_Q\Big\{\int_0^T \frac{1}{2}\sum_{ij\in E}(v_{ij})^2\theta_{ij}w_{ij}dt : d\rho_t = \rho_t Q_t dt,\ X(0)\sim\rho_0,\ X(T)\sim\rho_T\Big\}, \tag{7}$$

where $T = 1$ and the transition rate matrix Q_t may be written as

$$(Q_t)_{ii} = \frac{1}{2}\sum_{j\in N(i)} w_{ij}\frac{\theta_{ij}(\rho)}{\rho_i}v_{ij},$$

$$(Q_t)_{ji} = -\frac{1}{2}w_{ji}\frac{\theta_{ji}(\rho)}{\rho_j}v_{ji},$$

if $\theta_{ij} = \theta_{ji}$. In [18], the minimizer of the above discrete OT problem is called the geodesic random walk which is defined as a random walk whose marginal probability is supported on the set of geodesic paths on $\mathcal{P}(G)$, i.e., X_t is determined by the marginal distribution and the instantaneous transition rate matrix Q_t. However, examining the transition rate matrix, we can find that the geodesic random walk X_t may not be well-defined, because there may not exist such a stochastic process due to possible negative probability and transition probability (See Remark 3.2 for more details).

This example illustrates that when compared to the continuous case, where the Hamiltonian system (5) on the phase space corresponds to the Hamiltonian PDEs (4) on Wasserstein manifold, such a correspondence in discrete space can't be easily established, because the counterpart of (5) requires more careful treatments.

2.2. Inhomogeneous Markov process

In order to define a stochastic process which plays the role of the Hamiltonian mechanics (5) on a finite graph, we recall the definition of the **inhomogeneous Markov process** in [15]. The linear master equation

$$\frac{d\rho}{dt} = \rho Q$$

determines a linear Markov process. When $Q = Q(t)$, it corresponds to a time inhomogeneous Markov process. Here $Q(t)$ is a family of infinitesimal generators of the stochastic matrix or Kolmogorov matrix, namely, a square matrix which has non-positive (resp. non-negative) elements on the main diagonal (resp. off the main diagonal), and the sum of each row is zero. Among different types of inhomogeneous Markov process, the **nonlinear Markov processes** [15] whose transition rate matrix Q may depend not only on the current state x of the process but also on the current distribution ρ of the process is of particular interest to us.

Given an initial distribution ρ_0, a time inhomogeneous Markov process $\{X_t\}_{t\geq 0}$ can be defined as a process which has ρ_0 as the distribution of X_0 and $(s,t) \to P_{s,t}$ as its transition mechanism in the sense that

$$\mathbb{P}(X_0 = a_i) = \rho_i,\ \mathbb{P}(X_t = a_j | X_\sigma, \sigma \in [0,s]) = (P_{s,t})_{X(s)a_j},$$

where $(P_{s,t})_{a_i a_j} = \mathbb{P}(X_t = a_j | X_s = a_i)$. The corresponding forward Kolmogorov equation can be rewritten as

$$d_t P_{s,t} = P_{s,t} Q_t.$$

If $t \in [s,\infty) \mapsto \rho_t$ is continuously differentiable, then

$$\dot{\rho}_t = \rho_t Q_t,$$

is equivalent to

$$\rho_t = \rho_s P_{s,t},$$

for $t \geq s$. Given $(Q_t)_{t\geq 0}$, ρ_0, there exists an inhomogeneous Markov process X_t related to the transition rate matrix Q_t and the marginal distribution ρ_t. On the other hand, given an inhomogeneous Markov process with transition matrices $P_{s,t}$, it will induce the equation of ρ with Q_t (see e.g. [15]).

3. Hamiltonian process on a finite graph

As shown in Example 2.1, although it may not be possible to find a stochastic process for every discrete optimal transport problem, it reveals two key features that the density of such a stochastic process, if exists, satisfies the generalized master equation and that its Q_t-matrix is determined by a potential S_t, where S_t satisfies a discrete Hamiltonian Jacobi equation. Inspired by these properties, we introduce the definition of stochastic Hamiltonian process.

Definition 3.1. A stochastic process $\{X_t\}_{t\geq 0}$ is called a Hamiltonian process on the graph if

1. The density ρ_t of X_t satisfies the following generalized Master equation,

$$d_t\rho_t = \rho_t Q_t,$$

   with $(Q_t)_{ij} = w_{ji} f_{ji}(v_{ji}), (Q_t)_{ii} = -\sum_{j\in N(i)} w_{ij} f_{ji}(v_{ji})$, where the skew-matrix v is induced by a potential function S, i.e. $v = \nabla_G S + u$, with $div_G(\rho u) = 0$. And $f_{ij} : \mathbb{R} \to \mathbb{R}$, is a real-valued measurable function which is piecewise continuous in $x \in \mathbb{R}$ and may depend on t and ρ.
2. The density ρ and the potential S form a Hamiltonian system on the cotangent bundle of the density space.

The following theorem gives the structure of the Hamiltonian on the density manifold of the Hamiltonian process.

Theorem 3.1. *Suppose that the stochastic process $\{X_t\}_{t\geq 0}$ with density $\{\rho_t\}_{t\geq 0}$ and potential $\{S_t\}_{t\geq 0}$ defined in the Definition 3.1 forms a Hamiltonian process on the graph G. In addition assume that the antiderivative F_{ij} of f_{ij} exists for $ij \in E$. Then the Hamiltonian always has the form*

$$\mathcal{H}(\rho, S) = \sum_{i\in V}\sum_{j\in N(i)} \rho_i F_{ji}(S_j - S_i, \rho, t)w_{ji} + \mathcal{V}(\rho, t) \tag{8}$$

where $\mathcal{V}$ is a function depending ρ and t. Moreover, the Hamiltonian system on the cotangent bundle of $\mathcal{P}(G)$ can be written as:

$$\begin{aligned} \frac{\partial}{\partial t}\rho_i(t) &= \sum_{j\in N(i)} w_{ij} f_{ij}(S_i - S_j, \rho, t)\rho_j - w_{ji} f_{ji}(S_j - S_i, \rho, t)\rho_i \\ \frac{\partial}{\partial t}S_i(t) &= -\sum_{j\in N(i)} \left(w_{ji}F_{ji}(S_j - S_i, \rho, t) + \rho_i \frac{\partial}{\partial \rho_i} F_{ji}(S_j - S_i, \rho, t)w_{ji} \right) - \frac{\partial}{\partial \rho_i}\mathcal{V}(\rho, t). \end{aligned} \tag{9}$$

Proof. According to Definition 3.1, we have $\frac{\partial}{\partial t}\rho_i(t) = \sum_{j\in N(i)} w_{ij} f_{ij}(S_i - S_j, \rho, t)\rho_j - w_{ji} f_{ji}(S_j - S_i, \rho, t)\rho_i$. Since $\{\rho_t, S_t\}$ forms a Hamiltonian system, we are able to state

$$\frac{\partial}{\partial S_i}\mathcal{H}(\rho, S, t) = \sum_{j\in N(i)} w_{ij} f_{ij}(S_i - S_j, \rho, t)\rho_j - w_{ji} f_{ji}(S_j - S_i, \rho, t)\rho_i, \quad i \in V.$$

Considering the following quantity,

$$\mathcal{H}_0(\rho, S, t) = \sum_{i\in V}\sum_{j\in N(i)} \rho_i F_{ji}(S_j - S_i, \rho, t)w_{ji},$$

we can directly verify that $\frac{\partial}{\partial S}(\mathcal{H}(\rho, S, t) - \mathcal{H}_0(\rho, S, t)) = 0$. This suggests that there exists some function $\mathcal{V}$ depending on ρ and t such that $\mathcal{H}(\rho, S, t) - \mathcal{H}_0(\rho, S, t) = \mathcal{V}(\rho, t)$. This directly leads to form of Hamiltonian $\mathcal{H}(\rho, S, t) = \sum_{i\in V}\sum_{j\in N(i)} \rho_i F_{ji}(S_j - S_i, \rho, t) w_{ji} + \mathcal{V}(\rho, t)$. Furthermore, the discrete Hamiltonian Jacobi equation is derived as

$$\frac{\partial}{\partial t} S_t = -\frac{\partial}{\partial \rho}\mathcal{H}(\rho, S, t). \quad \square$$

As a direct consequence, we have the following properties of Hamiltonian process.

Proposition 3.1 *(Properties of Hamiltonian process). Assume that a stochastic process X_t on a finite graph is a Hamiltonian process. Then it holds that*

1. *there exists a Hamiltonian $\mathcal{H}$ on the density space such that its marginal distribution $\rho_t = X_t^{\#}\rho_0$ and the generator S_t of the transition rate matrix Q_t forms a Hamiltonian system;*
2. *the symplectic structure on the density space is preserved, i.e.,*

$$\omega_{g(\rho,S)}(g'(\rho, S)\xi, g'(\rho, S)\eta) = \omega_{(\rho,S)}(\xi, \eta),$$

where ω denotes the symplectic form on $T^*\mathcal{P}(G)$, $\xi, \eta \in T_{(\rho,S)}(T^*P(G))$ and $g'(\rho, S)$ is the Jacobi matrix of the Hamiltonian flow on the density space;

3. *$\mathcal{H}(t) = \mathcal{H}(0)$, if the Hamiltonian $\mathcal{H}$ is independent of t;*
4. *and X_t is mass-preserving, i.e., $\sum_{i=1}^N \rho_i(t) = \sum_{i=1}^N \rho_i(0)$.*

Remark 3.1 *(Particle-level properties of Hamiltonian process).* Consider the Hamiltonian with specific form

$$\mathcal{H}(\rho, S) = \sum_{i\in V}\sum_{j\in N(i)} \rho_j F_{ji}(S_j - S_i) w_{ji} + \sum_{i\in V} \rho_i V_i.$$

Suppose that $\{X(t)\}$ is a Hamiltonian process on G associated to the Hamiltonian $\mathcal{H}$. Then one can verify $\mathbb{E}[H(X(t), S(t))]$ with

$$H(X(t), S(t)) = \sum_{j\in N(X(t))} F_{jX(t)}(S_j(t) - S_{X(t)}(t)) w_{jX(t)} + V_{X(t)},$$

remains constant as time t evolves.

Based on the definition of Hamiltonian process, we are able to construct the discrete optimal transport problem which retains the property that the minimizer is a stochastic process on the graph for Example 2.1.

Proposition 3.2. *There always exists a density dependent weight θ such that the geodesic random walk in Example 2.1 is a Hamiltonian process.*

Proof. Define $\theta_{ij}^U = \theta_S^U(\rho_i, \rho_j)$, where $\theta_S^U(\rho_i, \rho_j) = \rho_i$ if $S_j > S_i$. Denote $(x)^+ = \max(0, x)$, $(x)^- = \min(0, x)$. Using the notations in Example 2.1, the geodesic random walk on G with the probability weight $\theta = \theta^U$ satisfies

$$d\rho_i = \sum_{j \in N(i)} w_{ij}(v_{ij})^+ \rho_j + \sum_{j \in N(i)} w_{ij}(v_{ij})^- \rho_i. \tag{10}$$

From the discrete Hodge decomposition on the graph [9], for any skew matrix v and probability density $\rho \in \mathcal{P}_o(G)$, there exists a decomposition $v = \nabla_G S + u$ with $div_G^\theta(\rho u) = 0$. Here $(\nabla_G S)_{ij} := (S_i - S_j)$ and $div_G^\theta(\rho u) := -(\sum_{j \in N(i)} u_{ij}\theta_{ij}^U(\rho))$. To see this fact, it suffices to prove that there exists a unique solution of S such that $div_G^\theta(\rho \nabla_G S) = div_G^\theta(\rho v)$. The connectivity of the graph and the fact that $\rho \in \mathcal{P}_o(G)$ implies that if

$$\langle div_G^\theta(\rho \nabla_G S), S\rangle = \frac{1}{2} \sum_{(i,j) \in E} ((S_i - S_j)^-)^2 \theta_{ij}(\rho) = 0,$$

then 0 must be a simple eigenvalue of $div_G^\theta(\rho \nabla_G)$ with eigenvector $(1, \cdots, 1)$. Thus S is unique up to a constant shift and the skew matrix $v_t = \nabla_G S_t + u$ satisfies

$$d(S_t)_i = -\frac{1}{2} \sum_{j \in N(i)} w_{ij}((S_i - S_j)^-)^2 + C(t),\ div_G^\theta(\rho u) = 0,$$

where $C(t)$ is independent of nodes. Meanwhile, f_{ij} can be selected to achieve $f_{ij}(S_i - S_j) = (S_i - S_j)^+$ and thus

$$(Q_t)_{ii} = \sum_{j \in N(i)} w_{ij}(S_i - S_j)^- = \sum_{j \in N(i)} w_{ij} f_{ji}(S_j - S_i),$$
$$(Q_t)_{ji} = w_{ji}(S_j - S_i)^+ = w_{ji} f_{ji}(S_j - S_i),\ ij \in E, \text{ otherwise } Q_{ji} = 0.$$

We can define a time inhomogeneous Markov process as follows by the transition matrix $\mathbb{P}(X_t = v_j | X_\tau, \tau \in [0, s]) = (P_{s,t})_{X(s)v_j}$. Given the past $\sigma(\{X_\tau : \tau \in [0, t]\})$ of X up to time $t \geq 0$, the probability of its having moved away from X_t at the time $t + h$ with h small enough can be approximated by $1 - (Q_t)_{X_t X_t} h$, i.e.,

$$\left| \mathbb{P}(X(t+h) = X_t | X_\tau, \tau \leq t) - 1 - (Q_t)_{X_t X_t} h \right| = o(h).$$

Here $\{-(Q_t)_{ii}\}_i$ is often called as the transition rate of X_t. Given the history that the jump appeared $\sigma(\{X_\tau : \tau \in [0, t]\} \cup \{X_{t+h} \neq X_t\})$, the probability that $X_{t+h} = a_j$ is approximately $(P_{t,t+h})_{X_t a_j}$, which implies that

$$\left| \mathbb{P}(X(t+h) = a_j | X_\tau, \tau \leq t) - h(Q_t)_{X_t a_j} \right| = o(h). \quad \square$$

Remark 3.2. It is worth mentioning that the Hamiltonian system on $\mathcal{P}(G)$ does not necessarily induce a stochastic process on G. This can also be illustrated by using the optimal transport problem introduced in Example 2.1. Let us take $w_{ij} = 1$ if $ij \in E$ for simplicity. In order to define a Hamiltonian process on G, the probability weight θ can not be chosen arbitrarily here. For example, if we take the probability weight $\theta_{ij} = \theta^A(\rho_i, \rho_j) = \frac{1}{2}(\rho_i + \rho_j)$ in [9], the density equation can be rewritten as

$$d_t \rho_t = \rho_t Q_t,$$

where

$$(Q_t)_{ii} = \frac{1}{2} \sum_{j \in N(i)} (S_i - S_j),$$

$$(Q_t)_{ij} = \frac{1}{2}(S_j - S_i),\ ij \in E,\ \text{otherwise } Q_{ij} = 0.$$

The function $f_{ij}(x) = \frac{1}{2}x$.

When $\theta_{ij} = \theta^L(\rho_i, \rho_j) = \frac{\rho_i - \rho_j}{\log(\rho_i) - \log(\rho_j)}$ in [7], the density equation can be rewritten as

$$d_t \rho_t = \rho_t Q_t,$$

where

$$(Q_t)_{ii} = \sum_{j \in N(i)} \frac{(S_i - S_j)}{\log(\rho_i) - \log(\rho_j)},$$

$$(Q_t)_{ij} = -\frac{(S_i - S_j)}{\log(\rho_i) - \log(\rho_j)},\ ij \in E,\ \text{otherwise } Q_{ij} = 0.$$

The function $f_{ij}(x) = \frac{x}{\log(\rho_i) - \log(\rho_j)}$.

In both cases, there is no guarantee that the off-diagonal of Q_t is non-positive. Hence, Q_t is unable to admit a stochastic process X_t which is time inhomogeneous Markov due to the appearance of negative transition probabilities. For valid choices of θ that may admit stochastic processes, we refer to [7], [19] and references therein.

Remark 3.3. If $\theta_{ij} > 0$ for all $ij \in E$, then the Hodge decomposition yields a unique potential S up to a constant which induces v. If there exists $ij \in E$ such that $\theta_{ij} = 0$, then the generator S may be not unique. Meanwhile, the Hamiltonian Jacobi equation may become one-side inequality,

$$v_{ij} = S_i - S_j,\ \partial_t S_i + \frac{\partial}{\partial \rho_i} \mathcal{H}(\rho, S) \le 0.$$

Remark 3.4. The initial value problem of the Hamiltonian system of ρ, S may develop singularity at a finite time $T > 0$, i.e., either $\lim_{t \to T} S_i(t) = \infty$ or $\lim_{t \to T} \rho_i \le 0$.

We would like to emphasize that a Hamiltonian process is not Markov in general. The sufficient and necessary conditions when a Hamiltonian process gives a Markov process are presented as follows.

Theorem 3.2. *Given a Hamiltonian process $\{X_t\}_{t \ge 0}$ on the graph with a Hamiltonian $\mathcal{H}(\rho, S) = \sum_{i=1}^N \sum_{j \in N(i)} F_{ij}(\rho, S) w_{ij} \rho_i$. If X_t is a Markov process, then (ρ, S) in Definition (3.1) satisfies the following system,*

$$
\begin{aligned}
&\frac{\partial^2 F_{ij}}{\partial S_i \partial \rho_i}\Big(\sum_{l\in N(i)} \frac{\partial F_{il}}{\partial S_i}\rho_i w_{il} + \sum_{l\in N(i)} \frac{\partial F_{li}}{\partial S_i}\rho_l w_{li}\Big) \\
&+ \frac{\partial^2 F_{ij}}{\partial S_i \partial \rho_j}\Big(\sum_{k\in N(j)} \frac{\partial F_{jk}}{\partial S_j}\rho_j w_{jk} + \sum_{k\in N(j)} \frac{\partial F_{kj}}{\partial S_j}\rho_l w_{kj}\Big) \\
&- \frac{\partial^2 F_{ij}}{\partial S_i \partial S_i}\Big(\sum_{l\in N(i)} \frac{\partial F_{il}}{\partial \rho_i}\rho_i w_{il} + \sum_{l\in N(i)} \frac{\partial F_{li}}{\partial \rho_i}\rho_l w_{li} + \sum_{l\in N(i)} (F_{il}w_{il} + F_{li}w_{li})\Big) \\
&- \frac{\partial^2 F_{ij}}{\partial S_i \partial S_j}\Big(\sum_{k\in N(j)} \frac{\partial F_{jk}}{\partial \rho_j}\rho_j w_{jk} + \sum_{k\in N(j)} \frac{\partial F_{kj}}{\partial \rho_j}\rho_k w_{kj} + \sum_{k\in N(j)} (F_{jk}w_{jk} + F_{ki}w_{ki})\Big) = 0
\end{aligned} \tag{11}
$$

for $i, j \in V$. Conversely, if (ρ, S) satisfies (11), then there exists a Markov process which is Hamiltonian.

Proof. Since X_t is a Hamiltonian process, the transition matrix is determined by $\rho_t Q_t = \frac{\partial H}{\partial S} = d_t \rho_t$. This implies that

$$
(\rho_t Q_t)_i = \sum_{j\in N(i)} \frac{\partial F_{ij}(\rho, S)}{\partial S_i} w_{ij}\rho_i + \sum_{j\in N(i)} \frac{\partial F_{ji}(\rho, S)}{\partial S_i} w_{ji}\rho_j .
$$

Therefore, $(Q_t)_{ii} = \sum_{j\in N(i)} \frac{\partial F_{ij}(\rho,S)}{\partial S_i} w_{ij}$, $(Q_t)_{ij} = \frac{\partial F_{ij}(\rho,S)}{\partial S_j} w_{ij}$. Since X_t preserves the mass, it holds that $\sum_{j\in N(i)} (\frac{\partial F_{ij}(\rho,S)}{\partial S_i} + \frac{\partial F_{ij}(\rho,S)}{\partial S_j}) w_{ij} = 0$ for every $i \le N$.

Notice that X_t is Markov implies that $d_t Q_{ij} = 0$, for $i, j \le N$, that is

$$
d_t \frac{\partial F_{ij}}{\partial S_i} = 0, d_t \frac{\partial F_{ji}}{\partial S_j} = 0.
$$

Direct calculation leads to

$$
\begin{aligned}
d_t \frac{\partial F_{ij}}{\partial S_i} &= \frac{\partial^2 F_{ij}}{\partial S_i \partial \rho_i} d_t \rho_i + \frac{\partial^2 F_{ij}}{\partial S_i \partial \rho_j} d_t \rho_j \\
&\quad + \frac{\partial^2 F_{ij}}{\partial S_i \partial S_i} d_t S_i + \frac{\partial^2 F_{ij}}{\partial S_i \partial S_j} d_t S_j \\
&= \frac{\partial^2 F_{ij}}{\partial S_i \partial \rho_i}\Big(\sum_{l\in N(i)} \frac{\partial F_{il}}{\partial S_i}\rho_i w_{il} + \sum_{l\in N(i)} \frac{\partial F_{li}}{\partial S_i}\rho_l w_{li}\Big) \\
&\quad + \frac{\partial^2 F_{ij}}{\partial S_i \partial \rho_j}\Big(\sum_{k\in N(j)} \frac{\partial F_{jk}}{\partial S_j}\rho_j w_{jk} + \sum_{k\in N(j)} \frac{\partial F_{kj}}{\partial S_j}\rho_l w_{kj}\Big) \\
&\quad - \frac{\partial^2 F_{ij}}{\partial S_i \partial S_i}\Big(\sum_{l\in N(i)} \frac{\partial F_{il}}{\partial \rho_i}\rho_i w_{il} + \sum_{l\in N(i)} \frac{\partial F_{li}}{\partial \rho_i}\rho_l w_{li} + \sum_{l\in N(i)} (F_{il}w_{il} + F_{li}w_{li})\Big)
\end{aligned}
$$

$$-\frac{\partial^2 F_{ij}}{\partial S_i \partial S_j}\Big(\sum_{k\in N(j)} \frac{\partial F_{jk}}{\partial \rho_j}\rho_j w_{jk} + \sum_{k\in N(j)} \frac{\partial F_{kj}}{\partial \rho_j}\rho_k w_{kj} + \sum_{k\in N(j)} (F_{jk} w_{jk} + F_{ki} w_{ki})\Big),$$

which yields the desired result. Conversely, if (ρ, S) satisfies (11), the previous arguments lead to the equation of ρ becomes a linear Master equation. Then there always exists a Markov process which is a stochastic representation of linear Master equation. Meanwhile, it can be verified that this Markov process satisfies all the conditions in Definition 3.1 and is Hamiltonian. □

Corollary 3.1. *Given a Hamiltonian $\mathcal{H}(\rho, S) = \sum_{i=1}^N \sum_{j\in N(i)} F_{ij}(\rho, S) w_{ij} \rho_i$. Assume that there exists $(\rho^*, S^*(t))$ satisfies the following conditions,*

1. $\sum_{j\in N(i)} \frac{\partial F_{ij}(\rho,S)}{\partial S_i} + \frac{\partial F_{ij}(\rho,S)}{\partial S_j} = 0,$
2. *ρ^* is independent of t and $(\rho^*, S^*(t))$ solves*

$$\sum_{l\in N(i)} \frac{\partial F_{il}}{\partial S_i}\rho_i w_{il} + \sum_{l\in N(i)} \frac{\partial F_{li}}{\partial S_i}\rho_l w_{li} = 0,$$

$$\frac{\partial^2 F_{ij}}{\partial S_i \partial S_i}\Big(\sum_{l\in N(i)} \frac{\partial F_{il}}{\partial \rho_i}\rho_i w_{il} + \sum_{l\in N(i)} \frac{\partial F_{li}}{\partial \rho_i}\rho_l w_{li} + \sum_{l\in N(i)} (F_{il} w_{il} + F_{li} w_{li})\Big)$$
$$+\frac{\partial^2 F_{ij}}{\partial S_i \partial S_j}\Big(\sum_{k\in N(j)} \frac{\partial F_{jk}}{\partial \rho_j}\rho_j w_{jk} + \sum_{k\in N(j)} \frac{\partial F_{kj}}{\partial \rho_j}\rho_k w_{kj} + \sum_{k\in N(j)} (F_{jk} w_{jk} + F_{ki} w_{ki})\Big) = 0$$

Then there exists a Hamiltonian process which is Markov and preserves the mass. Furthermore, the Hamiltonian process is invariant with respect to ρ^*.

4. Hamiltonian process via discrete SBP on graphs

Although the SBP [26] has a history close to 100 years, it has received revived attention from control theory and machine learning communities recently, see [17,24]. For convenience, the background of continuous SBP is presented in the appendix.

For the discrete counterpart of SBP on graph, there are two different treatments reported in the literature.

(A) One is to consider a reference path measure R (induced by a reversible random walk) on the graph and then study the optimization problem involving the relative entropy between the reference measure R and the path measure P with given initial and terminal distributions [17,18].) In this framework, the reference random walk is often related to a discrete version of (33) (For example, the linear discretization of the Laplacian gives the time homogeneous Markov chain as the reference in [5]).

(B) Another way is proposed by the discrete version of (30) or (32) directly [8].

We shall show that different treatments create differences on the structure and formulation of equations, in particular the discrete Laplacian operator. Each of these formulations can determine its corresponding Hamiltonian process on graph.

4.1. Discrete SBP based on relative entropy and reference Markov measure

In the following discussion, we always assume that $w_{ij} = 1$ if $ij \in E$ for conciseness of formulations. By using the discrete Girsanov theorem on graph, the discrete SBP in the form of relative entropy (A) becomes the following control problem

$$\min_{\widehat{m}^t \geq 0} \left\{ \int_0^1 \sum_{i \in V} \rho(i,t) \sum_{j \in N(i)} \left(\frac{\widehat{m}^t_{ij}}{m^t_{ij}} \log\left(\frac{\widehat{m}^t_{ij}}{m^t_{ij}}\right) - \frac{\widehat{m}^t_{ij}}{m^t_{ij}} + 1 \right) m^t_{ij} \, dt \right\} \tag{12}$$

$$\text{subject to:} \quad \frac{d}{dt}\rho(i,t) = \sum_{j \in N(i)} \widehat{m}^t_{ji}\rho_j - \widehat{m}^t_{ij}\rho_i \quad \rho(\cdot,0) = \rho^0, \ \rho(\cdot,1) = \rho^1,$$

where the reference measure R is determined by the master equation $d_t\widetilde{\rho}_t = \sum_{j \in N(i)} m^t_{ji}\widetilde{\rho}_j - m^t_{ij}\widetilde{\rho}_i$.

Remark 4.1. The formula for relative entropy between path measure P and reference path measure R is formulated as

$$H(P|R) = \int_0^1 \sum_{i \in V} \rho(i,t) \sum_{j \in N(i)} \left(\frac{\widehat{m}^t_{ij}}{m^t_{ij}} \log\left(\frac{\widehat{m}^t_{ij}}{m^t_{ij}}\right) - \frac{\widehat{m}^t_{ij}}{m^t_{ij}} + 1 \right) m^t_{ij} \, dt.$$

This result is provided in [17], [18]. A rigorous proof for this formula originates from Theorem 2.9 of [16].

Let us denote $u(x) = x \log x - x + 1$. By introducing Lagrange multiplier ψ, we obtain the following Lagrangian functional

$$\begin{aligned}
\mathcal{L}(\rho, \widehat{m}, \psi) &= \int_0^1 \sum_{i \in V} \rho(i,t) \sum_{j \in N(i)} u\left(\frac{\widehat{m}^t_{ij}}{m^t_{ij}}\right) m^t_{ij} \, dt \\
&+ \int_0^1 \sum_{i \in V} -\rho(i,t)\frac{\partial}{\partial t}\psi(i,t) - \psi(i,t) \left(\sum_{j \in N(i)} \widehat{m}^t_{ji}\rho_j - \widehat{m}^t_{ij}\rho_i \right) dt \\
&= \int_0^1 -\sum_{i \in V} \rho(i,t)\frac{\partial}{\partial t}\psi(i,t) - \frac{1}{2}\sum_{(i,j) \in E} \left[\frac{\widehat{m}_{ji}}{m_{ji}}(\psi(i,t) - \psi(j,t)) - u\left(\frac{\widehat{m}_{ji}}{m_{ji}}\right) \right] m_{ji}\rho(j,t) \\
&+ \left[\frac{\widehat{m}_{ij}}{m_{ij}}(\psi(j,t) - \psi(i,t)) - u\left(\frac{\widehat{m}_{ij}}{m_{ij}}\right) \right] m_{ij}\rho(i,t) \, dt.
\end{aligned}$$

When solving the above saddle point problem, we minimize over $\widehat{m}$ and get

$$\int_0^1 -\sum_{i\in V}\rho(i,t)\frac{\partial}{\partial t}\psi(i,t) - \frac{1}{2}\sum_{(i,j)\in E}[u^*(\psi(i,t)-\psi(j,t))m_{ji}\rho(j,t) + u^*(\psi(j,t)-\psi(i,t))m_{ij}\rho(i,t)]\,dt.$$

Here u^* is the Legendre dual of u: $u^*(x) = \sup_y\{x \cdot y - u(y)\}$, leading to $u^*(x) = e^x - 1$. By formulating the Lagrangian, we can identify the Hamiltonian of this control problem, which can be written as:

$$\mathcal{H}(\rho,\psi) = \sum_{i\in V}\sum_{j\in N(i)}(\exp(\psi(j,t)-\psi(i,t)) - 1)m_{ij}\rho(i,t). \tag{13}$$

Then the above control problem implies the following Hamiltonian system

$$\partial_t\rho = \frac{\partial\mathcal{H}(\rho,\psi)}{\partial\psi}, \quad \partial_t\psi = -\frac{\partial\mathcal{H}(\rho,\psi)}{\partial\rho},$$

that is,

$$\begin{aligned}\frac{\partial}{\partial t}\rho(i,t) &= \sum_{j\in N(i)} -e^{\psi(j,t)-\psi(i,t)}m_{ij}\rho(i,t) + e^{\psi(i,t)-\psi(j,t)}m_{ji}\rho(j,t), \\ \frac{\partial}{\partial t}\psi(i,t) &= -\sum_{j\in N(i)}(e^{\psi(j,t)-\psi(i,t)} - 1)m_{ij}.\end{aligned} \tag{14}$$

By using the Hopf-Cole transform, we can further verify that the discrete SBP problem determines a Hamiltonian process on the graph. Let us consider the following transform $\tau : T^*\mathcal{P}(G) \to T^*\mathcal{P}(G)$ as:

$$\tau[(\rho,\psi)] = (\rho, \psi - \frac{1}{2}\ln\rho) \tag{15}$$

Let us denote $g'(\rho,\psi) = D\tau(\rho,\psi)$. Then the symplectic form ω is unchanged in the sense that

$$\omega_{g(\rho,\psi)}(g'(\rho,\psi)\xi, g'(\rho,\psi)\eta) = \omega_{(\rho,\psi)}(\xi,\eta),$$

where $(\xi,\eta) \in T_{(\rho,\psi)}T^*\mathcal{P}(G)$. By using the symplectic submersion from $\mathcal{P}(G)$ to $\mathbb{R}^N$, the symplectic form can be represented by $(g'(\rho,\psi)\xi)^T Jg'(\rho,\psi)\eta = \xi^T J\eta$, where J is the standard symplectic matrix. Since $d_t\tau(\rho,\psi)^T = \tau' d_t(\rho,\psi)^T$ and that $(\tau')^T J\tau' = J$, we conclude that the Hopf–Cole transformation (15) is a symplectic transformation on the cotangent bundle of the density manifold. Denote (ρ, S) as the new coordinate. Then $\{\rho_t, S_t\}$ satisfies the following Hamiltonian system:

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

with

$$\tilde{\mathcal{H}}(\rho, S) = \mathcal{H}(\tau^{-1}(\rho, S)) = \sum_{i\in V}\sum_{j\in N(i)} e^{(S_j - S_i)} m_{ij}\sqrt{\rho_i\rho_j}, \tag{16}$$

that is

$$\begin{aligned} \frac{\partial S(i,t)}{\partial t} &= -m_{ii} - \frac{1}{2}\sum_{j\in N(i)} e^{S_j - S_i} m_{ij}\frac{\sqrt{\rho_j}}{\sqrt{\rho_i}} - \frac{1}{2}\sum_{j\in N(i)} e^{S_i - S_j} m_{ji}\frac{\sqrt{\rho_j}}{\sqrt{\rho_i}}, \\ \frac{\partial \rho(i,t)}{\partial t} &= \sum_{j\in N(i)} e^{S_i - S_j} m_{ji}\sqrt{\rho_j}\sqrt{\rho_i} - \sum_{j\in N(i)} e^{S_j - S_i} m_{ij}\sqrt{\rho_i}\sqrt{\rho_j}. \end{aligned} \tag{17}$$

As a consequence, we verify that, as reported in [17], the discrete SBP corresponds to a Hamiltonian process with the transition rate matrix Q ($Q_{ij} = \widehat{m}_{ij}$) defined by $Q_{ii} = -\sum_{j\in N(i)} e^{S_j - S_i}\frac{\sqrt{\rho_j}}{\sqrt{\rho_i}} m_{ij}$, $Q_{ij} = e^{S_i - S_j}\frac{\sqrt{\rho_i}}{\sqrt{\rho_j}} m_{ji}$ if $ij \in E$.

Using the above procedures, we can naturally extend the original SBP problem to the following generalized control problem

$$\begin{aligned} &\min_{\hat{m}^t \geq 0} \left\{ \int_0^1 \sum_{i\in V} \rho(i,t) \sum_{j\in N(i)} u\left(\frac{\hat{m}^t_{ij}}{m^t_{ij}}\right) m^t_{ij}\, dt \right\} \\ &\text{subject to:} \quad \frac{d}{dt}\rho(i,t) = \sum_{j\in N(i)} \hat{m}^t_{ji}\rho_j - \hat{m}^t_{ij}\rho_i \quad \rho(\cdot,0) = \rho^0,\ \rho(\cdot,1) = \rho^1. \end{aligned} \tag{18}$$

Here u is an arbitrary convex function. Then the Hamiltonian associated with this general control problem is

$$\mathcal{H}(\rho, \psi) = \sum_{i\in V}\sum_{j\in N(i)} u^*(\psi(j,t) - \psi(i,t)) m_{ij}\rho(i,t), \tag{19}$$

where $\lambda_{ij} = (u')^{-1}(\psi_j - \psi_i)$.

For the sake of completeness of our discussion, we also reveal the relations among the so-called Schrödinger system [10,11,3] and our derived systems (14) and (17). All three PDE systems are derived from the SBP. We introduce the Madelung Transform $\phi : T^*\mathcal{P}(G) \to T^*\mathcal{P}(G)$

$$(f, g) = \phi(\rho, S) = (\sqrt{\rho}e^{-S}, \sqrt{\rho}e^{S}), \tag{20}$$

or equivalently,

$$(f, g) = \tilde{\phi}(\rho, \psi) = (\rho e^{-\psi}, e^{\psi}). \tag{21}$$

Combining (20) with (17), or combining (21) with (14) yields the Schrödinger system:

$$\frac{\partial}{\partial t} f(i,t) = \sum_{j\in N(i)} (f(j,t) - f(i,t)) m_{ij}^t, \tag{22}$$
$$\frac{\partial}{\partial t} g(i,t) = -\sum_{j\in N(i)} (g(j,t) - g(i,t)) m_{ij}^t.$$

Similar to our previous analysis, we can verify that both transforms ϕ and $\tilde{\phi}$ preserves the symplectic form. And we know that (22) is a Hamiltonian system and its corresponding Hamiltonian is

$$\widehat{\mathcal{H}}(f,g) = \sum_{i\in V} \sum_{j\in N(i)} f_i g_j m_{ij}^t.$$

By applying the Theorem 3.2, we obtain the following result about the conditions under which the Hamiltonian process in SBP enjoys the stationary measure and Markov property.

Proposition 4.1. *Assume that the reference process is mass-preserving, i.e., $\sum_i \widetilde{\rho}(i,t) = \sum_i \widetilde{\rho}^0(i)$, and possesses a stationary measure ρ^*. Then there exists a stationary point (ρ^*, S^*) of the Hamiltonian system (17) on the density manifold.*

Proof. Take $\frac{\partial \tilde{H}}{\partial S} = 0$ and $\frac{\partial \tilde{H}}{\partial \rho} = 0$ such that (ρ, S) is independent of time. The equation of ρ leads to

$$\sum_{j\in N(i)} e^{S_i - S_j} m_{ji} \sqrt{\rho_j} = \sum_{j\in N(i)} e^{S_j - S_i} m_{ij} \sqrt{\rho_j}.$$

Due to $m_{ii} = -\sum_{j\in N(i)} m_{ij}$, the equation of S becomes

$$\frac{1}{2} \sum_{j\in N(i)} (e^{S_i - S_j} m_{ji} + e^{S_j - S_i} m_{ij}) \sqrt{\rho_j} = \sum_{j\in N(i)} m_{ij} \sqrt{\rho_i}.$$

Applying the above relationships, we obtain that

$$\sum_{j\in N(i)} e^{S_i - S_j} m_{ji} \sqrt{\rho_j} = \sum_{j\in N(i)} m_{ij} \sqrt{\rho_i}.$$

This immediately implies that

$$\sum_{j\in N(i)} e^{-S_j} m_{ji} \sqrt{\rho_j} = \sum_{j\in N(i)} e^{-S_i} m_{ij} \sqrt{\rho_i}.$$

Now by taking $e^{S_j^*} \sqrt{\rho_j^*} = e^{S_i^*} \sqrt{\rho_i^*}$ for all $ij \in E$, the first equation is reduced to

$$\sum_{j\in N(i)} m_{ji} \rho_j^* = \sum_{j\in N(i)} m_{ij} \rho_i^*.$$

This leads to

$$\sum_{j\in N(i)} m_{ji}\rho_j^* + m_{ii}\rho_i^* = 0,$$

which is the sufficient and necessary condition that the reference process admits the stationary measure ρ^*. From the above arguments, there always exists a stationary point (ρ^*, S^*) which refers to the reference process itself and $\rho^0 = \rho^1 = \rho^*$ in the SBP. □

In the following, we show that if the solution process of the SBP is Markov, then its density function ρ must be invariant with respect to time.

Corollary 4.1. *Assume there exists a Markov process solving the SBP and that the reference process is mass-preserving, then for all $ij \in E$, $c_{ij} = \frac{e^{S_i}\sqrt{\rho_i}}{e^{S_j}\sqrt{\rho_j}}$ is the solution of*

$$-\sum_{k\in N(i)} c_{ki}m_{ik} + \sum_{l\in N(j)} c_{lj}m_{jl} - m_{jj} + m_{ii} = 0. \tag{23}$$

Moreover, ρ is the invariant measure of the solution process in SBP.

Proof. Since the solution process of the SBP is time homogeneous Markov, we can verify that $\frac{e^{S_i}\sqrt{\rho_i}}{e^{S_j}\sqrt{\rho_j}} = c_{ij} > 0$ is independent of time and that

$$d_t\rho = \rho Q,$$

where $Q_{ii} = -\sum_{j\in N(i)} c_{ji}m_{ij}$, $Q_{ij} = c_{ji}m_{ij}$. Let $e^{\psi_i} = e^{S_i}\sqrt{\rho_i}$. Then it holds that

$$\frac{\partial}{\partial t}\rho(i,t) = \sum_{j\in N(i)} -e^{\psi(j,t)-\psi(i,t)}m_{ij}\rho(i,t) + e^{\psi(i,t)-\psi(j,t)}m_{ji}\rho(j,t)$$

$$\frac{\partial}{\partial t}\psi(i,t) = -\sum_{j\in N(i)} (e^{\psi(j,t)-\psi(i,t)} - 1)m_{ij}.$$

As a consequence, for $ij \in E$,

$$\begin{aligned} d_t c_{ij} &= d_t[(e^{\psi_i-\psi_j})] \\ &= c_{ij}(-\sum_{l\in N(i)} e^{\psi_l-\psi_i}m_{il} + \sum_{k\in N(j)} e^{\psi_k-\psi_j}m_{jk}) + c_{ij}(-m_{jj} + m_{ii}) \\ &= c_{ij}(-\sum_{l\in N(i)} c_{li}m_{il} + \sum_{k\in N(j)} c_{kj}m_{jk} - m_{jj} + m_{ii}) = 0. \end{aligned} \tag{24}$$

Since $c_{ij} > 0$ for $ij \in E$, we obtain (23). Next we show that the density function ρ is invariant with respect to time.

Notice that $e^{S_i-S_j}=\frac{\sqrt{\rho_j}}{\sqrt{\rho_i}}c_{ij}$ leads to

$$d(S_i-S_j)=\frac{1}{2}\frac{d\rho_j}{\rho_j}-\frac{1}{2}\frac{d\rho_i}{\rho_i}+d\ln(c_{ij})=\frac{1}{2}\frac{d\rho_j}{\rho_j}-\frac{1}{2}\frac{d\rho_i}{\rho_i}.$$

This implies that

$$\begin{aligned}
&-m_{ii}-\frac{1}{2}\sum_{k\in N(i)}e^{S_k-S_i}m_{ik}\frac{\sqrt{\rho_k}}{\sqrt{\rho_i}}-\frac{1}{2}\sum_{k\in N(i)}e^{S_i-S_k}m_{ki}\frac{\sqrt{\rho_k}}{\sqrt{\rho_i}}\\
&+m_{jj}+\frac{1}{2}\sum_{l\in N(j)}e^{S_l-S_j}m_{jl}\frac{\sqrt{\rho_l}}{\sqrt{\rho_j}}+\frac{1}{2}\sum_{l\in N(j)}e^{S_j-S_l}m_{lj}\frac{\sqrt{\rho_l}}{\sqrt{\rho_j}}\\
&=-\frac{1}{2\rho_i}(\sum_{k\in N(i)}e^{S_i-S_k}m_{ki}\sqrt{\rho_i\rho_k}-\sum_{k\in N(i)}e^{S_k-S_i}m_{ik}\sqrt{\rho_i\rho_k})\\
&+\frac{1}{2\rho_j}(\sum_{l\in N(j)}e^{S_j-S_l}m_{lj}\sqrt{\rho_j\rho_l}-\sum_{l\in N(j)}e^{S_l-S_j}m_{jl}\sqrt{\rho_j\rho_l}),
\end{aligned}$$

which is equivalent to (23). Using (24), it yields that

$$\sum_{k\in N(i)}c_{ik}m_{ki}\rho(k,t)-c_{ki}m_{ik}\rho(i,t)=\frac{1}{\rho_j}\Big(\sum_{l\in N(j)}c_{jl}m_{lj}\rho(l,t)-c_{lj}m_{jl}\rho(j,t)\Big),$$

that is,

$$d_t\rho_i=d_t\ln(\rho_j).$$

Similarly, we have $d_t\rho_j=d_t\ln(\rho_i)$, which implies that

$$d_t\rho_i=\rho_i\rho_jd_t\rho_i.$$

Now we claim that ρ must be invariant with respect to t. Indeed, if there exists $ij\in E$ such that $d\rho_i\neq 0$, then we have that $\rho_i\rho_j=1$. However, this contradicts the mass conservation $\sum_{i=1}^N\rho_i=1$. It follows that $d_t\rho_i=0$, and therefore ρ should be invariant with respect to time. We conclude that ρ must be the invariant measure of the solution process in the SBP. □

4.2. Discrete SBP based on minimum action with Fisher information

Another way (B) to describe the discrete SBP (see e.g. [8]) lies on the discretization of the variational problem (32). Consider the following control problem by directly discretizing the Fisher information $I(\rho)$ in (32):

$$J_1=\min_{\rho,v}\Big\{\int_0^1(\frac{1}{2}\langle v,v\rangle_\rho+\frac{1}{8}I(\rho))dt+\frac{1}{8}\sum_i(\rho^1(i)\log(\rho^1(i))-\rho^0\log(\rho^0(i)))\Big\},\quad (25)$$

where $\rho_i\in H^1((0,1))$, $v_{ij}\in L^2((0,1);\theta_{ij}(\rho))$ and

$$d_t\rho_t = \rho_t Q_t = -div_G^\theta(\rho_t v_t)$$

with $\rho^0, \rho^1 \in \mathcal{P}_o(G)$. In this case, we look for a stochastic process which obeys the above master equation and minimize the action with the Fisher information $I(\rho) := \frac{1}{2}\sum_{ij\in E}(\log(\rho_i) - \log(\rho_j))^2\widetilde{\theta}_{ij}(\rho)$. Here $\widetilde{\theta}$ is another density dependent weight which may be different from earlier defined θ on the graph G.

By using Lagrangian multiplier method, the critical point of the discrete variational approach should satisfy

$$\begin{aligned}
&v_{ij}(t) = (S_i(t) - S_j(t)),\\
&d_t\rho_i - \sum_{j\in N(i)}(S_i - S_j)\theta_{ij}(\rho) = 0,\\
&d_t S_i + \frac{1}{2}\sum_{j\in N(i)}(S_i - S_j)^2\frac{\partial\theta_{ij}}{\partial\rho_i} = \frac{1}{8}\frac{\partial}{\partial\rho_i}I(\rho).
\end{aligned} \tag{26}$$

It forms a Hamiltonian system on the density space with the Hamiltonian $\frac{1}{4}\sum_{i,j}(S_i - S_j)^2\theta_{ij}(\rho) - \frac{1}{8}I(\rho)$. In other words, the critical point gives a Hamiltonian process on the graph.

We can also reformulate the above system (26) in the form of Schrödinger system (33). By taking differential on $f = \sqrt{\rho}e^S$ and $g = \sqrt{\rho}e^{-S}$, we get

$$\begin{aligned}
d_t f &= e^{(\frac{1}{2}\log(\rho)+S)}(\frac{1}{2}\frac{d_t\rho}{\rho} + d_t S)\\
&= e^{(\frac{1}{2}\log(\rho)+S)}\Big(\frac{1}{2}\frac{\sum_{j\in N(i)} w_{ij}(S_i - S_j)\theta_{ij}(\rho)}{\rho} - \frac{1}{2}\sum_{j\in N(i)} w_{ij}(S_i - S_j)^2\frac{\partial\theta_{ij}}{\partial\rho_i} + \frac{1}{8}\frac{\partial}{\partial\rho_i}I(\rho)\Big),\\
d_t g &= e^{(\frac{1}{2}\log(\rho)-S)}(\frac{1}{2}\frac{d_t\rho}{\rho} - d_t S)\\
&= e^{(\frac{1}{2}\log(\rho)-S)}\Big(\frac{1}{2}\frac{\sum_{j\in N(i)} w_{ij}(S_i - S_j)\theta_{ij}(\rho)}{\rho} + \frac{1}{2}\sum_{j\in N(i)} w_{ij}(S_i - S_j)^2\frac{\partial\theta_{ij}}{\partial\rho_i} - \frac{1}{8}\frac{\partial}{\partial\rho_i}I(\rho)\Big).
\end{aligned}$$

Rewriting the above systems into compact form leads to

$$\begin{aligned}
d_t f &= -\frac{1}{2}\Delta_G f,\\
d_t g &= \frac{1}{2}\Delta_G g,
\end{aligned} \tag{27}$$

where Δ_G is the nonlinear discretization of the Laplacian operator,

$$\begin{aligned}
&(\Delta_G f)_j\\
&= -f_j\Bigg(\frac{1}{f_j g_j}\sum_{l\in N(j)}\Big(\widetilde{w}_{jl}(\log(f_j/g_j) - \log(f_l/g_l))\widetilde{\theta}_{ij}(fg)
\end{aligned}$$

$$+ w_{jl}(\log(f_j g_j) - \log(f_l g_l))\theta_{ij}(fg)\Big)$$

$$+ \sum_{l\in N(j)} \Big(\widetilde{w}_{jl}|\log(f_j/g_j) - \log(f_l/g_l)|^2 \frac{\partial\widetilde{\theta}_{ij}(fg)}{\partial f_j g_j} + w_{jl}|\log(f_j g_j)$$

$$- \log(f_l g_l)|^2 \frac{\partial\theta_{ij}(fg)}{\partial f_j g_j}\Big)\Bigg).$$

Remark 4.2. In approach (A), the Hamiltonian systems ((14), (17) and (22)) are corresponding to the control problem (12), which is derived from discretizing the relative entropy $H(P|R)$ in (29); In approach (B), the Hamiltonian systems ((26) and (27)) are corresponding to the control problem (25), which is derived via discretizing the Fisher information $I(\rho)$ in (30). It worth mentioning that under continuous cases, (29) and (30) are equivalent under the transform (31) and their corresponding Hamiltonian systems are also equivalent. However, this is not true for discrete cases. Discretizing the SBP at different stages leads to different Hamiltonian systems.

Remark 4.3 *(Nonlinear Markov process as reference process in approach (B)).* Let us recall that in continuous space $\mathbb{R}^d$, f, g solve the Schrödinger system

$$\frac{\partial}{\partial t} f_t = \mathcal{L}_t f_t, \quad \frac{\partial}{\partial t} g_t = -\mathcal{L}_t g_t. \quad \text{with } f_0,\ g_1 \text{ are given,}$$

with $\mathcal{L}_t$ corresponds to the generator of the reference process R (cf. Equation (32) of [17]).

By comparing the systems (22) and (27) related to f, g, it is observed that $\mathcal{L}_t$ in approach (A) can be viewed as a linear approximation of Laplacian operator, which is associated to the Markov reference process R with transition rate matrix $\{m^t_{ij}\}$; On the other hand, $\mathcal{L}_t = \Delta_G$ in approach (B) is a nonlinear approximation of Laplacian operator. We can thus interpret Δ_G as a nonlinear generator depending on both the state and the distribution. According to the definition of nonlinear Markov process mentioned in section 2.2, we can associate approach (B) with a nonlinear Markov reference process R generated by Δ_G even though such reference process is not needed in the original control formulation (25).

4.3. Periodic marginal distribution of Hamiltonian process in SBP

The periodic solution, as one classical topic of Hamiltonian systems, has been studied for many decades (see e.g. [6,25,21]). For our considered Hamiltonian process, the periodicity of the solution appears in the density evolution. Below, we present several examples of periodic reference process, and prove that if the periodic Hamiltonian process exists, it coincides with the reference process in SBP.

By using the Floquet theorem in [27], the fundamental matrix $X(t)$ satisfies $X(t+T) = X(t)\exp(LT)$, where $\exp(LT)$ is a non-singular constant matrix. The Floquet exponents of $d_t\rho = \rho Q_t$ are the eigenvalues $\mu_i, i \le k \le N$ of the matrix L. If there exists some i such that $\exp(\mu_i T) = 1$ or -1, then there exists periodic density function with period T or $2T$. As a consequence, we obtain the following results.

Lemma 4.1. *Assume that $\{Q_t\}_{t\ge 0}$ is transition rate matrix and Q_t is T-periodic. If there exists a Floquet exponent $\mu = \frac{k\pi\mathbf{i}}{T}, k \in \mathbb{Z}$, then $d_t\rho = \rho Q_t$ has a periodic density.*

Example 4.1. Consider a 2-nodes graph G. Given a reference measure which possesses the marginal distribution as follows,

$$d_t\rho_1 = \rho_1 m_{11} + \rho_2 m_{21},$$
$$d_t\rho_2 = \rho_1 m_{12} + \rho_2 m_{22},$$

where $m_{21} = -m_{11}, m_{22} = -m_{12}, m_{11} = -\frac{\frac{1}{2}-\frac{1}{4}\cos(t)+\frac{1}{8}\sin(t)-\frac{1}{16}\sin(t)\cos(t)}{(\frac{1}{2}+\frac{1}{4}\cos(t))^2}$ and $m_{22} = -\frac{1}{\frac{1}{2}+\frac{1}{4}\cos(t)}$.

There exists a nontrivial periodic solution $\rho_1(t) = \frac{1}{2} + \frac{1}{4}\cos(t)$, $\rho_2(t) = 1 - \rho_1(t)$. And the periods of ρ_1 and ρ_2 are both $T = 2\pi$. Therefore, there exists a time inhomogeneous Markov process X_t with periodic marginal distribution ρ_t on G with the transition rate matrix $Q_t = (m_{ij})_{i,j\le 2}$.

We can also show the existence of time inhomogeneous Markov process with periodic marginal distribution on any fully-connected graph.

Proposition 4.2. *Suppose G is a fully connected graph, and $\{\rho_t\}$ is a periodic density trajectory (with period T) in $\mathcal{P}_o(G)$, then we can always find a transition rate matrix $Q(t)$ such that ρ_t is the solution to the master equation $\dot{\rho}_t = \rho_t Q(t)$.*

Proof. Assume G contains n vertices. Let us assume the non-diagonal entries of $Q(t)$ to be $\{m_{ij}\}$, we rearrange these entries to form a $n(n-1)$ dimensional vector as:

$$m = (m_{12}, ..., m_{1n}, m_{21}, m_{23}, ..., m_{2n}, ..., m_{n1}, ..., m_{n\,n-1})^T.$$

Plugging m into the Master's equation, we derive the linear equation for m:

$$P(t)\, m = \begin{pmatrix} \dot{\rho}_1 & \dot{\rho}_2 & \dots & \dot{\rho}_n \end{pmatrix}^T. \tag{28}$$

Where P is an $n \times n(n-1)$ matrix defined as

$$P(t) = \left(P_1(t) \;\middle|\; P_2(t) \;\middle|\; \dots \;\middle|\; P_n(t) \right).$$

$$P_m(t) = \begin{pmatrix} \rho_m(t) I_m & 0_{m\times(n-m-1)} \\ -\rho_m(t)\boldsymbol{e}_m^T & -\rho_m(t)\boldsymbol{e}_{n-m-1}^T \\ 0_{(n-m-1)\times m} & \rho_m(t) I_{n-m-1} \end{pmatrix}_{n\times(n-1)} \quad \text{for } 1 \le m \le n$$

Here we denote $\boldsymbol{e}_m^T = \underbrace{(1, ..., 1)}_{m\ 1\text{s}}$. We can verify that

$$m^0 = \left(\frac{1}{(n-1)\rho_1}\boldsymbol{e}_{n-1}^T, \frac{1}{(n-1)\rho_2}\boldsymbol{e}_{n-1}^T, ..., \frac{1}{(n-1)\rho_n}\boldsymbol{e}_{n-1}^T \right)$$

belongs to the kernel of $P(t)$, and that $P(t)$ is a full rank matrix. There must exist a solution m^* to (28), where its entries are expressions of $\{\rho_i, \dot{\rho}_i\}_{i\in V}$. In other words, we can directly give such a solution. To be more specific, let's consider the transport process on the loop from vertex 1 to

2, 2 to 3,... n-1 to n and n to 1. This corresponds to setting m_{ij} to 0 except $m_{12}, m_{23}, ..., m_{n-1\ n}$, and m_{n1}. Now the equation (28) becomes:

$$\begin{pmatrix} -\rho_1 & & & \rho_n \\ & -\rho_2 & & \\ & & \ddots & \\ & & & \ddots & \\ & & & & -\rho_n \end{pmatrix} \begin{pmatrix} m_{12} \\ m_{23} \\ \vdots \\ m_{n-1\ n} \\ m_{n1} \end{pmatrix} = \begin{pmatrix} \dot\rho_1 \\ \dot\rho_2 \\ \vdots \\ \dot\rho_{n-1} \\ \dot\rho_n \end{pmatrix}$$

Therefore the solution is $(-\frac{\dot\rho_1-\dot\rho_n}{\rho_1}, -\frac{\dot\rho_2}{\rho_2}, \cdots, -\frac{\dot\rho_{n-1}}{\rho_{n-1}}, -\frac{\dot\rho_n}{\rho_n})^T$.

Then we can directly take $m(t) = Km^0(t) + m^*(t)$, since $\{\rho_t\}$ is in the interior of $\mathcal{P}(G)$, we can always find a large enough $K > 0$ that guarantees the entries of $m(t)$ to be always non negative. And $m(t)$ forms the transition rate matrix $Q(t)$ whose master equation admits the periodic solution $\{\rho_t\}$. □

Example 4.2. Consider the periodic marginal distribution ρ_t:

$$\rho_t = (\frac{\cos t}{2\sqrt{6}} + \frac{\sin t}{6\sqrt{2}} + \frac{1}{3}, -\frac{\cos t}{2\sqrt{6}} + \frac{\sin t}{6\sqrt{2}} + \frac{1}{3}, -\frac{\sin t}{3\sqrt{2}} + \frac{1}{3}),$$

which is a circle centered at $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ with radius $\frac{1}{2\sqrt{3}}$ on $\mathcal{P}(G)$. Following the idea of Proposition 4.2, one may take

$$\begin{aligned} m_{11}(t) &= -\frac{6\sqrt{2} + \sqrt{3}\sin t - 3\cos t}{\sqrt{3}\cos t + 4\sin t + 2\sqrt{2}},\ m_{12} = -m_{11},\ m_{13} = 0, \\ m_{22}(t) &= -\frac{24 - 4\sqrt{2}\cos t}{-\sqrt{6}\cos t + \sqrt{2}\sin t + 4},\ m_{21} = -\frac{1}{2}m_{22},\ m_{23} = -\frac{1}{2}m_{22}, \\ m_{33}(t) &= -\frac{3\sqrt{2}}{\sqrt{2} - \sin t}\ m_{13} = 0, m_{23} = -m_{33}, \end{aligned}$$

such that $d_t = \rho_t Q_t$ with $Q_t = (m_{ij})_{i,j\le 3}$.

Next we aim to use general SBP (18) to produce a Hamiltonian process with periodic marginal distribution on G. In particular, when the convex function $u = x\log(x) - x - 1$, by using the Nelson's transformation $\psi_i = \sqrt{\rho_i}e^{S_i}$, the Hamiltonian system can be also rewritten as

$$\begin{aligned} dS_i &= -m_{ii} - \frac{1}{2}\sum_{j\in N(i)} e^{S_j-S_i} m_{ij}(t)\frac{\sqrt{\rho_j}}{\sqrt{\rho_i}} - \frac{1}{2}\sum_{j\in N(i)} e^{S_i-S_j} m_{ji}(t)\frac{\sqrt{\rho_j}}{\sqrt{\rho_i}}, \\ d\rho_i &= \sum_{j\in N(i)} e^{S_i-S_j} m_{ji}(t)\sqrt{\rho_j}\sqrt{\rho_i} - \sum_{j\in N(i)} e^{S_j-S_i} m_{ij}(t)\sqrt{\rho_i}\sqrt{\rho_j}, \end{aligned}$$

with the Hamiltonian $\tilde{\mathcal{H}}(\rho, S, t) = \sum_{i\in V}\sum_{j\in N(i)} e^{(S_j-S_i)} m_{ij}(t)\sqrt{\rho_i\rho_j}$. Taking ψ as a time-independent potential and choosing ρ^0, ρ^1 as the initial and terminal distribution from the periodic solution, then the distribution of the solution process is exactly same as the reference process. Thus it induces a Hamiltonian system which is periodic in time. Therefore there exists SBP with the given ρ^0, ρ^1 such that the solution process is Hamiltonian and its marginal distribution is periodic in time.

In the following, we assume that the Legendre transformation u^* of u in (18) is continuous differentiable and satisfies

$$
\begin{aligned}
&u^*(x) \geq 0, \text{ if } x \leq 0, \; u^*(x) \leq 0, \text{ if } x \geq 0, \\
&\frac{\partial u^*}{\partial x}(0) = 1, \; \lim_{x \to -\infty} \left| \frac{\partial u^*}{\partial x}(x) \right| < \infty, \; \lim_{x \to +\infty} \frac{\partial u^*}{\partial x}(x) = +\infty.
\end{aligned}
$$

Now we are able to give the characterization of the periodic Hamiltonian process on finite graph via general SBP.

Theorem 4.1. *Assume that the reference process is periodic with the marginal distribution and its period $T > 0$. There always exists ρ^0, ρ^1 such that the critical point of the general SBP problem* (18) *is a Hamiltonian process and its marginal distribution is periodic in time.*

Proof. Notice that the critical point of SBP satisfies

$$
\begin{aligned}
&\frac{\partial}{\partial t}\rho(i,t) = \sum_{j \in N(i)} -\frac{\partial u^*}{\partial x}(\psi_j - \psi_i) m_{ij} \rho(i,t) + \frac{\partial u^*}{\partial x}(\psi_i - \psi_j) m_{ji} \rho(j,t), \\
&\frac{\partial}{\partial t}\psi(i,t) = -\sum_{j \in N(i)} u^*(\psi_j - \psi_i) m_{ij},
\end{aligned}
$$

where $\rho(0) = \rho^0, \rho(1) = \rho^1$. Choosing ρ^0, ρ^1 as two different distributions at different time of the reference process, and taking $\psi_i = \psi_j$, we get

$$
\begin{aligned}
&\frac{\partial}{\partial t}\rho(i,t) = \sum_{j \in N(i)} -m_{ij}\rho(i,t) + m_{ji}\rho(j,t), \\
&\frac{\partial}{\partial t}\psi(i,t) = 0.
\end{aligned}
$$

This implies that the critical point forms a Hamiltonian system with Hamiltonian $H(\rho, \psi, t) = \sum_{i,j} m_{ij}(t)\rho_i$. Due to the fact that the marginal distribution of reference process is periodic in time, the critical point is exactly equal to the reference process and its marginal distribution is periodic. □

One may wonder whether there exists certain Hamiltonian process whose marginal distribution is periodic but is not the reference process. We first use a 2-nodes graph example to point out it is not possible to get such Hamiltonian by using SBP when $u(x) = x\log(x) - x - 1$. Even worse, we show that for general finite graph, the periodic Hamiltonian process exists if and only if it equals to a reference process in general SBP.

Example 4.3. Given G consisting of 2 nodes. Assume the reference process with transition rate matrix m is periodic with period $T > 0$ and $\{t \in [0, T] | m_{ij}(t) = 0, ij \in E\}$ has Lebesgue measure zero. Notice that ρ, S of the Hamiltonian process $X(t)$ satisfies

$$\frac{\partial}{\partial t}\rho(1,t) = -e^{\psi(2,t)-\psi(1,t)} m_{12}\rho(1,t) + e^{\psi(1,t)-\psi(2,t)} m_{21}\rho(2,t),$$
$$\frac{\partial}{\partial t}(\psi(1,t) - \psi(2,t)) = -(e^{\psi(2,t)-\psi(1,t)} - 1)m_{12} + (e^{\psi(1,t)-\psi(2,t)} - 1)m_{21}.$$

Since $m_{12}, m_{21} \geq 0$, then $\psi(1) - \psi(2)$ equals to constant if and only if $\psi(1) = \psi(2)$. Meanwhile, if $\psi_1 - \psi_2 > 0$, then $\psi_1 - \psi_2$ is increasing to $+\infty$, and $\psi_1 - \psi_2$ is decreasing to $-\infty$ if $\psi_1 < \psi_2$. Then we claim that ρ_1 is not periodic in time. If we assume that ρ_1 is periodic with period T_1, then it holds true $\int_{kT_1}^{(k+1)T_1} -e^{\psi(2,t)-\psi(1,t)} m_{12}\rho(1,t) + e^{\psi(1,t)-\psi(2,t)} m_{21}\rho(2,t)dt = 0$. Without losing generality, let us assume that $\psi_1 - \psi_2 > 0$. It is not hard to see that $e^{\psi(1,t)-\psi(2,t)}$ is increasing to $+\infty$ and $e^{\psi(1,t)-\psi(2,t)}$ is decreasing to 0 as $t \to \infty$. The boundedness of $\rho(1,t)$, $\rho(2,t)$ yield that there exists large enough k such that

$$\int_{kT_1}^{(k+1)T_1} -e^{\psi(2,t)-\psi(1,t)} m_{12}\rho(1,t) + e^{\psi(1,t)-\psi(2,t)} m_{21}\rho(2,t)dt > 0,$$

which leads to a contradiction. Therefore, $\rho(t)$ is periodic in time if and only if $\psi_1 = \psi_2$. This implies that $X(t)$ is exactly the reference process.

Theorem 4.2. *Assume the reference process with transition rate matrix m is periodic with period $T > 0$ and $\{t \in [0, T] | m_{ij}(t) = 0, ij \in E\}$ has Lebesgue measure zero. Then the Hamiltonian process which has periodic density distribution in general SBP problem (18) is equal to the reference process which has the periodic density distribution.*

Proof. Assume that there is a maximum $\psi_{i^*} \geq \psi_i$, $i \neq i^*$ and $\psi_{i^*} > \psi_{i_{min}}$. Then according to the evolution of ψ,

$$\frac{\partial}{\partial t}\psi(i,t) = -\sum_{j\in N(i)} u^*(\psi(j,t) - \psi(i,t))m_{ij},$$

then the maximum principle holds, i.e., $\psi_{i^*}(t) \geq \psi_i(t) \geq \psi_{i_{min}}(t)$. Notice that

$$\frac{d}{dt}\rho(i,t) = \sum_{j\in N(i)} -\frac{\partial u^*}{\partial x}(\psi_j - \psi_i)m_{ij}\rho(i,t) + \frac{\partial u^*}{\partial x}(\psi_i - \psi_j)m_{ji}\rho(j,t).$$

The periodicity of ρ_i implies that there exists $T_1 > 0$ for any $k \in \mathbb{N}^+$ such that

$$\int_{kT_1}^{(k+1)T_1} \sum_{j\in N(i)} -\frac{\partial u^*}{\partial x}(\psi_j - \psi_i)m_{ij}\rho(i,t) + \frac{\partial u^*}{\partial x}(\psi_i - \psi_j)e^{\psi(i,t)-\psi(j,t)} m_{ji}\rho(j,t)dt = 0.$$

Due to the maximum principle, if there exists one node l with a local maximum of ψ_l connected with another node k with a local minimum of ψ_k, it will lead to $\psi_l - \psi_k \to +\infty$ as $t \to \infty$. This contradicts with the periodicity of ρ_k and ρ_l. If any node l with a local maximum of ψ_l is not connected with another node k with a local minimum of ψ_k, we pick a road $l, j_1, \cdots, j_w, k$ which connects l and k. Notice that $\psi_l \to +\infty$, $\psi_k \to -\infty$, $\psi_{j_m} \in (\psi_k, \psi_l), m \le w$. Then there must exists j_m such that m is the smallest number which satisfies $\psi_k - \psi_{j_m} \to -\infty$. Now consider the periodicity of ρ_{j_m}. There exists k' large enough such that

$$\int_{k'T_1}^{(k'+1)T_1} \sum_{j\in N(j_m)} -\frac{\partial u^*}{\partial x}(\psi_j - \psi_{j_m}) m_{j_m j}\rho(j_m, t) + \frac{\partial u^*}{\partial x}(\psi_{j_m} - \psi_j) m_{j j_m}\rho(j, t)dt > 0.$$

This leads to a contradiction, we complete the proof.

5. More examples and future work

In this section, we conclude the paper by presenting a few more examples of Hamiltonian processes on graph and more questions to be considered in the future.

Example 5.1. (Euler-Lagrangian equations [14]) Assume that the Lagrangian in density manifold is given by $\mathcal{L}(\rho_t, \dot\rho_t) = \frac{1}{2} g_W(\dot\rho_t, \dot\rho_t) - \mathcal{F}(\rho_t)$. Here $g_W(\sigma_1, \sigma_2) := -\sigma_1(\Delta_\rho)^+\sigma_2$ where $\sigma_k \in T_\rho \mathcal{P}_o(G), k = 1, 2$ and $(\Delta_\rho)^+$ is the pseudo inverse of the weight graph Laplacian matrix $\Delta_\rho(\cdot) := div_G^\theta(\rho\nabla_G(\cdot))$. Then the critical point of

$$\inf_{\rho_t} \int_0^T \mathcal{L}(\rho_t, \partial_t\rho_t)dt$$

with given ρ_0 and ρ_T satisfies the Euler-Lagrangian equation

$$\partial_t \frac{\delta}{\delta \partial_t \rho_t}\mathcal{L}(\rho_t, \partial_t\rho_t) = \frac{\delta}{\delta\rho_t}\mathcal{L}(\rho_t, \partial_t\rho_t) + C(t).$$

By introducing the Legendre transform $S_t = (-\Delta_{\rho_t})^+\partial_t\rho_t$, it can be rewritten as a Hamiltonian system. That is

$$\begin{aligned}
&\partial_t\rho_t + div_G^\theta(\rho\nabla_G S) = 0, \\
&\partial_t S_t + \frac{1}{4}\sum_{j\in N(i)} (S_i - S_j)^2(\partial_{\rho_i}\theta(\rho_i, \rho_j) + \partial_{\rho_i}\theta(\rho_j, \rho_i)) + \frac{\delta}{\delta\rho_t}\mathcal{F}(\rho_t) = C(t),
\end{aligned}$$

with the Hamiltonian $\mathcal{H}(\rho, S) = \frac{1}{4}\sum_{ij}(S_i - S_j)^2\theta_{ij}w_{ij} + \mathcal{F}(\rho_t)$. Therefore, if the transition rate matrix in generalized master equation is well-defined, the Euler-Lagrangian equation in density space determines a Hamiltonian process on G.

Example 5.2. (Madelung system [10]) The energy is given by

$$\mathcal{H}(\rho, S) = \frac{1}{4}\sum_{ij\in E}(S_i - S_j)^2\theta_{ij}w_{ij} + \mathcal{F}(\rho_t) + \beta I(\rho_t), \beta > 0.$$

Here $\mathcal{F}(\rho) = \sum_i \rho_i \mathbb{V}_i + \sum_{i,j}\rho_i\rho_j\mathbb{W}_{ij}$, and $I(\rho) = \frac{1}{2}\sum_{ij\in E}(\log(\rho_i) - \log(\rho_j))^2\widetilde{\theta}_{ij}$. Here $\widetilde{\theta}_{ij}$ is another density dependent weight on the graph that can be the same or different from θ_{ij}. The Madelung system is

$$\begin{aligned}
&\partial_t\rho_t + div_G^\theta(\rho\nabla_G S) = 0,\\
&\partial_t S_t + \frac{1}{4}\sum_{j\in N(i)}(S_i - S_j)^2(\partial_{\rho_i}\theta(\rho_i, \rho_j) + \partial_{\rho_i}\theta(\rho_j, \rho_i)) + \frac{\delta}{\delta\rho_t}\mathcal{F}(\rho_t) + \beta\frac{\delta}{\delta\rho_t}I(\rho_t) = C(t).
\end{aligned}$$

When taking $\theta = \theta^U$, the Madelung system in density space determines a Hamiltonian process on G. This system has a close relationship with the discrete Schrödinger equation [9].

Example 5.3. (L^p-Wasserstein distance) The L^p Wasserstein distance, $p \in (1, \infty)$, is related to the following minimization problem,

$$W_p^p(\rho^0, \rho^1) = \inf_v\{\int_0^1\sum_{i=1}^N\sum_{j\in N(i)}\frac{1}{2}\theta_{ij}(\rho)v_{ij}^p dt : \partial_t\rho + div_G^\theta(\rho v) = 0, \rho(0) = \rho^0, \rho(1) = \rho^1\}.$$

We refer to [13] for a continuous version of L^p-Wasserstein distance. Its critical point is related to the Hamiltonian system in density space

$$\begin{aligned}
&\partial_t\rho_t + div_G^\theta(\rho_t|\nabla_G S|^{q-2}\nabla_G S) = 0,\\
&\partial_t(S_i) + \frac{1}{2q}\sum_{j\in N(i)}|(\nabla_G S)_{ij}|^q(\partial_1\theta_{ij} + \partial_2\theta_{ji}) = 0,
\end{aligned}$$

with the Hamiltonian

$$\mathcal{H}(\rho, S) = \frac{1}{2q}\sum_{i,j}|\nabla_G S|^q\theta_{ij}, \frac{1}{q} + \frac{1}{p} = 1, p \in (1, \infty).$$

When the equation of ρ is determined by a transition rate matrix, this leads to a Hamiltonian process.

To end the discussion, we want to mention two problems that are worth to be studied further.

- As shown in [10], the classical Hamiltonian ODEs induce the Wasserstein–Hamiltonian flows on the density manifold. There are many special properties for Hamiltonian system in continuous space, such as conservation of energy, preservation of the volume etc. The particle-level counterpart on graph is the Hamiltonian process introduced in Definition 3.1.

In addition to the conservation property discussed in Remark 3.1, are there other quantities or structures being preserved by the Hamiltonian process on the graph G?

- As discussed in [22], stochastic differential equations can be well approximated by continuous time random walk on the lattices. Then it is natural to ask whether the proposed Hamiltonian process on a lattice can be used to approximate a Hamiltonian system in $\mathbb{R}^d$ or not. If so, how well is the approximation?

Acknowledgments

The authors are very grateful to Prof. Shui-Nee Chow for introducing this problem and his helpful suggestions. The research is 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.

Appendix A

A.1. The background of SBP

Denote $\Omega = C([0,1], \mathbb{R}^d)$. Given $R \in M^+(\Omega)$ the law of the reversible Brownian motion (here we consider the Brownian motion with the volume Lebesgue measure, denoted by Leb, as the initial distribution). Consider the relative entropy of any probability measure with respect to R,

$$H(P|R) = \int_{\Omega} \log(\frac{dP}{dR}) dP.$$

The SBP can be formulated as

$$\min H(P|R),\ P \in \mathcal{P}(\Omega) : P_0 = \mu_0,\ P_1 = \mu_1. \tag{29}$$

Here $P_0 := P(X_0 \in \cdot)$, $P_1 := P(X_1 \in \cdot)$ and $X_t(\omega) := \omega(t)$ is the canonical process with $\omega \in \Omega$. It is proven (see e.g. [17]) that if $H(\widetilde{\mu}_0|Leb) < \infty$ and $H(\widetilde{\mu}_1|Leb) < \infty$, the SBP has a unique solution $\widehat{P}$ which enjoys the following decomposition

$$\widehat{P} = f_0(X_0) g_1(X_1) R \in \mathcal{P}(\Omega),$$

where f_0, g_1 are nonnegative measurable functions such that

$$\mathbb{E}_R[f_0(X_0) g_1(X_1)] = 1.$$

Introduce the function f_t, g_t defined by

$$\begin{aligned} f_t(z) &:= \mathbb{E}_R[f_0(X_0)|X_t = z], \\ g_t(z) &:= \mathbb{E}_R[g_1(X_1)|X_t = z],\ P_t\text{-}a.e.,\ z \in \mathbb{R}^d, \end{aligned}$$

and the constraint

$$\widetilde{\mu}_0 = f_0 g_0 Leb,\ \widetilde{\mu}_1 = f_1 g_1 Leb.$$

Then the SBP (2) with $\hbar = 1$ is equivalent to the following minimal action problem, i.e.,

$$\begin{aligned} &\inf\{H(P|R) : P_0 = \widetilde{\mu}_0, P_1 = \widetilde{\mu}_1\} - H(\mu_0|Leb) \qquad (30)\\ &= \inf\Big\{ \int_0^1 \int_{\mathbb{R}^d} \frac{|v_t|^2}{2}\mu_t(dx)dt : (\partial_t - \frac{\Delta}{2})\mu + \nabla\cdot(v\mu) = 0,\\ &\quad P_0 = \mu_0, P_1 = \mu_1 \Big\} \end{aligned}$$

We denote ρ_t the density of μ_t with respect to the Lebesgue measure. In addition, with the assumption that μ_0, μ_1 have finite second moments, the critical point of the minimal action problem satisfies the following system

$$\begin{aligned} &(\partial_t - \frac{\Delta}{2})\rho + \nabla\cdot(\nabla\phi\rho) = 0,\ \rho(0) = \rho_0,\\ &(\partial_t + \frac{\Delta}{2})\phi + \frac{1}{2}|\nabla\phi|^2 = 0,\ \phi(1) = \log(g_1) \end{aligned}$$

with $v_t = \nabla\phi_t$. There is also a backward version of this PDE system, namely

$$\begin{aligned} &(-\partial_t - \frac{\Delta}{2})\rho + \nabla\cdot(\nabla\psi\rho) = 0,\ \rho(1) = \rho_1,\\ &(-\partial_t + \frac{\Delta}{2})\psi + \frac{1}{2}|\nabla\psi|^2 = 0,\ \psi(0) = \log(f_0). \end{aligned}$$

Here we have the relation $\nabla\psi_t + \nabla\phi_t = \nabla\log(\rho_t)$.

Applying the transformation

$$S_t = \phi_t - \frac{1}{2}\log(\rho_t) \qquad (31)$$

as being done in [23], we arrive at the Hamiltonian system on the density space,

$$\begin{aligned} &\frac{\partial}{\partial t}\rho + \nabla\cdot(\rho(t,x)\nabla S) = 0,\\ &\frac{\partial}{\partial t}S + \frac{1}{2}|\nabla S|^2 - \frac{1}{8}\frac{\delta}{\delta\rho_t}I(\rho_t) = 0. \end{aligned}$$

The corresponding Hamiltonian is $\mathcal{H}(\rho, S) = \frac{1}{2}\int_{\mathbb{R}^d}|\nabla S|^2\rho dx - \frac{1}{8}I(\rho)$ where $I(\rho) = \int_{\mathbb{R}^d}|\nabla\log(\rho)|^2\rho dx$ is the Fisher information. Meanwhile, the action minimizing problem (30) can be rewritten as

$$\inf_{v_t}\Big\{\int_0^1 \mathbb{E}[\frac{1}{2}v(t,X(t))^2]+\frac{1}{8}I(\rho(t))dt+\frac{1}{2}\int(\rho^1\log(\rho^1)-\rho^0\log(\rho^0))dx \tag{32}$$
$$\mid dX_t = v(t,X_t)dt,\ X(0)\sim\rho^0,\ X(1)\sim\rho^1\Big\}.$$

Here $\rho(t)$ is the density of the marginal distribution of X_t.

Next, by introducing the conjugate Madelung transformation $f=\sqrt{\rho}e^S$, $g=\sqrt{\rho}e^{-S}$ (also known as “Hopf-Cole” transformation), f and g satisfy so-called “Schrödinger system” (see e.g. [10,11,3]),

$$(\partial_t-\frac{\Delta}{2})g=0,\ g(0)=g_0, \tag{33}$$
$$(\partial_t+\frac{\Delta}{2})f=0,\ f(1)=f_1.$$

This also implies the following relationships

$$\phi=\log(f)=S+\frac{1}{2}\log(\rho),\ \psi=\log(g)=-S+\frac{1}{2}\log(\rho).$$

References

[1] V.I. Arnold, Mathematical Methods of Classical Mechanics, second edition, Graduate Texts in Mathematics, vol. 60, Springer-Verlag, New York, 1989, translated from the Russian by K. Vogtmann and A. Weinstein.

[2] J. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math. 84 (3) (2000) 375–393.

[3] A. Blaquiére, Controllability of a Fokker-Planck equation, the Schrödinger system, and a related stochastic optimal control, Dyn. Control 2 (1992) 235–253.

[4] E.A. Carlen, Conservative diffusions, Commun. Math. Phys. 94 (3) (1984) 293–315.

[5] Y. Chen, T. Georgiou, M. Pavon, A. Tannenbaum, Robust transport over networks, IEEE Trans. Autom. Control 62 (9) (2017) 4675–4682.

[6] T.M. Cherry, On periodic solutions of Hamiltonian systems of differential equations, Philos. Trans. R. Soc. Lond., Ser. A, Contain. Pap. Math. Phys. Character 227 (1928) 137–221.

[7] S. Chow, W. Huang, Y. Li, H. Zhou, Fokker-Planck equations for a free energy functional or Markov process on a graph, Arch. Ration. Mech. Anal. 203 (3) (2012) 969–1008.

[8] S. Chow, W. Li, C. Mou, H. Zhou, A discrete Schrodinger bridge problem via optimal transport on graphs, J. Dyn. Differ. Equ. 20 (33) (2020) 34.

[9] S. Chow, W. Li, H. Zhou, A discrete Schrödinger equation via optimal transport on graphs, J. Funct. Anal. 276 (8) (2019) 2440–2469.

[10] S. Chow, W. Li, H. Zhou, Wasserstein Hamiltonian flows, J. Differ. Equ. 268 (3) (2020) 1205–1219.

[11] G. Conforti, M. Pavon, Extremal flows in Wasserstein space, J. Math. Phys. 59 (6) (2018) 063502.

[12] J. Cui, L. Dieci, H. Zhou, Time discretizations of Wasserstein-Hamiltonian flows, arXiv:2006.09187, 2020.

[13] J. Dolbeault, B. Nazaret, G. Savaré, A new class of transport distances between measures, Calc. Var. Partial Differ. Equ. 34 (2) (2009) 193–231.

[14] W. Gangbo, W. Li, C. Mou, Geodesics of minimal length in the set of probability measures on graphs, ESAIM Control Optim. Calc. Var. 25 (2019) 78.

[15] V.N. Kolokoltsov, Nonlinear Markov Processes and Kinetic Equations, Cambridge Tracts in Mathematics, vol. 182, Cambridge University Press, Cambridge, 2010.

[16] C. Léonard, Girsanov theory under a finite entropy condition, in: Séminaire de Probabilités XLIV, in: Lecture Notes in Math., vol. 2046, Springer, Heidelberg, 2012, pp. 429–465.

[17] C. Léonard, A survey of the Schrödinger problem and some of its connections with optimal transport, Discrete Contin. Dyn. Syst. 34 (4) (2014) 1533–1574.

[18] C. Léonard, Lazy random walks and optimal transport on graphs, Ann. Probab. 44 (3) (2016) 1864–1915.
[19] J. Maas, Gradient flows of the entropy for finite Markov chains, J. Funct. Anal. 261 (8) (2011) 2250–2292.
[20] E. Madelung, Quanten theorie in hydrodynamischer form, Z. Phys. 40 (3–4) (1927) 322–326.
[21] J. Mawhin, M. Willem, Critical Point Theory and Hamiltonian Systems, Applied Mathematical Sciences, vol. 74, Springer-Verlag, New York, 1989.
[22] B.-R. Nawaf, V.-E. Eric, Continuous-time random walks for the numerical solution of stochastic differential equations, Mem. Am. Math. Soc. 256 (1228) (2018), v+124.
[23] E. Nelson, Derivation of the Schrödinger equation from Newtonian mechanics, Phys. Rev. 150 (4) (1966) 1079–1085.
[24] M. Pavon, Quantum Schrödinger bridges, in: Directions in Mathematical Systems Theory and Optimization, in: Lect. Notes Control Inf. Sci., vol. 286, Springer, Berlin, 2003, pp. 227–238.
[25] P.H. Rabinowitz, Periodic solutions of Hamiltonian systems, Commun. Pure Appl. Math. 31 (2) (1978) 157–184.
[26] E. Schrödinger, Über die Umkehrung der Naturgesetze, Sitz.ber Preuss. Akad. Wiss. Phys.-Math. Kl. Akad. Wiss. (1931).
[27] G. Teschl, Ordinary Differential Equations and Dynamical Systems, Graduate Studies in Mathematics, vol. 140, American Mathematical Society, Providence, RI, 2012.
[28] A.J. van der Schaft, Port-Hamiltonian systems: an introductory survey, in: International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 1339–1365.
[29] A.J. van der Schaft, B.M. Maschke, Port-Hamiltonian systems on graphs, SIAM J. Control Optim. 51 (2) (2013) 906–937.
[30] C. Villani, Optimal Transport: Old and New, Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 338, Springer-Verlag, Berlin, 2009.