



Dynamical Schrödinger Bridge Problems on Graphs

Shui-Nee Chow¹ · Wuchen Li² · Chenchen Mou³ · Haomin Zhou¹

Received: 28 February 2020 / Revised: 31 August 2020 / Accepted: 23 February 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

We study a discrete dynamical Schrödinger bridge problem (SBP) as a dynamical variational problem on a finite graph. We prove that the discrete SBP exists a unique minimizer, which satisfies a boundary value Hamiltonian flow on probability simplex equipped with L^2 -Wasserstein metric. In our formulation, we establish the connection between discrete SBP problems and Hamiltonian flows.

Keywords Optimal transport · Schrödinger bridge problem · Fisher information · Hamiltonian system · Graph

1 Introduction

In recent years, Schrödinger bridge problem (SBP) has been studied extensively in mathematics and engineer communities [19,30]. It plays important roles in applications, such as mean field games [2,4,17], Bayesian sampling problems [1] and machine learning [31,32]. The problem is proposed by Schrödinger [33], which describes the optimal value and trajectory in the space of probability densities for minimal kinetic energy transported by drift-diffusion processes. Nowadays, SBP can be viewed as a relaxation of optimal transport [29,34], which has both static and dynamical formulations. The static formulation refers to the entropic

This work is supported by University of South Carolina, start up funding. It is also partially supported by NSF Awards DMS-1620345, DMS-1830225, and ONR Award N00014-18-1-2852.

✉ Wuchen Li
wuchen@mailbox.sc.edu

Shui-Nee Chow
chow@math.gatech.edu

Chenchen Mou
chencmou@cityu.edu.hk

Haomin Zhou
hmzhou@math.gatech.edu

¹ School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA

² Department of Mathematics, University of South Carolina, Columbia, SA, USA

³ Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong

relaxation of linear programming problems, whose computation, known as the Sinkhorn's algorithm [13], has been widely used. The dynamical formulation of SBP studies an optimal control problem on density space constrained by the Fokker–Planck equation. In this formulation, the path of the minimizer is a Hamiltonian flow on the density space equipped with L^2 -Wasserstein metric. This interpretation has been found strong connections to the Nelson's variational problem, from which he derives Schrödinger equation. It is related to the Nelson's stochastic mechanics [5–7,28] and stochastic calculus of variations [21,35,36].

In this paper, we study a dynamical SBP on finite graphs. Here the graph represents the discrete states, which arises in numerical computations and modeling [24]. Our approach is mainly based on the recently developed theories on discrete dynamical optimal transport [9,25–27] and discrete Nelson's stochastic mechanics [10]. The SBP on a graph can be posed as a variational problem on the probability simplex constrained by the discrete Fokker–Planck equation. We prove that the minimizer of the SBP on the graph is a unique path, which satisfies a Hamiltonian system in the probability simplex w.r.t. discrete Wasserstein-2 metric. Furthermore, after applying the discrete version of “Nelson's transformation”, i.e. a canonical symplectic transform, we convert the Hamiltonian system into a different expression in term of the discrete Fisher information on the graph.

There exist many different models for the discrete SBP in the literature. Among them, Léonard proposed it as a dynamical variation problem based on random walks [20]. Chen et al. studied the problem based on forward-backward heat equations [8]. They are different from our consideration, which is based on the dynamical optimal transport on graphs. Our formulation naturally connects with Hamiltonian flows on discrete probability simplex, see related works in Wasserstein extreme flows by Conforti and Pavon [11,12]. We remark that our Hamiltonian flow on a graph has potential connections to the discrete Ricci curvature introduced by Erbar and Maas [14], see related discussions in [17]. In addition, the derived Hamiltonian flow can be used as a spatial discretization scheme to compute the minimizer of SBP. We use a few simple examples to illustrate the relation between the proposed discrete SBP problems and their graph structures.

We arrange the paper as follows. In Sect. 2, we briefly review the dynamical SBP on continuous space. In Sect. 3, we propose dynamical SBP on a graph and prove the existence of minimizer path. In Sect. 4, we prove the uniqueness of minimizer by Nelson transform. We end the paper by showing a few numerical examples.

2 Review of Dynamical Schrödinger Bridge Problem

In this sequel, we briefly review optimal transport and Schrödinger bridge problem (SBP); see more details in [19]. The SBP has many different, but equivalent formulations in the continuous sample space. We focus on its dynamical formulation as follows:

$$\inf_b \left\{ \int_0^1 \frac{1}{2} \mathbb{E}_{X_t \sim \rho_t} \|b(t, X_t)\|^2 dt : \dot{X}_t = b(t, X_t) + \sqrt{2\beta} \dot{B}_t, \quad X_0 \sim \rho^0, \quad X_1 \sim \rho^1 \right\}. \quad (1)$$

Here \mathbb{E} is the expectation operator and the infimum is taken over all possible drift function $b: [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}$, such that X_t is a stochastic process in \mathbb{R}^d with a standard Brownian motion B_t , $\beta > 0$ is a given scalar, and X_0, X_1 are random variables with given fixed probability densities $\rho^0(x), \rho^1(x)$.

Here the variational problem (1) can be reformulated in term of probability densities. Denote

$$\int_A \rho(t, x) dx = \Pr(X_t \in A), \quad \text{for any measurable set } A.$$

Then $\rho(t, x)$ satisfies the forward transition equation of X_t , i.e., the Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho b) - \beta \Delta \rho = 0.$$

And the objective functional of (1) can be rewritten as

$$\int_0^1 \mathbb{E}_{X_t \sim \rho_t} \|b(t, X_t)\|^2 dt = \int_0^1 \int_{\mathbb{R}^d} \|b(t, x)\|^2 \rho(t, x) dx dt.$$

One can reform (1) as an action minimization problem in the space of densities:

$$\inf_b \left\{ \int_0^1 \int_{\mathbb{R}^d} \frac{1}{2} \|b(t, x)\|^2 \rho(t, x) dx dt : \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho b) - \beta \Delta \rho = 0, \rho(0, x) = \rho^0(x), \rho(1, x) = \rho^1(x) \right\}. \quad (2)$$

The minimizer of (2) satisfies the following system of equations:

$$\begin{cases} b(t, x) = \nabla \Phi(t, x) \\ \frac{\partial \rho(t, x)}{\partial t} + \nabla \cdot (\rho(t, x) \nabla \Phi(t, x)) = \beta \Delta \rho(t, x) \\ \frac{\partial \Phi(t, x)}{\partial t} + \frac{1}{2} \|\nabla \Phi(t, x)\|^2 = -\beta \Delta \Phi(t, x). \end{cases} \quad (3)$$

Here the first PDE is the Fokker–Planck equation while the second PDE is the Hamilton–Jacobi equation, and $\Phi(t, x)$ is the Lagrange multiplier for variational problem (2). We notice that variational problem (2) and its minimizer (3) are very similar to the Benamou–Breiner variational formula for the 2-Wasserstein metric [3] and its geodesic equations. The differences are the two extra Laplacian terms in SBP.

More interestingly, there are more connections between SBP and optimal transport [7]. Denote

$$v(t, x) := b(t, x) - \beta \nabla \log \rho(t, x). \quad (4)$$

Substituting v into (2) and performing the integration by parts with respect to both time and spatial variables, SBP (2) can be rewritten as

$$\inf_v \left\{ \int_0^1 \left\{ \int_{\mathbb{R}^d} \frac{1}{2} v^2(t, x) \rho(t, x) dx + \mathcal{I}(\rho(t, \cdot)) \right\} dt + \beta \left[\int_{\mathbb{R}^d} \rho^1(x) \log \rho^1(x) - \rho^0(x) \log \rho^0(x) dx \right] \right\}, \quad (5)$$

where the infimum is taken over all Borel vector fields $v(t, x)$, such that

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

and

$$\mathcal{I}(\rho) := \int_{\mathbb{R}^d} \|\nabla \log \rho(x)\|^2 \rho(x) dx,$$

represents the Fisher information functional [15]. Hence the minimizer of (5) satisfies

$$\begin{cases} v(t, x) = \nabla S(t, x) \\ \frac{\partial \rho(t, x)}{\partial t} + \nabla \cdot (\rho(t, x) \nabla S(t, x)) = 0 \\ \frac{\partial S(t, x)}{\partial t} + \frac{1}{2} \|\nabla S(t, x)\|^2 = \frac{\beta^2}{2} \frac{\delta}{\delta \rho(t, x)} \mathcal{I}(\rho). \end{cases} \quad (6)$$

Here $\frac{\delta}{\delta \rho(t, x)}$ is the L^2 first variation operator w.r.t. density $\rho(t, x)$, and the first PDE of (6) is a continuity equation while the second PDE is an Hamilton–Jacobi equation with the L^2 differential of Fisher information functional.

We notice that (4) is the key technique used by Edward Nelson to derive Schrödinger equation [28]. So we call (4) the *Nelson transformation*. We will also perform this transformation discretely, and derive the discrete version of (3), (6) on finite graphs.

3 Schrödinger Bridge Problem on Graphs

In this section, we study a Schrödinger bridge problem on a graph. It is a discrete analog of variational problem (2).

3.1 Dynamical Optimal Transport on Graphs

We review some notations in optimal transport on graphs. Consider a weighted graph $G = (V, E, \omega)$, where $V = \{1, 2, \dots, n\}$ is the vertex set, E is the edge set, and ω is the set of weights on edges. The probability set (simplex) supported on all vertices of G is defined by

$$\mathcal{P}(G) = \left\{ (\rho_i)_{i=1}^n \in \mathbb{R}^n : \sum_{i=1}^n \rho_i = 1, \rho_i \geq 0 \right\},$$

where ρ_i is the discrete probability function at node i . Its interior is denoted by $\mathcal{P}_+(G)$.

For the convenience of notions, we define the following operations on graphs. A *vector field* b on G refers to a *skew-symmetric matrix*, $b: V \times V \rightarrow \mathbb{R}$:

$$b_{ij} = \begin{cases} -b_{ji} & \text{if } (i, j) \in E; \\ 0 & \text{otherwise.} \end{cases}$$

Given a function $\Phi: V \rightarrow \mathbb{R}$, a *potential vector field* $\nabla_G \Phi: V \times V \rightarrow \mathbb{R}$ is defined as

$$\nabla_G \Phi_{ij} = \begin{cases} \sqrt{\omega_{ij}}(\Phi_i - \Phi_j) & \text{if } (i, j) \in E; \\ 0 & \text{otherwise.} \end{cases}$$

Let $m: V \times V \rightarrow \mathbb{R}$ be an anti-symmetric flux function such that $m_{ij} = -m_{ji}$. The divergence of m , denoted as $\text{div}_G(m): S \rightarrow \mathbb{R}$, is defined by

$$\text{div}_G(m)_i = - \sum_{j \in N(i)} \sqrt{\omega_{ij}} m_{ij},$$

where $N(i) = \{j \in V : (i, j) \in E\}$ represents the adjacent set of i . Consider a particular flux function

$$m_{ij} := \theta_{ij}(\rho) b_{ij},$$

where $\theta_{ij}(\rho)$ represents the discrete probability (weight) on edge (i, j) , defined by

$$\theta_{ij}(\rho) = \begin{cases} \frac{\frac{\rho_i}{d_i} - \frac{\rho_j}{d_j}}{\log \frac{\rho_i}{d_i} - \log \frac{\rho_j}{d_j}} & \text{if } \rho_i > 0 \text{ and } \rho_j > 0; \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

with $d_i = \frac{\sum_{j \in N(i)} \omega_{ij}}{\sum_{i=1}^n \sum_{j \in N(i)} \omega_{ij}}$ representing the volume at node i .

We remark that the choice of θ_{ij} is not unique [25]. The other choice of θ_{ij} is the arithmetic mean

$$\theta_{ij}(\rho) = \frac{1}{2} \left(\frac{\rho_i}{d_i} + \frac{\rho_j}{d_j} \right).$$

For the simplicity of proof, we present the result by using θ_{ij} in (7). In fact, the proof can be adjusted to the arithmetic mean. More details are provided in Remark 4.

Given two vector fields $v = (v_{ij})_{(i,j) \in E}$, $\tilde{v} = (\tilde{v}_{ij})_{(i,j) \in E}$ on the graph and $\rho \in \mathcal{P}(G)$. The discrete *inner product* is defined by

$$(v, \tilde{v})_\rho := \frac{1}{2} \sum_{(i,j) \in E} v_{ij} \tilde{v}_{ij} \theta_{ij}(\rho).$$

Here the coefficient $\frac{1}{2}$ is due to the convention that the graph is undirected. Hence the term on each edge, e.g. (i, j) , (j, i) , is counted twice. The L^2 -Wasserstein metric on $\mathcal{P}(G)$ can be defined as follows.

Definition 1 For any $\rho^0, \rho^1 \in \mathcal{P}(G)$, define the Wasserstein distance $W: \mathcal{P}(G) \times \mathcal{P}(G) \rightarrow \mathbb{R}$ by

$$W(\rho^0, \rho^1)^2 := \inf_{\rho(t), b(t)} \left\{ \int_0^1 (v(t), v(t))_{\rho(t)} dt \right\}.$$

Here the infimum is taken over pairs $(\rho(t), v(t))$ with $\rho \in H^1((0, 1), \mathbb{R}^n)$ and $v_{ij} = -v_{ji}: [0, 1] \rightarrow \mathbb{R}$ measurable, satisfying

$$\frac{d}{dt} \rho(t) + \operatorname{div}_G(\rho(t)v(t)) = 0, \quad \rho(0) = \rho^0, \quad \rho(1) = \rho^1.$$

Variational problem in Definition 1 has an equivalent representation, which allow us to equip the probability simplex with a Riemannian structure. We show this by the following matrix function following graph Laplacian notations. We notice that these divergence, gradient operators in matrix forms are consistent with the ones in previous definitions.

Definition 2 (*Weighted Laplacian matrix*) Define the matrix function $L(\cdot): \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ by

$$L(a) = D^T \Theta(a) D, \quad a = (a_i)_{i=1}^n \in \mathbb{R}^n,$$

where

- $D \in \mathbb{R}^{|E| \times n}$ is the discrete gradient operator

$$D_{(i,j) \in E, k \in V} = \begin{cases} \sqrt{\omega_{ij}}, & \text{if } i = k, i > j \\ -\sqrt{\omega_{ij}}, & \text{if } j = k, i > j; \\ 0, & \text{otherwise} \end{cases}$$

- $-D^T \in \mathbb{R}^{n \times |E|}$ is the discrete divergence operator, which is the transpose of discrete gradient operator D ;
- $\Theta(a) \in \mathbb{R}^{|E| \times |E|}$ is a weight matrix depending on a ,

$$\Theta(a)_{(i,j) \in E, (k,l) \in E} = \begin{cases} \theta_{ij}(a) & \text{if } (i, j) = (k, l) \in E \\ 0 & \text{otherwise} \end{cases}.$$

Let $a = \rho \in \mathcal{P}_+(G)$, we next study the property of matrix $L(\rho)$, from which we shall build the Riemannian metric tensor of probability simplex.

Lemma 3 (Discrete Hodge decomposition) *Given $\rho \in \mathcal{P}_+(G)$, the following properties hold:*

- (i) $L(\rho)$ is a semi-positive matrix with zero being its simple eigenvalue. Denote the eigenvalue and corresponding orthonormal eigenvectors of $L(\rho)$ by $0 = \lambda_0(\rho) < \lambda_1(\rho) \leq \dots \leq \lambda_{n-1}(\rho)$, and $U(\rho) = (u_0, u_1(\rho), \dots, u_{n-1}(\rho))$, then $L(\rho)$ has the decomposition

$$L(\rho) = U(\rho) \begin{pmatrix} 0 & & & \\ & \lambda_1(\rho) & & \\ & & \ddots & \\ & & & \lambda_{n-1}(\rho) \end{pmatrix} U(\rho)^T,$$

with $u_0 = \frac{1}{\sqrt{n}}(1, \dots, 1)^T$.

- (ii) For any discrete vector field v and $\rho \in \mathcal{P}_+(G)$, there exists a unique discrete gradient vector field $\nabla_G \Phi \in \mathbb{R}^{|E|}$, such that

$$v_{ij} = \nabla_G \Phi_{ij} + \Psi_{ij}, \quad \text{div}_G(\rho \Psi) = 0.$$

In addition,

$$(v, v)_\rho = (\nabla_G \Phi, \nabla_G \Phi)_\rho + (\Psi, \Psi)_\rho.$$

Proof The proof is a direct extension of the classical graph Hodge decomposition with the probability weight function $\theta_{ij}(\rho)$. Given a discrete vector field v and $\rho \in \mathcal{P}_+(G)$, we shall show that there exists a unique gradient vector field $\nabla_G \Phi$, such that

$$-\text{div}_G(\rho \nabla_G \Phi) = L(\rho) \Phi = -\text{div}_G(\rho v).$$

Consider

$$\Phi^T L(\rho) \Phi = \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij} (\Phi_i - \Phi_j)^2 \theta_{ij}(\rho) = 0.$$

Since $\rho_i > 0$ for any $i \in V$ and the graph is connected, we find that $\Phi_1 = \dots = \Phi_n$ is the only solution of the above equation. Thus 0 must be the simple eigenvalue of $L(\rho)$ with eigenvector $(1, \dots, 1)^T$. Since $\text{div}_G(\rho v) \in \text{Ran}(L(\rho))$ and $\text{Ker}(L(\rho)) = \{u_0\}$. Thus there exists a unique solution of Φ up to a constant shift, i.e. $\nabla_G \Phi$ is unique. And $\Psi = v - \nabla_G \Phi$ satisfies $\text{div}_G(\rho \Psi) = \text{div}_G(\rho v) - \text{div}_G(\rho \nabla_G \Phi) = 0$. Let $v_{ij} = \nabla_G \Phi_{ij} + \Psi_{ij}$, where $\text{div}_G(\rho \Psi) = 0$. Then

$$\begin{aligned} (v, v)_\rho &= (\nabla_G \Phi, \nabla_G \Phi)_\rho + 2(\nabla_G \Phi, \Psi)_\rho + (\Psi, \Psi)_\rho \\ &= (\nabla_G \Phi, \nabla_G \Phi)_\rho + (\Psi, \Psi)_\rho, \end{aligned}$$

which finishes the proof. \square

From Lemma 3, for any discrete vector field v , there exists a unique pair $(\nabla_G \Phi, \Psi)$, such that $\operatorname{div}_G(\rho v) = \operatorname{div}_G(\rho \nabla_G \Phi)$ and

$$(v, v)_\rho = (\nabla_G \Phi, \nabla_G \Phi)_\rho + (\Psi, \Psi)_\rho \geq (\nabla_G \Phi, \nabla_G \Phi)_\rho.$$

Thus the proposed L^2 -Wasserstein metric W on graph is equivalent to

$$W(\rho^0, \rho^1)^2 = \inf_{\Phi(t)} \left\{ \int_0^1 (\nabla_G \Phi(t), \nabla_G \Phi(t))_{\rho(t)} dt : \frac{d\rho}{dt} + \operatorname{div}_G(\rho \nabla_G \Phi) = 0, \rho(0) = \rho^0, \rho(1) = \rho^1 \right\}. \quad (8)$$

Remark 1 We notice that the Wasserstein distance has several equivalent formulations in continuous sample space. One is the linear programming minimization over couplings with a given ground cost. The other is the dynamical formulations, as we stated in this paper, also known as Benamou–Breiner formula [34]. The discretization of these formulations results in different formulations in discrete space, despite their limit to continuous space are equivalent. We select the dynamical formulation because it fits our goal of making dynamical connections to Schrödinger equations and bridge problems on graphs.

3.2 Riemannian Manifold of Probability Simplex

Here, our goal is to demonstrate that (8) introduces a Riemannian metric tensor of the probability simplex in both primal and dual coordinates. The probability simplex $\mathcal{P}(G)$ is a manifold with boundary. To simplify the discussion, we focus on the interior $\mathcal{P}_+(G)$. For carefully geometric description of probability simplex, see [23, 25, 26]. And for more details about the geodesics on the boundary set, see [16].

Denote the tangent space at a point $\rho \in \mathcal{P}_+(G)$ by

$$T_\rho \mathcal{P}_+(G) = \left\{ (\sigma_i)_{i=1}^n \in \mathbb{R}^n : \sum_{i=1}^n \sigma_i = 0 \right\},$$

and the space of potential function on the set of vertices set by $\mathcal{F}(G) = \{(\Phi_i)_{i=1}^n \in \mathbb{R}^n\}$. Consider the quotient space

$$\mathcal{F}(G)/\mathbb{R} = \{[\Phi] \mid (\Phi_i)_{i=1}^n \in \mathbb{R}^n\},$$

where $[\Phi] = \{(\Phi(1) + c, \dots, \Phi(n) + c) \mid c \in \mathbb{R}\}$ are functions defined up to a shift of constants.

We introduce an identification map by the weighted Laplacian operator $L(\rho)$.

$$\mathbf{V}: \mathcal{F}(G)/\mathbb{R} \rightarrow T_\rho \mathcal{P}_+(G), \quad \mathbf{V}_\Phi = L(\rho)\Phi.$$

From Lemma 3, $\mathbf{V}: \mathcal{F}(G)/\mathbb{R} \rightarrow T_\rho \mathcal{P}_+(G)$ is a well defined linear and one to one map, i.e., $\mathcal{F}(G)/\mathbb{R} \cong T_\rho^* \mathcal{P}_+(G)$. Here $T_\rho^* \mathcal{P}_+(G)$ is the cotangent space of $\mathcal{P}_+(G)$.

This identification induces the following inner product on $T_\rho \mathcal{P}_+(G)$.

Definition 4 (*Inner product in dual coordinates*) Given $\rho \in \mathcal{P}_+(G)$, the inner product $g_W: T_\rho \mathcal{P}_+(G) \times T_\rho \mathcal{P}_+(G) \rightarrow \mathbb{R}$ takes any two tangent vectors \mathbf{V}_Φ and $\mathbf{V}_{\tilde{\Phi}} \in T_\rho \mathcal{P}_+(G)$ to

$$g_W(\mathbf{V}_\Phi, \mathbf{V}_{\tilde{\Phi}}) = (\nabla_G \Phi, \nabla_G \tilde{\Phi})_\rho. \quad (9)$$

The above equation is written in the dual coordinates of the Riemannian manifold, i.e. $\Phi \in \mathcal{F}(G)/\mathbb{R}$. Here $(\mathcal{P}_+(G), g_W)$ is a $(n-1)$ dimensional Riemannian manifold. As in [18], we call $(\mathcal{P}_+(G), g_W)$ *probability manifold*.

On this manifold, the heat flow is the gradient flow of negative Boltzmann–Shannon entropy given by,

$$\mathcal{H}(\rho) = \sum_{i=1}^n \rho_i \log \rho_i.$$

In other words,

$$\begin{aligned} \frac{d\rho}{dt} &= -(L(\rho)^\dagger)^\dagger (\nabla_{\rho_i} \mathcal{H}(\rho))_{i=1}^n \\ &= -L(\rho)(\log \rho + 1) \\ &= \operatorname{div}_G(\rho \nabla_G \log \rho), \end{aligned}$$

where the symbol \dagger represents the Moore–Penrose inverse operator. Thus the Fokker–Planck equations on a graph is given by

$$\frac{d\rho}{dt} + \operatorname{div}_G(\rho(b - \beta \nabla_G \log \rho)) = 0,$$

where b is the discrete drift vector and $\beta > 0$ is the noise level.

3.3 Discrete Schrödinger Bridge Problem

We are now ready to present the SBP on a graph G .

Definition 5 Given a graph $G = (V, E, \omega)$ with a scale $\beta > 0$, SBP on a graph is the following action minimization problem:

$$J := \inf_{\rho, b} \int_0^1 \frac{1}{2} (b, b)_\rho dt, \quad (10)$$

where the infimum is taken over $\rho_i(t) \in H^1((0, 1))$ and $b_{ij}(t) \in L^2(0, 1; \theta_{ij}(\rho))$, i.e., $\theta_{ij}(\rho(t))b_{ij}(t) \in L^2((0, 1))$, such that

$$\frac{d\rho}{dt} + \operatorname{div}_G(\rho(b - \beta \nabla_G \log \rho)) = 0, \quad \text{with } \rho(0), \rho(1) \text{ fixed in } \mathcal{P}(G).$$

In the following theorem, we demonstrate that the minimizer of (10) exists, and we characterize the minimizer by a pair of ODEs.

Theorem 6 *There exists a minimizer of problem (10), denoted by $(\rho^*(t), b^*(t))$ such that*

$$\rho_i^*(t) \in H^1((0, 1)), \quad \text{and } b_{ij}^*(t) \in L^2(0, 1; \theta_{ij}(\rho^0)).$$

In addition, $(\rho^(t), b^*(t))$ satisfies a pair of ODEs for a.e. $t \in [0, 1]$:*

$$\begin{cases} b_{ij} = \sqrt{\omega_{ij}}(\Phi_i - \Phi_j) \\ \frac{d\rho_i}{dt} - \sum_{j \in N(i)} \omega_{ij}(\Phi_i - \Phi_j)\theta_{ij}(\rho) = \beta \frac{1}{d_i} \sum_{j \in N(i)} \omega_{ij}(\rho_j - \rho_i) \\ \frac{d\Phi_i}{dt} + \frac{1}{2} \sum_{j \in N(i)} \omega_{ij}(\Phi_i - \Phi_j)^2 \frac{\partial \theta_{ij}}{\partial \rho_i} = -\beta \frac{1}{d_i} \sum_{j \in N(i)} \omega_{ij}(\Phi_j - \Phi_i). \end{cases} \quad (11)$$

Remark 2 We observe that ODE (11) can be viewed as equation (3) on graphs. Here the first ODE is the discrete Fokker–Planck equation, while the second ODE is the discrete Hamilton–Jacobi–Bellman equation.

3.4 Proof of Theorem 6

To study (10), we consider the following related minimization problem

$$\tilde{J} := \inf_{\rho, m} \int_0^1 \frac{1}{2} \sum_{(i,j) \in E} \alpha(\theta_{ij}(\rho), m_{ij}) dt, \quad (12)$$

where

$$\alpha(x, y) := \begin{cases} \frac{|y|^2}{2x} & x > 0 \\ 0 & y = 0, x = 0 \\ +\infty & \text{otherwise} \end{cases}$$

and the infimum is taken over $\rho_i \in H^1((0, 1))$ and $m_{ij} \in L^2((0, 1))$ such that

$$\frac{d\rho_i}{dt} + \sum_{j \in N(i)} \sqrt{w_{ij}} m_{ji} + \beta \frac{1}{d_i} \sum_{j \in N(i)} \omega_{ij} (\rho_i - \rho_j) = 0, \quad \rho(0), \rho(1) \text{ are fixed in } \mathcal{P}(G).$$

Here we point out the technical difficulties in action problem (10). In particular, the boundary of probability set provides the difficulties in characterizing the geodesics [16]. It is no longer an issue in studying minimizers of SBP (10). The proof is outlined as follows:

- Step 1: In Lemma 7, we show that there exists a feasible path. In Lemmas 8 and 9, we show that the minimization problems (10) and (12) are equivalent, from which we prove the existence of minimizer;
- Step 2: In Lemma 10, we prove that the minimizer path $\rho^*(t)$ almost surely lies in the interior of probability simplex;
- Step 3: In Lemmas 11 and 12, we characterize the minimizer path.

Lemma 7 For any $\rho^0, \rho^1 \in \mathcal{P}(G)$, there exists a path $\rho \in H^1((0, 1); \mathbb{R}^n)$ and $m \in L^2((0, 1); \mathbb{R}^{n \times n})$, such that $\rho(0) = \rho^0, \rho(1) = \rho^1$ with

$$\int_0^1 \frac{1}{2} \sum_{(i,j) \in E} \alpha(\theta_{ij}(\rho(t)), m_{ij}) dt < \infty. \quad (13)$$

Proof Let $\tilde{\rho} = (1/n, \dots, 1/n) \in \mathcal{P}(G)$. We now define

$$\rho(t) = \begin{cases} (1-2t)\rho^0 + 2t\tilde{\rho}, & t \in [0, 1/2); \\ (2-2t)\tilde{\rho} + (2t-1)\rho^1, & t \in [0, 1/2). \end{cases}$$

Since $\tilde{\rho} \in \mathcal{P}_+(G)$, then $L(\rho(t)): \mathbb{R}^n / \text{Span}\{(1, \dots, 1)\} \rightarrow T_{\rho} \mathcal{P}_+(G)$ is a bijection, and thus $L(\rho(t))$ and its pseudo inverse operator $L(\rho(t))^{\dagger}$, as in Lemma 3, are bounded linear operators in $(0, 1)$. By the construction of $\rho(t)$, there exists a constant $C > 0$ such that

$$\sup_{t \in [0, 1]} \left| \frac{d\rho}{dt} - \beta L(\rho(t)) \log \rho(t) \right| \leq C.$$

Let

$$\Phi(t) := L(\rho(t))^{\dagger} \left(\frac{d\rho}{dt} + L(\rho(t)) \log \rho(t) \right)$$

and

$$m(t) = (m_{ij}(t))_{ij} := (\theta_{ij}(\rho(t))\sqrt{w_{ij}}(\Phi_j(t) - \Phi_i(t)))_{ij}.$$

Using the boundedness property of $L(\rho(t))^\dagger$, we have $\sup_{t \in [0,1]} |m(t)| \leq C$. Then, by the choice of θ_{ij} , it is obvious that (13) holds. \square

Lemma 8 *The minimization problem (12) can be obtained by (ρ^*, m^*) which satisfies $\rho_i^* \in H^1((0, 1))$, $m_{ij}^* \in L^2((0, 1))$ and*

$$\mathcal{L}^1\left(\{t \in (0, 1); \theta_{ij}(\rho^*(t)) = 0, m_{ij}^*(t) \neq 0, \text{ for some } (i, j) \in E\}\right) = 0, \quad (14)$$

where \mathcal{L}^1 is the Lebesgue measure of \mathbb{R}^1 .

Proof We define

$$\mathcal{A}(\rho, m) := \int_0^1 \frac{1}{2} \sum_{(i,j) \in E} \alpha(\theta_{ij}(\rho(t)), m_{ij}(t)) dt$$

for any $\rho_i \in H^1((0, 1))$ and $m_{ij} \in L^2((0, 1))$.

We notice $\alpha(\theta_{ij}(\rho(t)), m_{ij}(t)) \geq \frac{m_{ij}^2}{2}$ if $\rho(t) \in \mathcal{P}(G)$. Now suppose that $\{(\rho^k, m^k)\}_{k=1}^{+\infty}$ is a minimizing sequence of minimization problem (12), i.e., $\tilde{J} = \lim_{k \rightarrow +\infty} \mathcal{A}(\rho^k, m^k)$. Therefore, there exists $K > 0$ such that $k \geq K$

$$\int_0^1 \sum_{(i,j) \in E} (m_{ij}^k(t))^2 dt < \tilde{J} + 1,$$

i.e., $\sup_{k \geq K} \|m^k\|_{L^2((0,1))} < \tilde{J} + 1$. There exists $m^* \in L^2((0, 1))$ such that m^k converges to m^* weakly in $L^2((0, 1))$. Consider the Fokker–Planck equation as follows:

$$\frac{d\rho_i^k}{dt} + \sum_{j \in N(i)} \sqrt{w_{ij}} m_{ji}^k = \beta \frac{1}{d_i} \sum_{j \in N(i)} \omega_{ij} (\rho_j^k - \rho_i^k). \quad (15)$$

Since $\|\rho^k\|_{L^2((0,1))} < \tilde{J} + 1$ and $\rho^k \in \mathcal{P}(G)$ if $k \geq K$, we have by Sobolev Embedding Theorem

$$\sup_{k \geq K} \|\rho^k(t)\|_{C^{\frac{1}{2}}((0,1))} \leq C \sup_{k \geq K} \|\rho^k(t)\|_{H^1((0,1))} \leq C(\tilde{J}),$$

where $C(\tilde{J})$ depends on \tilde{J} . By Arzelà–Ascoli Theorem, there exists $\rho^* \in C^{\frac{1}{2}}((0, 1))$ such that ρ^k converges to ρ^* in $L^\infty((0, 1))$ and ρ^k converges to ρ^* weakly in $H^1((0, 1))$ up to a subsequence. Now taking limit in (15), we get (ρ^*, m^*) satisfying

$$\frac{d\rho_i^*}{dt} + \sum_{j \in N(i)} \sqrt{w_{ij}} m_{ji}^* = \beta \frac{1}{d_i} \sum_{j \in N(i)} \omega_{ij} (\rho_j^* - \rho_i^*) \quad (16)$$

in the weak sense. Since α is a non-negative convex, lower semicontinuous function, by the standard theory of the calculus of variations we obtain \mathcal{A} is non-negative and lower semicontinuous on $L^2((0, 1)) \times L^2((0, 1))$ for the weak convergence. So it achieve its minimum at (ρ^*, m^*) . \square

Lemma 9 *Minimization problems (10) and (12) are equivalent.*

Proof By Lemma 8, the minimization problem (12) can be obtained by (ρ^*, m^*) satisfying $\rho_i^* \in H^1((0, 1))$, $m_{ij}^* \in L^2((0, 1))$ and (14) holds. We define

$$b_{ij}^*(t) := \begin{cases} 0 & \theta_{ij}(\rho^*(t)) = 0; \\ \frac{m_{ij}^*(t)}{\theta_{ij}(\rho^*(t))} & \theta_{ij}(\rho^*(t)) = 0. \end{cases} \quad (17)$$

Then $b_{ij}^* \in L^2(0, 1; \theta_{ij}(\rho^*))$. Since (16) holds, we have, by the definition of $b_{ij}^*(t)$ and (14),

$$\frac{d\rho_i^*}{dt}(t) + \sum_{j \in N(i)} \sqrt{\omega_{ij}} b_{ji}^*(t) \theta_{ij}(\rho^*(t)) = \beta \frac{1}{d_i} \sum_{j \in N(i)} \omega_{ij} (\rho_j^*(t) - \rho_i^*(t)) \quad (18)$$

and

$$\int_0^1 \frac{1}{4} \sum_{(i,j) \in E} \theta_{ij}(\rho^*(t)) b_{ij}^*(t)^2 dt = \int_0^1 \frac{1}{2} \sum_{(i,j) \in E} \alpha(\theta_{ij}(\rho^*(t)), m_{ij}^*(t)) dt.$$

Therefore

$$J \geq \tilde{J}.$$

For any (ρ, b) satisfying $\rho_i \in H^1((0, 1))$, $b_{ij} \in L^2(0, 1; \theta_{ij}(\rho))$ such that (18) holds, we define $m_{ij}(t) := \theta_{ij}(\rho(t)) b_{ij}(t)$. It is obvious that $m_{ij} \in L^2((0, 1))$ and (16) holds. Then

$$J \leq \tilde{J}.$$

Therefore we have

$$J = \tilde{J}.$$

□

Remark 3 Let b^* be given by (17). By the proof of Lemma 9, the minimization problem (10) can be obtained by (ρ^*, b^*) .

Lemma 10 Let (ρ^*, b^*) be given in Lemma 9. Then

$$\mathcal{L}^1\{t \in [0, 1]; \rho_i^*(t) = 0 \text{ for some } i \in V\} = 0. \quad (19)$$

Proof We define a set $A_i := \{t \in [0, 1]; \rho_i^*(t) = 0, \rho_j^*(t) > 0 \text{ for some } j \in N(i)\}$. Thus, we have $\theta_{ij}(\rho^*)(t) = 0$ for any $t \in A_i$ and $j \in N(i)$. We claim that $\mathcal{L}^1(A_i) = 0$. If not, we have $\mathcal{L}^1(A_i) > 0$. Since $\frac{d\rho_i^*}{dt} \in L^2((0, 1))$, then, for a.e. $t \in [0, 1]$,

$$\frac{d\rho_i^*}{dt}(t) = \lim_{r \rightarrow 0^+} \frac{\rho_i^*(t-r) - \rho_i^*(t)}{-r}.$$

Now we can choose a time $t_0 \in [0, 1]$ such that

$$\frac{d\rho_i^*}{dt}(t_0) = \lim_{r \rightarrow 0^+} \frac{\rho_i^*(t_0-r) - \rho_i^*(t_0)}{-r} \quad (20)$$

and $t_0 \in \{t \in [0, 1]; \rho_i^*(t) = 0, \rho_j^*(t) > 0, \text{ for some } j \in N(i)\}$. By (18), we have

$$\frac{d\rho_i^*}{dt}(t_0) > 0.$$

By (20), there exists $r_0 > 0$ such that $\rho_i^*(t_0 - r_0) < 0$. It contradicts with $\rho^*(t_0 - r_0) \in \mathcal{P}(G)$. Therefore,

$$\mathcal{L}^1(\{t \in [0, 1]; \rho_i^*(t) = 0, \rho_j^*(t) = 0, \text{ for all } j \in N(i)\}) = 1.$$

Since i is arbitrary and G is connected, (19) holds. \square

Lemma 11 *The minimization problem (10) can be obtained by $(\rho^*, \nabla_G \Phi^*)$ satisfying $\rho_i^* \in H^1((0, 1))$ and $\Phi_j^* - \Phi_i^* \in L^2(0, 1; \theta_{ij}(\rho^*))$.*

Proof By the proof of Lemma 9, (10) can be obtained by (ρ^*, b^*) where b^* is given by (17). Using Lemma 10, we know that $\rho^* \in \mathcal{P}_+(G)$ for a.e. $t \in [0, 1]$. For each $t \in [0, 1]$ such that $\rho^*(t) \in \mathcal{P}_+(G)$, we have, by Lemma 3, $\left(\sum_{j \in N(i)} \sqrt{\omega_{ij}} b_{ij}^*(t) \theta_{ij}(\rho^*(t))\right)_{i=1}^n \in \text{Ran}(L(\rho^*(t)))$, i.e., there exists $\Phi^*(t)$ such that

$$\left(\sum_{j \in N(i)} \sqrt{\omega_{ij}} b_{ij}^*(t) \theta_{ij}(\rho^*(t))\right)_{i=1}^n = \left(\sum_{j \in N(i)} \omega_{ij} (\Phi_j^*(t) - \Phi_i^*(t)) \theta_{ij}(\rho^*(t))\right)_{i=1}^n.$$

We let $u^*(t) := b^*(t) - \nabla_G \Phi^*(t)$. Then

$$\int_0^1 \frac{1}{2} (b^*, b^*)_{\rho^*} dt = \int_0^1 \frac{1}{2} (\nabla_G \Phi^*, \nabla_G \Phi^*)_{\rho^*} dt + \int_0^1 \frac{1}{2} (u^*, u^*)_{\rho^*} dt.$$

Therefore, $\Phi_j^* - \Phi_i^* \in L^2(0, 1; \theta_{ij}(\rho^*))$ and

$$\int_0^1 \frac{1}{2} (\nabla_G \Phi^*, \nabla_G \Phi^*)_{\rho^*} dt \leq \int_0^1 \frac{1}{2} (b^*, b^*)_{\rho^*} dt = J.$$

\square

Lemma 12 *The minimizer $(\rho^*, \nabla_G \Phi^*)$ of minimization problem (10) solves (11) weakly.*

Proof Since we prove that $\rho^*(t)$ is a.e. in $[0, 1]$, then we can apply the standard perturbation argument. We then obtain the minimizer $(\rho^*, \nabla_G \Phi^*)$ satisfying the ODEs in (11). See details in the proof of Theorem 3 at [10] \square

Finally, we are ready to present the proof of Theorem 6.

Proof of Theorem 6 For any $t_0 \in [0, 1]$ such that $\rho^*(t_0) \in \mathcal{P}_+(G)$, there exists $\delta_0 > 0$ such that $\rho^*(t) \in \mathcal{P}_+(G)$. Since $\int_0^1 (\nabla_G \Phi^*, \nabla_G \Phi^*)_{\rho^*} dt < +\infty$ and $\inf_{t \in (t_0 - \delta_0, t_0 + \delta_0)} \theta_{ij}(\rho^*(t)) > 0$ for all $(i, j) \in E$, then $\Phi_i^* - \Phi_j^* \in L^2((t_0 - \delta_0, t_0 + \delta_0))$ for any $(i, j) \in E$. Using Lemma 12, we have $\Phi^* \in W^{1,1}(t_0 - \delta_0, t_0 + \delta_0)$. Using a bootstrap argument, we have (ρ^*, Φ^*) is smooth and solves (10) classically in $(t_0 - \delta_0, t_0 + \delta_0)$. Thus Theorem 6 follows from Lemma 10. \square

4 Nelson's Transformation

In this sequel, we prove that there exists a unique minimizer for discrete SBP. Our main tool is based on Nelson's transformation as follows. Define a new vector field v on a graph

$$v_{ij} := b_{ij} - \beta \nabla_G (\log \rho)_{ij}.$$

Substituting v into (10), we obtain a new action minimization problem.

Definition 13 Given a graph $G = (V, E, \omega)$ with a scale $\beta > 0$, consider the following action minimization problem:

$$J_1 := \inf_{\rho, v} \int_0^1 \left\{ \frac{1}{2} (v, v)_{\rho(t)} + \frac{\beta^2}{2} \mathcal{I}(\rho(t)) \right\} dt, \quad (21)$$

where the infimum is taken over $\rho_i(t) \in H^1((0, 1))$ and $v_{ij}(t) \in L^2(0, 1; \theta_{ij}(\rho))$, such that

$$\frac{d\rho}{dt} + \operatorname{div}_G(\rho v) = 0, \quad \text{and} \quad \rho(0), \rho(1) \text{ are fixed in } \mathcal{P}(G).$$

Here $\mathcal{I}: \mathcal{P}(G) \rightarrow \mathbb{R}$ is the discrete Fisher information functional defined by

$$\mathcal{I}(\rho) := (\nabla_G \log \rho, \nabla_G \log \rho)_\rho = \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij} (\log \rho_i - \log \rho_j)^2 \theta_{ij}(\rho).$$

We use the convention that $\mathcal{I}(\rho) = +\infty$ if $\rho \in \mathcal{P}(G) \setminus \mathcal{P}_+(G)$.

Derivation of (21): first, the Fokker–Planck equation on a graph in (10) can be rewritten in term of v (continuity equation): $\frac{d\rho}{dt} + \operatorname{div}_G(\rho v) = 0$. Second, the Lagrangian in (10) forms

$$\begin{aligned} \frac{1}{2} (b, b)_\rho &= \frac{1}{2} (v + \beta \nabla_G \log \rho, v + \beta \nabla_G \log \rho)_\rho \\ &= \frac{1}{2} (v, v)_\rho + \frac{\beta^2}{2} (\nabla_G \log \rho, \nabla_G \log \rho)_\rho + \beta (\nabla_G \log \rho, v)_\rho. \end{aligned}$$

Notice

$$\int_0^1 (\nabla_G \log \rho, v)_\rho dt = \mathcal{H}(\rho(1)) - \mathcal{H}(\rho(0)) = \text{Constant}, \quad (22)$$

where $\mathcal{H}(\rho) = \sum_{i=1}^n \rho_i \log \rho_i$ is the discrete negative Boltzmann–Shannon entropy. Here (22) holds since

$$\begin{aligned} \int_0^1 (\nabla_G \log \rho, v)_\rho dt &= \int_0^1 \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij} (\log \rho_i - \log \rho_j) v_{ij} \theta_{ij}(\rho) dt \\ &= - \int_0^1 \sum_{i=1}^n \log \rho_i \left[\sum_{j \in N(i)} \omega_{ij} v_{ij} \theta_{ij}(\rho) \right] dt \\ &= \int_0^1 \sum_{i=1}^n \log \rho_i \frac{d\rho_i}{dt} dt \\ &= \sum_{i=1}^n \rho_i(t) \log \rho_i(t) \Big|_{t=0}^{t=1} - \int_0^1 \sum_{i=1}^n \rho_i \frac{d}{dt} \log \rho_i dt \\ &= \mathcal{H}(\rho(1)) - \mathcal{H}(\rho(0)) - \int_0^1 \sum_{i=1}^n \rho_i \frac{1}{\rho_i} \frac{d\rho_i}{dt} dt \\ &= \mathcal{H}(\rho(1)) - \mathcal{H}(\rho(0)), \end{aligned} \quad (23)$$

where the second equality is by discrete continuity equation, the third equality is based integration by parts w.r.t time and the last equality is from $\frac{d}{dt} \sum_{i=1}^n \rho_i = 0$. Combining

above three steps, we obtain (21). Following (21), several properties of SBP on a graph can be shown in next theorem.

Theorem 14 *The minimizer path $\rho^*(t)$ of SBP is unique.*

Proof Before presenting the proof, we follow the idea in (12) to reform (21). Define the discrete flux function as $m = (m_{ij})_{(i,j) \in E} := (v_{ij}\theta_{ij}(\rho))_{(i,j) \in E}$. Consider

$$\bar{J}(\rho(t), m(t)) := \inf_{\rho, m} \int_0^1 \frac{1}{2} \sum_{(i,j) \in E} \alpha(\theta_{ij}(\rho), m_{ij}) + \frac{\beta^2}{2} \mathcal{I}(\rho(t)) dt, \quad (24)$$

where the infimum is taken over $\rho_i \in H^1((0, 1))$ and $m_{ij} \in L^2((0, 1))$ such that

$$\frac{d\rho}{dt} + \operatorname{div}_G(m) = 0, \quad \rho(0), \rho(1) \text{ are fixed in } \mathcal{P}(G).$$

Claim 1 *Minimization problems (10), (12), (21), (24) are equivalent.*

Proof of Claim 1 Since the equivalence between (21) and (24) is similar to the one for (10) and (12) in Lemma. We only need to show minimization problems (10), (21) are equivalent. Let $(\rho^*, \nabla_G \Phi^*)$ be the minimizer of minimization problem (10). By Lemma 10, we know that the following discrete Nelson's transformation

$$v^* := \nabla_G \Phi^* - \beta \nabla_G \log \rho^*$$

is well defined a.e. in $[0, 1]$. Similarly, $S^* := \Phi^* - \log \rho^*$ is also well defined a.e. in $[0, 1]$. Thus, we have $\nabla_G S^* = v^*$. Since $\rho_i^* \in H^1((0, 1))$ and the continuity equation holds for the pair $(\rho^*, \nabla_G \Phi)$, we have (23):

$$\int_0^1 (\nabla_G \log \rho^*, v^*)_{\rho^*} dt = \mathcal{H}(\rho(1)) - \mathcal{H}(\rho(0)),$$

where $\mathcal{H}(\rho) = \sum_{i=1}^n \rho_i \log \rho_i$ is the discrete linear entropy. It is obvious that $\mathcal{H}(\rho(0))$ and $\mathcal{H}(\rho(1))$ are fixed finite constants. Then $v_{ij}^*(t) \in L^2(0, 1; \theta_{ij}(\rho^*))$ and

$$\frac{d\rho^*}{dt} + \operatorname{div}_G(\rho^* v^*) = 0.$$

Therefore, we have $J \geq J_1 + \mathcal{H}(\rho(1)) - \mathcal{H}(\rho(0))$.

Let (ρ, v) satisfy $\rho_i(t) \in H^1((0, 1))$, $v_{ij}(t) \in L^2(0, 1; \theta_{ij}(\rho))$,

$$\frac{d\rho}{dt} + \operatorname{div}_G(\rho v) = 0$$

and

$$\int_0^1 \frac{1}{2} (v, v)_{\rho} + \frac{\beta^2}{2} \mathcal{I}(\rho(t)) dt < +\infty. \quad (25)$$

We claim that

$$\mathcal{L}^1 \left\{ t \in [0, 1]; \rho_i(t) = 0 \text{ for some } i \in V \right\} = 0. \quad (26)$$

Otherwise there exists $i_0 \in V$ and $\epsilon_0 > 0$ such that

$$\mathcal{L}^1 \left\{ t \in [0, 1]; \rho_{i_0}(t) = 0 \right\} = \epsilon_0.$$

We define a set $A_{i_0} := \{t \in [0, 1]; \rho_{i_0}(t) = 0, \rho_j(t) > 0 \text{ for some } j \in N(i_0)\}$. Thus, we have $\mathcal{I}(\rho)(t) = +\infty$ for any $t \in A_{i_0}$. Since (25) holds, we have $\mathcal{L}^1(A_{i_0}) = 0$. Therefore,

$$\mathcal{L}^1\left\{t \in [0, 1]; \rho_{i_0}(t) = 0, \rho_j(t) = 0 \text{ for some } j \in N(i_0)\right\} = \epsilon_0.$$

Since the graph G is connected, we have

$$\mathcal{L}^1\left\{t \in [0, 1]; \rho_i(t) = 0, \text{ for all } i \in V\right\} = \epsilon_0.$$

It contradicts with $\rho \in \mathcal{P}(G)$. Therefore, (26) holds. Then

$$b := v + \beta \nabla_G \log \rho$$

is well defined a.e. in $[0, 1]$. By a similar calculation to (23), we have

$$\int_0^1 \frac{1}{2} (b, b)_\rho dt = \int_0^1 \frac{1}{2} (v, v)_\rho + \frac{\beta^2}{2} (\nabla_G \log \rho, \nabla_G \log \rho)_\rho dt + \mathcal{H}(\rho(1)) - \mathcal{H}(\rho(0)). \quad (27)$$

Using (25), we have $b_{ij} \in L^2(0, 1; \theta_{ij}(\rho))$ and

$$\frac{d\rho}{dt} + \operatorname{div}_G(\rho(b - \beta \nabla_G \log \rho)) = 0.$$

Therefore,

$$\int_0^1 \frac{1}{2} (b, b)_\rho dt \geq \int_0^1 \frac{1}{2} (\nabla_G \Phi^*, \nabla_G \Phi^*)_{\rho^*} dt.$$

By (23) and (27), we have

$$\begin{aligned} & \int_0^1 \frac{1}{2} (v, v)_\rho + \frac{\beta^2}{2} (\nabla_G \log \rho, \nabla_G \log \rho)_\rho dt \\ & \geq \int_0^1 \frac{1}{2} (v^*, v^*)_{\rho^*} + \frac{\beta^2}{2} (\nabla_G \log \rho^*, \nabla_G \log \rho^*)_{\rho^*} dt. \end{aligned}$$

Then $J \leq J_1 + \mathcal{H}(\rho(1)) - \mathcal{H}(\rho(0))$. Therefore, $J = J_1 + \mathcal{H}(\rho(1)) - \mathcal{H}(\rho(0))$. Moreover, $(\rho^*, \nabla_G \Phi^*)$ and $(\rho^*, \nabla_G S^*)$ are the minimizers of variational problems (10) and (21), respectively. \square

Our proof is based on formulation (24). If there are two minimizer paths $(\rho^1(t), m^1(t))$, $(\rho^2(t), m^2(t))$, we shall prove

$$\rho^1(t) \equiv \rho^2(t) \quad \text{for a.e. } t \in [0, 1]. \quad (28)$$

Here the uniqueness can be shown by the strictly convexity of discrete Fisher information functional:

Claim 2 $\mathcal{I}(\rho)$ is a strictly convex functional in $\mathcal{P}_+(G)$.

Assume the claim is true and suppose (28) is not true, $\rho^1(t) \neq \rho^2(t)$ a.e. $t \in [0, 1]$. Then for a fixed $\lambda \in [0, 1]$,

$$\begin{aligned} & \bar{J}(\lambda \rho^1 + (1 - \lambda) \rho^2, \lambda m^1 + (1 - \lambda) m^2) \\ & = \int_0^1 \frac{1}{2} \sum_{(i,j) \in E} \alpha(\theta_{ij}(\lambda \rho^1 + (1 - \lambda) \rho^2), \lambda m_{ij}^1 + (1 - \lambda) m_{ij}^2) dt \end{aligned}$$

$$\begin{aligned}
 & + \int_0^1 \frac{\beta^2}{2} \mathcal{I}(\lambda \rho^1(t) + (1-\lambda) \rho^2(t)) dt \\
 & < \lambda \int_0^1 \frac{1}{2} \sum_{(i,j) \in E} \alpha(\theta_{ij}(\rho^1), m_{ij}^1) dt + (1-\lambda) \int_0^1 \frac{1}{2} \sum_{(i,j) \in E} \alpha(\theta_{ij}(\rho^1), m_{ij}^1) dt \\
 & + \lambda \int_0^1 \frac{\beta^2}{2} \mathcal{I}(\rho^1(t)) dt + (1-\lambda) \int_0^1 \frac{\beta^2}{2} \mathcal{I}(\rho^2(t)) dt \\
 & = \lambda \bar{J}(\rho^1, m^1) + (1-\lambda) \bar{J}(\rho^2, m^2),
 \end{aligned}$$

where the inequality is from both α and \mathcal{I} are convex function. While the strictly inequality is from Claim 2 and Lemma 10, in which the minimizer path $\rho^1(t)$, $\rho^2(t)$ are positive a.e. Clearly, $(\lambda \rho^1 + (1-\lambda) \rho^2, \lambda m^1 + (1-\lambda) m^2)$ is with smaller cost functional than the one in $(\rho^1(t), m^1(t))$, which is a contradiction. In the end, we prove Claim 2.

Proof of Claim 2

$$\bar{\lambda}(\rho) := \min_{\sigma \in T_{\rho} \mathcal{P}_+(G)} \left\{ \sigma^T \text{Hess}_{\mathbb{R}^n} \mathcal{I}(\rho) \sigma : \sigma^T \sigma = 1, \sum_{i=1}^n \sigma_i = 0 \right\} > 0. \quad (29)$$

Since

$$\frac{\partial^2}{\partial \rho_i \partial \rho_j} \mathcal{I}(\rho) = \begin{cases} -\frac{1}{\rho_i \rho_j} \omega_{ij} t_{ij} & \text{if } j \in N(i); \\ \frac{1}{\rho_i^2} \sum_{k \in N(i)} \frac{1}{d_{ik}^2} t_{ik} & \text{if } i = j; \\ 0 & \text{otherwise,} \end{cases}$$

where

$$t_{ij} = \rho_i + \rho_j > 0.$$

Then

$$\begin{aligned}
 \sigma^T \text{Hess}_{\mathbb{R}^n} \mathcal{I}(\rho) \sigma &= \frac{1}{2} \sum_{(i,j) \in E} t_{ij} \left\{ \left(\frac{\sigma_i}{\rho_i} \right)^2 + \left(\frac{\sigma_j}{\rho_j} \right)^2 - 2 \frac{\sigma_i}{\rho_i} \frac{\sigma_j}{\rho_j} \right\} \\
 &= \frac{1}{2} \sum_{(i,j) \in E} t_{ij} \left(\frac{\sigma_i}{\rho_i} - \frac{\sigma_j}{\rho_j} \right)^2 \geq 0.
 \end{aligned}$$

Suppose (29) is not true, there exists a unit vector σ^* , such that

$$\bar{\lambda}(\rho) = \sigma^{*T} \text{Hess}_{\mathbb{R}^n} \mathcal{I}(\rho) \sigma^* = \frac{1}{2} \sum_{(i,j) \in E} t_{ij} \left(\frac{\sigma_i^*}{\rho_i} - \frac{\sigma_j^*}{\rho_j} \right)^2 = 0.$$

Then $\frac{\sigma_1^*}{\rho_1} = \frac{\sigma_2^*}{\rho_2} = \dots = \frac{\sigma_n^*}{\rho_n} = 0$. Combining with $\sum_{i=1}^n \sigma_i^* = 0$, we have the fact $\sigma_1^* = \sigma_2^* = \dots = \sigma_n^* = 0$, which contradicts that σ^* is a unit vector. \square

Combining Claims 1 and 2, we finish the proof. \square

Problem (21) characterizes the other formulation of minimizer.

Corollary 15 *The minimizer of (21), $(\rho^*(t), v^*(t))$, satisfies the following ODE classically, for a.e. $t \in [0, 1]$*

$$\begin{cases} v_{ij}(t) = \sqrt{\omega_{ij}}(S_i(t) - S_j(t)); \\ \frac{d\rho_i}{dt} - \sum_{j \in N(i)} \omega_{ij}(S_i - S_j)\theta_{ij}(\rho) = 0; \\ \frac{dS_i}{dt} + \frac{1}{2} \sum_{j \in N(i)} \omega_{ij}(S_i - S_j)^2 \frac{\partial \theta_{ij}}{\partial \rho_i} = \frac{\beta^2}{2} \frac{\partial}{\partial \rho_i} \mathcal{I}(\rho). \end{cases} \quad (30)$$

Proof The derivation of (30) is similarly to the one in (11). We omit the proof here. \square

Here the ODEs (11) and (30) represent the same minimizer path under a change of variable

$$S_i = \Phi_i - \beta \log \rho_i. \quad (31)$$

We observe that ODEs (11), (30) can be both written into the following symplectic forms:

$$\frac{d}{dt} \begin{pmatrix} \rho \\ \Phi \end{pmatrix} = \begin{pmatrix} 0 & \mathbb{I} \\ -\mathbb{I} & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \rho} \\ \frac{\partial}{\partial \Phi} \end{pmatrix} \mathcal{H}(\rho, \Phi), \quad \frac{d}{dt} \begin{pmatrix} \rho \\ S \end{pmatrix} = \begin{pmatrix} 0 & \mathbb{I} \\ -\mathbb{I} & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \rho} \\ \frac{\partial}{\partial S} \end{pmatrix} \tilde{\mathcal{H}}(\rho, S),$$

where $\mathbb{I} \in \mathbb{R}^{n \times n}$ is an identity matrix, and $\mathcal{H}, \tilde{\mathcal{H}}$ are Hamiltonians,

$$\mathcal{H}(\rho, \Phi) := \frac{1}{2} (\nabla_G \Phi, \nabla_G \Phi)_\rho - \beta (\nabla_G \log \rho, \nabla_G \Phi)_\rho,$$

and

$$\tilde{\mathcal{H}}(\rho, S) := \frac{1}{2} (\nabla_G S, \nabla_G S)_\rho - \frac{1}{2} \beta^2 (\nabla_G \log \rho, \nabla_G \log \rho)_\rho.$$

It is clear that the change of variable (31) is a canonical transformation.

Remark 4 For $\theta_{ij}(\rho) = \frac{\rho_i/d_i + \rho_j/d_j}{2}$, the proof for characterization of the minimizer is different. We shall prove the existence of minimizer path from variational problem (21). In this approach, we mainly use the fact that the Fisher information is infinity on the boundary of probability simplex. Based on it, we show that the minimizer path is in the interior of probability simplex almost surely for $t \in [0, 1]$. From the Nelson transformation, the minimizer path of (10) and (21) are equivalent. Hence we characterize the minimizer path in (10).

Remark 5 It is worth mentioning that our dynamical Schrödinger bridge problem on graphs naturally connects with the Hamiltonian flows in density space. This is because our problems come from the geometric action in probability simplex with a Fisher information regularization. Hence our formulation is automatically connected with entropic Ricci curvature defined by [14]. In other words, one can prove energy splitting type functional inequalities, using the lower bound of Hessian operator of entropy in discrete Wasserstein space. See related discussions in Theorem 2 of [22].

5 Numerical Examples

In this section, we demonstrate SBP on graphs (10) by several examples. We mainly use the build-in function `bvp4c` in MATLAB to solve the problem (30) numerically. In our computations, we assume $\omega_{ij} = 1$ for all edges and further let $\theta_{ij} = \frac{\rho_i + \rho_j}{2}$.

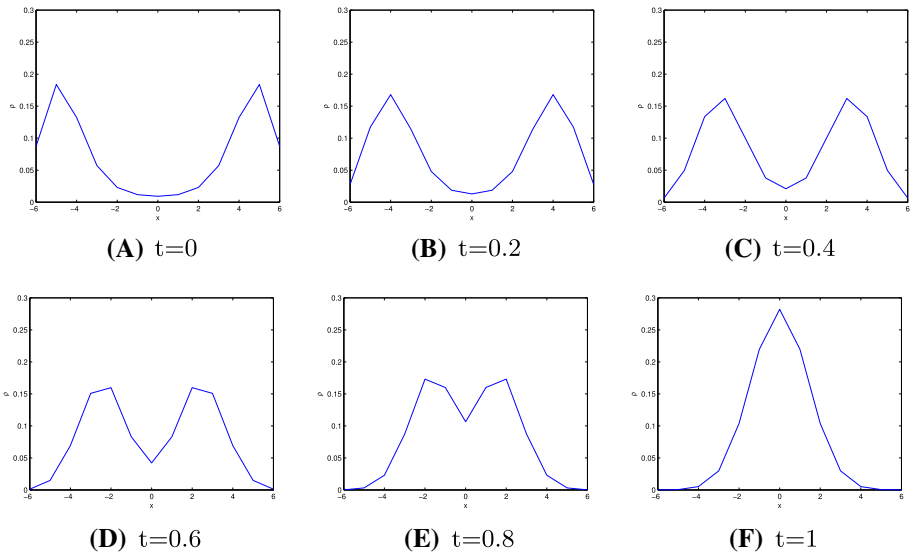


Fig. 1 Snapshot of probability function $\rho(t)$ connecting ρ^0, ρ^1

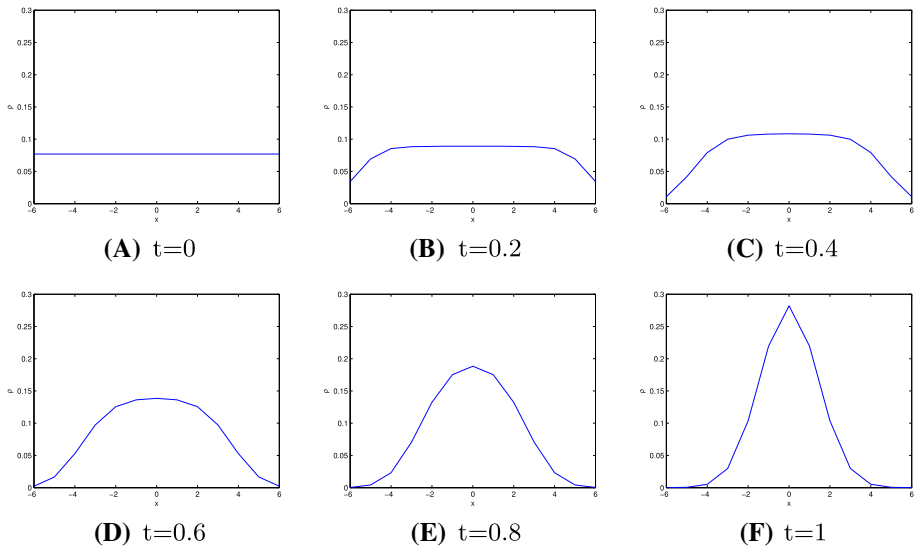


Fig. 2 Snapshot of probability function $\rho(t)$ connecting ρ^0, ρ^1

Fig. 3 The optimal value of (10) is 7.9522

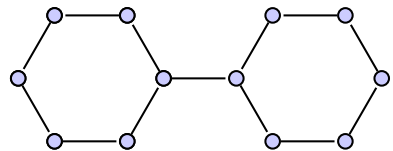


Fig. 4 The optimal value of (10) is 30.4671

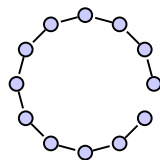
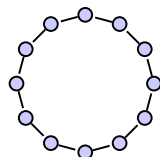
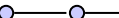
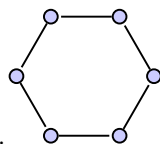


Fig. 5 The optimal value of (10) is 29.2792



Example 1 (Lattice graph) Our first example is the SBP (10) on a lattice graph $G = L_n$, i.e.  Let $n = 13$, $\frac{\beta^2}{2} = 10^{-4}$, $\rho_i^0 = K_0 e^{-\frac{1}{2}(x(i)^2/2 - x(i)^4/96)}$ and

$\rho_i^1 = K_1 e^{-x(i)^2/4}$, where $x(i) = -6 + (i - 1) * \Delta x$, $\Delta x = 1$, K_0, K_1 are normalization constants such that $\sum_{i=1}^n \rho_i^0 = \sum_{i=1}^n \rho_i^1 = 1$. The optimal value of (10) is 52.9057 and the snapshots of the optimal path is demonstrated in Fig. 1.



Example 2 (Cycle graph) Here we consider (10) on a cycle graph $G = C_n$, i.e.

Let $n = 13$, $\frac{\beta^2}{2} = 10^{-4}$, $\rho_i^0 = \frac{1}{n}$ and $\rho_i^1 = K_1 e^{-x(i)^2/4}$, where $x(i) = -6 + (i - 1) * \Delta x$, $\Delta x = 1$, K_1 is a normalization constants such that $\sum_{i=1}^n \rho_i^0 = \sum_{i=1}^n \rho_i^1 = 1$. The optimal value of (10) is 40.1917 and the snapshots of the minimizer is demonstrated in Fig. 2.

Example 3 (Effect of graph structures) In the last example, we illustrate how the graph structure affects the optimal value of SBP (10). Consider a configuration similar as example 2 with $n = 12$. We introduce three graphs in Figs. 3, 4 and 5. Our computation indicates that the minimal values for the corresponding SBPs are significantly different from each other, which depends on the structures of graphs.

Acknowledgements We would like to thank Professor Wilfrid Gangbo for many discussions on the related topics.

References

1. Bernton, E., Heng, J., Doucet, A., Jacob, P.E.: Schrödinger bridge samplers [arXiv:1912.13170](https://arxiv.org/abs/1912.13170) (2019)
2. Backhoff, J., Conforti, G., Gentil, I., Leonard, C.: The mean field Schrodinger problem: ergodic behavior, entropy estimates and functional inequalities. Probab. Theory Relat. Fields 1–56 (2020)
3. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numer. Math. **84**(3), 375–393 (2000)
4. Benamou, J., Carlier, G., Di Marino, S., Nenna, L.: An entropy minimization approach to second-order variational mean-field games. Math. Models Methods Appl. Sci. **29**(08), 1553–1583 (2019)
5. Carlen, E.: Conservative diffusions. Commun. Math. Phys. **94**, 293–315 (1984)
6. Carlen, E.: Stochastic mechanics: a look back and a look ahead. Diffusion Quantum Theory Radically Elem. Math. **47**, 117–139 (2014)

7. Chen, Y., Georgiou, T., Pavon, M.: On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. *J. Optim. Theory Appl.* **169**(2), 671–691 (2016)
8. Chen, Y., Georgiou, T., Pavon, M., Tannenbaum, A.: A new approach to robust transportation over networks. In: *IEEE 55th Conference on Decision and Control (CDC)* (2016)
9. Chow, S.-N., Huang, W., Li, Y., Zhou, H.: Fokker–Planck equations for a free energy functional or Markov process on a graph. *Arch. Ration. Mech. Anal.* **203**(3), 969–1008 (2012)
10. Chow, S.-N., Li, W., Zhou, H.: A discrete Schrödinger equation via optimal transport on graphs (2017)
11. Conforti, G.: A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *Probab. Theory Relat. Fields* (2019)
12. Conforti, G., Pavon, M.: Extremal curves in Wasserstein space. In: *International Conference on Geometric Science of Information*, pp. 91–99 (2017)
13. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. *Adv. Neural Inf. Process. Syst.* (2013)
14. Erbar, M., Maas, J.: Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.* **206**(3), 997–1038 (2012)
15. Frieden, B.R.: *Science from Fisher Information: A Unification*. Cambridge University Press, Cambridge (2004)
16. Gangbo, W., Li, W., Mou, C.: Geodesic of minimal length in the set of probability measures on graphs. In: *Control, Optimisation and Calculus of Variations, ESAIM* (2019)
17. Gentil, I., Leonard, C., Ripani, L.: Dynamical aspects of generalized Schrödinger problem via Otto calculus—a heuristic point of view [arXiv:1806.01553](https://arxiv.org/abs/1806.01553) (2018)
18. Lafferty, D.: The density manifold and configuration space quantization. *Trans. Am. Math. Soc.* (1988)
19. Léonard, C.: A survey of the Schrödinger problem and some of its connections with optimal transport. *Discret. Contin. Dyn. Syst. A* (2014)
20. Léonard, C.: Lazy random walks and optimal transport on graphs. *Ann. Probab.* **44**(3), 1864–1915 (2016)
21. Leger, F.: A geometric perspective on regularized optimal transport. *J. Dyn. Differ. Equ.* (2018)
22. Leger, F., Li, W.: Hopf–Cole transform on generalized Schrödinger bridge problem. Preprint at [arXiv:1901.09051](https://arxiv.org/abs/1901.09051)
23. Li, W.: Transport information geometry I: Riemannian calculus on probability simplex. [arXiv:1803.06360](https://arxiv.org/abs/1803.06360) (2018)
24. Li, W., Yin, P., Osher, S.: Computations of optimal transport distance with Fisher information regularization. *J. Sci. Comput.* 1581–1595 (2018)
25. Maas, J.: Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.* **261**(8), 2250–2292 (2011)
26. Mielke, A.: A gradient structure for reaction–diffusion systems and for energy–drift–diffusion. *Nonlinearity* **24**(4), 13–29 (2011)
27. Mielke, A.: Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial. Differ. Equ.* **48**(1–2), 1–31 (2013)
28. Nelson, E.: *Quantum Fluctuations*. Princeton University Press, Princeton (1985)
29. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. *Commun. Partial Differ. Equ.* **26**(1–2) (2001)
30. Pavon, M.: Quantum Schrödinger bridges. *Dir. Math. Syst. Theory Optim.* 227–238 (2003)
31. Pavon, M., Tabak, E., Trigila, G.: The data-drive Schroedinger bridge. [arXiv:1806.01364](https://arxiv.org/abs/1806.01364) (2018)
32. Peyre, G., Cuturi, M.: *Computational Optimal Transport* (2018)
33. Schrödinger, E.: Über die Umkehrung der Naturgesetze, Sitzungsberichte der Preuss Akad. Wissen. Berlin Phys. Math. Klasse 144–153 (1931)
34. Villani, C.: *Optimal Transport: Old and New*, vol. 338. Springer, Berlin (2008)
35. Yasue, K.: Stochastic calculus of variations. *J. Funct. Anal.* **41**(3), 327–340 (1981)
36. Zambrini, J.-C., Yasue, K.: Semi-classical quantum mechanics and stochastic calculus of variations. *Ann. Phys.* (143)1 (1982)