

ENTROPY DISSIPATION OF FOKKER-PLANCK EQUATIONS ON GRAPHS

SHUI-NEE CHOW, WUCHEN LI, AND HAOMIN ZHOU

ABSTRACT. We study the nonlinear Fokker-Planck equation on graphs, which is the gradient flow in the space of probability measures supported on the nodes with respect to the discrete Wasserstein metric. The energy functional driving the gradient flow consists of a Boltzmann entropy, a linear potential and a quadratic interaction energy. We show that the solution converges to the Gibbs measures exponentially fast. The continuous analog of this asymptotic rate is related to the Yano's formula.

1. INTRODUCTION

Optimal transport theory reveals many deep connections between partial differential equations and geometry. For example, in the seminal work [17], it is proved that the linear Fokker-Planck equation (FPE) is the gradient flow of a free energy in the probability space equipped with Wasserstein metric [14, 24, 30, 31]. This gradient flow interpretation has been extended to mean field settings, in which the free energy contains an interaction energy [1]. Many studies have been carried out showing that the solution of FPE converges to its equilibrium in an exponential rate, and this is known as the entropy dissipation in the literature [4, 9, 21].

The goal of this paper is studying the entropy dissipation of FPE in discrete settings, for example on finite graphs. Such a consideration is motivated by recent applications in evolutionary game theory [8] and numerical schemes for partial differential equations (PDEs) [6]. The optimal transport metric on graphs has been established by several groups independently [7, 19, 22]. The gradient flow structure based on such a metric has attracted a lot of attention in recent years. For example, Mass and Erbar studied the discrete heat flow, and further gave the Ricci curvature lower bound in the irreducible Markov kernel on a finite set [12]. More generalizations are followed in [13, 15, 20]. Mielke proposed the discrete reaction diffusion equation [23]. Erbar, Fathi, Laschos and Schlichting introduced a discrete McKean-Vlasov equation [11], which is the evolution equation for the probability density function of the mean field Markov process.

Various properties of these metrics and gradient flows have been brought under a lot of attention as well [5, 12, 15, 10, 16]. In addition, the discrete Wasserstein distance on graphs has attracted communities interested in data analysis and more specially in describing clustering and point clouds [27, 28, 29].

This work is partially supported by NSF Awards DMS-1419027, DMS-1620345, and ONR Award N000141310408.

Following the setups in [7], we further study the dynamical properties of the gradient flows in the discrete Wasserstein geometry in this paper. Special attention is given to a free energy containing a quadratic interaction energy, a linear potential and the Boltzmann entropy. In this case, the gradient flow can be viewed as the nonlinear FPE on graphs, which is a set of ordinary differential equations (ODEs). We show that the solution of FPE converges to, the unique or one of the multiple when the free energy is non-convex, Gibbs measure exponentially fast, which mimics the entropy dissipation property, but in a discrete space. We further provide an explicit formula that bounds the asymptotic convergence rate. The continuous analog of the asymptotic rate is related to Yano's formula in Riemannian geometry [32, 33].

It is worth mentioning that the gradient flow considered in this paper is closely related to the novel work of [11]. The difference is at the various goals of studies, which result in different definitions for the 2-Wasserstein metric, so as the gradient flows. The authors of [11] aim at modeling the McKean-Vlasov Markov process on discrete states, in which the Logarithm mean of probability weight (a concept defined in section 2) is used. They show that the gradient flow structure arises as the limit of the N -particle dynamics when N goes to infinity. In this paper, we study the dynamical properties of mean field gradient flows motivated by numerical considerations and population games [6, 8], in which the probability weight is often assumed to be linear.

The structure of this paper is arranged as follows. We review the discrete 2-Wasserstein metric and Fokker-Planck equations on graphs in the next section, then study its convergence in Section 3. In Section 4, we discuss some properties stemming from the convergence rate, including the connection with Yano's formula.

2. OPTIMAL TRANSPORT ON FINITE GRAPHS

In this section, we briefly review the constructions of the 2-Wasserstein metric and corresponding FPE on a graph. We mainly follow the approaches given in [7, 19], with some modified notations for a simpler presentation.

Consider a weighted connected finite graph $G = (V, E, \omega)$, where $V = \{1, 2, \dots, n\}$ is the vertex set, E is the edge set, and $\omega = (\omega_{ij})_{i,j \in V}$ contains the weight of each edge,

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

We assume that G is undirected and contains no self loops or multiple edges. The adjacency set of vertex $i \in V$ is denoted by

$$N(i) = \{j \in V \mid (i, j) \in E\}.$$

The probability set (simplex) supported on all vertices of G is defined by

$$\mathcal{P}(G) = \{(\rho_i)_{i=1}^n \in \mathbb{R}^n \mid \sum_{i=1}^n \rho_i = 1, \quad \rho_i \geq 0, \quad \text{for any } i \in V\},$$

where ρ_i is the discrete probability function at node i . Its interior is denoted by $\mathcal{P}_o(G)$. We introduce the following notations and operations on G and $\mathcal{P}(G)$ and use them for the construction of the discrete 2-Wasserstein metric.

A *vector field* $v = (v_{ij})_{i,j \in V} \in \mathbb{R}^{n \times n}$ on G is a *skew-symmetric matrix* on the edge set E :

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

Given a function $\Phi = (\Phi_i)_{i=1}^n \in \mathbb{R}^n$ defined on the nodes of G , a *potential vector field* $\nabla_G \Phi = (\nabla_G \Phi_{ij})_{i,j \in V} \in \mathbb{R}^{n \times n}$ refers to

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

For a given probability function $\rho \in \mathcal{P}(G)$ and a vector field v , we define the product $\rho v \in \mathbb{R}^{n \times n}$, called *flux function* on G , by

$$\rho v := (v_{ij} \theta_{ij}(\rho))_{(i,j) \in E},$$

where $\theta_{ij}(\rho)$ are chosen as

$$\theta_{ij}(\rho) = \frac{\rho_i + \rho_j}{2}, \quad \text{for any } (i, j) \in E.$$

Remark 1. There are many other choices of θ_{ij} for modeling and computations, such as Logarithm mean [7, 19] and up-wind type [7]; see details in appendix.

We define the *divergence* of ρv on G by

$$\text{div}_G(\rho v) := - \left(\sum_{j \in N(i)} \sqrt{\omega_{ij}} v_{ij} \theta_{ij}(\rho) \right)_{i=1}^n \in \mathbb{R}^n.$$

Given two vector fields v, w on a graph and $\rho \in \mathcal{P}(G)$, the discrete *inner product* is defined by

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

The coefficient $1/2$ in front of the summation accounts for the fact that every edge in G is counted twice. In particular, we have

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

With these definitions, we introduce an integration by parts formula on graphs that will be used throughout this paper: *For any vector field v and potential function Φ on a graph, the following properties hold*

$$- \sum_{i=1}^n \text{div}_G(\rho v)|_i \Phi_i = (v, \nabla_G \Phi)_\rho,$$

and

$$\sum_{i=1}^n \operatorname{div}_G(\rho v)|_i = 0 .$$

Proof. From $v_{ij} + v_{ji} = 0$, we write

$$\begin{aligned} -\sum_{i=1}^n \operatorname{div}_G(\rho v)|_i \Phi_i &= \sum_{i=1}^n \sum_{j \in N(i)} \sqrt{\omega_{ij}} v_{ij} \theta_{ij} \Phi_i \\ &= \frac{1}{2} \left(\sum_{(i,j) \in E} \sqrt{\omega_{ij}} v_{ij} \Phi_i \theta_{ij} + \sum_{(j,i) \in E} \sqrt{\omega_{ij}} v_{ji} \Phi_j \theta_{ji} \right) \\ &= \frac{1}{2} \sum_{(i,j) \in E} v_{ij} \sqrt{\omega_{ij}} (\Phi_i - \Phi_j) \theta_{ij} \\ &= (\nabla_G \Phi, v)_\rho . \end{aligned}$$

Let $\Phi = (1, \dots, 1)^T$, then $\sum_{i=1}^n \operatorname{div}_G(\rho v)|_i = -\sum_{i=1}^n (v, \nabla_G 1)_\rho = 0$.

□

We prefer not to replace θ_{ij} by its explicit formula, as done in [7, 19], to emphasize the freedom of using different θ_{ij} , which can result in different definitions for the flux function, divergence operator and inner product, and hence lead to different definitions for the discrete 2-Wasserstein metric.

2.1. 2-Wasserstein metric on a graph. The discrete analogue of 2-Wasserstein metric \mathcal{W}_2 on probability set $\mathcal{P}_o(G)$ can be given as following. For any given $\rho^0, \rho^1 \in \mathcal{P}_o(G)$, define

$$\mathcal{W}_2^2(\rho^0, \rho^1) := \inf_v \left\{ \int_0^1 (v(t), v(t))_{\rho(t)} dt : \dot{\rho}(t) + \operatorname{div}_G(\rho(t)v(t)) = 0, \quad \rho(0) = \rho^0, \quad \rho(1) = \rho^1 \right\} , \quad (1)$$

where $\dot{\rho}(t) = \frac{d}{dt} \rho(t)$ and the infimum is taken over all vector fields v on a graph, and ρ is a continuously differentiable curve $\rho: [0, 1] \rightarrow \mathcal{P}_o(G)$.

This is the corresponding Benamou-Brenier formula [2] in discrete space. Modifying a similar proof as given in [19], one can show the following lemma; see details in [18].

Lemma 1. *Given a vector field on a graph $v = (v_{ij})_{(i,j) \in E}$ with $v_{ij} = -v_{ji}$, and a measure $\rho \in \mathcal{P}_o(G)$, there exists a unique decomposition, such that*

$$v = \nabla_G \Phi + u, \quad \text{and} \quad \operatorname{div}_G(\rho u) = 0 ,$$

where Φ is a function defined on V . In addition, the following property holds,

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

One may view Lemma 1 as a discrete analogue of the well-known Hodge decomposition. Using it, the metric (1) can be proven equivalent to

$$(\mathcal{W}_2(\rho^0, \rho^1))^2 = \inf_{\Phi} \left\{ \int_0^1 (\nabla_G \Phi, \nabla_G \Phi)_\rho dt : \dot{\rho} + \operatorname{div}_G(\rho \nabla_G \Phi) = 0, \rho(0) = \rho^0, \rho(1) = \rho^1 \right\}, \quad (2)$$

where the infimum is taken over all potentials $\Phi: [0, 1] \rightarrow \mathbb{R}^n$.

Let us denote the tangent space at $\rho \in \mathcal{P}_o(G)$ as

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

We define a weighted graph Laplacian matrix $L(\rho) \in \mathbb{R}^{n \times n}$:

$$L(\rho) = -D^T \Theta(\rho) D,$$

where $D \in \mathbb{R}^{|E| \times |V|}$ is the discrete gradient matrix

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

and $\Theta \in \mathbb{R}^{|E| \times |E|}$ is the diagonal weighted matrix

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

We would like to emphasize that the weights in $L(\rho)$ depend on the distribution ρ .

Lemma 2. *For any given $\sigma \in T_\rho \mathcal{P}_o(G)$, there exists a unique function Φ , up to a constant shift, satisfying*

$$\sigma = L(\rho) \Phi = -\operatorname{div}_G(\rho \nabla_G \Phi).$$

Proof. If $\rho \in \mathcal{P}_o(G)$, all diagonal entries of the weighted matrix $\Theta(\rho)$ are nonzero. Consider

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

then $\Phi_i = \Phi_j$ for any $(i,j) \in E$. Since G is connected, $\Phi_i = \text{constant}$ for any $i \in V$. Thus 0 is a simple eigenvalue of $L(\rho)$ and $L(\rho)(1, \dots, 1)^T = 0$, i.e. $(1, \dots, 1)^T \in \ker(L(\rho))$. Hence $\dim(\mathbb{R}^n / \ker(L(\rho))) = \dim(\operatorname{Ran}(L(\rho))) = \dim(\mathcal{T}_\rho \mathcal{P}_o(G)) = n - 1$. Since $\sum_{i=1}^n \operatorname{div}_G(\rho \nabla_G \Phi)_i = 0$, we have $\operatorname{Ran}(L(\rho)) \subset \mathcal{T}_\rho \mathcal{P}_o(G)$. Therefore

$$(\mathbb{R}^n / \ker(L(\rho))) \cong \operatorname{Ran}(L(\rho)) = \mathcal{T}_\rho \mathcal{P}_o(G),$$

which proves the lemma. \square

Based on Lemma 2, we write

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

where $0 < \lambda_{\sec}(L(\rho)) \leq \dots \leq \lambda_{\max}(L(\rho))$ are eigenvalues of $L(\rho)$ arranged in ascending order, and T is its corresponding eigenvector matrix. We denote the pseudo-inverse of $L(\rho)$ by

$$L^{-1}(\rho) = T \begin{pmatrix} 0 & & & \\ & \frac{1}{\lambda_{\sec} L(\rho)} & & \\ & & \ddots & \\ & & & \frac{1}{\lambda_{\max} L(\rho)} \end{pmatrix} T^{-1} .$$

Then matrix $L^{-1}(\rho)$ endows an inner product on $T_\rho \mathcal{P}_o(G)$.

Definition 3. For any two tangent vectors $\sigma^1, \sigma^2 \in T_\rho \mathcal{P}_o(G)$, define the inner product $g : T_\rho \mathcal{P}_o(G) \times T_\rho \mathcal{P}_o(G) \rightarrow \mathbb{R}$ by

$$g(\sigma^1, \sigma^2) := (\Phi^1)^T L(\rho) (\Phi^2) = (\sigma^1)^T L^{-1}(\rho) \sigma^2 ,$$

where $\sigma^1 = L(\rho) \Phi^1$ and $\sigma^2 = L(\rho) \Phi^2$.

Hence metric (1) is equivalent to

$$(\mathcal{W}_2(\rho^0, \rho^1))^2 = \inf \left\{ \int_0^1 \dot{\rho}^T L^{-1}(\rho) \dot{\rho} dt : \quad \rho(0) = \rho^0, \quad \rho(1) = \rho^1, \quad \rho \in \mathcal{C} \right\} , \quad (3)$$

where \mathcal{C} is the set of all continuously differentiable curves $\rho(t) : [0, 1] \rightarrow \mathcal{P}_o(G)$. From (3), it is clear that $(\mathcal{P}_o(G), \mathcal{W}_2)$ is a finite dimensional Riemannian manifold.

2.2. Gradient flows on finite graphs. We now consider the gradient flow of $\mathcal{F} : \mathcal{P}(G) \rightarrow \mathbb{R}$ on the Riemannian manifold $(\mathcal{P}_o(G), \mathcal{W}_2)$.

Theorem 4 (Gradient flows). For a finite graph G and a constant $\beta > 0$, the gradient flow of any functional $\mathcal{F}(\rho) \in C^2(\mathcal{P}(G))$ on $(\mathcal{P}_o(G), \mathcal{W}_2)$ is

$$\dot{\rho} = -L(\rho) \nabla_\rho \mathcal{F}(\rho) ,$$

i.e.

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} \omega_{ij} \theta_{ij}(\rho) \left(\frac{\partial}{\partial \rho_j} \mathcal{F}(\rho) - \frac{\partial}{\partial \rho_i} \mathcal{F}(\rho) \right) . \quad (4)$$

Proof. For any $\sigma \in T_\rho \mathcal{P}_o(G)$, there exists $\Phi \in \mathbb{R}^n$, such that $\sigma = -\text{div}_G(\rho \nabla_G \Phi) = L(\rho) \Phi$. By Definition 3,

$$(\dot{\rho}, \sigma)_\rho = \dot{\rho}^T L^{-1}(\rho) \sigma = \sum_{i=1}^n \frac{d\rho_i}{dt} \Phi_i . \quad (5)$$

On the right hand side,

$$\begin{aligned} d\mathcal{F}(\rho) \cdot \sigma &= \sum_{i=1}^n \frac{\partial}{\partial \rho_i} \mathcal{F}(\rho) \cdot \sigma_i = F(\rho)^T L(\rho) \Phi \\ &= \Phi^T L(\rho) F(\rho) = - \sum_{i=1}^n \Phi_i \operatorname{div}_G(\rho \nabla_G F(\rho))_i, \end{aligned} \quad (6)$$

where we denote $F(\rho) = (F_i(\rho))_{i=1}^n = (\frac{\partial}{\partial \rho_i} \mathcal{F}(\rho))_{i=1}^n$. Recall the definition of gradient flow on a manifold,

$$(\dot{\rho}, \sigma)_\rho + d\mathcal{F}(\rho) \cdot \sigma = 0,$$

for any $\sigma = \operatorname{div}_G(\rho \nabla_G \Phi) \in T_\rho \mathcal{P}_o(G)$. Substituting (5) and (6) into the above definition, we have

$$\sum_{i=1}^n \left\{ \frac{d\rho_i}{dt} - \operatorname{div}_G(\rho \nabla_G F(\rho)) \right\} \Phi_i = 0.$$

Since $(\Phi_i)_{i=1}^n \in \mathbb{R}^n$ is arbitrary, we must have

$$\frac{d\rho_i}{dt} + \sum_{j \in N(i)} \omega_{ij} \theta_{ij}(\rho) (F_i(\rho) - F_j(\rho)) = 0$$

for all $i \in V$, which is (4). \square

Clearly, (4) is the discrete analog of Wasserstein gradient flow in continuous space

$$\frac{\partial \rho}{\partial t} = \nabla \cdot \left(\rho \nabla \frac{\delta}{\delta \rho} \mathcal{F}(\rho) \right),$$

where $\frac{\delta}{\delta \rho} \mathcal{F}$ is the first variation of \mathcal{F} . In what follows, we consider a particular free energy, which contains a quadratic interaction energy, a linear potential and the Boltzmann entropy:

$$\mathcal{F}(\rho) = \frac{1}{2} \rho^T \mathbb{W} \rho + \mathbb{V}^T \rho + \beta \sum_{i=1}^n \rho_i \log \rho_i,$$

where $\mathbb{V} \in \mathbb{R}^n$, and $\mathbb{W} \in \mathbb{R}^{n \times n}$ is a symmetric matrix. Its gradient flow becomes

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} \omega_{ij} \theta_{ij}(\rho) (\mathbb{V}_j - \mathbb{V}_i + (\mathbb{W}\rho)_j - (\mathbb{W}\rho)_i) + \sum_{j \in N(i)} \omega_{ij} \theta_{ij}(\rho) (\log \rho_j - \log \rho_i), \quad (7)$$

which is the discrete analog of nonlinear FPE

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [\rho \nabla (\mathbb{V}(x) + \int_{\mathbb{R}^d} \mathbb{W}(x, y) \rho(t, y) dy)] + \Delta \rho.$$

So we call (7) nonlinear FPE on graphs. A particular attention is given to

$$\sum_{j \in N(i)} \omega_{ij} \theta_{ij}(\rho) (\log \rho_j - \log \rho_i),$$

which can be viewed as a nonlinear representation of Laplacian operator for ρ .

3. ENTROPY DISSIPATION

In this section, we focus on the convergence properties of FPE (7). Denote the nonlinear Gibbs measure

$$\rho_i^\infty = \frac{1}{K} e^{-\frac{(\mathbf{W}\rho^\infty)_i + \mathbf{V}_i}{\beta}}, \quad \text{where} \quad K = \sum_{j=1}^n e^{-\frac{(\mathbf{W}\rho^\infty)_j + \mathbf{V}_j}{\beta}}.$$

It is easy to verify that the Gibbs measure is the equilibrium of (7). Our main theorem here is to show how fast $\rho(t)$, the solution of FPE (7), converges to ρ^∞ .

We start with the result that $\mathcal{F}(\rho)$ is strictly convex in $\mathcal{P}_o(G)$. It ensures that the equilibrium of gradient flow is unique. In this case, we find a convergence rate for this unique equilibrium. If it is semi-positive definite, as in classical ODE system, it may exist multiple, or a set of equilibrium. An asymptotic statement is provided in the next section.

Theorem 5. *Assume $\rho^0 \in \mathcal{P}_o(G)$ and $\mathcal{F}(\rho)$ is strictly positive definite in $\mathcal{P}(G)$, then there exists a constant $C > 0$, such that*

$$\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty) \leq e^{-Ct} (\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)). \quad (8)$$

Furthermore,

$$C = 2m(\rho^0) \lambda_{\sec}(\hat{L}) \lambda_{\min}(\text{Hess}\mathcal{F}) \frac{1}{(r+1)^2},$$

where

$$r = \sqrt{2} \text{Deg}(G) \max_{(i,j) \in E} \omega_{ij} \frac{\|\text{Hess}\mathcal{F}\|_1}{\lambda_{\min}(\text{Hess}\mathcal{F})^{\frac{3}{2}}} \frac{1 - m(\rho^0)}{m(\rho^0)^2} \frac{\lambda_{\max}(\hat{L})}{\lambda_{\sec}(\hat{L})^2} \sqrt{\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)},$$

$\text{Deg}(G)$ is the maximal degree of the graph, $\hat{L} = D^T D$ is the graph Laplacian matrix, $\lambda_{\sec}(\hat{L})$ and $\lambda_{\max}(\hat{L})$ are the second smallest and the largest eigenvalues of \hat{L} respectively,

$$\|\text{Hess}\mathcal{F}\|_1 = \sup_{\rho \in \mathcal{P}(G)} \|\text{Hess}\mathcal{F}(\rho)\|_1, \quad \lambda_{\min}(\text{Hess}\mathcal{F}) = \min_{\rho \in \mathcal{P}(G)} \lambda_{\min}(\text{Hess}\mathcal{F}(\rho)),$$

and

$$m(\rho_0) = \frac{1}{2} \left(\frac{1}{1 + (2M)^{\frac{1}{\beta}}} \right)^{n-2} \min \left\{ \frac{1}{1 + (2M)^{\frac{1}{\beta}}}, \min_{i \in V} \rho_i^0 \right\} > 0,$$

with

$$M = e^{2 \sup_{i \in V, j \in V} |\mathbf{V}_i| + |\mathbf{W}_{ij}|}. \quad (9)$$

Before giving the complete proof, we want to point out the main difficulties that we must overcome. Since $\mathcal{F}(\rho)$ is strictly convex and ρ^∞ is its unique minimizer, it is not hard to show $\rho(t)$ converging to ρ^∞ . In general, the rate of convergence is determined by comparing the ratio between the first and second derivative of $\mathcal{F}(\rho(t))$ along the gradient flow. If one can find a constant $C > 0$, such that

$$\frac{d^2}{dt^2} \mathcal{F}(\rho(t)) \geq -C \frac{d}{dt} \mathcal{F}(\rho(t)) \quad (10)$$

holds for all $t \geq 0$, one can obtain, by integration,

$$\frac{d}{dt} [\mathcal{F}(\rho^\infty) - \mathcal{F}(\rho(t))] \geq -C [\mathcal{F}(\rho^\infty) - \mathcal{F}(\rho(t))].$$

Then (8) is proved following Gronwall's inequality.

For FPE (7), the first derivative of \mathcal{F} along (7) gives

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = F(\rho)^T \dot{\rho} = -F(\rho)^T L(\rho) F(\rho) = -\dot{\rho}^T L^{-1}(\rho) \dot{\rho} ,$$

while the second derivative is

$$\begin{aligned} \frac{d^2}{dt^2}\mathcal{F}(\rho(t)) &= 2 \dot{\rho}^T \text{Hess}\mathcal{F}(\rho) L(\rho) F(\rho) - F(\rho)^T L(\dot{\rho}) F(\rho) \\ &= 2 \dot{\rho}^T \text{Hess}\mathcal{F}(\rho) \dot{\rho} - \dot{\rho}^T L^{-1}(\rho) L(\dot{\rho}) L^{-1}(\rho) \dot{\rho} , \end{aligned}$$

where $L(\dot{\rho}) = D^T \text{diag}(\theta_{ij}(\dot{\rho})) D$.

Comparing $\frac{d}{dt}\mathcal{F}(\rho(t))$ with $\frac{d^2}{dt^2}\mathcal{F}(\rho(t))$, we find

$$C := \inf_{\rho \in B(\rho^0)} \frac{\dot{\rho}^T \text{Hess}\mathcal{F}(\rho) \dot{\rho}}{\dot{\rho}^T L^{-1}(\rho) \dot{\rho}} - \frac{\dot{\rho}^T L^{-1}(\rho) L(\dot{\rho}) L^{-1}(\rho) \dot{\rho}}{\dot{\rho}^T L^{-1}(\rho) \dot{\rho}} . \quad (11)$$

Quadratic Cubic

However, it is not simple to get an estimation of C . In the continuous case, there are only a few examples [4], depending on special interaction potentials \mathbb{W} , that allow us to find C explicitly. In the discrete space, we overcome this difficulty by borrowing techniques from dynamical systems. If ρ is close enough to the equilibrium ($\dot{\rho}$ is near zero), estimating C in (11) becomes possible. This is because the cubic term of $\dot{\rho}$ in (11) becomes one order smaller than $\dot{\rho}^T L^{-1}(\rho) \dot{\rho}$, and the dominating quadratic term can be estimated by a solvable eigenvalue problem.

Following this idea, the sketch of the proof is as follows: In lemma 6, we first show that the solution of FPE (7) is well defined, and it converges to ρ^∞ . In fact, it can be shown $\rho \in B(\rho^0)$, a compact subset in $\mathcal{P}_o(G)$. Then we estimate the convergence rate in $B(\rho^0)$ by two parts, depending on a parameter $x > 0$ controlling the closeness between ρ and ρ^∞ . If $\rho(t)$ is far away from ρ^∞ , the dissipation formula $\frac{d}{dt}\mathcal{F}(\rho) = -F(\rho)^T L(\rho) F(\rho) < 0$ gives one convergence rate $r_1(x)$; If $\rho(t)$ is close to ρ^∞ , estimating (11) is possible. Thus (10) implies another rate $r_2(x)$. Combing the two together, we find a lower bound of dissipation rate C by calculating

$$\max_{x>0} \min\{r_1(x), r_2(x)\} .$$

Lemma 6. *For any initial condition $\rho^0 \in \mathcal{P}_o(G)$, equation (7) has a unique solution $\rho(t) : [0, \infty) \rightarrow \mathcal{P}_o(G)$. Moreover,*

(i) *There exists a constant $m(\rho_0) > 0$, such that*

$$\rho_i(t) \geq m(\rho_0) > 0 ,$$

for all $i \in V$ and $t \geq 0$.

(ii)

$$\lim_{t \rightarrow +\infty} \rho(t) = \rho^\infty .$$

Remark 2. It is worth mentioning in the proof of Lemma, we uses the requirement of graph being connected. The requirement is needed to match many properties in continuous states. It is used to show that $\rho(t)$ stays in the interior of probability simplex and the equilibria of gradient flow are the critical points of energy in the probability set. This is consistent with the continuous settings. In addition, it is used to guarantee consistent semi-discrete schemes for gradient flows [6] and Wasserstein metric [16].

Proof. First, we prove (i) by constructing a compact set $B(\rho^0) \subset \mathcal{P}_o(G)$. Denote a sequence of constants ϵ_l , $l = 0, 1, \dots, n$,

$$\epsilon_1 = \frac{1}{2} \min\left\{\frac{1}{1 + (2M)^{\frac{1}{\beta}}}, \min_{i \in V} \rho_i^0\right\} \quad \text{and} \quad \epsilon_l = \frac{\epsilon_{l-1}}{1 + (2M)^{\frac{1}{\beta}}}, \quad \text{for } l = 2, \dots, n,$$

where M is defined in (9). Then we define

$$B(\rho^0) = \{(\rho_i)_{i=1}^n \in \mathcal{P}(G) \mid \sum_{r=1}^l \rho_{i_r} \leq 1 - \epsilon_l, \text{ for any } l \in \{1, \dots, n-1\}, \\ \text{and } 1 \leq i_1 < \dots < i_l \leq n\}.$$

We shall show that if $\rho^0 \in B(\rho^0)$, then $\rho(t) \in B(\rho^0)$ for all $t \geq 0$. In other words, the boundary of $B(\rho^0)$ is a repeller for the ODE (7). Assume $\rho(t_1) \in \partial B(\rho^0)$ at time t_1 , this means that there exist indices i_1, \dots, i_l with $l \leq n-1$, such that

$$\sum_{r=1}^l \rho_{i_r}(t_1) = 1 - \epsilon_l. \quad (12)$$

We will show

$$\frac{d}{dt} \sum_{r=1}^l \rho_{i_r}(t)|_{t=t_1} < 0.$$

Let $A = \{i_1, \dots, i_l\}$ and $A^c = V \setminus A$. On the one hand, for any $j \in A^c$,

$$\rho_j(t_1) \leq 1 - \sum_{r=1}^l \rho_{i_r}(t_1) = \epsilon_l. \quad (13)$$

On the other hand, since $\rho(t_1) \in B(\rho_0)$, for any $i \in A$, then $\sum_{k \in A \setminus \{i\}} \rho_k(t_1) \leq 1 - \epsilon_{l-1}$, and from the assumption (12), $\rho_i(t_1) + \sum_{k \in A \setminus \{i\}} \rho_k(t_1) = 1 - \epsilon_l$, we obtain

$$\rho_i(t_1) \geq 1 - \epsilon_l - (1 - \epsilon_{l-1}) = \epsilon_{l-1} - \epsilon_l. \quad (14)$$

Combining equations (13) and (14), we know that for any $i \in A$ and $j \in A^c$,

$$\begin{aligned} F_j(\rho) - F_i(\rho) &= (\mathbb{V}_j + (\mathbb{W}\rho)_j) - (\mathbb{V}_i + (\mathbb{W}\rho)_i) + \beta(\log \rho_j - \log \rho_i) \\ &\leq 2 \sup_{i,j \in V} |\mathbb{V}_i + \mathbb{W}_{ij}| + \beta(\log \epsilon_l - \log(\epsilon_{l-1} - \epsilon_l)) \\ &\leq -\log 2, \end{aligned} \quad (15)$$

where the last inequality is from $\epsilon_l = \frac{\epsilon_{l-1}}{1 + (2M)^{\frac{1}{\beta}}}$ and $M = \sup_{i,j \in V} e^{2(|\mathbb{V}_i| + |\mathbb{W}_{ij}|)}$.

Since the graph is connected, there exists $i_* \in A$, $j_* \in A^c \cap N(i_*)$ such that

$$\sum_{i \in A} \sum_{j \in A^c \cap N(i_*)} \theta_{ij}(\rho(t_1)) \geq \theta_{i_* j_*}(\rho(t_1)) > 0. \quad (16)$$

By combining (15) and (16), we have

$$\begin{aligned} \frac{d}{dt} \sum_{r=1}^l \rho_{i_r}(t)|_{t=t_1} &= \sum_{i \in A} \sum_{j \in N(i)} \theta_{ij}(\rho) [F_j(\rho) - F_i(\rho)]|_{\rho=\rho(t_1)} \\ &= \sum_{i \in A} \left\{ \sum_{j \in A \cap N(i)} \theta_{ij}(\rho) [F_j(\rho) - F_i(\rho)] \right. \\ &\quad \left. + \sum_{j \in A^c \cap N(i)} \theta_{ij}(\rho) [F_j(\rho) - F_i(\rho)] \right\}|_{\rho=\rho(t_1)} \\ &= \sum_{i \in A} \sum_{j \in A^c \cap N(i)} \theta_{ij}(\rho) [F_j(\rho) - F_i(\rho)]|_{\rho=\rho(t_1)} \\ &\leq -\log 2 \sum_{i \in A} \sum_{j \in A^c \cap N(i)} \theta_{ij}(\rho(t_1)) \\ &\leq -\log 2 \theta_{i_* j_*}(\rho(t_1)) < 0, \end{aligned}$$

where the third equality is from $\sum_{(i,j) \in A} \theta_{ij}(F_j - F_i) = 0$. Therefore, we have $\rho(t) \in B(\rho^0)$, thus $\min_{i \in V, t > 0} \rho(t) \geq m(\rho^0)$. (ii) can be proved similarly as in [7], so we omit it here. \square

Lemma 7. For $\rho \in \mathcal{P}_o(G)$, then

$$\lambda_{\sec}(\hat{L}) \cdot \min_{i \in V} \rho_i \leq \lambda_{\sec}(L(\rho)) \leq \lambda_{\max}(L(\rho)) \leq \max_{i \in V} \rho_i \cdot \lambda_{\max}(\hat{L}),$$

and

$$\frac{1}{\max_{i \in V} \rho_i \cdot \lambda_{\max}(\hat{L})} \leq \lambda_{\sec}(L^{-1}(\rho)) \leq \lambda_{\max}(L^{-1}(\rho)) \leq \frac{1}{\min_{i \in V} \rho_i \cdot \lambda_{\sec}(\hat{L})}.$$

Proof. Since

$$\min_{i \in V} \rho_i \cdot \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij}(\Phi_i - \Phi_j)^2 \leq \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij}(\Phi_i - \Phi_j)^2 \theta_{ij}(\rho) \leq \max_{i \in V} \rho_i \cdot \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij}(\Phi_i - \Phi_j)^2,$$

and the Laplacian matrix \hat{L} has the simple eigenvalue 0 with eigenvector $(1, \dots, 1)$, then for any vector $\Phi \in \mathbb{R}^n$ with $\sum_{i=1}^n (\Phi_i - \frac{1}{n} \sum_{j=1}^n \Phi_j)^2 = 1$, we have

$$\min_{i \in V} \rho_i \cdot \lambda_{\sec}(\hat{L}) \leq \Phi^T L(\rho) \Phi \leq \max_{i \in V} \rho_i \cdot \lambda_{\max}(\hat{L}).$$

This implies that

$$\min_{i \in V} \rho_i \cdot \lambda_{\sec}(\hat{L}) \leq \lambda_{\sec}(L(\rho)) \leq \lambda_{\max}(L(\rho)) \leq \max_{i \in V} \rho_i \cdot \lambda_{\max}(\hat{L}).$$

By the definition of $L^{-1}(\rho)$, we can prove the other inequality. \square

Remark 3. The result is also true for general θ_{ij} by assuming $\min(\rho_i, \rho_j) \leq \theta_{ij} \leq \max(\rho_i, \rho_j)$. Write the weighted Laplacian matrix $L(\rho)$ into quadratic form, i.e.

$$\frac{1}{2} \sum_{(i,j) \in E} (\Phi_i - \Phi_j)^2 \cdot \min_{i \in V} \rho_i \leq \Phi^T L(\rho) \Phi = \frac{1}{2} \sum_{(i,j) \in E} (\Phi_i - \Phi_j)^2 \theta_{ij} \leq \frac{1}{2} \sum_{(i,j) \in E} (\Phi_i - \Phi_j)^2 \cdot \max_{i \in V} \rho_i .$$

It is clear that the eigenvalue of $L(\rho)$ depends on the upper and lower bound of probability weight θ_{ij} .

We are now ready to prove the main result.

Proof of Theorem 5. Given a parameter $x > 0$, we divide $B(\rho^0)$ into two parts:

$$B(\rho^0) = \underbrace{\{\rho \in B(\rho^0) : \mathcal{F}(\rho) - \mathcal{F}(\rho^\infty) \geq x\}}_{B_1} \cup \underbrace{\{\rho \in B(\rho^0) : \mathcal{F}(\rho) - \mathcal{F}(\rho^\infty) \leq x\}}_{B_2}$$

We consider the convergence rate in B_1 first.

Lemma 8. Denote $r_1(x) = C_1 x$, where

$$C_1 = 2m(\rho^0) \lambda_{\sec}(\hat{L}) \lambda_{\min}(\text{Hess}\mathcal{F}) \frac{1}{\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)} .$$

Then

$$\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^0) \leq e^{-r_1(x)t} (\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)) , \quad (17)$$

for any $t \leq T = \inf\{\tau > 0 : \mathcal{F}(\rho(\tau)) - \mathcal{F}(\rho^\infty) = x\}$.

Proof. We shall show

$$\frac{\min_{\rho \in B_1} \{F(\rho)^T L(\rho) F(\rho)\}}{\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)} \geq C_1 x .$$

If this is true, then for $t \leq T$,

$$\begin{aligned} \frac{d}{dt} \mathcal{F}(\rho(t)) &= -F(\rho)^T L(\rho) F(\rho) \\ &\leq -\min_{\rho \in B_1} F(\rho)^T L(\rho) F(\rho) \frac{\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty)}{\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty)} \\ &\leq -\frac{\min_{\rho \in B_1} F(\rho)^T L(\rho) F(\rho)}{\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)} [\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty)] \\ &\leq -C_1 x [\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty)] . \end{aligned}$$

From Gronwall's inequality, (17) is proven.

By Taylor expansion on ρ , we have

$$\mathcal{F}(\rho^\infty) = \mathcal{F}(\rho) + F(\rho) \cdot (\rho^\infty - \rho) + \frac{1}{2} (\rho^\infty - \rho)^T \text{Hess}\mathcal{F}(\bar{\rho}) (\rho^\infty - \rho) ,$$

where $\bar{\rho} = \rho + s(\rho^\infty - \rho)$, for some constant $s \in (0, 1)$. Denote the Euclidean projection matrix onto $\mathcal{T}_\rho \mathcal{P}_o(G)$ by

$$\mathbb{P} = \mathbb{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^T ,$$

where $\mathbf{1} = [1, \dots, 1]^T$ and $\mathbb{I} \in \mathbb{R}^{n \times n}$ is the identity matrix. Since $\mathbb{P}F(\rho) \cdot (\rho - \rho^\infty) = F(\rho) \cdot (\rho - \rho^\infty)$, then

$$\begin{aligned} x &\leq \mathcal{F}(\rho) - \mathcal{F}(\rho^\infty) = \mathbb{P}F(\rho) \cdot (\rho - \rho^\infty) - \frac{1}{2}(\rho - \rho^\infty)^T \text{Hess}\mathcal{F}(\bar{\rho})(\rho - \rho^\infty) \\ &\leq \|\mathbb{P}F(\rho)\|_2 \|\rho - \rho^\infty\|_2 - \frac{1}{2} \lambda_{\min}(\text{Hess}\mathcal{F}) \|\rho - \rho^\infty\|_2^2. \end{aligned}$$

The above implies

$$\begin{aligned} \|\mathbb{P}F(\rho)\|_2 &\geq \frac{x}{\|\rho - \rho^\infty\|_2} + \frac{1}{2} \lambda_{\min}(\text{Hess}\mathcal{F}) \|\rho - \rho^\infty\|_2 \\ &\geq \sqrt{2x \lambda_{\min}(\text{Hess}\mathcal{F})}. \end{aligned}$$

Thus by the Poincaré inequality on graph, i.e.

$$\begin{aligned} F(\rho)^T L(\rho) F(\rho) &= \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij} (F_i(\rho) - F_j(\rho))^2 \theta_{ij}(\rho) \\ &\geq \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij} (F_i(\rho) - F_j(\rho))^2 m(\rho^0) \\ &= \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij} \left[\left(F_i(\rho) - \frac{1}{n} \sum_{k=1}^n F_k(\rho) \right) - \left(F_j(\rho) - \frac{1}{n} \sum_{k=1}^n F_k(\rho) \right) \right]^2 m(\rho^0) \\ &= m(\rho^0) (\mathbb{P}F(\rho))^T \hat{L} (\mathbb{P}F(\rho)) \\ &\geq m(\rho^0) \lambda_{\sec}(\hat{L}) \|\mathbb{P}F(\rho)\|_2^2 \\ &\geq 2m(\rho^0) \lambda_{\min}(\text{Hess}\mathcal{F}) \lambda_{\sec}(\hat{L}) x, \end{aligned} \tag{18}$$

which finishes the proof. \square

Next we give the convergence rate in B_2 .

Lemma 9. Denote $r_2(x) = C_2 - C_3 \sqrt{x}$, where

$$C_2 = 2m(\rho^0) \lambda_{\sec}(\hat{L}) \lambda_{\min}(\text{Hess}\mathcal{F}),$$

and

$$C_3 = 2\sqrt{2} \text{Deg}(G) \max_{(i,j) \in E} \omega_{ij} \frac{\|\text{Hess}\mathcal{F}\|_1}{\sqrt{\lambda_{\min}(\text{Hess}\mathcal{F})}} \frac{1 - m(\rho^0)}{m(\rho^0)} \frac{\lambda_{\max}(\hat{L})}{\lambda_{\sec}(\hat{L})}.$$

Then

$$\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty) \leq e^{-r_2(x)(t-T)} (\mathcal{F}(\rho(T)) - \mathcal{F}(\rho^\infty)), \tag{19}$$

for any $t \geq T = \inf\{\tau > 0: \mathcal{F}(\rho(\tau)) - \mathcal{F}(\rho^\infty) = x\}$.

Proof. We shall show

$$\min_{\rho \in B_2} \left\{ \frac{2\rho^T \text{Hess}\mathcal{F}(\rho) \dot{\rho} - \dot{\rho}^T L^{-1}(\rho) L(\dot{\rho}) L^{-1}(\rho) \dot{\rho}}{\dot{\rho}^T L^{-1}(\rho) \dot{\rho}} \right\} \geq r_2(x).$$

Suppose it is true, then

$$\frac{d^2}{dt^2}\mathcal{F}(\rho(t)) \geq -r_2(x)\frac{d}{dt}\mathcal{F}(\rho(t))$$

holds for all $t \geq T$. Integrating this formula in $[t, +\infty)$, we obtain

$$\frac{d}{dt}[\mathcal{F}(\rho^\infty) - \mathcal{F}(\rho(t))] \geq -r_2(x)[\mathcal{F}(\rho^\infty) - \mathcal{F}(\rho(t))] .$$

By Gronwall's inequality, (19) is proven.

We come back to estimate $r_2(x)$. Since $F(\rho^\infty) = c[1, \dots, 1]^T$ is a constant vector, by Taylor expansion, we have

$$\begin{aligned} x &\geq \mathcal{F}(\rho) - \mathcal{F}(\rho^\infty) \\ &= F(\rho^\infty) \cdot (\rho - \rho^\infty) + \frac{1}{2}(\rho - \rho^\infty)^T \text{Hess}\mathcal{F}(\bar{\rho})(\rho - \rho^\infty) \\ &= \frac{1}{2}(\rho - \rho^\infty)^T \text{Hess}\mathcal{F}(\bar{\rho})(\rho - \rho^\infty) \\ &\geq \frac{1}{2}\lambda_{\min}(\text{Hess}\mathcal{F})\|\rho - \rho^\infty\|_2^2 . \end{aligned}$$

Thus

$$\|\rho - \rho^\infty\|_2 \leq \sqrt{\frac{2}{\lambda_{\min}(\text{Hess}\mathcal{F})}}\sqrt{x}.$$

Since

$$\begin{aligned} \dot{\rho}_i &= (L(\rho)F(\rho))_i \\ &\leq \sum_{j \in N(i)} \omega_{ij}|F_i(\rho) - F_j(\rho)|\theta_{ij}(\rho) \\ &\leq \text{Deg}(G) \max_{(i,j) \in E} \omega_{ij} \max_{(i,j) \in E} |F_i(\rho) - F_j(\rho)| \max_i \rho_i \\ &\leq \text{Deg}(G) \max_{(i,j) \in E} \omega_{ij} \max_{(i,j) \in E} |F_i(\rho) - F_j(\rho)|(1 - m(\rho^0)) \end{aligned}$$

and

$$\begin{aligned} F_i(\rho) - F_j(\rho) &= F_i(\rho^\infty) + \nabla_\rho F_i(\bar{\rho}) \cdot (\rho - \rho^\infty) - F_j(\rho^\infty) - \nabla_\rho F_j(\bar{\rho}) \cdot (\rho - \rho^\infty) \\ &= (\nabla_\rho F_i(\bar{\rho}) - \nabla_\rho F_j(\bar{\rho})) \cdot (\rho - \rho^\infty) \\ &\leq \|\nabla_\rho F_i(\bar{\rho}) - \nabla_\rho F_j(\bar{\rho})\|_2 \|\rho - \rho^\infty\|_2 \\ &\leq 2 \sup_{i \in V, \rho \in \mathcal{P}(G)} \|\nabla_\rho F_i(\rho)\|_2 \|\rho - \rho^\infty\|_2 \\ &\leq 2 \sup_{i \in V, \rho \in \mathcal{P}(G)} \|\nabla_\rho F_i(\rho)\|_1 \|\rho - \rho^\infty\|_2 \\ &= 2\|\text{Hess}\mathcal{F}\|_1 \|\rho - \rho^\infty\|_2 , \end{aligned}$$

where $\bar{\rho}, \tilde{\rho}$ are two discrete densities between the line segment of ρ and ρ^∞ .

Combining these two estimates, we get

$$\|\dot{\rho}\|_\infty \leq 2 \cdot \text{Deg}(G) \max_{(i,j) \in E} \omega_{ij}(1 - m(\rho^0))\|\text{Hess}\mathcal{F}\|_1 \sqrt{\frac{2x}{\lambda_{\min}(\text{Hess}\mathcal{F})}} .$$

Denote

$$L^{-\frac{1}{2}}(\rho) = T \begin{pmatrix} 0 & & & \\ & (\frac{1}{\lambda_{\text{sec}} L(\rho)})^{\frac{1}{2}} & & \\ & & \ddots & \\ & & & (\frac{1}{\lambda_{\text{max}} L(\rho)})^{\frac{1}{2}} \end{pmatrix} T^{-1},$$

and $\sigma = \frac{1}{\|L^{-\frac{1}{2}}(\rho)\dot{\rho}\|_2} L^{-\frac{1}{2}}(\rho)\dot{\rho}$, thus

$$\begin{aligned} & \frac{2\dot{\rho}^T \text{Hess}\mathcal{F}(\rho)\dot{\rho}}{\dot{\rho}^T L^{-1}(\rho)\dot{\rho}} - \frac{\dot{\rho}^T L^{-1}(\rho)L(\dot{\rho})L^{-1}(\rho)\dot{\rho}}{\dot{\rho}^T L^{-1}(\rho)\dot{\rho}} \\ & \geq 2 \frac{\dot{\rho}^T \text{Hess}\mathcal{F}(\rho)\dot{\rho}}{\dot{\rho}^T L^{-1}(\rho)\dot{\rho}} - \|\dot{\rho}\|_\infty \frac{\dot{\rho}^T L^{-1}(\rho) \cdot \hat{L} \cdot L^{-1}(\rho)\dot{\rho}}{\dot{\rho}^T L^{-1}(\rho)\dot{\rho}} \\ & = 2\sigma^T L^{\frac{1}{2}}(\rho) \text{Hess}\mathcal{F}(\rho) L^{\frac{1}{2}}(\rho)\sigma - a(x)\sigma^T L^{-\frac{1}{2}}(\rho) \hat{L} L^{-\frac{1}{2}}(\rho)\sigma \\ & \geq 2\lambda_{\min}(\text{Hess}\mathcal{F})\sigma^T L(\rho)\sigma - a(x)\lambda_{\max}(\hat{L})\sigma^T L^{-1}(\rho)\sigma \\ & \geq 2\lambda_{\min}(\text{Hess}\mathcal{F})\lambda_{\text{sec}}(L(\rho)) - a(x)\lambda_{\max}(\hat{L})\lambda_{\max}(L^{-1}(\rho)) \\ & \geq 2\lambda_{\min}(\text{Hess}\mathcal{F})m(\rho^0)\lambda_{\text{sec}}(\hat{L}) - \frac{a(x)\lambda_{\max}(\hat{L})}{m(\rho^0)\lambda_{\text{sec}}(\hat{L})} \\ & = C_2 - C_3\sqrt{x}, \end{aligned}$$

where the last inequality comes from $\lambda_{\text{sec}}(L(\rho)) \geq m(\rho^0)\lambda_{\text{sec}}(\hat{L})$ and $\lambda_{\max}(L^{-1}(\rho)) \leq \frac{1}{m(\rho^0)\lambda_{\text{sec}}(\hat{L})}$ in Lemma 7. \square

We are ready to find the overall convergence rate. By Lemma 8 and Lemma 9, one can show that for any $t \geq 0$,

$$\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty) \leq e^{-\min\{r_1(x), r_2(x)\}t} (\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)),$$

for any $x > 0$. We estimate a constant rate C by showing

$$\max_{x>0} \min\{C_1 x, C_2 - C_3\sqrt{x}\} \geq C.$$

It is clear that the maximizer $x^* > 0$ is achieved at $C_1 x^* = C_2 - C_3\sqrt{x^*}$, i.e. $\sqrt{x^*} = \frac{-C_3 + \sqrt{C_3^2 + 4C_1 C_2}}{2C_1}$. Thus

$$\begin{aligned} C_1 x^* &= \frac{(-C_3 + \sqrt{C_3^2 + 4C_1 C_2})^2}{4C_1} = \frac{(C_3^2 + 4C_1 C_2 - C_3^2)^2}{4C_1(C_3 + \sqrt{C_3^2 + 4C_1 C_2})^2} \\ &\geq \frac{16C_1^2 C_2^2}{4C_1 \cdot 4(C_3 + \sqrt{C_1 C_2})^2} = C_2 \frac{1}{(\frac{C_3}{\sqrt{C_1 C_2}} + 1)^2}, \end{aligned}$$

which finishes the proof. \square

Remark 4. For general choice $\theta_{ij}(\rho) \in C^1$, especially the Logarithm mean used in [11], the rate can also be established. Following the proof in Lemma 9, one only needs to replace r by $\bar{r} = r \max_{\rho \in B(\rho^0), (i,j) \in E} \frac{\partial \theta_{ij}}{\partial \rho_i}$. Since $B(\rho^0)$ stays in a compact set away from

the boundary of probability set, then $\bar{r} < \infty$ is a finite number. Thus the convergence rate for generalized θ_{ij} is given by

$$C = 2m(\rho^0)\lambda_{\text{sec}}(\hat{L})\lambda_{\min}(\text{Hess}\mathcal{F})\frac{1}{(\bar{r}+1)^2}.$$

In addition, when $\mathbb{W} = 0$, Theorem 5 gives the exponential convergence of linear FPE on graphs, for any potential $\mathbb{V} \in \mathbb{R}^n$.

It is also important to understand how the constants scale with the graph size, i.e. the number of nodes and edges. The general rate depends on the smallest value of ρ^0 , which can be very small if it is given near the boundary of probability simplex. Due to the definition of the discrete Laplacian, this eigenvalue problem depends on the size of the graph. Our numerical calculations, in [6], indicate that the constants are smaller for graphs with more nodes, but they become larger if more edges are presented (denser graph).

In some cases, we can estimate the asymptotic rate analytically, an example is given Remark 5.

3.1. Inequalities. In the literature, it is well known that the convergence of FPE can be used to prove the so called Log-Sobolev inequality and a few others. We mimic this result on graphs and further extend the inequality to the case that includes the nonlinear interaction energy. For simplicity, we take $\beta = 1$ and consider $\mathcal{F}(\rho) = \frac{1}{2}\rho^T \mathbb{W} \rho + \mathbb{V}^T \rho + \sum_{i=1}^n \rho_i \log \rho_i$, which is strictly convex in $\mathcal{P}(G)$. Again, we denote ρ^∞ as the Gibbs measure.

The Log-Sobolev inequality describes a relationship between two functionals named relative entropy and relative Fisher information, which can be expressed using our notations in the following formulas,

$$\mathcal{H}(\rho|\rho^\infty) := \mathcal{F}(\rho) - \mathcal{F}(\rho^\infty) \quad \text{Relative entropy ;} \quad (20)$$

and

$$\begin{aligned} \mathcal{I}(\rho|\rho^\infty) &:= F(\rho)^T L(\rho) F(\rho) \quad \text{Relative Fisher information} \\ &= \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij} \left(\log \frac{\rho_i}{e^{-(\mathbb{W}\rho)_i - \mathbb{V}_i}} - \log \frac{\rho_j}{e^{-(\mathbb{W}\rho)_j - \mathbb{V}_j}} \right)^2 \theta_{ij}(\rho). \end{aligned} \quad (21)$$

Corollary 10. *If $\mathcal{F}(\rho)$ is strictly convex in $\mathcal{P}(G)$, then there exists a constant $\lambda > 0$, such that*

$$\mathcal{H}(\rho|\rho^\infty) \leq \frac{1}{2\lambda} \mathcal{I}(\rho|\rho^\infty).$$

We want to point out that when $\mathbb{W} = 0$, corollary (10) is reduced to the standard Log-Sobolev inequality. In this case, functionals (20) and (21) can be written as

$$\mathcal{H}(\rho) = \sum_{i=1}^n \rho_i \log \frac{\rho_i}{\rho_i^\infty}, \quad \mathcal{I}(\rho) = \frac{1}{2} \sum_{(i,j) \in E} \omega_{ij} \left(\log \frac{\rho_i}{\rho_i^\infty} - \log \frac{\rho_j}{\rho_j^\infty} \right)^2 \theta_{ij}(\rho).$$

Their continuous counterparts are

$$\mathcal{H}(\rho) = \int_{\mathbb{R}^d} \rho(x) \log \frac{\rho(x)}{\rho^\infty(x)} dx, \quad \mathcal{I}(\rho) = \int_{\mathbb{R}^d} (\nabla \log \frac{\rho(x)}{\rho^\infty(x)})^2 \rho(x) dx.$$

Proof. We use the fact that the dissipation of relative entropy is the relative Fisher information along FPE (7),

$$\mathcal{I}(\rho(t)) = F(\rho)^T L(\rho) F(\rho) = -\frac{d}{dt} \mathcal{H}(\rho(t)|\rho^\infty).$$

Similar as in Theorem 5, we divide $\mathcal{P}(G)$ into two regions based on a given parameter $x > 0$:

$$\mathcal{P}(G) = \underbrace{\{\rho \in \mathcal{P}(G) : \mathcal{H}(\rho|\rho^\infty) \leq x\}}_{D_1} \cup \underbrace{\{\rho \in \mathcal{P}(G) : \mathcal{H}(\rho|\rho^\infty) \geq x\}}_{D_2}$$

We shall show two upper bounds of $\frac{\mathcal{H}(\rho)}{\mathcal{I}(\rho)}$ in D_1 and D_2 respectively.

On one hand, consider FPE (7), with $\rho(t)$ starting from an initial measure $\rho \in D_1$. Since $\mathcal{H}(\rho(t)|\rho^\infty)$ is a Lyapunov function, then $\rho(t) \in D_1$ for all $t > 0$. Following Lemma 9, there exists $r_2(x) > 0$, such that

$$\frac{d^2}{dt^2} \mathcal{H}(\rho(t)|\rho^\infty) \geq -r_2(x) \frac{d}{dt} \mathcal{H}(\rho(t)|\rho^\infty),$$

which implies

$$\int_0^\infty \frac{d^2}{d\tau^2} \mathcal{H}(\rho(\tau)|\rho^\infty) d\tau \geq \int_0^\infty -r_2(x) \frac{d}{d\tau} \mathcal{H}(\rho(\tau)|\rho^\infty) d\tau,$$

i.e.

$$\mathcal{I}(\rho|\rho^\infty) = \frac{d}{d\tau} \mathcal{H}(\rho(\tau)|\rho^\infty)|_{\tau=0}^{\tau=\infty} \geq r_2(x) (\mathcal{H}(\rho(\tau)|\rho^\infty)|_{\tau=0}^{\tau=\infty}) = r_2(x) \mathcal{H}(\rho|\rho^\infty),$$

where $\lim_{t \rightarrow \infty} \frac{d}{dt} \mathcal{H}(\rho(t)|\rho^\infty) = \lim_{t \rightarrow \infty} \mathcal{H}(\rho(t)|\rho^\infty) = 0$. Thus

$$\lambda_1 = \sup_{\rho \in D_1} \frac{\mathcal{H}(\rho|\rho^\infty)}{\mathcal{I}(\rho|\rho^\infty)} \leq \frac{1}{r_2(x)} < \infty.$$

On the other hand, if $\rho \in D_2$, we shall show

$$\lambda_2 = \sup_{\rho \in D_2} \frac{\mathcal{H}(\rho|\rho^\infty)}{\mathcal{I}(\rho|\rho^\infty)} \leq \frac{\sup_{\rho \in D_2} \mathcal{H}(\rho|\rho^\infty)}{\inf_{\rho \in D_2} \mathcal{I}(\rho|\rho^\infty)} < \infty.$$

It is trivial that $\mathcal{H}(\rho|\rho^\infty)$ is bounded above. We only need to show $\inf_{\rho \in D_2} \mathcal{I}(\rho|\rho^\infty) > 0$. Assume this is not true, i.e. $\inf_{\rho \in D_2} \mathcal{I}(\rho|\rho^\infty) = 0$. Since $\mathcal{I}(\cdot|\rho^\infty)$ is a lower semi continuous function in $\mathcal{P}(G)$, $\mathcal{I}(\cdot|\rho^\infty)$ is infinity on $\mathcal{P}(G) \setminus \mathcal{P}_o(G)$, and D_2 is a compact set, there exists $\rho^* \in D_2 \cap \mathcal{P}_o(G)$, such that

$$\mathcal{I}(\rho^*|\rho^\infty) = F(\rho^*)^T L(\rho^*) F(\rho^*).$$

This implies $F_i(\rho^*) = F_j(\rho^*)$ for any $(i, j) \in E$. Since G is connected, then $\rho^* = \rho^\infty = \arg \min_{\rho \in \mathcal{P}(G)} \mathcal{F}(\rho)$, which contradicts $\rho^* \in D_2$. By choosing $\frac{1}{2\lambda} = \max\{\lambda_1, \lambda_2\}$, we prove the result. \square

3.2. Asymptotic properties. If \mathbb{W} is not a positive definite matrix, there may exist multiple Gibbs measures. The asymptotic convergence rate can be established whenever the solution is near a equilibrium. In what follows, we study such an asymptotic rate.

Assume that the initial measure ρ^0 is in a basin of attraction of an equilibrium ρ^∞ , meaning

$$(A) \quad \lim_{t \rightarrow \infty} \rho(t) = \rho^\infty \quad \text{and} \quad \rho^\infty \text{ is an isolated equilibrium} .$$

Theorem 11. *Let (A) hold and*

$$\lambda = \lambda_{\text{sec}}(L(\rho^\infty) \cdot \text{Hess}\mathcal{F}(\rho^\infty)) > 0 .$$

Then for any sufficiently small $\epsilon > 0$ satisfying $(\lambda - \epsilon) > 0$, there exists a time $T > 0$, such that when $t > T$,

$$\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty) \leq e^{-2(\lambda - \epsilon)(t - T)}(\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)) .$$

Remark 5. The asymptotical rate clearly depends on the graph size, i.e. the number of nodes and edges. We use an example to illustrate this. Consider the energy only containing the entropy. In this case, $\rho^\infty = (\frac{1}{n}, \dots, \frac{1}{n})$, and the asymptotic rate is given by

$$\lambda_{\text{sec}}(L(\rho^\infty)\text{Hess}\mathcal{F}(\rho^\infty)) = \lambda_{\text{sec}}(\hat{L}) .$$

As in the standard spectral graph theory, the second smallest eigenvalue of a standard Laplacian metric \hat{L} converges to the one in continuous state when the graph becomes denser.

Proof. Since $\lim_{t \rightarrow \infty} \rho(t) = \rho^\infty$, for sufficient small $\epsilon > 0$, there exists $t > T$, such that

$$\lambda_{\text{sec}}(L(\rho) \cdot \text{Hess}\mathcal{F}(\rho)) \geq \lambda - \frac{1}{2}\epsilon ,$$

and

$$\|\dot{\rho}\|_\infty = \|L(\rho)F(\rho)\|_\infty \leq \epsilon \frac{m(\rho^0) \cdot \lambda_{\text{sec}}(\hat{L})}{\lambda_{\text{max}}(\hat{L})} .$$

Similar to the proof of Lemma 9, we have

$$\begin{aligned} \frac{\frac{d^2}{dt^2}\mathcal{F}(\rho)}{\frac{d}{dt}\mathcal{F}(\rho)} &= \frac{2\dot{\rho}^T \text{Hess}\mathcal{F}(\rho)\dot{\rho} - \dot{\rho}^T L^{-1}(\rho)L(\dot{\rho})L^{-1}(\rho)\dot{\rho}}{\dot{\rho}^T L^{-1}(\rho)\dot{\rho}} \\ &\geq 2\lambda_{\text{sec}}(L(\rho) \cdot \text{Hess}\mathcal{F}(\rho)) - \|\dot{\rho}\|_\infty \cdot \frac{\lambda_{\text{max}}(\hat{L})}{m(\rho^0) \cdot \lambda_{\text{sec}}(\hat{L})} \\ &\geq 2(\lambda - \epsilon) . \end{aligned}$$

Following strategies in (10), we prove the result. \square

The techniques used in this proof can also be applied to some non-gradient flows, for example, the FPEs with a non-symmetric interaction potential \mathbb{W} . In this approach, the free energy $\mathcal{F}(\rho)$ no longer exists. However, the relative Fisher information always exists, which is used to measure the closeness between $\rho(t)$ and ρ^∞ .

Corollary 12. *Let (A) hold and*

$$\lambda = \lambda_{\text{sec}}(L(\rho^\infty) \cdot (\mathcal{J}F^T + \mathcal{J}F)(\rho^\infty)) > 0 ,$$

where $\mathcal{J}F$ is the Jacobi operator on the vector function $F(\rho)$. Then for any sufficiently small $\epsilon > 0$ satisfying $(\lambda - \epsilon) > 0$, there exists a time $T > 0$, such that when $t > T$,

$$\mathcal{I}(\rho(t)|\rho^\infty) \leq e^{-2(\lambda-\epsilon)(t-T)} \mathcal{I}(\rho(T)|\rho^\infty) .$$

Proof. Since

$$-\frac{d}{dt} \mathcal{I}(\rho(t)|\rho^\infty) = \dot{\rho}^T (\mathcal{J}F(\rho)^T + \mathcal{J}F(\rho)) \dot{\rho} - \dot{\rho}^T L^{-1}(\rho) L(\dot{\rho}) L^{-1}(\rho) \dot{\rho} .$$

Following the proof in Theorem 11, it is straightforward to show that if $t > T$, there exists $\epsilon > 0$, such that

$$\frac{d}{dt} \mathcal{I}(\rho(t)|\rho^\infty) \leq -2(\lambda - \epsilon) \mathcal{I}(\rho(t)|\rho^\infty) .$$

By the Gronwall's equality, we prove the result. \square

In the end, we shall give an explicit formula for the quadratic form in (11), i.e.

$$\lambda = \min_{\sigma \in \mathcal{T}_\rho \mathcal{P}_o(G)} \frac{\sigma^T \text{Hess} \mathcal{F} \sigma}{\sigma^T L^{-1}(\rho) \sigma} .$$

From Lemma 2, there exists a unique $\Phi \in \mathbb{R}^n$, up to constant shift, such that $\sigma = L(\rho)\Phi$. Thus

$$\begin{aligned} \lambda &= \min_{\Phi \in \mathbb{R}^n} \frac{\Phi^T \cdot L(\rho) \cdot \text{Hess} \mathcal{F} \cdot L(\rho) \cdot \Phi}{\Phi^T L(\rho) \Phi} \\ &= \min_{\Phi \in \mathbb{R}^n} \{ \Phi^T \cdot L(\rho) \cdot \text{Hess} \mathcal{F} \cdot L(\rho) \cdot \Phi : \Phi^T L(\rho) \Phi = 1 \} . \end{aligned} \quad (22)$$

We can rewrite the formula (22) explicitly. Introducing

$$h_{ij,kl} = \left(\frac{\partial^2}{\partial \rho_i \partial \rho_k} + \frac{\partial^2}{\partial \rho_j \partial \rho_l} - \frac{\partial^2}{\partial \rho_i \partial \rho_l} - \frac{\partial^2}{\partial \rho_j \partial \rho_k} \right) \mathcal{F}(\rho) , \quad \text{for any } i, j, k, l \in V ,$$

we have

$$\lambda = \min_{\Phi \in \mathbb{R}^n} \frac{1}{4} \sum_{(i,j) \in E} \sum_{(k,l) \in E} \frac{1}{d_{ij}^2 d_{kl}^2} h_{ij,kl} (\Phi_i - \Phi_j) \theta_{ij} (\Phi_k - \Phi_l) \theta_{kl} , \quad (23)$$

s.t.

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

In fact, it is not hard to show that λ is the eigenvalue problem of Hessian operator at the equilibrium in $(\mathcal{P}_o(G), \mathcal{W}_2)$. In the next section, we shall present what (23) suggests in its continuous analog.

4. CONNECTION WITH WASSERSTEIN GEOMETRY

We exploit the meaning of $h_{ij,kl}$ by examining its continuous analog in this section. Our calculation indicates a nice relation to a famous identity in Riemannian geometry, known as Yano's formula [32, 33].

Consider a smooth finite dimensional Riemannian manifold \mathcal{M} . We assume that \mathcal{M} is oriented, compact and has no boundary. We denote $\mathcal{P}(\mathcal{M})$ the space of density functions supported on \mathcal{M} , $T_\rho\mathcal{P}(\mathcal{M})$ the tangent space at $\rho \in \mathcal{P}(\mathcal{M})$, i.e. $T_\rho\mathcal{P}(\mathcal{M}) = \{\sigma(x) : \int_{\mathcal{M}} \sigma(x)dx = 0\}$. Following Otto calculus in [25, 31], for any $\sigma(x) \in T_\rho\mathcal{P}(\mathcal{M})$, there exists a function $\Phi(x)$ satisfying $\sigma(x) = -\nabla \cdot (\rho \nabla \Phi(x))$. This correspondence and the 2-Wasserstein metric endow an scalar inner product on $T_\rho\mathcal{P}(\mathcal{M})$

$$(\sigma(x), \tilde{\sigma}(x)) = (\nabla \Phi, \nabla \tilde{\Phi})_\rho := \int_{\mathcal{M}} \nabla \Phi \cdot \nabla \tilde{\Phi} \rho dx .$$

Now consider a smooth free energy $\mathcal{F} : \mathcal{P}(\mathcal{M}) \rightarrow \mathbb{R}$. We assume that $\rho^* \in \mathcal{P}(\mathcal{M})$ is an equilibrium satisfying

$$\rho^*(x) > 0 , \quad \nabla \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho)|_{\rho^*} = 0 , \quad (24)$$

where $\frac{\delta}{\delta \rho(x)}$ is the first variation operator in L_2 metric.

To understand $h_{ij,kl}$, we calculate the Hessian of \mathcal{F} at ρ^* with respect to the 2-Wasserstein metric, and show

$$(\text{Hess}_{\mathcal{W}_2} \mathcal{F} \cdot \nabla \Phi, \nabla \Phi)_{\rho^*} = \int_{\mathcal{M}} \int_{\mathcal{M}} (D_x D_y \frac{\delta^2}{\delta \rho(x) \delta \rho(y)} \mathcal{F}(\rho)|_{\rho^*} \nabla \Phi(x), \nabla \Phi(y)) \rho^*(x) \rho^*(y) dx dy , \quad (25)$$

where $\frac{\delta^2}{\delta \rho(x) \delta \rho(y)} \mathcal{F}(\rho)$ is the second variation of functional $\mathcal{F}(\rho)$ in L_2 metric, D_x and D_y are the covariant derivatives in x and y respectively, and Φ an arbitrary smooth function.

It is known that the Hessian can be computed by differentiating the function twice along the geodesic. Then the Hessian at the equilibrium ρ^* satisfies

$$(\text{Hess}_{\mathcal{W}_2} \mathcal{F} \cdot \nabla \Phi, \nabla \Phi)_\rho = \frac{d^2}{dt^2} \mathcal{F}(\rho_t)|_{t=0} .$$

where ρ_t and $\nabla \Phi_t$ are time dependent functions satisfying the geodesic equation [25, 31],

$$\begin{cases} \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t \nabla \Phi_t) = 0 \\ \frac{\partial \Phi_t}{\partial t} + \frac{1}{2} (\nabla \Phi_t)^2 = 0 \end{cases} ,$$

with an initial measure $\rho|_{t=0} = \rho^*$ and velocity $\nabla \Phi_t|_{t=0} = \nabla \Phi$. Following the geodesic, we have

$$\begin{aligned} \frac{d}{dt} \mathcal{F}(\rho_t) &= \int_{\mathcal{M}} \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho_t) \frac{\partial \rho_t}{\partial t} dx = - \int_{\mathcal{M}} \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho_t) \nabla \cdot (\rho_t \nabla \Phi_t) dx \\ &= \int_{\mathcal{M}} \nabla \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho_t) \cdot \nabla \Phi_t(x) \rho_t(x) dx , \end{aligned} \quad (26)$$

where the third equality holds by the integration by parts formula and the fact that \mathcal{M} has no boundary.

Similarly, we obtain the second order derivative along the geodesic,

$$\frac{d^2}{dt^2}\mathcal{F}(\rho_t) = \int_{\mathcal{M}} \frac{d}{dt} \left[\nabla \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho_t) \right] \cdot \nabla \Phi_t(x) \rho_t(x) dx \quad (A_1)$$

$$+ \int_{\mathcal{M}} \nabla \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho_t) \cdot \frac{\partial}{\partial t} (\nabla \Phi_t(x) \rho_t(x)) dx. \quad (A_2)$$

Because $\rho|_{t=0} = \rho^*$ and (24) holds, $(A_2)|_{t=0} = 0$. Thus

$$\begin{aligned} \frac{d^2}{dt^2}\mathcal{F}(\rho_t)|_{t=0} &= (A_1)|_{t=0} = \int_{\mathcal{M}} \frac{d}{dt} \left[\nabla \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho_t) \right] |_{t=0} \cdot \nabla \Phi(x) \rho(x) dx \\ &= \int_{\mathcal{M}} \nabla \frac{d}{dt} \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho_t) |_{t=0} \cdot \nabla \Phi(x) \rho(x) dx \\ &= - \int_{\mathcal{M}} \frac{d}{dt} \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho_t) |_{t=0} \nabla \cdot (\nabla \Phi(x) \rho(x)) dx. \end{aligned} \quad (27)$$

In addition, we compute the first order derivative of $\frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho)$ along the geodesic and obtain

$$\begin{aligned} \frac{d}{dt} \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho_t) |_{t=0} &= \int_{\mathcal{M}} \frac{\delta^2}{\delta \rho(x) \delta \rho(y)} \mathcal{F}(\rho) \frac{\partial \rho(t, y)}{\partial t} |_{t=0} dy \\ &= - \int_{\mathcal{M}} \frac{\delta^2}{\delta \rho(x) \delta \rho(y)} \mathcal{F}(\rho) \nabla \cdot (\nabla \Phi(y) \rho(y)) dy. \end{aligned} \quad (28)$$

Substituting (28) into (27), we get

$$\begin{aligned} \frac{d^2}{dt^2}\mathcal{F}(\rho_t) |_{t=0} &= \int_{\mathcal{M}} \int_{\mathcal{M}} \frac{\delta^2}{\delta \rho(x) \delta \rho(y)} \mathcal{F}(\rho) |_{\rho^*} \nabla \cdot (\rho^*(x) \nabla \Phi(x)) \nabla \cdot (\rho^*(y) \nabla \Phi(y)) dx dy \\ &= \int_{\mathcal{M}} \int_{\mathcal{M}} (D_x D_y \frac{\delta^2}{\delta \rho(x) \delta \rho(y)} \mathcal{F}(\rho) |_{\rho^*} \nabla \Phi(x), \nabla \Phi(y)) \rho^*(x) \rho^*(y) dx dy, \end{aligned} \quad (29)$$

where the second equality is achieved by the integration by parts with respect to x and y . Hence, we obtain (25).

Through (25), we find the continuous analog of $h_{ij,kl}$ as

$$D_x D_y \frac{\delta^2}{\delta \rho(x) \delta \rho(y)} \mathcal{F}(\rho)$$

We next illustrate this analog in a particular situation, namely the linear entropy

$$\mathcal{H}(\rho) = \int_{\mathcal{M}} \rho(x) \log \rho(x) dx.$$

In this case, the unique equilibrium (minimizer) $\rho^*(x) = 1$ is a uniform measure on \mathcal{M} , where the total volume of \mathcal{M} is assumed to be 1. Hence (29) becomes,

$$(\text{Hess}_{\mathcal{W}_2} \mathcal{H} \cdot \nabla \Phi, \nabla \Phi)_{\rho^*} = \int_{\mathcal{M}} [\nabla \cdot (\rho^* \nabla \Phi(x))]^2 \frac{1}{\rho^*(x)} dx = \int_{\mathcal{M}} [\nabla \cdot (\nabla \Phi(x))]^2 dx. \quad (30)$$

The optimal transport theory [25, 31] gives another formulation of Hessian:

$$(\text{Hess}_{\mathcal{W}_2} \mathcal{H} \cdot \nabla \Phi, \nabla \Phi)_{\rho} = \int_{\mathcal{M}} [\text{Ric}(\nabla \Phi(x), \nabla \Phi(x)) + \text{tr}(D^2 \Phi(x) D^2 \Phi(x)^T)] \rho(x) dx, \quad (31)$$

where ρ is an arbitrary density function, D^2 is the second covariant derivative and Ric is the Ricci curvature tensor on \mathcal{M} . Evaluating (31) at the equilibrium $\rho^* = 1$ and comparing it with (30), we observe

$$\int_{\mathcal{M}} [\nabla \cdot (\nabla \Phi)]^2 dx = \int_{\mathcal{M}} [\text{Ric}(\nabla \Phi, \nabla \Phi) + \text{tr}(D^2 \Phi D^2 \Phi^T)] dx ,$$

which is the well-known Yano's formula with vector field $\nabla \Phi$.

APPENDIX

As we mentioned in Section 2, one has the freedom to select different $\theta_{ij} := \theta(\rho_i, \rho_j)$ in the definition of flux function on a graph, as long as the selection satisfies certain properties. In our study of the gradient flow on graphs, we generally require θ_{ij} to satisfy the following two conditions:

- (i)
$$\min\{\rho_i, \rho_j\} \leq \theta_{ij}(\rho) \leq \max\{\rho_i, \rho_j\} ;$$
- (ii)
$$\theta_{ij}(\rho) = \theta_{ji}(\rho) .$$

In practice, there are many choices of θ_{ij} other than the simple average used in this paper. Two examples are up-wind type

$$\theta_{ij}(\rho) := \begin{cases} \rho_i & \text{if } \frac{\partial}{\partial \rho_i} \mathcal{F}(p) > \frac{\partial}{\partial \rho_j} \mathcal{F}(p) , \ j \in N(i) \\ \rho_j & \text{if } \frac{\partial}{\partial \rho_i} \mathcal{F}(p) < \frac{\partial}{\partial \rho_j} \mathcal{F}(p) , \ j \in N(i) , \\ \frac{\rho_i + \rho_j}{2} & \text{if } \frac{\partial}{\partial \rho_i} \mathcal{F}(p) = \frac{\partial}{\partial \rho_j} \mathcal{F}(p) , \ j \in N(i) \end{cases}$$

or Logarithm mean type

$$\theta_{ij}(\rho) := \frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} .$$

Acknowledgement: This paper is based on Chapter 3 of Wuchen Li's thesis, which was written under the supervision of Professor Shui-Nee Chow and Professor Haomin Zhou. We would like to thank Professor Wilfrid Gangbo for many fruitful and inspirational discussions on the related topics.

REFERENCES

- [1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2006.
- [2] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3): 375–393
- [3] Eric Carlen and Wilfrid Gangbo. Constrained steepest descent in the 2-Wasserstein metric. *Annals of mathematics*, 807–846, 2003.
- [4] José A Carrillo, Robert J McCann and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matemática Iberoamericana*, 19(3):971–1018, 2003.
- [5] R. Che, W. Huang, Y. Li and P. Tetali. Convergence to global equilibrium for FokkerPlanck equations on a graph and Talagrand-type inequalities. *Journal of Differential Equations*, 261(4), pp.2552–2583.

- [6] Shui-Nee Chow, Luca Dieci, Wuchen Li and Haomin Zhou. Entropy dissipation semi-discretization schemes for Fokker-Planck equations. *arXiv:1608.02628*, 2016.
- [7] Shui-Nee Chow, Wen Huang, Yao Li and Haomin Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. *Archive for Rational Mechanics and Analysis*, 203(3):969–1008, 2012.
- [8] Shui-Nee Chow, Wuchen Li, Jun Lu and Haomin Zhou. Population games and discrete optimal transport. *arXiv:1704.00855*, 2017.
- [9] Laurent Desvillettes and Cédric Villani. Entropic methods for the study of the long time behavior of kinetic equations. *Transport Theory and Statistical Physics*, pages 155–168, 2001.
- [10] M. Erbar. Gradient flow of the entropy for jump processes. *Ann. Inst. H. Poincaré Probab. Statist.*, (50): 920–945, 2014.
- [11] M. Erbar, M. Fathi, V. Laschos and A. Schlichting. Gradient flow structure for McKean-Vlasov equations on discrete spaces. *Discrete and Continuous Dynamical System*, 2016.
- [12] Matthias Erbar and Jan Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Archive for Rational Mechanics and Analysis*, 206(3): 997–1038, 2012.
- [13] Matthias Erbar and Jan Maas. Gradient flow structures for discrete porous medium equations. *arXiv:1212.1129*, 2012.
- [14] Matthias Erbar, Kazumasa Kuwada, and Karl-Theodor Sturm. On the equivalence of the entropic curvature-dimension condition and Bochners inequality on metric measure spaces. *Inventiones mathematicae* 201.3: 993–1071, 2015.
- [15] Max Fathi and Jan Maas. Entropic Ricci curvature bounds for discrete interacting systems, *arXiv:1501.00562*, 2016.
- [16] N. Gigli and J. Maas. Gromov–Hausdorff Convergence of Discrete Transportation Metrics. *SIAM Journal on Mathematical Analysis*, 45(2), 879899, 2013.
- [17] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.
- [18] Wuchen Li. A study of stochastic differential equations and Fokker-Planck equations with applications. *Phd thesis*, 2016.
- [19] Jan Maas. Gradient flows of the entropy for finite Markov chains. *Journal of Functional Analysis*, 261(8):2250–2292, 2011.
- [20] Jan Maas and Daniel Matthes. Long-time behavior of a finite volume discretization for a fourth order diffusion equation, *arXiv preprint arXiv:1505.03178*, 2015.
- [21] Peter A Markowich and Cédric Villani. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp.*, pages 1–29, 2000.
- [22] Alexander Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Dedicated to Herbert Gajewski on the occasion of his 70th birthday. Nonlinearity*. 24(4): 13–29, 2011.
- [23] Alexander Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calculus of Variations and Partial Differential Equations*, 48(1-2):1–31, 2013.
- [24] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 26(1-2), 2001.
- [25] Felix Otto and Cédric Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis* 173 (2): 361–400, 2000.
- [26] Justin Solomon, Raif Rustamov, Leonidas Guibas, Adrian Butscher. Continuous-Flow Graph Transportation Distances. *arxiv.org/abs/1603.06927*, 2016.
- [27] N.G. Trillos and D. Slepcev. Continuum Limit of Total Variation on Point Clouds. *Archive for Rational Mechanics and Analysis*, 220(1): 193241, 2016.
- [28] N.G. Trillos and D. Slepcev. A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis*, 2016.
- [29] N.G. Trillos. Gromov-Hausdorff limit of Wasserstein spaces on point clouds. *arXiv:1702.03464*, 2017.
- [30] Cédric Villani. *Topics in Optimal transportation*. Number 58. American Mathematical Soc., 2003.
- [31] Cédric Villani. *Optimal transport: old and new*. Volume 338. Springer Science & Business Media, 2008.
- [32] Kentaro Yano. Some remarks on tensor fields and curvature. *Annals of Mathematics*, pages 328–347, 1952.

- [33] Kentaro Yano. Some integral formulas and their applications. *The Michigan Mathematical Journal*, 5(1): 63–73, 1958.

SCHOOL OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES.

E-mail address: chow@math.gatech.edu

E-mail address: wcli@math.ucla.edu

E-mail address: hmzhou@math.gatech.edu