FOKKER-PLANCK EQUATIONS FOR A FREE ENERGY FUNCTIONAL OR MARKOV PROCESS ON A GRAPH

SHUI-NEE CHOW, WEN HUANG, YAO LI AND HAOMIN ZHOU

Abstract. The classical Fokker-Planck equation is a linear parabolic equation which describes the time evolution of probability distribution of a stochastic process defined on a Euclidean space. Corresponding to the stochastic process, there often exists a free energy functional which is defined on the space of probability distributions and is a linear combination of a potential and an entropy. In recent years, it has been shown that Fokker-Planck equation is the gradient flow of the free energy functional defined on the Riemannian manifold of probability distributions whose inner product is generated by a 2-Wasserstein distance. In this paper, we consider similar matters for a free energy functional or Markov process defined on a graph with a finite number of vertices and edges. If $N \geq 2$ is the number of vertices of the graph, we show that the corresponding Fokker-Planck equation is a system of N nonlinear ordinary differential equations defined on a Riemannian manifold of probability distributions. However, in contrast to the case of stochastic processes defined on Euclidean spaces, situation is more subtle for discrete spaces. We have different choices for inner products on the space of probability distributions resulting in different Fokker-Planck equations for the same process. It is shown that there is a strong connection but also substantial differences between the systems of ordinary differential equations and the classical Fokker-Planck equation on Euclidean spaces. Furthermore, each of these systems of ordinary differential equations is a gradient flow for the free energy functional defined on a Riemannian manifold whose metric is closely related to certain Wasserstein metrics. Some examples will also be discussed.

1. Introduction

In this paper, we are concerned with the relationships among three concepts defined on graphs: free energy functional, Fokker-Planck equation and stochastic process. These concepts have been intensively used and studied on continuous state space $\mathbb{R}^N$ in many disciplines and applications. We begin by recalling some of the well known facts about them.

1991 *Mathematics Subject Classification.* 37H10,60J27,60J60.

Key words and phrases. Markov process, Fokker-Planck equation, free energy, Gibbs density, graph.

The second author Wen Huang is partially supported by NSFC, Fok Ying Tung Education Foundation, FANEDD(Grant 200520) and the Fundamental Research Funds for the Central Universities. The third author Yao Li is partially supported by NSF grant DMS0708331. The fourth author Haomin Zhou is supported in part by NSF Faculty Early Career Development (CAREER) Award DMS-0645266.

Consider a stochastic process defined by the following randomly perturbed differential equation,

$$dx = -\nabla\Psi(x)dt + \sqrt{2\beta}dW_t,\ x \in \mathbb{R}^N, \tag{1.1}$$

where $\Psi(x)$ is a given scalar-valued function, β a constant, and dW_t the white noise. This stochastic differential equation (SDE) has been serving as one of the primary and most effective tools in many practical problems that involve uncertainty or incomplete information. Examples can be found in many different disciplines such as finance, physics, chemistry, biology and engineering. Obviously, its solutions (or trajectories) are stochastic processes that are no longer deterministic. Hence, it is more desirable to know their probability distribution properties.

The classical Fokker-Planck equation is a partial differential equation describing the time evolution of the probability density function $\rho(x,t)$ of the trajectories of the SDE (1.1). It has the form

$$\frac{\partial \boldsymbol{\rho}(x,t)}{\partial t} = \nabla\cdot(\nabla\Psi(x)\boldsymbol{\rho}(x,t)) + \beta\Delta\boldsymbol{\rho}(x,t), \tag{1.2}$$

where $\nabla\cdot(\nabla\Psi(x)\boldsymbol{\rho}(x,t))$ is called drift term, and $\Delta\boldsymbol{\rho}(x,t)$ is the diffusion term that is generated by white noise. This is why the SDE (1.1) is also called a diffusion process in stochastic literature. Fokker-Planck equation plays a prominent role in physics, chemical, and biological systems [17, 38, 40], and has been intensively studied.

Unlike the case that there exists a clear relationship between the stochastic process (1.1) and Fokker-Planck equation (1.2), their connections to the free energy functional is much less obvious. The notion of free energy is widely used in many different subjects, and it usually means different things in different contexts. For example, free energy in thermodynamics is related to the maximal amount of work that can be extracted from a system. The concept of free energy is also used in other fields, such as statistical mechanics, biology, chemistry, image processing, Markov network. Readers are referred to [27, 41, 48] for more references.

Free energy functionals may carry different names too. For instance, the free energy is called Helmholtz free energy by physicists and Gibbs free energy by chemists. Although free energy functionals refer to different things in different areas, they often have similar formation, namely it is a scalar valued function defined on probability distributions and composed by a combination of a potential energy U and an entropy functional S, i.e. the free energy is expressed as

$$F(\boldsymbol{\rho}) = U(\boldsymbol{\rho}) - \beta S(\boldsymbol{\rho}), \tag{1.3}$$

where β is a constant coefficient called temperature, and $\boldsymbol{\rho}$ is a probability density function defined on a state space X, which may be "continuous", such as $X = \mathbb{R}^N$, or "discrete" $X = \{a_1, \cdots, a_N\}$. For a system with state space $\mathbb{R}^N$, the potential energy functional is defined by

$$U(\boldsymbol{\rho}) := \int_{\mathbb{R}^N} \Psi(x)\boldsymbol{\rho}(x)dx,$$

where $\Psi(x)$ is a given potential function. The entropy, also called Gibbs-Boltzmann entropy, is given by

$$S(\boldsymbol{\rho}) := -\int_{\mathbb{R}^N} \boldsymbol{\rho}(x) \log \boldsymbol{\rho}(x) dx,$$

which measures the complexity of the system.

It is well known that the global minimizer of the free energy F is a probability distribution called Gibbs distribution

$$\boldsymbol{\rho}^*(x) = \frac{1}{K} e^{-\Psi(x)/\beta}, \qquad \text{where } K = \int_{\mathbb{R}^N} e^{-\Psi(x)/\beta} \, \mathrm{d}x. \tag{1.4}$$

Here, we note that in order for equation (1.4) to be well defined, Ψ must grow rapidly enough to ensure that K is finite. In this paper, we only consider potentials satisfying this condition.

Although historical developments of the free energy and Fokker-Planck equation are not directly related, there are many studies that reveal some connections between them. Here, we list some well known results concerning with the relationships among Fokker-Planck equation, the free energy functional and Gibbs distribution:

(1) The free energy (1.3) is a Lyapunov functional of Fokker-Planck equation (1.2), i.e, if the probability density $\boldsymbol{\rho}(t, x)$ is a solution of (1.2), then $F(\boldsymbol{\rho}(t, x))$ is a decreasing function in time.
(2) Gibbs distribution (1.4), the global minimizer of (1.3), is a stationary solution of Fokker-Planck equation (1.2) [17, 38].

These classical results can also be found in [13, 15, 22, 23, 29].

In recent years, there has been many studies investigating connections among free energy, Fokker-Planck equation, abstract Ricci curvature and optimal transport theory for continuous state space. For example, a remarkable result has been reported in [22, 29] that Fokker-Planck equation is the gradient flow of the free energy functional on a Riemannian manifold that is defined by a space of probability distributions with a 2-Wasserstein metric on it. More precisely, let the state space X be a suitable complete metric space with distance d, and $\mathcal{P}(X)$ be the space of Borel probability measures on X. For any given two elements $\mu_1, \mu_2 \in \mathcal{P}(X)$, the 2-Wasserstein distance between μ_1 and μ_2 is defined by

$$W_2(\mu_1, \mu_2)^2 = \inf_{\lambda \in \mathcal{M}(\mu_1, \mu_2)} \int_{X \times X} d(x, y)^2 d\lambda(x, y), \tag{1.5}$$

where $\mathcal{M}(\mu_1, \mu_2)$ is the collection of Borel probability measures on $X \times X$ with marginals μ_1 and μ_2 respectively. Then $(\mathcal{P}(X), W_2)$ forms a Riemannian manifold and Fokker-Planck equation (1.2) is the gradient flow of the free energy (1.3) on this manifold. Clearly, we have two metric spaces (X, d) and $(\mathcal{P}(X), W_2)$, and there exists an isometric embedding given as $X \to \mathcal{P}(X)$ by $x \to \delta_x$. For the origin of Wasserstein distance, we refer to [12, 47]; and for the modern theory and further discussions on Wasserstein distance, we refer to the articles [2, 7, 8, 9, 14, 16, 26, 30, 45, 46] and references therein.

More recently, it is found that the 2-Wasserstein distance gives minimal energy curves on $X = \mathbb{R}^N$ [31], and the convexity of the entropy on $(\mathcal{P}(X), W_2)$ is equivalent

to the nonnegativity of Ricci curvature, which induces the definition of abstract Ricci curvature on length spaces (spaces that curves can be defined)[11, 36, 37, 42, 43, 44]. Furthermore, it is proved in [21] that if (X, d) is a length space, so is the manifold with the 2-Wasserstein metric $(\mathcal{P}(X), W_2)$.

To summarize, we use Figure 1 to illustrate the relationships among the free energy, Fokker-Planck equation and the stochastic process in the state space $\mathbb{R}^N$. From the free energy point of view, Fokker-Planck equation is the gradient flow of the free energy on the probability space with 2-Wasserstein metric. From the viewpoint of the stochastic process, Fokker-Planck equation describes the time evolution of the probability density function. Therefore, Fokker-Planck equation can be derived from both ends of Figure 1. Furthermore, if we know any one of the three concepts, the other two can be derived from it.

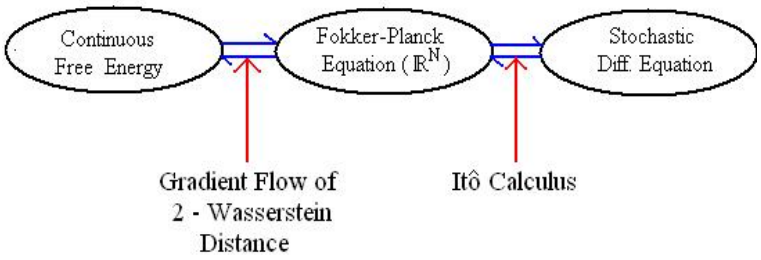


FIGURE 1. Interrelations among the free energy, Fokker-Planck equation and the stochastic differential equation in $\mathbb{R}^N$.

In this paper, we will study similar matters for a discrete state space, such as a graph. For a system with a discrete state space $X = \{a_1, a_2, \cdots, a_N\}$, we denote $\boldsymbol{\rho} = \{\rho_i\}_{i=1}^N$ as a probability distribution on X, i.e.,

$$\sum_{i=1}^{N} \rho_i = 1 \quad \rho_i \geq 0,$$

where ρ_i is the probability of state a_i. Then the free energy functional has the following expression:

$$F(\boldsymbol{\rho}) = \sum_{i=1}^{N} \Psi_i \rho_i + \beta \sum_{i=1}^{N} \rho_i \log \rho_i, \tag{1.6}$$

where Ψ_i is the potential at the state a_i. Obviously, the potential energy functional is given by

$$U(\boldsymbol{\rho}) := \sum_{i=1}^{N} \Psi_i \rho_i,$$

and the entropy is

$$S(\boldsymbol{\rho}) := -\sum_{i=1}^{N} \rho_i \log \rho_i.$$

The free energy functional has a global minimizer, called Gibbs density, is given by

$$\rho_i^* = \frac{1}{K} e^{-\Psi_i/\beta}, \qquad \text{where } K = \sum_{i=1}^{N} e^{-\Psi_i/\beta}. \tag{1.7}$$

Despite of the remarkable developments on the subject on a continuous state space, much less is known if the state space is discrete, especially when X is a graph. There are studies reporting results on the mass transport theory for discrete spaces [3, 28, 39]. However, to the best of our knowledge, Fokker-Planck equation on a graph has not been established. The notion of "white noise" is also not clear for a Markov process defined on the graph. They are the main subjects for this paper.

Due to the developments in the continuous state space, it is natural to apply spatial discretization schemes, such as the well known central difference scheme, to Fokker-Planck equation (1.2) to obtain its counterpart for a discrete state space. This is particularly intuitive if the discrete space is a lattice. The resulting equation for the discrete state space is a coupled system of ordinary differential equations. However, a number of problems arise with this approach. For instance, commonly used discretization schemes often lead to steady states that are different from Gibbs density (1.7), which is the global minimizer of the free energy. This suggests that the equations obtained by the discretization schemes do not capture the real energy landscape of the free energy on the discrete space, and it is not the desired Fokker-Planck equation. To better illustrate this problem, we give a simple, but detailed example in the next section .

Inspired by Figure 1, we can define Fokker-Planck equation on a graph X by two different strategies: (1) From the free energy viewpoint, we will endow a Riemannian metric d, which depends on the potential as well as the structure of the graph, on the probability space $\mathcal{P}(X)$. Then Fokker-Planck equation can be derived as the gradient flow of the free energy F on the Rienmannian manifold $(\mathcal{P}(X), d)$. (2) From the stochastic process viewpoint, we will introduce a new interpretation of white noise perturbations to a Markov process on X, and derive Fokker-Planck equation as the time evolution equation for its probability density function. We must note that unlike the continuous state space case, in which two approaches obtain the same Fokker-Planck equation, we obtain two different Fokker-Planck equations on the graph following these approaches. It seems one of the reasons we obtain different Fokker-Planck equations is that graphs are not length spaces in general.

To be more precise on the approaches, we consider a graph $G = (V, E)$, where $V = \{a_1, \cdots, a_N\}$ is the set of vertices, and E the set of edges. We denote the neighborhood of a vertex $a_i \in V$ as $N(i)$: $N(i) = \{j \in \{1, 2, \cdots, N\} | \{a_i, a_j\} \in E\}$ We further assume that the graph G is a simple graph, i.e., there are no self loops or multiple edges, and G is connected with $|V| \geq 2$. It is worth to mention that

the results reported in the paper still hold after nominal modifications when these assumptions on the graph G are invalid.

We let $\mathbf{\Psi} = (\Psi_i)_{i=1}^N$ be a given potential function on V, where Ψ_i is the potential on a_i, $\beta \geq 0$ be the strength of "white noise". We denote

$$\mathcal{M} = \{\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathbb{R}^N | \sum_{i=1}^N \rho_i = 1 \text{ and } \rho_i > 0 \text{ for } i = 1, 2, \cdots, N\},$$

as the space of all positive probability distributions on V. Then from the free energy viewpoint as shown in Theorem 4.1, we have a Fokker-Planck equation on $\mathcal{M}$:

$$\begin{aligned} \frac{d\rho_i}{dt} = & \sum_{j \in N(i), \Psi_j > \Psi_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_j \\ & + \sum_{j \in N(i), \Psi_j < \Psi_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_i \\ & + \sum_{j \in N(i), \Psi_j = \Psi_i} \beta(\rho_j - \rho_i) \end{aligned} \tag{1.8}$$

for $i = 1, 2 \cdots, N$. If we take the stochastic process viewpoint, we will show in Theorem 5.2 another Fokker-Planck equation on $\mathcal{M}$:

$$\begin{aligned} \frac{d\rho_i}{dt} = & \sum_{j \in N(i), \bar{\Psi}_j > \bar{\Psi}_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_j \\ & + \sum_{j \in N(i), \bar{\Psi}_j < \bar{\Psi}_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_i \end{aligned} \tag{1.9}$$

for $i = 1, 2, \cdots, N$, where $\bar{\Psi}_i = \Psi_i + \beta \log \rho_i$ for $i = 1, 2, \cdots, N$. For convenience, we call equations (1.8) and (1.9) Fokker-Planck equation I (1.8) and II (1.9) respectively.

On one side, Fokker-Planck equation I (1.8) is the gradient flow of the free energy (1.6) on the Riemannian manifold $(\mathcal{M}, d_{\mathbf{\Psi}})$, where $d_{\mathbf{\Psi}}$ is a Riemannian metric on $\mathcal{M}$ induced by Ψ. A key step is to define the Riemannian metric $d_{\mathbf{\Psi}}$ generated by inner products which will be given in Section 3. On the other side, Fokker-Planck equation II (1.9) is derived from a Markov process on G subject to a "white noise" perturbation. Although they are different, both (1.8) and (1.9) share similar properties for $\beta > 0$:

- Free energy F decreases along solutions of both equations.
- Both equations are gradient flows of the same free energy on the same probability space $\mathcal{M}$, but with different metrics.
- Gibbs distribution $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$ is the stable stationary solution of both equations.
- Near Gibbs distribution, the difference between two equations is small.
- For both equations, given any initial condition $\boldsymbol{\rho}^0 \in \mathcal{M}$, there exists a unique solution $\boldsymbol{\rho}(t, \boldsymbol{\rho}^0)$ for $t \geq 0$, and $\boldsymbol{\rho}(t, \boldsymbol{\rho}^0) \to \boldsymbol{\rho}^*$, as $t \to +\infty$.

However, there are differences between equations (1.8) and (1.9). Fokker-Planck equation I (1.8) is obtained from the gradient flow of the free energy F on the

Riemannian metric space $(\mathcal{M}, d_{\mathbf{\Psi}})$. However, its connection to a Markov process on the graph is not clear. On the other hand, Fokker-Planck equation II (1.9) is obtained from a Markov process subject to "white noise" perturbations. This equation can also be considered as a "gradient flow" of the free energy on another metric space $(\mathcal{M}, d_{\bar{\mathbf{\Psi}}})$. However, the geometry of $(\mathcal{M}, d_{\bar{\mathbf{\Psi}}})$ is not smooth. In fact, we will show that in this case, $\mathcal{M}$ is divided into finite segments, and $d_{\bar{\mathbf{\Psi}}}$ is only smooth on each segments. We also note that the manner we derive Fokker-Planck equation II (1.9) seems to be related to Onsager's flux [33, 34, 35].

Formally, if the graph is a lattice, Fokker-Planck equation I (1.8) and II (1.9) can be viewed as special upwind schemes of a Fokker-Planck equation on the continuous state space (1.2). However, they are not commonly used schemes, especially the diffusion term is discretized by a surprising consistent scheme, which, to the best of our knowledge, has not been reported before. It is worth to mention that most of the commonly used consistent and stable schemes lead to unexpected problems similar to the case of the central difference scheme as demonstrated in Section 2.

We also want to mention that results obtained in this paper is largely inspired by the recent developments in Fokker-Planck equation and 2-Wasserstein metric, especially the theory reported in [22, 23, 29]. Our results are also influenced by the upwind schemes for shock capturing in conservation laws [5, 25], as well as the recent studies on Parrondo's paradox [18, 19] and flashing ratchet models for molecular motors [1, 20]. In fact, we will use the flashing ratchet model as an example to demonstrate that our Fokker-Planck equations can be used to explain how the ratchet can turn two energy losing processes into an energy gaining process.

This paper is organized as follows: In Section 2, we give a toy example to compare Fokker-Planck equations I (1.8), II (1.9) and the equation obtained by the standard central difference discretization. Some basic geometric properties of $\mathcal{M}$ are shown in Section 3. In Section 4, we prove that Fokker-Planck equation I (1.8) is the gradient flow of free energy, and show some related properties. In Section 5, we show how we interpret "white noise" in the Markov process to obtain Fokker-Planck equation II (1.9). In Section 6, we explain the upwind structure in Fokker-Planck equations I (1.8) and II (1.9). In the last section, we consider the flashing ratchet model as an application.

2. A Toy Example

In this section, we consider a toy example to compare Fokker-Planck equations I (1.8) and II (1.9) with an equation obtained by discretizing Fokker-Planck equation (1.2) with central finite difference scheme. We shall see that the free energy decreases with time along our Fokker-Planck equations I (1.8) and II (1.9), and Gibbs distribution is the stationary solution of them. While the equation obtained by central difference scheme does not have these properties.

Let us consider a continuous potential function $\Psi : \mathbb{R} \to \mathbb{R}$ shown in Figure 2 (A.1). In this example $\Psi(x)$ is a piecewise polynomial function with continuous

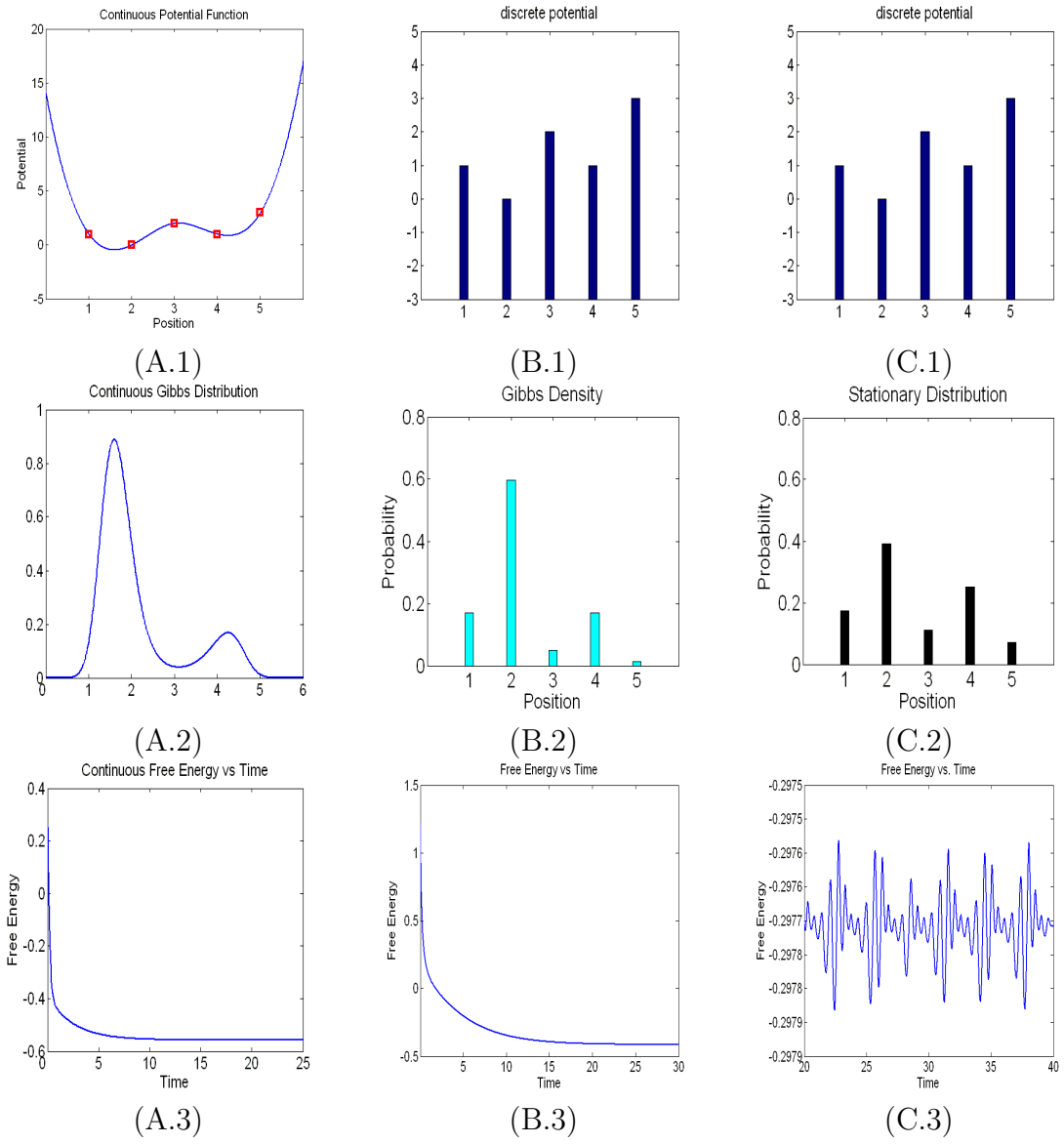


FIGURE 2. (A.1) Potential Ψ in Fokker-Planck equation (1.2). (A.2) Gibbs distribution of Fokker-Planck equation (1.2) for $\beta = 0.8$. (A.3) The free energy decreases with time along Fokker-Planck equation (1.2). (B.1) and (C.1) Potentials on $a_1 = 1, a_2 = 2, a_3 = 3, a_4 = 4, a_5 = 5$. (B.2) Gibbs distribution of Fokker-Planck equation (2.1). (B.3) The free energy decreases with time along Fokker-Planck equation (2.1). (C.2) Stationary distribution of equation (2.3). (C.3) The discrete free energy does not decrease with time along equation (2.3).

second order derivative:

$$\Psi(x) = \begin{cases} -1.5x^3 + 10.5x^2 - 22x + 14 & \text{if} \quad x \leq 3 \\ 1.5x^3 - 16.5x^2 + 59x - 67 & \text{if} \quad x > 3 \end{cases}$$

We fix the temperature, the level of noise, $\beta = 0.8$ through out this example. Before discussing the discrete example, we show some results for the continuous state space. In Figure 2, we plot the steady state solution (see (A.2)) of Fokker-Planck equation (1.2), which is Gibbs distribution. The free energy is a decreasing function along the solution of Fokker-Planck equation as shown in (A.3). These observations match well with the existing theory.

We now consider the discrete space X as a 1-D lattice consisting of five points

$$a_1 = 1, a_2 = 2, a_3 = 3, a_4 = 4, a_5 = 5.$$

In other words, we have the graph $G = (V, E)$, where $V = \{a_1, a_2, \cdots, a_5\}$ and $E = \{\{a_i, a_{i+1}\} : i = 1, 2, 3, 4\}$. The potential function at each point is given by:

$$\Psi_1 = 1, \Psi_2 = 0, \Psi_3 = 2, \Psi_4 = 1, \Psi_5 = 3$$

First, we apply Fokker-Planck equation I (1.8) to G to obtain the following equation:

$$\begin{cases} \dfrac{d\rho_1}{dt} = -\rho_1 + 0.8(\log\rho_2 - \log\rho_1)\rho_1 \\ \dfrac{d\rho_2}{dt} = 2\rho_3 + \rho_1 + 0.8((\log\rho_3 - \log\rho_2)\rho_3 - (\log\rho_2 - \log\rho_1)\rho_1) \\ \dfrac{d\rho_3}{dt} = -3\rho_3 + 0.8((\log\rho_4 - \log\rho_3)\rho_3 - (\log\rho_3 - \log\rho_2)\rho_3) \\ \dfrac{d\rho_4}{dt} = \rho_3 + 2\rho_5 + 0.8((\log\rho_5 - \log\rho_4)\rho_5 - (\log\rho_4 - \log\rho_3)\rho_3) \\ \dfrac{d\rho_5}{dt} = -2\rho_5 - 0.8(\log\rho_5 - \log\rho_4)\rho_5 \end{cases} \tag{2.1}$$

We compute the solution of this equation and found that the free energy decreases in time along the solution (shown in Figure 2 (B.3)), and the solution converges to Gibbs distribution (see Figure 2 (B.2)) as $t \to \infty$.

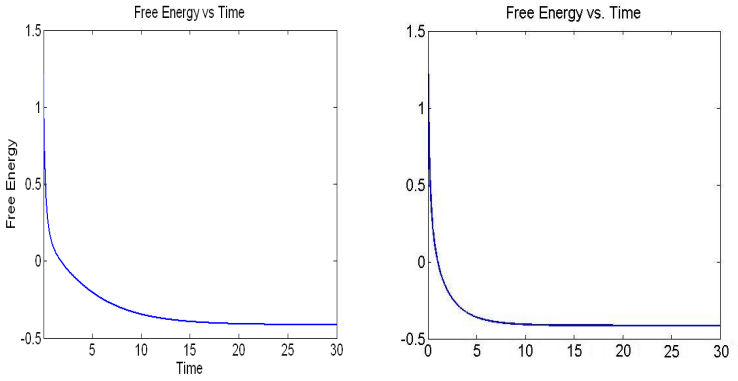
Next, we apply equation (1.9) to G to obtain Fokker-Planck Equation II:

$$\begin{cases} \dfrac{\mathrm{d}\rho_1}{dt} = (0.8(\log\rho_2 - \log\rho_1) - 1)c_{12}(-1.25) \\ \dfrac{\mathrm{d}\rho_2}{dt} = (0.8(\log\rho_2 - \log\rho_1) - 1)c_{12}(-1.25) + (0.8(\rho_3 - \rho_2) + 2)c_{23}(2.5) \\ \dfrac{\mathrm{d}\rho_3}{dt} = (0.8(\rho_2 - \rho_3) - 2)c_{23}(2.5) + (0.8(\rho_4 - \rho_3) - 1)c_{34}(-1.25) \\ \dfrac{\mathrm{d}\rho_4}{dt} = (0.8(\rho_3 - \rho_4) + 1)c_{34}(-1.25) + (0.8(\rho_5 - \rho_4) + 2)c_{45}(2.5) \\ \dfrac{\mathrm{d}\rho_5}{dt} = (0.8(\rho_4 - \rho_5) - 2)c_{45}(2.5), \end{cases} \tag{2.2}$$

where $c_{ij}(x)$ is a number depends on the value of ρ_i and ρ_j

$$c_{ij}(x) = \begin{cases} \rho_i & \text{if} \qquad \rho_i > e^x\rho_j \\ \rho_j & \text{if} \qquad \rho_i < e^x\rho_j \end{cases}$$

This equation has similar properties as equation (2.1), the free energy decreases in time along its solution, and its solution also converges to Gibbs distribution as $t \to \infty$. For comparison purpose, we plot the free energy decaying pattern along the solutions of equation (2.1) and equation (2.2) in the left (L) and right (R) of Figure 2 respectively. Clearly, both curves decrease to 0 as $t \to \infty$.



(L) (R)

FIGURE 3. (L) free energy vs time along Equation (2.1). (R) free energy vs time along Equation (2.2).

We also apply a commonly used finite difference scheme for reaction-diffusion equations to discretize Fokker-Planck equation (1.2). To be more precise, we use upwind scheme for the drift term and central difference scheme for the diffusion term. This yields the following equation,

$$
\begin{cases}
\dfrac{d\rho_1}{dt} = -\rho_1 + 0.8(\rho_2 - \rho_1) \\
\dfrac{d\rho_2}{dt} = 2\rho_3 + \rho_1 + 0.8(\rho_3 - 2\rho_2 + \rho_1) \\
\dfrac{d\rho_3}{dt} = -3\rho_3 + 0.8(\rho_4 - 2\rho_3 + \rho_2) \\
\dfrac{d\rho_4}{dt} = \rho_3 + 2\rho_5 + 0.8(\rho_5 - 2\rho_4 + \rho_3) \\
\dfrac{d\rho_5}{dt} = -2\rho_5 - 0.8(\rho_5 - \rho_4)
\end{cases}
\tag{2.3}
$$

We compute the solution of equation (2.3) and found that the discrete free energy does not decrease in time (see Figure 2 (C.3)), and its stationary distribution (see Figure 2 (C.2)) is not Gibbs distribution. Obviously, these results are not satisfactory and equation (2.3) should not be considered as Fokker-Planck equation for the lattice X.

Of course, one may comment that equation (2.3) can produce expected results if the mesh (lattice) is refined enough and the grid size is sufficiently small. In that case, the discrete solution converges to the solution of Fokker-Planck equation (1.2) for continuous state space. Indeed, this is true. However, in this paper, we

consider problems only defined on discrete spaces. They may not be formulated by discretizing problems of continuous space, and they may not be refined, such as graphs.

Finally, we remark that we have tried many other commonly used discretization schemes for this example. All of them have similar problems as equation (2.3).

3. Metrics on $\mathcal{M}$ and Riemannian manifold

In the introduction, we have claimed that Fokker-Planck equation I (1.8) and Fokker-Planck II (1.9) can be seen as the gradient flows of the free energy with respect to two specific metrics $d_{\mathbf{\Psi}}, d_{\bar{\mathbf{\Psi}}}$ on $\mathcal{M}$. We will give the definitions of $d_{\mathbf{\Psi}}, d_{\bar{\mathbf{\Psi}}}$ in this section. Note that $d_{\mathbf{\Psi}}, d_{\bar{\mathbf{\Psi}}}$ are dependent on the potential $\mathbf{\Psi}$ on V and $\beta > 0$. We will also provide another two Riemannian metrics d_m and d_M on $\mathcal{M}$, which are independent of the potential $\mathbf{\Psi}$ on V and $\beta > 0$, and are upper and lower bounds of the metrics $d_{\mathbf{\Psi}}, d_{\bar{\mathbf{\Psi}}}$. The Riemannian inner products of $d_{\mathbf{\Psi}}$, $d_{\bar{\Psi}}$, d_m and d_M at $\rho \in \mathcal{M}$ are denoted by $g_\rho^{\mathbf{\Psi}}$, $g_\rho^{\bar{\mathbf{\Psi}}}$, g_ρ^m and g_ρ^M respectively. For simplicity in notations, we may omit the sub-index $\boldsymbol{\rho}$ or super-index $\mathbf{\Psi}$ (or $\bar{\mathbf{\Psi}}$) and simply let g or g^Ψ be $g_{\boldsymbol{\rho}}^{\Psi}$ (and similarly let g or $g^{\bar{\mathbf{\Psi}}}$ be $g_\rho^{\bar{\mathbf{\Psi}}}$) if there is no confusion.

Given a graph $G = (V, E)$ with $V = \{a_1, a_2, \cdots, a_N\}$, we consider all positive probability distributions on V:

$$\mathcal{M} = \left\{ \boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathbb{R}^N | \sum_{i=1}^N \rho_i = 1 \text{ and } \rho_i > 0 \text{ for } i \in \{1, 2, \cdots, N\} \right\}.$$

We define its closure as,

$$\overline{\mathcal{M}} = \left\{ \boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathbb{R}^N | \sum_{i=1}^N \rho_i = 1 \text{ and } \rho_i \geq 0 \text{ for } i \in \{1, 2, \cdots, N\} \right\},$$

and denote $\partial\mathcal{M}$ as the boundary of $\mathcal{M}$:

$$\partial\mathcal{M} = \left\{ \boldsymbol{\rho} = \{\rho_i\}_{i=1}^N \in \mathbb{R}^N | \sum_{i=1}^N \rho_i = 1, \rho_i \geq 0 \text{ and } \prod_{i=1}^N \rho_i = 0 \right\}$$

The tangent space $T_\rho\mathcal{M}$ at $\rho \in \mathcal{M}$ is defined by

$$T_\rho\mathcal{M} = \left\{ \boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \in \mathbb{R}^N | \sum_{i=1}^N \sigma_i = 0 \right\}$$

Let d be the standard Euclidean metric on $\mathbb{R}^N$. It is clear that d is also a Riemannian metric on $\mathcal{M}$. Now let

$$\mathbf{\Phi} : (\mathcal{M}, d) \to (\mathbb{R}^N, d) \tag{3.1}$$

with $\mathbf{\Phi}(\boldsymbol{\rho}) = (\Phi_i(\boldsymbol{\rho}))_{i=1}^N$, $\boldsymbol{\rho} \in \mathcal{M}$ be a given smooth map. Next, we will endow a metric $d_{\mathbf{\Phi}}$ on $\mathcal{M}$ which is dependent on $\mathbf{\Phi}$ and the structure of G. Later we will provide precise definitions of $\mathbf{\Phi}$ in different cases. In the following, we just consider $\mathbf{\Phi}$ as an arbitrary but given smooth map.

For two values $r_1 \geq 0$ and $r_2 \geq 0$, we consider the function:

$$e(r_1, r_2) = \begin{cases} \frac{r_1 - r_2}{\log r_1 - \log r_2} & \text{if } r_1 \neq r_2 \text{ and } r_1 r_2 > 0 \\ 0 & \text{if } r_1 r_2 = 0 \\ r_1 & \text{if } r_1 = r_2 \end{cases}.$$

It is not hard to show that e is a continuous function on

$$\{(r_1, r_2) \in \mathbb{R}^2 : r_1 \geq 0, r_2 \geq 0\}$$

and for every $r_1 \geq 0, r_2 \geq 0$,

$$\min\{r_1, r_2\} \leq e(r_1, r_2) \leq \max\{r_1, r_2\}.$$

This says that the following function:

$$\frac{r_1 - r_2}{\log r_1 - \log r_2}$$

can be extended by continuity from the open first quadrant in the plane to its closure. For simplicity, we will use its original form instead of the function $e(r_1, r_2)$.

Given $\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathcal{M}$, we now endow an inner product on $T_{\boldsymbol{\rho}}\mathcal{M}$. We begin by considering the following equivalence relation "$\sim$ " in $\mathbb{R}^N$:

$$\boldsymbol{p} \sim \boldsymbol{q} \quad \text{if and only if} \quad p_1 - q_1 = p_2 - q_2 = \cdots = p_N - q_N.$$

We denote $\mathcal{W}$ as the vector space $\mathbb{R}^N / \sim$. In other words, for $\boldsymbol{p} \in \mathbb{R}^N$ we consider its equivalent class

$$[\boldsymbol{p}] = \{(p_1 + c, p_2 + c, \cdots, p_N + c) : c \in \mathbb{R}\},$$

and all such equivalent classes form the vector space $\mathcal{W}$.

For a given $\boldsymbol{\Phi}$ (see (3.1)) , we now define an identification $\tau_{\boldsymbol{\Phi}}$ from $\mathcal{W}$ to $T_{\boldsymbol{\rho}}\mathcal{M}$. Let $[\boldsymbol{p}] = [(p_i)_{i=1}^N] \in \mathcal{W}$, define $\tau_{\boldsymbol{\Phi}}([\boldsymbol{p}]) = (\sigma_i)_{i=1}^N \in T_{\boldsymbol{\rho}}\mathcal{M}$:

$$\begin{aligned} \sigma_i = & \sum_{j \in N(i), \Phi_j > \Phi_i} (p_i - p_j)\rho_j + \sum_{j \in N(i), \Phi_j < \Phi_i} (p_i - p_j)\rho_i \\ & + \sum_{j \in N(i), \Phi_j = \Phi_i} (p_i - p_j) \frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} \end{aligned} \tag{3.2}$$

for $i = 1, 2, \cdots, N$, where $\Phi_k = \Phi_k(\boldsymbol{\rho})$ for $k \in \{1, 2, \cdots, N\}$. By the identification $\tau_{\boldsymbol{\Phi}}$, we can express $\boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$ by $[\boldsymbol{p}] := \tau_{\boldsymbol{\Phi}}^{-1}(\boldsymbol{\sigma}) \in \mathcal{W}$. When $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^N$ is identified with $[(p_i)_{i=1}^N]$ by $\tau_{\boldsymbol{\Phi}}(\boldsymbol{\sigma}) = [\boldsymbol{p}]$, we write $\boldsymbol{\sigma} \simeq [(p_i)_{i=1}^N]$. It is clear that such identification is dependent on $\boldsymbol{\Phi}$, the probability distribution $\boldsymbol{\rho}$ and the structure of the graph G. Now we show that the identification (3.2) is well defined.

Lemma 3.1. *If each σ_i satisfies* (3.2)*, then the map $\tau_\Phi : [(p_i)_{i=1}^N] \in \mathcal{W} \mapsto \boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \in T_\rho\mathcal{M}$ is a linear isomorphism.*

Proof. It is clear that

$$\tau_{\boldsymbol{\Phi}} : [(p_i)_{i=1}^N] \in \mathcal{W} \mapsto \tau_{\boldsymbol{\Phi}}([(p_i)_{i=1}^N]) = (\sigma_i)_{i=1}^N \in T_\rho\mathcal{M}$$

is a well defined linear map. We note that $\mathcal{W}$ and $T_{\boldsymbol{\rho}}\mathcal{M}$ are both $(N-1)$-dimensional real linear spaces. In order to show the map $\tau_{\boldsymbol{\Phi}}$ is isomorphism, it is sufficient to show

that the map $\tau_{\boldsymbol{\Phi}}$ is injective, which is equivalent to the fact that if $\boldsymbol{p} = \{p_i\}_{i=1}^N \in \mathbb{R}^N$ satisfies

$$\begin{aligned} &\sum_{j\in N(i),\Phi_j>\Phi_i} (p_i - p_j)\rho_j + \sum_{j\in N(i),\Phi_j<\Phi_i} (p_i - p_j)\rho_i \\ &+ \sum_{j\in N(i),\Phi_j=\Phi_i} (p_i - p_j)\frac{\rho_i - \rho_j}{\log\rho_i - \log\rho_j} = 0 \end{aligned} \tag{3.3}$$

for $i = 1, 2, \cdots, N$, then $p_1 = p_2 = \cdots = p_N$.

Let $(p_i)_{i=1}^N \in \mathbb{R}^N$ satisfy (3.3). For $\{a_i, a_j\} \in E$, we set

$$C(\{a_i, a_j\}) = \begin{cases} \rho_j & \text{if } \Phi_i < \Phi_j \\ \rho_i & \text{if } \Phi_i > \Phi_j \\ \frac{\rho_i - \rho_j}{\log\rho_i - \log\rho_j} & \text{if } \Phi_i = \Phi_j \end{cases}.$$

Then

$$\sum_{j\in N(i)} C(\{a_i, a_j\})(p_i - p_j) = 0$$

for $i = 1, 2, \cdots, N$. This implies

$$p_i = \frac{\sum\limits_{j\in N(i)} C(\{a_i, a_j\})p_j}{\sum\limits_{j\in N(i)} C(\{a_i, a_j\})} \tag{3.4}$$

for $i = 1, 2, \cdots, N$. Let $c = \max\{p_i : i = 1, 2, \cdots, N\}$. Now we claim $p_i = c$ for all $i = 1, 2, \cdots, N$, that is, $p_1 = p_2 = \cdots = p_N$. If this is not true, then we can find $\{a_\ell, a_k\} \in E$ such that $p_\ell = c$, $p_k < c$, since the graph G is connected. However, by (3.4),

$$\begin{aligned} c = p_\ell &= \frac{\sum_{j\in N(\ell)} C(\{a_\ell, a_j\})p_j}{\sum_{j\in N(\ell)} C(\{a_\ell, a_j\})} \\ &= c + \frac{\sum_{j\in N(\ell)} C(\{a_\ell, a_j\})(p_j - c)}{\sum_{j\in N(\ell)} C(\{a_\ell, a_j\})} \\ &\le c - \frac{C(\{a_\ell, a_k\})(c - p_k)}{\sum_{j\in N(\ell)} C(\{a_\ell, a_j\})} \\ &< c. \end{aligned}$$

This is a contradiction. And the proof is completed. □

Definition 3.2. By the above identification (3.2), we endow an inner product on $T_{\boldsymbol{\rho}}\mathcal{M}$ as below:

$$g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) \;=\; \sum_{i=1}^N p_i^1\sigma_i^2 = \sum_{i=1}^N p_i^2\sigma_i^1.$$

Note that the above definition is equivalent to

$$\begin{aligned} g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}(\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2) = &\sum_{\{a_i,a_j\}\in E,\Phi_i<\Phi_j} \rho_j(p^1_i-p^1_j)(p^2_i-p^2_j) \\ &+ \sum_{\{a_i,a_j\}\in E,\Phi_i=\Phi_j} (p^1_i-p^1_j)(p^2_i-p^2_j)\frac{\rho_i-\rho_j}{\log\rho_i-\log\rho_j}, \end{aligned} \tag{3.5}$$

for $\boldsymbol{\sigma}^1=(\sigma^1_i)_{i=1}^N, \boldsymbol{\sigma}^2=(\sigma^2_i)_{i=1}^N \in T_\rho\mathcal{M}$, where $[(p^1_i)_{i=1}^N], [(p^2_i)_{i=1}^N] \in \mathcal{W}$ satisfy

$$\boldsymbol{\sigma}^1=(\sigma^1_i)_{i=1}^N \simeq [(p^1_i)_{i=1}^N] \text{ and } \boldsymbol{\sigma}^2=(\sigma^2_i)_{i=1}^N \simeq [(p^2_i)_{i=1}^N].$$

In particular,

$$\begin{aligned} g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}(\boldsymbol{\sigma},\boldsymbol{\sigma}) = &\sum_{\{a_i,a_j\}\in E,\Phi_i<\Phi_j} \rho_j(p_i-p_j)^2 \\ &+ \sum_{\{a_i,a_j\}\in E,\Phi_i=\Phi_j} (p_i-p_j)^2\frac{\rho_i-\rho_j}{\log\rho_i-\log\rho_j} \end{aligned} \tag{3.6}$$

for $\boldsymbol{\sigma}\in T_\rho\mathcal{M}$, where $\boldsymbol{\sigma}\simeq[(p_i)_{i=1}^N]$.

We note that the map $\boldsymbol{\rho}\in\mathcal{M}\mapsto g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}$ is bounded and measurable, but may not be continuous. If $\boldsymbol{\Phi}$ is a constant map, then the map $\boldsymbol{\rho}\in\mathcal{M}\mapsto g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}$ is smooth.

Since $\boldsymbol{\rho}\in\mathcal{M}\mapsto g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}$ is measurable, Using the inner product $g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}$, we can define the distance between two points $\boldsymbol{\rho}^1$ and $\boldsymbol{\rho}^2$ in $\mathcal{M}$ by

$$d_{\boldsymbol{\Phi}}(\boldsymbol{\rho}^1,\boldsymbol{\rho}^2) = \inf_\gamma L(\gamma(t)) \tag{3.7}$$

where $\gamma:[0,1]\to\mathcal{M}$ ranges over all continuously differentiable curve with $\gamma(0)=\boldsymbol{\rho}^1$, $\gamma(1)=\boldsymbol{\rho}^2$, and

$$L(\gamma(t)) = \int_0^1 \sqrt{g^{\boldsymbol{\Phi}}_{\gamma(t)}(\dot\gamma(t),\dot\gamma(t))}dt$$

is the arc length of γ . Although $g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}$ is not a smooth inner product with respect to $\boldsymbol{\rho}$, the length of any smooth curve is still well defined since $\boldsymbol{\rho}\in\mathcal{M}\mapsto g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}$ is measurable. It can be shown by a lemma (Lemma 3.4 towards the end of this section) that $d_{\boldsymbol{\Phi}}$ is a metric on $\mathcal{M}$. Thus we have a metric space $(\mathcal{M},d_{\boldsymbol{\Phi}})$. If $\boldsymbol{\Phi}$ is a constant map, then $(\mathcal{M},g^{\boldsymbol{\Phi}})$ is a Riemannian manifold and the metric $d_{\boldsymbol{\Phi}}$ is a Riemannian metric on $\mathcal{M}$ since the map $\boldsymbol{\rho}\in\mathcal{M}\mapsto g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}$ is smooth.

Remark 3.3. This identification (3.2) is motivated by a similar identification introduced by F. Otto in [29] for the case of a continuous state space. We first replace a corresponding differential operator in [29] by a combination of finite differences because our state space V is discrete. Next, motivated by the upwind scheme in the study of numerical methods for entropy solutions in conservation laws and the structure of Kolmogorov equation (5.1) in section 5, we obtain the identification (3.2) and call the inner product $g^{\boldsymbol{\Phi}}_{\boldsymbol{\rho}}$ the upwind inner product induced by $\boldsymbol{\Phi}$.

Given $\boldsymbol{\rho}=(\rho_i)_{i=1}^N\in\mathcal{M}$. Let the following three matrices

$$A=[A(i,j)]_{N\times N},\ A_m=[A_m(i,j)]_{N\times N}\ ,\ \text{and } A_M=[A_M(i,j)]_{N\times N}$$

be defined as follows: If $i \neq j$, then

$$A(i,j) = \begin{cases} -\rho_j & \text{if } j \in N(i), \Phi_j > \Phi_i \\ -\rho_i & \text{if } j \in N(i), \Phi_j < \Phi_i \\ -\frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} & \text{if } j \in N(i), \Phi_j = \Phi_i \\ 0 & \text{otherwise} \end{cases},$$

$$A_m(i,j) = \begin{cases} -\max\{\rho_i, \rho_j\} & \text{if } \{a_i, a_j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

and

$$A_M(i,j) = \begin{cases} -\min\{\rho_i, \rho_j\} & \text{if } \{a_i, a_j\} \in E \\ 0 & \text{otherwise} \end{cases}.$$

If $i = j$, then

$$\begin{cases} A(i,i) = -\sum_{k \neq i} A(i,k) \\ A_m(i,i) = -\sum_{k \neq i} A_m(i,k) \\ A_M(i,i) = -\sum_{k \neq i} A_M(i,k) \end{cases}.$$

Thus the identification (3.2) can be expressed by

$$\boldsymbol{\sigma}^T = A\boldsymbol{p}^T$$

where $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \in T_{\boldsymbol{\rho}}\mathcal{M}$ and $\boldsymbol{p} = (p_i)_{i=1}^N \in \mathbb{R}^N$.

Now we consider two new identifications

$$\sigma^T = A_m \boldsymbol{p}^T \tag{3.8}$$

and

$$\sigma^T = A_M \boldsymbol{p}^T. \tag{3.9}$$

Similar to Lemma 3.1, the identifications (3.8) and (3.9) are both linear isomorphisms between $T_\rho\mathcal{M}$ and $\mathcal{W}$. We recall that the identification (3.2) induces the inner product $g_\rho^{\boldsymbol{\Phi}}(\cdot,\cdot)$ on $T_\rho\mathcal{M}$. Using the inner product $g^{\boldsymbol{\Phi}}$, we get the distance $d_{\boldsymbol{\Phi}}(\cdot,\cdot)$ on $\mathcal{M}$. Similarly, the identification (3.8)(resp. (3.9)) induces an inner product $g_{\boldsymbol{\rho}}^m(\cdot,\cdot)$ (resp. $g_{\boldsymbol{\rho}}^M(\cdot,\cdot)$) on $T_\rho\mathcal{M}$. It is not hard to see that the map $\boldsymbol{\rho} \mapsto g_{\boldsymbol{\rho}}^m$ (resp. $\boldsymbol{\rho} \mapsto g_{\boldsymbol{\rho}}^M$) is smooth. And using the inner product g^m (resp. g^M), we obtain distance $d_m(\cdot,\cdot)$ (resp. $d_M(\cdot,\cdot)$) on $\mathcal{M}$. Note that $(\mathcal{M}, g^m)$ and $(\mathcal{M}, g^M)$ are smooth Riemannian manifolds.

Lemma 3.4. *For any smooth map $\boldsymbol{\Phi} : (\mathcal{M}, d) \to (\mathbb{R}^N, d)$ and for any $\boldsymbol{\rho}^1, \boldsymbol{\rho}^2 \in \mathcal{M}$,*

$$d_m(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) \leq d_{\boldsymbol{\Phi}}(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) \leq d_M(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2).$$

Proof. Let $\boldsymbol{\Phi} : (\mathcal{M}, d) \to (\mathbb{R}^N, d)$ be a smooth map. Given $\boldsymbol{\rho} \in \mathcal{M}$, the identification (3.2) is given by

$$\boldsymbol{\sigma}^T = A\boldsymbol{p}^T$$

where $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \in T_{\boldsymbol{\rho}}\mathcal{M}$ and $\boldsymbol{p} = (p_i)_{i=1}^N \in \mathbb{R}^N$. Since

$$\sum_{i=1}^N \sigma_i = 0$$

for $(\sigma_i)_{i=1}^N \in T_{\boldsymbol{\rho}}\mathcal{M}$, we delete the last row and column of matrix A to get a symmetric diagonally dominant $(N-1)\times(N-1)$-matrix B and the identify (3.2) becomes

$$\boldsymbol{\sigma}_*^T = B\boldsymbol{p}_*^T$$

where $\boldsymbol{\sigma}_* = (\sigma_i)_{i=1}^{N-1}$ and $\boldsymbol{p}_* = (p_i - p_N)_{i=1}^{N-1}$.

Similarly we can get symmetric diagonally dominant matrices B_m and B_M from A_m and A_M respectively. Moreover, B, B_m and B_M are all irreducible matrices since the graph G is connected. They are all nonsingular and positive definite, because they all have positive diagonally entries.

The inner products are given as

$$\begin{aligned} g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma},\boldsymbol{\sigma}) &= \boldsymbol{\sigma}\boldsymbol{p}^T = \boldsymbol{\sigma}_*\boldsymbol{p}_*^T = \boldsymbol{\sigma}_* B^{-1}\boldsymbol{\sigma}_*^T \\ g_{\boldsymbol{\rho}}^{m}(\boldsymbol{\sigma},\boldsymbol{\sigma}) &= \boldsymbol{\sigma}\boldsymbol{p}^T = \boldsymbol{\sigma}_*\boldsymbol{p}_*^T = \boldsymbol{\sigma}_* B_m^{-1}\boldsymbol{\sigma}_*^T \\ g_{\boldsymbol{\rho}}^{M}(\boldsymbol{\sigma},\boldsymbol{\sigma}) &= \boldsymbol{\sigma}\boldsymbol{p}^T = \boldsymbol{\sigma}_*\boldsymbol{p}_*^T = \boldsymbol{\sigma}_* B_M^{-1}\boldsymbol{\sigma}_*^T \end{aligned}$$

for $\boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$. It is well known that a symmetric diagonally dominant real matrix with nonnegative diagonal entries is positive semidefinite. Since $B_m - B$, $B - B_M$ are still symmetric diagonally dominant matrices with nonnegative diagonal entries, we have $B_m - B$ and $B - B_M$ are positive semidefinite. Now we claim that: for $\boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$,

$$g_{\boldsymbol{\rho}}^{m}(\boldsymbol{\sigma},\boldsymbol{\sigma}) \le g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma},\boldsymbol{\sigma}) \le g_{\boldsymbol{\rho}}^{M}(\boldsymbol{\sigma},\boldsymbol{\sigma}).$$

In order to prove this claim, we first prove

$$g_{\boldsymbol{\rho}}^{m}(\boldsymbol{\sigma},\boldsymbol{\sigma}) \le g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma},\boldsymbol{\sigma}).$$

We note that

$$g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma},\boldsymbol{\sigma}) - g_{\boldsymbol{\rho}}^{m}(\boldsymbol{\sigma},\boldsymbol{\sigma}) = \boldsymbol{\sigma}_*(B^{-1} - B_m^{-1})\boldsymbol{\sigma}_*^T.$$

Therefore, to show that

$$g_{\boldsymbol{\rho}}^{m}(\boldsymbol{\sigma},\boldsymbol{\sigma}) \le g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma},\boldsymbol{\sigma}),$$

it is sufficient to show $(B^{-1} - B_m^{-1})$ is positive semi-definite.

Since B is a positive definite symmetric matrix, B^{-1} is also positive definite symmetric. Combing this with the fact that $(B_m - B)$ is positive semi-definite, we know that $(B^{-1} - B_m^{-1})$ is positive semi-definite from the following equality

$$B^{-1} - B_m^{-1} = B_m^{-1}\{(B_m - B)B^{-1}(B_m - B)^T + (B_m - B)\}(B_m^{-1})^T.$$

In a similar fashion, we prove $g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma},\boldsymbol{\sigma}) \le g_{\boldsymbol{\rho}}^{M}(\boldsymbol{\sigma},\boldsymbol{\sigma})$.

Thus we obtain

$$d_m(\boldsymbol{\rho}^1,\boldsymbol{\rho}^2) \le d_{\boldsymbol{\Phi}}(\boldsymbol{\rho}^1,\boldsymbol{\rho}^2) \le d_M(\boldsymbol{\rho}^1,\boldsymbol{\rho}^2)$$

for any $\boldsymbol{\rho}^1, \boldsymbol{\rho}^2 \in \mathcal{M}$. □

In the remaining part of this section, we focus on the function $\mathbf{\Phi}$ used in the metric derivations. There are two different choices for $\mathbf{\Phi}$ which are related to Fokker-Planck equation I (1.8) and II (1.9) respectively. We consider the given potential $\mathbf{\Psi} = (\Psi_i)_{i=1}^N$ on V and $\beta \geq 0$, where Ψ_i is the potential on vertex a_i.

In the first case for Fokker-Planck equation I (1.8), we let

$$\mathbf{\Phi}(\boldsymbol{\rho}) \equiv \mathbf{\Psi},$$

where $\boldsymbol{\rho} \in \mathcal{M}$. In this case, for a given $\boldsymbol{\rho} \in \mathcal{M}$, the identification (3.2)

$$\boldsymbol{\sigma} \simeq [(p_i)_{i=1}^N]$$

is given by the identification

$$\begin{aligned} \sigma_i = &\sum_{j\in N(i),\Psi_j>\Psi_i} (p_i - p_j)\rho_j + \sum_{j\in N(i),\Psi_j<\Psi_i} (p_i - p_j)\rho_i \\ &+ \sum_{j\in N(i),\Psi_j=\Psi_i} (p_i - p_j)\frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} \end{aligned} \tag{3.10}$$

the norm (3.6) is given by

$$\begin{aligned} g^{\mathbf{\Psi}}_{\boldsymbol{\rho}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) = &\sum_{\{a_i,a_j\}\in E,\Psi_i<\Psi_j} \rho_j(p_i - p_j)^2 \\ &+ \sum_{\{a_i,a_j\}\in E,\Psi_i=\Psi_j} (p_i - p_j)^2\frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} \end{aligned} \tag{3.11}$$

for $\boldsymbol{\sigma} \in T_\rho\mathcal{M}$, where $\boldsymbol{\sigma} \simeq [(p_i)_{i=1}^N]$.

We note that the map $\boldsymbol{\rho} \in \mathcal{M} \mapsto g^{\mathbf{\Psi}}_{\boldsymbol{\rho}}$ is smooth, the inner product $g^{\mathbf{\Psi}}$ generates a Riemannian metric space $(\mathcal{M}, d_{\mathbf{\Psi}})$, where $d_{\mathbf{\Psi}}$ comes from (3.7). We claim that Fokker-Planck equation I (1.8) is the gradient flow of free energy on the Riemannian manifold $(\mathcal{M}, g^{\mathbf{\Psi}})$, which will be proved in section 4.

The second case is for Fokker-Planck equation II (1.9), we have

$$\mathbf{\Phi}(\boldsymbol{\rho}) \equiv \bar{\mathbf{\Psi}}(\boldsymbol{\rho})$$

where $\bar{\mathbf{\Psi}}(\boldsymbol{\rho}) = (\bar{\Psi}_i(\boldsymbol{\rho}))_{i=1}^N$ which is defined by

$$\bar{\Psi}_i(\boldsymbol{\rho}) = \Psi_i + \beta \log \rho_i.$$

In this case, for a given $\boldsymbol{\rho} \in \mathcal{M}$, the identification (3.2)

$$\boldsymbol{\sigma} \simeq [(p_i)_{i=1}^N]$$

is given by the identification

$$\begin{aligned} \sigma_i = &\sum_{j\in N(i),\bar{\Psi}_j>\bar{\Psi}_i} (p_i - p_j)\rho_j + \sum_{j\in N(i),\bar{\Psi}_j<\bar{\Psi}_i} (p_i - p_j)\rho_i \\ &+ \sum_{j\in N(i),\bar{\Psi}_j=\bar{\Psi}_i} (p_i - p_j)\frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} \end{aligned} \tag{3.12}$$

where $\bar{\Psi}_i = \bar{\Psi}_i(\boldsymbol{\rho})$ and the inner product (3.5) on $T_{\boldsymbol{\rho}}\mathcal{M}$ is given by

$$\begin{aligned} g^{\bar{\boldsymbol{\Psi}}}_{\boldsymbol{\rho}}(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) &= \sum_{i=1}^N p_i^1 \sigma_i^2 = \sum_{i=1}^N p_i^2 \sigma_i^1 \\ &= \sum_{\{a_i,a_j\}\in E, \bar{\Psi}_i < \bar{\Psi}_j} \rho_j (p_i^1 - p_j^1)(p_i^2 - p_j^2) \\ &\quad + \sum_{\{a_i,a_j\}\in E, \bar{\Psi}_i = \bar{\Psi}_j} (p_i^1 - p_j^1)(p_i^2 - p_j^2) \frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} \end{aligned} \tag{3.13}$$

for $\boldsymbol{\sigma}^1 = (\sigma_i^1)_{i=1}^N, \boldsymbol{\sigma}^2 = (\sigma_i^2)_{i=1}^N \in T_{\boldsymbol{\rho}}\mathcal{M}$, where $[(p_i^1)_{i=1}^N], [(p_i^2)_{i=1}^N] \in \mathcal{W}$ satisfy $\boldsymbol{\sigma}^1 \simeq [(p_i^1)_{i=1}^N]$ and $\boldsymbol{\sigma}^2 \simeq [(p_i^2)_{i=1}^N]$ by the identification (3.12). Particularly,

$$\begin{aligned} g^{\bar{\boldsymbol{\Psi}}}_{\boldsymbol{\rho}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) &= \sum_{\{a_i,a_j\}\in E, \bar{\Psi}_i < \bar{\Psi}_j} \rho_j (p_i - p_j)^2 \\ &\quad + \sum_{\{a_i,a_j\}\in E, \bar{\Psi}_i = \bar{\Psi}_j} (p_i - p_j)^2 \frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} \end{aligned} \tag{3.14}$$

for $\boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$, where $\boldsymbol{\sigma} \simeq [(p_i)_{i=1}^N]$.

The inner product $g^{\bar{\boldsymbol{\Psi}}}$ gives a metric space $(\mathcal{M}, d_{\bar{\boldsymbol{\Psi}}})$, where $d_{\bar{\boldsymbol{\Psi}}}$ comes from (3.7). The map $\boldsymbol{\rho} \in \mathcal{M} \mapsto g^{\bar{\boldsymbol{\Psi}}}_{\boldsymbol{\rho}}$ is not continuous due to the fact that $\bar{\boldsymbol{\Psi}}$ depends on $\boldsymbol{\rho}$. Then $(\mathcal{M}, g^{\bar{\boldsymbol{\Psi}}})$ is not a Riemannian manifold. However $g^{\bar{\boldsymbol{\Psi}}}_{\boldsymbol{\rho}}$ is a piecewise smooth function with respect to $\boldsymbol{\rho}$, thus the space $(\mathcal{M}, g^{\bar{\boldsymbol{\Psi}}})$ comprises finitely many smooth Riemannian manifolds.

More precisely, the space $\mathcal{M}$ is divided into segments by $N(N-1)/2$ sub-manifolds

$$S_{r,t} = \{(\rho_i)_{i=1}^N \in \mathcal{M} : \Psi_r + \beta \log \rho_r = \Psi_t + \beta \log \rho_t\},\ 1 \le r < t \le N.$$

The inner product $g^{\bar{\boldsymbol{\Psi}}}$ is smooth in each component divided by $\{S_{r,t}\}_{1\le r<t\le N}$, and gives a smooth Riemannian distance in each and every component. Moreover, all sub-manifolds $S_{r,t}$ intersect at one point which is Gibbs distribution $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$, where

$$\rho_i^* = \frac{1}{K} e^{-\frac{\Phi_i}{\beta}}, \qquad K = \sum_{i=1}^N e^{-\frac{\Phi_i}{\beta}}.$$

With above understandings, Fokker-Planck equation II (1.9) can also be seen as the "gradient flow" of free energy on the metric space $(\mathcal{M}, d_{\bar{\boldsymbol{\Psi}}})$, which will be shown in section 5, where by the "gradient flow", we mean that for every $\boldsymbol{\rho} \in \mathcal{M}$ that is not in any S_{ij}, there exists a small neighborhood, such that the solution of Fokker-Planck equation II (1.9) though $\boldsymbol{\rho}$ is the gradient flow of the free energy with respect to the metric $d_{\bar{\boldsymbol{\Psi}}}$.

By Lemma 3.4, $d_{\boldsymbol{\Psi}}$ and $d_{\bar{\boldsymbol{\Psi}}}$ are bounded by d_m and d_M. In general, the explicit expressions of these distances $d_{\boldsymbol{\Psi}}, d_{\bar{\boldsymbol{\Psi}}}, d_m$ and d_M are hard to obtain. In the following example 3.5, we show one explicit expression of the distance function $d_{\boldsymbol{\Psi}}$.

Example 3.5.

We consider a star graph $G = (V, E)$ with

$$V = \{a_1, \cdots, a_N, a_{N+1}\}$$

and

$$E = \{\{a_i, a_{N+1}\} : i = 1, 2, \cdots, N+1\}.$$

Let the potential $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^{N+1}$ on V be defined such that

$$\Psi_i > \Psi_{N+1}, \qquad i = 1, \cdots, N.$$

Then the identification (3.10) on the tangent space $T_{\boldsymbol{\rho}}\mathcal{M}$ is

$$\begin{aligned} \sigma_1 &= (p_1 - p_{N+1})\rho_1 \\ &\cdots \\ \sigma_N &= (p_N - p_{N+1})\rho_N \\ \sigma_{N+1} &= -\sum_{i=1}^{N}(p_i - p_{N+1})\rho_i, \end{aligned}$$

where $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^{N+1} \in T_{\boldsymbol{\rho}}\mathcal{M} \simeq [(p_i)_{i=1}^{N+1}] \in \mathcal{W}$.

By such identification and (3.11), the norm is given by

$$g_{\boldsymbol{\rho}}^{\boldsymbol{\Psi}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) = \sum_{i=1}^{N}(p_i - p_{N+1})^2\rho_i = \sum_{i=1}^{N}\frac{\sigma_i^2}{\rho_i},$$

for $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^{N+1} \in T_{\boldsymbol{\rho}}\mathcal{M}$, where $\boldsymbol{\sigma} \simeq [(p_i)_{i=1}^{N+1}]$.

Given $\boldsymbol{\rho}^1 = (\rho_1^1, \cdots, \rho_{N+1}^1), \boldsymbol{\rho}^2 = (\rho_1^2, \cdots, \rho_{N+1}^2) \in \mathcal{M}$, we suppose that

$$\gamma(t) = (\rho_1(t), \cdots, \rho_N(t), \rho_{N+1}(t)) : [0, 1] \to \mathcal{M}$$

is a continuously differentiable curve from $\boldsymbol{\rho}^1$ to $\boldsymbol{\rho}^2$. Then

$$\begin{aligned} L(\gamma) &= \int_0^1 \sqrt{g_{\gamma(t)}^{\boldsymbol{\Psi}}(\dot{\gamma}(t), \dot{\gamma}(t))}dt \\ &= \int_0^1 \sqrt{\sum_{i=1}^{N}\frac{(\rho_i'(t))^2}{\rho_i(t)}}dt \\ &= \int_0^1 \sqrt{\sum_{i=1}^{N}(x'(t))^2}dt, \end{aligned}$$

where we use the substitution $x_i(t) = \sqrt{\rho_i(t)}$ for $i = 1, \cdots, N$. Let $\alpha(t) = (x_1(t), x_2(t), \cdots, x_N(t))$ for $t \in [0, 1]$ and

$$D = \{(x_1, x_2, \cdots, x_N) \in \mathbb{R}^N : \sum_{i=1}^{N} x_i^2 < 1 \text{ and } x_i > 0 \text{ for } i = 1, 2, \cdots, N\}.$$

Then D is a convex subset of unit open ball in $\mathbb{R}^N$ and α is a continuously differentiable curve in D from $\eta^1 = (\sqrt{\rho_i^1})_{i=1}^N$ to $\eta^2 = (\sqrt{\rho_i^2})_{i=1}^N$. Clearly, we have

$$\begin{aligned} L(\gamma) &= \int_0^1 \sqrt{\sum_{i=1}^N (x'(t))^2 dt} \\ &\geq \|\eta^1 - \eta^2\| \\ &= \sqrt{\sum_{i=1}^N \left(\sqrt{\rho_i^1} - \sqrt{\rho_i^2}\right)^2}. \end{aligned}$$

On the other hand, we take

$$\alpha^*(t) = (x_1^*(t), \cdots, x_N^*(t)) := t\eta^1 + (1-t)\eta^2$$

for $t \in [0,1]$. In fact α^* is the straight line segment in D from $\eta^1 = (\sqrt{\rho_i^1})_{i=1}^N$ to $\eta^2 = (\sqrt{\rho_i^2})_{i=1}^N$. Let $\rho_i^*(t) = (x_i^*(t))^2$ for $i = 1, 2, \cdots, N$ and $\rho_{N+1}^*(t) = 1 - (\sum_{i=1}^N \rho_i^*(t))$. Then

$$\gamma^*(t) = (\rho_i^*(t))_{i=1}^{N+1} : [0,1] \to \mathcal{M}$$

is a continuously differentiable curve from $\boldsymbol{\rho}^1$ to $\boldsymbol{\rho}^2$. This implies that

$$L(\gamma^*) = \int_0^1 \sqrt{\sum_{i=1}^N ((x_i^*(t))')^2 dt} = \|\eta^1 - \eta^2\| = \sqrt{\sum_{i=1}^N \left(\sqrt{\rho_i^1} - \sqrt{\rho_i^2}\right)^2}.$$

Therefore, we obtain

$$d_{\boldsymbol{\Psi}}(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) = \inf_\gamma L(\gamma) = \sqrt{\sum_{i=1}^N \left(\sqrt{\rho_i^1} - \sqrt{\rho_i^2}\right)^2}$$

for $\boldsymbol{\rho}^1 = (\rho_1^1, \cdots, \rho_{N+1}^1), \boldsymbol{\rho}^2 = (\rho_1^2, \cdots, \rho_{N+1}^2) \in \mathcal{M}$. Finally, we note that the metric $d_{\boldsymbol{\Psi}}$ can be extended naturally to the space $\overline{\mathcal{M}}$, i.e., .

$$d_{\boldsymbol{\Psi}}(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) = \sqrt{\sum_{i=1}^N \left(\sqrt{\rho_i^1} - \sqrt{\rho_i^2}\right)^2}$$

for $\boldsymbol{\rho}^1 = (\rho_1^1, \cdots, \rho_N^1), \boldsymbol{\rho}^2 = (\rho_1^2, \cdots, \rho_{N+1}^2) \in \overline{\mathcal{M}}$.

4. Fokker-Planck Equation I

In this section, we first show that Fokker-Planck equation I (1.8) defined on a graph $G = (V, E)$ with potentials $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ on V and $\beta \geq 0$ is the gradient flow of free energy F on Riemannian manifold $(\mathcal{M}, g^{\boldsymbol{\Psi}})$ which is introduced in section 3. Then some basic properties of Fokker-Planck equation I (1.8) will be given.

With fixed $\beta \geq 0$, we have the free energy F on space $\overline{\mathcal{M}}$:

$$F(\boldsymbol{\rho}) = \sum_{i=1}^N \Psi_i \rho_i + \beta \sum_{i=1}^N \rho_i \log \rho_i \tag{4.1}$$

where $\boldsymbol{\rho} = \{\rho_i\}_{i=1}^N \in \overline{\mathcal{M}}$. Thus, we have the gradient flow of F on $(\mathcal{M}, g^{\boldsymbol{\Psi}})$ given by,

$$\frac{d\boldsymbol{\rho}}{dt} = -\text{grad}F(\boldsymbol{\rho}), \tag{4.2}$$

where $\text{grad}F(\boldsymbol{\rho})$ is in the tangent space $T_{\boldsymbol{\rho}}\mathcal{M}$. Thus we shall prove that equation (4.2) gives us Fokker-Planck equation I (1.8) on $\mathcal{M}$.

If the differential of F (which is in the cotangent space) is denoted by $\text{diff}F$, then (4.2) could be expressed as

$$g^{\boldsymbol{\Psi}}_{\boldsymbol{\rho}}(\frac{d\boldsymbol{\rho}}{dt}, \boldsymbol{\sigma}) = -\text{diff}F(\boldsymbol{\rho}).\boldsymbol{\sigma} \qquad \forall \boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}. \tag{4.3}$$

It is clear that

$$\text{diff}F((\rho_i)_{i=1}^N) = (\Phi_i + \beta(1 + \log \rho_i))_{i=1}^N \tag{4.4}$$

for $(\rho_i)_{i=1}^N \in \mathcal{M}$. Finally, by (4.3) and the identification (3.10) we will obtain the explicit expression of the vector field on $\mathcal{M}$.

Now we are ready to show our first main result.

Theorem 4.1. *Given a graph $G = (V, E)$ with its vertex set $V = \{a_1, a_2, \cdots, a_N\}$, edge set E, a potential $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ on V and a constant $\beta \geq 0$, let the neighborhood set of a vertex a_i be*

$$N(i) = \{j \in \{1, 2, \cdots, N\} | \{a_i, a_j\} \in E\},$$

then

(1) *The gradient flow of free energy F,*

$$F(\boldsymbol{\rho}) = \sum_{i=1}^N \Psi_i \rho_i + \beta \sum_{i=1}^N \rho_i \log \rho_i$$

on the Riemannian manifold $(\mathcal{M}, g^{\boldsymbol{\Psi}})$ of probability densities $\boldsymbol{\rho}$ on V is

$$\begin{aligned} \frac{d\rho_i}{dt} &= \sum_{j \in N(i), \Psi_j > \Psi_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\, \rho_j \\ &+ \sum_{j \in N(i), \Psi_j < \Psi_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\, \rho_i \\ &+ \sum_{j \in N(i), \Psi_j = \Psi_i} \beta(\rho_j - \rho_i) \end{aligned}$$

for $i = 1, 2, \cdots, N$, which is Fokker-Planck equation I (1.8).

(2) *Gibbs distribution $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$ given by*

$$\rho_i^* = \frac{1}{K} e^{-\Psi_i/\beta} \text{ with } K = \sum_{i=1}^N e^{-\Psi_i/\beta} \tag{4.5}$$

is the unique stationary distribution of equation (1.8) in $\mathcal{M}$. Furthermore, the free energy F attains its global minimum at Gibbs distribution.

(3) *There exists a unique solution*

$$\boldsymbol{\rho}(t) : [0, \infty) \to \mathcal{M}$$

of equation (1.8) *with initial value* $\boldsymbol{\rho}^0 \in \mathcal{M}$*, and* $\boldsymbol{\rho}(t)$ *satisfies:*
(a) *the free energy* $F(\boldsymbol{\rho}(t))$ *decreases as time* t *increases,*
(b) $\boldsymbol{\rho}(t) \to \boldsymbol{\rho}^*$ *under the Euclidean metric of* $\mathbb{R}^N$ *as* $t \to +\infty$*.*

Remark 4.2. Given $\boldsymbol{\rho}^0 \in \overline{\mathcal{M}}$ and a continuous function

$$\boldsymbol{\rho}(t) : [0, c) \to \overline{\mathcal{M}}$$

for some $0 < c \leq +\infty$, we call such a function a generalized solution of equation (1.8) with initial value $\boldsymbol{\rho}^0$, if $\boldsymbol{\rho}(0) = \boldsymbol{\rho}^0$ and $\boldsymbol{\rho}(t) \in \mathcal{M}$ satisfies equation (1.8) for $t \in (0, c)$. In Appendix B, we will give an example of a graph G and a free energy to show that there may not exist a generalized solution to (1.8) for some $\boldsymbol{\rho}^0 \in \partial\mathcal{M} := \overline{\mathcal{M}} \setminus \mathcal{M}$. We also note that the equation (1.8) is not well defined on the boundary $\partial\mathcal{M} := \overline{\mathcal{M}} \setminus \mathcal{M}$.

As a direct consequence, we have the following result.

Corollary 4.3. *Given the graph* $G = (V, E)$ *with* $V = \{a_1, a_2, \cdots, a_N\}$ *and potential* $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ *on* V*, we have*

(1) *If the noise level* $\beta = 0$*, then Fokker-Planck equation I* (1.8) *for the discrete state space is*

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i), \Psi_j > \Psi_i} (\Psi_j - \Psi_i)\rho_j + \sum_{j \in N(i), \Psi_j < \Psi_i} (\Psi_j - \Psi_i)\rho_i \tag{4.6}$$

for $i = 1, 2, \cdots, N$*.*

(2) *In a special case when the potential is a constant at each vertices, this equation is the master equation:*

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} \beta(\rho_j - \rho_i) \tag{4.7}$$

for $i = 1, 2, \cdots, N$*.*

Remark 4.4. Equation (4.6) describes the time evolution of probability distribution due to the potential energy and is also the probability distribution of a time homogeneous Markov process on the graph G. The master equation is a first order differential equation that describes the time evolution of the probability distribution at every vertex in the discrete states space. Its entropy increases along with the master equation. In some sense, Fokker-Planck equation I (1.8) is a generalization of master equation. We refer to [4] for more details on master equation.

Proof of Theorem 4.1. 1). We know that the gradient flow of free energy F on $(\mathcal{M}, g^{\boldsymbol{\Psi}})$ is given by equation (4.3),

$$g^{\boldsymbol{\Psi}}_{\boldsymbol{\rho}}(\frac{d\boldsymbol{\rho}}{dt}, \boldsymbol{\sigma}) = -\mathrm{diff}F(\boldsymbol{\rho}).\boldsymbol{\sigma} \quad \forall \boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}.$$

The left hand side of equation (4.3) is

$$g^{\mathbf{\Psi}}_{\boldsymbol{\rho}}(\frac{d\boldsymbol{\rho}}{dt},\boldsymbol{\sigma}) = \sum_{i=1}^{N} \frac{d\rho_i}{dt} p_i \tag{4.8}$$

where $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \simeq [(p_i)_{i=1}^N]$ (see identification (3.10)). By (4.4), the right hand side of equation (4.3) is

$$-\mathrm{diff}F(\rho).\boldsymbol{\sigma} = -\sum_{i=1}^{N}(\Psi_i + \beta(1+\log\rho_i))\sigma_i. \tag{4.9}$$

Using the identification (3.10), we have

$$\begin{aligned}
&\sum_{i=1}^{N}(\Psi_i + \beta(1+\log\rho_i))\sigma_i \\
=\ &\sum_{i=1}^{N}(\Psi_i + \beta\log\rho_i)\sigma_i \\
=\ &\sum_{i=1}^{N}(\Psi_i + \beta\log\rho_i)(\sum_{j\in N(i),\Psi_j>\Psi_i}(p_i-p_j)\rho_j + \sum_{j\in N(i),\Psi_j<\Psi_i}(p_i-p_j)\rho_i \\
&+\sum_{j\in N(i),\Psi_j=\Psi_i}(p_i-p_j)\frac{\rho_i-\rho_j}{\log\rho_i-\log\rho_j}) \\
=\ &\sum_{\{a_i,a_j\}\in E,\Psi_i<\Psi_j}\{(\Psi_i-\Psi_j)+\beta(\log\rho_i-\log\rho_j)\}\rho_j(p_i-p_j) \\
&+\beta\sum_{\{a_i,a_j\}\in E,\Psi_i=\Psi_j}(\rho_i-\rho_j)(p_i-p_j) \\
=\ &\sum_{i=1}^{N}\{\sum_{j\in N(i),\Psi_j>\Psi_i}((\Psi_i-\Psi_j)\rho_j+\beta(\log\rho_i-\log\rho_j)\rho_j) \\
&+\sum_{j\in N(i),\Psi_j<\Psi_i}((\Psi_i-\Psi_j)\rho_i+\beta(\log\rho_i-\log\rho_j)\rho_i) \\
&+\beta\sum_{j\in N(i),\Psi_j=\Psi_i}(\rho_i-\rho_j)\}p_i
\end{aligned}$$

This gives

$$
\begin{aligned}
&\sum_{i=1}^{N}(\Psi_i+\beta(1+\log\rho_i))\sigma_i\\
=&\sum_{i=1}^{N}\{\sum_{j\in N(i),\Psi_j>\Psi_i}((\Psi_i-\Psi_j)\rho_j+\beta(\log\rho_i-\log\rho_j)\rho_j)\\
&+\sum_{j\in N(i),\Psi_j<\Psi_i}((\Psi_i-\Psi_j)\rho_i+\beta(\log\rho_i-\log\rho_j)\rho_i)\\
&+\beta\sum_{j\in N(i),\Psi_j=\Psi_i}(\rho_i-\rho_j)\}p_i.
\end{aligned}
\tag{4.10}
$$

Combing this equation with equations (4.3), (4.8) and (4.9), we have

$$
\begin{aligned}
\sum_{i=1}^{N}\frac{d\rho_i}{dt}p_i=&\sum_{i=1}^{N}\{\sum_{j\in N(i),\Psi_j>\Psi_i}((\Psi_j-\Psi_i)\rho_j+\beta(\log\rho_j-\log\rho_i)\rho_j)\\
&+\sum_{j\in N(i),\Psi_j<\Psi_i}((\Psi_j-\Psi_i)\rho_i+\beta(\log\rho_j-\log\rho_i)\rho_i)\\
&+\beta\sum_{j\in N(i),\Psi_j=\Psi_i}(\rho_j-\rho_i)\}p_i.
\end{aligned}
$$

Since the above equality stands for any $(p_i)_{i=1}^N\in\mathbb{R}^N$, we have

$$
\begin{aligned}
\frac{d\rho_i}{dt}=&\sum_{j\in N(i),\Psi_j>\Psi_i}((\Psi_j-\Psi_i)\rho_j+\beta(\log\rho_j-\log\rho_i)\rho_j)\\
&+\sum_{j\in N(i),\Psi_j<\Psi_i}((\Psi_j-\Psi_i)\rho_i+\beta(\log\rho_j-\log\rho_i)\rho_i)\\
&+\sum_{j\in N(i),\Psi_j=\Psi_i}\beta(\rho_j-\rho_i)
\end{aligned}
$$

for $i=1,2,\cdots,N$, which is exactly Fokker-Planck equation I (1.8). This completes the proof of (1).

2). It is well known that F attains its minimum at Gibbs density. By a simple computation, it is easy to see that Gibbs distribution (4.5) is a stationary distribution of equation (1.8) in $\mathcal{M}$.

Let $\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathcal{M}$ be a stationary distribution of equation (1.8). Then

$$\begin{aligned} 0 &= \sum_{j\in N(i),\Psi_j>\Psi_i} ((\Psi_i-\Psi_j)\rho_j + \beta(\log\rho_i - \log\rho_j)\rho_j) \\ &\quad + \sum_{j\in N(i),\Psi_j<\Psi_i} ((\Psi_i-\Psi_j)\rho_i + \beta(\log\rho_i - \log\rho_j)\rho_i) \\ &\quad + \sum_{j\in N(i),\Psi_j=\Psi_i} \beta(\rho_i-\rho_j), \end{aligned}$$

for any $i \in \{1,2,\cdots,N\}$. For $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \in T_\rho\mathcal{M}$, we let $\boldsymbol{\sigma} \simeq [(p_i)_{i=1}^N]$ for some $(p_i)_{i=1}^N \in \mathbb{R}^N$. Then by equation (4.10), we have

$$\begin{aligned} &\sum_{i=1}^N (\Psi_i + \beta(1+\log\rho_i))\sigma_i \\ &= \sum_{i=1}^N \{ \sum_{j\in N(i),\Psi_j>\Psi_i} ((\Psi_i-\Psi_j)\rho_j + \beta(\log\rho_i - \log\rho_j)\rho_j) \\ &\quad + \sum_{j\in N(i),\Psi_j<\Psi_i} ((\Psi_i-\Psi_j)\rho_i + \beta(\log\rho_i - \log\rho_j)\rho_i) \\ &\quad + \sum_{j\in N(i),\Psi_j=\Psi_i} \beta(\rho_i-\rho_j)\} p_i \\ &= 0. \end{aligned}$$

Thus gives

$$\sum_i (\Psi_i + \beta(1+\log\rho_i))\sigma_i = 0 \tag{4.11}$$

for any $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \in T_{\boldsymbol{\rho}}\mathcal{M}$. We note that for any $(\sigma_i)_{i=1}^{N-1} \in \mathbb{R}^{N-1}$, if we take

$$\sigma_N = -\sum_{i=1}^{N-1} \sigma_i$$

then $(\sigma_i)_{i=1}^N \in T_{\boldsymbol{\rho}}\mathcal{M}$. Thus by (4.11), one has

$$\sum_{i=1}^{N-1} \{(\Psi_i + \beta(1+\log\rho_i)) - (\Psi_N + \beta(1+\log\rho_N))\}\sigma_i = 0$$

for any $(\sigma_i)_{i=1}^{N-1} \in \mathbb{R}^{N-1}$. This implies

$$(\Psi_i + \beta\log\rho_i) - (\Psi_N + \beta\log\rho_N) = 0,$$

which is

$$\rho_i = e^{\frac{\Psi_N - \Psi_i}{\beta}} \rho_N$$

for $i = 1,2,\cdots,N-1$.

Combing this fact with $\sum_{i=1}^N \rho_i = 1$, we have $\rho_i = \frac{1}{K} e^{-\Psi_i/\beta} = \rho_i^*$ for $i = 1,2,\cdots,N$, where $K = \sum_{i=1}^N e^{-\frac{\Psi_i}{\beta}}$. This completes the proof of (2).

3). A continuous function

$$\boldsymbol{\rho}(t) : [0, c) \to \mathcal{M}$$

for some $0 < c \leq +\infty$ is a solution of equation (1.8) with initial value $\boldsymbol{\rho}^0 \in \mathcal{M}$, if $\boldsymbol{\rho}(0) = \boldsymbol{\rho}^0$ and satisfies equation (1.8) for $t \in [0, c)$. For any $\boldsymbol{\rho}^0 \in \mathcal{M}$, there exists a maximal interval of existence $[0, c(\boldsymbol{\rho}^0))$ and $0 < c(\boldsymbol{\rho}^0)) \leq +\infty$. Next we will show that for any $\boldsymbol{\rho}^0$, $c(\boldsymbol{\rho}^0) = +\infty$. In fact, this follows from the claim,

Claim: Given $\boldsymbol{\rho}^0 \in \mathcal{M}$, there exists a compact subset B of $\mathcal{M}$ with respect to Euclidean metric such that $\boldsymbol{\rho}^0 \in \text{int}(B)$, where $\text{int}(B)$ is the interior of B in $\mathcal{M}$ with respect to Euclidean metric. For any $\boldsymbol{\rho} \in B$, if

$$\boldsymbol{\rho}(t) : [0, c(\boldsymbol{\rho})) \to \mathcal{M}$$

is the solution of the equation (1.8) with initial value $\boldsymbol{\rho}$ on its maximal interval of existence, then $c(\boldsymbol{\rho}) = +\infty$ and $\boldsymbol{\rho}(t) \in \text{int}(B)$ for $t > 0$.

Proof of Claim. Let $\boldsymbol{\rho}^0 = (\rho_i^0)_{i=1}^N \in \mathcal{M}$ be fixed. Firstly, we construct a compact subset B of $\mathcal{M}$ with respect to Euclidean metric such that $\boldsymbol{\rho}^0 \in \text{int}(B)$. Then it is sufficient to show for any $\boldsymbol{\rho} \in B$, the solution $\boldsymbol{\rho}(t)$ through $\boldsymbol{\rho}$ remains in $\text{int}(B)$ for small $t > 0$. Let us denote

$$M = \max\{e^{2|\Psi_i|} : i = 1, 2, \cdots, N\},$$

$$\epsilon_0 = 1,$$

and

$$\epsilon_1 = \frac{1}{2} \min\{\frac{\epsilon_0}{(1 + (2M)^{\frac{1}{\beta}})}, \min\{\rho_i^0 : i = 1, \cdots, N\}\}.$$

For $\ell = 2, 3, \cdots, N-1$, we let

$$\epsilon_\ell = \frac{\epsilon_{\ell-1}}{1 + (2M)^{\frac{1}{\beta}}}.$$

We define

$$B = \{\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathcal{M} : \sum_{r=1}^{\ell} \rho_{i_r} \leq 1 - \epsilon_\ell \text{ where } \ell \in \{1, \cdots, N-1\},$$
$$1 \leq i_1 < \cdots < i_\ell \leq N\}.$$

Then B is a compact subset of $\mathcal{M}$ with respect to Euclidean metric,

$$\text{int}(B) = \{\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathcal{M} : \sum_{r=1}^{\ell} \rho_{i_r} < 1 - \epsilon_\ell, \text{ where } \ell \in \{1, \cdots, N-1\},$$
$$1 \leq i_1 < \cdots < i_\ell \leq N\}.$$

and $\boldsymbol{\rho}^0 \in \text{int}(B)$.

Let $\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in B$ and $\boldsymbol{\rho}(t) : [0, c(\boldsymbol{\rho})) \to \mathcal{M}$ be the solution to the equation (1.8) with initial value $\boldsymbol{\rho}$ on its maximal interval of existence. In order to show

$\boldsymbol{\rho}(t) \in \text{int}(B)$ for small $t > 0$, it is sufficient to show that for any $\ell \in \{1, 2, \cdots, N-1\}$ and $1 \le i_1 < i_2 < \cdots i_\ell \le N$, one has

$$\sum_{r=1}^{\ell} \rho_{i_r}(t) < 1 - \epsilon_\ell$$

for sufficiently small $t > 0$.

Given $\ell \in \{1, 2, \cdots, N-1\}$ and $1 \le i_1 < i_2 < \cdots i_\ell \le N$, since $\boldsymbol{\rho} \in B$, we have

$$\sum_{r=1}^{\ell} \rho_{i_r} \le 1 - \epsilon_\ell.$$

Then there are two cases. The first one is

$$\sum_{r=1}^{\ell} \rho_{i_r} < 1 - \epsilon_\ell.$$

It is clear that

$$\sum_{r=1}^{\ell} \rho_{i_r}(t) < 1 - \epsilon_\ell$$

for enough small $t > 0$ by continuity.

The second case is

$$\sum_{r=1}^{\ell} \rho_{i_r} = 1 - \epsilon_\ell.$$

Let $A = \{i_1, i_2, \cdots, i_\ell\}$ and $A^c = \{1, 2, \cdots, N\} \setminus A$. Then for any $j \in A^c$,

$$\rho_j \le 1 - (\sum_{r=1}^{\ell} \rho_{i_r}) = \epsilon_\ell. \tag{4.12}$$

Since $\boldsymbol{\rho} \in B$, we have

$$\sum_{j=1}^{\ell-1} \rho_{s_j} \le 1 - \epsilon_{\ell-1},$$

for any $1 \le s_1 < s_2 < \cdots < s_{\ell-1} \le N$. Hence for each $i \in A$,

$$\rho_i \ge 1 - \epsilon_\ell - (1 - \epsilon_{\ell-1}) = \epsilon_{\ell-1} - \epsilon_\ell. \tag{4.13}$$

Combing equations (4.12),(4.13) and the fact

$$\epsilon_\ell \le \frac{\epsilon_{\ell-1}}{1 + (2M)^{\frac{1}{\beta}}},$$

one has, for any $i \in A, j \in A^c$,

$$\Psi_j - \Psi_i + \beta(\log \rho_j - \log \rho_i) \le \Psi_j - \Psi_i + \beta(\log \epsilon_\ell - \log(\epsilon_{\ell-1} - \epsilon_\ell)) \le -\log 2. \tag{4.14}$$

For $\{a_i, a_j\} \in E$, we set

$$C(\{a_i,a_j\}) = \begin{cases} \rho_j & \text{if } \Psi_i < \Psi_j \\ \rho_i & \text{if } \Psi_i > \Psi_j \\ \frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} & \text{if } \Psi_i = \Psi_j \end{cases}. \tag{4.15}$$

Clearly, $C(\{a_i, a_j\}) > 0$ for $\{a_i, a_j\} \in E$. Since the graph G is connected, there exists $i_* \in A, j_* \in A^c$ such that $\{a_{i_*}, a_{j_*}\} \in E$. Thus

$$\sum_{i\in A, j\in A^c, \{a_i,a_j\}\in E} C(\{a_i,a_j\}) \geq C(\{a_{i_*},a_{j_*}\}) > 0. \tag{4.16}$$

Now by (4.14) and (4.16), one has

$$\begin{aligned}
\frac{\mathrm{d}}{dt}\sum_{r=1}^{\ell}\rho_{i_r}(t)\mid_{t=0} &= \sum_{i\in A}\left(\sum_{j\in N(i)} C(\{a_i,a_j\})\left(\Psi_j - \Psi_i + \beta(\log\rho_j - \log\rho_i)\right)\right) \\
&= \sum_{i\in A}\{\sum_{j\in A\cap N(i)} C(\{a_i,a_j\})\left(\Psi_j - \Psi_i + \beta(\log\rho_j - \log\rho_i)\right) + \\
&\quad \sum_{j\in A^c\cap N(i)} C(\{a_i,a_j\})\left(\Psi_j - \Psi_i + \beta(\log\rho_j - \log\rho_i)\right)\} \\
&= \sum_{i\in A}\left(\sum_{j\in A^c\cap N(i)} C(\{a_i,a_j\})\left(\Psi_j - \Psi_i + \beta(\log\rho_j - \log\rho_i)\right)\right) \\
&\leq \sum_{i\in A}\left(\sum_{j\in A^c\cap N(i)} -C(\{a_i,a_j\})\log 2\right) \\
&= -\log 2\left(\sum_{i\in A, j\in A^c, \{a_i,a_j\}\in E} C(\{a_i,a_j\})\right) \\
&\leq -C(\{a_{i_*},a_{j_*}\})\log 2 < 0.
\end{aligned}$$

Combing this with the fact

$$\sum_{r=1}^{\ell}\rho_{i_r} = 1 - \epsilon_\ell,$$

it is clear that

$$\sum_{r=1}^{\ell}\rho_{i_r}(t) < 1 - \epsilon_\ell$$

for sufficiently small $t > 0$. This completes the proof of Claim. □

Given $\boldsymbol{\rho}^0 \in \mathcal{M}$, by the above Claim, there exists a unique solution

$$\boldsymbol{\rho}(t) : [0,\infty) \to \mathcal{M}$$

to equation (1.8) with initial value $\boldsymbol{\rho}^0$, and we can find a compact subset B of $\mathcal{M}$ with respect to Euclid metric such that $\{\boldsymbol{\rho}(t) : t \in [0, +\infty)\} \subset B$. For $t \in (0, +\infty)$,

$$\begin{aligned}\frac{dF(\boldsymbol{\rho}(t))}{dt} &= \mathrm{diff}F(\boldsymbol{\rho}(t)).\frac{d\boldsymbol{\rho}}{dt}(t)\\ &= -g^{\boldsymbol{\Psi}}_{\rho(t)}(\frac{d\boldsymbol{\rho}}{dt}(t), \frac{d\boldsymbol{\rho}}{dt}(t))\\ &\leq 0\end{aligned}$$

and thus

$$\frac{dF(\boldsymbol{\rho}(t))}{dt} = 0$$

if and only if

$$\frac{d\boldsymbol{\rho}(t)}{dt} = 0.$$

This is equivalent to $\boldsymbol{\rho}(t) = (\boldsymbol{\rho}^*_i)_{i=1}^N$ by (2). This implies that the free energy $F(\boldsymbol{\rho}(t))$ decreases when time t increases.

Finally, we show that $\boldsymbol{\rho}(t) \to \boldsymbol{\rho}^*$ under the Euclid metric of $\mathbb{R}^N$ when $t \to +\infty$. We let

$$\omega(\boldsymbol{\rho}^0) = \{\boldsymbol{\rho} \in \mathbb{R}^N : \exists\, t_i \to +\infty \text{ such that } \lim_{i\to+\infty} \boldsymbol{\rho}(t_i) = \boldsymbol{\rho} \quad \text{Euclidean metric}\}$$

be the ω-limit set of $\boldsymbol{\rho}^0$. Clearly, $\omega(\boldsymbol{\rho}^0) \subset B$ is a compact set of $\mathbb{R}^N$ with respect to Euclidean metric.

To show $\boldsymbol{\rho}(t) \to \boldsymbol{\rho}^*$ under the Euclidean metric of $\mathbb{R}^N$ when $t \to +\infty$, it is sufficient to show that $\omega(\boldsymbol{\rho}^0) = \{\boldsymbol{\rho}^*\}$. Since $\omega(\boldsymbol{\rho}^0)$ is a compact set and the free energy F is continuous on $\overline{\mathcal{M}}$, we can find $\boldsymbol{\rho}^1 \in \omega(\boldsymbol{\rho}^0)$ such that $F(\boldsymbol{\rho}^1) = \max\{F(\boldsymbol{\rho}) : \boldsymbol{\rho} \in \omega(\boldsymbol{\rho}^0)\}$. Then there exists $t_i \to +\infty$ such that $\lim_{i\to+\infty} \boldsymbol{\rho}(t_i) = \boldsymbol{\rho}^1$ and $\lim_{i\to+\infty} \boldsymbol{\rho}(t_i - 1) = \boldsymbol{\rho}^2$ for some $\boldsymbol{\rho}^2 \in M$. If we let $\boldsymbol{\rho}^2(t)$ be the solution to equation (1.8) with initial value $\boldsymbol{\rho}^2$, then $\boldsymbol{\rho}^2(0) = \boldsymbol{\rho}^2$ and $\boldsymbol{\rho}^2(1) = \boldsymbol{\rho}(1)$. Note that

$$\begin{aligned}\frac{dF(\boldsymbol{\rho}^2(t))}{dt} &= \mathrm{diff}F(\boldsymbol{\rho}^2(t)).\frac{d\boldsymbol{\rho}^2}{dt}(t)\\ &= -g^{\boldsymbol{\Psi}}_{\boldsymbol{\rho}^2(t)}(\frac{d\boldsymbol{\rho}^2}{dt}(t), \frac{d\boldsymbol{\rho}^2}{dt}(t))\\ &\leq 0\end{aligned}$$

and thus

$$\frac{dF(\boldsymbol{\rho}^2(t))}{dt} = 0$$

if and only if

$$\frac{d\boldsymbol{\rho}^2(t)}{dt} = 0,$$

which is equivalent to $\boldsymbol{\rho}^2(t) = \boldsymbol{\rho}^*$ by (2). Hence if $\boldsymbol{\rho}^1 \neq \boldsymbol{\rho}^*$, then $F(\boldsymbol{\rho}^2) > F(\boldsymbol{\rho}^1)$, a contradiction with $F(\boldsymbol{\rho}^1) = \max\{F(\boldsymbol{\rho}) : \boldsymbol{\rho} \in \omega(\boldsymbol{\rho}^0)\}$. So $\boldsymbol{\rho}^1 = \boldsymbol{\rho}^*$. Thus,

$$\max\{F(\boldsymbol{\rho}) : \boldsymbol{\rho} \in \omega(\boldsymbol{\rho}^0)\} = F(\boldsymbol{\rho}^*).$$

It is well known that $\boldsymbol{\rho}^*$ is the unique minimal value point of F. This implies $\omega(\boldsymbol{\rho}^0) = \{\boldsymbol{\rho}^*\}$. This completes the proof of (3). □

There are many reasons why we consider Fokker-Planck equation I (1.8) on the manifold $\mathcal{M}$ instead of its closure. One of the main reasons is that $\overline{\mathcal{M}}$ is a manifold with boundary and its tangent space is only well defined in its interior. Another reason is that the free energy is not differentiable when $\rho_i = 0$ for some i. It is also not clear how the Riemannian metric $\mathrm{d}_{\boldsymbol{\Psi}}$ on $\mathcal{M}$ can be extended to $\overline{\mathcal{M}}$. Moreover, even if the distance is well defined on $\overline{\mathcal{M}}$, there may not be a solution to the equation (1.8) with initial value on the boundary ∂M (see Appendix B, Example B.1).

Theorem 4.1 (3) guarantees the solution of (1.8) can never attain the boundary $\partial\mathcal{M}$ if the initial value is in $\mathcal{M}$. In practice, we still need an equation describe the transient process if the initial value is on the boundary. Thus, we may need some procedure to "kick" the solution into $\mathcal{M}$ in order for one to use the equation (1.8).

Let us go back to the discrete free energy

$$F(\boldsymbol{\rho}) = \sum_{i=1}^{N} \Psi_i \rho_i + \beta \sum_{i=1}^{N} \rho_i \log \rho_i.$$

The gradient of free energy is

$$\nabla F(\boldsymbol{\rho}) = (\Psi_1 + \beta + \beta \log \rho_1, \cdots, \Psi_N + \beta + \beta \log \rho_N).$$

There is a natural question on what will happen if the probability distribution $\boldsymbol{\rho}$ is on the boundary $\partial\mathcal{M} := \overline{\mathcal{M}} \setminus \mathcal{M}$ of $\mathcal{M}$? To answer this, we give some definitions first.

Given $\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \overline{\mathcal{M}}$, in order to consider the set of probability densities with at least one of their components is zero, we let

$$\mathcal{Z}(\boldsymbol{\rho}) = \{i \in \{1, 2, \cdots, N\} | \rho_i = 0\}.$$

We also let $\overline{\mathcal{Z}}(\boldsymbol{\rho})$ denote the "neighborhood" of $\mathcal{Z}(\boldsymbol{\rho})$ as

$$\overline{\mathcal{Z}}(\boldsymbol{\rho}) = \{i \in \{1, 2, \cdots, N\} | \text{ there exists } \{a_i, a_j\} \in E \text{ such that } \rho_i = 0 \text{ or } \rho_j = 0\}.$$

When $\boldsymbol{\rho}$ is on the boundary of $\mathcal{M}$,

$$\Psi_i + \beta + \beta \log \rho_i = -\infty$$

for $i \in \mathcal{Z}(\boldsymbol{\rho})$. In this case, we may ignore all the potentials, and all these edges connecting two nonzero terms. Thus when $\boldsymbol{\rho}$ is on the boundary of $\mathcal{M}$, we consider the following equation

$$\frac{d\rho_i}{dt} = \begin{cases} \sum\limits_{j \in N(i)} (\rho_j - \rho_i) & \text{if } i \in \mathcal{Z}(\boldsymbol{\rho}) \\ \sum\limits_{j \in N(i) \cap \mathcal{Z}(\boldsymbol{\rho})} (\rho_j - \rho_i) & \text{if } i \in \overline{\mathcal{Z}}(\boldsymbol{\rho}) \setminus \mathcal{Z}(\boldsymbol{\rho}) \\ 0 & \text{if } i \in \{1, 2, \cdots, N\} \setminus \overline{\mathcal{Z}}(\boldsymbol{\rho}) \end{cases} \tag{4.17}$$

as Fokker-Planck equation.

Remark 4.5. The above computation implies the following: 1) Equation (4.17) has a unique solution; 2) In an infinitesimal time interval, the solution of (4.17) goes into $\mathcal{M}$ if its initial value $\boldsymbol{\rho}_0 \in \partial\mathcal{M}$; 3) The free energy decreases along the solution of (4.17) in a sufficient small time interval.

5. Fokker-Planck Equation II

In this section, we show how Fokker-Planck equation II (1.9) can be derived from a stochastic process on the graph G. The related geometric properties are also discussed.

Given a graph $G = (V, E)$ with $V = \{a_1, a_2, \cdots, a_N\}$, it is known that Fokker-Planck equation describes the time evolution of probability density function of a stochastic process that comes from the gradient flow subject to a white noise perturbation. For the graph G, we consider a time homogeneous Markov process $X(t)$ generated by the potential function $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ on V as a "gradient flow". More precisely, the process is the one generated by the potential function $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ and is a time-homogeneous Markov process $X(t), t \geq 0$ that takes values on the set $V = \{a_1, a_2, \cdots, a_N\}$. If the process starts at a vertex or state a_i at time t, then the transition probability to the state a_j at time $t + h$ is given by

$$\begin{aligned} & Pr(X(t+h) = a_j | X(t) = a_i) \\ = & \begin{cases} (\Psi_i - \Psi_j)h + o(h) & \text{if} \quad j \in N(i), \Psi_j < \Psi_i \\ 1 - \sum\limits_{k \in N(i), \Psi_k < \Psi_i} (\Psi_i - \Psi_k)h + o(h) & \text{if} \quad j = i \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

where $o(h)$ represents the usual quantity which goes to zero faster than h. Hence, over a sufficiently small interval of time, the probability of a particular transition (between different states) is roughly proportional to the duration of the time interval.

The generating matrix $Q = [Q_{ij}]_{N \times N}$ for the Markov process is defined by: if $i \neq j$,

$$Q_{ij} = \begin{cases} \Psi_i - \Psi_j & \text{if } \{a_i, a_j\} \in E, \Psi_j < \Psi_i \\ 0 & \text{otherwise,} \end{cases}$$

which is the transition rate from i to j. Otherwise $Q_{ii} = -\sum_{j \neq i} Q_{ij}$.

Let $\boldsymbol{\rho}(t) = (\rho_i(t))_{i=1}^N$, where $\rho_i(t) = Pr(X(t) = a_i)$ for $i = 1, 2, \cdots, N$, the time evolution of probability distribution $\boldsymbol{\rho}(t)$ is given by a system of first-order ordinary differential equations (forward Kolmogorov Equation)

$$\begin{cases} \frac{\partial \boldsymbol{\rho}}{\partial t}(t) = \boldsymbol{\rho}(t) Q \\ \boldsymbol{\rho}(0) = \{\rho_i^0\}_{i=1}^N, \end{cases}$$

which is,

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i), \Psi_j > \Psi_i} (\Psi_j - \Psi_i)\rho_j + \sum_{j \in N(i), \Psi_j < \Psi_i} (\Psi_j - \Psi_i)\rho_i \tag{5.1}$$

for $i = 1, 2, \cdots, N$.

We remark that this is just an analog of a gradient flow in a continuous state space:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\nabla\Psi(x), x \in \mathbb{R}^N \tag{5.2}$$

and the corresponding degenerate Fokker-Planck equation (without any noise)

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\nabla\Psi\rho). \tag{5.3}$$

We call the Markov process $X(t)$ the gradient Markov process generated by discrete potential function $\mathbf{\Psi}$ on graph G.

The equation (5.1) could also be considered as a discretization of equation (5.3), with the upwind scheme [25]. The upwind scheme for solving Fokker-Planck equation numerically is very natural, not only by the numerical reason, but also because it is consistent with the forward Kolmogorov equation (5.1).

We recall that on the continuous state space $\mathbb{R}^N$, Fokker-Planck equation describes the time evolution of a gradient dynamical system perturbed by white noise. More precisely, when we add white noise $\sqrt{2\beta}dW_t$ into equation (5.2), we obtain the stochastic differential equation

$$dx = -\nabla\Psi(x)dt + \sqrt{2\beta}dW_t,\ x \in \mathbb{R}^N. \tag{5.4}$$

The time evolution of probability density function of the stochastic differential equation (5.4) is

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\nabla\Psi\rho) + \beta\Delta\rho. \tag{5.5}$$

In other words, Fokker-planck equation (5.5) can be obtained by adding white noise to the degenerate Fokker-Planck equation (5.3) As an analog, the Fokker-Planck equation on a discrete space can also be obtained by adding a white noise with strengthen $\sqrt{2\beta}$ into the Markov process or equation (5.1). Then, it is natural to ask what is the white noise to a Markov process or equation (5.1)?

To answer the question, we look for hints in the continuous state space. We rewrite Fokker-Planck equation (5.5) in the following fashion:

$$\begin{aligned} \rho_t &= \nabla \cdot (\nabla\Psi\rho) + \beta\Delta\rho \\ &= \nabla \cdot [\nabla(\Psi + \beta\log\rho)\rho]. \end{aligned}$$

and the corresponding free energy:

$$\begin{aligned} F &= \int_{\mathbb{R}^N} \Psi\rho + \beta\int_{\mathbb{R}^N} \rho\log\rho dx \\ &= \int_{\mathbb{R}^N} (\Psi + \beta\log\rho)\rho dx \end{aligned}$$

By comparing the above equalities, we conclude that if white noise $\sqrt{2\beta}dW_t$ is add to the system on $\mathbb{R}^N$, we have a new potential

$$\bar{\Psi}(x,t) = \Psi(x) + \beta\log\rho(x,t).$$

As an analogue, if white noise is added to the time homogeneous Markov process $X(t)$, we have a new potential defined as follows:

$$\bar{\Psi}_i(t) = \Psi_i + \beta \log \rho_i(t)$$

for $i = 1, 2, \cdots, N$.

Remark 5.1. On continuous state space, the new potential function $\Psi + \beta \log \rho(x,t)$ is called Onsager's potential, and its derivative is called Onsager thermodynamic flux [6, 33].

We now have a new Markov process $X_\beta(t)$ which is time inhomogeneous, and could be considered as a "white noise" perturbation from the original Markov process $X(t)$. Given that the process starts in a state a_i at time t, then the transition to state a_j at time $t+h$ with probability is given by

$$\begin{aligned} & Pr(X_\beta(t+h) = a_j | X_\beta(t) = a_i) \\ = & \begin{cases} (\bar{\Psi}_i(t) - \bar{\Psi}_j(t))h + o(h), & \text{if } j \in N(i), \bar{\Psi}_j(t) < \bar{\Psi}_i(t) \\ 1 - \sum\limits_{k \in N(i), \bar{\Psi}_k(t) < \bar{\Psi}_i(t)} (\bar{\Psi}_i(t) - \bar{\Psi}_k(t))h + o(h) & \text{if } j = i \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

where $\bar{\Psi}_k(t) = \Psi_k + \beta \log \rho_k(t)$ and $\rho_k(t) = Pr(X_\beta(t) = a_k)$ for $k \in \{1, 2, \cdots, N\}$. The time evolution of probability distribution $(\rho_i(t))_{i=1}^N$ of $X_\beta(t)$ satisfies the following Kolmogorov forward equation:

$$\begin{aligned} \frac{d\rho_i}{dt} = & \sum_{j \in N(i), \bar{\Psi}_j > \bar{\Psi}_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_j \\ & + \sum_{j \in N(i), \bar{\Psi}_j < \bar{\Psi}_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_i, \end{aligned}$$

which is exactly Fokker-Planck equation II (1.9).

For a given potential $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ on V and $\beta \geq 0$, we have obtained the the inner product $g^{\bar{\boldsymbol{\Psi}}}$ with distance $d_{\bar{\boldsymbol{\Psi}}}$ in section 3. The free energy F on the space $\mathcal{M}$ is given by:

$$F(\boldsymbol{\rho}) = \sum_{i=1}^N \Psi_i \rho_i + \beta \sum_{i=1}^N \rho_i \log \rho_i$$

where $\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathcal{M}$.

Since $\boldsymbol{\rho} \in \mathcal{M} \mapsto g_{\boldsymbol{\rho}}^{\bar{\boldsymbol{\Psi}}}$ may not be continuous, we may have no gradient flow of F on $(\mathcal{M}, g^{\bar{\boldsymbol{\Psi}}})$. However, we may consider a *a generalized gradient flow* of F on $(\mathcal{M}, g^{\bar{\boldsymbol{\Psi}}})$ of probability densities $\boldsymbol{\rho}$ on V because of the special relationship between $\mathcal{M}$ and its tangent spaces. Note that the derivative $\frac{d\boldsymbol{\rho}(t)}{dt} \in T_{\boldsymbol{\rho}(t)}\mathcal{M}$ is the same as the one computed using Euclidean metric. Thus, we can consider the the following equation as our generalized gradient flow:

$$g_{\boldsymbol{\rho}}^{\bar{\boldsymbol{\Psi}}}(\frac{d\boldsymbol{\rho}}{dt}(t), \boldsymbol{\sigma}) = -\mathrm{diff} F(\boldsymbol{\rho}(t)).\boldsymbol{\sigma} \qquad \forall \boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}, \tag{5.6}$$

where $\mathrm{diff}F$ is the differential of F (see (4.4)). This also implies:

$$\frac{dF(\boldsymbol{\rho}(t))}{dt} = \mathrm{diff}F(\boldsymbol{\rho}(t)).\frac{d\boldsymbol{\rho}(t)}{dt} = -g^{\bar{\boldsymbol{\Psi}}}_{\boldsymbol{\rho}(t)}(\frac{d\boldsymbol{\rho}(t)}{dt}, \frac{d\boldsymbol{\rho}(t)}{dt}). \tag{5.7}$$

We have the following theorem:

Theorem 5.2. *Given the graph $G = (V, E)$ with $V = \{a_1, a_2, \cdots, a_N\}$, potential $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ on V and $\beta \geq 0$, we consider the gradient Markov process $X(t)$ generated by the potential $\boldsymbol{\Psi}$. Then we have*

(1) *Fokker-Planck equation II* (1.9)

$$\begin{aligned}\frac{d\rho_i}{dt} &= \sum_{j\in N(i),\bar{\Psi}_j>\bar{\Psi}_i} ((\Psi_j + \beta\log\rho_j) - (\Psi_i + \beta\log\rho_i))\rho_j \\ &+ \sum_{j\in N(i),\bar{\Psi}_j<\bar{\Psi}_i} ((\Psi_j + \beta\log\rho_j) - (\Psi_i + \beta\log\rho_i))\rho_i\end{aligned}$$

describes the time evolution of probability distribution of $X_\beta(t)$, where $X_\beta(t)$ is the time inhomogeneous Markov process perturbed by "white noise" from the original Markov process $X(t)$.

(2) *Fokker-Planck equation II* (1.9) *is the generalized gradient flow* (5.6) *of the free energy F on the metric space $(\mathcal{M}, g^{\bar{\boldsymbol{\Psi}}})$ of probability densities $\boldsymbol{\rho}$ on V.*

(3) *Gibbs distribution $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$*

$$\rho_i^* = \frac{1}{K}e^{-\Psi_i/\beta} \text{ with } K = \sum_{i=1}^N e^{-\Psi_i/\beta}.$$

is the unique stationary distribution of equation (1.9) *in $\mathcal{M}$, and the free energy F attains minimum at Gibbs distribution.*

(4) *There exists a unique solution $\boldsymbol{\rho}(t) : [0, \infty) \to \mathcal{M}$ to equation* (1.9) *with initial value $\boldsymbol{\rho}^0 \in \mathcal{M}$, and $\boldsymbol{\rho}(t)$ satisfies*
   (a) *The free energy $F(\boldsymbol{\rho}(t))$ decreases when time t increases.*
   (b) *$\boldsymbol{\rho}(t) \to \boldsymbol{\rho}^*$ under the Euclid metric of $\mathbb{R}^N$ when $t \to +\infty$.*

Proof. (1). Result in (1) comes from the discussion in the beginning of this section.

(2). A continuously differentiable function $\boldsymbol{\rho}(t) : [0, c) \to \mathcal{M}$ for some $c > 0$ or $c = +\infty$ is called a solution of equation (1.9) with initial value $\boldsymbol{\rho} \in \mathcal{M}$, if $\boldsymbol{\rho}(0) = \boldsymbol{\rho}$ and $\boldsymbol{\rho}(t)$ satisfies equation (1.9) for $t \in [0, c)$. Since the Fokker-Planck equation II (1.9) is Lipschitz continuous, by the existence and uniqueness theorem of ordinary differential equations, it is clear that for any $\boldsymbol{\rho} \in \mathcal{M}$, there exists a maximal interval $[0, c(\boldsymbol{\rho}))$ in which the solution to the equation (1.9) with initial value ρ is uniquely defined with $c(\rho) > 0$ or $c(\boldsymbol{\rho}) = +\infty$.

Let $\boldsymbol{\rho}(t) : [0, c) \to \mathcal{M}$ be a solution of (1.9). For $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \in T_{\boldsymbol{\rho}(t)}\mathcal{M}$, we take $(p_i)_{i=1}^N \in \mathbb{R}^N$ such that $\boldsymbol{\sigma} \simeq [(p_i)_{i=1}^N]$ by identification (3.12). Then by (4.4) and

identification (3.12), we have

$$\begin{aligned}
&\sum_{i=1}^{N}(\Psi_i+\beta(1+\log\rho_i))\sigma_i=\sum_i(\Psi_i+\beta\log\rho_i)\sigma_i\\
&=\sum_{i=1}^{N}(\Psi_i+\beta(1+\log\rho_i))(\sum_{j\in N(i),\bar\Psi_j>\bar\Psi_i}(p_i-p_j)\rho_j+\sum_{j\in N(i),\bar\Psi_j<\bar\Psi_i}(p_i-p_j)\rho_i\\
&\qquad+\sum_{j\in N(i),\bar\Psi_j=\bar\Psi_i}(p_i-p_j)\frac{\rho_i-\rho_j}{\log\rho_i-\log\rho_j})\\
&=\sum_{\{a_i,a_j\}\in E,\bar\Psi_i<\bar\Psi_j}\{(\Psi_i-\Psi_j)+\beta(\log\rho_i-\log\rho_j)\}\rho_j(p_i-p_j)\\
&\qquad+\beta\sum_{\{a_i,a_j\}\in E,\bar\Psi_i=\bar\Psi_j}(\bar\Psi_i-\bar\Psi_j)(p_i-p_j)\\
&=\sum_i\{\sum_{j\in N(i),\bar\Psi_j>\bar\Psi_i}((\Psi_i-\Psi_j)\rho_j+\beta(\log\rho_i-\log\rho_j)\rho_j)\\
&\qquad+\sum_{j\in N(i),\bar\Psi_j<\bar\Psi_i}((\Psi_i-\Psi_j)\rho_i+\beta(\log\rho_i-\log\rho_j)\rho_i)\}p_i
\end{aligned}$$

where $\rho_k=\rho_k(t)$ and $\bar\Psi_k=\Psi_k+\beta\log\rho_k(t)$ for $k=1,2,\cdots,N$. That is

$$\begin{aligned}
&\sum_{i=1}^{N}(\Psi_i+\beta(1+\log\rho_i))\sigma_i\\
&=\sum_{i=1}^{N}\{\sum_{j\in N(i),\bar\Psi_j>\bar\Psi_i}((\Psi_i-\Psi_j)\rho_j+\beta(\log\rho_i-\log\rho_j)\rho_j)\\
&\qquad+\sum_{j\in N(i),\bar\Psi_j<\bar\Psi_i}((\Psi_i-\Psi_j)\rho_i+\beta(\log\rho_i-\log\rho_j)\rho_i)\}p_i.
\end{aligned}\tag{5.8}$$

Combing this equation with Fokker-Planck equation II (1.9), and by identifications (3.12) and (3.14), we have

$$\begin{aligned}
&\mathrm{diff}F(\boldsymbol{\rho}(t)).\boldsymbol{\sigma}=\sum_{i=1}^{N}(\Psi_i+\beta(1+\log\rho_i))\sigma_i\\
&=\sum_{i=1}^{N}\{\sum_{j\in N(i),\bar\Psi_j>\bar\Psi_i}((\Psi_i-\Psi_j)\rho_j+\beta(\log\rho_i-\log\rho_j)\rho_j)\\
&\qquad+\sum_{j\in N(i),\bar\Psi_j<\bar\Psi_i}((\Psi_i-\Psi_j)\rho_i+\beta(\log\rho_i-\log\rho_j)\rho_i)\}p_i\\
&=-\sum_{i=1}^{N}\frac{d\rho_i}{dt}p_i=-g^{\bar{\boldsymbol{\Psi}}}_{\boldsymbol{\rho}(t)}(\frac{d\boldsymbol{\rho}(t)}{dt},\boldsymbol{\sigma}).
\end{aligned}$$

Hence (5.6) is true, and this finishes the proof of (2).

(3). Using (5.8) and replacing the identification (3.10) by identification (3.12), the proof of (3) is completely similar to the proof of (2) in Theorem 4.1.

(4). We note that Fokker-Planck equation II (1.9) is Lipschitz continuous. Using (5.7) and replacing the $C(\{a_i, a_j\})$ in (4.15) by

$$C(\{a_i, a_j\}) = \begin{cases} \rho_j & \text{if } \bar{\Psi}_i < \bar{\Psi}_j \\ \rho_i & \text{if } \bar{\Psi}_i > \bar{\Psi}_j \\ \frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} & \text{if } \bar{\Psi}_i = \bar{\Psi}_j, \end{cases} .$$

the proof of (4) is similar to the proof of (3) in Theorem 4.1. □

In above discussions, we have considered the graph $G = (V, E)$ with potential $\mathbf{\Psi} = (\Psi_i)_{i=1}^N$ on V. We have given a time-homogeneous Markov process as the gradient Markov process generated by $\mathbf{\Psi}$ on graph G, and obtained the Fokker-Planck equation II (1.9). On the other side, if the potential is not given, and we are only given a time-homogeneous Markov process, we will show that we can still define a potential function in certain cases, especially when the transition graph of Markov process is a weighted directed "gradient like" graph. Let us consider a weighted directed simple graph G without self-loop or multi-edge. If every directed path from a_i to a_j have the same total weight for any two vertices a_i and a_j, then we say the graph G is *gradient like*. In this case, a potential energy function can be defined on this graph, and it is unique up to a constant. To better explain it, we use the following example.

Example 5.3.

Given a graph $G = (V, E)$ as shown in Figure 4, which is directed and weighted. In this graph, $V = \{a_1, a_2, a_3, a_4, a_5\}$ and the number in each directed edge is the transition rate. The transition probability rate matrix Q of this Markov process

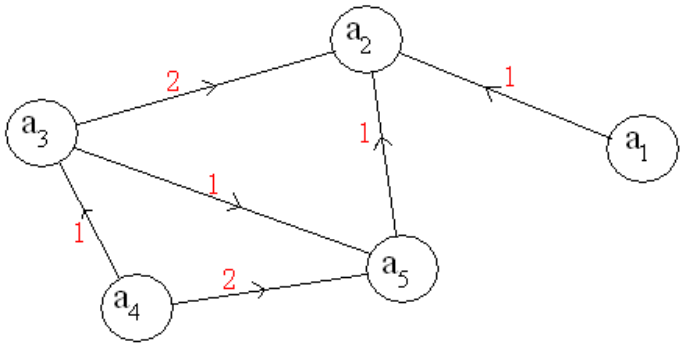


FIGURE 4

induced by Figure 4 is:

$$\mathbf{Q} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Since the underlying weighted directed graph is gradient like, we may associate the potential energy on each vertex as in Figure 5:

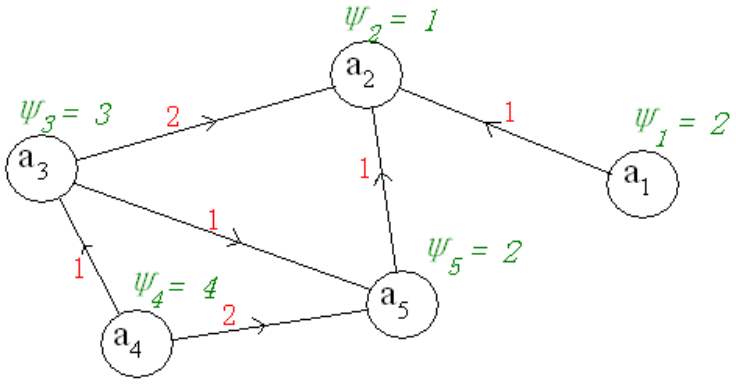


FIGURE 5

With this potential function, we can find its free energy and Fokker-Planck equation.

6. Upwind Scheme

From the discussion in previous sections, it is clear that Fokker-Planck equations (1.8) and (1.9) on a graph are not typical discretizations of Fokker-Planck equation in continuous state space. Moreover, one cannot obtain Fokker-Planck equation in discrete state space with desired properties by simply discretizing Fokker-Planck equation in continuous state space with commonly used finite difference schemes. However, it is worth to mention that Fokker-Planck equations we have derived are in fact motivated by well-known strategies in finite difference methods, in particular the upwind schemes for hyperbolic equations. For readers' convenience, we give a brief introduction to the upwind schemes for hyperbolic equations in the Appendix. In this section, we explain those connections in detail, and show that our discrete Fokker-Planck equations are actually consistent to Fokker-Planck equation in continuous state space provided that the discrete space can be refined and converges to a continuous state space.

For simplicity, let us demonstrate the connections on a 1-D lattice $G = (V, E)$ with vertex set $V = \{a_1, \cdots, a_N\}$ $(N \geq 3)$and $E = \{\{a_i, a_{i+1}\} : i = 1, 2, \cdots, N-1\}$. Since we want to show the relation between Fokker-Planck equation I and II with the finite difference scheme, we view the lattice as an equal partition points of the

interval $[0,1]$ with mesh size h. We also assume that the potential function on V does not have equal values on adjacent points. Then our discrete Fokker-Planck equation I at vertex a_i, $i \in \{2,3,\cdots,N-1\}$, is given by,

$$\begin{aligned}
\frac{d\rho_i}{dt} &= ((\Psi_{i+1}-\Psi_i)\rho_{i+1}-(\Psi_i-\Psi_{i-1})\rho_i)\\
&\quad +\beta((\log\rho_{i+1}-\log\rho_i)\rho_{i+1}-(\log\rho_i-\log\rho_{i-1})\rho_i)\\
&\quad \text{if}\quad \Psi_{i+1}>\Psi_i>\Psi_{i-1},\\
\frac{d\rho_i}{dt} &= ((\Psi_{i+1}-\Psi_i)\rho_i-(\Psi_i-\Psi_{i-1})\rho_{i-1})\\
&\quad +\beta((\log\rho_{i+1}-\log\rho_i)\rho_i-(\log\rho_i-\log\rho_{i-1})\rho_{i-1})\\
&\quad \text{if}\quad \Psi_{i+1}<\Psi_i<\Psi_{i-1},\\
\frac{d\rho_i}{dt} &= ((\Psi_{i+1}-\Psi_i)\rho_{i+1}-(\Psi_i-\Psi_{i-1})\rho_{i-1})\\
&\quad +\beta((\log\rho_{i+1}-\log\rho_i)\rho_{i+1}-(\log\rho_i-\log\rho_{i-1})\rho_{i-1}\\
&\quad \text{if}\quad \Psi_{i+1}>\Psi_i,\Psi_{i-1}<\Psi_i,\\
\frac{d\rho_i}{dt} &= ((\Psi_{i+1}-\Psi_i)\rho_i-(\Psi_i-\Psi_{i-1})\rho_i)\\
&\quad +\beta((\log\rho_{i+1}-\log\rho_i)\rho_i-(\log\rho_i-\log\rho_{i-1})\rho_i\\
&\quad \text{if}\quad \Psi_{i+1}<\Psi_i,\Psi_{i-1}<\Psi_i.
\end{aligned}$$

First, we consider the drift terms, the ones involve the potentials, on the right hand side of the equations. It is obvious that when the potential is increasing at vertex a_i, which corresponds to the first scenario as $\Psi_{i+1}>\Psi_i>\Psi_{i-1}$, the term $(\Psi_{i+1}-\Psi_i)\rho_{i+1}-(\Psi_i-\Psi_{i-1})\rho_i$ involves density values ρ_{i+1} and ρ_i, which are from the right side of position i. If one views the differences in potentials $(\Psi_i-\Psi_{i-1})$ and $(\Psi_{i+1}-\Psi_i)$, which are all positive, as the convection coefficient to determine the "wind blowing" direction, then it is from right to the left. Thus the right hand side of the equation only involves information from the upwind (higher potential) direction. Similarly, the upwind direction for the decreasing potential case with $\Psi_{i+1}<\Psi_i<\Psi_{i-1}$ is from left to right. And Fokker-Planck equation only relies on the values ρ_i and ρ_{i-1}. In the other two cases, there are no clear up wind directions and therefore central differences are used. Moreover, one can see that the evolution of ρ_i only depends on its neighboring values with higher potentials. And this is also true for general graphs.

The appearance of upwind directions in Fokker-Planck equations I and II becomes natural if we take a closer look at the working mechanism of the drift terms. If we ignore the diffusion terms by taking $\beta=0$, then the probability density ρ_i evolves according to the gradient descent direction. The consequence is that the probability is clustered on local minima of the potentials. And therefore, their corresponding density functions are combinations of Dirac Delta functions sitting on the minima. This is very similar to the shock (discontinuities in solutions) formation in nonlinear hyperbolic conservation laws, in which upwind idea is considered as a fundamental

strategy in designing shock capturing schemes. On the other hand, if one uses central differences in shock formation, numerical oscillations, which are called Gibbs' phenomenon, are inevitable. This may explain that the central difference discretization can not achieve decreasing free energy in the toy example in the previous section. For more discussion on numerical schemes for nonlinear conservation laws, readers are referred to books such as [5, 25].

7. The Parrondo's Paradox

In the last section, we demonstrate an interesting application of Fokker-Planck equation I (1.8) to explain Parrondo's paradox in game theory. Roughly speaking, the paradox states that it is possible to construct a winning strategy by playing two losing strategies alternately [19]. There exists an extensive literature on the paradox. And it has been used for many different problems such as a flashing ratchet model for the molecular motors. We refer to [1, 20, 34, 35] and references therein for more discussions on the subject. Here, we explain the paradox for the flashing ratchet model from the free energy point of view.

We begin by reviewing Parrondo's paradox in a coin toss game [18, 19, 32], in which a player wins 1 dollar if the head side faces up, and loses 1 dollar otherwise. Let us assume that we have two strategies:

(1) Game A is to toss a biased coin, the probability of winning is 0.49, and the probability of losing is 0.51. Hence the expectation is -0.02. Obviously, this is a losing strategy.
(2) Game B is more complex. If the current capital is a multiple of 3, one toss a biased coin with winning probability 0.09 and losing probability 0.91. If the current capital is not a multiple of 3, then the playing toss another biased coin with winning probability 0.74, losing probability 0.26. The expectation of Game B is -0.0174, so it is also a losing strategy too.

However, if one plays game A and B randomly, or iteratively plays Game A twice then plays Game B twice, then it may form a winning strategy. Figure 6 shows the expectations of capitals by playing the games 200 times. The expectations are computed by taking the average of 1,000,000 trials. In Figure 6, the red line is the expectation after each toss when playing strategy A only; the blue line is the expectation when playing strategy B only. Clearly, both are losing strategies. The black line shows the expectation of the capital after each toss if playing A and B alternatively as "AABBAABB...". And the green line is the expectation of capital after each toss by playing A and B randomly. Surprisingly, the latter two strategies are winning strategies.

Parrondo's paradox has been used to explain the flashing ratchet model for molecular motors. Here, we use our Fokker-Planck equation I to compute the evolution of the probability distribution and the free energy in the model.

We consider a graph $G = (V, E)$ having 23 vertices:

$$V = \{a_1, a_2, \cdots, a_{23}\} \text{ and } E = \{\{a_i, a_{i+1}\} : i = 1, \cdots, 22\}.$$

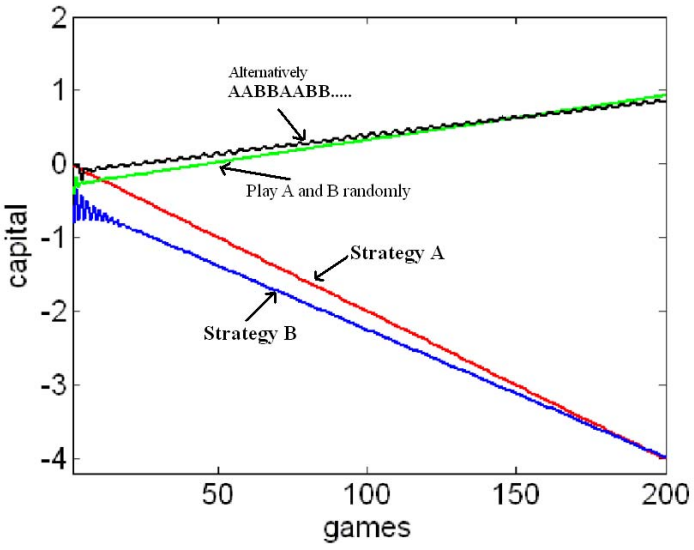


FIGURE 6. Capital vs time. Red line is playing Game A 200 times; Blue line is playing Game B 200 times; Black line is playing Game A twice, then playing Game B twice, and iterating 50 times; Green line is playing Game A and B randomly for 200 times.

It is a 1-D lattice. We give two different potential functions Ψ^A and Ψ^B on the graph: Ψ^A is define on V as in Figure 7 and $\Psi^B_i = 0$ for $i = 1, 2, \cdots, 23$. The values for Ψ^A are list in Table 1,

TABLE 1. Potential function Ψ^A

a_i	1	2	3	4	5	6	7	8	9	10	11	12
Ψ^A_i	5	3.4	2.2	2.5	2.8	3.1	1.9	2.2	2.5	2.8	1.6	1.9
a_i	13	14	15	16	17	18	19	20	21	22	23	
Ψ^A_i	2.2	2.5	1.3	1.6	1.9	2.2	1	1.3	1.6	1.9	4	

We fix the temperature $\beta = 0.05$. Then for a probability density $\rho = (\rho_i)_{i=1}^{23}$ on V, the free energy are

$$F_A(\rho) = \sum_{i=1}^{23} \Psi^A_i \rho_i + 0.05 \sum_{i=1}^{23} \rho_i \log \rho_i$$

$$F_B(\rho) = \sum_{i=1}^{23} \Psi^B_i \rho_i + 0.05 \sum_{i=1}^{23} \rho_i \log \rho_i = 0.05 \sum_{i=1}^{23} \rho_i \log \rho_i.$$

Applying Theorem 4.1 (1) to G with Ψ^A and $\beta = 0.05$, we obtain a discrete Fokker-Planck equation, denoted as Equation A for Process A (Here we omit the detailed expression of Equation A for simply). Similarly, applying Theorem 4.1 (1) to G with

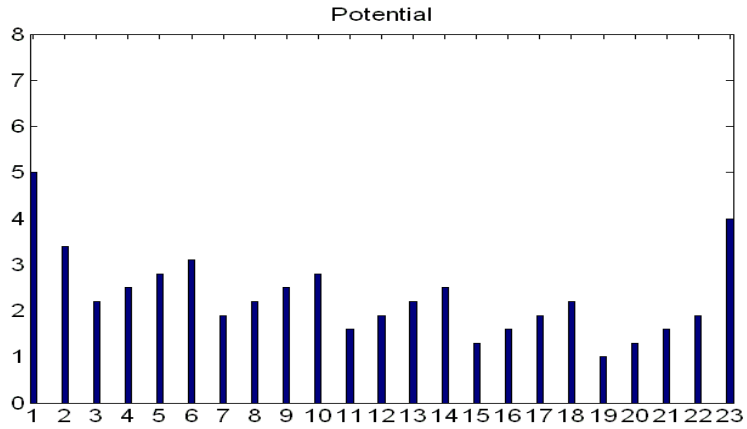


FIGURE 7. Potential function Ψ^A.

Ψ^B and $\beta = 0.05$, we obtain Equation B for Process B:

$$\begin{cases} \dfrac{d\rho_1}{dt} = 0.05(\rho_2 - \rho_1) \\ \dfrac{d\rho_i}{dt} = 0.05(\rho_{i+1} + \rho_{i-1} - 2\rho_i), \quad 1 < i < 23. \\ \dfrac{d\rho_{23}}{dt} = 0.05(\rho_{22} - \rho_{23}) \end{cases}$$

By Theorem 4.1 (3), the free energy F_A decreases with time monotonously along equation A and the free energy F_B decreases with time monotonously along equation B. Thus both processes along Equations A and B are energy dissipative processes, and the free energy decreases monotonously. However, if we apply two processes A and B alternatively, then eventually we may observe an energy gaining process.

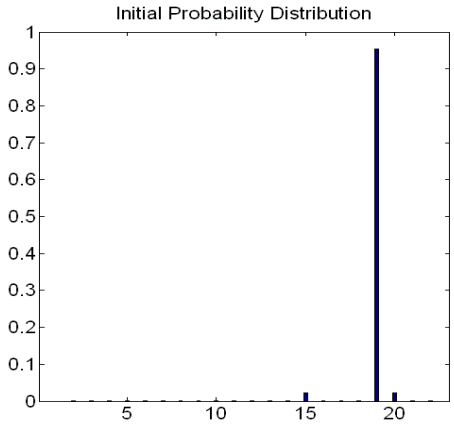


FIGURE 8. Initial Distribution

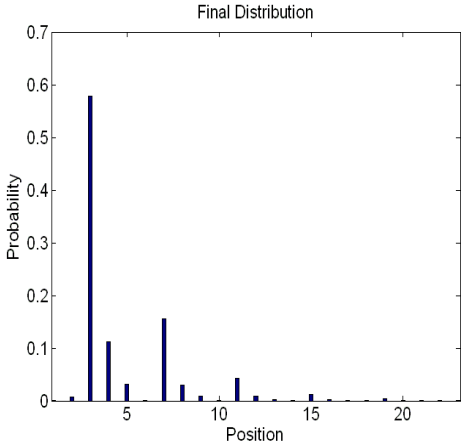


FIGURE 9. Final distribution

More precisely, we choose an initial probability distribution ρ^0 as shown in Figure 8. The peak of ρ^0 was on the right. Next, we choose time interval length $T = 0.3$ and we use Equation A when $0 \leq t < T$, and Equation B when $T \leq t < 2T$. Then we repeat the processes. After taking $ABABAB\cdots$ for 400-times, the peak of probability distribution moved to the left hand side (Figure 9). This indicates a directed motion from the lower potential places to higher potential regions, which can be used to explain the directed motions by molecular motors.

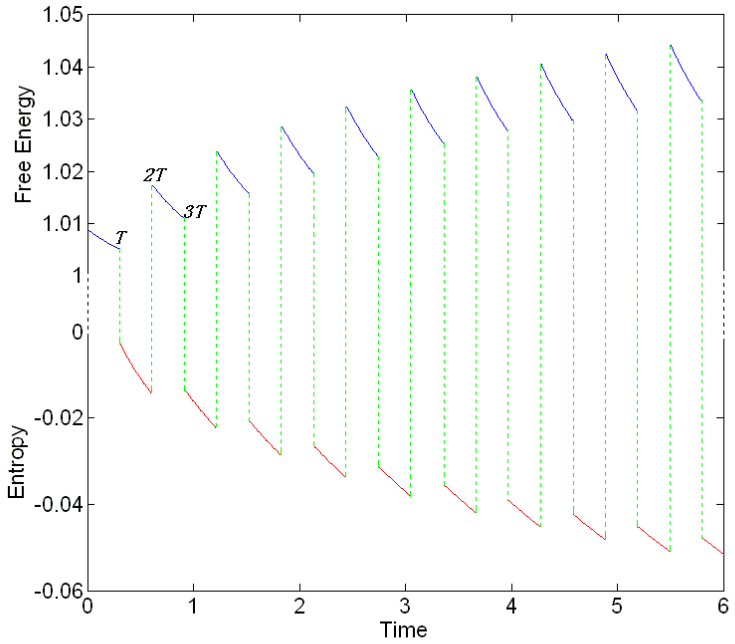


FIGURE 10. first 10 process

To better illustrate the processes, we show the free energy changes in Figure 10 and 11. Figure 10 shows the free energy in first 10 iterations. Process A ends at time T, and Process B begins. At time $2T$, Process B ends, and another Process A starts. These steps are repeated. Although free energy decreased on each process A and B, but the free energy at time $3T$ is still higher that that at time T. So Applying the two processes A and B alternatively, we observe an energy gaining process at the end of process A of each iteration at times T, $3T$, $5T, \cdots$. We show this energy gaining phenomenon in Figure 11. For a similar example in a continuous state space, see [10].

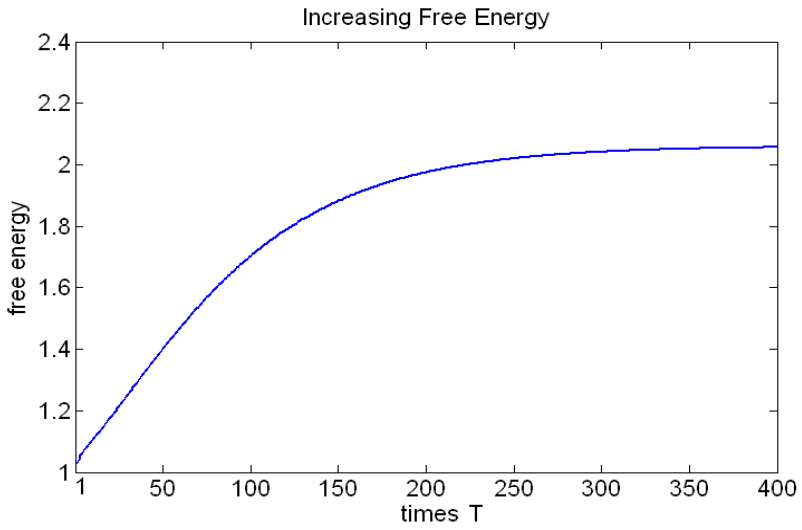


FIGURE 11. The free energy at the end of A process for the first 400 iterations

APPENDIX A. UP WIND SCHEMES FOR HYPERBOLIC EQUATIONS

Here we give a brief introduction to the upwind schemes for hyperbolic equations. Let us start with a simple linear one-way wave propagation equation,

$$u_t + au_x = 0, \quad x \in \mathbb{R},$$

with initial condition given by $u(0,x) = u_0(x)$. The convection coefficient a may be a constant or a function of x. If a is a constant, the solution can be easily verified as

$$u(t,x) = u_0(x - at), \tag{A.1}$$

where $x = x_0 + at$ for any $x_0 \in \mathbb{R}$ is the so called characteristic line of the solution. One salient feature is that the solution are constants along the characteristic lines. In other words, the solution value at one point (t,x) only depends on the initial condition from which the characteristic line is coming. The characteristic directions (reverse in time) is often called the upwind directions. To solve the initial value problem numerically, finite difference methods, including the upwind schemes, have been one of the most popular choices in practice. For simplicity, let us start with a discretization of the problem by introducing the spatial and temporal mesh points $x_i = i * h$ $(i \in \mathbb{Z})$ and $t_n = n * k$ $(n \in \mathbb{Z}^+)$, where $h > 0$ and $k > 0$ are spatial and temporal mesh sizes respectively. We denote u_i^n as the numerical solution approximating the exact solution $u(t,x)$ at (t_n, x_i). Then the upwind scheme is given by

$$u_i^{n+1} = u_i^n - a(x_i)\frac{k}{h}\begin{cases} u_i^n - u_{i-1}^n & \text{if} \quad a(x_i) \geq 0 \\ u_{i+1}^n - u_i^n & \text{if} \quad a(x_i) < 0. \end{cases}$$

Here we note that if the coefficient a is a function that changes sign, then the upwind directions are different from point to point.

The characteristic lines also exist for nonlinear hyperbolic conservation laws. Let us consider the following equation,

$$u_t + f(u)_x = 0, \tag{A.2}$$

where f is a given function, and $u(0,x) = u_0(x)$ is the initial condition. In this case, the characteristic line is determined by

$$\frac{dX(t)}{dt} = f'(u(X(t),t)),$$

which depends on the solution u. It is well-known that singularities (often called shocks in fluid dynamics) may develop, due to the intersections of the characteristic lines, at any time even with smooth initial data. The presents of discontinuities in the solution imposes much tougher challenges in their numerical simulations. Because most of the finite difference methods are based on polynomial interpolations of the solution on discrete values, it is almost inevitable to generate oscillations if one attempts to interpolate the solution across the discontinuities. The oscillations are related to Gibbs' phenomenon. In this case, the upwind strategy, which only interpolates the solution from one-side of the discontinuities, plays a more crucial role in their numerical simulations. Interesting readers are referred to the book by LeVeque [25] for more in-depth discussions.

Appendix B. A counterexample

Applying Theorem 4.1 (1) to a graph G with $\mathbf{\Psi}$ and $\beta > 0$, we obtain Fokker-Planck equation (1.8) on $\mathcal{M}$. For $\boldsymbol{\rho}^0 \in \overline{\mathcal{M}}$, a continuous function $\boldsymbol{\rho}(t) : [0, c) \to \overline{\mathcal{M}}$ for some $c > 0$ or $c = +\infty$ is a solution of equation (1.8) with initial value $\boldsymbol{\rho}^0$, if $\boldsymbol{\rho}(0) = \boldsymbol{\rho}^0$ and $\boldsymbol{\rho}(t) \in \mathcal{M}$ satisfying equation (1.8) for $t \in (0, c)$. Here we give an example to show that for certain graphs and potentials, there may not exist solutions to equation (1.8) with initial value $\boldsymbol{\rho}^0$ for some $\boldsymbol{\rho}^0 \in \partial\mathcal{M} := \overline{\mathcal{M}} \setminus \mathcal{M}$.

Example B.1.

We consider a graph with 3 vertices:

$$V = \{a_1, a_2, a_3\} \text{ and } E = \{\{a_1, a_3\}, \{a_2, a_3\}\}.$$

In this case, we have

$$\mathcal{M} = \{\boldsymbol{\rho} = (\rho_i)_{i=1}^3 \in \mathbb{R}^3 : \rho_1 + \rho_2 + \rho_3 = 1 \text{ and } \rho_i > 0 \text{ for } i = 1, 2, 3\}$$

and

$$\overline{\mathcal{M}} = \{\boldsymbol{\rho} = (\rho_i)_{i=1}^3 \in \mathbb{R}^3 : \rho_1 + \rho_2 + \rho_3 = 1 \text{ and } \rho_i \geq 0 \text{ for } i = 1, 2, 3\}.$$

We assign potential $\mathbf{\Psi} = (\Psi_i)_{i=1}^3$ on V with $\Psi_1 > \Psi_3$ and $\Psi_2 > \Psi_3$, and fix $\beta > 0$. Applying Theorem 4.1 (1) for G, Ψ_i and β, we obtain the Fokker-Planck equation I

(1.8) on $\mathcal{M}$ as follow:

$$\text{(B.1)}\qquad \begin{cases} \dfrac{d\rho_1}{dt} = (\Psi_3 - \Psi_1 + \beta(\log\rho_3 - \log\rho_1))\,\rho_1 \\ \dfrac{d\rho_2}{dt} = (\Psi_3 - \Psi_2 + \beta(\log\rho_3 - \log\rho_2))\,\rho_2 \\ \dfrac{d\rho_3}{dt} = \displaystyle\sum_{i=1}^{2} (\Psi_i - \Psi_3 + \beta(\log\rho_i - \log\rho_3))\,\rho_i \end{cases}$$

Now let $\boldsymbol{\rho}^0 = (0,1,0) \in \partial\mathcal{M}$. Then we claim that there is no solution to equation (B.1) with initial value $\boldsymbol{\rho}^0$.

In fact, if the claim is not true, then there exists a continuous function

$$\boldsymbol{\rho}(t) = (\rho_1(t), \rho_2(t), \rho_3(t)) : [0, c) \to \overline{\mathcal{M}}$$

for some $c > 0$ or $c = +\infty$ such that $\boldsymbol{\rho}(0) = \boldsymbol{\rho}^0$ and $\boldsymbol{\rho}(t) \in \mathcal{M}$ satisfying equation (B.1) for $t \in (0, c)$. By (B.1), one has

$$\begin{aligned} \frac{\mathrm{d}\log\rho_1(t)}{dt} &= (\Psi_3 - \Psi_1) + \beta(\log\rho_3(t) - \log\rho_1(t)) \\ \frac{\mathrm{d}\log\rho_2(t)}{dt} &= (\Psi_3 - \Psi_2) + \beta(\log\rho_3(t) - \log\rho_2(t)) \end{aligned}$$

for $t \in (0, c)$. Let $x(t) = \log\rho_1(t) - \log\rho_2(t)$ for $t \in (0, c)$, then one gets

$$\frac{\mathrm{d}x(t)}{dt} = \Psi_2 - \Psi_1 - \beta x(t)$$

for $t \in (0, c)$. Fix $T \in (0, c)$. It is clear that for any $0 < s < T$,

$$\text{(B.2)}\qquad e^{\beta T} x(T) = e^{\beta s} x(s) + \int_s^T e^{(\Psi_2 - \Psi_1)t} dt.$$

Since $\lim_{s\searrow 0} e^{\beta s} x(s) = -\infty$ and $\lim_{s\searrow 0} \int_s^T e^{(\Psi_2-\Psi_1)t} dt = \int_0^T e^{(\Psi_2-\Psi_1)t} dt$, if one lets $s \searrow 0$ in (B.2), then one has $e^{\beta T} x(T) = -\infty$, i.e., $\rho_1(T) = -\infty$. This is a contradiction with $(\rho_1(T), \rho_2(T), \rho_3(T)) \in \mathcal{M}$.

References

[1] Ait-Haddou, R. and Herzog, W., Brownian Ratchet Models of Molecular Motors, Cell Biochemistry and Biophysics 38 (2) (2003) 191-213.

[2] Ambrosio, L., Gigli, N. and Savare, G., Gradient Flows: In Metric Spaces and in the Space of Probability Measures, Lectures in Mathematics ETH Zürich Birkhäuser Verlag, Basel, 2008.

[3] Bonciocat, A. I. and Sturm, K. -T., Mass transportation and rough curvature bounds for discrete spaces, J. Funct. Anal. 256 (9) (2009) 2944-2966.

[4] Van den Broeck, C., The master equation and some applications in physics, Stochastic processes applied to physics, World Sci. Publishing, Singapore, 1985, 1-28.

[5] Buet, C. and Cordier, S., Numerical Analysis of Conservative and Entropy Schemes for the Fokker-Planck-Landau Equation, SIAM J. Numer. Anal. 36 (3) (1999) 953-973.

[6] Bergmann, P. G. and Lebowitz, J. L., New Approach to Nonequilibrium Processes, Phys. Rev. 99, 578 (1955).

[7] Carlen, E. A. and Gangbo, W., Constrained steepest descent in the 2-Wasserstein metric, Ann. of Math. (2) 157 (3) (2003) 807-846.
[8] Carlen, E. A. and Gangbo, W., On the solution of a model Boltzmann equation via constrained steepest ascent in a Wasserstein metric. Arch. Rational Mech. Ana., 172, no. 1, 21-64, 2004.
[9] Cordero, D., Gangbo, W. and Houdre, C., Inequalities for generalized entropy and optimal transportation. Comteporary Mathematics, AMS Vol 353, 2004.
[10] Chow, S.-N., Huang, W., Li, Y. and Zhou, H. M., A free energy based mathematical study for molecular motors, 2010, submitted.
[11] Cordero-Erausquin, D., McCann, R. J. and Schmuckenschläger, M., A Riemannian interpolation inequality á la Borell, Brascamp and Lieb., Invent. Math. 146 (2) (2001) 219-257.
[12] Dobrushin, R. L., Prescribing a system of random variables by conditional distributions, Theor. Probab. Appl. 15 (1970), 458-486.
[13] Ellis, R. S., Entropy, Large Deviations and Statistical Mechanics, Grundlehren der Mathematischen Wissenschaften, vol. 271, Springer-Verlag, New-York, 1985.
[14] Evans, L. C., Partial differential equations and Monge-Kantorovich mass transfer, Current developments in mathematics, 1997, 65-126, Int. Press, Boston, MA, 1999.
[15] Evans, L. C., Entropy and Partial Differential Equations, Lecture Notes for a Graduate Course, see "http://math.berkeley.edu/ evans/".
[16] Evans, L. C. and Gangbo, W., Differential equations methods for the Monge-Kantorovich mass transfer problem, Mem. Amer. Math. Soc. 137 (1999), no. 653.
[17] Gardiner, C. W., Handbook of stochastic methods. For physics, chemistry and the natural sciences. Second edition. Springer series in Synergetics, vol. 13, Springer-Verlag, Berlin, 1985.
[18] Harmer, G. P. and Abbott, D., Losing strategies can win by Parrondo's paradox, Nature, 402 (6764) (1999) 864.
[19] Harmer, G. P., Abbott, D., A review of Parrondo's Pararox, Fluctuation and Noise Letters 2 (2) (2002) 71-107.
[20] Heath, D., Kinderlehrer, D. and Kowalczyk, M., Discrete and continuous ratchets: from coin toss to molecular motor, Discrete Contin. Dyn. Syst. Ser. B 2 (2) (2002) 153-167.
[21] Lott, J. and Villani, C., Ricci curvature for metric-measure spaces via optimal transport, Ann. of Math. (2) 169 (3) (2009) 903-991.
[22] Jordan, R., Kinderlehrer, D. and Otto, F., The variational formulation of the Fokker-Planck equation, SIAM J. Math. Anal. 29 (1) (1998) 1-17.
[23] Kinderlehrer, D. and Otto, F., Free energy and the Fokker-Planck equation. Landscape paradigms in physics and biology (Los Alamos, NM, 1996), Phys. D 107 (2-4) (1997) 265-271.
[24] Kinderlehrer, D. and Kowalczyk, M., Diffusion-Mediated Transport and the Flashing Ratchet, Arch. Rational Mech. Anal. 161 (2002) 149-179.
[25] LeVeque, R., Numerical Methods for Conservation Laws, Lectures in Mathematics ETH Zürich, Birkhäuser, Verlag, 1992.
[26] Lott, J., Some Geometric Calculations on Wasserstein Space, Commun. Math. Phys. 277 (2) (2008) 423-437.
[27] Mathews, D. H. and Turner, D. H., Prediction of RNA secondary structure by free energy minimization, Current Opinion in Structural Biology 16 (3) (2006) 270-278.
[28] McCann, R. J., Exact solutions to the transportation problem on the line, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. 455 (1999), no.1984, 1341-1380.
[29] Otto, F., The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations 26 (1-2) (2001) 101-174.
[30] Otto, F. and Villani, C., Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal. 173 (2) (2000) 361-400.
[31] Otto, F. and Westdickenberg, M., Eulerian calculus for the contraction in the Wasserstein distance, SIAM J. Math. Anal. 37 (4) (2005) 1227-1255.
[32] Percus1, O. E. and Percus, J. K., Can Two Wrongs Make a Right? Coin-Tossing Games and Parrondos Paradox, The Mathematical Intelligencer, Volume 24, Number 3 (2002) 68-72.

[33] Qian, H., Motor protein with nonequilibrium potential: its thermodynamics and efficiency, Phys. Rev. E 69, 012901 (2004).
[34] Qian, H., Cycle kinetics, steady state thermodynamics and motors - a paradigm for living matter physics, Journal of Physics: Condensed Matter, 17(2005) 3783-3794.
[35] Qian, M., Zhang X., Wilson R. J. and Feng, J., Efficiency of Browian motors in terms of entropy production rate, EPL 84 10014 (2008).
[36] Von Renesse, M. -K. and Sturm, K. -T., Transport inequalities, gradient estimates, entropy, and Ricci curvature, Comm. Pure Appl. Math. 58 (7)(2005) 923-940.
[37] Von Renesse, M. -K., Sturm, K. -T., Entropic measure and Wasserstein diffusion, Ann. Probab. 37 (3)(2009) 1114-1191.
[38] Risken, H., The Fokker-Planck Equation: Methods of Solution and Applications, Second edition, Springer Series in Synergetics, vol. 18, Springer-Verlag, Berlin, 1989.
[39] Sammer, M. and Tetali, P., Concentration on the discrete torus using transportation, Combin. Probab. Comput. 18 (2009), no. 5,
[40] Schuss, Z., Singular perturbation methods in stochastic differential equations of mathematical physics, SIAM Rev. 22 (2)(1980) 119-155.
[41] Smolka, B. and Wojciechowski, K. W., Contrast enhancement of badly illuminated images based on Gibbs distribution and random walk model, Lecture Notes in Computer Science, vol. 1296, Springer Berlin, 1997.
[42] Sturm, K. -T., Generalized Ricci bounds and convergence of metric measure spaces, C. R. Math. Acad. Sci. Paris 340 (3) (2005) 235-238.
[43] Sturm, K. -T., On the geometry of metric measure spaces I, Acta Math. 196 (1) (2006) 65-131.
[44] Sturm, K. -T., On the geometry of metric measure spaces II, Acta Math. 196 (1)(2006) 133-177.
[45] Villani, C., Optimal Transport. Old and New, Grundlehren der Mathematischen Wissenschaften, 338. Springer-Verlag, Berlin, 2009.
[46] Villani, C., Topics in Optimal Transportation. Graduate Studies in Mathematics 58, American Mathematical Society, Providence (2003).
[47] Wasserstein, L.N., Markov processes over denumerable products of spaces describing large systems of automata, Probl. Inform. Transmission 5 (3)(1969) 47-52.
[48] Wu, Y., Hua, G. and Yu, T., Tracking Articulated Body by Dynamic Markov Network, the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set.

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332
E-mail address: chow@math.gatech.edu, yli@math.gatech.edu,hmzhou@math.gatech.edu

Department of mathematics, University of science and Technology of china, Hefei Anhui 230026, P. R. China
E-mail address: wenh@mail.ustc.edu.cn