Free Energy, Fokker-Planck Equations, and Random walks on a Graph with Finite Vertices

> *Haomin Zhou* Georgia Institute of Technology

Jointly with *S.-N. Chow* (Georgia Tech) *Wen Huang* (USTC) *Yao Li* (NYU)

Research Supported Partially by NSF/DTRA, ONR

# **Outline**

- **Introduction and Motivation**
- Main results
- Idea of proof
- Examples

#### Introduction: continuous media

•Randomly perturbed gradient system:

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

•Time evolution of the probability density function, the Fokker-Planck equation:

$$
\rho_t(x,t) = \nabla \cdot (\nabla \Psi(x)\rho(x,t)) + \beta \Delta \rho(x,t)
$$

•Invariant distribution at steady state -- Gibbs distribution:

$$
\rho^*(x) = \frac{1}{K} e^{-\Psi(x)/\beta} \qquad K = \int_{\mathbb{R}^N} e^{-\Psi(x)/\beta} dx
$$

### Introduction: free energy view

•Free energy  $F(\rho) = U(\rho) - \beta S(\rho)$ 

•**Potential** 
$$
U(\rho) = \int_{\mathbb{R}^N} \Psi(x) \rho(x) dx
$$

•**Gibbs-Boltzmann Entropy**  

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

•Fokker-Planck equation is the gradient flow of the free energy under 2-Wasserstein metric on the manifold of probability space.

•Gibbs distribution is the global attractor of the gradient system.

#### Introduction: Wasserstein metric  $T$  original transport problem, problem, problem, problem, problem, as as  $\alpha$ introduction: vvasserstein metric of work. In modern parlance, we are given two nonnegative Radon measures µ<sup>±</sup> on R<sup>n</sup>

•Related to mass transport (Monge-Kantorovich): so that cords that cords the work required to move a unit mass from the position  $\alpha$  except a unit mass from the position  $\alpha$ • Related to mass transport (Monge-Kantorovich): simply proportional to the distance moved.)



Our problem is therefore to find and characterize an optimal mass transfer s<sup>∗</sup> ∈ A which

measure µ<sup>+</sup> onto µ<sup>−</sup> and, among all such mappings, minimizes I[·]. We will later see that a

This is even now, over two hundred years later, a difficult mathematical problem, owing

really remarkable array of interesting mathematical and physical interpretations follow.

The total work corresponding to a mass rearrangement plan s ∈ A is thus the total work corresponding to a is t<br>A is thus thus the total work corresponding to a is thus the total work corresponding to a isolate to a isolat

$$
\mu(\mathbb{R}) - \mu(\mathbb{R}) < \infty,
$$
\n
$$
\sum_{\mathbf{x} = \mathbf{spt}(\mathbb{R}^+)} \mathsf{Cost functional: } I[\mathbf{s}] := \int_{\mathbb{R}^n} c(x, \mathbf{s}(x)) d\mu^+(x).
$$
\n
$$
\sum_{\mathbf{x} = \mathbf{spt}(\mathbb{R}^+)} \mathsf{Optimal transport: } s^* = \arg \min_{s \in A} I[s]
$$

**X=spt(**µ**+) Y=spt(**µ**-)** 

In other words, we wish to construct a one-to-one mapping <sup>s</sup><sup>∗</sup> : <sup>R</sup><sup>n</sup> <sup>→</sup> <sup>R</sup><sup>n</sup> which pushes the

h(y) dµ−(y) (1.3)

 $\mathcal{L}=\mathcal{$ 

I[s]. (1.5)

 $\mu + (\mathbb{D}^n) = \mu - (\mathbb{D}^n) < \infty$ 

<sup>c</sup> : <sup>R</sup><sup>n</sup> <sup>×</sup> <sup>R</sup><sup>n</sup> <sup>→</sup> [0, <sup>∞</sup>);

•2-Wasserstein metric: Rn n metric: minimal cost to tran  $\mathcal{L} = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{i} \sum_{j=1}^{n} \frac{1}{j} \sum$  $\sqrt{2}$  =  $\sqrt{2}$ minimal cost to transport  $\rho^1$  to  $\rho^2$ ,

$$
W_2(\rho^1, \rho^2)^2 = \inf_{\mu \in \mathcal{P}(\rho^1, \rho^2)} \int d(x, y)^2 d\mu(x, y)
$$

minimizes the work:

### Introduction

•Fokker-Planck equations + 2-Wasserstein metric: Otto, Kinderlehrer, Villani, McCann, Carlen, Lott, Strum, Gangbo,Jordan,Evans, Brenier, Benamou, and many many more,

- •Related weak KAM: Mather, E, Fathi, Evans, ...  $\mathbf{F}$  of the Theorems  $\mathbf{F}$  of the Theorems  $\mathbf{F}$
- •Related to linear programming, manifold learning, image processing, **Exercice Commeation Construction** 
	- •Complete picture in continuous media:



### Motivation

- Our Goal: establish Fokker-Planck equations on graphs with finite vertices.
	- 1, Laplace operator on graphs,
	- 2, White noise to a Markov process on graphs,
- Why on graphs: Physical space (number of sites or states) is finite, not necessary from a spatial discretization such as a lattice.
- Applications: game theory, RNA folding, logistic, chemical reactions, machine learning, Markov networks, numerical schemes, ...
- Mathematics: Graph theory, Mass transport, Dynamical systems, Stochastic Processes, PDE's, ...
- Many Recent Developments: Erbar, Mielke, Mass, Gigli, Ollivier, Villani, Tetali, Qian,...

#### Motivation: basic setup and the equipotential set of ai as a set of ai as<br>In the extent of ai as a set of ai

#### Graph with finite vertices:  $\mathcal{C}^{\text{max}}_{\text{max}}$  and  $\mathcal{C}^{\text{max}}_{\text{max}}$  and  $\mathcal{C}^{\text{max}}_{\text{max}}$  and  $\mathcal{C}^{\text{max}}_{\text{max}}$

we define the set of predecessors of  $\sigma$  and  $\sigma$  are defined by a set of ai as  $\sigma$ 

the set of successors of a<sup>i</sup> as

$$
G = (V, E), \quad V = \{a_1, \cdots, a_N\} \qquad E \text{ the edges of } G
$$

Riemannian metric g) is  $\overline{\phantom{a}}$ Neighbors of a vertex  $C^{\infty}$  and  $C^{\infty}$  are  $\mathcal{A}^{\infty}$  and  $\mathcal{A}^{\infty}$  and  $\mathcal{A}^{\infty}$  is the  $X$ ,  $\mathcal{A}^{\infty}$  $\textsf{lex} \; : N(i) = \{a_j \in V | \{a_i, a_j\} \in E\}$ 

Free energy: Potential Entropy  $\overline{P}$  $\blacksquare$  $(i=1)$   $(i=1)$  $F(\rho) = \sum$ *N i*=1  $\Psi_i \rho_i + \beta$  $\sum$ *N i*=1  $\rho_i \log \rho_i$ 

 $\left\{ \rho_{i}\right\} _{i=1}^{N}$  Probability density function defined on the graph G **density funct**  $j$  ects  $\frac{1}{2}$  $\overline{a}$  defined on the graph  $\overline{a}$ 

 $\blacksquare$ Shuite Chow, Wen Huang, Yao Li, Hanno Chow, Yao Li, Hanno Chow, Yao Li, Hanno On Graph with Finite Vertices Vertices  $\blacksquare$  $\overline{C}$  :  $\overline{C}$   $\over$ Gibbs distribution:

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

is the unique stationary distribution of equation (2.8) in M, and the free energy F attains

 $e^{i\theta}$  is the neighborhood of air  $e^{i\theta}$  is the neighborhood of air  $e^{i\theta}$  is the neighborhood of air  $e^{i\theta}$ 

Our Motivation

Then

#### Motivation: A toy example Varion Propriet of the Theorems of the Theorems of the Theorems of the Theorems of the Theorem **N** LUY

#### Intuition: it is seamless from continuous to discrete.

Examples

Consider this potential function again:



Figure: Potential Energy

#### .<br>Г Discretization

..  $3, a_4 = 4, a_5 = 5$ . We make a discretization at five points  ${a_1 = 1, a_2 = 2, a_3 =$ 

#### . **Noise**

..  $\beta = 0.8$ We set the noise strength

. . . . . .

### Motivation: A toy example

#### **Continuous**



Discrete (central- difference scheme):





Figure: Central Difference



 $\mathcal{F}_{\mathcal{F}}$  for  $\mathcal{F}_{\mathcal{F}}$  and  $\mathcal{F}_{\mathcal{F$ 

### **Motivation**

#### Challenges:

• Common discretizations of continuous equations often lead to incorrect results,

*Theorem*: Any given linear discretization of the continuous equation can be written as

$$
\frac{d\rho_i}{dt} = \sum_j \left( \left( \sum_k e_{jk}^i \Phi_k \right) + c_j^i \right) \rho_j.
$$

Let

$$
A = \{ \Phi \in \mathbb{R}^N : \sum_j ((\sum_k e_{jk}^i \Phi_k) + c_j^i) e^{-\frac{\Phi_j}{\beta}} = 0 \}.
$$

Then A is a zero measure set.

• Graphs are not length spaces and many of the essential techniques cannot be used anymore,

• The notion of random perturbation (white noise) of a Markov process on discrete spaces is not clear.

#### Our Strategies Proof of the Theorems

Derive Fokker-Planck equations on graphs in two different ways edges; Neighborhood of the new setting in the new setting of the new setting in the new setting of an interval of  $\sim$ 



New ideas:

Our Motivation

- •White Noise for Markov processes on Graphs, .<br>... ie for Markov<br>Ieme  $\mathbf{L}$ pi occases  $\overline{a}$
- •Upwind scheme,  $-1$
- •ODEs for Fokker-Planck equations,  $r_{\rm i}$
- •Gradient flows on Riemannian Manifolds.

# Our Strategies

•Inspired by:

upwind scheme for numerical solutions of hyperbolic conservation laws

•Motivated by:

another interpretation of white noise on graphs

•Infuenced by:

- Remarkable result of Jordan, Kinderlehrer, Otto (1998)  $\bullet$
- Heath, Kinderlehrer, Kowalczyk (2002) ; discrete and continuous ratchets  $\bullet$
- Parrondo paradox (1996); review article by Harmer, Abbott (2002)  $\bullet$
- Carlen, Gangbo (2003); nonlinear Fokker Planck, constrained gradient flow  $\bullet$

# Our Strategies

#### Approach the problem from two different ways

•Define Riemannian Manifold (*M, d*)

$$
\mathcal{M} = \left\{ \{ \rho_i \}_{i=1}^N \in \mathbb{R}^N \mid \sum_{i=1}^N \rho_i = 1; \rho_i > 0 \right\}
$$

 $C_{\rm eff}$  and  $C_{\rm eff}$  and  $V_{\rm eff}$  and  $V_{\rm eff}$  and  $V_{\rm eff}$  and  $V_{\rm eff}$  is the  $X_{\rm eff}$  is the  $X_{\rm eff}$ 

 $\cdot$ n is the  $\cdot$ Main Results Fokker-Planck equation is the gradient flow of the free energy on the manifold.

• Add "white" noise to Markov processes on the graph.

Remark 1 Fokker-Planck equation describes the dynamics of the transition probability density function.



Both equations (2) and (3) are spatial discretizations of

#### Main Results <u>Main Results in the set of the set</u> Proof of the Theorems **Manual**

Give Given free energy  ${\mathcal F}(\rho)=\sum_{i=1}^N\Psi_i\rho_i+\beta\rho_i\log\rho_i$  on a graph  $G = (V, E).$ 

#### Theorem 1 Theorem I G = (V, E).<br>G = (V, E). we have a Fokker-Planck equation in the form of the state o<br>Planck equation is a form of the state of the

we have a Fokker-Planck equation i idiich cy  $(\bullet + \bullet \bullet \bullet$ 

$$
\frac{d\rho_i}{dt} = \sum_{j \in N(i); \Psi_j > \Psi_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i)) \rho_j
$$
  
+ 
$$
\sum_{j \in N(i); \Psi_i > \Psi_j} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i)) \rho_i
$$
  
+ 
$$
\sum_{j \in N(i); \Psi_i = \Psi_j} \beta(\rho_j - \rho_i)
$$

#### Main Results IMAIN main Results

Given a graph  $G = (V, E)$ , a "gradient Markov process" on graph G generated by potential  $\{\Psi_i\}_{i=1}^N$ , suppose the process is subjected to "white noise" with strength  $\beta > 0$ Given a graphy

Introduction

Examples

#### Theorem II we have a Fokker-Planck equation (different from the fokker-Planck equation (different from the focker-Planck e<br>The form the form th

nsition probabilit.<br>seatisfies the folle Sa ((Ψ<sup>j</sup> + β log ρj) − (Ψ<sup>i</sup> + β log ρi))ρ<sup>j</sup>  $\frac{1}{2}$ subjected to "with strength  $\frac{1}{2}$  or  $\frac{1}{2}$  or  $\frac{1}{2}$   $\frac{1}{2}$ u ansierum probability density function of the perturbe ess salisties lite followit The transition probability density function of the perturbed Markov process satisfies the following "Fokker-Planck" equation

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

$$
+ \sum_{j \in N(i); \bar{\Psi}_j > \bar{\Psi}_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_i
$$

 $\mathcal{L}$  and  $\mathcal{L}$  and  $\mathcal{L}$  and  $\mathcal{L}$   $\$ where  $\overline{\Psi}_i = \Psi_i + \beta \log \rho_i$ 

Equation in Theorem II is different from the equation in Theorem I

Shui-Nee Chow, Wen Huang, Yao Li, Haomin Zhou Fokker-Planck Equation on Graph with Finite Vertices Tuesday, October 15, 2013

#### Properties of the Equations Examples Theorem 1  $\blacksquare$   $\blacksquare$   $\blacksquare$   $\blacksquare$   $\blacksquare$   $\blacksquare$   $\blacksquare$   $\blacksquare$

#### The equations in both Theorem I and Theorem II have the following common properties Introduction Proposed the Theorems of the Theorems of the Theorem set of the Theorem is the Theorem of the Theorem is the T<br>The Theorem is the T ang i ne Proof of the Theorems **Porpr** llowing common pr  $\mathbf{r}$ J∪LI I Theorem Lond Theorem 1 perties<br>}

Introduction

j∈P(i)



Gibbs density  $\rho^*$  is a stable stationary solution and is the global minimum of free energy  $\mathsf{Gibbs}$  density a  $\bigwedge$  in  $\bigwedge$ (2) The Gibbs distribution of the Gibbs distribution of the Gibbs distribution of the Gibbs distribution of th<br>The Gibbs distribution of the Gibbs distribution of the Gibbs distribution of the Gibbs distribution of the Gi

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

Given any initial data  $\rho_0 \in \mathcal{M}$ , we have a unique  $\rho(t):[0,\infty)\to \mathcal{M}$  with initial value  $\rho_0$ , and Given any initial data  $\rho_0 \in \mathcal{M}$ , we have a unique solution  $\rho(t):[0,\infty)\rightarrow \mathcal{M}$  with initial value  $\rho_0$ , and limt→∞ ρ(t) = ρ<sup>∗</sup>  $\epsilon$ is the unique stationary distribution of equation (2.8) in M, and the free energy  $\epsilon$ minimum at the Given any initial da  $\rho(t):[0,\infty)\rightarrow \mathcal{M}$  with initial value  $\rho_0$ , and with initial value  $\mathcal{N}$  in  $\mathcal{N}$  is and  $\mathcal{N}$  satisfies and  $\mathcal{N}$  satisfies and  $\mathcal{N}$ 

$$
\lim_{t \to \infty} \rho(t) = \rho^*
$$

 $c = \frac{1}{\sqrt{2}}$  is called a solution of equation of equation (2.8) with initial value  $p$ 

The boundary of  $\mathcal M$  is a repeller **Remarks 2.4. 1. A continuous function provided** provided provided provided provided provided provided provided p<br>Provided provided provided

Then

## Remarks

• Both equations are consistent, but non-standard, discretization of the continuous Fokker-Planck equation by upwind schemes.

• Both equations are gradient flows of the free energy w.r.t. different metrics.

• Those metrics, depending on the potential function, on the Riemannian manifolds are bounded by two metrics that are independent of the potential.

• Near Gibbs distribution (steady state), two equations are almost the same. The difference is small.

# Proof

#### Idea of Proof for Theorem I

• Construct the Riemannian manifold (*M, d*).

$$
\mathcal{M} = \left\{ {\rho_i} \}_{i=1}^N \in \mathbb{R}^N \mid \sum_{i=1}^N \rho_i = 1; \rho_i > 0 \right\}
$$

difficulties: how to define *d*?

• Compute the gradient flow of the free energy on the Riemannian manifold,

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

$$
\frac{d\rho}{dt} = -\text{grad} F|_{\rho}
$$

# Proof for Theorem II

 $S_{\rm eff}$  ,  $N_{\rm eff}$   $\sim$   $L_{\rm eff}$   $\sim$   $L_{\rm eff}$   $\sim$   $L_{\rm eff}$   $\sim$   $L_{\rm eff}$  with  $T_{\rm eff}$  with  $T_{\rm eff}$  with  $T_{\rm eff}$   $\sim$   $L_{\rm eff}$   $\sim$ 

Theorem 1

Main Results

#### Ideas of proof Iday of Pro



- "gradient flow" on the graph
- "gradient flow" subject to "white noise" perturbation
- **•** Fokker-Planck equation in Theorem 2

# Proof for Theorem II

Key observations

Free Energy

$$
F = \sum \Psi_i \rho_i + \beta \sum \rho_i \log \rho_i = \sum (\Psi_i + \beta \log \rho_i) \rho_i
$$

Continuous Fokker-Planck equation

$$
\rho_t = \nabla \cdot (\nabla \Psi \rho) + \beta \Delta \rho = \nabla \cdot (\nabla \Psi \rho + \beta \nabla \rho)
$$
  
= 
$$
\nabla \cdot [(\nabla \Psi + \beta \nabla \rho/\rho)\rho] = \nabla \cdot [\nabla (\Psi + \beta \log \rho)\rho]
$$

New potential  $\overline{\Psi} = (\Psi + \beta \log \rho)$ 

Kolmgorov forward equation with the new potential leads to Fokker-Planck equation in Theorem II

#### Markov Process on Graphs Introduction VIIV DO C

#### "Gradient Flow" on Graphs

We call the following Markov process  $X(t)$  a gradient Markov process generated by potential  $\{\Psi_i\}_{i=1}^N$ : For  $\{a_i, a_j\} \in E$ , if  $\Psi_i > \Psi_j$ , the transition rate  $q_{ij}$  from *i* to *j* is  $\Psi_i - \Psi_i$ .

Examples



### White Noise Perturbations

Markov process  $X(t)$  on the graph with transition rate  $q_{ij}$ 

$$
P(X(t + h) = a_j | X(t) = a_i) = q_{ij}h + o(h)
$$

Kolmgorov forward equation for probability density function

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

White noise to the Markov process can be viewed as a perturbation to its potential (transition rate)

$$
\Psi_i \to (\Psi_i + \beta \log \rho_i)
$$

### Laplace Operator on a Graph

Laplace operator for a positive function  $\rho$  defined on  $G$  can be given by

$$
\Delta \rho_i = \sum_{j \in N(i), \rho_j > \rho_i} (\log \rho_j - \log \rho_i) \rho_j + \sum_{j \in N(i), \rho_j < \rho_i} (\log \rho_j - \log \rho_i) \rho_i
$$

On an 1-D lattice with  $\rho_{i-1} < \rho_i < \rho_{i+1}$ , it becomes

$$
\Delta \rho_i = (\log \rho_{i+1} - \log \rho_i)\rho_{i+1} + (\log \rho_{i-1} - \log \rho_i)\rho_i
$$

# Example : Discrete Flashing Ratchet

- Main idea: Two energy dissipative processes may lead to free energy increasing if used alternatively or randomly,
- Related to Parrondo's Paradox: two losing game strategies may lead to a winning strategy if used alternatively or randomly,
- Used to explain working mechanism of molecular motors,
- There exists an extensive literature: Parrondo, Harmer, Abbott, Heath, Kinderlehrer, Kowalczyk, Ait-Haddou,Herzog, ...

# Flashing Ratchet

Two energy decreasing processes

A: randomly perturbed gradient flow of the potential function, governed by Fokker-Planck equation in Theorem I,

$$
\frac{d\rho_i}{dt} = \sum_{j \in N(i); \Psi_j > \Psi_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i)) \rho_j
$$
  
+ 
$$
\sum ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i)) \rho_i + \sum
$$

*j*∈*N*(*i*);Ψ*i*=Ψ*<sup>j</sup>*  $\beta(\rho_j - \rho_i)$ 

B: randomly diffusion on the graph, governed by the Fokker-Planck equation in Theorem I with constant potential,

#### Discrete Potential

 $\mathcal{F}_{\mathcal{A}}$  for example, lett $\mathcal{A}$  and discrete potential function on  $1$ 



 $S_{\rm eff}$  ,  $N_{\rm eff}$   $\sim$   $N_{\rm eff}$   $\sim$   $N_{\rm eff}$   $\sim$   $N_{\rm eff}$   $\sim$   $N_{\rm eff}$  with  $\sim$   $N_{\rm eff}$   $\sim$   $N_{\rm eff}$ 

. . . . . .

 $rac{d\rho_i}{dt} = \sum_{i \in \mathcal{N}_i}$ *j*∈*N*(*i*)  $\beta(\rho_j - \rho_i)$ 

*j*∈*N*(*i*);Ψ*i>*Ψ*<sup>j</sup>*

In both cases, the probability density function moves to the right (lower energy states).

# Flashing Racket

Discrete Flashing Ratchet

 $P$  is the Theorems of the Th

Motion from lower potential to higher potential And directed motion could be observed: And directed motion could be observed:



Shui-Nee Chow, Wen Huang, Yao Li, Haomin Zhou Fokker-Planck Equation on Graph with Finite Vertices

. . . . . .

Directed Motion

Flashing Ratchet Examples

#### Use A, B alternatively as: ABABAB...



Increase and the control of the control of

**Flashing Ratchet Property** Examples  $D$ otokot

#### Free energy plot of the first 10 steps Evergy vs. Time  $T$  is the free energy in first 10  $\mu$  processes and B proce



### Influence Predictions in Networks

Given some cascades (observations of information propagating in a network, for which the structure may not be even known) up to a certain time.

Goal: predict the influence region at a later time.



### Example : Parrondo's Paradox

# Quote NYT, Jan 2000

The paradox may shed light on social interactions and voting behaviors, Dr. Abbott said. For example, President Clinton, who at first denied having a sexual affair with Monica S. Lewinsky (game A) saw his popularity rise when he admitted that he had lied (game B.) The added scandal created more good for Mr. Clinton.

### The End

# Thank You!