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:

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

Introduction: free energy view

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

- Potential $$U(\rho) = \int_{\mathbb{R}^N} \Psi(x)\rho(x)\mathrm{d}x$$

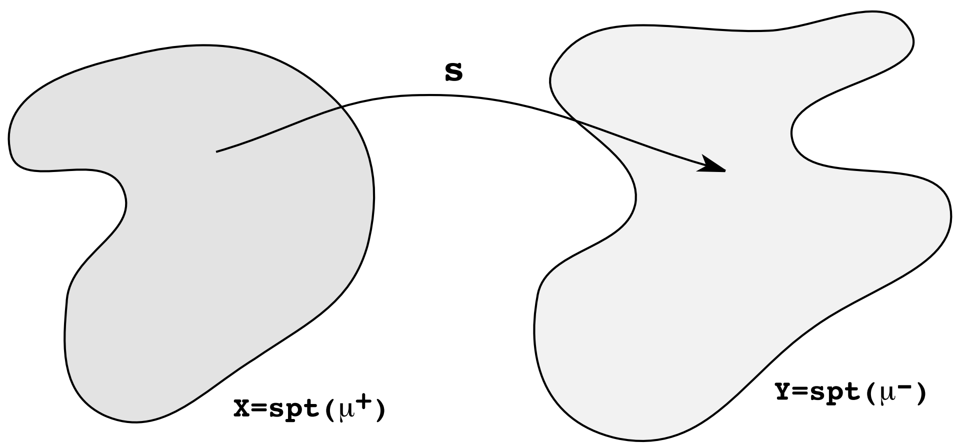
- Gibbs-Boltzmann Entropy
$$S(\rho) = -\int_{\mathbb{R}^N} \rho(x)\log\rho(x)\mathrm{d}x$$

- 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

• Related to mass transport (Monge-Kantorovich):



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

Cost functional: $I[\mathrm{s}] := \int_{\mathbb{R}^n} c(x, \mathrm{s}(x))\, d\mu^+(x).$

Optimal transport: $s^* = arg \min_{s \in A} I[s]$

• 2-Wasserstein metric: minimal cost to transport ρ^1 to ρ^2,

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

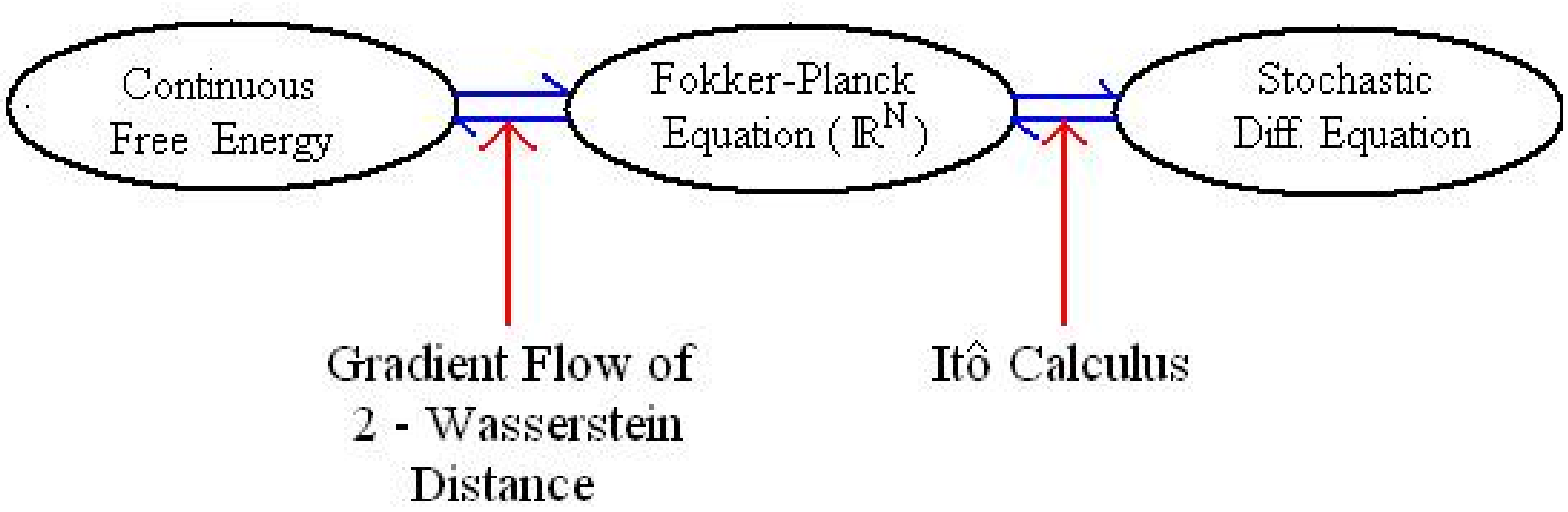
Introduction

•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, ...

•Related to linear programming, manifold learning, image processing,

•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

Graph with finite vertices:

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

Neighbors of a vertex : $N(i) = \{a_j \in V | \{a_i, a_j\} \in E\}$

Free energy:

$$F(\rho) = \underbrace{\sum_{i=1}^{N} \Psi_i \rho_i}_{\text{Potential}} + \beta \underbrace{\sum_{i=1}^{N} \rho_i \log \rho_i}_{\text{Entropy}}$$

$\{\rho_i\}_{i=1}^{N}$ Probability density function defined on the graph G

Gibbs distribution:

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

Motivation: A toy example

Intuition: it is seamless from continuous to discrete.

Consider this potential function again:

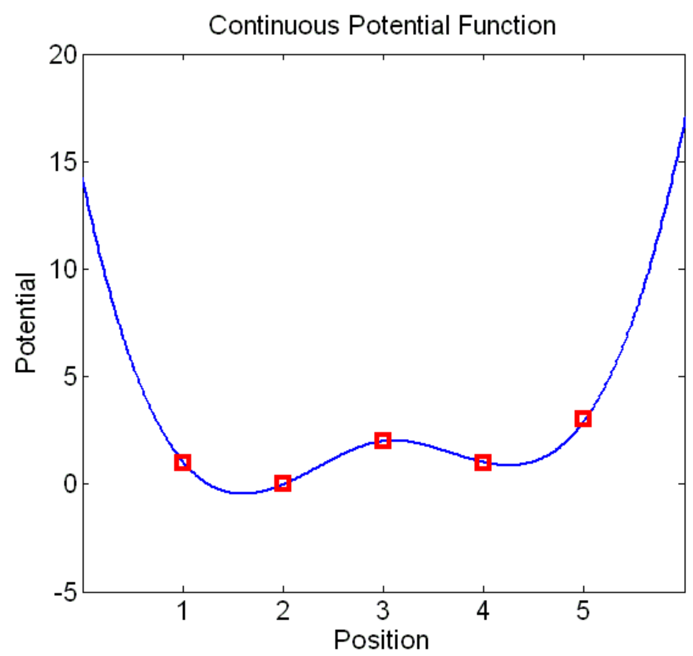


Figure: Potential Energy

Discretization

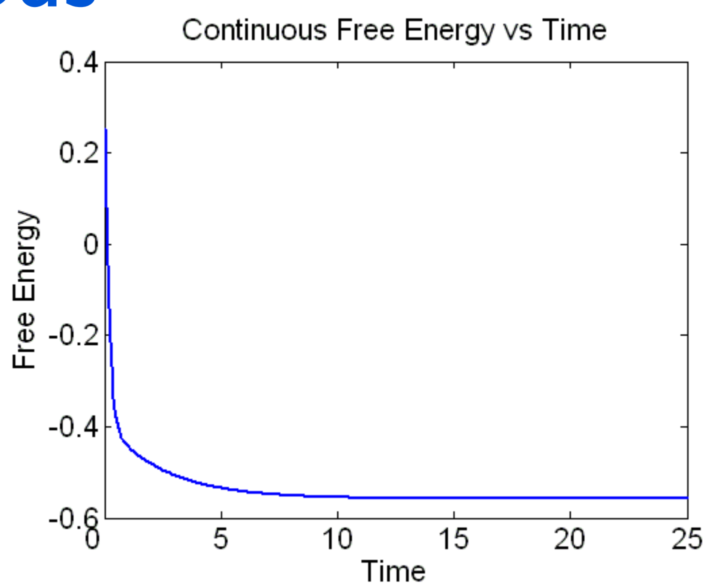
We make a discretization at five points $\{a_1 = 1, a_2 = 2, a_3 = 3, a_4 = 4, a_5 = 5\}$.

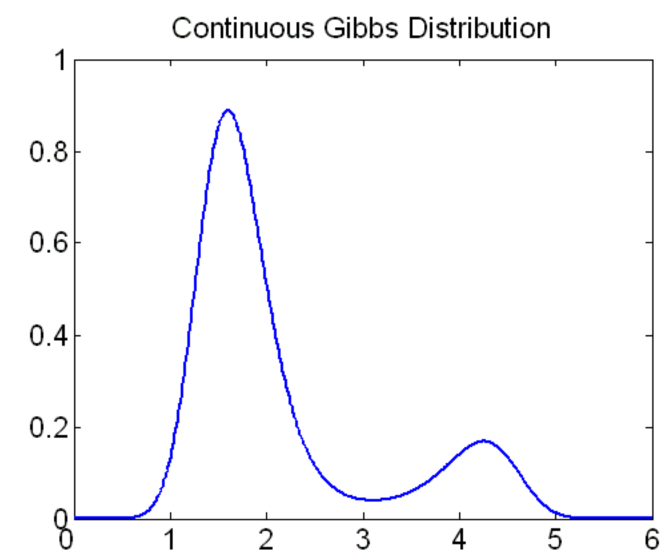
Noise

We set the noise strength $\beta = 0.8$

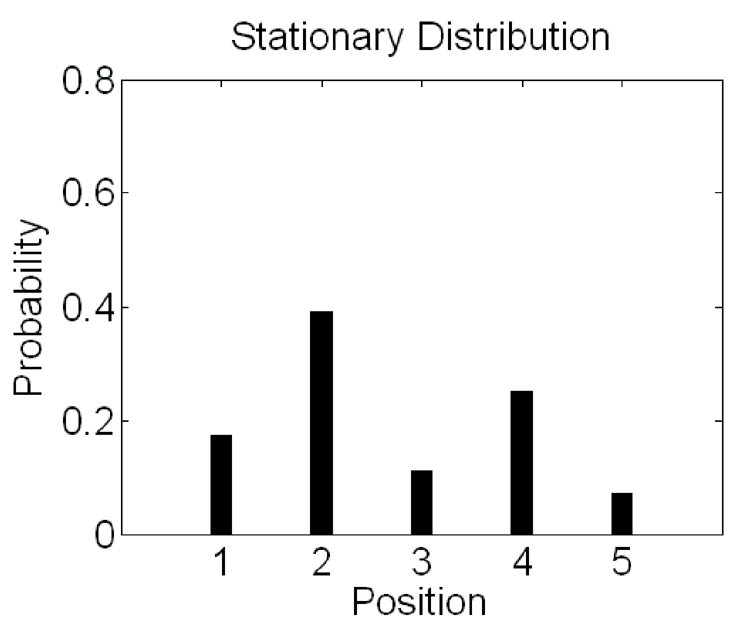
Motivation: A toy example

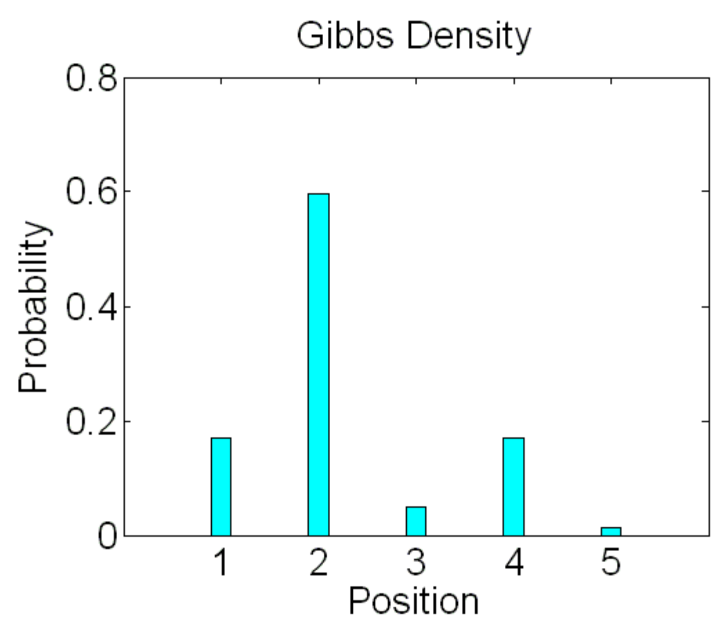
Continuous





Discrete (central- difference scheme): Wrong answer





Motivation

Challenges:

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

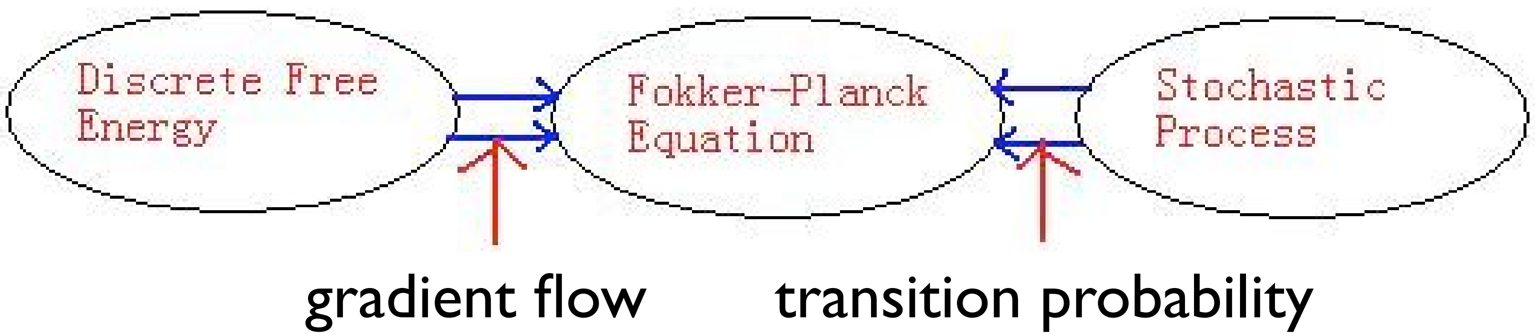
  Theorem: Any given linear discretization of the continuous equation can be written as
  $$\frac{d\rho_i}{dt} = \sum_j ((\sum_k e^i_{jk}\Phi_k) + c^i_j)\rho_j.$$
  Let
  $$A = \{\Phi \in \mathbb{R}^N : \sum_j ((\sum_k e^i_{jk}\Phi_k) + c^i_j)e^{-\frac{\Phi_j}{\beta}} = 0\}.$$
  Then A is a zero measure 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

Derive Fokker-Planck equations on graphs in two different ways



New ideas:

- White Noise for Markov processes on Graphs,
- Upwind scheme,
- ODEs for Fokker-Planck equations,
- 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)
  - Heath, Kinderlehrer, Kowalczyk (2002) ; discrete and continuous ratchets
  - Parrondo paradox (1996); review article by Harmer, Abbott (2002)
  - Carlen, Gangbo (2003); nonlinear Fokker Planck, constrained gradient flow

Our Strategies

Approach the problem from two different ways

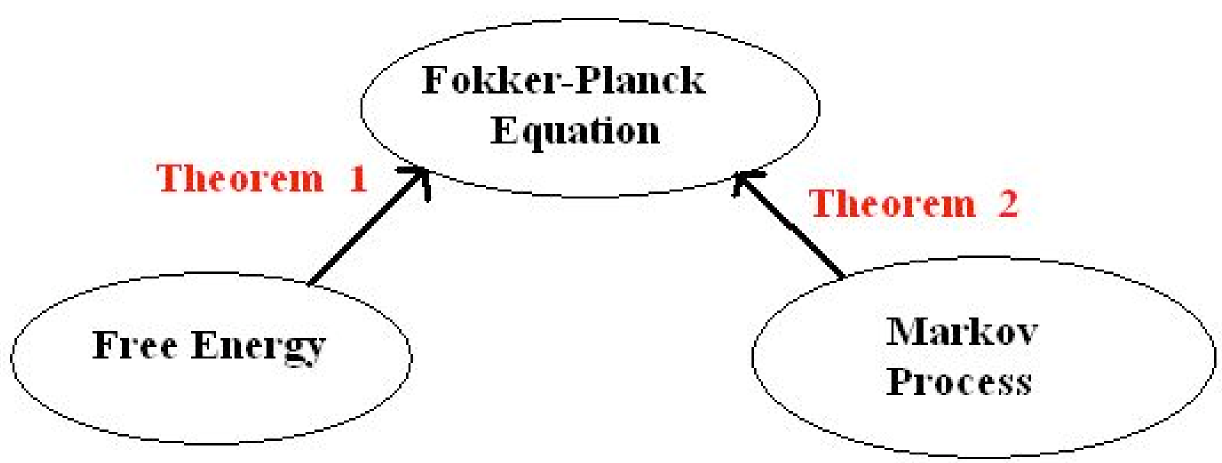
- Define Riemannian Manifold $(\mathcal{M}, d)$

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

Fokker-Planck equation is the gradient flow of the free energy on the manifold.

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

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



Main Results

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

Theorem 1

we have a Fokker-Planck equation

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

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 $\beta > 0$

Theorem II

The transition probability density function of the perturbed Markov process satisfies the following "Fokker-Planck" equation

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

where $\bar{\Psi}_i = \Psi_i + \beta \log \rho_i$

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

Properties of the Equations

The equations in both Theorem I and Theorem II have the following common properties

- Gibbs density ρ^* is a stable stationary solution and is the global minimum of free energy

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

- Given any initial data $\rho_0 \in \mathcal{M}$, we have a unique solution $\rho(t) : [0, \infty) \to \mathcal{M}$ with initial value ρ_0, and

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

- The boundary of $\mathcal{M}$ is a repeller

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 $(\mathcal{M}, d)$.

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

$$\text{difficulties: how to define } d?$$

- 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

Ideas of proof

Continuous case

- $$\frac{dx}{dt} = -\nabla\Psi(x)$$
- $$\frac{dx}{dt} = -\nabla\Psi(x) + \sqrt{2\beta}dW_t$$
- $$\frac{\partial\rho}{\partial t} = \nabla\cdot(\nabla\Psi\rho) + \beta\Delta\rho$$

Discrete case

- "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

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

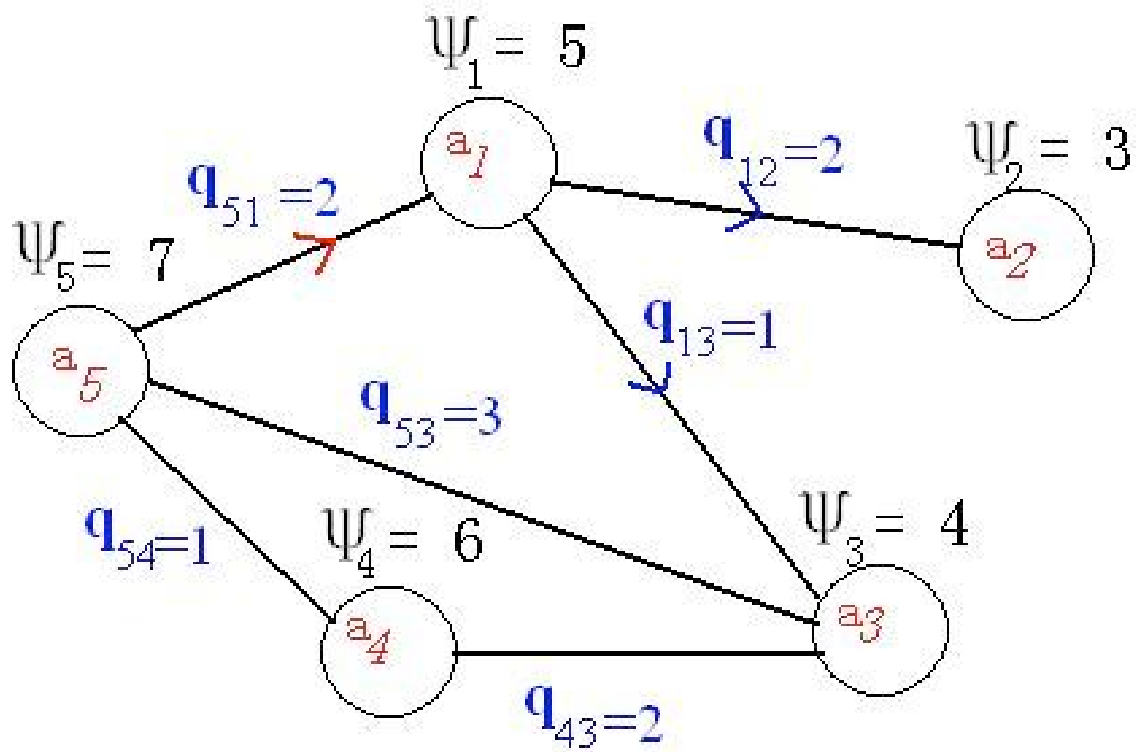
New potential $\bar{\Psi} = (\Psi + \beta \log \rho)$

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

Markov Process on Graphs

“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 $\Psi_i - \Psi_j$.



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)

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

Laplace Operator on a Graph

Laplace operator for a positive function ρ 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

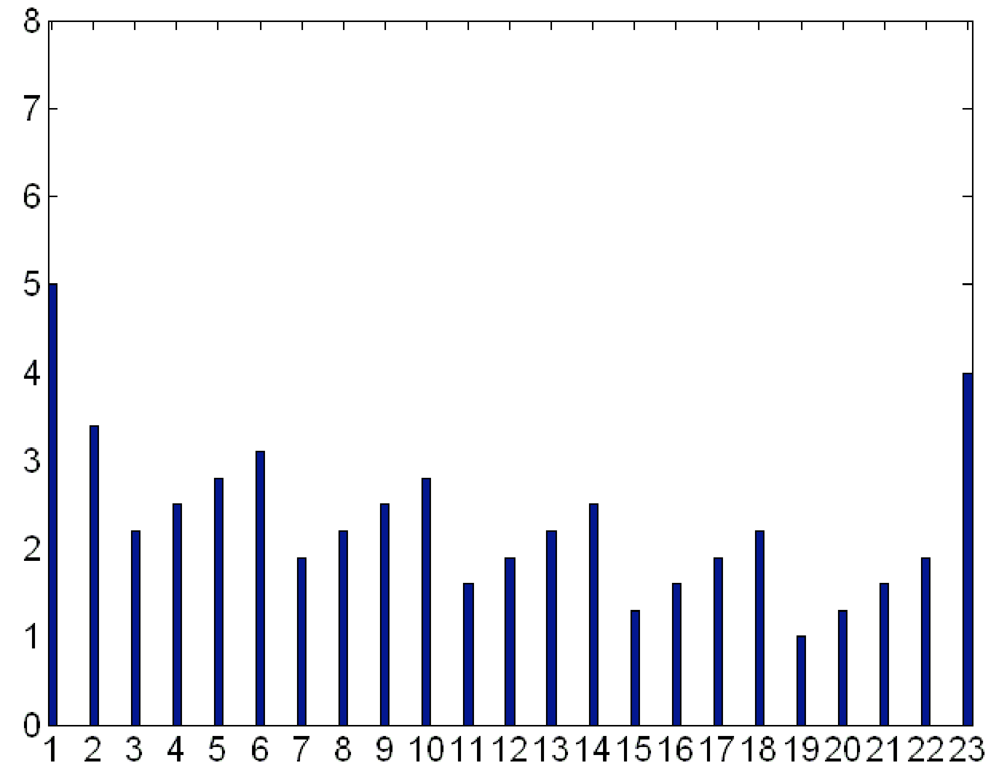
A: randomly perturbed gradient flow of the potential function, governed by Fokker-Planck equation in Theorem I,

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

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

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

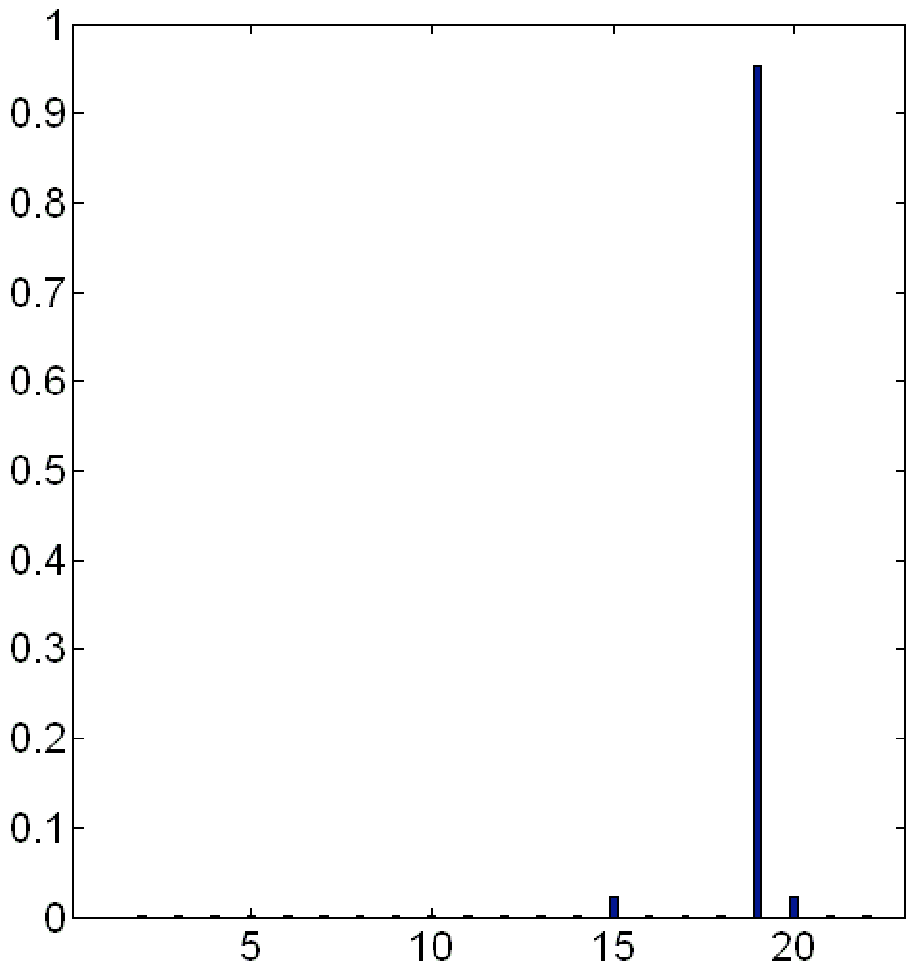
Discrete Potential



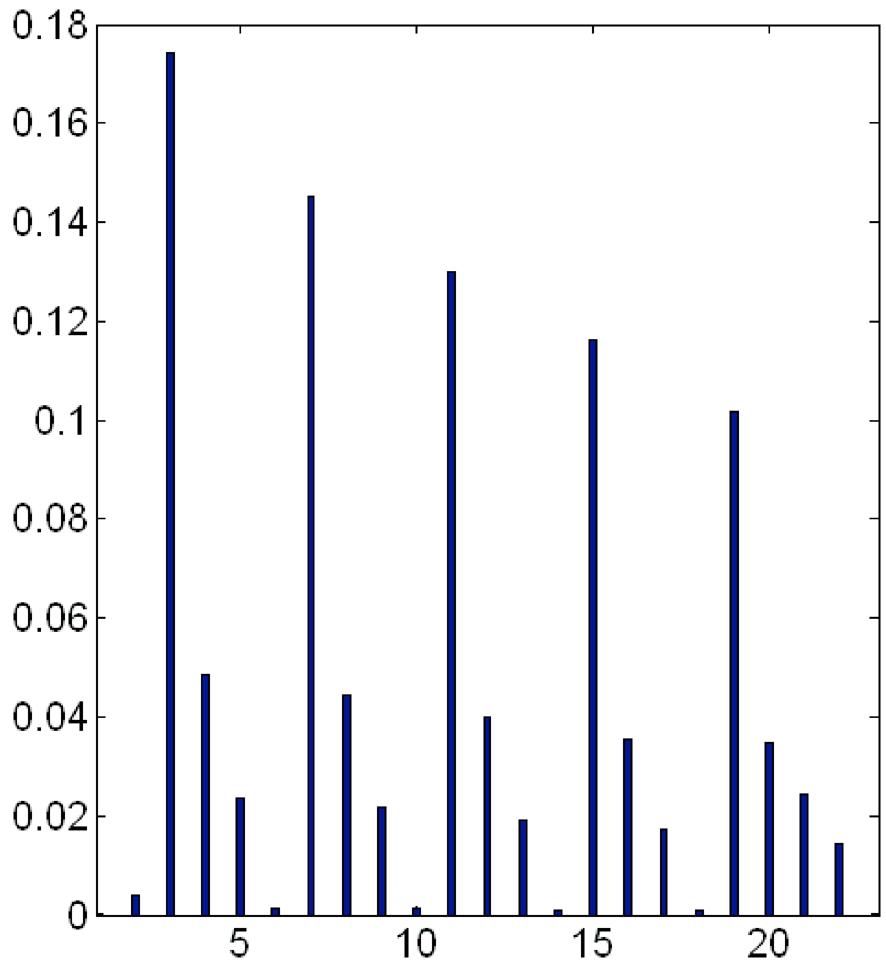
In both cases, the probability density function moves to the right (lower energy states).

Flashing Racket

Motion from lower potential to higher potential



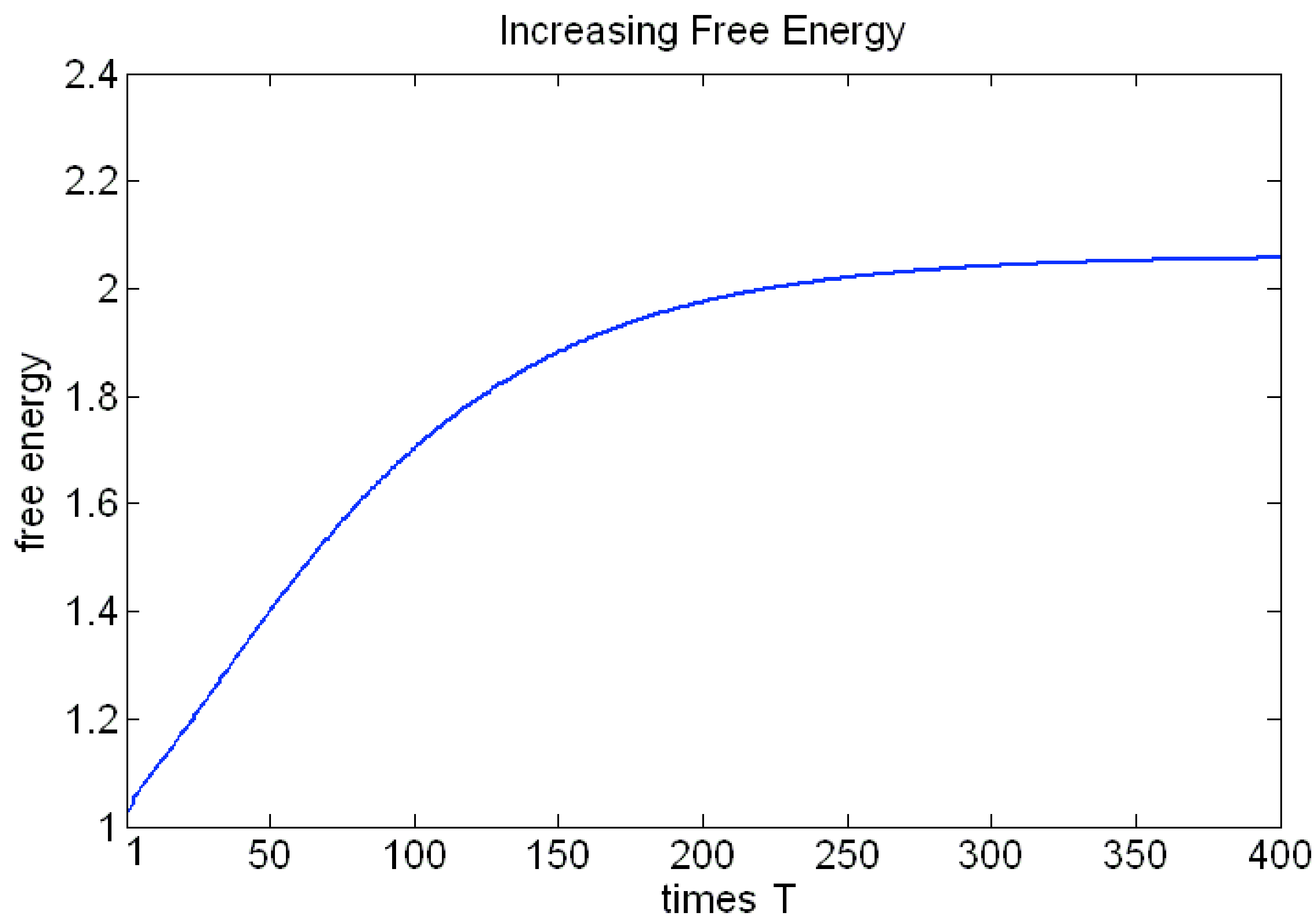
Initial distribution



Final distribution

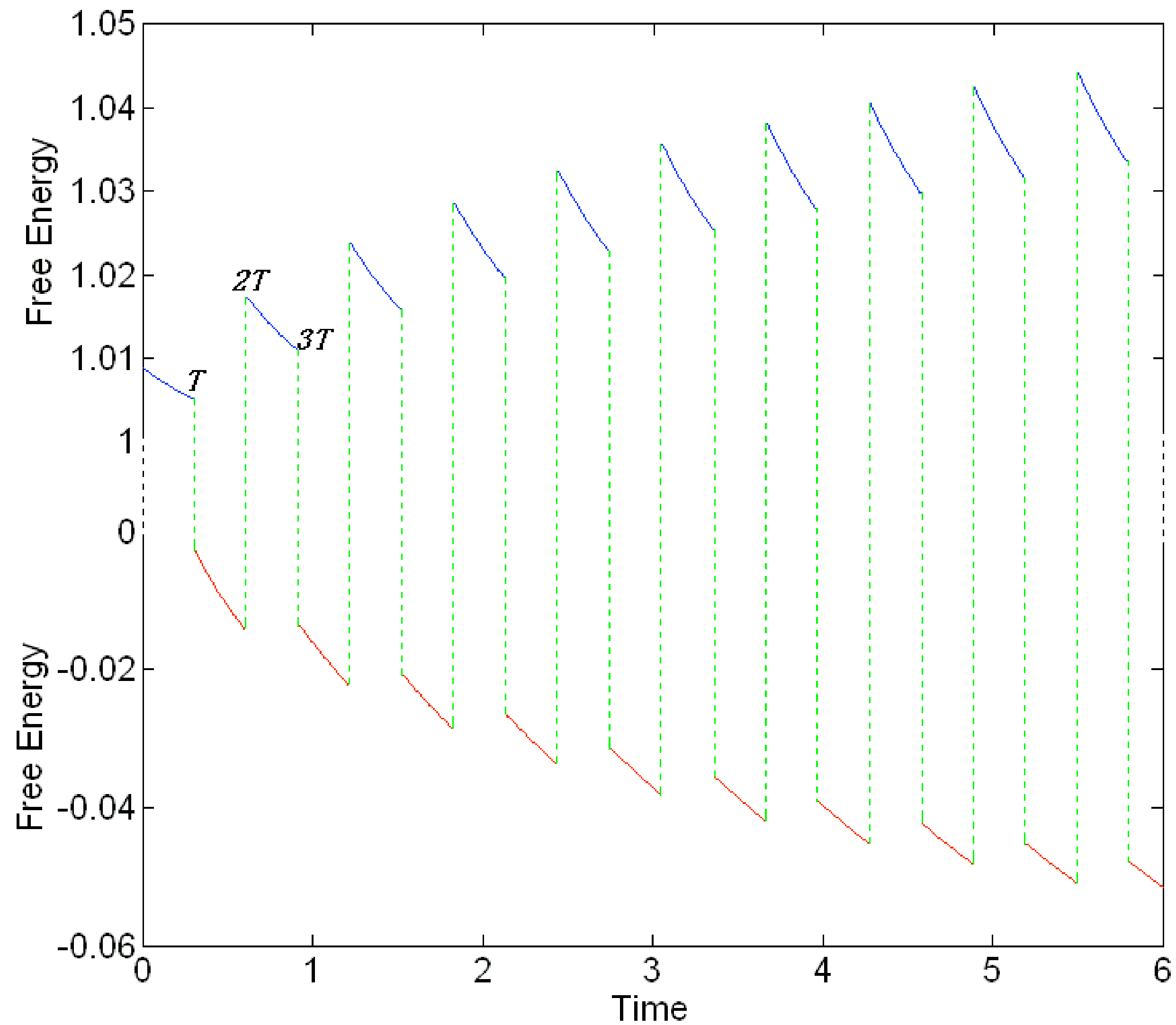
Flashing Ratchet

Use A, B alternatively as: ABABAB ...



Flashing Ratchet

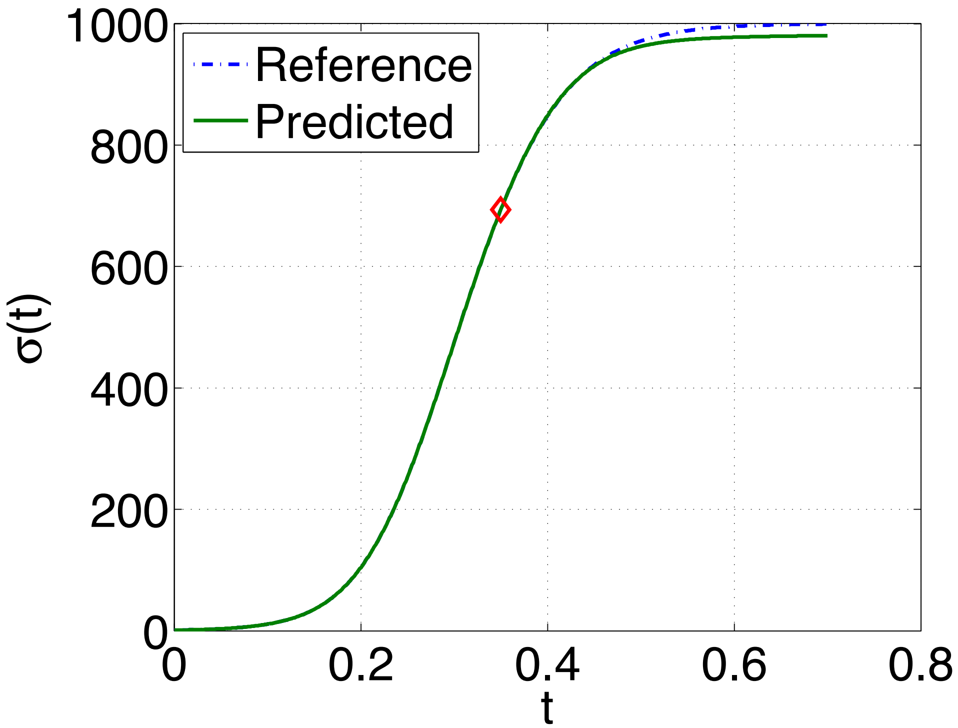
Free energy plot of the first 10 steps

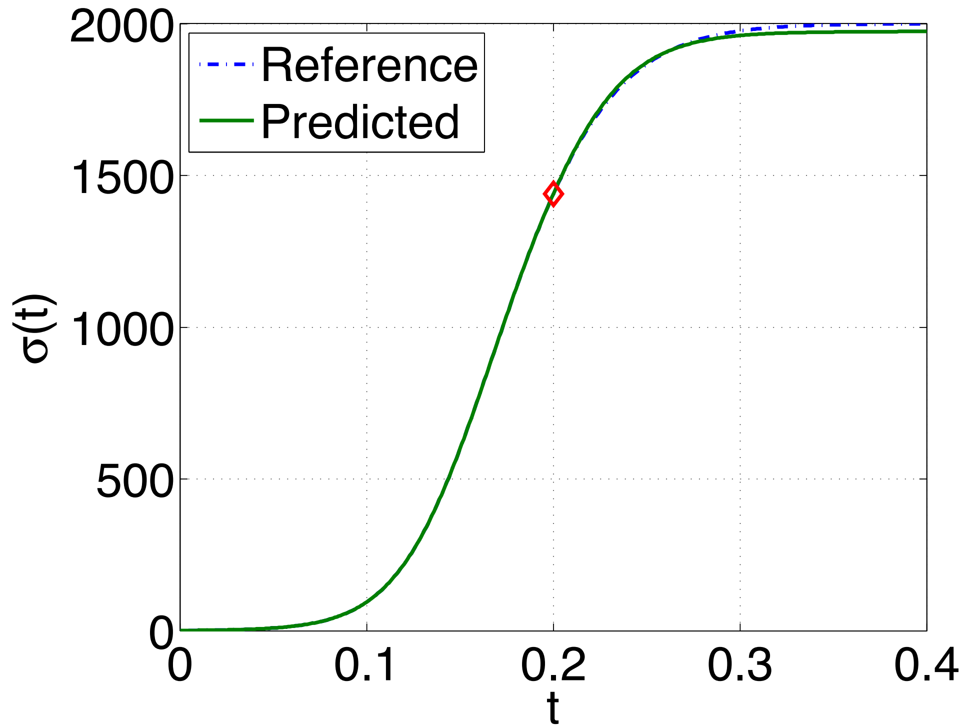


Influence Predictions in Networks

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 !