A STUDY ON STOCHASTIC DIFFERENTIAL EQUATIONS AND FOKKER-PLANCK EQUATIONS WITH APPLICATIONS

A Thesis
Presented to
The Academic Faculty

by

Wuchen Li

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology
May 2016

A STUDY ON STOCHASTIC DIFFERENTIAL EQUATIONS AND FOKKER-PLANCK EQUATIONS WITH APPLICATIONS

Approved by:

Professor Shui-Nee Chow
School of Mathematics
Georgia Institute of Technology

Professor Luca Dieci
Committee chair, Advisor, School of Mathematics
Georgia Institute of Technology

Professor Magnus Egerstedt
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Wilfrid Gangbo
School of Mathematics
Georgia Institute of Technology

Professor Haomin Zhou
Advisor, School of Mathematics
Georgia Institute of Technology

Date Approved: March 31, 2016

To

My Parents and *wife Siyu*

ACKNOWLEDGEMENTS

When I came to Georgia tech on August 8th, 2011, I would not imagine to be able to visit the beautiful applied mathematics world as I have seen today. I could not finish this thesis without the help and support of people around me.

Firstly, I express my deepest gratitude to my advisor, Professor Haomin Zhou. I'm grateful to him for helping me figure out the research directions and leading me into the exciting applied mathematics world. His meticulous caring and support make my five years PhD life productive and interesting. I thank him for being a mentor not only in mathematics, but also in life. I can not sufficiently express my thanks.

Many thanks go to my co-advisor, Professor Luca Dieci. I am deeply impressed by his optimistic attitude towards difficulties, and often inspired by his thoughtful comments in thousands of discussions. He provides me continuing support during every aspects of my PhD life. Both of my advisors are best mentors in the world. How lucky I am!

I am also appreciative of Professor Shui-Nee Chow. I am indebted to him for many sharp perspectives discussions. He is the one who can always see the big picture. Every judgement he make turns out to be true, which is stated in this thesis. I'm grateful to him for giving me uncountable encouragements and incredibly useful papers.

I would like to thank Professor Wilfrid Gangbo and Professor Magnus Egerstedt for many insightful discussions and enjoyable collaborations. These discussions and collaborations not only give me opportunities to see the beautiful work in pure mathematics, but also help me understand the real-world problems.

I also want to thank my fellow graduate students and friends. Special thanks to

Yin Ke, Guang Zhe, Weizhe Zhang, Jun Lu, Jing Hu, Haoyan Zhou, Yunlong He, Ruidong Wang, Tongzhou Chen, Xiaojing Ye and Lei Zhang for years of friendship and many interesting discussions, both in mathematics and life.

Lastly, I want to thank my parents and wife Siyu for their constant caring and support. I will treasure every moment in this 5 years journey.

<div align="right">Written in Sugar Hill, February 21, 2016.</div>

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

Fokker-Planck equations, along with stochastic differential equations, play vital roles in physics, population modeling, game theory and optimization (finite dimensional or infinite dimensional). In this thesis, we study three topics connected to them, both theoretically and computationally.

(1) Optimal transport on finite graphs [34, 36, 38];

(2) Numerical algorithms for constrained optimal control [35, 37, 62];

(3) Analysis of stochastic oscillators [42].

This thesis is arranged as follows:

Chapter II gives the necessary mathematics background, which contains a brief survey of Fokker-Planck equations, gradient flows, optimal control and optimal transport. Through them, we design and analyze practical algorithms for real world problems.

Chapter III is the theoretical **heart** of the thesis. In recent years, optimal transport has been considered by many authors and is essential in geometry and partial differential equations. We consider a similar setting for discrete states which are on a finite but arbitrary graph. By defining a discrete 2-Wasserstein metric, we derive gradient flows of discrete free energies. We name gradient flows as Fokker-Planck equations on graphs, which are ordinary differential equations. Furthermore, we obtain exponential convergence result for such gradient flows. This derivation provides tools for graphs' functional inequalities, "geometry" analysis of graphs, modeling in game theory (Chapter IV) and numerics for nonlinear partial differential equations (Chapter V).

Chapter VI is mainly on the computational part. It proposes a new algorithm, called method of evolving junctions (MEJ), to compute optimal solutions for a class of constrained optimal control problems. The main idea is that through the geometric structures of optimal solutions, we convert the infinite dimensional minimization problem into finite dimensional optimizations. Then we apply the intermittent diffusion, a stochastic differential equation based global optimization method, to find the global optimal solution. By numerical examples, MEJ can effectively solve many problems in Robotics, including the optimal path planning problem in dynamical environments and differential games.

Chapter VII concerns on modeling problem for stochastic oscillator. We introduce a new type of noise for the stochastic van der Pol oscillator. We show that the perturbed solutions under this new noise are globally bounded. Furthermore, we derive a pair of Fokker-Planck equations for the new noise model.

CHAPTER I

INTRODUCTION

1.1 Optimal transport on finite graphs

Consider an infinite dimensional minimization problem in kinectic mechanics [2, 94, 95]

$$\min_{\rho(x)} \int_{\mathbb{R}^d} V(x)\rho(x)dx + \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} W(x,y)\rho(x)\rho(y)dxdy + \beta\int_{\mathbb{R}^d}\rho(x)\log\rho(x)dx$$

subject to

$$\int_{\mathbb{R}^d}\rho(x)dx = 1, \quad \rho(x) \geq 0, \quad x \in \mathbb{R}^d,$$

where the variable $\rho(x)$ is a probability density function supported on $\mathbb{R}^d$, $V : \mathbb{R}^d \to \mathbb{R}$, $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ are given functions with $W(x,y) = W(y,x)$ for any $x, y \in \mathbb{R}^d$.

Recently, using the viewpoint of optimal transport, the above minimization problem has an interesting interpretation. Equipping the probability space $\mathcal{P}(\mathbb{R}^d)$ with the 2-Wasserstein metric, the gradient flow of the above objective functional (named free energy) forms a nonlinear partial differential equation (PDE), which is the Fokker-Planck equation [4, 58, 79]

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

Problem 1: *Can we establish a similar approach on discrete states?*

In details, we consider a simple finite graph $G = (V, E)$, where $V = \{1, 2, \cdots, n\}$ is the vertex set and E is the edge set. The graph G is associated with a probability set

$$\mathcal{P}(G) = \{(\rho_i)_{i=1}^n \mid \sum_{i=1}^n \rho_i = 1, \quad \rho_i \geq 0\},$$

and an objective function, named discrete free energy

$$\mathcal{F}(\rho) = \sum_{i=1}^n v_i\rho_i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n w_{ij}\rho_i\rho_j + \sum_{i=1}^n \rho_i\log\rho_i,$$

1

where $(v_i)_{i=1}^n$ is a constant vector and $(w_{ij})_{1 \leq i,j \leq n}$ is a given symmetric matrix.

By this setting, how can we derive the Fokker-Planck equation as the gradient flow of discrete free energy on probability set? In other words, consider the optimization

$$\min_\rho \sum_{i=1}^n v_i \rho_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \rho_i \rho_j + \sum_{i=1}^n \rho_i \log \rho_i \quad s.t. \quad \sum_{i=1}^n \rho_i = 1, \quad \rho_i \geq 0, \quad i \in V.$$

What is the gradient flow of the objective function associated with graph G's structure?

In the literature, the optimal transportation on discrete states is certainly not a new concept. Ollivier introduces a 1-Wasserstein metric [77], which can not be applied to Fokker-Planck equation directly; Erbar and Maas [43, 66], Mielke [71] consider a 2-Wasserstein metic on discrete states, which is essential to analyze linear Markov processes. However, if we consider the discrete interaction energy $\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \rho_i \rho_j$, all of above approaches can not answer problem 1 clearly.

In this thesis, motivated by [27], we fully understand problem 1 from the dynamical viewpoint. Based on a new "discrete 2-Wasserstein metric", we derive the gradient flow of $\mathcal{F}(\rho)$:

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} (F_j(\rho) - F_i(\rho))_+ \rho_j - \sum_{j \in N(i)} (F_i(\rho) - F_j(\rho))_+ \rho_i, \tag{1}$$

where $F_i(\rho) = \frac{\partial}{\partial \rho_i} \mathcal{F}(\rho)$ and $(h)_+ = \max\{h, 0\}$. We call (1) the Fokker-Planck equation on a finite graph.
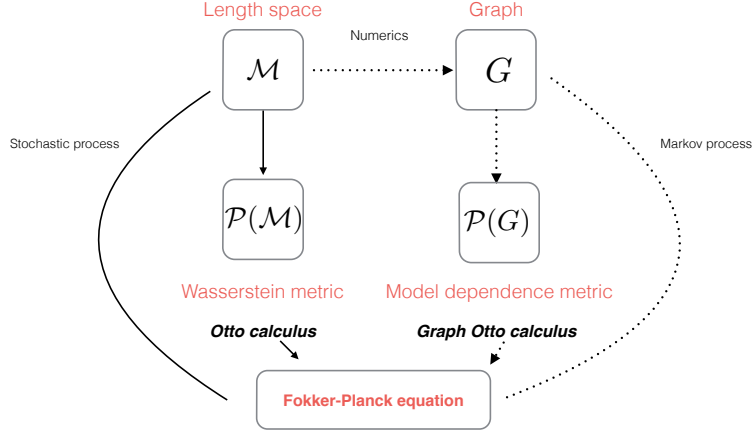
Figure 1: Derivation

There are many reasons why we say that (1) is a gradient flow: (i) $\mathcal{F}(\rho)$ is a Lyapunov function of (1),

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = -\sum_{(i,j)\in E}(F_i(\rho) - F_j(\rho))_+^2 \rho_i \leq 0;$$

(ii) The minimizers of $\mathcal{F}(\rho)$, discrete Gibbs measures[1], are equilibria of (1). Thus, a natural question arises: if ρ^0 converges to a strict local minimizer ρ^∞, how fast is the convergence?

Problem 2: *Can we analyze the convergence speed of* (1) *to a Gibbs measure?*

From differential geometry, we know that the convergence rate of gradient flow depends on the Hessian operator of the free energy $\mathcal{F}(\rho)$ on manifold. Unfortunately,

---

[1]ρ^* is a discrete Gibbs measure, if it solves the fixed point problem:

$$\rho_i^* = \frac{1}{K}e^{-\frac{v_i + \sum_{j=1}^n w_{ij}\rho_j^*}{\beta}}, \quad \text{where} \quad K = \sum_{i=1}^n e^{-\frac{v_i + \sum_{j=1}^n w_{ij}\rho_j^*}{\beta}}.$$

this is an open problem in continuous states[2]. For a special choice of interaction potential $W(x,y) := W(x-y)$, Carrillo, McCann and Villani [23] discover a nice formula for the Hessian operator, from which they prove an exponential convergence result.

In this thesis, motivated by [23], Villani's open problem and the dynamical viewpoint, we fully solve the above problem 2. We derive a formula, which plays the role of Hessian matrix at discrete Gibbs measures w.r.t discrete 2-Wasserstein metric on finite graphs.

Definition 1 *Let* $f_{ij} := \frac{\partial^2}{\partial \rho_i \partial \rho_j} \mathcal{F}(\rho)$ *and*

$$h_{ij,kl} := f_{ik} + f_{jl} - f_{il} - f_{jk} \quad \text{for any } i, j, k, l \in V.$$

We define

$$\lambda_{\mathcal{F}}(\rho) := \min_{(\Phi_i)_{i=1}^n} \sum_{(i,j)\in E} \sum_{(k,l)\in E} h_{ij,kl} (\Phi_i - \Phi_j)_+ \rho_i (\Phi_k - \Phi_l)_+ \rho_k$$

s.t.

$$\sum_{(i,j)\in E} (\Phi_i - \Phi_j)_+^2 \rho_i = 1.$$

Based on Definition 1, we show that if the Gibbs measure ρ^∞ is a strictly local minimizer of free energy and ρ^0 is in the attraction region of ρ^∞, then

$$\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty) \leq e^{-Ct}(\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)), \tag{2}$$

where C is a positive constant depending on ρ^0 and graph G. Moreover, we prove that the discrete free energy decreases exponentially with asymptotic dissipation rate $2\lambda_{\mathcal{F}}(\rho^\infty)$.

We address problem 1 and 2 as follows. In chapter 3, we discuss mainly the theoretical derivation and convergence result of (1). Besides those, we study the following problems.

---

[2]See Villani's open problem 15.11 in [95]: Find a **nice** formula for the Hessian of the functional $\mathcal{F}(\rho)$.

4

- In section 3.2.4, we show the Hodge decomposition of discrete 2-Wasserstein metric; In section 3.5, we use the convergence result to prove several functional inequalities on finite graphs.

- In section 3.8, motived by the convergence result and Villani's open problem, we derive a formula for Hessian operator of free energy at the Gibbs measure in continuous states.

Furthermore, we demonstrate that (1) provides tools for many applications.

- In chapter 4, we derive new evolutionary dynamics in game theory;

- In chapter 5, we introduce new numerical schemes for a certain type of nonlinear PDEs, including nonlinear Fokker-Planck equations and nonlinear diffusion equations.

1.2   A new algorithm for constrained optimal control

Consider the other type of infinite dimensional optimization problem

$$\min_{x,u} \int_0^T L(x(t), u(t), t)dt + \psi(x(T), T), \tag{3}$$

where the state, control variable $x(t)$, $u(t)$ are subject to a dynamic system and phase, control constraints

$$\dot{x} = f(x(t), u(t), t), \quad t \in [0, T]; \quad x(0) = x_0, \quad x(T) = x_T;$$

$$\phi(x(t), t) \geq 0, \quad \varphi(u(t), t) \geq 0, \quad t \in [0, T].$$

Many engineering problems, including path-planning problem in Robotics, can be formulated into the framework of (3), known as optimal control problems with constraints. Because the complexity of those applications, few of them can be solved analytically. Thus numerical methods are often employed. Traditionally, the methods are divided into three categories, (1) state-space (Hamilton-Jacobi-Bellman equations) [12, 74]; (2) indirect (Pontryagin's Maximum Principle) [9, 21, 64, 70]; (3)

direct methods (Nonlinear programming) [40, 45, 49, 75]. However, there are some well-known limitations of the above three general methods. Namely, HJB approach, which gives the global solution, can be computationally expensive and suffers from the notorious problem known as "curse of dimensionality". Indirect methods, which find local optimal solutions, is numerically painful to handle constraints; Direct methods require finer discretization (smaller time steps) if better accuracy is expected, and this leads higher computational cost.

Problem 3: *Can we derive a new fast alogrthim for* (3)*?*

In this thesis, for a special class of (3), we design a new method, called Method of Evolving Junctions (MEJ), to find the global optimal path. MEJ is built on the following facts. All local and global optimal paths share a similar geometric structure called **Separable**, meaning the path can be partitioned into a finite number of segments, on which the constraints are either active or inactive. There is no switching between active and inactive inside each segment. We call the partition points junctions. Using those junctions, we can reduce the optimal control to a finite dimensional optimization. Such a reduction allows us to find global solution(s) by SDEs with given initial values.

In chapter 6, we demonstrate MEJ though several numerical examples, including linear quadratic controls, optimal path planning problems in dynamical environments, shortest path problems and differential games.

1.3 Analysis for stochastic oscillators

Classical theories [8, 46] predict that solutions of differential equations will leave any neighborhood of a stable limit cycle, if white noise is added to the system. In reality, many engineering systems modeled by second order differential equations, like the

van der Pol oscillator

$$\begin{cases} dx = ydt \\ dy = [\alpha(1 - x^2)y - x]dt + \epsilon dW_t \end{cases}$$

show incredible robustness against noise perturbations, and the perturbed trajectories remain in the neighborhood of a stable limit cycle for all times of practical interest [26].

Problem 4: Can we propose a new model of noise to bridge this discrepancy between theory and practice?

In this thesis, we introduce a new model: The key is to consider a new event set:

$$B = \{ \omega \mid \sup_{|t-s| \leq T} |W_t(\omega) - W_s(\omega)| \leq M \} .$$

where T and M are two given positive constants, t and s are any two instants of time at most T-apart. Restricting to perturbations within this new class of noise, we consider stochastic perturbations of second order differential systems that –in the unperturbed case– admit asymptotically stable limit cycles. We show that the perturbed solutions are globally bounded and remain in a tubular neighborhood of the underlying deterministic periodic orbit. In addition, we define stochastic Poincaré map(s), and further derive Fokker-Planck equations under the new noise. We show all these results in chapter 7.

CHAPTER II

PRELIMINARY IN MATHEMATICS

In this chapter, we briefly introduce the mathematics needed in this thesis, which contains Fokker-Planck equations, gradient flows, optimal control and optimal transportation. These are highly related topics, which are widely used in applied mathematics. Based on them, we design algorithms for real world problems. To simply the illustration, we don't address on their regularity issues and just perform formal calculations.

2.1 Fokker-Planck equations

This thesis mainly focus on the Fokker-Planck equation, which is basic in many subjects, including probability, physics, and modeling. It has the form

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (f(x)\rho) = \beta \nabla \cdot (AA^T \nabla \rho), \quad x \in \mathbb{R}^d, \tag{4}$$

where $AA^T = A(x)A(x)^T$ is a nonnegative definite (diffusion) matrix and $f(x) \in \mathbb{R}^d$ is a (drift) vector function on x. Here the unknown $\rho(t, x)$ is a probability density function for given time t, which keeps non-negativity and conserves the total probability.

Underlying (4) is the stochastic differential equation

$$dX_t = f(X_t)dt + \sqrt{2\beta}A(X_t)dW_t, \quad X_t \in \mathbb{R}^d,$$

where W_t is a standard Wiener process (Brownian motion). The Fokker-Planck equation describes the evolution of the transition probability of Markov process X_t. Here

$$\rho(t, x)dx := \Pr(X_t \in dx | X_0); \quad \forall x \in \mathbb{R}^d, \ t > 0 \ .$$

The underlying state Ω of the Fokker-Planck equation can be a variety other than $\mathbb{R}^d$. For example, Ω can be a bounded open set, where the boundary can be handled with either zero-flux condition or periodicity conditions; Ω can aslo be equipped with a differential structure, such as the Riemannian manifold.

2.2 Gradient flows

Gradient flows are fundamental evolutionary systems associated with (finite or infinite dimensional) optimization problems, which provide the basis for numerical intituation methods, known as the gradient descent method. In this thesis, we will discuss this concept a lot.

Connections with optimization Consider an optimization problem

$$\min_{x \in \mathbb{R}^d} V(x),$$

where $V(x) \in C^2(\mathbb{R}^d)$ is called energy (objective) function. A natural way to solve this minimization is through the gradient flow

$$dx_t = -\nabla V(x_t)dt. \tag{5}$$

Notice that $V(x)$ is a Lyapunov function of (5).

$$\frac{d}{dt}V(x_t) = -(\nabla V(x_t), \nabla V(x_t)) \leq 0.$$

So if $V(x)$ is a strictly convex function, then the gradient flow (5) converges to the minimizer. In numerical methods, the steepest descent method arises from this property.

But, things are not always perfect. In applications, one often wishes to find the global minimizer of a non-convex energy function. Now, even assuming that a global minimizer exists, can we guarantee a numerical method finding it? Unfortunately, there is no way obtaining a global minimizers other than by comparison of all local ones.

However, the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} - \nabla \cdot (\rho \nabla V(x)) = \beta \Delta \rho \tag{6}$$

connects this question by using the probability. This connection can be understood at two levels. On one hand, the SDE associated with (6) is a gradient flow with stochastic perturbation

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta}dW_t.$$

The solution of this SDE has a positive probability to jump out of any basins of attraction of local minimizers; On the other hand, the equilibrium of (6) ($\frac{\partial \rho}{\partial t} = 0$), named Gibbs measure, has an explicit formula

$$\rho^* = \frac{1}{K}e^{-\frac{V(x)}{\beta}}, \quad K = \int_{\mathbb{R}^d} e^{-\frac{V(x)}{\beta}} dx.$$

The asymptotic ($\beta \to 0$) behavior of the Gibbs measure is Dirac mass, which is concentrated at global minimizers. Based on the above two hints, people have designed many global optimization techniques.

What's more, there is an intrinsic connection between (6) and the optimization problem. That is the PDE (6) is a gradient flow of a stochastic optimization in "probability manifold"[1]. Here the stochastic optimization means

$$\min_{\text{r.v. } X \in \mathbb{R}^d} \quad \mathbb{E}V(X) + \beta \mathcal{H}(X)$$

where X is a random variable with probability density function $\rho(x) \in \mathcal{P}(\mathbb{R}^d)$ and

$$\mathbb{E}V(X) = \int_{\mathbb{R}^d} V(x)\rho(x)dx, \quad \mathcal{H}(X) = \int_{\mathbb{R}^d} \rho(x) \log \rho(x)dx.$$

Riemannian structure To understand this intrinsic connection, we review the definition of gradient flows in a finite dimensional smooth Riemannian manifold $(\mathcal{M}, g)$, where g defines a scalar inner product on the tangent space $T_x\mathcal{M}$ with $x \in \mathcal{M}$.

---

[1]Probability set $\mathcal{P}(\mathbb{R}^d)$ with a 2-Wasserstein metric.

The differential structure induces gradient flows, since the inner product can identify the gradient of energy $V : \mathcal{M} \to \mathbb{R}$ on the manifold via two formulas

- Tangency condition:

$$\text{grad}_{\mathcal{M}} V(x) \in T_x \mathcal{M};$$

- Duality condition:

$$g(\text{grad}_{\mathcal{M}} V(x), \sigma) = dV(x) \cdot \sigma,$$

where $\sigma \in T_x \mathcal{M}$, $dV(x)$ is the differential of $V(x)$ and the dot in R.H.S means the direction of derivative along σ.

Then the gradient flow of V on $(\mathcal{M}, g)$ forms

$$dx_t = -\text{grad}_{\mathcal{M}} V(x_t) dt.$$

2.3 Optimal control

Optimal control is an infinite dimensional optimization problem with applications in physics and engineering. It seeks to determine the input (control) to a dynamical system that optimizes a given performance functional with certain constraints. We briefly illustrate several techniques for solving it via a special example.

Consider an optimal control problem

$$\inf_{v} \{ \int_0^1 L(t, x, v) dt \ : \ \dot{x}(t) = v(t), \quad x(0) = x_0, \quad x(1) = x_1 \},$$

where the minimizer is among all smooth curves, $x \ : \ [0, 1] \to \mathbb{R}^d$, and the running cost $L \ : \ [0, 1] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is smooth, which is called Lagrangian. Denote $\mathcal{J}(x) = \int_0^1 L(t, x, v) dt$ as the cost functional.

There are two angles to solve this problem. The first angle is calculus of variation, through which we obtain an ODE system for a local minimizer. Let x be a minimizer and $h(t)$ be an arbitrary function with $h(0) = h(1) = 0$. Substituting $x_\epsilon(t) = x(t) +$

$\epsilon h(t)$ into the cost functional, $\mathcal{J}(x_\epsilon)$ becomes a function of ϵ, whose stationary point is at $\epsilon = 0$.

$$0 = \frac{d}{d\epsilon}\mathcal{J}(x_\epsilon)|_{\epsilon=0} = \int_0^1 \left(\nabla_x L(t, x, v) \cdot h(t) + \nabla_v L(t, x, v) \cdot \dot{h}(t)\right)dt$$

$$= \int_0^1 \left(\nabla_x L(t, x, v) - \frac{d}{dt}\nabla_v L(t, x, v)\right) \cdot h(t)dt.$$

Since h is an arbitrary function, the minimizer x satisfies

$$\nabla_x L(t, x, v) - \frac{d}{dt}\nabla_v L(t, x, v) = 0. \tag{7}$$

(7) is known as the **Euler-Lagrange equation**. The L.H.S. of (7) is also called the **first variation formula** of $\mathcal{J}$ with respect to x at point t, denoted as $\frac{\delta}{\delta x(t)}J(x)$.

Moreover, we can transfer the Euler-Lagrange equation into a **Hamiltonian system**, which enables to access its rich mathematical structure. The Hamiltonian H is constructed via the Legendre transform:

$$p = \nabla_v L(t, x, v), \quad H(t, p, x) = \sup_{v \in \mathbb{R}^d}(p \cdot v - L(t, x, v)).$$

Through it, (7) is equivalent to

$$\dot{p}(t) = -\nabla_x H(t, p(t), x(t))$$
$$\dot{x}(t) = \nabla_p H(t, p(t), x(t)). \tag{8}$$

(8) is a special case of Pontryagin maximum principle.

The second angle is dynamic programming, from which we derive a nonlinear partial differential equation, named **Hamilton-Jacobi** equation, for the global minimizer. Define an optimal cost-to-go function

$$\Phi(t, x) = \inf_v\{\int_0^t L(t, x(s), v(s))dt : \quad x(t) = x, \quad x(0) = x_0\}.$$

By Bellman's principle of optimality, the sub-arc of optimal path is also optimal. Going from time $t - \Delta t$ to t,

$$\Phi(t, x) = \inf_v\{\int_{t-\Delta t}^t L(t, x(s), v(s))dt + \Phi(t - \Delta t, x(t - \Delta t))\}$$

$$= \inf_v\{L(t, x, v)\Delta t + \Phi(t, x) - \frac{\partial \Phi}{\partial t}\Delta t - \nabla_x \Phi \cdot v\Delta t\} + o(\Delta t)$$

12

where the second equality is by Taylor expansion. If we cancel out $\Phi(t, x)$, divide Δt and let $\Delta t \to 0$, then we obtain an equation

$$\frac{\partial \Phi}{\partial t} + \sup_v \{\nabla_x \Phi \cdot v - L\} = 0. \tag{9}$$

Notice that (9) is a special case of Hamilton-Jacobi-Bellman equation (HJB).

We illustrate these two viewpoints by a simple Lagrangian, $L(t, x, v) = v^2$. Then the Euler-Lagrange equation (7) becomes

$$\dot{x}(t) = v(t), \quad \dot{v}(t) = 0;$$

While the Hamilton-Jacobi equation forms

$$\begin{cases} \dot{x}(t) = \nabla_x \Phi(t, x), \\ \frac{\partial \Phi}{\partial t} + \frac{1}{2}(\nabla \Phi)^2 = 0. \end{cases} \tag{10}$$

Notice that the above two system give the same solution: $x(t) = x_0 + \frac{d(x_0, x_1)}{t_1 - t_0}$ and $\Phi(t, x) = \frac{d(x_0, x)}{t}$, where d is the Euclidean distance.

2.4 Optimal transport

Optimal transport is the other infinite dimensional optimization problem. It seeks to find the optimal cost functional between two probability measures. Nowadays, it provides more and more powerful tools in both pure and applied mathematics.

The problem[2] is to find the optimal transportation plan between two probability measures

$$C(\mu, \nu) = \inf_\pi \{ \int c(x, y) d\pi(x, y) \; : \; \pi \text{ with marginals } \mu \text{ and } \nu \} \tag{11}$$

where $c(x, y)$ is the cost function for transporting one unit of mass from x to y and μ, ν are two probability measures supported on continuous states. The continuous

---

[2]Kantorovich's formulation

state can be a variety, including $\mathbb{R}^d$, Riemannian manifold, and a metric space. For concreteness, we use $\mathbb{R}^d$ to illustrate.

Before introducing the associated theory, we illustrate (11) in its "discrete" version. The optimal transportation between discrete measures is a linear programing problem

$$\min_{\pi_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \pi_{ij} \quad \text{s.t.} \quad \sum_{i=1}^{n} \pi_{ij} = \mu_i, \quad \sum_{j=1}^{n} \pi_{ij} = \nu_i, \quad \pi_{ij} \geq 0. \tag{12}$$

From the knowledge of optimization, (12)'s dual problem becomes

$$\max_{\Phi, \Psi} \sum_{i=1}^{n} \Psi_i \nu_i - \sum_{j=1}^{n} \Phi_j \mu_j \quad \text{s.t.} \quad \Psi_i - \Phi_j \leq c_{ij}.$$

So it is not hard to guess that the duality problem for (11)

$$C(\mu, \nu) = \sup_{\Psi, \Phi} \{ \int \Psi d\nu - \int \Phi d\mu \ : \ \Psi(x) - \Phi(y) \leq c(x, y) \}.$$

The mathematical structure of the dual problem paves enough regularities to show the existence of (11)'s minimizer. But there is no end. Duality provides more insight for the original problem (12). We explain this by a special cost function, which is a square distance function

$$c(x, y) = d(x, y)^2,$$

where $d(x, y)$ is the Euclidean distance.

2.4.1   2-Wasserstein metric

Under this cost function,

$$C(\mu, \nu) = \inf_{\pi} \{ \int d(x, y)^2 \pi(x, y) \ : \ \pi \text{ with marginals } \mu \text{ and } \nu \}.$$

$W_2(\mu, \nu) := \sqrt{C(\mu, \nu)}$ introduces a metric on the probability set $\mathcal{P}(\mathbb{R}^d)$. Unlike the "discrete" problem (12), the continuous underlying space provides a way to rewrite the transport problem (11) into a time-dependent version.

We explain this by one simple observation: The distance between two Dirac measures (point measure) is the two points' Euclidean distance. Recall that the Euclidean distance has a a time-dependent version,

$$d(x, y)^2 = \inf\{\int_0^1 v(t)^2 dt \ : \ \dot{x}(t) = v(t), \ x(0) = x, \ x(1) = y\},$$

whose solution satisfies (10).

Can we generalize more from the observation of Dirac measures? The answer is yes. "The geodesic in law space is the law of geodesic in underlying spaces"[3]. More precisely, 2-Wasserstein metric has a time-dependent version, which satisfies

$$W_2(\mu, \nu)^2 = \inf\{\int_0^1 \int v^2 \rho \, dx dt \ : \ \frac{\partial \rho_t}{\partial t} + \nabla \cdot (v\rho) = 0, \ \rho_0 \sim \mu, \ \rho_1 \sim \nu\}, \qquad (13)$$

where ρ_t represents the probability density function at time t. From the duality of (13) and (11), it is known that the statical and time dependence definitions are equivalent. Furthermore, the minimizer of (13) satisfies a pair of PDEs

$$\begin{cases} \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\nabla \Phi_t \rho_t) = 0 \\ \frac{\partial \Phi_t}{\partial t} + \frac{1}{2}(\nabla \Phi_t)^2 = 0 \end{cases} \qquad (14)$$

where $\rho_0 \sim \mu$, $\rho_1 \sim \nu$; the first equation is the continuity equation with velocity $\nabla \Phi_t$ and the second equation is a HJB equation, providing the equation for the velocities. (14) can be viewed as geodesic equations in the probability manifold, which connects measures μ and ν.

2.4.2 Otto calculus

The 2-Wasserstein metric provides a way to treat probability set $\mathcal{P}(\mathbb{R}^d)$ as an infinite dimensional "Riemannian" manifold. Through that, we can connect a certain type of Fokker-Planck equations as gradient flows of scalar functional (named free energies). The derivation is named as Otto calculus [95].

---

[3]Here the law means the probability measure.

Simply put, we identify the tangent space of $\mathcal{P}(\mathbb{R}^d)$ at $\rho(x)$ with potential functions on $\mathbb{R}^d$ (modulo additive constants). In short, a potential function $\Phi(x)$ represents a formula $-\nabla \cdot (\rho \nabla \Phi)$ in tangent space $T_\rho \mathcal{P}(\mathbb{R}^d)$. And the inner product in probability manifold is endowed by

$$(\Phi_1, \Phi_2)_\rho := \int \nabla \Phi_1 \cdot \nabla \Phi_2 \rho(x) dx.$$

This inner product leads a way to define the gradient of scalar functional on probability manifold.

We illustrate the above process by a special example. Consider a scalar functional

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^d} V(x)\rho(x)dx + \beta \int_{\mathbb{R}^d} \rho(x) \log \rho(x)dx.$$

Then the gradient of this functional on the probability manifold satisfies

$$\frac{d}{dt}\mathcal{F}(\rho_t)|_{t=0} = (\mathrm{grad}_{\mathcal{P}(\mathbb{R}^d)}\mathcal{F}(\rho), \Phi)_\rho$$

where ρ_t is the solution of equation (14) with initial measure ρ. Then

$$\frac{d}{dt}\mathcal{F}(\rho_t)|_{t=0} = \int_{\mathbb{R}^d} \frac{\partial}{\partial t}\{V(x)\rho_t\}dx + \beta \int_{\mathbb{R}^d} \frac{\partial}{\partial t}\{\rho_t \log \rho_t\}dx|_{t=0}$$
$$= \int_{\mathbb{R}^d} \nabla \Phi \cdot \nabla(V(x) + \beta \log \rho(x))\rho(x)dx,$$

where the second equality is through integration by parts. From the identification of inner product, $\mathrm{grad}_{\mathcal{P}(\mathbb{R}^d)}\mathcal{F}(\rho)$ is associated with the potential function $V(x) + \log \rho(x)$, representing formula $-\nabla \cdot [\rho \nabla(V(x) + \beta \log \rho)]$. In other words, the gradient flow of $\mathcal{F}(\rho)$ is

$$\frac{\partial \rho}{\partial t} = -\mathrm{grad}_{\mathcal{P}(\mathbb{R}^d)}\mathcal{F}(\rho) = \nabla \cdot [\rho \nabla(V(x) + \beta \log \rho)],$$

which is exactly (6).[4]

---

[4]We summarize the above viewpoint as "The law of gradient flow is the gradient flow in law space".

In general, for any functional $\mathcal{F} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$, $\text{grad}_{\mathcal{P}(\mathbb{R}^d)}\mathcal{F}$ is associated with $\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)$[5], meaning that the gradient flow of $\mathcal{F}(\rho)$ is

$$\frac{\partial\rho}{\partial t} = \nabla \cdot (\rho\nabla\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)). \tag{15}$$

(15) contains a class of PDEs, including nonlinear Fokker-Planck equations.

---

[5]It is the first variational formula, (L.H.S. of (7)). Especially, if $\mathcal{F}(\rho) = \int_{\mathbb{R}^d} L(x, \rho)dx$,

$$\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho) = \frac{\partial}{\partial\rho}L(x, \rho).$$

CHAPTER III

PART 1: OPTIMAL TRANSPORT ON FINITE GRAPHS

This chapter aims to connect Fokker-Planck equations and optimal transport on discrete underlying states, which are finite graphs.

We briefly review some facts on optimal transport theory [2, 94, 95]. Consider a minimization problem

$$\min_{\rho(x)\in\mathcal{P}(\mathbb{R}^d)} \mathcal{F}(\rho) = \int_{\mathbb{R}^d} V(x)\rho(x)dx + \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} W(x,y)\rho(x)\rho(y)dxdy + \beta\int_{\mathbb{R}^d}\rho(x)\log\rho(x)dx.$$

Here the variable $\rho(x)$ is a probability density function supported on $\mathbb{R}^d$ and the objective scalar functional is called free energy, with $V : \mathbb{R}^d \to \mathbb{R}$, $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $W(x,y) = W(y,x)$ for any $x,y \in \mathbb{R}^d$. Recently, by equipping probability set $\mathcal{P}(\mathbb{R}^d)$ with a 2-Wasserstein metric, the gradient flow of the above minimization problem forms a nonlinear Fokker-Planck equation [4], which is commonly considered in granular media [13, 23]

$$\frac{\partial\rho}{\partial t} = \nabla\cdot[\rho\nabla(V(x) + \int_{\mathbb{R}^d} W(x,y)\rho(t,y)dy)] + \beta\Delta\rho. \tag{16}$$

There are many ways to observe (16)'s gradient flow structures. For example, the free energy is (16)'s Lyapunov function,

$$\frac{d}{dt}\mathcal{F}(\rho(t,x)) = -\int_{\mathbb{R}^d}(\nabla F(x,\rho))^2\rho(t,x)dx, \quad F(x,\rho) = \frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho),$$

where $\frac{\delta}{\delta\rho(x)}$ is the first variational formula; Under suitable conditions of V and W, the solution $\rho(t,x)$ converges to the local minimizer of free energy, named Gibbs measure.[1]

---

[1]$\rho^*(x)$ is a Gibbs measure if it solves the fixed point problem

$$\rho^*(x) = \frac{1}{K}e^{-\frac{V(x)+\int_{\mathbb{R}^d} W(x,y)\rho^*(y)dy}{\beta}}, \quad \text{where} \quad K = \int_{\mathbb{R}^d} e^{-\frac{V(x)+\int_{\mathbb{R}^d} W(x,y)\rho^*(y)dy}{\beta}}dx.$$

In addition, Carrillo, McCann and Villani show that (16)'s solution converges to the Gibbs measure exponentially [23].

In this chapter, we plan to establish a similar approach on a finite graph. Here we consider a simple finite undirected graph $G = (V, E)$, where $V = \{1, 2, \cdots, n\}$ is the vertex set, E is the edge set which contains no self loops or multiple edges. We consider the probability set supported on all vertices of G

$$\mathcal{P}(G) = \{(\rho_i)_{i=1}^n \mid \sum_{i=1}^n \rho_i = 1, \quad \rho_i \geq 0\},$$

and a discrete free energy $\mathcal{F} : \mathcal{P}(G) \to \mathbb{R}$. For illustration, we consider mainly

$$\mathcal{F}(\rho) = \mathcal{V}(\rho) + \mathcal{W}(\rho) + \mathcal{H}(\rho),$$

where $\mathcal{V}(\rho)$, $\mathcal{W}(\rho)$, $\mathcal{H}(\rho)$ represents the discrete linear potential energy, interaction potential energy and linear entropy, respectively, which means

$$\mathcal{V}(\rho) = \sum_{i=1}^n v_i \rho_i, \quad \mathcal{W}(\rho) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \rho_i \rho_j, \quad \mathcal{H}(\rho) = \sum_{i=1}^n \rho_i \log \rho_i.$$

Here $(v_i)_{i=1}^n$ and $(w_{ij})_{1 \leq i,j \leq n}$ is a given constant vector and constant symmetric matrix. Under this setting, can we derive the gradient flow of discrete free energy $\mathcal{F}(\rho)$ on $\mathcal{P}(G)$? In other words, we consider the optimization problem

$$\min_{\rho} \mathcal{F}(\rho) \quad \text{s.t.} \quad \sum_{i=1}^n \rho_i = 1, \quad \rho_i \geq 0.$$

What is the gradient flow for the above optimization associated with the graph G's structure?

Despite optimal transport theory has been developed on continuous states, not much is known on discrete states. There are naturally two difficulties. One is that, although it is possible to define a metric on the discrete state, the graph is not a length space, so we can not connect with gradient flow directly; the other is that the finite graph often introduces more complicated neighborhood information. For example, consider two sphere like graphs connected by one edge.
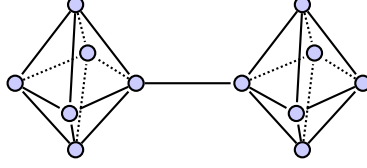
In this chapter, we derive the gradient flow based on a discrete "2-Wasserstein metric". Because we are not able to follow the static metric's definition in continuous states, we adopt another way in the time dependent version, Benamou-Brenier formula and up-wind scheme [27], to give the metric's definition.
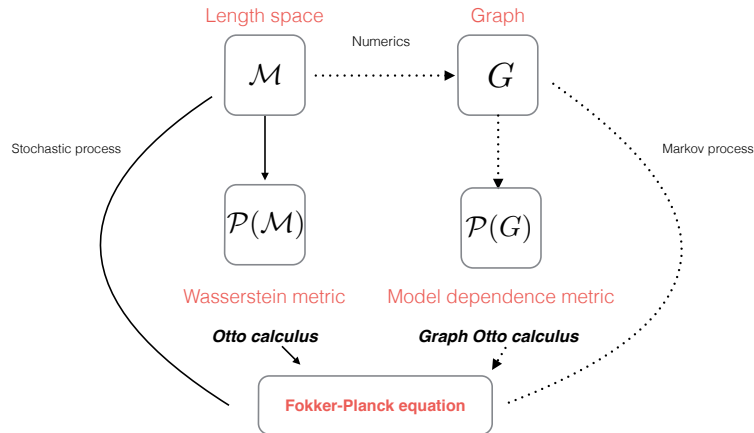
With the metric in hand, we derive the gradient flow of the discrete free energy

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} (F_j(\rho) - F_i(\rho))_+ \rho_j - \sum_{j \in N(i)} (F_i(\rho) - F_j(\rho))_+ \rho_i, \quad \text{for any } i \in V, \quad (17)$$

where $F_i(\rho) := \frac{\partial}{\partial \rho_i} \mathcal{F}(\rho)$, $\{\cdot\}_+ = \max\{\cdot, 0\}$ and $N(i)$ is all adjacent vertices (neighborhood) of i.[2] Since this process is motivated by Jordan, Kinderlehrer, and Otto [58, 79], we name (17) as Fokker-Planck equations on finite graphs, which are ordinary differential equations. There are several interesting questions associated with the gradient flow (17):

---

[2] $N(i) = \{j \in V \mid (i,j) \in E\}$, where (i,j) means that there is an edge connecting vertices i and j.

(i) Does the free energy decrease along (17)? What are the equilibria of (17)?

(ii) If the gradient flow converges to a strictly local minimizer ρ^∞, how fast is the convergence rate?

Question (i) is simple. We observe that $\mathcal{F}(\rho)$ is a Lyapunov function of (17),

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = -\sum_{(i,j)\in E}(F_i(\rho) - F_j(\rho))_+^2 \rho_i \leq 0.$$

It is easy to show that discrete Gibbs measures[3] are equilibria of (17). However, question (ii) is tricky to answer. The optimization of discrete free energy may have multiple minimizers (Gibbs measures), since the interaction energy can be non convex. So, in general, it is not possible to find a uniform convergence rate for all initial measures. It is natural to think of question (ii) in a dynamical way:

(iii) What is the asymptotic convergence rate for a given Gibbs measure?

We adopt the entropy dissipation method to answer (iii) [23, 68]. The concept of "entropy", introduced in [23], refers to the difference between two measures' free energies. However, we can not apply this method directly on a general free energy. Since the method requires the explicitly formula of the Hessian operator of free energy on probability manifold, it is still an open problem for general interaction energy[4].

We apply the **dynamical viewpoint** to conquer this difficulty. That is we find a formula $\lambda_{\mathcal{F}}(\rho)$ on finite graphs, see Definition 4 and Lemma 6, which plays the role of the smallest eigenvalue of free energy's "Hessian matrix" at **Gibbs measure** on probability manifold. Based on it, we show that if the Gibbs measure ρ^∞ is a strictly

---

[3]ρ^* is a discrete Gibbs measure, if it solves the fixed point problem:

$$\rho_i^* = \frac{1}{K}e^{-\frac{v_i + \sum_{j=1}^n w_{ij}\rho_j^*}{\beta}}, \quad \text{where} \quad K = \sum_{i=1}^n e^{-\frac{v_i + \sum_{j=1}^n w_{ij}\rho_j^*}{\beta}}.$$

[4]Problem 15.11 in [95]: Find a **nice** formula for the Hessian of $\mathcal{F}(\rho)$.

local minimizer of free energy and ρ^0 is in the attraction region of ρ^∞, then

$$\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty) \leq e^{-Ct}(\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)), \tag{18}$$

where C is a positive constant depending on ρ^0 and graph G's structure. Moreover, we prove that the discrete free energy decreases exponentially with asymptotic dissipation rate $2\lambda_\mathcal{F}(\rho^\infty)$.

In previous works, Ollivier introduces a 1-Wasserstein metric [77], which assumes that there is a metric on graph. It can not connect with Fokker-Planck equations. Erbar and Maas [43, 66], and Mielke [71] consider a similar 2-Wasserstein metic on discrete states, in which the probability set forms a smooth Riemannian manifold. They provide tools to analyze some linear Markov processes and numerical schemes for linear Fokker-Planck equations. However, our metric is different from them on two levels. (1) Our metric is only piecewise smooth, which doesn't satisfy A_1, A_5 condition in [66]. (2) The Fokker-Planck equation associated with our metric keeps the $\log \rho_i$ term, which deeply impacts the effect of the Laplacian operator. One one hand, it keeps the gradient flow's solution inside the probability manifold $\mathcal{P}_o(G)$, and the solution of (17) converges to the Gibbs measure automatically; One the other hand, it induces a discrete version of Hessian matrix of free energy on metric manifold at Gibbs measure.

We explain the plan of this chapter. We summarize our main results in section 3.1. We introduce the 2-Wasserstein metric in section 3.2. Based on such metric, we derive the gradient flow of free energy in section 3.3. By dynamical viewpoint, we show the exponential convergence result to Gibbs measures in section 3.4. Several examples are discussed in 3.5.

3.1 *Main results*

In this section, we briefly introduce our main results.

Discrete 2-Wasserstein metric First, we build a discrete 2-Wasserstein metric from the time-dependent viewpoint, which is a discrete version of the Benamou-Brenier formula:

Definition 2 (Discrete 2-Wasserstein metric) *For any ρ^0, $\rho^1 \in \mathcal{P}_o(G)$, we define*

$$W_{2;\mathcal{F}}(\rho^0, \rho^1)^2 := \inf \ \frac{1}{2} \int_0^1 \sum_{(i,j) \in E} g_{ij}(\rho)(\Phi_i - \Phi_j)^2 dt,$$

where the infimum is taken over piecewise C^1 curves $\rho : [0,1] \to \mathcal{P}_o(G)$ satisfying continuity equation introduced by measurable vector function $(\Phi_i(t))_{i=1}^n : [0,1] \to \mathbb{R}^n$,

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} g_{ij}(\rho)(\Phi_i - \Phi_j), \quad \rho(0) = \rho^0, \quad \rho(1) = \rho^1,$$

where $F_i(\rho) := \frac{\partial}{\partial \rho_i} \mathcal{F}(\rho)$ and

$$g_{ij}(\rho) := \begin{cases} \rho_i & \text{if } F_i(\rho) > F_j(\rho), \ j \in N(i); \\ \rho_j & \text{if } F_i(\rho) < F_j(\rho), \ j \in N(i); \\ \frac{\rho_i + \rho_j}{2} & \text{if } F_i(\rho) = F_j(\rho), \ j \in N(i). \end{cases}$$

Derivation of gradient flow Base on the 2-Wasserstein metric, we derive the gradient flow of free energy on probability manifold.

Theorem 3 (Derivation) *Given a simple finite graph $G = (V, E)$ and free energy $\mathcal{F}(\rho)$.*

(i) The gradient flow of discrete free energy $\mathcal{F}(\rho)$ on the metric manifold $(\mathcal{P}_o(G), W_{2;\mathcal{F}})$ is

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} \rho_j (F_j(\rho) - F_i(\rho))_+ - \sum_{j \in N(i)} \rho_i (F_i(\rho) - F_j(\rho))_+$$

for any $i \in V$. Recall that $F_i(\rho) = \frac{\partial}{\partial \rho_i} \mathcal{F}(\rho)$.

(ii) For any initial measure $\rho^0 \in \mathcal{P}_o(G)$, there exists a unique solution $\rho(t)$: $[0, \infty) \to \mathcal{P}_o(G)$ to equation (17) . Moreover, there is a constant $\epsilon > 0$ depending on ρ^0, such that $\rho_i(t) \geq \epsilon$ for all $i \in V$ and $t > 0$.

(iii) The free energy $\mathcal{F}(\rho)$ is a Lyapunov function of (17). If $\rho(t)$ is a solution of (17) with initial measure $\rho^0 \in \mathcal{P}_o(G)$, then

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = - \sum_{(i,j)\in E} (F_i(\rho) - F_j(\rho))_+^2 \rho_i \leq 0.$$

Moreover, if $\rho^\infty = \lim_{t\to\infty} \rho(t)$ exits, then ρ^∞ is a Gibbs measure, meaning that ρ^∞ solves the fixed point problem:

$$\rho_i^\infty = \frac{1}{K}e^{-\frac{v_i + \sum_{j=1}^n w_{ij}\rho_j^\infty}{\beta}}, \quad where \quad K = \sum_{i=1}^n e^{-\frac{v_i + \sum_{j=1}^n w_{ij}\rho_j^\infty}{\beta}}.$$

Remark 1 *(i) introduces an explicit formula for the gradient flow; (ii) demonstrate that (17) is a well defined ODE system; (iii) describes (17)'s gradient flow structure.*

Remark 2 *(ii) says more than that (17) is well defined.* ***(ii) shows that the boundary of $\mathcal{P}(G)$ is a repeller for (17), meaning that $\rho_i(t)$ stays positive for all $i \in V$ and $t > 0$. This property is the key for the convergence result.***

Convergence result How fast is convergence occuring? Since there may exist multiple minimizers of the free energy, it is not possible to find a unique convergence rate for all initial measures. Instead, we adopt a dynamical viewpoint. We introduce a formula, which induces the asymptotic convergence rate of (17) on finite graphs.

Definition 4 *Let $f_{ij} := \frac{\partial^2}{\partial\rho_i\partial\rho_j}\mathcal{F}(\rho)$ and*

$$h_{ij,kl} := f_{ik} + f_{jl} - f_{il} - f_{jk} \quad for\ any\ i,\ j,\ k,\ l \in V.$$

We define

$$\lambda_{\mathcal{F}}(\rho) := \min_\Phi \sum_{(i,j)\in E}\sum_{(k,l)\in E} h_{ij,kl}(\Phi_i - \Phi_j)_+\rho_i(\Phi_k - \Phi_l)_+\rho_k$$

24

where the minimum is taken among all $(\Phi_i)_{i=1}^n \in \mathbb{R}^n$ *with*

$$\sum_{(i,j)\in E} (\Phi_i - \Phi_j)_+^2 \rho_i = 1.$$

Based on this quantity $\lambda_{\mathcal{F}}(\rho)$, we derive (17)'s convergence result. More precisely, we assume that ρ^0 is in the attraction basin of the Gibbs measure ρ^∞, meaning that

$$\rho(0) = \rho^0 \quad \text{implies} \quad \lim_{t\to\infty} \rho(t) = \rho^\infty. \qquad (A)$$

Theorem 5 (Convergence) *Assume (A) holds and $\lambda_{\mathcal{F}}(\rho^\infty) > 0$, then there exists a constant $C > 0$, which depends on the initial measure ρ^0 and graph structure G, such that*

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

Moreover, the asymptotic convergence rate is $2\lambda_{\mathcal{F}}(\rho^\infty)$. In other words, for any sufficient small $\epsilon > 0$, there exists a constant time $T > 0$, such that when $t > T$,

$$\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty) \le e^{-2(\lambda_{\mathcal{F}}(\rho^\infty)-\epsilon)(t-T)}(\mathcal{F}(\rho(T)) - \mathcal{F}(\rho^\infty)).$$

Remark 3 $\lambda_{\mathcal{F}}(\rho)$ **plays the role of the smallest eigenvalue of free energy's "Hessian" matrix at Gibbs measure on metric manifold** $\mathcal{P}_o(G)$**.**

Analysis of dissipation rate What is the condition for $\lambda_{\mathcal{F}}(\rho^\infty) > 0$? We answer this question by giving another expression for $\lambda_{\mathcal{F}}(\rho)$.

Lemma 6

$$\lambda_{\mathcal{F}}(\rho) = \min_{\Phi}\{\left(\tilde{div}_G(\rho\nabla_G\Phi)\right)^T Hess_{\mathbb{R}^n}\mathcal{F}(\rho)\,\tilde{div}_G(\rho\nabla_G\Phi) \; : \; \sum_{(i,j)\in E}(\Phi_i - \Phi_j)_+^2\rho_i = 1\},$$

where $Hess_{\mathbb{R}^n}\mathcal{F}(\rho) := (\frac{\partial^2}{\partial\rho_i\partial\rho_j}\mathcal{F}(\rho))_{1\le i,j\le n}$ *and*

$$\tilde{div}_G(\rho\nabla_G\Phi) := \Big(\sum_{j\in N(i)}(\Phi_i - \Phi_j)_+\rho_i - \sum_{j\in N(i)}(\Phi_j - \Phi_i)_+\rho_j \Big)_{i=1}^n.$$

Lemma 6 induces a clear relation between (17)'s convergence result and the convexity of discrete free energy in $\mathbb{R}^n$.

Lemma 7 *If the matrix $Hess_{\mathbb{R}^n}\mathcal{F}(\rho)$ is positive definite at $\rho \in \mathcal{P}_o(G)$, then $\lambda_{\mathcal{F}}(\rho) > 0$. In particular, if we consider a linear entropy:*

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

Then for any $\rho \in \mathcal{P}_o(G)$,

$$\lambda_{\mathcal{H}}(\rho) = \min_{\Phi}\{\sum_{i=1}^{n} \frac{1}{\rho_i}(div_G(\rho\nabla_G\Phi)|_i)^2 \ : \ \sum_{(i,j)\in E}(\Phi_i - \Phi_j)_+^2\rho_i = 1\} > 0.$$

Lemma 7 implies the effect of linear entropy, based on which we prove the convergence results of linear and nonlinear Fokker-Planck equations on graphs.

Corollary 8 (Linear Fokker-Planck equation) *Consider the gradient flow* (17) *of free energy*

$$\mathcal{F}(\rho) = \sum_{i=1}^{n} v_i\rho_i + \beta\sum_{i=1}^{n} \rho_i \log \rho_i,$$

whose unique Gibbs measure is

$$\rho_i^\infty = \frac{1}{K}e^{-\frac{v_i}{\beta}}, \quad K = \sum_{i=1}^{n} e^{-\frac{v_i}{\beta}}.$$

Then there exists a constant $C > 0$, such that (18) *holds with the asymptotic rate $2\lambda_{\mathcal{F}}(\rho^\infty)$.*

Corollary 9 (Nonlinear Fokker-Planck equation) *Consider the gradient flow* (17) *of free energy*

$$\mathcal{F}(\rho) = \sum_{i=1}^{n} v_i\rho_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\rho_i\rho_j + \beta\sum_{i=1}^{n} \rho_i \log \rho_i,$$

with a semi positive definite matrix $(w_{ij})_{1\le i,j\le n}$. Then there exists a unique Gibbs measure

$$\rho_i^\infty = \frac{1}{K}e^{-\frac{v_i+\sum_{j\in N(i)} w_{ij}\rho_j^\infty}{\beta}}, \quad K = \sum_{i=1}^{n} e^{-\frac{v_i+\sum_{j\in N(i)} w_{ij}\rho_j^\infty}{\beta}}.$$

Furthermore, there exists a constant $C > 0$, such that (18) *holds with the asymptotic rate $2\lambda_{\mathcal{F}}(\rho^\infty)$.*

3.2 Discrete 2-Wasserstein metric

In this section, we define a discrete 2-Wasserstein metric, which is basic for deriving the Fokker-Planck equations on graphs.

3.2.1 Motivation

We consider a discrete potential vector field on a graph G:

$$\nabla_G \Phi := (\Phi_i - \Phi_j)_{(i,j) \in E}, \quad \text{where } (\Phi_i)_{i \in V} \in \mathbb{R}^n.$$

We shall build a discrete Benamou-Brenier formula[5]. For $\rho \in \mathcal{P}_o(G)$, we introduce a discrete **divergence** of $\nabla_G \Phi$ with respect to ρ:

$$\text{div}_G(\rho \nabla_G \Phi) := \Big(- \sum_{j \in N(i)} (\Phi_i - \Phi_j) g_{ij}(\rho) \Big)_{i=1}^n,$$

and an inner product of $\nabla_G \Phi$ with respect to ρ:

$$(\nabla_G \Phi, \nabla_G \Phi)_\rho := \frac{1}{2} \sum_{(i,j) \in E} (\Phi_i - \Phi_j)^2 g_{ij}(\rho).$$

Here for any $i, j \in V$, we define scalar functions $g_{ij}(\rho)$ satisfying

$$g_{ij}(\rho) = g_{ji}(\rho) \quad \text{and} \quad \min\{\rho_i, \rho_j\} \le g_{ij}(\rho) \le \max\{\rho_i, \rho_j\}. \tag{19}$$

Notice that there are many choices of g_{ij} satisfying (19). In this chapter, we are particularly interested in one that depends on the given free energy. We define

$$g_{ij}(\rho) := \begin{cases} \rho_i & \text{if } F_i(\rho) > F_j(\rho),\ j \in N(i); \\[2mm] \rho_j & \text{if } F_i(\rho) < F_j(\rho),\ j \in N(i); \\[2mm] \frac{\rho_i + \rho_j}{2} & \text{if } F_i(\rho) = F_j(\rho),\ j \in N(i), \end{cases} \tag{20}$$

where $F_i(\rho) = \frac{\partial}{\partial \rho_i} \mathcal{F}(\rho)$.

---

[5]We build analogs of $\nabla \cdot (\rho \nabla \Phi)$, $\int (\nabla \Phi)^2 \rho dx$.

By these settings, we can formally define a discrete 2-Wasserstein metric. For any two measures, ρ^0, $\rho^1 \in \mathcal{P}_o(G)$, we consider

$$W_{2;\mathcal{F}}(\rho^0, \rho^1) := \inf\{ \Big(\int_0^1 (\nabla_G \Phi, \nabla_G \Phi)_\rho dt \Big)^{\frac{1}{2}} \; : \; \frac{d\rho}{dt} + \text{div}_G(\rho \nabla_G \Phi) = 0, \; \rho \in \mathcal{C}\},$$

where $\mathcal{C} = \{\rho(t)$ is a piecewise C^1 curve with $\rho(0) = \rho^0$, $\rho(1) = \rho^1\}$. Notice that the above formula's explicit form is Definition 2.

3.2.2   Metric

In the sequel, we justify that the discrete 2-Wasserstein metric is well defined.

To do so, we look at the geometry angle[6]. That is we will endow $\mathcal{P}_o(G)$ with an inner product on the tangent space

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

Consider an equivalence relation "$\sim$" in $\mathbb{R}^n$ as modulo additive constants, whose quotient space means

$$\mathbb{R}^n / \sim = \{[\Phi] \mid (\Phi_i)_{i=1}^n \in \mathbb{R}^n\}, \quad \text{where} \quad [\Phi] = \{(\Phi_1 + c, \cdots, \Phi_n + c) \mid c \in \mathbb{R}^1\}.$$

We introduce an identification map

$$\tau : \; \mathbb{R}^n / \sim \to T_\rho \mathcal{P}_o(G), \qquad \tau([\Phi]) := -\text{div}_G(\rho \nabla_G \Phi).$$

Lemma 10 *The map $\tau : \; \mathbb{R}^n / \sim \to T_\rho \mathcal{P}_o(G)$ is a well defined, linear one to one map.*

Proof 1 *At the beginning, we show that τ is well defined. We denote*

$$\sigma_i = -div_G(\rho \nabla_G \Phi)|_i = \sum_{j \in N(i)} (\Phi_i - \Phi_j) g_{ij}(\rho).$$

---

[6]Our approach is a discrete version of Otto calculus. We strongly recommend readers, who are not familiar with Otto calculus, to learn it in [95], so as to understand the following justification.

It is equivalent to show $\sum_{i=1}^{n} \sigma_i = 0$. Indeed,

$$\sum_{i=1}^{n} \sigma_i = \sum_{i=1}^{n} \sum_{j \in N(i)} (\Phi_i - \Phi_j) g_{ij}(\rho)$$

$$= \sum_{(i,j) \in E} \Phi_i g_{ij}(\rho) - \sum_{(i,j) \in E} \Phi_j g_{ij}(\rho)$$

Relabel i and j on the first formula

$$= \sum_{(i,j) \in E} \Phi_j g_{ji}(\rho) - \sum_{(i,j) \in E} \Phi_j g_{ij}(\rho)$$

Notice $g_{ij} = g_{ji}$

$$= 0.$$

Hence, we know that the map τ is a well-defined linear map. Since $T_\rho \mathcal{P}_o(G)$ and $\mathbb{R}^n / \sim$ are $(n-1)$ dimensional linear spaces, we only need to prove τ is injective. I.e. if

$$\sigma_i = \sum_{j \in N(i)} g_{ij}(\rho)(\Phi_i - \Phi_j) = 0, \quad for\ any\ i \in V,$$

then $[\Phi] = 0$, meaning that $\Phi_1 = \Phi_2 = \cdots = \Phi_n$.

Assume this is not true. Let $c = \max_{i \in V} \Phi_i$. Since the graph G is connected, there exists an edge $(k,l) \in E$, such that $\Phi_l = c$ and $\Phi_k < c$. By $\sigma_l = 0$, we know that

$$\Phi_l = \frac{\sum_{j \in N(l)} g_{lj}(\rho) \Phi_j}{\sum_{j \in N(l)} g_{lj}(\rho)} = c + \frac{\sum_{j \in N(l)} g_{lj}(\rho)(\Phi_j - c)}{\sum_{j \in N(l)} g_{lj}(\rho)} < c,$$

which contradicts $\Phi_l = c$.

This identification map induces a scalar inner product on $\mathcal{P}_o(G)$.

Definition 11 *For any tangent vector $\sigma^1, \sigma^2 \in T_\rho \mathcal{P}_o(G)$, we define an inner product $g(\cdot, \cdot)\colon T_\rho \mathcal{P}_o(G) \times T_\rho \mathcal{P}_o(G) \to \mathbb{R}$,*

$$g(\sigma^1, \sigma^2) := \frac{1}{2} \sum_{(i,j) \in E} g_{ij}(\rho)(\Phi_i^1 - \Phi_j^1)(\Phi_i^2 - \Phi_j^2),$$

where $[\Phi^1]$, $[\Phi^2] \in \mathbb{R}^n / \sim$, such that $\sigma^1 = \tau([\Phi^1])$, $\sigma^2 = \tau([\Phi^2])$. Moreover, we can also denote the inner product by:

$$(\nabla_G \Phi^1, \nabla_G \Phi^2)_\rho := g(\sigma^1, \sigma^2).$$

Similarly as Otto calculus[7], one can check that the inner product in Definition 11 also has other formations:

$$g(\sigma^1, \sigma^2) = \sum_{i=1}^{n} \Phi_i^1 \sigma_i^2 = \sum_{i=1}^{n} \Phi_i^2 \sigma_i^1. \tag{21}$$

Proof 2 (Proof of (21)**)** *Without loss of generality, we shall prove*

$$g(\sigma^1, \sigma^2) = \sum_{i=1}^{n} \Phi_i^1 \sigma_i^2.$$

Since $\sigma^2 = \tau([\Phi^2]) = (\sum_{j \in N(i)} (\Phi_i^2 - \Phi_j^2) g_{ij})_{i=1}^n,$

$$\sum_{i=1}^{n} \Phi_i^1 \sigma_i^2 = \sum_{i=1}^{n} \Phi_i^1 \sum_{j \in N(i)} (\Phi_i^2 - \Phi_j^2) g_{ij}$$

$$= \frac{1}{2} \left(\sum_{i=1}^{n} \Phi_i^1 \sum_{j \in N(i)} (\Phi_i^2 - \Phi_j^2) g_{ij} + \sum_{i=1}^{n} \Phi_i^1 \sum_{j \in N(i)} (\Phi_i^2 - \Phi_j^2) g_{ij} \right)$$

$$\text{Relabel } i \text{ and } j \text{ on the second formula}$$

$$= \frac{1}{2} \left(\sum_{(i,j) \in E} \Phi_i^1 (\Phi_i^2 - \Phi_j^2) g_{ij} + \sum_{(i,j) \in E} \Phi_j^1 (\Phi_j^2 - \Phi_i^2) g_{ji} \right)$$

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

Under the above setting, we can justify the metric. Notice that the metric is equivalent to

$$W_{2;\mathcal{F}}(\rho^0, \rho^1)^2 = \inf\{ \int_0^1 g(\sigma, \sigma) dt \; : \; \frac{d\rho}{dt} = \sigma, \; \sigma \in T_\rho \mathcal{P}_o(G), \; \rho \in \mathcal{C} \}.$$

Since g_{ij} is a measurable function, we know $W_{2;\mathcal{F}}$ is a well defined metric. More details are stated in [27].

3.2.3 Piecewise smooth manifold

However, things are not perfect. Fix i and $j \in V$, $g_{ij}(\rho)$ may be discontinuous with respect to ρ, so the inner product $g(\cdot, \cdot)$ doesn't induce a smooth Riemannian manifold.

---

[7]Consider two potential functions Φ^1, Φ^2, then $\int \Phi^1(-\nabla \cdot (\rho \nabla \Phi^2)) dx = \int \nabla \Phi^1 \nabla \Phi^2 \rho dx.$

More precisely, the set $\mathcal{P}_o(G)$ is divided into finite (at most $\frac{n(n-1)}{2}$) many of smooth components. Each one's boundary is a sub-manifold

$$\mathcal{P}_{i,j} = \{\rho \in \mathcal{P}_o(G) \mid F_i(\rho) = F_j(\rho), \ (i,j) \in E\}.$$

We observe that all these $\mathcal{P}_{i,j}$ intersect at a set, which contains all of Gibbs measures:

$$\cap_{(i,j)\in E}\mathcal{P}_{i,j} = \{\rho^* \in \mathcal{P}_o(G) \mid \rho^* \text{ is one of Gibbs measures}\}. \tag{22}$$

Proof 3 (Proof of (22)) *Consider a measure $\rho^* \in \cap_{(i,j)\in E}\mathcal{P}_{i,j}$. Since the graph G is connected,*

$$F_1(\rho^*) = F_2(\rho^*) = \cdots = F_n(\rho^*).$$

Define

$$C := F_i(\rho^*) = v_i + \sum_{j=1}^{n} w_{ij}\rho_j^* + \beta \log \rho_i^*, \quad \text{for any } i \in V.$$

Letting $K = e^{-\frac{C}{\beta}}$ and using the fact $\sum_{i=1}^{n} \rho_i^* = 1$, we have

$$\rho_i^* = \frac{1}{K}e^{-\frac{v_i + \sum_{j=1}^{n} w_{ij}\rho_j^*}{\beta}}, \quad K = \sum_{i=1}^{n} e^{-\frac{v_i + \sum_{j=1}^{n} w_{ij}\rho_j^*}{\beta}},$$

meaning that ρ^* is a Gibbs measure.

We demonstrate the above facts by a simple example.

Example 1 *Consider a three vertex's lattice graph G and suppose there exists a unique Gibbs measure for given free energy.*

$$G \qquad\qquad \mathcal{P}(G)$$

31
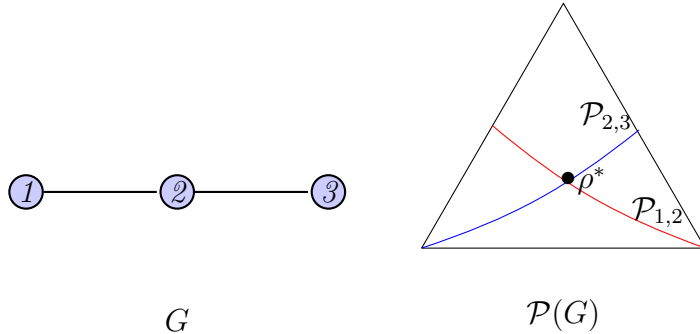
Here $\mathcal{P}_{1,2}$, $\mathcal{P}_{2,3}$ divide the probability manifold (simplex) $\mathcal{P}_o(G)$ into 4 pieces:

$$\{\rho \ : \ F_1(\rho) > F_2(\rho), \ F_2(\rho) > F_3(\rho)\}; \quad \{\rho \ : \ F_1(\rho) < F_2(\rho), \ F_2(\rho) > F_3(\rho)\};$$

$$\{\rho \ : \ F_1(\rho) > F_2(\rho), \ F_2(\rho) < F_3(\rho)\}; \quad \{\rho \ : \ F_1(\rho) < F_2(\rho), \ F_2(\rho) < F_3(\rho)\}.$$

Their boundaries intersect at a point $\{\rho^* \ : \ F_1(\rho) = F_2(\rho) = F_3(\rho)\}$, which is the Gibbs measure.

3.2.4 Hodge decomposition

As in optimal transport theory [4], we can justify the discrete 2-Wasserstein metric that the minimizer of Kinect energy is attached at the potential field on finite graphs.

In details, we consider a discrete vector field on a graph G:

$$v := (v_{ij})_{(i,j)\in E}, \quad \text{where} \sum_{j \in N(i)} g_{ij}(\rho)\frac{v_{ij} + v_{ji}}{2} = 0.$$

We shall build a discrete Benamou-Brenier formula for discrete vector field. For $\rho \in \mathcal{P}_o(G)$, we introduce a discrete divergence of v with respect to ρ:

$$\text{div}_G(\rho v) := \Big(- \sum_{j \in N(i)} v_{ij} g_{ij}(\rho) \Big)_{i=1}^{n},$$

and an inner product of v with respect to ρ:

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

We are going to show that the discrete 2-Wasserstein metric can be formally rewritten as

$$W_{2;\mathcal{F}}(\rho^0,\rho^1) = \inf_v \Big\{ \int_0^1 (v,v)_\rho dt \ : \ \frac{d\rho}{dt} + \text{div}_G(\rho v) = 0, \ \rho(0) = \rho^0, \ \rho(1) = \rho^1 \Big\}.$$

Theorem 12 (Hodge decomposition) *If ρ^0, $\rho^1 \in \mathcal{P}_o(G)$, then*

$$W_{2;\mathcal{F}}(\rho^0,\rho^1)^2 = \inf_{[\mathbf{v}],\rho} \frac{1}{2} \int_0^1 \sum_{(i,j)\in E} g_{ij}(\rho)v_{ij}^2 dt,$$

where the infimum is taken over all C^1 curves $\rho : [0,1] \to \mathcal{P}_o(G)$ satisfying continuity equation introduced by measurable matrix function $v(t) = (v_{ij}(t))_{(i,j)\in V\times V} : [0,1] \to \mathbb{R}^{n^2}$,

$$\frac{d\rho_i}{dt} = \sum_{j\in N(i)} g_{ij}(\rho)v_{ij}, \quad \sum_{j\in N(i)} g_{ij}(\rho)\frac{v_{ij}+v_{ji}}{2} = 0, \quad for\ any\ i \in V,$$

$$\rho(0) = \rho^0, \quad \rho(1) = \rho^1.$$

The proof is based on the following two lemmas.

Lemma 13 *Consider any skew matrix*[8] $(v_{ij})_{(i,j)\in V\times V} \in \mathbb{R}^{n^2}$, $(\Phi)_{i=1}^n \in \mathbb{R}^n$ *satisfying*

$$\sum_{j\in N(i)} g_{ij}(\rho)(\Phi_i - \Phi_j) = \sum_{j\in N(i)} g_{ij}(\rho)v_{ij}, \quad for\ any\ i \in V.$$

Then

$$\sum_{(i,j)\in E} g_{ij}(\rho)v_{ij}^2 \geq \sum_{(i,j)\in E} g_{ij}(\rho)(\Phi_i - \Phi_j)^2.$$

Proof 4 *Since*

$$\sum_{(i,j)\in E} g_{ij}(\rho)v_{ij}^2$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)(v_{ij} - (\Phi_i - \Phi_j) + \Phi_i - \Phi_j)^2 \tag{23}$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)[(v_{ij} - (\Phi_i - \Phi_j))^2 + (\Phi_i - \Phi_j)^2 + 2(v_{ij} - (\Phi_i - \Phi_j))(\Phi_i - \Phi_j)].$$

To prove the main result, it is sufficient to show the following claim.

Claim:

$$\sum_{(i,j)\in E} g_{ij}(\rho)[v_{ij} - (\Phi_i - \Phi_j)](\Phi_i - \Phi_j) = 0, \tag{24}$$

---

[8] *It means $v_{ij} = -v_{ji}$.*

Proof 5 (Proof of Claim) *Since $g_{ij} = g_{ji}$ and $v_{ij} = -v_{ji}$, we have*

$$\sum_{(i,j)\in E} g_{ij}(\rho)[v_{ij} - (\Phi_i - \Phi_j)](\Phi_i - \Phi_j)$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)v_{ij}(\Phi_i - \Phi_j) - \sum_{(i,j)\in E} g_{ij}(\rho)(\Phi_i - \Phi_j)^2$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)v_{ij}\Phi_i - \sum_{(i,j)\in E} g_{ij}(\rho)v_{ij}\Phi_j - 2(\nabla_G\Phi, \nabla_G\Phi)_\rho$$

$$= \sum_{i=1}^{n}\sum_{j\in N(i)} g_{ij}(\rho)v_{ij}\Phi_i - \sum_{(j,i)\in E} g_{ji}(\rho)v_{ji}\Phi_i - 2(\nabla_G\Phi, \nabla_G\Phi)_\rho$$

$$= \sum_{i=1}^{n}\sum_{j\in N(i)} g_{ij}(\rho)v_{ij}\Phi_i + \sum_{i=1}^{n}\sum_{j\in N(i)} g_{ij}(\rho)v_{ij}\Phi_i - 2(\nabla_G\Phi, \nabla_G\Phi)_\rho$$

$$= 2(\nabla_G\Phi, \nabla_G\Phi)_\rho - 2(\nabla_G\Phi, \nabla_G\Phi)_\rho = 0.$$

Combining the fact (23) and (24), we have

$$\sum_{(i,j)\in E} g_{ij}(\rho)v_{ij}^2 - \sum_{(i,j)\in E} g_{ij}(\rho)(\Phi_i - \Phi_j)^2$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)[v_{ij} - (\Phi_i - \Phi_j)]^2 \geq 0,$$

which finishes the proof.

The above result can be extended into a more general form.

Corollary 14 *For any matrix $(v_{ij})_{(i,j)\in V} \in \mathbb{R}^{n^2}$, $(\Phi_i)_{i=1}^{n} \in \mathbb{R}^n$, satisfying*

$$\sum_{j\in N(i)} g_{ij}(\rho)\frac{v_{ij} + v_{ji}}{2} = 0, \quad \text{for any } i \in V,$$

and

$$\sum_{j\in N(i)} g_{ij}(\rho)(\Phi_i - \Phi_j) = \sum_{j\in N(i)} g_{ij}(\rho)v_{ij}, \quad \text{for any } i \in V.$$

Then

$$\sum_{(i,j)\in E} g_{ij}(\rho)v_{ij}^2 \geq \sum_{(i,j)\in E} g_{ij}(\rho)(\Phi_i - \Phi_j)^2.$$

Proof 6 *It is same to prove a new claim.*

New claim:

$$\sum_{(i,j)\in E} g_{ij}(\rho)[v_{ij} - (\Phi_i - \Phi_j)](\Phi_i - \Phi_j) = 0.$$

Proof 7 (Proof of New claim) : *To show that, since* $v_{ij} = \frac{v_{ij}-v_{ji}}{2} + \frac{v_{ij}+v_{ji}}{2}$ *and assumption, for any* $i \in V$, *we have*

$$\sigma_i = \sum_{j\in N(i)} g_{ij}(\rho)v_{ij}$$

$$= \sum_{j\in N(i)} g_{ij}(\rho)[\frac{v_{ij}-v_{ji}}{2} + \frac{v_{ij}+v_{ji}}{2}]$$

$$= \sum_{j\in N(i)} g_{ij}(\rho)\frac{v_{ij}-v_{ji}}{2}.$$

From lemma 33 and $g_{ij} = g_{ji}$, *we have*

$$\sum_{(i,j)\in E} g_{ij}(\rho)[v_{ij} - (\Phi_i - \Phi_j)](\Phi_i - \Phi_j)$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)[\frac{v_{ij}-v_{ji}}{2} + \frac{v_{ij}+v_{ji}}{2} - (\Phi_i - \Phi_j)](\Phi_i - \Phi_j)$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)[\frac{v_{ij}-v_{ji}}{2} - (\Phi_i - \Phi_j)](\Phi_i - \Phi_j) + \sum_{(i,j)\in E} g_{ij}(\rho)\frac{v_{ij}+v_{ji}}{2}(\Phi_i - \Phi_j)$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)\frac{v_{ij}+v_{ji}}{2}(\Phi_i - \Phi_j)$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)\frac{v_{ij}+v_{ji}}{2}\Phi_i - \sum_{(i,j)\in E} g_{ij}(\rho)\frac{v_{ij}+v_{ji}}{2}\Phi_j$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)\frac{v_{ij}+v_{ji}}{2}\Phi_i - \sum_{(j,i)\in E} g_{ji}(\rho)\frac{v_{ij}+v_{ji}}{2}\Phi_i = 0.$$

By a direct calculation, we know that

$$\sum_{(i,j)\in E} g_{ij}(\rho)v_{ij}^2 - \sum_{(i,j)\in E} g_{ij}(\rho)(\Phi_i - \Phi_j)^2$$

$$= \sum_{(i,j)\in E} g_{ij}(\rho)[(v_{ij} - (\Phi_i - \Phi_j))^2 + (\Phi_i - \Phi_j)^2 + 2(v_{ij} - (\Phi_i - \Phi_j))(\Phi_i - \Phi_j)]$$

$$+2 \sum_{(i,j)\in E} g_{ij}(\rho)[v_{ij} - (\Phi_i - \Phi_j)](\Phi_i - \Phi_j) \quad \textit{By the new claim}$$

$$\geq 0,$$

which finishes the proof.

From Corollary 14, we have proved Theorem 12.

3.3 Fokker-Planck equation on a graph

In this sequel, we derive the gradient flow of $\mathcal{F}(\rho)$ on manifold $(\mathcal{P}_o(G), W_{2;\mathcal{F}})$. Since the metric manifold is not smooth, we derive an ODE system, which satisfies the definition of gradient flow on each components of smooth Riemannian manifolds. We name such ODE system as the Fokker-Planck equation on a graph.

3.3.1 Derivation

Before starting the derivation, we briefly review the definition of gradient flow on a smooth Riemannian manifold. Abstractly, the gradient flow is defined as

$$\frac{d\rho}{dt} = -\text{grad}_{\mathcal{P}_o(G)}\mathcal{F}(\rho).$$

Here the gradient is in the tangent space $T_\rho \mathcal{P}_o(G)$, defined by the duality condition:

$$g(\text{grad}_{\mathcal{P}_o(G)}\mathcal{F}(\rho), \sigma) = d\mathcal{F}(\rho) \cdot \sigma, \quad \text{for any } \sigma \in T_\rho \mathcal{P}_o(G),$$

where the dot in R.H.S. represents "$d\mathcal{F}$ applies to σ": $d\mathcal{F} \cdot \sigma = \sum_{i=1}^n \frac{\partial}{\partial \rho_i}\mathcal{F}(\rho)\sigma_i$. In all, the gradient flow satisfies

$$(\frac{d\rho}{dt}, \sigma)_\rho + d\mathcal{F}(\rho) \cdot \sigma = 0, \quad \text{for any } \sigma \in T_\rho \mathcal{P}_o(G). \tag{25}$$

Based on (25), we are ready to show Theorem 3.[9]

Proof 8 (Proof of Theorem 3 (i)) *We show the derivation of* (17). *Note that any* $\sigma \in T_\rho \mathcal{P}_o(G)$, *there exists* $[\Phi] \in \mathbb{R}^n / \sim$, *such that* $\tau([\Phi]) = \sigma = \left(\sum_{j \in N(i)} g_{ij}(\rho)(\Phi_i - \Phi_j) \right)_{i=1}^n$. *Since*

$$
\begin{aligned}
d\mathcal{F}(\rho) \cdot \sigma &= \sum_{i=1}^n \frac{\partial}{\partial \rho_i} \mathcal{F}(\rho) \cdot \sigma_i \\
&= \sum_{i=1}^n F_i(\rho) \sum_{j \in N(i)} g_{ij}(\rho)(\Phi_i - \Phi_j) \\
&= \sum_{i=1}^n \sum_{j \in N(i)} g_{ij}(\rho) F_i(\rho) \Phi_i - \sum_{i=1}^n \sum_{j \in N(i)} g_{ij}(\rho) F_i(\rho) \Phi_j
\end{aligned}
$$

Relabel i *and* j *on the second formula above.*

$$
\begin{aligned}
&= \sum_{i=1}^n \sum_{j \in N(i)} g_{ij}(\rho) F_i(\rho) \Phi_i - \sum_{i=1}^n \sum_{j \in N(i)} g_{ji}(\rho) F_j(\rho) \Phi_i \\
&= \sum_{i=1}^n \sum_{j \in N(i)} g_{ij}(\rho) \big(F_i(\rho) - F_j(\rho) \big) \Phi_i.
\end{aligned}
$$

Combining the above formula and (21) *into* (25), *we have*

$$
(\frac{d\rho}{dt}, \sigma)_\rho + d\mathcal{F}(\rho) \cdot \sigma = \sum_{i=1}^n \{ \frac{d\rho_i}{dt} + \sum_{j \in N(i)} g_{ij}(\rho) \big(F_i(\rho) - F_j(\rho) \big) \} \Phi_i = 0.
$$

Since the above formula is true for all $(\Phi_i)_{i=1}^n \in \mathbb{R}^n$,

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

holds for all $i \in V$. *From the definition of* g_{ij} *in* (20), *we finish the derivation of* (17).

3.3.2   Existence and uniqueness

Secondly, we prove the existence and uniqueness of the solution for ODE (17).

---

[9]This proof is mainly based on the result of [27], in which the authors only consider linear Fokker-Planck equations. We generalize the result to the nonlinear case.

Proof 9 (Proof of Theorem 3 (ii)) *We show the results by the following claim:*

*   ***Claim:** Given $\rho^0 \in \mathcal{P}_o(G)$, there exists a compact set $B(\rho^0) \subset \mathcal{P}_o(G)$ in Euclidean metric, such that $\rho^0 \in B(\rho^0)$ and the solution of (17) exists, with $\rho(t) : [0, \infty) \to B(\rho^0)$.*

Proof 10 (Proof of the claim) *At the beginning, we construct a compact set $B(\rho^0) \subset \mathcal{P}_o(G)$. Denote*

$$M = e^{2 \sup_{i \in V, \ \rho \in \mathcal{P}(G)} |v_i + \sum_{j=1}^n w_{ij} \rho_j|}$$

and a sequence of constant ϵ_l, $l = 0, 1, \cdots, n$, where $\epsilon_0 = 1$ with

$$\epsilon_1 = \frac{1}{2} \min\{\frac{\epsilon_0}{1 + (2M)^{\frac{1}{\beta}}}, \min_{i \in V} \rho_i^0\} \quad and \quad \epsilon_l = \frac{\epsilon_{l-1}}{1 + (2M)^{\frac{1}{\beta}}}, \quad for \ l = 2, \cdots, n \ .$$

Then we define

$$B(\rho^0) = \{(\rho_i)_{i=1}^n \in \mathcal{P}_o(G) \mid \sum_{r=1}^l \rho_{i_r} \le 1 - \epsilon_l, \ for \ any \ l \ indexes \ 1 \le i_1 < \cdots < i_l \le n,$$

$$with \ l \in \{1, \cdots, n-1\}\}.$$

Notice that $B(\rho^0)$ is a compact subset of $\mathcal{P}_o(G)$ in Euclidean metric.

Firstly, we observe that the R.H.S of (17) is a Lipchitz function when $\rho \in B(\rho^0)$, since $(F_i(\rho) - F_j(\rho))_+ = \frac{|F_i(\rho) - F_j(\rho)| + F_i(\rho) - F_j(\rho)}{2}$. Then (17) has a unique solution as long as $\rho(t) \in B(\rho^0)$.

Secondly, we show that if $\rho^0 \in B(\rho^0)$, then $\rho(t) \in B(\rho^0)$ for all $t \ge 0$. In other words, the boundary of $B(\rho^0)$ is a repeller for the ODE (17). Consider a time t_1 with $\rho(t_1) \in \partial B(\rho^0)$, meaning that there exists indices $i_1, \cdots, i_l$ with $l \le n - 1$, such that

$$\sum_{r=1}^l \rho_{i_r}(t_1) = 1 - \epsilon_l. \tag{26}$$

We are going to show

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

We begin the proof by letting $A = \{i_1, \cdots, i_l\}$ and $A^c = V \setminus A$. On one hand, for any $j \in A^c$,

$$\rho_j(t_1) \leq 1 - \sum_{r=1}^{l} \rho_{i_r}(t_1) = \epsilon_l. \tag{27}$$

On the other hand, since $\rho(t_1) \in B(\rho_0)$, for any $i \in A$, then $\sum_{k \in A \setminus \{i\}} \rho_k(t_1) \leq 1 - \epsilon_{l-1}$ and from assumption (26), $\rho_i(t_1) + \sum_{k \in A \setminus \{i\}} \rho_k(t_1) = 1 - \epsilon_l$, we obtain

$$\rho_i(t_1) \geq 1 - \epsilon_l - (1 - \epsilon_{l-1}) = \epsilon_{l-1} - \epsilon_l. \tag{28}$$

Combining equations (27) and (28), we know that for any $i \in A$ and $j \in A^c$,

$$F_j(\rho) - F_i(\rho) = (v_j + \sum_{j \in N(i)} w_{ij}\rho_i) - (v_i + \sum_{j=1}^{n} w_{ij}\rho_j) + \beta(\log \rho_j - \log \rho_i)$$

$$\leq 2 \sup_{i \in V, \ \rho \in \mathcal{P}(G)} |v_i + \sum_{j \in N(i)} w_{ij}\rho_j| + \beta(\log \epsilon_l - \log(\epsilon_{l-1} - \epsilon_l)) \tag{29}$$

$$\leq -\log 2,$$

where the last inequality is from $\epsilon_l = \frac{\epsilon_{l-1}}{1+(2M)^{\frac{1}{\beta}}}$ and $M = \sup_{i \in V, \ \rho \in \mathcal{P}(G)} |v_i + \sum_{j \in N(i)} w_{ij}\rho_j|$.

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

$$\sum_{i \in A} \sum_{j \in A^c \cap N(i^*)} g_{ij}(\rho(t_1)) \geq g_{i_*j_*}(\rho(t_1)) > 0. \tag{30}$$

By combining (29) and (30), we have

$$\frac{d}{dt} \sum_{r=1}^{l} \rho_{i_r}(t)|_{t=t_1} = \sum_{i \in A} \sum_{j \in N(i)} g_{ij}(\rho)[F_j(\rho) - F_i(\rho)]|_{\rho=\rho(t_1)}$$

$$= \sum_{i \in A} \{ \sum_{j \in A \cap N(i)} g_{ij}(\rho)[F_j(\rho) - F_i(\rho)]$$

$$+ \sum_{j \in A^c \cap N(i)} g_{ij}(\rho)[F_j(\rho) - F_i(\rho)]\}|_{\rho=\rho(t_1)}$$

$$= \sum_{i \in A} \sum_{j \in A^c \cap N(i)} g_{ij}(\rho)[F_j(\rho) - F_i(\rho)]|_{\rho=\rho(t_1)}$$

$$\leq -\log 2 \sum_{i \in A} \sum_{j \in A^c \cap N(i)} g_{ij}(\rho(t_1))$$

$$\leq -\log 2 \ g_{i_*j_*}(\rho(t_1)) < 0,$$

where the third equality is from $\sum_{(i,j) \in A} g_{ij}(F_j - F_i) = 0$. In all, we finish the proof.

From the claim, if the initial measure $\rho^0 \in \mathcal{P}_o(G)$, there exists a unique solution

$$\rho(t) : [0, +\infty) \to \mathcal{P}_o(G)$$

to equation (17). Moreover, since $\rho(t) \in B(\rho^0)$, there exists a constant $\epsilon > 0$, such that $\rho_i(t) \geq \epsilon$, for any $i \in V$.

3.3.3 Gradient flow structure

Lastly, we show that (17) has the gradient flow structure.

Proof 11 (Proof of Theorem 3 (iii)) *Firstly, we show that $\mathcal{F}(\rho)$ is a Lyapunov function of (17). Since*

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = \sum_{i=1}^{n} F_i(\rho) \cdot \frac{d\rho_i}{dt}$$

$$= \sum_{i=1}^{n} \sum_{j \in N(i)} F_i(\rho)(F_j(\rho) - F_i(\rho))_+ \rho_j - \sum_{i=1}^{n} \sum_{j \in N(i)} F_i(\rho)(F_i(\rho) - F_j(\rho))_+ \rho_i$$

Switch i, j on the first formula

$$= \sum_{i=1}^{n} \sum_{j \in N(i)} F_j(\rho)(F_i(\rho) - F_j(\rho))_+ \rho_i - \sum_{i=1}^{n} \sum_{j \in N(i)} F_i(\rho)(F_i(\rho) - F_j(\rho))_+ \rho_i$$

$$= - \sum_{(i,j) \in E} (F_i(\rho) - F_j(\rho))_+^2 \rho_i \leq 0.$$

Next, we show that if ρ^∞ is an equilibrium, then ρ^∞ is a Gibbs measure. At the beginning, since $\lim_{t \to \infty} \rho(t) = \rho^\infty$ we claim that

$$\lim_{t \to \infty} \frac{d}{dt}\mathcal{F}(\rho(t)) = \sum_{i=1}^{n} \sum_{j \in N(i)} (F_i(\rho^\infty) - F_j(\rho^\infty))_+^2 \rho_i^\infty = 0. \qquad (31)$$

Assume this is not true, we can easily obtain that $\inf_{t \geq 0} \mathcal{F}(\rho(t)) = -\infty$, which contradicts the fact that $\mathcal{F}(\rho)$ is bounded when $\rho \in \mathcal{P}(G)$.

From $\rho(t) \in B(\rho^0)$, we know that $\rho^\infty \in B(\rho^0)$. Combining this with (31), we have $F_i(\rho^\infty) = F_j(\rho^\infty)$, for any $(i, j) \in E$. Since the graph is connected, we obtain $F_1(\rho^\infty) = \cdots = F_n(\rho^\infty)$. Define

$$C := v_i + \sum_{j=1}^{n} w_{ij}\rho_j^\infty + \beta \log \rho_i^\infty, \quad \text{for any } i \in V.$$

Letting $K = e^{-\frac{C}{\beta}}$ and using the fact $\sum_{i=1}^{n} \rho_i^\infty = 1$, we have

$$\rho_i^\infty = \frac{1}{K} e^{-\frac{v_i + \sum_{j=1}^{n} w_{ij} \rho_j^\infty}{\beta}}, \quad K = \sum_{i=1}^{n} e^{-\frac{v_i + \sum_{j=1}^{n} w_{ij} \rho_j^\infty}{\beta}},$$

which finishes the proof.

3.4   Convergence results

In this section, we show the convergence results for the gradient flow (17).

Motivation Our proof is based on the structure of gradient flow. We illustrate the idea by a simple example. Consider a energy function $g(x) \in C^2(\mathbb{R}^n)$, whose gradient flow is

$$\frac{dx_t}{dt} = -\nabla g(x_t), \quad x_t \in \mathbb{R}^n.$$

By a simple computation of the first and second derivative of $g(x_t)$ with respect to t, we have

$$\frac{d}{dt} g(x_t) = -(\nabla g(x_t), \nabla g(x_t)), \quad \frac{d^2}{dt^2} g(x_t) = 2(\text{Hess}_{\mathbb{R}^n} g(x_t) \cdot \nabla g(x_t), \nabla g(x_t)).$$

Suppose the energy function $g(x)$ is λ-convex, $\text{Hess}_{\mathbb{R}^n} g(x) \geq \lambda I$ for all $x \in \mathbb{R}^n$, we have the comparison between the first and second derivative

$$\frac{d^2}{dt^2} g(x_t) \geq -2\lambda \frac{d}{dt} g(x_t).$$

Such comparison induces the convergence result. Taking integration on time interval $[t, +\infty)$ on the above inequality, we have

$$\frac{d}{dt} [g(x_t) - g(x_\infty)] \leq -2\lambda [g(x_t) - g(x_\infty)];$$

Applying Gronwall's inequality on the above formula, we obtain

$$g(x_t) - g(x_\infty) \leq e^{-2\lambda t} (g(x_0) - g(x_\infty)),$$

which shows that the energy decreases exponentially.

This convergence result can be extended by the **dynamical viewpoint**. We obtain a similar result if g is **locally** strictly convex at the equilibrium. In other words, if x_0 is in the attraction region of a strictly convex local minimizer x_∞, then there exists a constant $C > 0$, such that

$$g(x_t) - g(x_\infty) \leq e^{-Ct}(g(x_0) - g(x_\infty)).$$

3.4.1 Entropy dissipation method

In this sequel, we prove Theorem 5 similarly. The proof is divided into three parts:

- In *preliminary computation*, we estimate $\frac{d^2}{dt^2}\mathcal{F}(\rho(t))$, the second derivative of free energy along gradient flow (17);

- In *comparison*, we use $\frac{\frac{d^2}{dt^2}\mathcal{F}(\rho(t))}{\frac{d}{dt}\mathcal{F}(\rho(t))}$, the ratio between the first and second derivative to show the asymptotic convergence result;

- In *Main results*, we prove the convergence result (18) based on the dynamical viewpoint.

3.4.1.1 Preliminary computation

In first part, we need an explicit formula for the second derivative. We argue that

$$\frac{d^2}{dt^2}\mathcal{F}(\rho(t)) = 2 \sum_{(i,j)\in E} \sum_{(k,l)\in E} h_{ij,kl}(F_i - F_j)_+\rho_i(F_k - F_l)_+\rho_k + o(\frac{d}{dt}\mathcal{F}(\rho(t))), \quad (32)$$

where $\lim_{h\to 0} \frac{o(h)}{h} = 0$, $h_{ij,kl} = f_{ik} + f_{lj} - f_{il} - f_{jk}$ and $f_{ij} = \frac{\partial^2}{\partial\rho_i\partial\rho_j}\mathcal{F}(\rho)$.

Calculation: Recall that the gradient flow (17) is

$$\frac{d\rho_i}{dt} = \sum_{j\in N(i)} (F_j - F_i)_+\rho_j - \sum_{j\in N(i)} (F_i - F_j)_+\rho_i.$$

We calculate the first derivative of free energy along (17). From Theorem 3 (iii),

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = -\sum_{i=1}^n \sum_{j\in N(i)} (F_i - F_j)_+^2\rho_i. \quad (33)$$

42

And we compute the second derivative of free energy along (17).[10] By using the product rule on (33),

$$\frac{d^2}{dt^2}\mathcal{F}(\rho(t)) = \quad -\sum_{i=1}^{n}\sum_{j\in N(i)}(F_i - F_j)_+^2\frac{d\rho_i}{dt} \qquad \clubsuit$$

$$-2\sum_{i=1}^{n}\sum_{j\in N(i)}(\frac{dF_i}{dt} - \frac{dF_j}{dt})(F_i - F_j)_+\rho_i. \quad \spadesuit$$

Now, we plan to show (32) by two steps.

Step 1 $\clubsuit$ is the high order term of the first derivative (33):

$$\clubsuit = o(\frac{d}{dt}\mathcal{F}(\rho(t)));$$

Step 2 $\spadesuit$ has the same order of the first derivative (33):

$$\spadesuit = 2\sum_{(i,j)\in E}\sum_{(k,l)\in E}h_{ij,kl}(F_i - F_j)_+\rho_i(F_k - F_l)_+\rho_k.$$

Step 1: We start with the estimation of $\clubsuit$. Observe that when $t \to \infty$,

$$\frac{\clubsuit}{\frac{d}{dt}\mathcal{F}(\rho(t))} = \frac{\sum_{i=1}^{n}\sum_{j\in N(i)}(F_i - F_j)_+^2\frac{d\rho_i}{dt}}{\sum_{i=1}^{n}\sum_{j\in N(i)}(F_i - F_j)_+^2\rho_i} \to 0,$$

since $\frac{d\rho_i}{dt} \to 0$ and $\rho_i(t) \to \rho_i^\infty > 0$ in Theorem 3 (ii).

Lemma 15 *There always exists a constant $m(\rho) = \max_{(i,j)\in E}(F_i - F_j)_+ \max_{i\in V} deg(i)\frac{\max_{i\in V}\rho_i}{\min_{i\in V}\rho_i}$ where $deg(i)$ represents the degree[11] of vertex i, such that*

$$\clubsuit \geq m(\rho(t))\frac{d}{dt}\mathcal{F}(\rho(t)).$$

Proof 12 *The proof is based on a direct estimation. Denote $\frac{d\rho_i}{dt} = \sum_{k\in N(i)}\rho_k(F_k -$*

---

[10] $\frac{d^2}{dt^2}\mathcal{F}(\rho(t))$ exists for all $t \geq 0$, since $(F_i(\rho) - F_j(\rho))_+^2$ is differentiable everywhere with respect to ρ.
[11] *Number of vertices in $N(i)$.*

$F_i)_+ - \sum_{k \in N(i)} \rho_i (F_i - F_k)_+$. *Then*

$$\clubsuit = - \sum_{i=1}^{n} \sum_{j \in N(i)} (F_i - F_j)^2_+ \{ \sum_{k \in N(i)} \rho_k (F_k - F_i)_+ - \sum_{k \in N(i)} \rho_i (F_i - F_k)_+ \}$$

$$\geq - \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{k \in N(i)} (F_i - F_j)^2_+ (F_k - F_i)_+ \rho_k$$

$$\geq - \max_{(i,j) \in E} (F_i - F_j)_+ \max_{i \in V} \rho_i \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{k \in N(i)} (F_i - F_j)_+ (F_k - F_i)_+$$

$$\geq - \max_{(i,j) \in E} (F_i - F_j)_+ \max_{i \in V} \rho_i \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{k \in N(i)} \frac{(F_i - F_j)^2_+ + (F_k - F_i)^2_+}{2}$$

$$\geq - \max_{(i,j) \in E} (F_i - F_j)_+ \max_{i \in V} \rho_i \{ \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{k \in N(i)} \frac{(F_i - F_j)^2_+}{2} + \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{k \in N(i)} \frac{(F_k - F_i)^2_+}{2} \}$$

$$\geq - \max_{(i,j) \in E} (F_i - F_j)_+ \max_{i \in V} \rho_i \cdot \max_{i \in V} deg(i) \cdot \sum_{i=1}^{n} \sum_{j \in N(i)} (F_i - F_j)^2_+$$

$$\geq - \max_{(i,j) \in E} (F_i - F_j)_+ \max_{i \in V} deg(i) \frac{\max_{i \in V} \rho_i}{\min_{i \in V} \rho_i} \sum_{i=1}^{n} \sum_{j \in N(i)} (F_i - F_j)^2_+ \rho_i.$$

By letting $m = \max_{(i,j) \in E} (F_i - F_j)_+ \max_{i \in V} deg(i) \frac{\max_{i \in V} \rho_i}{\min_{i \in V} \rho_i}$, we finish the proof.

Step 2: Let's estimate $\spadesuit$ though a direct calculation.

Lemma 16

$$\spadesuit = 2 \sum_{(i,j) \in E} \sum_{(k,l) \in E} h_{ij,kl} (F_i - F_j)_+ \rho_i (F_k - F_l)_+ \rho_k.$$

Proof 13

$$-\frac{1}{2}\spadesuit = \sum_{i=1}^{n}\sum_{j\in N(i)}(F_i - F_j)_+\rho_i(\frac{d}{dt}F_i(\rho(t)) - \frac{d}{dt}F_j(\rho(t)))$$

$$= \sum_{i=1}^{n}\sum_{j\in N(i)}(F_i - F_j)_+\rho_i(\sum_{k=1}^{n}\frac{\partial F_i}{\partial\rho_k}\frac{d\rho_k}{dt} - \sum_{k=1}^{n}\frac{\partial F_j}{\partial\rho_k}\frac{d\rho_k}{dt})$$

$$Recall\ \frac{\partial F_i}{\partial\rho_k} = \frac{\partial^2}{\partial\rho_i\partial\rho_k}\mathcal{F}(\rho) = f_{ik}$$

$$= \sum_{i=1}^{n}\sum_{j\in N(i)}(F_i - F_j)_+\rho_i\sum_{k=1}^{n}(f_{ik} - f_{kj})\frac{d\rho_k}{dt}$$

$$Denote\ \frac{d\rho_k}{dt} = \sum_{l\in N(k)}(F_l - F_k)_+\rho_l - \sum_{l\in N(k)}(F_k - F_l)_+\rho_k$$

$$= \sum_{i=1}^{n}\sum_{j\in N(i)}(F_i - F_j)_+\rho_i\sum_{k=1}^{n}(f_{ik} - f_{kj})[\sum_{l\in N(k)}(F_l - F_k)_+\rho_l - \sum_{l\in N(k)}(F_k - F_l)_+\rho_k]$$

$$= \sum_{i=1}^{n}\sum_{j\in N(i)}(F_i - F_j)_+\rho_i\{\sum_{k=1}^{n}\sum_{l\in N(k)}(f_{ik} - f_{kj})(F_l - F_k)_+\rho_l$$

$$- \sum_{k=1}^{n}\sum_{l\in N(k)}(f_{ik} - f_{kj})(F_k - F_l)_+\rho_k\}$$

$$Relabel\ k,\ l\ for\ the\ first\ formula$$

$$= \sum_{i=1}^{n}\sum_{j\in N(i)}(F_i - F_j)_+\rho_i\{\sum_{k=1}^{n}\sum_{l\in N(k)}(f_{il} - f_{lj})(F_k - F_l)_+\rho_k$$

$$- \sum_{k=1}^{n}\sum_{l\in N(k)}(f_{ik} - f_{kj})(F_k - F_l)_+\rho_k\}$$

$$= \sum_{i=1}^{n}\sum_{j\in N(i)}\sum_{k=1}^{n}\sum_{l\in N(k)}(f_{il} - f_{lj} - f_{ik} + f_{kj})(F_i - F_j)_+\rho_i(F_k - F_l)_+\rho_k$$

$$= \sum_{(i,j)\in E}\sum_{(k,l)\in E}(f_{il} - f_{lj} - f_{ik} + f_{kj})(F_i - F_j)_+\rho_i(F_k - F_l)_+\rho_k.$$

Let $h_{ij,kl} = f_{ik} + f_{lj} - f_{il} - f_{jk}$, we finish the proof.

Formula (32) is shown by Lemma 15 and 16.

3.4.1.2   Comparison

In part two, we prove the asymptotic convergence result of (17).

Firstly, we obtain the comparison between the first and second derivative.

Consider

$$\text{Ratio}(\rho) := \frac{\sum_{(i,j)\in E}\sum_{(k,l)\in E} h_{ij,kl}(F_i - F_j)_+\rho_i(F_k - F_l)_+\rho_k}{\sum_{(i,j)\in E}(F_i - F_j)_+^2\rho_i}, \tag{34}$$

where $\sum_{(i,j)\in E}(F_i - F_j)_+^2\rho_i > 0$. By replacing $(F_i)_{i=1}^n$ by $(\Phi_i)_{i=1}^n \in \mathbb{R}^n$, we arrive at the Definition 4, which gives the lower bound of the ratio function. Recall that

$$\lambda_{\mathcal{F}}(\rho) = \min_{\Phi} \sum_{(i,j)\in E}\sum_{(k,l)\in E} h_{ij,kl}(\Phi_i - \Phi_j)_+\rho_i(\Phi_k - \Phi_l)_+\rho_k, \tag{35}$$

where the minimum is taken among all $(\Phi_i)_{i=1}^n \in \mathbb{R}^n$ with

$$\sum_{(i,j)\in E}(\Phi_i - \Phi_j)_+^2\rho_i = 1. \tag{36}$$

Lemma 17 *If $\rho \in \mathcal{P}_o(G)$, then*

$$Ratio(\rho) \geq \lambda_{\mathcal{F}}(\rho).$$

Proof 14 *For any given vector $F = (F_i)_{i=1}^n \in \mathbb{R}^n$ with*

$$C(F) := \sum_{(i,j)\in E}(F_i - F_j)_+^2\rho_i > 0,$$

consider a vector $\bar{\Phi} = (\bar{\Phi}_i)_{i=1}^n$ with $\bar{\Phi}_i = \frac{F_i}{\sqrt{C(F)}}$. Then we have

$$\sum_{(i,j)\in E}(\bar{\Phi}_i - \bar{\Phi}_j)_+^2\rho_i = 1.$$

From (34),

$$Ratio(\rho) = \sum_{(i,j)\in E}\sum_{(k,l)\in E} h_{ij,kl}(\bar{\Phi}_i - \bar{\Phi}_j)_+\rho_i(\bar{\Phi}_k - \bar{\Phi}_l)_+\rho_k$$

$$\leq \min_{\Phi} \sum_{(i,j)\in E}\sum_{(k,l)\in E} h_{ij,kl}(\Phi_i - \Phi_j)_+\rho_i(\Phi_k - \Phi_l)_+\rho_k = \lambda_{\mathcal{F}}(\rho).$$

Thus, we obtain the comparison between the first and second derivative.

Lemma 18

$$\frac{d^2}{dt^2}\mathcal{F}(\rho(t)) \geq -\big(2\lambda_\mathcal{F}(\rho(t)) - m(\rho(t))\big)\frac{d}{dt}\mathcal{F}(\rho(t)),$$

where $m(\rho(t))$ is defined in Lemma 15.

Proof 15 *From Lemma 16 and definition 4, we have*

$$\spadesuit = 2\sum_{(i,j)\in E}\sum_{(k,l)\in E} h_{ij,kl}(F_i - F_j)_+\rho_i(F_k - F_l)_+\rho_k$$

$$\geq 2\lambda_\mathcal{F}(\rho(t))\sum_{(i,j)\in E}(F_i - F_j)_+^2\rho_i$$

$$= -2\lambda_\mathcal{F}(\rho(t))\frac{d}{dt}\mathcal{F}(\rho(t)).$$

Then combining Lemma 15 and 16, we know

$$\frac{d^2}{dt^2}\mathcal{F}(\rho(t)) = \clubsuit + \spadesuit \geq -\big(2\lambda_\mathcal{F}(\rho(t)) - m(\rho(t))\big)\frac{d}{dt}\mathcal{F}(\rho(t)).$$

Secondly, we show $\lim_{t\to\infty}\lambda_\mathcal{F}(\rho(t)) = \lambda_\mathcal{F}(\rho^\infty)$. The asymptotic comparison rate is determined by the given Gibbs measure.

Lemma 19 *$\lambda_\mathcal{F}(\rho)$ is a continuous function with respect to $\rho \in \mathcal{P}_o(G)$.*

Proof 16 *We observe that (35), (36) remains the same for Φ modulo any additive constant. Without loss of generality, we let $\Phi_n = 0$, thus (35), (36) is uniquely determined by $(\Phi_i)_{i=1}^{n-1}$. In other words, if we denote*

$$D = \{(\Phi_i)_{i=1}^{n-1} \in \mathbb{R}^{n-1} \mid (36) \text{ holds with } \Phi_n = 0\},$$

and

$$\alpha(\rho, \Phi) = \sum_{(i,j)\in E}\sum_{(k,l)\in E} h_{ij,kl}(\Phi_i - \Phi_j)_+\rho_i(\Phi_k - \Phi_l)_+\rho_k,$$

then, from Lemma 17, $\lambda_\mathcal{F}(\rho) = \min_{\Phi\in D}\alpha(\rho, \Phi)$.

For any $\rho \in \mathcal{P}_o(G)$, we consider a compact region $B \subset \mathcal{P}_o(G)$ in Euclidean metric, such that $\rho \in B$. To prove the continuity of $\lambda_\mathcal{F}(\rho)$, we need to show that α is uniformly continuous on $B \times D$. Since α is a continuous function, it is sufficient to show that D is a compact set.

Proof 17 (Proof of D being a compact set.) *Notice that D is a closed set. It is sufficient to show D is a bounded set in $\mathbb{R}^{n-1}$. Since (36) holds, for any fixed $i, j \in V$ with $(i, j) \in E$,*

$$|\Phi_i - \Phi_j| \leq \max\{\sqrt{\frac{1}{\rho_i}}, \sqrt{\frac{1}{\rho_j}}\} \leq \sqrt{\frac{1}{\min_{k \in V} \rho_k}}.$$

Since G is connected, there exists a finite sequence of edges that connect vertices i and n. In other words, there exists vertices $k_l \in V$, with $1 \leq l \leq m$, such that $k_1 = i$, $k_m = n$.

$$|\Phi_i| = |\Phi_i - \Phi_n| \leq \sum_{l=1}^{m-1} |\Phi_{k_{l+1}} - \Phi_{k_l}| \leq (m-1)\sqrt{\frac{1}{\min_{k \in V} \rho_k}} < \infty,$$

which finishes the proof.

Then, we show that $\lambda_{\mathcal{F}}$ is continuous from the uniform continuity of α. For any $\epsilon > 0$, there exists a constant $\delta > 0$ with

$$\alpha(\Phi^1, \rho^1) > \alpha(\Phi^2, \rho^2) - \epsilon,$$

when $\|\Phi^1 - \Phi^2\| < \delta$ and $\|\rho^1 - \rho^2\| < \delta$. Here $\|\cdot\|$ is an Euclidean norm. For fixed $\rho^1 \in B$, there exists a point $\Phi^1 \in D$ with $\lambda_{\mathcal{F}}(\rho^1) = \alpha(\Phi^1, \rho^1)$. Hence

$$\lambda_{\mathcal{F}}(\rho^1) > \alpha(\Phi^2, \rho^2) - \epsilon \geq \min_{\Phi \in D} \alpha(\Phi, \rho^2) - \epsilon = \lambda_{\mathcal{F}}(\rho^2) - \epsilon,$$

for all ρ^1, ρ^2 with $\|\rho^1 - \rho^2\| < \delta$. By symmetric ρ^1, ρ^2, we know

$$|\lambda_{\mathcal{F}}(\rho^1) - \lambda_{\mathcal{F}}(\rho^2)| < \epsilon,$$

when $|\rho^1 - \rho^2| < \delta$, which finishes the proof.

In all, we show the asymptotical convergence result.

Lemma 20 *Assume (A) holds and $\lambda_{\mathcal{F}}(\rho^\infty) > 0$, then for any sufficient small $\epsilon > 0$, there exists a constant time $T > 0$, such that when $t > T$,*

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

Proof 18 *We know that $\lim_{t\to\infty} \rho(t) = \rho^\infty$ implies two facts. On one hand, from Lemma 19, the continuity of $\lambda_{\mathcal{F}}(\rho)$ implies: for $0 < \epsilon << \lambda_{\mathcal{F}}(\rho^\infty)$, there exists a time T_1, when $t > T_1$,*

$$\lambda_{\mathcal{F}}(\rho(t)) \geq \lambda_{\mathcal{F}}(\rho^\infty) - \frac{\epsilon}{2}. \tag{37}$$

On the other hand,

$$\rho(t) \to \rho^\infty \quad \Rightarrow \quad \lim_{t\to\infty} \max_{(i,j)\in E} (F_i(\rho(t)) - F_j(\rho(t)))_+ = 0,$$

which further implies

$$\lim_{t\to\infty} m(\rho(t)) = \max_{(i,j)\in E} (F_i - F_j)_+ \max_{i\in V} deg(i) \frac{\max_{i\in V} \rho_i}{\min_{i\in V} \rho_i} = 0.$$

It means that there exists a time T_2, such that when $t > T_2$, $m(\rho(t)) \leq \epsilon$.

Let $T = \max\{T_1, T_2\}$ and consider $t \in [T, \infty)$. Since (37) and Lemma 18 holds,

$$\frac{d^2}{dt^2} \mathcal{F}(\rho(t)) \geq - \left(2\lambda_{\mathcal{F}}(\rho(t)) - m(\rho(t))\right) \frac{d}{dt} \mathcal{F}(\rho(t))$$
$$\geq - 2(\lambda_{\mathcal{F}}(\rho^\infty) - \epsilon) \frac{d}{dt} \mathcal{F}(\rho(t)). \tag{38}$$

Similar as in motivation, integrating (38) on $t \in [T, \infty)$,

$$\frac{d}{dt} [\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty)] \leq -2(\lambda_{\mathcal{F}}(\rho^\infty) - \epsilon)(\mathcal{F}(\rho(T)) - \mathcal{F}(\rho^\infty)).$$

Following the Gronwall's inequality of (17) with initial condition $\rho(T)$, we have

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

3.4.1.3   Main result

In part three, we present the proof of main results.

Proof 19 (Sketch of Theorem 5 proof) *Our proof is based on the dynamical viewpoint. Let T be defined in Lemma 20. We consider the convergence result in two time zones. If $t \leq T$, since the gradient flow can not arrive at the minimizer in finite time, the first derivative's lower bound gives one convergence rate C_1; If $t > T$, Lemma 20 has already provided the other exponential convergence rate $C_2 = 2(\lambda_{\mathcal{F}}(\rho^\infty) - \epsilon)$. Combing the above two facts, we obtain the overall convergence rate $C = \min\{C_1, C_2\}$.*

Proof 20 (Proof of Theorem 5) *If $\rho^0 = \rho^\infty$, the convergence result (18) is trivial. From now on, we consider $\rho^0 \neq \rho^\infty$.*

We shall discuss two zones, $[T, \infty)$ and $[0, T]$, where T is defined in Lemma 20.

For case one, consider $t \in [0, T]$, we show that (18) holds for a constant C_1. Denote

$$m_2 = \min_{0 \leq t \leq T} \sum_{(i,j) \in E} (F_i(\rho(t)) - F_j(\rho(t)))_+^2.$$

We show that $m_2 > 0$. Assume this is not true, suppose $m_2 = 0$. Since $\rho(t)$ is continuous and $[0, T]$ is a bounded region, there exists a time $T_0 \in [0, T]$, such that

$$\sum_{(i,j) \in E} (F_i(\rho(T_0)) - F_j(\rho(T_0)))_+^2 = 0.$$

Since $\rho_i(T_0) > 0$ for all $i \in V$, we have $F_1(\rho(T_0)) = \cdots = F_n(\rho(T_0))$. It implies $\frac{d\rho_i}{dt}|_{T_0} = 0$, for any $i \in V$, meaning that the equilibrium of ODE (17) is arrived at a finite time, which is impossible. Hence

$$
\begin{aligned}
\frac{d}{dt}(\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty)) &= -\sum_{i=1}^n \sum_{j \in N(i)} (F_i(\rho) - F_j(\rho))_+^2 \rho_i \\
&\leq -\sum_{(i,j) \in E} (F_i(\rho(t)) - F_j(\rho(t)))_+^2 \cdot \min_{t \geq 0,\ i \in V} \rho_i(t) \\
&\leq -\min_{0 \leq t \leq T_0} \sum_{(i,j) \in E} (F_i(\rho(t)) - F_j(\rho(t)))_+^2 \cdot \min_{t \geq 0,\ i \in V} \rho_i(t) \\
&\leq -m_2 \min_{t \geq 0,\ i \in V} \rho_i(t) \\
&= -\frac{m_2 \min_{t \geq 0,\ i \in V} \rho_i(t)}{\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty)}(\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty)) \\
&\leq -\frac{m_2 \min_{t \geq 0,\ i \in V} \rho_i(t)}{\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)}(\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty)),
\end{aligned}
$$

where the last inequality is from $\mathcal{F}(\rho(t)) \geq \mathcal{F}(\rho^0)$. From Gronwall's inequality,

$$\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty) \leq e^{-C_1 t}(\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)), \tag{39}$$

where

$$C_1 = \frac{m_2 \min_{t \geq 0,\ i \in V} \rho_i(t)}{\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)}.$$

For case two, we consider $t \in (T, \infty)$. Denote

$$C_2 = 2(\lambda_{\mathcal{F}}(\rho^\infty) - \epsilon).$$

We show that (18) holds for a constant $C = \min\{C_1, C_2\}$. Since

$$
\begin{aligned}
\mathcal{F}(\rho(t)) - \mathcal{F}(\rho^\infty) &\leq e^{-C_2(t-T)}(\mathcal{F}(\rho(T)) - \mathcal{F}(\rho^\infty)) \qquad \textit{Lemma 20} \\
&\leq e^{-C(t-T)}e^{-C_1 T}(\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)) \quad \textit{Since (39) holds} \\
&= e^{-Ct}e^{-(C_1-C)T}(\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)) \\
&\leq e^{-Ct}(\mathcal{F}(\rho^0) - \mathcal{F}(\rho^\infty)),
\end{aligned}
$$

where the last inequality is from $C_1 \geq C$, $e^{-(C_1-C)T} < 1$.

By combining the above two steps, we know that (18) holds for all $t \geq 0$.

3.4.2 Analysis of dissipation rate

In last section, we show that if $\lambda_{\mathcal{F}}(\rho^\infty) > 0$, the convergence result (18) holds. What is the explicit condition for $\lambda_{\mathcal{F}}(\rho^\infty) > 0$?

In this section, we give a clear answer to this question. That is we find the relation between the Hessian matrix of free energy in $\mathbb{R}^n$ and asymptotic convergence rate $\lambda_{\mathcal{F}}(\rho)$. The relation is shown by the following formula:

$$\sum_{(i,j)\in E}\sum_{(k,l)\in E} h_{ij,kl}(\Phi_i - \Phi_j)_+\rho_i(\Phi_k - \Phi_l)_+\rho_k = \left(\tilde{\mathrm{div}}_G(\rho\nabla_G\Phi)\right)^T \mathrm{Hess}_{\mathbb{R}^n}\mathcal{F}(\rho)\tilde{\mathrm{div}}_G(\rho\nabla_G\Phi),$$

$$(40)$$

where we recall $h_{ij,kl} = f_{ik} + f_{jl} - f_{il} - f_{jk}$ and denote

$$\tilde{\mathrm{div}}_G(\rho\nabla_G\Phi) := \Big(\sum_{j\in N(i)} (\Phi_i - \Phi_j)\tilde{g}_{ij} \Big)_{i=1}^n \quad \text{with} \quad \tilde{g}_{ij} := \begin{cases} \rho_i & \text{if } \Phi_i > \Phi_j; \\ \rho_j & \text{if } \Phi_i < \Phi_j. \end{cases}$$

Let's prove (40) by a direct calculation.

Proof 21 (Proof of Lemma 6) *Notice $\mathrm{Hess}_{\mathbb{R}^n}\mathcal{F}(\rho) = (f_{ik})_{i\in V, k\in V}$.*

Then

$$\left(\tilde{div}_G(\rho\nabla_G\Phi)\right)^T Hess_{\mathbb{R}^n}\mathcal{F}(\rho)\tilde{div}_G(\rho\nabla_G\Phi)$$

$$=\sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}\,\tilde{div}_G(\rho\nabla_G\Phi)|_i\,\tilde{div}_G(\rho\nabla_G\Phi)|_k$$

$$=\sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}[-\sum_{j\in N(i)}(\Phi_i-\Phi_j)\tilde{g}_{ij}][-\sum_{l\in N(k)}(\Phi_k-\Phi_l)\tilde{g}_{kl}]$$

$$=\sum_{(i,j)\in E}(\Phi_i-\Phi_j)\tilde{g}_{ij}\sum_{(k,l)\in E} f_{ik}(\Phi_k-\Phi_l)\tilde{g}_{kl}$$

$$=\sum_{(i,j)\in E}(\Phi_i-\Phi_j)\tilde{g}_{ij}\{\sum_{(k,l)\in E,\ \Phi_k>\Phi_l} f_{ik}(\Phi_k-\Phi_l)\rho_k+\sum_{(k,l)\in E,\ \Phi_k<\Phi_l} f_{ik}(\Phi_k-\Phi_l)\rho_l\}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{Relabel } k \textit{ and } l \textit{ for the second formula}$$

$$=\sum_{(i,j)\in E}(\Phi_i-\Phi_j)\tilde{g}_{ij}\{\sum_{(k,l)\in E,\ \Phi_k>\Phi_l} f_{ik}(\Phi_k-\Phi_l)\rho_k-\sum_{(k,l)\in E,\ \Phi_k>\Phi_l} f_{il}(\Phi_k-\Phi_l)\rho_k\}$$

$$=\sum_{(i,j)\in E}\sum_{(k,l)\in E}(f_{ik}-f_{il})(\Phi_i-\Phi_j)\tilde{g}_{ij}(\Phi_k-\Phi_l)_+\rho_k$$

$$=\sum_{(i,j)\in E,\ \Phi_i>\Phi_j}\sum_{(k,l)\in E}(f_{ik}-f_{il})(\Phi_i-\Phi_j)\rho_i(\Phi_k-\Phi_l)_+\rho_k$$

$$+\sum_{(i,j)\in E,\ \Phi_i<\Phi_j}\sum_{(k,l)\in E}(f_{ik}-f_{il})(\Phi_i-\Phi_j)\rho_j(\Phi_k-\Phi_l)_+\rho_k$$

$$\qquad\qquad\qquad\qquad \textit{Relabel } i \textit{ and } j \textit{ for the second formula}$$

$$=\sum_{(i,j)\in E,\ \Phi_i>\Phi_j}\sum_{(k,l)\in E}(f_{ik}-f_{il})(\Phi_i-\Phi_j)\rho_i(\Phi_k-\Phi_l)_+\rho_k$$

$$-\sum_{(i,j)\in E,\ \Phi_i>\Phi_j}\sum_{(k,l)\in E}(f_{jk}-f_{jl})(\Phi_i-\Phi_j)\rho_i(\Phi_k-\Phi_l)_+\rho_k$$

$$=\sum_{(i,j)\in E}\sum_{(k,l)\in E}(f_{ik}+f_{jl}-f_{il}-f_{jk})(\Phi_i-\Phi_j)_+\rho_i(\Phi_k-\Phi_l)_+\rho_k.$$

(40) immediately induces the relation between the convexity of free energy and convergence result.

Proof 22 (Proof of Lemma 7) *We are going to show*

$$\min_{\Phi\in D}\left(\tilde{div}_G(\rho\nabla_G\Phi)\right)^T Hess_{\mathbb{R}^n}\mathcal{F}(\rho)\tilde{div}_G(\rho\nabla_G\Phi)>0.$$

Assume this is not true, suppose

$$\min_{\Phi \in D} \left(\tilde{div}_G(\rho \nabla_G \Phi) \right)^T Hess_{\mathbb{R}^n} \mathcal{F}(\rho) \tilde{div}_G(\rho \nabla_G \Phi) = 0,$$

where

$$D = \{ (\Phi_i)_{i=1}^{n-1} \in \mathbb{R}^{n-1} \mid (36) \text{ holds with } \Phi_n = 0 \}.$$

Since D is a compact set in $\mathbb{R}^{n-1}$, there exists a $\Phi^* \in D$ such that

$$\left(\tilde{div}_G(\rho \nabla_G \Phi^*) \right)^T Hess_{\mathbb{R}^n} \mathcal{F}(\rho) \tilde{div}_G(\rho \nabla_G \Phi^*) = 0.$$

*Since $Hess_{\mathbb{R}^n} \mathcal{F}(\rho)$ is a positive definite matrix, $div_G(\rho \nabla_G \Phi^*) = 0$. Similarly as in the proof of Lemma 10, we **claim**: $\Phi_1^* = \Phi_2^* = \cdots = \Phi_n^*$. If the claim is true, then*

$$\sum_{(i,j) \in E} (\Phi_i^* - \Phi_j^*)_+^2 \rho_i = 0,$$

which contradicts $\Phi^* \in D$.

Proof 23 (Proof of claim) *Suppose it is not true. Let $c = \max_{i \in V} \Phi_i^*$. Since the graph G is connected, there exists $(k,l) \in E$, such that $\Phi_l^* = c$ and $\Phi_k^* < c$. By $\sum_{j \in N(l)} (\Phi_j^* - \Phi_l^*) \tilde{g}_{lj} = 0$, we have*

$$\Phi_l^* = \frac{\sum_{j \in N(l)} \tilde{g}_{lj}(\rho) \Phi_j^*}{\sum_{j \in N(l)} \tilde{g}_{lj}(\rho)} = c + \frac{\sum_{j \in N(l)} \tilde{g}_{lj}(\rho)(\Phi_j^* - c)}{\sum_{j \in N(l)} \tilde{g}_{lj}(\rho)} < c,$$

which contradicts $\Phi_l^* = c$.

Moreover, since $Hess_{\mathbb{R}^n} \mathcal{H} = diag(\frac{1}{\rho_i})_{1 \le i \le n}$ is a positive definite matrix, we know $\lambda_{\mathcal{H}}(\rho) > 0$ from the above argument.

From Theorem 5 and Lemma 7, we show convergence results for linear and nonlinear Fokker-Planck equations on graphs.

Proof 24 (Proof of corollary 8 and 9) *Since $Hess_{\mathbb{R}^n} \mathcal{F}(\rho) = \beta diag(\frac{1}{\rho_i})_{1 \le i \le n} > 0$ or $(w_{ij})_{1 \le i,j \le n} + \beta diag(\frac{1}{\rho_i})_{1 \le i \le n} > 0$, there exists a unique minimizer ρ^∞. From Lemma 7, we have $\lambda_{\mathcal{F}}(\rho^\infty) > 0$. From Theorem 5, we have the convergence results of linear and nonlinear Fokker-Planck equations.*

3.5 Functional inequalities

In this section, we recover several famous functional inequalities on finite graphs, which are based on the convergence result of Fokker-Planck equation on graphs.

In the literatures, these inequalities have been investigated for a long time, see [16, 41]. Because the lack of 2-Wasserstein metric on discrete states, many other methods have been adopted. Here we use a different way, which is a direct analog of continuous state [68, 80] via the discrete 2-Wasserstein metric. In short, we apply the convergence result to recover graph modified Log-Sobolev inequality (GLSI), which further implies the graph modified Talagrand's inequality (GTI) and Poincare inequality (GP).

In details, we introduce three concepts to measure the closeness of discrete measures. For any $\mu = (\mu_i)_{i=1}^n$, $\nu = (\nu_i)_{i=1}^n \in \mathcal{P}_o(G)$, we consider

- Graph relative entropy (H):

$$\mathcal{H}_\nu(\mu) := \sum_{i=1}^n \mu_i \log \frac{\mu_i}{\nu_i};$$

- Graph relative Fisher information (I):

$$\mathcal{I}_\nu(\mu) := \sum_{(i,j)\in E} (\log \frac{\mu_i}{\nu_i} - \log \frac{\mu_j}{\nu_j})_+^2 \mu_i;$$

- Graph 2-Wasserstein metric (W):

$$W_{2;\mathcal{H}_\nu}(\mu,\nu) := \inf\{\sqrt{\int_0^1 (\Phi,\Phi)_{\bar{\mu}}dt} \ : \ \frac{d}{dt}\bar{\mu}+\mathrm{div}_G(\bar{\mu}\nabla_G\Phi) = 0, \ \bar{\mu}(0) = \mu, \ \bar{\mu}(1) = \nu\}.$$

We prove several inequalities between H, I and W.

Theorem 21 *For any finite simple graph G, there exists a constant $\lambda > 0$, such that the following inequality holds.*

- *Graph modified Log-Sobolev-inequality:*

$$\mathcal{H}_\nu(\mu) \leq \frac{1}{2\lambda}\mathcal{I}_\nu(\mu) \quad (GLSI);$$

- *Graph modified Talagrand inequality:*

$$W_{2;\mathcal{H}_\nu}(\mu, \nu) \leq \sqrt{\frac{2\mathcal{H}_\nu(\mu)}{\lambda}} \quad (GT);$$

- *Graph modified Poincare inequality: for any* $(f_i)_{i=1}^n \in \mathbb{R}^n$ *with* $\sum_{i=1}^n f_i \nu_i = 0$,

$$\sum_{i=1}^n f_i^2 \nu_i \leq \frac{1}{\lambda} \sum_{(i,j) \in E} (f_i - f_j)_+^2 \nu_i \quad (GP).$$

Moreover, for a sufficient small $\epsilon > 0$, *there exists a open set* D *containing* ν *in Euclidean metric, such that if* $\mu \in D$, *(GLSI), (GT) hold for* $\lambda = \lambda_{\mathcal{H}}(\nu) - \epsilon$.

Remark 4 *Because the convergence rate is only found in asymptotic sense, we are not able to find the optimal bound for all inequalities. Instead, we provide an explicit local bound around a special point, which is the Gibbs measure.*

Before showing the proof, we observe that

$$\mathcal{H}_\nu(\mu) = \sum_{i=1}^n \mu_i \log \mu_i - \sum_{i=1}^n \log \nu_i \mu_i$$

is a summation of entropy and linear potential energy, whose minimizer is a Gibbs measure ν. Notice that the gradient flow of $\mathcal{H}_\nu(\mu)$ is

$$\frac{d\mu_i}{dt} = \sum_{j \in N(i)} (\log \frac{\mu_j}{\nu_j} - \log \frac{\mu_i}{\nu_i})_+ \mu_j - \sum_{j \in N(i)} (\log \frac{\mu_i}{\nu_i} - \log \frac{\mu_j}{\nu_j})_+ \mu_i. \tag{41}$$

Along the gradient flow (41), we observe that the "Fisher" information satisfies

$$\mathcal{I}_\nu(\mu(t)) = -\frac{d}{dt}\mathcal{H}_\nu(\mu(t)). \tag{42}$$

Under this observation, we show the connections between H, W, I.

Proof 25 (Outline of proof) *At the beginning, we prove GLSI by the convergence result, meaning that the convergence rate near the equilibrium* ν *recovers the GLSI inequality. Secondly, we use GLSI to show GT. Our proof follows the idea in Theorem 3 of [68]. That is the special calculation law between the metric and gradient flow. Lastly, we use GLSI to show GP. It follows the linearization idea of Rothaus [83, 68].*

Proof 26 (Proof of Theorem 21) *We begin with the proof of GLSI. We shall show*

$$\frac{1}{2\lambda} = \sup_{\mu \in \mathcal{P}_o(G)} \frac{\mathcal{H}_\nu(\mu)}{\mathcal{I}_\nu(\mu)} < +\infty.$$

We will prove this by dividing $\mathcal{P}(G)$ into two regions. I.e.

$$\mathcal{P}(G) = D \cup (\mathcal{P}(G) \setminus D).$$

Here D is constructed as follows: Given a sufficiently small $\epsilon > 0$, there exists a constant δ,

$$D = \{\mu \in \mathcal{P}_o(G) \ : \ \mathcal{H}_\nu(\mu) < \delta\},$$

such that when $\mu \in D$, $\lambda_{\mathcal{H}}(\mu) > \lambda_{\mathcal{H}}(\nu) - \epsilon$.

One one hand, for $\mu \in D$, we consider the gradient flow (41), with $\mu(t)$ starting at initial measure μ. Notice that $\mathcal{H}_\nu(\mu)$ is a Lyapunov function, $\mu(t) \in D$ for all $t > 0$. Following the convergence result, we have

$$\frac{d^2}{dt^2}\mathcal{H}_\nu(\mu(t)) \geq -2(\lambda_{\mathcal{H}}(\nu) - \epsilon)\frac{d}{dt}\mathcal{H}_\nu(\mu(t)).$$

Integrating on both sides for time interval (t, ∞), we obtain

$$\frac{d}{dt}\mathcal{H}_\nu(\mu(t))|_{t=\infty} - \frac{d}{dt}\mathcal{H}_\nu(\mu(t)) \geq 2(\lambda_{\mathcal{H}}(\nu) - \epsilon)(\mathcal{H}_\nu(\mu(t))|_{t=\infty} - \mathcal{H}_\nu(\mu(t))).$$

Notice that $\frac{d}{dt}\mathcal{H}_\nu(\mu(t))|_{t=\infty} = \mathcal{H}_\nu(\mu(t))|_{t=\infty} = 0$, the above formula forms

$$\mathcal{I}_\nu(\mu(t)) = -\frac{d}{dt}\mathcal{H}_\nu(\mu(t)) \geq 2(\lambda_{\mathcal{H}}(\nu) - \epsilon)\mathcal{H}_\nu(\mu(t)),$$

which implies

$$\lambda_1 = \sup_{\mu \in D} \frac{\mathcal{H}_\nu(\mu)}{\mathcal{I}_\nu(\mu)} \leq \frac{1}{2(\lambda_{\mathcal{H}}(\nu) - \epsilon)} < \infty.$$

On the other hand, for $\mu \in \mathcal{P}(G) \setminus D$. Since $\mathcal{H}_\nu, \mathcal{I}_\nu$ are continuous functions with respect to μ, it is not hard to check that $\mathcal{I}_\nu(\mu)$ is bounded below 0 and $\mathcal{H}_\nu(\mu)$ are bounded above. Then

$$\lambda_2 = \sup_{\mu \in \mathcal{P}(G) \setminus D} \frac{\mathcal{H}_\nu(\mu)}{\mathcal{I}_\nu(\mu)} \leq \frac{\sup_{\mu \in \mathcal{P}(G) \setminus D} \mathcal{H}_\nu(\mu)}{\inf_{\mu \in \mathcal{P}(G) \setminus D} \mathcal{I}_\nu(\mu)} < \infty.$$

Let $\frac{1}{2\lambda} = \max\{\lambda_1, \lambda_2\}$, we prove GLSI.

Secondly, we shall prove GT by using GLSI. We construct

$$\psi(t) = W_{2;\mathcal{H}_\nu}(\mu(t), \mu) + \sqrt{\frac{2\mathcal{H}_\nu(\mu(t))}{\lambda}},$$

where $\mu(t)$ is the solution of gradient flow (41) with $\mu(0) = \mu$. If we show $\psi(t)$ is a decreasing function, we finish the proof of GT. Since $\psi(0) \le \psi(\infty)$, $\psi(0) = \sqrt{\frac{2\mathcal{H}_\nu(\mu)}{\lambda}}$ and $\psi(\infty) = W_{2;\mathcal{H}_\nu}(\mu, \nu)$.

In order to show $\psi(t)$ is a decreasing function, we need the following **claims**:

Claim 1: $W_{2;\mathcal{H}_\nu}(\mu(t), \mu)$ is a Lipschitz continuous function with respect to t.

Proof 27 (Proof of Claim 1) For any time $a < b$, we observe that

$$\frac{W_{2;\mathcal{H}_\nu}(\mu(b), \mu) - W_{2;\mathcal{H}_\nu}(\mu(a), \mu)}{b - a} \le \frac{W_{2;\mathcal{H}_\nu}(\mu(a), \mu(b))}{b - a}.$$

And since

$$W_{2;\mathcal{H}_\nu}(\mu(a), \mu(b))^2$$
$$= \inf\{\int_0^1 (\Phi, \Phi)_{\bar{\mu}(s)} ds : \frac{d\bar{\mu}}{ds} + div_{\bar{\mu}}(\bar{\mu}\nabla_G\Phi) = 0, \ \bar{\mu}(0) = \mu(a), \ \bar{\mu}(1) = \mu(b)\},$$

means the minimum is taken among all possible continuity equation. We consider a particular continuity equation by letting $t = a + s(b - a)$, $\Phi(t) = \log\frac{\mu_i(t)}{\nu_i}$,

$$\frac{d\mu_i}{dt} = (b - a)\{\sum_{j \in N(i)} (\log\frac{\mu_j}{\nu_j} - \log\frac{\mu_i}{\nu_i})_+ \mu_j - \sum_{j \in N(i)} (\log\frac{\mu_i}{\nu_i} - \log\frac{\mu_j}{\nu_j})_+ \mu_i\}, \qquad (43)$$

Notice that (43) is slight modification of (41), which changes the time variable.

Hence

$$W_{2;\mathcal{H}_\nu}(\mu(a), \mu(b))^2 \le (b - a)^2 \int_0^1 (\log\frac{\mu_i(s)}{\nu_i}, \log\frac{\mu_i(s)}{\nu_i})_{\mu(s)} ds$$
$$= (b - a)^2 \int_0^1 \mathcal{I}_\nu(\mu(s)) ds$$
$$\le \sup_{s \in [0,1]} \mathcal{I}_\nu(\mu(s)).$$

Since we have proved that for initial condition μ, there exists a compact region $B_o \subset \mathcal{P}_o(G)$, such that $\mu(t) \in B_o$, for any $t > 0$. Hence

$$\sup_{s \in [0,1]} \mathcal{I}_\nu(\mu(s)) \leq \sup_{\mu \in B_o} \mathcal{I}_\nu(\mu) < \infty.$$

Hence

$$\frac{W_{2;\mathcal{H}_\nu}(\mu(b),\mu) - W_{2;\mathcal{H}_\nu}(\mu(a),\mu)}{b-a} \leq \frac{W_{2;\mathcal{H}_\nu}(\mu(a),\mu(b))}{b-a} \leq \sup_{\mu \in B_o} \mathcal{I}_\nu(\mu) < \infty,$$

which proves the claim.

Claim 2:

$$\frac{d}{dt}\Big|^+ W_{2;\mathcal{H}_\nu}(\mu(t),\mu) \leq \sqrt{\mathcal{I}_\nu(\mu(t))}, \quad \text{for } t \text{ a.e.}$$

Proof 28 (Proof of claim 2) Since the function $W_{2;\mathcal{H}_\nu}(\mu(t),\mu_0)$ is Lipschitz continuous with respect to t, it is also absolutely continuous. We only need to consider time t, such that $\frac{d}{dt}\big|^+ W_{2;\mathcal{H}_\nu}(\mu(t),\mu_0)$ exists. Then

$$\frac{d}{dt}\Big|^+ W_{2;\mathcal{H}_\nu}(\mu(t),\mu_0) = \lim_{h \to 0} \frac{W_{2;\mathcal{H}_\nu}(\mu(t+h),\mu) - W_{2;\mathcal{H}_\nu}(\mu(t),\mu)}{h}$$
$$\leq \limsup_{h \to 0} \frac{W_{2;\mathcal{H}_\nu}(\mu(t+h),\mu(t))}{h}.$$

To show the claim, we shall prove

$$\frac{W_{2;\mathcal{H}_\nu}(\mu(t+h),\mu(t))}{h} \leq \sqrt{\mathcal{I}_\nu(\mu(t))}.$$

This can be shown by the definition of metric. Since

$$W_{2;\mathcal{H}_\nu}(\mu(t+h),\mu(t))^2$$
$$= \inf\Big\{ \int_0^1 (\Phi,\Phi)_{\bar{\mu}(s)} ds : \frac{d\bar{\mu}}{ds} + div_G(\bar{\mu}\nabla_G\Phi) = 0, \ \bar{\mu}(0) = \mu(t), \ \bar{\mu}(1) = \mu(t+h) \Big\},$$

means the minimum is taken among all possible continuity equation. Here we consider a particular continuity equation by letting $t = sh$, $\Phi = \log\frac{\mu}{\nu} = (\log\frac{\mu_i(t)}{\nu_i})_{i=1}^n$

$$\frac{d}{dt}\mu_h - h\,div_G(\mu_h \nabla_G \log\frac{\mu_h}{v}) = 0,$$

58

which is a time rescaling version of (41).

Hence

$$W_{2;\mathcal{H}_\nu}(\mu(t+h),\mu(t))^2 \leq h^2 \int_0^1 (\log\frac{\mu_h(s)}{\nu}, \log\frac{\mu_h(s)}{\nu})_{\mu_h(s)}ds$$
$$=h^2 \int_0^1 \mathcal{I}_\nu(\mu_h(s))ds.$$

Notice that $\mathcal{I}_\nu(\mu)$ is a continuous function with respect to μ and $\mu_h(s) = \mu(t+sh)$. Then for any $\epsilon > 0$, there exists a $\bar{h} > 0$, such that when $0 < h < \bar{h}$,

$$\mathcal{I}_\nu(\mu_h(s)) \leq \mathcal{I}_\nu(\mu(t)) + \epsilon,$$

which implies

$$\frac{W_{2;\mathcal{H}_\nu}(\mu(t+h),\mu(t))^2}{h^2} \leq \int_0^1 \mathcal{I}_\nu(\mu(s))ds = \mathcal{I}_\nu(\mu(t)) + \epsilon.$$

Thus

$$\frac{d}{dt}|^+W_{2;\mathcal{H}_\nu}(\mu(t),\mu) \leq \limsup_{h\to 0} \frac{W_{2;\mathcal{H}_\nu}(\mu(t+h),\mu(t))^2}{h^2} \leq \mathcal{I}_\nu(\mu(t)) + \epsilon.$$

Because ϵ is arbitrary, we finish the proof.

Claim 3: *If GLSI holds, then*

$$\sqrt{\mathcal{I}_\nu(\mu)} \leq \frac{\mathcal{I}_\nu(\mu)}{\sqrt{2\lambda\mathcal{H}_\nu(\mu)}}.$$

Proof 29

$$\sqrt{\mathcal{I}_\nu(\mu)} = \frac{\mathcal{I}_\nu(\mu)}{\sqrt{\mathcal{I}_\nu(\mu)}} \leq \frac{\mathcal{I}_\nu(\mu)}{\sqrt{2\lambda H_\nu(\mu)}}.$$

In all, we are ready to show $\psi(t)$ is a decreasing function. Since $\psi(t)$ is an absolute continuous function from claim 1, for any $a > b > 0$,

$$\psi(b) - \psi(a) = \int_a^b (\frac{d}{dt}|^+W_{2;\mathcal{H}_\nu}(\mu(t),\mu) + \frac{d}{dt}\sqrt{\frac{2\mathcal{H}_\nu(\mu(t))}{\lambda}})dt$$
$$\leq \int_a^b (\sqrt{\mathcal{I}_\nu(\mu(t))} + \frac{1}{2}\mathcal{H}_\nu(\mu(t))^{-\frac{1}{2}} \cdot \sqrt{\frac{2}{\lambda}}\frac{d}{dt}\mathcal{H}_\nu(\mu(t)))dt \quad By \ claim \ 2$$
$$\leq \int_a^b (\sqrt{\mathcal{I}_\nu(\mu(t))} - \frac{\mathcal{I}_\nu(\mu(t))}{\sqrt{2\lambda\mathcal{H}_\nu(\mu(t))}})dt \quad By \ (43)$$
$$\leq \int_a^b 0dt = 0 \quad By \ claim \ 3.$$

Hence we finish the proof.

Lastly, we prove GP by the linearization of GLSI. Let's construct $\mu^\epsilon = (\mu_i^\epsilon)_{i=1}^n = ((1 + \epsilon f_i)\nu_i)_{i=1}^n$, with $\sum_{i=1}^n f_i \nu_i = 0$. Clearly, $\mu^\epsilon \in \mathcal{P}_o(G)$. As ϵ goes to 0, we show **claim 4:**

$$\frac{\mathcal{H}_\nu(\mu^\epsilon)}{\epsilon^2} \to \frac{1}{2} \sum_{i=1}^n f_i^2 \nu_i,$$

and

$$\frac{\mathcal{I}_\nu(\mu^\epsilon)}{\epsilon^2} \to \frac{1}{2} \sum_{(i,j)\in E} (f_i - f_j)_+^2 \nu_i.$$

From claim 4, GLSI implies GP.

Let's show claim 4 in details. By using the Taylor expansion, $\log(1 + \epsilon f_i) = \epsilon f_i - \frac{1}{2}(\epsilon f_i)^2 + o(\epsilon^2)$,

$$\begin{aligned}
\frac{\mathcal{H}_\nu(\mu^\epsilon)}{\epsilon^2} &= \sum_{i=1}^n \mu_i \log \frac{\mu_i^\epsilon}{\nu_i} \\
&= \sum_{i=1}^n \frac{(1 + \epsilon f_i)\nu_i \log(1 + \epsilon f_i)}{\epsilon^2} \\
&= \sum_{i=1}^n \frac{(1 + \epsilon f_i)(\epsilon f_i - \frac{1}{2}(\epsilon f_i)^2)}{\epsilon^2} \nu_i + O(\epsilon) \\
&= \sum_{i=1}^n \frac{1}{\epsilon} f_i \nu_i + \frac{1}{2} \sum_{i=1}^n f_i^2 \nu_i + O(\epsilon) \\
&= \frac{1}{2} \sum_{i=1}^n f_i^2 \nu_i + O(\epsilon),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\mathcal{I}_\nu(\mu^\epsilon)}{\epsilon^2} &= \sum_{(i,j)\in E} (\log \frac{\mu_i^\epsilon}{\nu_i} - \log \frac{\mu_j^\epsilon}{\nu_j})_+^2 \mu_i^\epsilon \\
&= \sum_{(i,j)\in E} (\frac{\log(1 + \epsilon f_i) - \log(1 + \epsilon f_j)}{\epsilon})_+^2 (1 + \epsilon f_i)\nu_i \\
&= \sum_{(i,j)\in E} (\frac{\epsilon f_i - \epsilon f_j + O(\epsilon^2)}{\epsilon})_+^2 (1 + \epsilon f_i)\nu_i \\
&= \sum_{(i,j)\in E} (f_i - f_j)_+^2 \nu_i + O(\epsilon).
\end{aligned}$$

60

3.6   Examples

In this section, we demonstrate Fokker-Planck equations on graphs through several examples.

3.6.1   Toy examples

Example 2 (Multiple Gibbs measures) *Consider a lattice graph with three vertices:*

We consider the free energy

$$\mathcal{F}(\rho) = \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} w_{ij} \rho_i \rho_j + \beta \sum_{i=1}^{3} \rho_i \log \rho_i.$$

with Gibbs measures

$$\rho_i^\infty = \frac{1}{K} e^{-\frac{\sum_{j=1}^3 w_{ij} \rho_j^\infty}{\beta}}, \quad K = \sum_{i=1}^{n} e^{-\frac{\sum_{j=1}^3 w_{ij} \rho_j^\infty}{\beta}}.$$

Let β=0.1 and

$$(w_{ij})_{1 \leq i,j \leq 3} = - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

In this case, since $(w_{ij})_{1 \leq i,j \leq 3}$ is semi negative definite, there are two minimizers (Gibbs measures) of free energy, see Figure 2.

Figure 2: Plot of two Gibbs measures: one is $(0.0001, 0.4729, 0.5270)$, the other is $(0.9986, 0.0007, 0.0007)$.

61

Example 3 (Gradient flow) *In example 2's setting, we consider the gradient flow* (17), *see Figure 3:*

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} \rho_j (\sum_{i=1}^{n} w_{ij}\rho_i - \sum_{j=1}^{n} w_{ij}\rho_j + \beta \log \rho_j - \beta \log \rho_i)_+$$

$$- \sum_{j \in N(i)} \rho_i (\sum_{j=1}^{n} w_{ij}\rho_j - \sum_{i=1}^{n} w_{ij}\rho_i + \beta \log \rho_i - \beta \log \rho_j)_+.$$
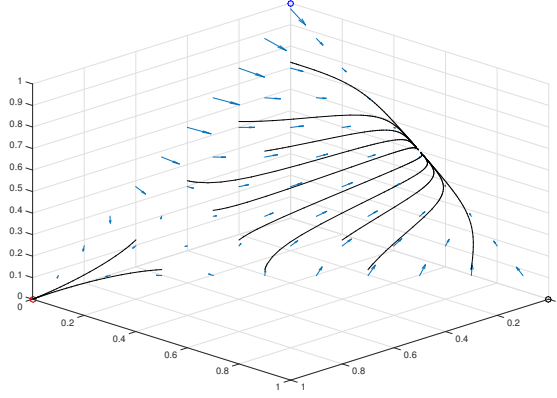
Figure 3: $\beta = 0.1$, vector field of (17).

Example 4 (Convergence rate) *Let's use Theorem 31 to demonstrate the convergence result in example 3. Because this graph has two edges, we know*

$$h_{12,12} = w_{11} + w_{22} - 2w_{12} + \beta(\frac{1}{\rho_1} + \frac{1}{\rho_2})$$

$$h_{12,23} = w_{12} + w_{23} - w_{13} - w_{22} + \beta(-\frac{1}{\rho_2})$$

$$h_{23,23} = w_{22} + w_{33} - 2w_{23} + \beta(\frac{1}{\rho_2} + \frac{1}{\rho_3}).$$

From Definition 4, we have

$$\lambda_{\mathcal{F}}(\rho) = \min_{\Phi} \ h_{12,12}(\Phi_1 - \Phi_2)^2 g_{12}^2 + 2h_{12,23}(\Phi_1 - \Phi_2)g_{12}(\Phi_2 - \Phi_3)g_{23}$$

$$+ h_{23,23}(\Phi_2 - \Phi_3)^2 g_{23}^2$$

s.t.

$$(\Phi_1 - \Phi_2)^2 g_{12} + (\Phi_2 - \Phi_3)^2 g_{23} = 1.$$

Here

$$g_{12} = \begin{cases} \rho_1 & \textit{if } \Phi_1 \geq \Phi_2, \\ \rho_2 & \textit{if } \Phi_1 < \Phi_2, \end{cases} \qquad g_{23} = \begin{cases} \rho_2 & \textit{if } \Phi_2 \geq \Phi_3, \\ \rho_3 & \textit{if } \Phi_2 < \Phi_3. \end{cases}$$

By solving the above optimization explicitly, we have

$$\lambda_{\mathcal{F}}(\rho) \geq \min_{g_{12}, g_{23}} 2 g_{12} g_{23} \frac{h_{12,12} h_{23,23} - h_{12,23}^2}{h_{12,12} g_{12} + h_{23,23} g_{23} + \sqrt{(h_{12,12} g_{12} - h_{23,23} g_{23})^2 + 4 h_{12,23}^2 g_{12} g_{23}}}. \quad (a)$$

From Theorem 4, if $h_{12,12} h_{23,23} - h_{12,23}^2 > 0$, the solution of (17) converges to ρ^∞ exponentially with asymptotic rate no less than $2(a)$.

3.6.2 Graph structure

In this sequel, we are curious about how the structure of graph affects the convergence rate. In particular, we consider a linear entropy

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

whose gradient flow, "Heat flow on a graph", is

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} (\log \rho_j - \log \rho_i)_+ \rho_j - \sum_{j \in N(i)} (\log \rho_i - \log \rho_j)_+ \rho_i, \qquad (44)$$

with the unique Gibbs measure $\rho^\infty = 1 = (\frac{1}{n}, \cdots, \frac{1}{n})$.

We design a numerical way to find the asymptotic convergence rate of (44). Fix a large enough time T, let $\rho(T)$ be the solution of (44) with initial condition ρ^0. Numerically,

$$\lambda_{\mathcal{H}}(1) \approx \frac{1}{2} \min_{\rho^0 \in \mathcal{P}_o(G)} \log \frac{\mathcal{H}(\rho(T)) - \mathcal{H}(1)}{\mathcal{H}(\rho(T+1)) - \mathcal{H}(1)}.$$

By the above numerical formula, we investigate how the structure of graph affects the asymptotical convergence rate. In next three examples, we numerically solve the convergence rates for three well know graphs.

Example 5 *Consider a circular graph C_n.*
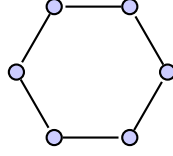
$$\lambda_{\mathcal{H}}(1) \approx \frac{4\pi^2}{n^2}.$$

Consider a star graph S_n.
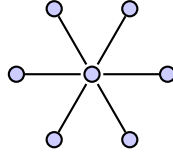
$$\lambda_{\mathcal{H}}(1) \approx 1.$$

Consider a complete graph K_n.
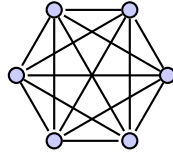
$$\lambda_{\mathcal{H}}(1) \approx n.$$

Remark 5 *Above convergence rates are numerically checked for $10 \leq n \leq 30$.*

Furthermore, we investigate how the change of graph structure, adding or deleting certain edges, affects the convergence rate.

Example 6 *One example is that, we consider a 6 vertices' graph with both cycle and star graph edges.*
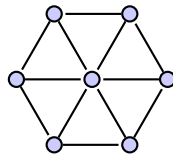
$$\lambda_{\mathcal{H}}(1) \approx 2.055.$$

By deleting one specify edge, we obtain two different rates.
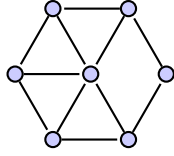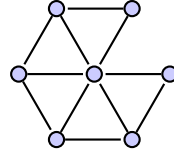
$$\lambda_{\mathcal{H}}(1) \approx 1.509. \qquad\qquad \lambda_{\mathcal{H}}(1) \approx 1.386.$$

The other example is that, we consider a graph with two cycles connected by one edge.
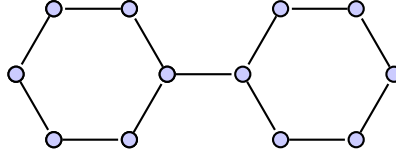
$$\lambda_{\mathcal{H}}(1) \approx 0.1580.$$

By deleting one specify edge, we obtain three different rates.
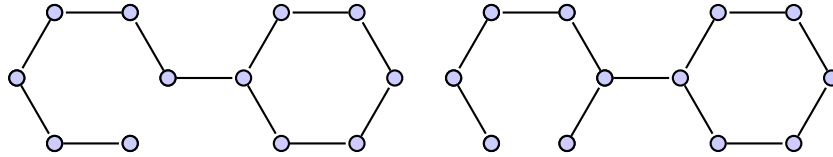
(a) $\lambda_{\mathcal{H}}(1) \approx 0.0918.$ \qquad\qquad (b) $\lambda_{\mathcal{H}}(1) \approx 0.1164.$
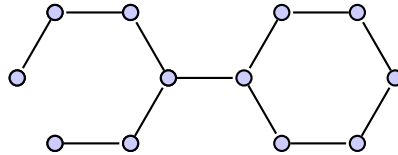
(c) $\lambda_{\mathcal{H}}(1) \approx 0.1516.$

Figure 4

Another example is that, we consider the following graph:
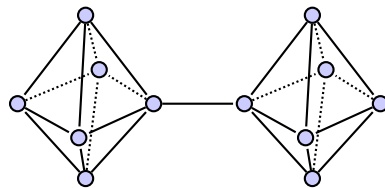
$$\lambda_{\mathcal{H}}(1) \approx 0.2426.$$

By deleting one specify edge, we obtain two different rates.

(a) $\lambda_{\mathcal{H}}(1) \approx 0.2276$.   (b) $\lambda_{\mathcal{H}}(1) \approx 0.2406$.
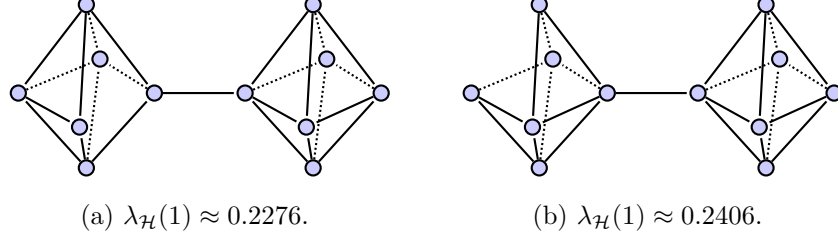
Figure 5

3.7   Conclusions

In this section, we summarize all results. Facing the optimization problem

$$\min_{\rho}\{\mathcal{F}(\rho) = v^T\rho + \frac{1}{2}\rho^T W\rho + \beta \sum_{i=1}^{n} \rho_i \log \rho_i \ : \ \sum_{i=1}^{n} \rho_i = 1, \quad \rho_i \geq 0\}, \qquad (45)$$

where $v = (v_i)_{i=1}^n$ is a constant vector and $W = (w_{ij})_{1\leq i,j\leq n}$ is a constant symmetric matrix, we introduce a gradient flow of (45) associated with graph structure G.

In details, we define the divergence operator with respect to ρ on finite graphs, such that the gradient flow of free energy (17) forms

$$\frac{d\rho}{dt} = \mathrm{div}_G(\rho \nabla_G F(\rho)), \quad F(\rho) = (\frac{\partial}{\partial \rho_i}\mathcal{F}(\rho))_{i=1}^n.$$

(17) is a gradient flow, since it has the following properties:

(a) $\mathcal{F}(\rho)$ is a Lyapunov function of (17):

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = -(F(\rho), F(\rho))_\rho = -g(\frac{d\rho}{dt}, \frac{d\rho}{dt});$$

(b) The minimizers of $\mathcal{F}(\rho)$, named Gibbs measures are equilibria of (17);

(c) If a Gibbs measure ρ^∞ is a strictly local minimizer of $\mathcal{F}(\rho)$, then

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

where C is a positive constant depending on ρ^0 and graph G's structure.

Importantly, this approach reflects the **effect of entropy** along with (17).

(i) The entropy induces the gradient flow (17)'s boundary repeller property. Notice that this property helps optimization (45) handles the boundary;

(ii) The entropy is the key to (17)'s convergence result. Recall that

$$\frac{d^2}{dt^2}\mathcal{F}(\rho(t)) = 2(\frac{d\rho}{dt})^T\big[W + \beta\mathrm{diag}(\frac{1}{\rho_i})_{1\leq i\leq n}\big]\frac{d\rho}{dt} + o(\frac{d}{dt}\mathcal{F}(\rho(t))).$$

One one hand, the entropy improves the convergence rate of (17), since

$$\mathrm{Hess}_{\mathbb{R}^n}\mathcal{H}(\rho) = \mathrm{diag}(\frac{1}{\rho_i})_{1\leq i\leq n}$$

is a positive definite matrix; On the other hand, the small order term is a result of boundary repeller property (i), which is crucial for convergence result!

To emphasize the impact of the difference of $\log\rho$ term in (17), we define a new Laplace operator[12] on finite graphs:

$$\Delta_G\rho := \mathrm{div}_G(\rho\nabla_G\log\rho),$$

where $\nabla_G\log\rho = (\log\rho_i - \log\rho_j)_{(i,j)\in E}$ and $g_{ij}(\rho)$ is defined in (19). We name Δ_G as Log-Laplacian, whose behavior in modeling and numerics are studied in next two chapters.

Remark 6 *The Log-Laplacian*

$$\Delta_G\rho = \sum_{j\in N(i)}(\log\rho_i - \log\rho_j)g_{ij}(\rho),$$

is clearly different from the currently known graph Laplacian, $\sum_{j\in N(i)}(\rho_j - \rho_i)$.

3.8 Relation with Villani's open problem

In this sequel, motivated by the convergence result in Theorem 4, we plan to find its analog in continuous states. Amazingly, this analog is related to Villani's open problem 15.11 in [95]:

---

[12]In continuous state, $\Delta\rho = \nabla\cdot(\rho\nabla\log\rho)$.

Find a **nice** formula for the Hessian of the functional $\mathcal{F}(\rho)$.

In general, it is hard to find a nice formula to represent the Hessian of nonlinear free energy with respect to all measures. However, for a special measure, such as a Gibbs measure, it is possible to do so. In other words, we find a formula to represent the Hessian of nonlinear free energy at its critical point. This formula gives us enough reasons, that the definition 4 plays the role of Hessian operator at Gibbs measure on discrete states.

3.8.1 Hessian operator of free energy at Gibbs measure

In this sequel, we shall derive the formula directly. We start with some notations: $\mathcal{M}$ is the underlying state, which is a smooth finite (d) dimensional Riemannian manifold. $\mathcal{P}_2(\mathcal{M})$ is the probability manifold supported on $\mathcal{M}$ embedded with 2-Wasserstein metric. The smooth functional, named free energy, is $\mathcal{F}: \mathcal{P}(\mathcal{M}) \to \mathbb{R}$. We consider the optimization problem

$$\min_{\mathcal{P}(\mathcal{M})} \mathcal{F}(\rho).$$

The local minimizer lying in the interior of $\mathcal{P}(\mathcal{M})$ is denoted as ρ^*, i.e.

$$\nabla \frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho)|_{\rho=\rho^*} = 0, \quad \text{for any } x \in \mathcal{M}.$$

In this setting, we will calculate the Hessian operator of objective functional $\mathcal{F}(\rho)$ at the local minimizer ρ^* on metric manifold $\mathcal{P}_2(\mathcal{M})$. To do so, we recall the Otto calculus in [95]. The geodesic on $\mathcal{P}_2(\mathcal{M})$ satifies

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

Here the continuity equation describes the motion of the measure and $\nabla \Phi_t$ can be understood as the velocity of geodesic. It is known that the gradient and Hessian of

free energy on $\mathcal{P}_2(\mathcal{M})$ is derived as follows:

$$(\mathrm{grad}_{\mathcal{P}_2(\mathcal{M})}\mathcal{F}, \nabla\Phi)_\rho := \frac{d}{dt}\mathcal{F}(\rho_t)|_{t=0}$$

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

Theorem 22

$$(\mathit{Hess}_{\mathcal{P}_2(\mathcal{M})}\mathcal{F} \cdot \nabla\Phi, \nabla\Phi)_{\rho^*}$$

$$= \int_{\mathcal{M}} \int_{\mathcal{M}} \frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)\nabla \cdot (\rho^*(x)\nabla\Phi(x))\nabla \cdot (\rho^*(y)\nabla\Phi(y))dxdy \qquad (46)$$

$$= \int_{\mathcal{M}} \int_{\mathcal{M}} (D_{xy}\frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)\nabla\Phi(x), \nabla\Phi(y))\rho^*(x)\rho^*(y)dxdy,$$

where $\frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho)$ is the second variational formula of functional $\mathcal{F}(\rho)$, $(\cdot, \cdot)$ is Euclidean inner product in $\mathbb{R}^d$.

Proof 30 *Let ρ_t satisfy the geodesic equation with initial measure ρ^*. Then the first derivative is a well known formula in Otto calculus:*

$$\frac{d}{dt}\mathcal{F}(\rho_t) = -\int_{\mathcal{M}} \frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho_t)\nabla \cdot (\nabla\Phi_t(x)\rho_t(x))dx$$

$$= \int_{\mathcal{M}} \nabla\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho_t) \cdot \nabla\Phi_t(x)\rho_t(x)dx. \qquad (47)$$

In addition, the second derivative is derived by the product law on (47):

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

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

Since $\rho_0 = \rho^*$, $\nabla\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho_t)|_{t=0} = 0$ implies $A_2|_{t=0} = 0$, we obtain

$$\frac{d^2}{dt^2}\mathcal{F}(\rho_t)|_{t=0} = (A_1) = \int_{\mathcal{M}} \frac{d}{dt}[\nabla\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho_t)] \cdot \nabla\Phi_t(x)\rho_t(x)dx$$

$$= \int_{\mathcal{M}} \nabla\frac{d}{dt}\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho_t) \cdot \nabla\Phi_t(x)\rho_t(x)dx \qquad (48)$$

$$= -\int_{\mathcal{M}} \frac{d}{dt}\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho_t)\nabla \cdot (\nabla\Phi_t(x)\rho_t(x))dx$$

69

Notice that $\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho)$ is also a smooth functional. Run the Otto calculus on this new functional, we have

$$\frac{d}{dt}\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho_t) = -\int_{\mathcal{M}}\frac{\delta^2}{\delta\rho(x)\delta\rho(y)}\mathcal{F}(\rho_t)\nabla\cdot(\nabla\Phi_t(y)\rho_t(y))dy. \tag{49}$$

Substitute (49) into (48), we finish the first equality. Through integration by parts with respect to x and y respectively, we can prove the second equality. In all, we finish the proof.

We illustrate the formula (46) by three examples.

Example 7 *Consider a free energy with linear potential energy and linear entropy with underlying state $\mathbb{R}^d$.*

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^d}V(x)\rho(x)dx + \beta\int_{\mathbb{R}^d}\rho(x)\log\rho(x)dx.$$

Here the Gibbs measure is

$$\rho^*(x) = \frac{1}{K}e^{-\frac{V(x)}{\beta}}, \quad K = \int_{\mathbb{R}^d}e^{-\frac{V(x)}{\beta}}dx.$$

The formula (46) shows that

$$(Hess_{\mathcal{P}_2(\mathbb{R}^d)}\mathcal{F}\cdot\nabla\Phi,\nabla\Phi)_\rho = \beta\int_{\mathbb{R}^d}(\nabla\cdot(\rho^*\nabla\Phi))^2\frac{1}{\rho^*(x)}dx. \tag{50}$$

We demonstrate that (50) is a new representation of the well known Hessian operator.

Lemma 23

$$\int_{\mathbb{R}^d}[(D^2V\cdot\nabla\Phi,\nabla\Phi) + \beta\,tr(D^2\Phi D^2\Phi^T)]\rho^*(x)dx = \beta\int_{\mathbb{R}^d}(\nabla\cdot(\rho^*\nabla\Phi))^2\frac{1}{\rho^*(x)}dx, \tag{51}$$

where ρ^* is the Gibbs measure

$$\rho^*(x) = \frac{1}{K}e^{-\frac{V(x)}{\beta}}, \quad K = \int_{\mathbb{R}^d}e^{-\frac{V(x)}{\beta}}dx,$$

and $(\cdot,\cdot)$ is an inner product in $\mathbb{R}^d$.

Remark 7 *L.H.S. of the above formula is a well known Hessian operator in optimal transport [95].*

Proof 31 *To simplify this validation, we let $\beta = 1$. It is not hard to check that the latter proof works for any $\beta > 0$. Since*

$$\int_{\mathbb{R}^d} (\nabla \cdot (\rho^* \nabla \Phi))^2 \frac{1}{\rho^*(x)} dx = \int_{\mathbb{R}^d} (\Delta \Phi \rho^*(x) + \nabla \rho^* \cdot \nabla \Phi)^2 \frac{1}{\rho^*(x)} dx$$

$$= \int_{\mathbb{R}^d} (\Delta \Phi)^2 \rho^*(x) + 2\Delta \Phi (\nabla \rho^* \cdot \nabla \Phi) + \frac{1}{\rho^*(x)} (\nabla \rho^* \cdot \nabla \Phi)^2 dx$$

$$\text{Since } \nabla \rho^* = -\nabla V \rho^*$$

$$= \int_{\mathbb{R}^d} (\Delta \Phi)^2 \rho^*(x) + 2\Delta \Phi (\nabla \rho^* \cdot \nabla \Phi) + (\nabla V \cdot \nabla \Phi)^2 \rho^* dx.$$

$$(a) \qquad\qquad (b) \qquad\qquad (c)$$

Let's calculate (a), (b) separately. We start with (a).

$$(a) = \int_{\mathbb{R}^d} (\Delta \Phi)^2 \rho^*(x) dx = \int_{\mathbb{R}^d} \nabla \cdot (\nabla \Phi) \Delta \Phi \rho^* dx$$

$$= -\int_{\mathbb{R}^d} \nabla \Phi \cdot \nabla (\Delta \Phi) \rho^* dx - \int_{\mathbb{R}^d} (\nabla \Phi \cdot \nabla \rho^*) \Delta \Phi dx.$$

Since[13]

$$-\nabla \Phi \cdot \nabla (\Delta \Phi) = -\Delta \frac{|\Delta \Phi|^2}{2} + tr(D^2 \Phi D^2 \Phi^T),$$

we have

$$(a) = \int_{\mathbb{R}^d} \left(-\Delta \frac{|\nabla \Phi|^2}{2} + tr(D^2 \Phi D^2 \Phi^T)\right) \rho^* dx - \int_{\mathbb{R}^d} (\nabla \Phi \cdot \nabla \rho^*) \Delta \Phi dx$$

$$= -\int_{\mathbb{R}^d} (D^2 \Phi \cdot \nabla \Phi, \nabla \rho^*) dx + \int_{\mathbb{R}^d} tr(D^2 \Phi D^2 \Phi^T) \rho^* dx - \frac{1}{2}(b).$$

Let's estimate (b).

$$\frac{1}{2}(b) = \int_{\mathbb{R}^d} \nabla \cdot (\nabla \Phi)(\nabla \rho^* \cdot \nabla \Phi) dx$$

$$= -\int_{\mathbb{R}^d} \nabla \Phi \cdot \nabla (\nabla \rho^* \cdot \nabla \Phi) dx$$

$$= -\int_{\mathbb{R}^d} [(D^2 \rho^* \cdot \nabla \Phi, \nabla \Phi) + (D^2 \Phi \cdot \nabla \rho^*, \nabla \Phi)] dx$$

$$= \int_{\mathbb{R}^d} [(D^2 V \cdot \nabla \Phi, \nabla \Phi) \rho^* - (\nabla V \cdot \nabla \Phi)^2 \rho^*] + (D^2 \Phi \cdot \nabla \rho^*, \nabla \Phi) dx,$$

---

[13] *It is a Bochner's formula in $\mathbb{R}^d$ with the Ricci curvature tensor as 0.*

where the last equality is from $D^2\rho^* = -D^2V\rho^* + \nabla V \otimes \nabla V \rho^*$.

In all,

$$(a) + (b) + (c) = \int_{\mathbb{R}^d} [(D^2V\nabla\Phi, \nabla\Phi) + tr(D^2\Phi D^2\Phi^T)]\rho^*(x)dx$$

which finishes the proof.

Example 8 *We consider a free energy with linear, interaction potential energy and linear entropy with underlying state $\mathbb{R}^d$.*

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^d} V(x)\rho(x)dx + \frac{1}{2}\int_{\mathbb{R}^d \times \mathbb{R}^d} W(x,y)\rho(x)\rho(y)dxdy + \beta \int_{\mathbb{R}^d} \rho(x)\log\rho(x)dx.$$

Here the Gibbs measure is

$$\rho^*(x) = \frac{1}{K}e^{-\frac{V(x)+\int_{\mathbb{R}^d} W(x,y)\rho^*(y)dy}{\beta}}, \quad K = \int_{\mathbb{R}^d} e^{-\frac{V(x)+\int_{\mathbb{R}^d} W(x,y)\rho^*(y)dy}{\beta}}dx.$$

(46) shows that

$$(Hess_{\mathcal{P}_2(\mathbb{R}^d)}\mathcal{F} \cdot \nabla\Phi, \nabla\Phi)_{\rho^*} = \int_{\mathbb{R}^d \times \mathbb{R}^d} (D_{xy}W(x,y)\nabla\Phi(x), \nabla\Phi(y))\rho^*(x)\rho^*(y)dxdy$$

$$+ \beta \int_{\mathbb{R}^d} (\nabla \cdot (\rho^*\nabla\Phi))^2 \frac{1}{\rho^*(x)}dx.$$

3.8.2 Connections with Yano's formula

In this sequel, we show that (46) connects with an important equality in Riemannian geometry, named Yano's formula [97, 98].

$$\int_{\mathcal{M}} [\text{Ric}(X, X) + \sum_{i,j} \nabla_j X_i \nabla_i X_j]dx = \int_{\mathcal{M}} (\nabla \cdot X)^2 dx. \tag{52}$$

The formula is valid for any vector field X in a compact orientable finite dimensional Riemannian manifold $\mathcal{M}$, ∇_i is the operator of covariant differentiation for ∇_{∂_i}, Ric is a Ricci curvature tensor and dx is the volume element of the space.

In details, we consider a free energy containing only linear entropy.

$$\mathcal{F}(\rho) = \int_{\mathcal{M}} \rho(x)\log\rho(x)dx,$$

where the underlying state is a Riemannian manifold $\mathcal{M}$. The minimizer of this free energy, Gibbs measure ρ^*, is a uniform measure on $\mathcal{M}$. I.e.

$$\rho^*(x) = \frac{1}{\text{vol}(\mathcal{M})}, \quad \text{for any } x \in \mathcal{M}.$$

In this case, if we look at the Hessian operator of free energy on $\mathcal{P}_2(\mathcal{M})$ at the Gibbs measure, we observe an interesting equality

$$(\text{Hess}_{\mathcal{P}_2(\mathcal{M})}\mathcal{F} \cdot \nabla\Phi, \nabla\Phi)_{\rho^*}$$
$$= \int_{\mathcal{M}} [\text{Ric}(\nabla\Phi, \nabla\Phi) + \text{tr}(D^2\Phi D^2\Phi^T)]\rho^*(x)dx$$
$$= \int_{\mathcal{M}} [\nabla \cdot (\rho^*\nabla\Phi)]^2 \frac{1}{\rho^*(x)}dx.$$

Here the first equality is known by optimal transport [95], while the second equality is proved by Theorem 46. Since ρ^* is a uniform measure, then the above formula means nothing but

$$\int_{\mathcal{M}} [\text{Ric}(\nabla\Phi, \nabla\Phi) + \text{tr}(D^2\Phi D^2\Phi^T)]dx = \int_{\mathcal{M}} [\nabla \cdot (\nabla\Phi)]^2 dx. \tag{53}$$

Interestingly, if we denote $\nabla\Phi = X$, (53) reflects the famous Yano's formula (52).

In all, there is a subtle relationship between the Hessian operator at Gibbs measure and the geometry of underlying space.

CHAPTER IV

APPLICATION I: EVOLUTIONARY DYNAMICS

4.1   Introduction

In this chapter, we illustrate the first application of Fokker-Planck equations on graphs, which is the evolutionary game theory.

Games play fundamental roles in many real world problems [19, 73, 96], including Social Networks, Virus, Biology species, Trading, Cancer. Game theory study a situation that all players are selfish, who want to maximize their own payoffs. Currently, people model games in two ways. One is through some statical equilibria, e.g. Nash equilibria. The other is through a dynamical model. It means that all players are making decisions through a Markov process, whose transition law are governed by a dynamical system. Nowadays, the second approach is adopted vastly, which is known as evolutionary game theory [55, 76, 84, 86].

In the literature, people propose many different dynamics models, e.g. Replicator dynamics [1], Best response dynamics [69] and Smith dynamics [87]. However, one fundamental question is still unclear: In games with discrete strategies, how can we model uncertainties in player's decision process?

In continuous strategy games, the question is easy to answer. The white noise is widely used to model uncertainties. For example, players' decisions are characterize by SDEs, whose transition laws are governed by Fokker-Planck equations, see examples in Mean field games [3, 14, 15, 22, 39, 48, 61]. However, these theories can not be applied to the discrete strategy directly.

In this chapter, we bridge this gap by using Fokker-Planck equations on graphs, which provide evolutionary dynamics to model finite players' game, population game

and spatial population game.

4.2 Review in Game theory

In this sequel, we briefly review some concepts in game theory. Quantitively speaking, the game contains three components: players, strategy sets, and payoffs. It refers a situation that each player picks up a choice in strategy set. The player receives his own payoff depending on all others' choices. The goal of the game is to investigate how players make decisions under this setting.

In the literature, many different types of games are discussed. Depending on the number of players, the game can be either finite players' or infinite players' (population game); Depending on strategy sets, the game can be with either continuous or discrete strategy; Depending on payoffs, the game can be either statical or dynamical. In this chapter, we focus on the statical game, meaning that there is no time variable in the description of games.

4.2.1 Games

We begin with describing a **finite players'** game. It describes a situation where finite players try to find a strategy in their own strategy set with the "best" payoff. More precisely, player v picks a choice x_v in pure strategy set S_v, then he receives a payoff depending on all others, $F_v : S_1 \times \cdots \times S_N \to \mathbb{R}$. The game forms a multiple goal optimization problem [1]

$$\max_{x_v \in S_v} F_v(x_v, x_{-v}), \quad v \in \{1, \cdots, N\}.$$

Compare with the optimization, there is no unique objective function for all players. So other than looking at the "maximizer", people describe a special status for all players, which is named Nash equilibrium, meaning that no player is willing to change his current strategy unilaterally.

---

[1] We use a convention that, $x_{-v} = (x_1, \cdots, x_{v-1}, x_{v+1}, \cdots)$ for all players' choices other than player i's, $u_i(x_i, x_{-i}) := u_i(x_1, \cdots, x_N)$.

Definition 24 *A strategy profile x^* is a Nash equilibrium (NE) if*

$$F_v(x_v^*, x_{-v}^*) \geq F_v(x_v, x_{-v}^*) \quad \text{for any player } v \text{ with } x_v \in S_v.$$

In general, the strategy set can be either continuous set (Borel set) or discrete set. For example, consider a game with continuous strategy set.

Example 9 *Let $S_1 = S_2 = \mathbb{R}^1$. Player 1 or 2 wants to maximize his own payoff.*

$$\max_{x_1 \in \mathbb{R}^1} F_1(x_1, x_2) = -(x_1 - 1)^2 - x_2,$$

and

$$\max_{x_2 \in \mathbb{R}^1} F_2(x_1, x_2) = -x_2^2.$$

It is easy to check that $(1, 0)$ is a NE.

In this chapter, we mainly consider the discrete strategy set, in which games are called **normal games**. Because of strategy sets being discrete, the payoff functions naturally form matrices. We adopt a two players' game to illustrate. Suppose the strategy set is $S_1 = \{1, \cdots, m\}$, $S_2 = \{1, \cdots, n\}$. If player 1 chooses $i \in S_1$ and player 2 picks $j \in S_2$, they receive payoff values $F_1(i, j)$ and $F_2(i, j)$. It is customary to denote payoffs as a bi-matrix form (A, B^T), where matrix $A = (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}$ and $B^T = (b_{ji})_{1 \leq i \leq n, 1 \leq j \leq m}$, with $a_{ij} = F_1(i, j)$ and $b_{ji} = F_2(j, i)$.

For example, we consider the "Prisoner's dilemma".

Example 10 *Two members of a criminal gang are arrested and imprisoned. Each prisoner is given the opportunity to cooperate or defect. Their payoff matrixes are given by*

	player 2 C	player 2 D
player 1 C	(-1, -1)	(-3, 0)
player 1 D	(0, -3)	(-2, -2)

In this case, the strategy set is $S = \{C, D\}$, where C represents "Cooperation" and D represents "Defection". Here the game's payoff matrix is

$$A = \begin{pmatrix} -1 & -3 \\ 0 & -2 \end{pmatrix}, \quad B^T = \begin{pmatrix} -1 & 0 \\ -3 & -2 \end{pmatrix}.$$

By comparing matrix values, we know that (D, D) is a NE.

Secondly, we consider **population game**, which contains countably infinite many identical players [84].

To illustrate, we begin by considering a special finite (N) players' game, named **autonomous** game. It means that the player's payoff doesn't rely on the player's identity. More precisely, suppose all players are with the same strategy set S and each player's payoff function $F_v : S^N \to \mathbb{R}$ is specially symmetric:

$$F_v^N(x_1, \cdots, x_N) = F_{\sigma(v)}^N(x_{\sigma(1)}, \cdots, x_{\sigma(N)}),$$

for all permutations σ on $\{1, \cdots, N\}$. Suppose this game contains a large number of players, meaning that N is large enough. Under this setting, we assume that the payoff function $F^N : S \times S^{N-1} \to \mathbb{R}$ can be generalized to a map $F : S \times \mathcal{P}(S) \to \mathbb{R}$. Here $\mathcal{P}(S)$ is the probability set supported on S, meaning

$$F(y, \rho^N) := F_v^N(x_1, \cdots, x_{v-1}, y, x_{v+1}, \cdots, x_N), \quad \text{for any } (x_1, \cdots, x_N) \in S^N,$$

and $\rho^N = \frac{1}{N-1} \sum_{l \neq v} \delta_{x_l}$ is the empirical distribution for other $N-1$ players. Imagine that as N goes to infinity, the empirical measure approaches a limit, say $\rho^N \to \rho$ and the payoff forms $F(y, \rho^N) \to F(y, \rho)$. This limiting process defines a population game, where the strategy set is S, players forms the probability set $\mathcal{P}(S)$ with payoff function $F : S \times \mathcal{P}(S) \to \mathbb{R}$.

Similarly, one can define the Nash equilibrium in population game.

Definition 25 *ρ^* is a NE if*

$$\textit{Support of } \rho^* \subset \arg\max_{y \in S} F(y, \rho^*).$$

Again, NE tells the fact that no player can improve his payoff by switching strategies in population games.

In particular, we are interested in the discrete strategy set. Let the strategy set be $S = \{1, \cdots, n\}$, whose probability set is a simplex,

$$\mathcal{P}(S) = \{(\rho_i)_{i=1}^n \in \mathbb{R}^n \mid \sum_{i=1}^n \rho_i = 1, \ \rho_i \geq 0\}.$$

with payoff function $F(\rho) = (F_i(\rho))_{i=1}^n$, representing that a player in game choosing a strategy i receives the payoff $F_i(\rho)$. Again, we describe a NE

Definition 26 *Population state ρ^* is a Nash equilibrium if*

$$\rho_i^* > 0 \text{ implies that } F_i(\rho^*) \geq F_j(\rho^*), \quad \text{for all } j \in S.$$

Let's illustrate this game by an example.

Example 11 *Suppose infinite many players play Prisoner's Dilemmas. Each player is random matched to play the game; The player choosing strategy i receives the payoff by the expectation of all other players. I.e. $F_i(\rho) = \sum_{j \in S} a_{ij} \rho_j$.*

In Example 10, the population state is $\rho = (\rho_C, \rho_D)$. If a player in the game chooses strategy C, he will receive the payoff $F_C(\rho) = -3\rho_C - \rho_D$. Similarly, $F_D(\rho) = -\rho_D$. By the definition, $(0, 1)$ is a NE, meaning that all players choose the strategy D.

4.2.2 Potential games

Althrough the game is very different from optimization, we introduce a particular type of game, potential game, to bridge them [54, 72]. The potential game means that there exists an objective function, named potential, which is the maximization goal of all players.

Let's illustrate potential games by various types of games. We start with finite players' games. Let the strategy set be S, payoff functions be $F_v : S^N \to \mathbb{R}$, $v \in \{1, \cdots, N\}$.

If the strategy set is continuous, for example $S = \mathbb{R}^1$:

Definition 27 *A game is a potential game, if there exists a function $\phi(x) \in C^1(\mathbb{R}^N)$, such that*

$$\frac{\partial}{\partial x_v}\phi(x) = \frac{\partial}{\partial x_v}F_v(x), \quad \text{for any } v \in \{1, \cdots, N\}.$$

Example 12 *Let $S_1 = S_2 = \mathbb{R}^1$. Player 1 and 2 wants to maximize their own payoffs.*

$$\max_{x_1 \in \mathbb{R}} F_1(x_1, x_2) = -x_1^2,$$

and

$$\max_{x_2 \in \mathbb{R}} F_2(x_1, x_2) = -x_2^2.$$

It is a potential game with potential

$$\phi(x_1, x_2) = -(x_1^2 + x_2^2).$$

In mathematics, it is easy to check that maximizers of potentials are NEs.

If the strategy set is discrete, for example $S = \{1, \cdots, n\}$:

Definition 28 *A game is a potential game if there exists a function $\phi : S^N \to \mathbb{R}$ such that for any $x_v^1, x_v^2 \in S$, $x_{-v} \in S^{N-1}$,*

$$\phi(x_v^1, x_{-v}) - \phi(x_v^2, x_{-v}) = F_v(x_v^1, x_{-v}) - F_v(x_v^2, x_{-v}).$$

Example 13 *Prisoner dilemma in Example 10 is a potential game with*

$$\phi(x) = -\frac{F_1(x_1, x_2) + F_2(x_1, x_2)}{2}, \quad x = (x_1, x_2) \in \{(C,C), (C,D), (D,C), (D,D)\}.$$

Again, we can easily observe that maximizers of potentials $\phi(x)$ are NEs.

Secondly, we consider potential games in population games. Let the strategy set be S, payoff functions be $F : S \times \mathcal{P}(S) \to \mathbb{R}$. If the strategy is continuous, say $S = \mathbb{R}^1$:

Definition 29 *The population game is a potential game, if there exists a potential functional $\mathcal{F} : \mathcal{P}(S) \to \mathbb{R}$, such that* [2]

$$\frac{\delta}{\delta \rho(x)} \mathcal{F}(\rho) = F(x, \rho), \quad \text{for any } x \in S.$$

One can show directly that maximizers of potential functional $\mathcal{F}(\rho)$ are NEs.

If the strategy set is discrete, $S = \{1, \cdots, n\}$:

Definition 30 *The population game is a potential game, if there exists a differentiable potential function $\mathcal{F} : \mathcal{P}(S) \to \mathbb{R}$, such that*

$$\frac{\partial}{\partial \rho_i} \mathcal{F}(\rho) = F_i(\rho), \quad \text{for any } i \in S.$$

We illustrate this concept by an example.

Example 14 *If we consider a population game in Example 11 and the payoff matrix A is a symmetric matrix, then the game becomes a potential game with potential function $\mathcal{F}(\rho) = \frac{1}{2} \rho^T A \rho$, since*

$$\nabla \mathcal{F}(\rho) = A\rho = F(\rho).$$

Again, by first order conditions, one can show that maximizers of $\mathcal{F}(\rho)$ are NEs.

4.2.3 Fokker-Planck equations and Evolutionary dynamics

In above, we have discussed the concept of games and NEs, which are statical descriptions and special statuses of the game. In the real world, players are making decisions dynamically. In order to model such behaviors, people introduce the dynamics to investigate games. This is known as Evolutionary game theory. Meanwhile, Fokker-Planck equations, along with SDEs, are fundamental tools for modeling. In this sequel, we explain the connection between Fokker-Planck equation and evolutionary game theory through continuous strategy set.

---

[2] $\frac{\delta}{\delta \rho}$ *is the notation of first variational formula.*

At the beginning, we consider a finite players' game. To better illustrate, we assume that the game contains N players; Each player v chooses a strategy in $S_v = \mathbb{R}^d$ with a payoff function $F_v^N : S_1 \times \cdots \times S_N \to \mathbb{R}$. In other words, the game describes a situation:

$$\max_{x_v \in \mathbb{R}^1} F_v^N(x_1, \cdots, x_v, \cdots, x_N), \quad v \in \{1, \cdots, N\}.$$

To build a dynamical model, we make following assumptions.

- Players are making decisions dynamically;

- Each player is myopic and greedy. He chooses a direction that increases his current payoff most rapidly;

- There are some inevitable uncertainties when players are making decisions.

In mathematics, we consider a SDE system

$$dx_v = \nabla_{x_v} F_v^N(x)dt + \sqrt{2\beta}dW_t^v, \quad v \in \{1, \cdots, N\}, \tag{54}$$

where $\beta > 0$ and $W_t^1, \cdots, W_t^N$ are N independent Wiener processes. (54) reflects all quantitive assumptions: Its solution, a stochastic process $x(t) = (x_v(t))_{i=1}^N$ indicates that players are dynamically making decisions; The "most rapidly" direction implies the gradient direction of each player's payoff, $\nabla_{x_v} F_v^N$ and the white noise effects represent uncertainties.

Things become more interesting if we look at the probability. The Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \left(\rho (\nabla_{x_v} F_v^N(x))_{v=1}^d \right) = \beta \Delta \rho, \tag{55}$$

describes (54)'s probability transition equation. Here the unknown $\rho(t, x)$ is a probability density function. Under suitable conditions, as the time variable t goes to infinity, (55) converges to a stationary solution, named invariant measure, which tells us more information about Nash equilibria.

Secondly, the interplay between SDE and Fokker-Planck equation can also be applied to model population games. Recall that the population game is a special symmetric game with the number of players goes to infinity. As in last section, the empirical measure of all players ρ^N converges to one probability measure ρ, with

$$F(x, \rho^N) = F^N(x_1, \cdots, x_N) \to F(x, \rho).$$

By the assumption in "propagation of chaos", the limit process of SDE system

$$dx_v = \nabla_{x_v} F_v^N(x) dt + \sqrt{2\beta} dW_{vt}, \quad v \in \{1, \cdots, N\},$$

becomes a nonlinear SDE, meaning that its transition law[3] satisfies

$$d\bar{x} = \nabla_{\bar{x}} F(\bar{x}, \rho(t, \bar{x})) dt + \sqrt{2\beta} dW_t$$
$$\rho(t, \cdot) \sim \text{Law}(\bar{x}(t)).$$

In other words, the transition equation satisfies a "mean field" equation, which is a nonlinear Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \nabla_{\bar{x}} F(\bar{x}, \rho)) = \beta \Delta_{\bar{x}} \rho,$$

where the unknown $\rho(t, \bar{x})$ is probability density function supported on $\mathbb{R}^d$ and the "mean field" reflects that one uses a "mean" payoff as the approximation of larger number of players', i.e. $F(x, \rho) \approx F^N(x_1, \cdots, x_N)$.

In all, Fokker-Planck equation, along with the SDE system introduces many interesting properties of limiting behaviors in games.

4.2.4 Gradient flows on strategy sets

This derivation is more natural if we look at a special type of game, potential game. The best-reply dynamics is nothing but gradient flows.

---

[3]We call law as the transition probability

In finite player's game, the potential game means that there exists a function $\phi(x)$, such that $\nabla\phi(x) = (\nabla_{x_v}F_v(x))_{v=1}^d$. Hence (54) is nothing but a perturbed gradient flow in $\mathbb{R}^{Nd}$

$$dx = \nabla\phi(x)dt + \sqrt{2\beta}dW_t,$$

whose transition law obeys the Fokker-Planck equation

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\nabla\phi(x)) = \beta\Delta\rho.$$

In population game, there exists a functional $\mathcal{F}(\rho)$, such that $\frac{\delta}{\delta\rho(x)}\mathcal{F}(\rho) = F(x,\rho)$. Hence the SDE is also a perturbed gradient flow [24]

$$d\bar{x} = \nabla_{\bar{x}}F(\bar{x},\rho)dt + \sqrt{2\beta}dW_t$$

$$\rho(t,\cdot) \sim \mathrm{Law}(\bar{x}(t)).$$

Its transition density satisfies a nonlinear Fokker-Planck equation

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\nabla_{\bar{x}}F(\bar{x},\rho)) = \beta\Delta\rho.$$

Notice that, in above two cases, all players is making decisions according gradient flows in strategy sets. In the viewpoint of optimal transport, we can say more. All players' probability transition equation, Fokker-Planck equation, can also be viewed as gradient flows in probability set.

As we can see in above, there are strong connections between Fokker-Planck equations and evolutionary dynamics on continuous strategy set. Later on, we will build a similar connection on discrete strategy sets.

4.3   Finite players' games

In this sequel, we focus on finite (N) players' games on discrete strategy sets. We will establish Fokker-Planck equations on graphs as evolutionary dynamics.

Recall that the game is described as follows: Player v picks a choice x_v in a discrete strategy set

$$S_v = \{\mathbf{s}_1^v, \ldots, \mathbf{s}_{n_v}^v\},$$

with payoff matrix $F_v : S_1 \times \cdots \times S_N \to \mathbb{R}$. Notice that the game forms

$$\max_{x_v \in S_v} F_v(x_1, \cdots, x_v, \cdots, x_N), \quad v \in \{1, \cdots, N\}.$$

Our goal is to derive this game's evolutionary dynamics.

4.3.1 Gradient flows on strategy graphs

We start with considering potential games, meaning that there exists a potential function $\phi : S \to \mathbb{R}$, under which the game forms

$$\max_{(x_1, \cdots, x_N) \in S_1 \times \cdots \times S_N} \phi(x_1, \cdots, x_N).$$

A natural dynamics to connect this optimization is gradient flow. The meaning of gradient flow is based on three assumptions:

(i) All players don't obtain a "far" viewpoint. They don't know the "best" strategies immediately. As an alternative, all players are making decisions dynamically and simultaneously;

(ii) At the decision time, the player knows all other players' choices. The player chooses his "best" strategy in current "available strategy set".

(iii) All players are not purely "rational". There is always some "uncertainties" that affects players' decision procedures.

We explain these assumptions in details. (i) is the fundamental assumption. Since the strategy set is discrete, all players can't make their decisions "purely" as differential equations. As an alternative, all players are making decisions by continuous time stochastic processes, where the dynamics means the transition equation of probability measures.

(ii) introduces the concept of "Available strategy set", which requires a discrete set's topology (neighborhood information). To model that, we introduce a strategy

graph, meaning that the strategy set is settled on a finite graph. "Available strategy set" represents the adjacency set on the graph.

More precisely, assume that player v's strategy set S_v is on graph $G_v = (S_v, E(S_v))$, $v \in \{1, \cdots, N\}$. The joint strategy set is settled on a Cartesian product graph $G = (S, E(S)) = G_1 \square \cdots \square G_N$, where the vertex set is

$$S = S_1 \times \cdots \times S_N = \{x = (x_1, \cdots, x_N) \mid x_v \in S_v, \ v \in \{1, \cdots, N\}\},$$

and the edge set is $E(S) = E(S_1) \times \cdots \times E(S_N)$. Under this setting, we denote the "available strategy set" as

$$N(x) = \{y \in S \mid (x, y) \in E(S)\},$$

where the notation (x, y) is short for an edge on G connecting vertices x and y.

Example 15 *Let's consider a two players' Prisoner's Dilemma, where $S_1 = S_2 = \{C, D\}$. We connect the strategy set with graph:*
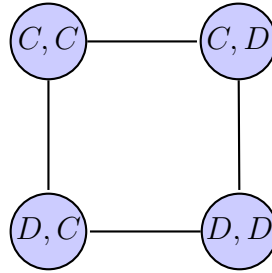
(iii) considers the "uncertainties" among players. To quantify that, we borrow a concept from kinetic mechanics in physics, named free energy [4]

$$\mathcal{F}(\rho) = -\sum_{x \in S} \phi(x)\rho(x) + \beta \sum_{x \in S} \rho(x) \log(x).$$

---

[4]Its full notation is

$$\mathcal{F}(\rho) = -\sum_{v=1}^{N} \sum_{x_v \in S_v} \phi(x_1, \cdots, x_N)\rho(x_1, \cdots, x_N) + \beta \sum_{v=1}^{N} \sum_{x_v \in S_v} \rho(x_1, \cdots, x_N) \log \rho(x_1, \cdots, x_N).$$

It is a summation of negative potential and Boltzmann-Shannon entropy from left to right. The notation of negative potential is just for mathematical convenient, which is to transfer "maximizing payoff" to "minimizing cost". The highlight here is the usage of Boltzmann-Shannon entropy (short as linear entropy), which is a quantity to model the total disorder of all players' decisions, with a positive constant β representing the strength of disorder.

In mathematics, we shall derive the gradient flow of free energy $\mathcal{F}(\rho)$ associated with the strategy graph G. By the optimal transport on graphs, we derive the new evolutionary dynamics:

Theorem 31 *Given a potential game with a strategy graph $G = (S, E(S))$, a potential $\phi(x)$ and a constant $\beta \geq 0$. Then the gradient flow of $\mathcal{F}(\rho)$,*

$$\mathcal{F}(\rho) = -\sum_{x \in S} \phi(x)\rho(x) + \beta \sum_{x \in S} \rho(x) \log \rho(x),$$

on the metric space $(\mathcal{P}_o(G), W_{2;\mathcal{F}})$ is

$$\begin{aligned}
\frac{d\rho(t,x)}{dt} &= \sum_{y \in N(x)} \rho(t,y)[\phi(x) - \phi(y) + \beta(\log \rho(t,y) - \log \rho(t,x))]_+ \\
&- \sum_{y \in N(x)} \rho(t,x)[\phi(y) - \phi(x) + \beta(\log \rho(t,x) - \log \rho(t,y))]_+ .
\end{aligned} \tag{56}$$

4.3.2 Markov process

As is well known, Fokker-Planck equation is a transition equation of Markov process. In this section, we shall connect a Markov process underlying (56).

In details, we introduce a continuous time stochastic process $X_\beta(t)$ on a finite state S. Its transition law, the transition probability from state x to state y, is as

follows:

$$\Pr(X_\beta(t+h) = y \mid X_\beta(t) = x)$$

$$= \begin{cases} (\bar{\phi}(x) - \bar{\phi}(y))_+ h + o(h) & \text{if } j \in N(i); \\ 1 - \sum_{j \in N(i)} (\bar{\phi}(x) - \bar{\phi}(y))_+ h + o(h) & \text{if } j = i; \\ 0 & \text{otherwise}, \end{cases}$$

where $\lim_{h \to 0} \frac{o(h)}{h} = 0$ and $\bar{\phi}(x) = -\phi(x, \rho) + \beta \log \rho(x)$.

Notice that $X_\beta(t)$ is a nonlinear Markov process, whose generating matrix $Q(\rho) = (Q_{ij}(\rho))_{1 \le i, j \le n}$ is defined as follows. If $i \ne j$

$$Q_{ij}(\rho) = \begin{cases} (\bar{\phi}(x) - \bar{\phi}(y))_+ & \text{if } (i, j) \in E, \\ 0 & \text{otherwise}, \end{cases}$$

and $Q_{ii} = -\sum_{j=1, j \ne i}^{n} Q_{ij}$. Let $\rho(t) = (\rho(t, x))_{i=1}^n$, $\rho(t, x) = \Pr(X_\beta(t) = x)$. Then the time evolution of $\rho(t)$ satisfies the Kolmogorov forward equation

$$\frac{d\rho}{dt} = \rho Q(\rho),$$

whose explicit formula is (56).

Interestingly, $X_\beta(t)$ gives a nice explanation of assumptions (i), (ii), (iii). Firstly, (i) is explained by the definition of continuous time Markov process. It means that in probability sense, all players are making decisions continuously on time. Secondly, (ii) is demonstrated by a "greedy" transition kernel Q. Whenever new strategies with better payoff are available (in strategy neighbor), the player will switch to them with probabilities proportional to the benefits (the difference of payoffs). Such behavior fills a "gradient" logic that the player improves his mean payoff "most rapidly", which is in the sense of

$$\frac{d}{dt}\mathbb{E} \text{ payoff} = -\mathbb{E} (\text{Benefit}^2),$$

meaning that

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = -\sum_{(x,y) \in E} (\bar{\phi}(x) - \bar{\phi}(y))_+^2 \rho(x).$$

87

Last and most interestingly, (iii) introduces a quantitative description of "uncertainties" in discrete states, which is through the Log-Laplacian. Heuristically, the uncertainties' logic is as follows: "The more precious the strategy is, the more players are willing to choose." In formulas,

"strategy x is precious" $\Rightarrow$ $\rho(x)$ is small $\Rightarrow$ "payoff" $\phi(x) - \log \rho(x)$ is large.

As a consequence, even if x's true payoff $\phi(x)$ is not better than others, players are still willing to switch their strategies towards x.

4.3.3   Fokker-Planck equations on strategy graphs

(56) guides evolutionary dynamics for general games. Notice that potential game means

$$\phi(x) - \phi(y) = F_v(x) - F_v(y), \quad \text{if } y \in N_v(x).$$

where $N_v(x)$ is the adjacent set of graph $G_v = (S_v, E(S_v))$. By this setting, (56) can be rewritten as

$$\frac{d\rho(t,x)}{dt} = \sum_{v=1}^{N} \sum_{y \in N_v(x)} [\bar{F}_v(y) - \bar{F}_v(x)]_+ \rho(t,y) - \sum_{v=1}^{N} \sum_{y \in N_v(x)} [\bar{F}_v(x) - \bar{F}_v(y)]_+ \rho(t,x), \quad (57)$$

where $\bar{F}_v(x) = -F_v(x) + \beta \log \rho(t,x)$. (57) is also a transition equation for Markov process $X_\beta(t)$,

$$\Pr(X_\beta(t+h) = y \mid X_\beta(t) = x)$$

$$= \begin{cases} (\bar{F}_v(x) - \bar{F}_v(y))_+ h & \text{if } y \in N_v(x); \\ 1 - \sum_{v=1}^{N} \sum_{y \in N_v(x)} (\bar{F}_v(x) - \bar{F}_v(y))_+ h + o(h) & \text{if } y = x; \\ 0 & \text{otherwise.} \end{cases}$$

We can check that (57) doesn't depend on the existence of potential. Because of this special relationship, we call (57) as the **Fokker-Planck equation on graphs**.

(57), along with Markov process $X_\beta(t)$ provides many interesting asymptotic behaviors of games. For example, the Fokker-Planck equation provides many vital

informations, including the "order" of NEs. To illustrate, we start with considering potential game. In such game, it is natural to use potential to give an "order" of Nash equilibria. In other words[5], if $x^1, \cdots, x^k \in S$ are distinct NEs, we define

$$x^1 \prec x^2 \cdots \prec x^k, \quad \text{if} \quad \phi(x^1) \le \cdots \le \phi(x^k). \tag{58}$$

The unique equilibrium of Fokker-Planck equation (57) always implies this order, since

$$\rho^*(x) = \frac{1}{K} e^{\frac{\phi(x)}{\beta}}, \quad x \in S.$$

It tells that the better (larger) is the potential, the larger is the probability in $\rho^*(x)$. So the above definition is equivalent to

$$x^1 \prec x^2 \cdots \prec x^k, \quad \text{if} \quad \rho^*(x^1) \le \cdots \le \rho^*(x^k). \tag{59}$$

Let's consider a general game, which doesn't have the potential. In this case, the order in (58) is not valid. However, the order in (59) still holds. In a word, we adopt the equilibrium of Fokker-Planck equation (57) to rank the "order" of NEs.

4.3.4 Examples

We explain several examples to demonstrate Fokker-Planck equations on strategy graphs.

Example 16 *Let's consider the Prisoner's Dilemma with the payoff matrix*

$$A = B = - \begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix}.$$

Here the strategy set is $S_1 \times S_2 = \{(C, C), (C, D), (D, C), (D, D)\}$. In this setting, the game is a potential game with potential

$$\phi(x) = -\frac{F_1(x_1, x_2) + F_2(x_1, x_2)}{2}, \quad \text{where } (x_1, x_2) \in S_1 \times S_2.$$

We connect this game with graph $K_2 \square K_2$

---

[5] $x \prec y$ is to say strategy y is better than strategy x.

89
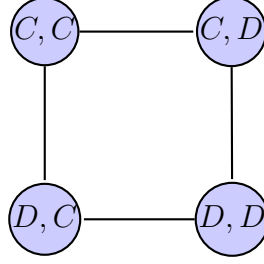
In this case, the transition measure function is

$$\rho(t) = (\rho_{CC}(t), \rho_{CD}(t), \rho_{DC}(t), \rho_{DD}(t))^T,$$

which satisfies Fokker-Planck equation (57)

$$
\begin{cases}
\dot{\rho}_{CC} = & [\bar{F}_1(D,C) - \bar{F}_1(C,C)]_+ \rho_{DC} + [\bar{F}_2(C,D) - \bar{F}_2(C,C)]_+ \rho_{CD} \\[2mm]
& -[\bar{F}_1(C,C) - \bar{F}_1(D,C)]_+ \rho_{CC} - [\bar{F}_2(C,C) - \bar{F}_2(C,D)]_+ \rho_{CC} \\[2mm]
\dot{\rho}_{CD} = & [\bar{F}_1(D,D) - \bar{F}_1(C,D)]_+ \rho_{DD} + [\bar{F}_2(C,C) - \bar{F}_2(C,D)]_+ \rho_{CC} \\[2mm]
& -[\bar{F}_1(C,D) - \bar{F}_1(D,D)]_+ \rho_{CD} - [\bar{F}_2(C,D) - \bar{F}_2(C,C)]_+ \rho_{CD} \\[2mm]
\dot{\rho}_{DC} = & [\bar{F}_1(C,C) - \bar{F}_1(D,C)]_+ \rho_{CC} + [\bar{F}_2(D,D) - \bar{F}_2(D,C)]_+ \rho_{DD} \\[2mm]
& -[\bar{F}_1(D,C) - \bar{F}_1(C,C)]_+ \rho_{DC} - [\bar{F}_2(D,C) - \bar{F}_2(D,D)]_+ \rho_{DC} \\[2mm]
\dot{\rho}_{DD} = & [\bar{F}_1(C,D) - \bar{F}_1(D,D)]_+ \rho_{CD} + [\bar{F}_2(D,C) - \bar{F}_2(D,D)]_+ \rho_{DC} \\[2mm]
& -[\bar{F}_1(D,D) - \bar{F}_1(C,D)]_+ \rho_{DD} - [\bar{F}_2(D,D) - \bar{F}_2(D,C)]_+ \rho_{DD}
\end{cases}
$$

where $\bar{F}_v(x_1, x_2) = -F_v(x_1, x_2) + \beta \rho_{x_1 x_2}$, $v = 1, 2$. By numerically solving (57) for

$$\rho^* = \lim_{\beta \to 0} \lim_{t \to \infty} \rho(t),$$

we obtain a unique measure ρ^* for any initial condition $\rho(0)$, see Figure 6.
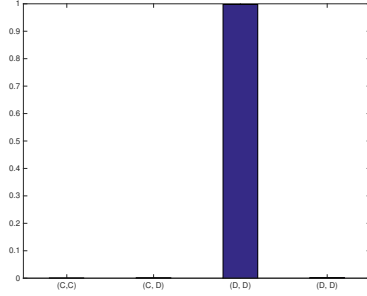
Figure 6: Two player's game: Prisoner's Dilemma

In this case, $\rho^* \approx (0, 1)$, which implies that two players will choose (D, D) eventually.

Example 17 *Let's consider a non autonomous game, meaning that players' payoff depends on his identity, i.e. $A \neq B$. For example, let $A = -\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ and $B = -\begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}$. We connect the game with graph $K_2 \square K_2$.*
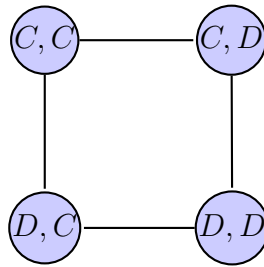
Again, by numerically solving (57) *for*

$$\rho^* = \lim_{\beta \to 0} \lim_{t \to \infty} \rho(t),$$

we obtain a unique measure ρ^* for any initial measure $\rho(0)$, which is demonstrated in Figure 7.
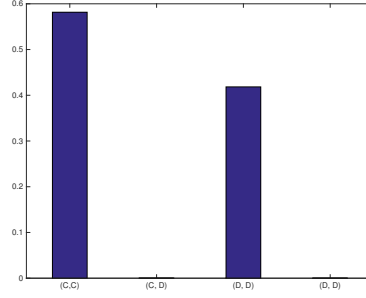
Figure 7: Two player's game: Asymmetric game

In this case, ρ^* only supports at (C, C) and (D, D), which are Nash equilibria of the game. Moreover, ρ^*_{CC} is larger than ρ^*_{DD}, which implies that (C, C) is more "stable" than (D, D). This result reflect the intuition of the game. Look at the situation player 1, 2 is at strategy (C, D). player 2 is more willing to change than player 1. Because if doing so, player 2 gains more benefits than player 1, i.e. $F_2(C, D) - F_2(C, C) = 2 > 1 = F_1(C, D) - F_1(D, D)$.

Example 18 *Let's consider the Rock-Scissors-Paper. Each player plays against others with strategies: Rock, Scissor and paper, which is short as r, s, p. Depending on win or lose, he receives a payoff 1 or -1. In other words, the game is with strategy sets $S_1 = S_2 = \{r, s, p\}$ and payoff matrixes*

$$A = B = - \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$$

We connect the game with the strategy graph $K_3 \square K_3$.

Again, by numerically solving (57),

$$\rho^* = \lim_{\beta \to 0} \lim_{t \to \infty} \rho(t).$$

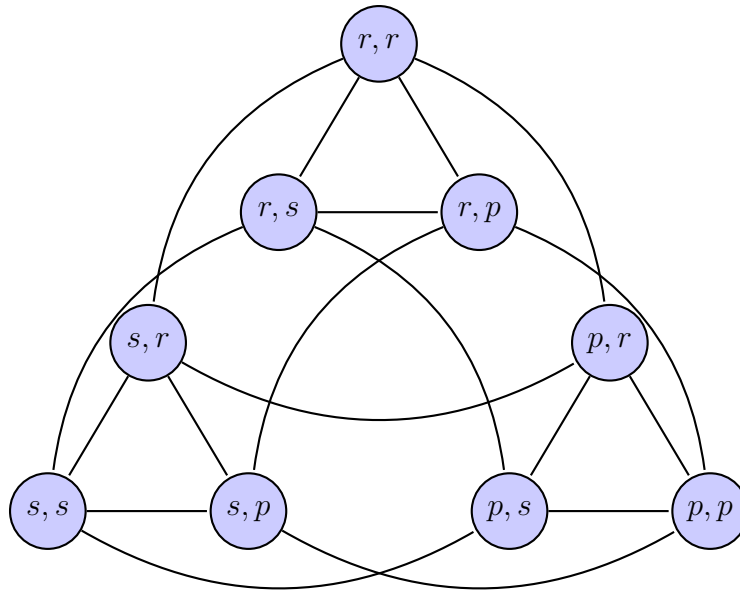*We obtain a unique measure ρ^* for any initial measure $\rho(0)$, which is demonstrated in Figure 8.*
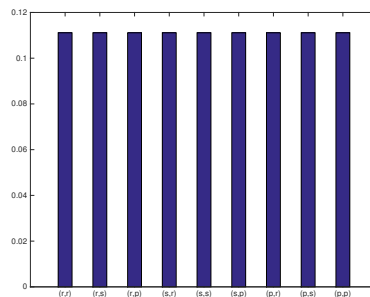
Figure 8: Two player's game: Rock-Paper-Scissors

In this case, ρ^* is a uniform mass function, which implies that two players will eventually choose their strategy uniformly.

4.4 Population games

In this section, we focus on population games with discrete strategy sets. We develop Fokker-Planck equations on graphs as new evolutionary dynamics. The game is described as follows: The strategy set is

$$S = \{1, \cdots, n\}.$$

The infinite players (population state) form a probability manifold

$$\mathcal{P}(S) = \{(\rho_i)_{i=1}^n \mid \rho_i \geq 0, \ \sum_{i=1}^n \rho_i = 1, \ i \in S\}$$

with payoff vector function $F(\rho) = (F_i(\rho))_{i=1}^n$. To better illustrate, we consider norm game as in example 11. It means that $F(\rho) = A\rho$, where $A = (a_{ij})_{1\leq i,j\leq n}$ is the interaction matrix.

4.4.1 Gradient flows on strategy graphs

We start with considering a potential game. The potential is $\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij}\rho_i\rho_j$, where A is a symmetric matrix.

A natural dynamics to connect this optimization is gradient flow. Similarly in finite player games, the gradient flow is based on following assumptions.

(i) All players don't obtain a "far" viewpoint. They don't know the "best" strategies immediately. As an alternative, all players are making decisions dynamically and simultaneously;

(ii) At the decision time, the player knows all other players' choices. The player chooses his "best" strategy in current "available strategy set".

(iii) All players are not purely "rational". There is always some "uncertainties" that affects players' decision procedures.

94

In details, (i) and (ii) are similarly to the finite games' case. More precisely, the strategy graph is a finite graph $G = (S, E)$, where S, E is the graph vertex and edge set. The "available strategy set" is

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

where the notation (i, j) is short for an edge on G connecting vertices i and j.

(iii) introduces "uncertainties" among the population. Similarly as finite players' games, we borrow the concept of free energy in kinetic mechanics

$$\mathcal{F}(\rho) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \rho_i \rho_j + \beta \sum_{i=1}^{n} \rho_i \log \rho_i.$$

It is a summation of negative potential and Boltzmann-Shannon entropy from left to right. Again, the highlight here is the usage of linear entropy, which is a quantity to model the total disorder of population, with a positive constant β representing the strength of disorder.

In all, we shall derive an evolutionary dynamics, which is the gradient flow of free energy $\mathcal{F}(\rho)$ associated with the strategy graph G.

Theorem 32 *Given a potential game with a strategy graph $G = (S, E(S))$ and a constant $\beta \geq 0$. Then the gradient flow of free energy $\mathcal{F}(\rho)$,*

$$\mathcal{F}(\rho) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \rho_i \rho_j + \beta \sum_{i=1}^{n} \rho_i \log \rho_i,$$

on the metric space $(\mathcal{P}_o(G), W_{2;\mathcal{F}})$ is

$$
\begin{aligned}
\frac{d\rho_i}{dt} = &\sum_{j \in N(i)} \rho_j \Big(\sum_{i=1}^{n} a_{ij} \rho_i - \sum_{j=1}^{n} a_{ij} \rho_j + \beta \log \rho_j - \beta \log \rho_i \Big)_+ \\
&- \sum_{j \in N(i)} \rho_i \Big(\sum_{j=1}^{n} a_{ij} \rho_j - \sum_{i=1}^{n} a_{ij} \rho_i + \beta \log \rho_i - \beta \log \rho_j \Big)_+.
\end{aligned}
\tag{60}
$$

4.4.2 Markov process

In this section, we shall connect a Markov process underlying (60), through which we explain (60)'s meaning in modeling level.

More precisely, we introduce a continuous time stochastic process $X_\beta(t)$ on a finite state S. Its transition law, the transition probability from state i to state j, is as follows:

$$\Pr(X_\beta(t + h) = j \mid X_\beta(t) = i)$$

$$= \begin{cases} (\bar{F}_i(\rho) - \bar{F}_j(\rho))_+ h & \text{if } j \in N(i); \\ 1 - \sum_{j \in N(i)} (\bar{F}_i(\rho) - \bar{F}_j(\rho))_+ h + o(h) & \text{if } j = i; \\ 0, & \text{otherwise,} \end{cases} \tag{61}$$

where $\lim_{h \to 0} \frac{o(h)}{h} = 0$ and $\bar{F}_i(\rho) = -F_i(\rho) + \beta \log \rho_i$.

Notice that $X_\beta(t)$ is a nonlinear Markov process, whose generating matrix $Q(\rho) = (Q_{ij}(\rho))_{1 \le i, j \le n}$ is

$$Q_{ij} := \begin{cases} (\bar{F}_i(\rho) - \bar{F}_j(\rho))_+ & \text{if } (i, j) \in E, \\ 0 & \text{if } (i, j) \notin E, \ i \ne j. \end{cases}, \quad \text{and} \quad Q_{ii} := -\sum_{j=1, j \ne i}^{n} Q_{ij}.$$

Let $\rho(t) = (\rho_i(t))_{i=1}^n$, $\rho_i = \Pr(X_\beta(t) = i)$. Then the time evolution of $\rho(t)$ satisfies the Kolmogorov forward equation

$$\frac{d\rho}{dt} = \rho Q(\rho).$$

whose explicit formula is exactly (60),

$$\frac{d\rho_i}{dt} = \sum_{j \in N(i)} \rho_j(\bar{F}_j(\rho) - \bar{F}_i(\rho))_+ - \sum_{j \in N(i)} \rho_i(\bar{F}_i(\rho) - \bar{F}_j(\rho))_+.$$

Let's focus on the modeling explanations of $X_\beta(t)$, whose transition law gives a detailed description of assumptions (i), (ii), (iii). Notice that this understanding can be viewed as the limiting behavior of finite players' games. Firstly, (i) is explained by the definition of continuous time Markov process. Secondly, (ii) is showed by a "greedy" decision rule. Whenever new strategies with better payoff are available (in strategy neighbor), the player will switch to them with probabilities proportional to

96

the benefits (the difference of payoffs). Such behavior fills a "gradient" logic: All players are to improve his mean payoff "most rapidly", in the sense of

$$\frac{d}{dt}\mathbb{E} \text{ payoff} = -\mathbb{E} \text{ (Benefit}^2),$$

meaning that

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = -\sum_{(i,j)\in E} (\bar{F}_i(\rho) - \bar{F}_j(\rho))_+^2 \rho_i.$$

Last and most interestingly, (iii) introduces a quantitative description of "uncertainties" in discrete states, which is through the Log-Laplacian term. Heuristically, the uncertainties' logic is as follows: "The more precious the strategy is, the more players are willing to choose." In formula,

strategy i is precious $\Rightarrow$ ρ_i is small $\Rightarrow$ "payoff" $F_i(\rho) - \log \rho_i$ is large.

As a consequence, even if i's true payoff $F_i(\rho)$ is not better than others, the player is still willing to switch their strategies towards i.

4.4.3 Fokker-Planck equations on strategy graphs

In this sequel, we consider Fokker-Planck equations on graphs for general population games.

Notice that (60) is always a well defined flow in $\mathcal{P}(S)$; $X_\beta(t)$ is always a Markov process underlying (60). They don't depend on the existence of potential. So we apply them as the dynamics and Markov process to model general games:

We name (60) **as a Fokker-Planck equation on a strategy graph**

$$\frac{d\rho_i}{dt} = \sum_{j\in N(i)} (\bar{F}_j(\rho) - \bar{F}_i(\rho))_+ \rho_j - \sum_{j\in N(i)} (\bar{F}_i(\rho) - \bar{F}_j(\rho))_+ \rho_i.$$

(60), along with $X_\beta(t)$, provides many interesting asymptotic behaviors of games from the variation of parameter β on Log-laplacian. If the game is a potential game, (60) is a well-known gradient system, whose equilibria are Gibbs measures. Moreover,

if the game doesn't have potential, there may exhibit a more complicit phenomenal other than Gibbs measures. For example, there is a situation with Hopf bifurcation in Example 21.

In addition, the concept of "order" among NEs can be introduced in population games. But it is different from finite players' games. There are two things to be noticed. One is that there may not exist the unique equilibrium (invariant measure) of (60). For example, in potential games, there is a case with multiple Gibbs measures. The other is that the probability measure itself is a NE. To conquer these conceptual differences, we need to consider a flow in the "probability" of probability sets, $\mathcal{P}(\mathcal{P}(S))$ to discuss the "order" of measures. It is certainly beyond the scope of this thesis, which will be studied in the future work.
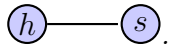
4.4.4 Examples

In this section, we demonstrate (60) on several population games.

Example 19 *("Irrationality") We start with the Stag hunt. It is a normal game with payoff matrix*

$$A = \begin{pmatrix} h & h \\ 0 & s \end{pmatrix}$$

The game is described as follows. Each player faces a choice, hunting for a hare (h) or a stag (s). The stag is worth more that hare, e.g. $s = 3$, $h = 2$. The game is with strategy set $\{h, s\}$, population state $\rho = (\rho_h, \rho_s)$ and payoff functions $F_h(\rho) = 2$, $F_s(\rho) = 3\rho_s$. It implies three Nash equilibria : $(0, 1)$, $(1, 0)$, and $(\frac{1}{3}, \frac{2}{3})$.

We apply the evolutionary dynamics (60) with the strategy graph: (h)——(s)*.*

$$\begin{cases} \dot{\rho}_h = \rho_s[2 - 3\rho_s + \beta \log \rho_s - \beta \log \rho_h]_+ - \rho_h[-2 + 3\rho_s + \beta \log \rho_h - \beta \log \rho_s]_+ \\ \dot{\rho}_s = \rho_h[3\rho_s - 2 + \beta \log \rho_h - \beta \log \rho_s]_+ - \rho_s[-3\rho_s + 2 + \beta \log \rho_s - \beta \log \rho_h]_+. \end{cases}$$

We explain (60)'s asymptotic property through its vector field on the probability manifold $\mathcal{P}(S)$ (line segment).
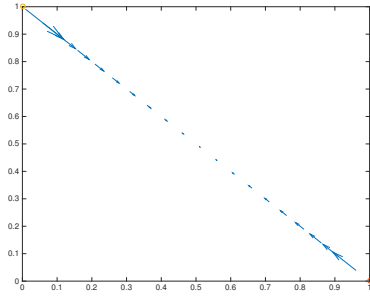
(a) $\beta = 5$
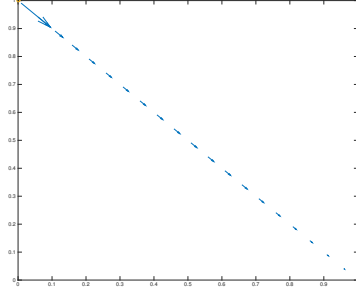
(b) $\beta = 0.5$
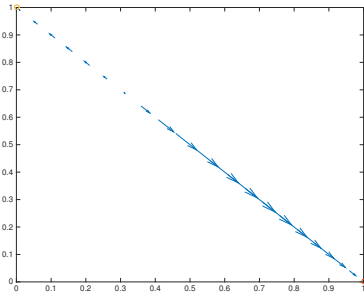
(c) $\beta = 0.1$
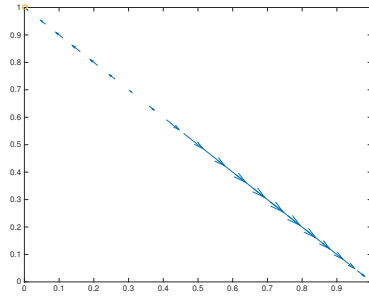
(d) $\beta = 0$

Figure 9: Population game: Stag hunt.

We show some interesting behaviors of the game by varying the parameter β in (60). If β is too large, as in Figure (A), the asymptoticly population state is $(\frac{1}{2}, \frac{1}{2})$. It means that all players are totally irrational, they flip a fair coin to decide what to hunt; If β is certainly large, as in Figure (B), all players will choose to hunt a hare (NE $(1, 0)$). It means that all players are partially rational. They know that the hare is always a safe choice, in the sense that they will get a hare not matter how the others choose; If β is small, as in Figure (C) and (D), all players choose a stag $(0, 1)$ or a hare $(1, 0)$, depending on initial state. It means that all players are rational enough, such that each player will make decisions according to the others.

Example 20 *We consider the Rock-Scissors-Paper played by the population. Its pay-off matrix is*

$$A = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}.$$

The game is with strategy set $S = \{r, s, p\}$, population state $\rho = (\rho_r, \rho_s, \rho_p)$ and payoff functions $F_r(\rho) = -\rho_s + \rho_p$, $F_s(\rho) = \rho_s - \rho_p$, $F_p(\rho) = -\rho_r + \rho_s$. It is with the unique Nash equilibrium $\rho^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.
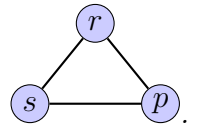
Again, let's look at the evolutionary dynamics (60) with the strategy graph:

We demonstrate (60)'s vector field on the probability manifold (triangular) in Figure 10.

(a) $\beta = 0$　　　　　　　　　(b) $\beta = 0.1$
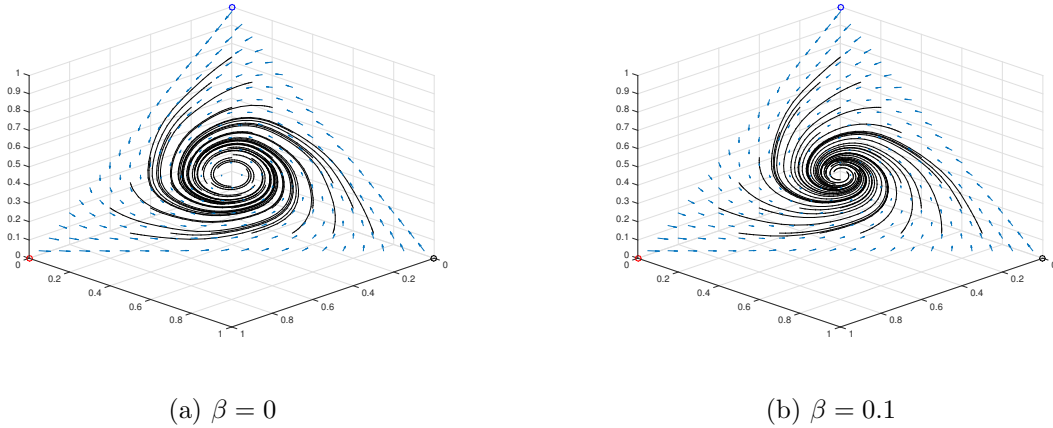
Figure 10: Population game: Rock-Paper-Scissors

In this case, the asymptoticly behavior of (60) is around $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. There is no much difference by varying parameter β.

Example 21 (Hopf) *We consider the Bad Rock-Paper-Scissors, whose payoff matrix is*

$$A = \begin{pmatrix} 0 & -2 & 1 \\ 1 & 0 & -2 \\ -2 & 1 & 0 \end{pmatrix}$$

which is slight different from the Rock-Paper-Scissors. In this case, the game is with strategy set $S = \{r, s, p\}$, population state $\rho = (\rho_r, \rho_s, \rho_p)$, and payoff functions $F_r(\rho) = -2\rho_s + \rho_p$, $F_s(\rho) = \rho_s - 2\rho_p$, $F_p(\rho) = -2\rho_r + \rho_s$. By the same setting of Example 20, we demonstrate (60)'s vector field.
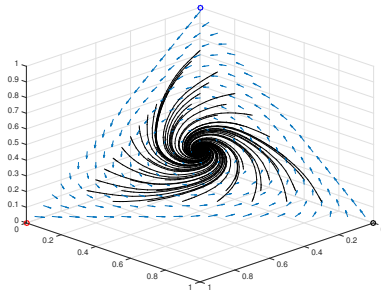
(a) $\beta = 0.5$
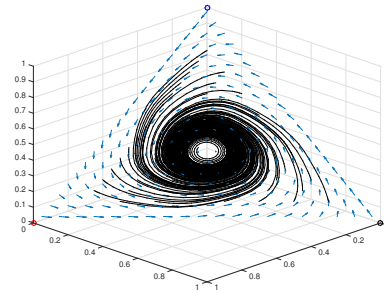
(b) $\beta = 0.1$
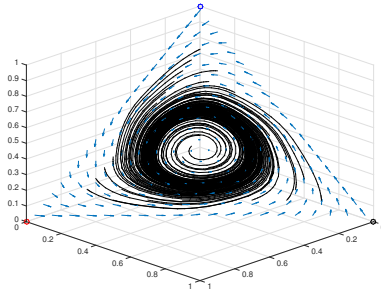
(c) $\beta = 0$

Figure 11: Population game: Bad Rock-Paper-Scissors

Observe that there is a Hopf bifurcation of (60) for parameter β, see Figure 11. If β is large, there is a unique equilibrium of (60) around $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$; If β goes to 0,

(60)'s solution approaches to a stable limit cycle.

Example 22 (Multiple Gibbs measures) *We consider a potential game with the payoff matrix*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Here the game is with strategy set $S = \{1, 2, 3\}$, population state $\rho = (\rho_1, \rho_2, \rho_3)$ and payoff functions $F_1(\rho) = \rho_1$, $F_2(\rho) = \rho_2 + \rho_3$ and $F_3(\rho) = \rho_2 + \rho_3$. It contains three sets of Nash equilibria :

$$\{\rho \mid \rho_1 = \frac{1}{2}\} \cup \{(1, 0, 0)\} \cup \{\rho \mid \rho_1 = 0\},$$

where the first and third one are lines on $\mathcal{P}(S)$. By applying (60) with a complete graph, we obtain two Gibbs measures near

$$\{(0, \frac{1}{2}, \frac{1}{2})\} \cup \{(1, 0, 0)\}.$$

See (60)'s vector field in Figure 12.
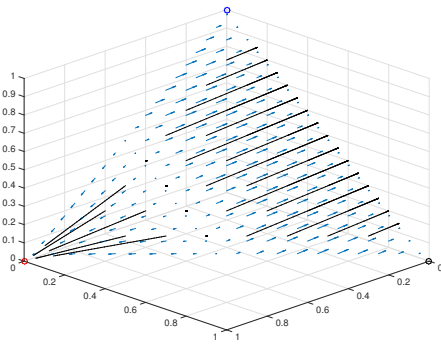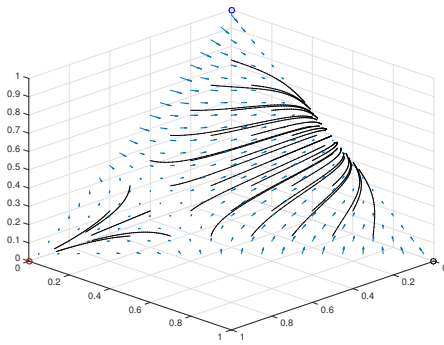
(a) $\beta = 0$ (b) $\beta = 0.1$

Figure 12: Population game: Multiple Gibbs measures

Example 23 (Unique Gibbs measure) *Let's consider the other potential game*

with the payoff matrix

$$A = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

The game is with strategy set $S = \{1, 2, 3\}$, population state $\rho = (\rho_1, \rho_2, \rho_3)$ and payoff functions $F_1(\rho) = \frac{1}{2}\rho_1$, $F_2(\rho) = \rho_2 + \rho_3$, $F_3(\rho) = \rho_2 + \rho_3$. It is with three sets of Nash equilibria

$$\{\rho \mid 1 - \frac{1}{2}\rho_1 = \rho_2 + \rho_3\} \cup \{(1, 0, 0)\} \cup \{\rho \mid 1 = \rho_2 + \rho_3\}.$$

By applying (60) *on a complete graph, there is the unique Gibbs measures near*

$$(0, \frac{1}{2}, \frac{1}{2}).$$

See (60)*'s vector fields in Figure 13.*

(a) $\beta = 0$ (b) $\beta = 0.1$
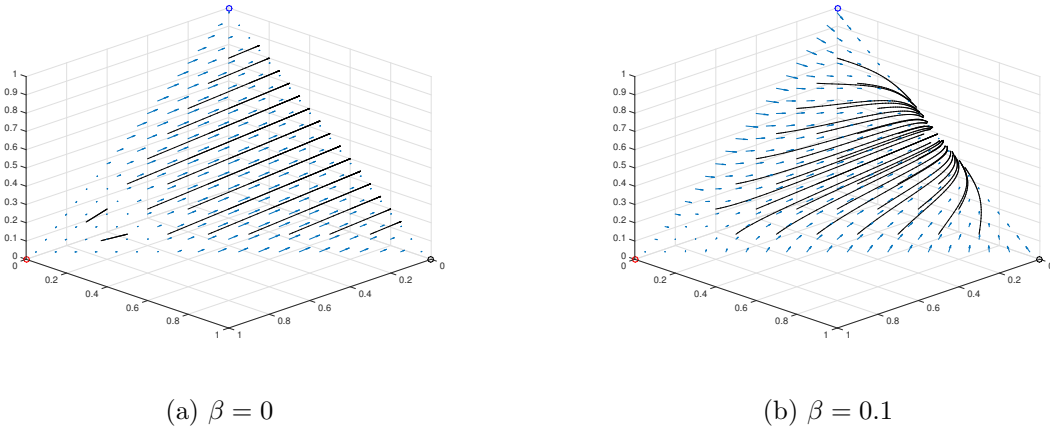
Figure 13: Population game: Unique Gibbs measure

4.5 Spatial population games

Spatial population games consider population games with spatial structures, which are widely used in population models, including crimes, disease spreading and biology etc. In this sequel, we build Fokker-Planck equations on spatial-strategy graphs to model this game.

The game is described as follows: infinite identical players are settled in vertices of a spatial graph. Each vertex of such spatial graph represents a place, where many players stay. The individual player plays games with his spatial neighbors, and receives a payoff vector depending on all. In this game, individual player tries to move his position and change his strategy, so as to improve his own payoff vectors.

In order to characterize the game quantitively, we discuss the strategy set, players and payoff in details.

For strategy set, we consider a strategy-spatial graph. Consider a population game with strategy graph $G_S = (S, E(S))$, where the vertex set is $S = \{\mathbf{s}_1, \cdots, \mathbf{s}_n\}$ and the edge set is $E(S)$. Suppose the population is settled on a spatial graph $G_L = (L, E(L))$, where the vertex set is $L = \{\mathbf{l}_1, \cdots, \mathbf{l}_m\}$ and the edge set is $E(L)$. The spatial-strategy graph is a graph $G = G_L \Box G_S = (V, E)$, where $\Box$ means the cartesian product of graphs and

$$V = L \times S, \quad E = E(S) \times E(L).$$

Example 24 *Let's consider a "Rock(r)-Scissors(s)-Paper(p)" game played by population in a spatial graph. Let the strategy space S be a complete graph k_3, and the spatial space L be a 2×2 lattice graph with $S = \{r, s, p\}$ and $L = \{\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3, \mathbf{l}_4\}$. We connect the game with the spatial-strategy graph:*
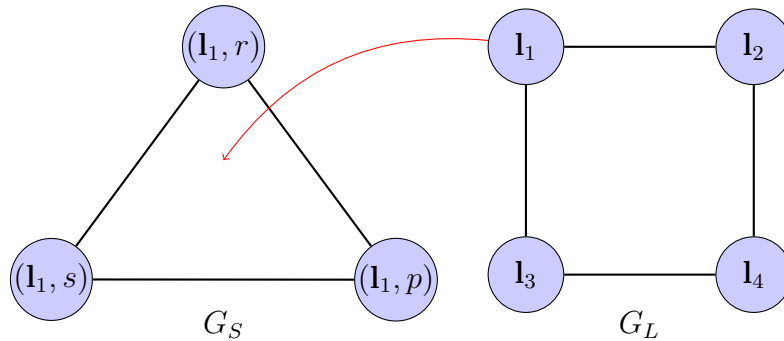
For players, we consider the population forming probability set supported on both

spatial and strategy set:

$$\mathcal{P}(G) = \{(\rho_{ij})_{1 \leq i \leq n, \ 1 \leq j \leq m} \mid \sum_{(l_i, s_j) \in L \times S} \rho_{ij} = 1, \quad \rho_{ij} \geq 0\},$$

where ρ_{ij} represents the proportion of people choosing spatial l_i and strategy s_j. Each player at position $\mathbf{l}_i \in L$ choosing strategy $\mathbf{s}_j \in S$ receives a vector function $(E_{ij}(\rho), F_{ij}(\rho))$, which are associated with spatial and strategy graphs.

For payoff functions, we consider a special case: suppose a normal game is played on a spatial graph; the payoff functions for spatial and strategy are same, which is according to the average of all the players' spatial neighbors[6], i.e. the individual player at position $\mathbf{l}_i$, choosing strategy $\mathbf{s}_j$ receives payoff

$$E_{ij}(\rho) = F_{ij}(\rho) = \sum_{(k,s) \in N(i,j)} a_{sj} \rho_{ks} + \sum_{s \in N_S(j)} a_{sj} \rho_{is} + a_{jj} \rho_{ij}.$$

Here $A = (a_{ls})_{(\mathbf{s}_l, \mathbf{s}_s) \in S \times S}$ is a payoff matrix and $N_L(i)$, $N_S(j)$, $N(i,j)$ represents adjunct set of $\mathbf{l}_i$, $\mathbf{s}_j$ on graphs G_L, G_S, G, meaning

$$N_L(i) = \{l_k \mid (l_i, l_k) \in E(L)\}, \quad N_S(j) = \{s_l \mid (s_j, s_l) \in E(S)\}, \quad N(i,j) = N_L(i) \times N_S(j).$$

In addition, we introduce a special type, potential game. If A is a symmetric matrix, then the game is a potential game with potential $\frac{1}{2} \sum_{(l_i, \mathbf{s}_j) \in L \times S} F_{ij}(\rho) \rho_{ij}$. In other words, the game describes a maximization problem

$$\max_{\rho \in \mathcal{P}(G)} \frac{1}{2} \sum_{(\mathbf{l}_i, \mathbf{s}_j) \in L \times S} F_{ij}(\rho) \rho_{ij}.$$

4.5.1 Gradient flows on spatial-strategy graphs

In this sequel, we shall derive a new evolutionary dynamics, which is the gradient flow of free energy associated with the spatial-strategy graph $G = G_L \square G_S$.

---

[6]Here we consider the self-interaction case. It means that the play also plays with others who lives in the same spatial node.

Theorem 33 *Given a potential game with a spatial-strategy graph $G = G_L \square G_S$ and a constant $\beta \geq 0$. Then the gradient flow of free energy*

$$\mathcal{F}(\rho) = -\frac{1}{2}\sum_{\mathbf{l}_i \in L}\sum_{\mathbf{s}_j \in S} F_{ij}(\rho)\rho_{ij} + \beta \sum_{(\mathbf{l}_i, \mathbf{s}_j) \in L \times S} \rho_{ij}\log\rho_{ij}$$

on metric space $(\mathcal{P}_o(G), W_{2;\mathcal{F}})$ is

$$
\begin{aligned}
\frac{d\rho_{ij}}{dt} &= \sum_{(k,s)\in N(i,j)} \rho_{ks}(F_{ij}(\rho) - F_{ks}(\rho) + \beta\log\rho_{ks} - \beta\log\rho_{ij})_+ \\
&\quad - \sum_{(k,s)\in N(i,j)} \rho_{ij}(F_{ks}(\rho) - F_{ij}(\rho) + \beta\log\rho_{ij} - \beta\log\rho_{ks})_+.
\end{aligned}
\tag{62}
$$

4.5.2 Markov process

In this section, we build a joint Markov process underlying (62).

Let's introduce a joint Markov process $(L_\beta(t), X_\beta(t))$ on a finite state $L \times S$, whose transition law, the transition probability from state $(\mathbf{l}_i, \mathbf{s}_j)$ to state $(\mathbf{l}_k, \mathbf{s}_s)$, is as follows:

$$P(L_\beta(t+h) = \mathbf{l}_k, X_\beta(t+h) = \mathbf{s}_s \mid L_\beta(t) = \mathbf{l}_i, X_\beta(t) = \mathbf{s}_j)$$

$$
= \begin{cases}
(\bar{F}_{ij}(\rho) - \bar{F}_{ks}(\rho))_+ h & \text{if } (k,s) \in N(i,j); \\
1 - \sum_{(k,s)\in N(i,j)}(\bar{F}_{ij}(\rho) - \bar{F}_{ks}(\rho))_+ h + o(h) & \text{if } (k,s) = (i,j); \\
0 & \text{otherwise},
\end{cases}
$$

where $F_{ij}(\rho) = -F_{ij}(\rho) + \beta\rho_{ij}$, for any $\mathbf{l}_i \in L$, $\mathbf{s}_j \in S$. Simiarly, $(L_\beta(t), X_\beta(t))$'s transition function is given by the Kolmogorov forward equation

$$\frac{d\rho_{ij}}{dt} = \sum_{(k,s)\in N(i,j)} \rho_{ks}(\bar{F}_{ks}(\rho) - \bar{F}_{ij}(\rho))_+ - \sum_{(k,s)\in N(i,j)} \rho_{ij}(\bar{F}_{ij}(\rho) - \bar{F}_{ks}(\rho))_+,$$

which is same as (62).

4.5.3 Fokker-Planck equations on spatial-strategy graphs

In this sequel, we derive an evolutionary dynamics on general spatial games. In other words, if a spatial game is not a potential game, the evolutionary dynamics is just a flow in $\mathcal{P}(G)$, not a gradient flow.

106

Replacing the payoff vector (F_{ij}, F_{ij}) to a general form (E_{ij}, F_{ij}) in (62), we obtain

$$\frac{d\rho_{ij}}{dt} = \sum_{k \in N_L(i)} (\bar{E}_{kj}(\rho) - \bar{E}_{ij}(\rho))_+ \rho_{kj} - \sum_{k \in N_L(i)} (\bar{E}_{ij}(\rho) - \bar{E}_{kj}(\rho))_+ \rho_{ij}$$
$$+ \sum_{s \in N_S(j)} (\bar{F}_{is}(\rho) - \bar{F}_{ij}(\rho))_+ \rho_{is} - \sum_{s \in N_S(j)} (\bar{F}_{ij}(\rho) - \bar{F}_{is}(\rho))_+ \rho_{ij} , \tag{63}$$

where $\bar{E}_{ij} = -E_{ij} + \beta \log \rho_{ij}$, $\bar{F}_{ij} = -F_{ij} + \beta \log \rho_{ij}$ for any $(\mathbf{l}_i, \mathbf{s}_j) \in L \times S$. We notice that (63) is an extension of (62), which doesn't depend on potentials. We call (63) as the **Fokker-Planck equation on spatial-strategy graph**. It connects a joint Markov process $(L_\beta(t), X_\beta(t))$, whose transition law is given by

$$P(L_\beta(t+h) = k, X_\beta(t+h) = s \mid L_\beta(t) = i, X_\beta(t) = j)$$

$$= \begin{cases} (\bar{E}_{ij}(\rho) - \bar{E}_{kj}(\rho))_+ h & \text{if } k \in N_L(i), \ s = j; \\ (\bar{F}_{ij}(\rho) - \bar{F}_{is}(\rho))_+ h & \text{if } s \in N_S(j), \ k = i; \\ 1 - \sum_{k \in N_L(i)} (\bar{E}_{ij}(\rho) - \bar{E}_{kj}(\rho))_+ h \\ \quad - \sum_{l \in N_S(j)} (\bar{F}_{ij}(\rho) - \bar{F}_{is}(\rho))_+ h + o(h) & \text{if } (k, s) = (i, j); \\ 0 & \text{otherwise.} \end{cases}$$

(63), along with Markov process $(L_\beta(t), X_\beta(t))$, provides many interesting asymptotic behaviors of games including spatial structures.

4.5.4 Examples

In this section, we demonstrate (63) by several spatial population games.

Example 25 (Spatial Prisoner's dilemma) *Here the Prisoner's dilemma [86] is with strategy set $S = \{C, D\}$, representing "Cooperation" and "Defection", and payoff matrix $A = \begin{pmatrix} -1 & -3 \\ 0 & -2 \end{pmatrix}$. And we assume the spatial graph is a Lattice graph.*

We apply the Fokker-Planck equation (63) to model this game. Here we want to investigate how a defector invades the cooperators.[7] I.e. we consider a special initial

---

[7] *The cooperator means a player who chooses C while the defector means a player who chooses D.*

condition of (63), meaning that there is a vertex $\mathbf{l}_{i^*} \in L$ with

$$\rho_{i^*D}(0) = \frac{1}{m} - \eta, \quad \rho_{i^*C}(0) = \eta, \quad and \quad \rho_{iC}(0) = \frac{1}{m} - \eta, \quad \rho_{iD}(0) = \eta, \ for \ all \ i \neq i^*,$$

(64)

where m is the number of vertices in G_L and η is a sufficient small value.

At the beginning, we consider a 3×3 lattice spatial graph. Let the initial condition satisfy (64), with $\beta = 0.01$, $\eta = 10^{-4}$, $m = 9$, where i^* is the left corner of lattice. We plot $\rho(t) = (\rho_{ij}(t))_{(i,j) \in L \times S}$ at $t = 0.3$, see Figure 14. Here the red, green graphs represent $\rho_{iC}(t)$, $\rho_{iD}(t)$, for any $\mathbf{l}_i \in L$.

(a) C $t = 0$ (b) D $t = 0$

(c) C $t = 0.3$ (d) D $t = 0.3$
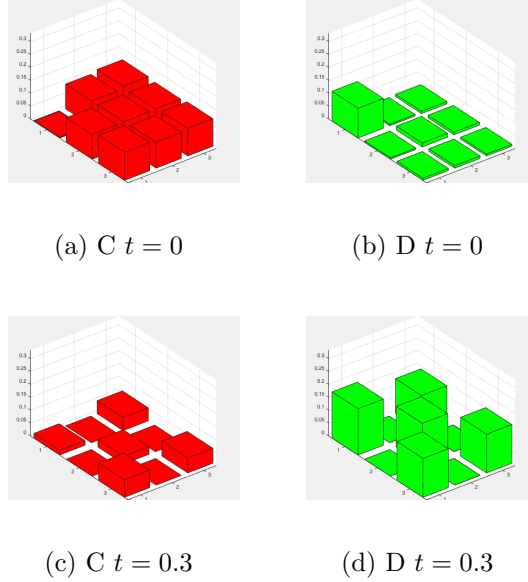
Figure 14: Spatial Prisoner's Dilemma, 3×3 spatial lattice

Secondly, we show that there exists multiple equilibria of (63). For example, we consider an initial measure (64) with $\beta = 0.01$, $\eta = 10^{-4}$, $m = 36$, i^* being at the left corner (Figure 15) or middle (Figure 16) of lattice graph. We show that there are two equilibria with respect to different initial conditions.

(a) C $t = 0$ (b) D $t = 0$

(c) C $t = 0.75$ (d) D $t = 0.75$
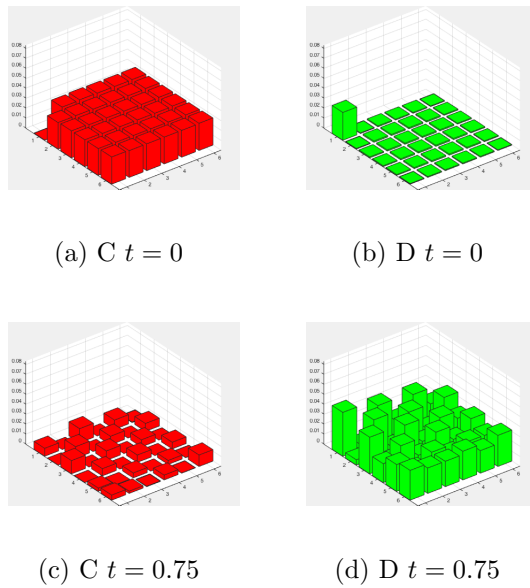
Figure 15: Spatial Prisoner's Dilemma, 6×6 spatial lattice I

(a) C $t = 0$ (b) D $t = 0$

(c) C $t = 1$ (d) D $t = 1$
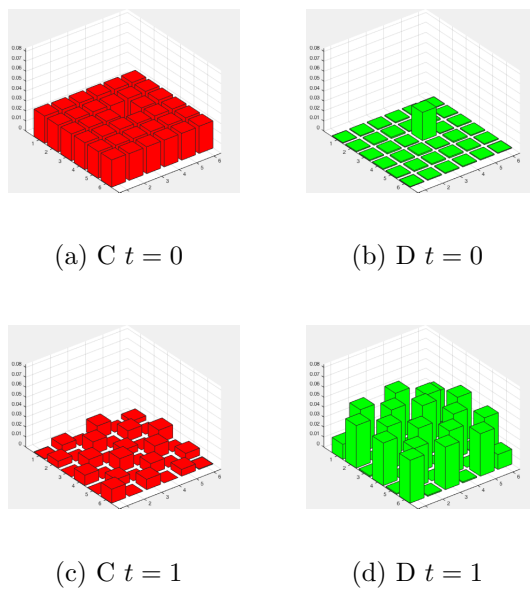
Figure 16: Spatial Prisoner's Dilemma, 6×6 spatial lattice II

Example 26 (Spatial Hawk-Dove game) *We consider a spatial Hawk-Dove game [60]. Let the payoff matrix be $A = \begin{pmatrix} -1 & -3 \\ -5 & 0 \end{pmatrix}$ and the spatial graph be a 3×3 lattice.*

We apply Fokker-Planck equation (63) to model this game. As in spatial Prisoner's Dilemma, let the initial condition (64) be with $\eta = 10^{-4}$, $m = 9$, and i^* be the left corner of lattice, we demonstrate the equilibrium of (63), see Figure 17.

(a) C $t = 2$ (b) D $t = 2$
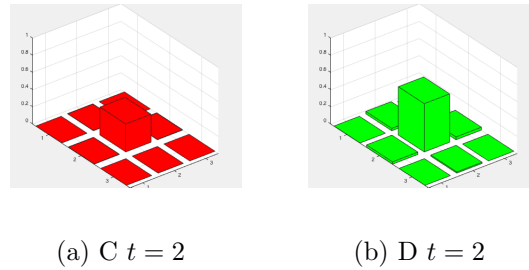
Figure 17: Spatial Hawk-Dove game, 3×3 spatial lattice

Example 27 (Potential games) *We consider a potential game. Let the payoff matrix be $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ and the spatial graph be a 6×6 lattice. In this case, Figure 18 shows that there is one unique equilibrium of* (63). *This is true for considering different initial position $\mathbf{l}_{i*}$ in* (64) *with $\eta = 10^{-4}$, $m = 36$.*

(a) C $t = 2$ (b) D $t = 2$
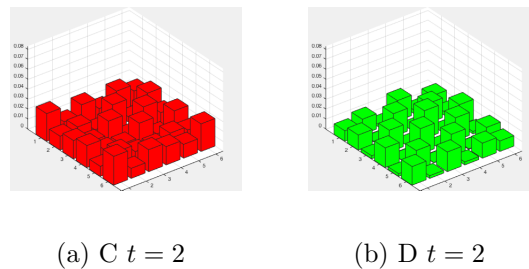
Figure 18: Spatial potential game, 6×6 spatial lattice

CHAPTER V

APPLICATION II: NUMERICAL SCHEMES FOR FOKKER-PLANCK EQUATIONS

5.1   Introduction

In this chapter, we introduce the second application of Fokker-Planck equations on finite graphs, which is new numerical scheme for a certain type of drift diffusion equations.

Consider a nonlinear Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [\rho \nabla (V(x) + \int_{\mathbb{R}^d} W(x,y)\rho(t,y)dy)] + \beta \Delta \rho. \tag{65}$$

Here the solution $\rho(t,x)$ is a probability density function supported on $\mathbb{R}^d$, which maintains positivity and conserves the total probability. And $V : \mathbb{R}^d \to \mathbb{R}$, $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ are functions with $W(x,y) = W(y,x)$ for any $x,y \in \mathbb{R}^d$. From the viewpoint of optimal transport [4, 95], (65) is a gradient flow of the following scalar functional, named free energy

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^d} V(x)\rho(x)dx + \frac{1}{2}\int_{\mathbb{R}^d \times \mathbb{R}^d} W(x,y)\rho(x)\rho(y)dxdy + \beta \int_{\mathbb{R}^d} \rho(x)\log\rho(x)dx, \tag{66}$$

There are many gradient flow structures of (65). For example, the free energy is the Lyapunov function of (65):

$$\frac{d}{dt}\mathcal{F}(\rho) = -\int_{\mathbb{R}^d} (\nabla F(x,\rho))^2 \rho(t,x)dx;$$

111

The minimizer of free energy, Gibbs measure is the equilibrium of (65);[1] Under suitable conditions, $\rho(t, x)$ converges to a Gibbs measure exponentially [23].

In this chapter, we derive a semi scheme [2] for (65) with a gradient flow structure. In details, we shall consider a finite graph $G = (V, E)$ to discretize the domain, where V is a vertex set $\{1, 2, \cdots, n\} \subset \mathbb{R}^d$ and E is an edge set. For concreteness, we will assume that G is a lattice graph corresponding to a uniform discretization of the domain with equal space.

We consider a discrete probability set supported on all vertices of G:

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

Notice that $(\rho_i)_{i=1}^n$ with graph G is a finite volume discretization of $\mathcal{P}(\mathbb{R}^d)$. In other words, ρ_i indicts a discrete probability measure

$$\rho_i = \int_{C_i} \rho(t, x) dx,$$

where C_i is a cube in $\mathbb{R}^d$ centered at point i and of width $2\Delta x$.

We consider a discrete free energy of $\mathcal{F}(\rho)$[3]

$$\mathcal{F}(\rho) = \sum_{i=1}^n v_i \rho_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \rho_i \rho_j + \beta \sum_{i=1}^n \rho_i \log \rho_i,$$

where $v_i = V(i)$ and $w_{ij} = W(i, j)$.

In this setting, the semi scheme is nothing but the gradient flow of discrete free energy with respect to discrete 2-Wasserstein metric. Similarly to chapter 3, the semi flow is

$$\frac{d\rho_i}{dt} = \frac{1}{\Delta x^2} \{ \sum_{j \in N(i)} \rho_j (F_j(\rho) - F_i(\rho))_+ - \sum_{j \in N(i)} \rho_i (F_i(\rho) - F_j(\rho))_+ \}, \qquad (67)$$

---

[1] $\rho^*(x)$ is a Gibbs measure, if it solves the fixed point problem

$$\rho^*(x) = \frac{1}{K} e^{-\frac{V(x) + \int_{\mathbb{R}^d} W(x,y) \rho^*(y) dy}{\beta}}, \quad \text{where} \quad K = \int_{\mathbb{R}^d} e^{-\frac{V(x) + \int_{\mathbb{R}^d} W(x,y) \rho^*(y) dy}{\beta}} dx.$$

[2] We only discretize the spatial variable, not the time variable

[3] It is the first order discretization of (66).

where $F_i(\rho) = \frac{\partial}{\partial \rho_i} \mathcal{F}(\rho)$, for any $i \in V$ and $(\cdot)_+ = \max\{\cdot, 0\}$.

As in continuous states, we can demonstrate (67)'s gradient flow structure. Firstly, the free energy is a Lyapunov function of (67), since

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = -\sum_{(i,j)\in E} (\frac{F_i(\rho) - F_j(\rho)}{\Delta x})^2_+ \rho_i \leq 0;$$

Secondly, (67)'s equilibrium is a discrete Gibbs measure

$$\rho_i^\infty = \frac{1}{K} e^{-\frac{v_i + \sum_{j=1}^n w_{ij}\rho_j^\infty}{\beta}}, \quad K = \sum_{i=1}^n e^{-\frac{v_i + \sum_{j=1}^n w_{ij}\rho_j^\infty}{\beta}};$$

Last and most importantly, we investigate the convergence speed to the discrete Gibbs measure. We show

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

where C is a positive constant. We can say more for this convergence, that is the asymptotic convergence rate is determined by the Hessian matrix of free energy on "Wasserstein" metric manifold at Gibbs measure.

This chapter is arranged as follows. In section 5.2, we derive (67) based on the chapter 3. Furthermore, we introduce a semi discretization for general Fokker-Planck equations in section 5.3. Several examples are demonstrated in section 5.4.

5.2 Gradient flows

Let's begin with the derivation of gradient flow. By modifying the inner product $\frac{1}{\Delta x^2}$ in chapter 3,

$$\Delta x^2 g_{ij}(\rho) := \begin{cases} \rho_i & \text{if } F_i(\rho) > F_j(\rho), \ j \in N(i); \\[2mm] \rho_j & \text{if } F_i(\rho) < F_j(\rho), \ j \in N(i); \\[2mm] \frac{\rho_i + \rho_j}{2} & \text{if } F_i(\rho) = F_j(\rho), \ j \in N(i), \end{cases}$$

we derive the semi-discretization in Theorem 34.

Theorem 34 (Derivation) *Given a graph G and a constant $\beta > 0$. Then the "generalized" gradient flow of discrete free energy $\mathcal{F}(\rho)$ on metric manifold $(\mathcal{P}_o(G), W_{2;\mathcal{F}})$ is*

$$\frac{d\rho_i}{dt} = \frac{1}{\Delta x^2}\{ \sum_{j \in N(i)} \rho_j(F_j(\rho) - F_i(\rho))_+ - \sum_{j \in N(i)} \rho_i(F_i(\rho) - F_j(\rho))_+\},$$

for any $i \in V$. Here

$$F_i(\rho) = v_i + \sum_{j=1}^{n} w_{ij}\rho_j + \beta \log \rho_i + \beta.$$

(i) *For any initial $\rho^0 \in \mathcal{P}_o(G)$, there exists a unique solution $\rho(t) : [0, \infty) \to \mathcal{P}_o(G)$ to equation (67) . Moreover, there is a constant $\epsilon > 0$ depending on ρ^0, such that $\rho_i(t) \geq \epsilon$ for all $i \in V$ and $t > 0$.*

(ii) *The free energy $\mathcal{F}(\rho)$ is a Lyapunov function of (67): If $\rho(t)$ is a solution of (67), then*

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = - \sum_{(i,j) \in E} (\frac{F_i(\rho) - F_j(\rho)}{\Delta x})_+^2 \rho_i.$$

Moreover, if ρ^∞ is an equilibrium of (67), then ρ^∞ is a Gibbs measure.

Proof 32 *For any $\sigma \in T_\rho \mathcal{P}_o(G)$, there exists $[\Phi] \in \mathbb{R}^n/ \sim$, such that $\tau([\Phi]) = \sigma$. Since*

$$d\mathcal{F}(\rho) \cdot \sigma = \sum_{i=1}^{n} \frac{\partial}{\partial \rho_i}\mathcal{F}(\rho) \cdot \sigma_i$$

$$= \frac{1}{\Delta x^2} \sum_{i=1}^{n} F_i(\rho) \sum_{j \in N(i)} g_{ij}(\rho)(\Phi_i - \Phi_j)$$

$$= \frac{1}{\Delta x^2}\{\sum_{i=1}^{n} \sum_{j \in N(i)} g_{ij}(\rho)F_i(\rho)\Phi_i - \sum_{i=1}^{n} \sum_{j \in N(i)} g_{ij}(\rho)F_i(\rho)\Phi_j\}$$

Relabel i and j on second formula

$$= \frac{1}{\Delta x^2}\{\sum_{i=1}^{n} \sum_{j \in N(i)} g_{ij}(\rho)F_i(\rho)\Phi_i - \sum_{i=1}^{n} \sum_{j \in N(i)} g_{ji}(\rho)F_j(\rho)\Phi_i\}$$

$$= \frac{1}{\Delta x^2}\{\sum_{i=1}^{n} \sum_{j \in N(i)} g_{ij}(\rho)\big(F_i(\rho) - F_j(\rho)\big)\Phi_i\}.$$

Combining the above formula into the definition of gradient flow, we have

$$\left(\frac{d\rho}{dt}, \sigma\right)_\rho + d\mathcal{F}(\rho) \cdot \sigma$$

$$= \sum_{i=1}^n \left\{\frac{d\rho_i}{dt} + \frac{1}{\Delta x^2} \sum_{j \in N(i)} g_{ij}(\rho)\big(F_i(\rho) - F_j(\rho)\big)\right\}\Phi_i$$

$$=0.$$

Since the above formula is true for all $(\Phi_i)_{i=1}^n \in \mathbb{R}^n$,

$$\frac{d\rho_i}{dt} + \frac{1}{\Delta x^2} \sum_{j \in N(i)} g_{ij}(\rho)\big(F_i(\rho) - F_j(\rho)\big) = 0$$

holds for all $i \in V$. From the definition of g_{ij}, we derive (67).

(i)'s proof is similarly to chapter 3. Here we only show (ii), which is to motivate readers to why (67) forms a gradient system. We show that $\mathcal{F}(\rho)$ is a Lyapunov function:

$$\frac{d}{dt}\mathcal{F}(\rho(t)) = \sum_{i=1}^n F_i(\rho) \cdot \frac{d\rho_i}{dt}$$

$$= \frac{1}{\Delta x^2}\left\{\sum_{i=1}^n \sum_{j \in N(i)} F_i(\rho)(F_j(\rho) - F_i(\rho))_+\rho_j - \sum_{i=1}^n \sum_{j \in N(i)} F_i(\rho)(F_i(\rho) - F_j(\rho))_+\rho_i\right\}$$

Switch i, j on the first formula

$$= \frac{1}{\Delta x^2}\left\{\sum_{i=1}^n \sum_{j \in N(i)} F_j(\rho)(F_i(\rho) - F_j(\rho))_+\rho_i - \sum_{i=1}^n \sum_{j \in N(i)} F_i(\rho)(F_i(\rho) - F_j(\rho))_+\rho_i\right\}$$

$$= -\sum_{(i,j) \in E} \left(\frac{F_i(\rho) - F_j(\rho)}{\Delta x}\right)^2_+ \rho_i \le 0.$$

We show that if $\rho^\infty = \lim_{t\to\infty} \rho(t)$ exists, then ρ^∞ is a Gibbs measure. Since $\mathcal{F}(\rho)$ is bounded in $\mathcal{P}(G)$, then $\lim_{t\to\infty} \frac{d}{dt}\mathcal{F}(\rho(t)) = 0$. And from (i), we know $\rho^\infty \ge \epsilon > 0$, then

$$\sum_{i=1}^n \sum_{j \in N(i)} (F_i(\rho^\infty) - F_j(\rho^\infty))^2_+ \rho_i^\infty = 0,$$

which implies $F_i(\rho^\infty) = F_j(\rho^\infty)$ for any $(i,j) \in E$. Since the graph is connected,

$$F_i(\rho^\infty) = F_j(\rho^\infty), \quad \text{for any } i, j \in V.$$

Let

$$C := v_i + \sum_{j=1}^{n} w_{ij} \rho_j^\infty + \beta \log \rho_i^\infty, \quad \text{for any } i \in V,$$

$K = e^{-C}$ *and use the fact* $\sum_{i=1}^{n} \rho_i^\infty = 1$, *we have*

$$\rho_i^\infty = \frac{1}{K} e^{-\frac{v_i + \sum_{j=1}^{n} w_{ij} \rho_j^\infty}{\beta}}, \quad K = \sum_{j=1}^{n} e^{-\frac{v_i + \sum_{j=1}^{n} w_{ij} \rho_j^\infty}{\beta}}.$$

Hence ρ^∞ *is a Gibbs measure, which finishes the proof.*

An interesting question associated with gradient flow arises. How fast is the convergence towards to the Gibbs measure? In this part, we answer the question from dynamical viewpoint. We introduce a quantity playing the role of smallest eigenvalue of Hessian matrix at the Gibbs measure:

Definition 35 *Let*

$$h_{ij,kl} = f_{ik} + f_{jl} - f_{il} - f_{jk} \quad \text{for any } i, j, k, l \in V.$$

We define

$$\lambda_{\mathcal{F}}(\rho) = \min_{\Phi} \frac{1}{\Delta x^4} \sum_{(i,j) \in E} \sum_{(k,l) \in E} h_{ij,kl} (\Phi_i - \Phi_j)_+ \rho_i (\Phi_k - \Phi_l)_+ \rho_k$$

where the minimum is taken among all $(\Phi_i)_{i=1}^n \in \mathbb{R}^n$ *with*

$$\sum_{(i,j) \in E} \left(\frac{\Phi_i - \Phi_j}{\Delta x} \right)_+^2 \rho_i = 1.$$

Theorem 36 (Convergence) *If* ρ^∞ *is a strict minimizer of* $\mathcal{F}(\rho)$, *then there exists a constant* $C > 0$, *such that*

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

Moreover, the asymptotic convergence rate is $2\lambda_{\mathcal{F}}(\rho^\infty)$. *I.e. for any sufficient small* $\epsilon > 0$, *there exists a time* $T > 0$, *such that when* $t > T$,

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

The proof of above theorem can be found in chapter 3.

5.3 Semi-discretizations

Motived by (67), we introduce a semi-discretization for general Fokker-Planck equation.

In details, we consider a general Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [\rho(f_v(x, \rho))_{v=1}^d]. \tag{68}$$

If (68) is a gradient flow, the vector field $(f_v)_{v=1}^d$ is a gradient field, which implies that there exists a scalar functional $F(x, \rho)$, such that

$$\nabla F(x, \rho) := (f_v(x, \rho))_{v=1}^d.$$

But for general vector fields, such F doesn't exist. However, there always exists a vector functional $(u_v(x, \rho))_{i=1}^d$, such that

$$\nabla_{x_v} u_v(x, \rho) = f_v(x, \rho), \quad \text{for } v \in \{1, \cdots, d\}.$$

In others words, the gradient flow

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [\rho(\nabla_{x_v} F(x, \rho))_{v=1}^d],$$

is a special case of flow (68)

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [\rho(\nabla_{x_v} u_v(x, \rho))_{v=1}^d].$$

This observation still holds for the semi-discretization. Notice that the gradient flow (30)[4]

$$\frac{d\rho}{dt} = \frac{1}{\Delta x^2} \{ \sum_{v=1}^d \sum_{j \in N_v(i)} \rho_j(F_j(\rho) - F_i(\rho))_+ - \sum_{v=1}^d \sum_{j \in N_v(i)} \rho_i(F_i(\rho) - F_j(\rho))_+ \},$$

---

[4]Here N_v is the adjacency set for the discretization of dimension v. Notice that G is a cartesian product of d's one dimensional lattice, $G = G_1 \square \cdots \square G_d$ with $G_v = (V_v, E_v)$. We denote a node by $i = (i_1, \cdots, i_d)$. Then

$$N_v(i) = \{(i_1, \cdots, i_{v-1}, j_v, i_{v+1}, \cdots, i_d) \mid (i_v, j_v) \in E_v\}.$$

is a special case of formulation

$$\frac{d\rho_i}{dt} = \frac{1}{\Delta x^2}\{\sum_{v=1}^{d}\sum_{j\in N_v(i)}[u_v(j,\rho)-u_v(i,\rho)]_+\rho_j - \sum_{v=1}^{d}\sum_{j\in N_v(i)}[u_v(i,\rho)-u_v(j,\rho)]_+\rho_i\}. \quad (69)$$

In all, we derive a new semi-discretization (69). We demonstrate an example of this new semi-discretization.

Example 28 (van der Pol oscillator) *Consider a 2 dimensional Fokker-Planck equation:*

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \begin{pmatrix} x_2 \\ \alpha(1-x_1^2)-x_2 \end{pmatrix}) + \frac{\partial^2 \rho}{\partial x_2^2}$$

$$= \nabla \cdot (\rho \begin{pmatrix} -x_2 \\ -\alpha(1-x_1^2)+x_2+\nabla_{x_2}\log\rho \end{pmatrix}).$$

where $f_1(x,\rho) = -x_2$ and $f_2(x,\rho) = -\alpha(1-x_1^2)+x_2+\nabla_{x_2}\log\rho$. We let

$$u_1(x,\rho) = \int f_1(x,\rho)dx_1 = -x_1 x_2,$$

and

$$u_2(x,\rho) = \int f_2(x,\rho)dx_2 = -\alpha(1-x_1^2)x_2 + \frac{1}{2}x_2^2 + \log\rho(x_1,x_2).$$

Hence the semi-discretization (69) becomes

$$\frac{d\rho_i}{dt} = \frac{1}{\Delta x^2}\{\sum_{j\in N_1(i)}\rho_j[u_1(j,\rho)-u_1(i,\rho)]_+ - \sum_{j\in N_1(i)}\rho_i[u_1(i,\rho)-u_1(j,\rho)]_+$$

$$+ \sum_{j\in N_2(i)}\rho_j[u_2(j,\rho)-u_2(i,\rho)]_+ - \sum_{j\in N_2(i)}\rho_i[u_2(i,\rho)-u_2(j,\rho)]_+\},$$

where $u_1(i,\rho) = u_1(x(i),\rho)$, $u_2(i,\rho) = u_2(x(i),\rho)$.

Remark 8 (Handling boundaries of PDE) *The semi-discretization works on a generality of graphs, which leads to a way to handle various boundaries of the PDE (65). We investigate mainly three cases, the domain of (65) is (i) $\mathbb{R}^d$; (ii) a bounded open set, with a zero-flux condition; (iii) a bounded open set with periodicity condition.*

The discretization graph can cover all three cases: (i) The graph is a large enough lattice in $\mathbb{R}^d$; (ii)The graph is a discretization of open set with equal distance; (iii) The graph is similar to (ii), by considering periodic points as one node.

As a completion, we show that (69) is a consistent semi-discretization.

Theorem 37 *The semi-discretization* (69) *is a consistent finite volume scheme for the PDE* (68).

Proof 33 *Let's prove the consistency. For any $v = 1, \cdots, d$, e_v denote the vector $e_v = (0, \cdots, 1, \cdots, 0)^T$, where 1 is in the v-th position. Recall that $i \in \mathbb{R}^d$ represents the position of the point x_i. Notice that $N_v(i)$ contains points $x_i - e_v \Delta x$, $x_i + e_v \Delta x$ on $\mathbb{R}^d$, and C_i is a cube in $\mathbb{R}^d$ centered at point i with equal width $2\Delta x$.*

Without loss of generality, we assume $u_v(x_i + e_v \Delta x, \rho) \geq u_v(x_i, \rho) \geq u_v(x_i -$

$e_v\Delta x, \rho$). *By using Taylor expansion on R.H.S of* (69) *in the direction* e_i, *we obtain*

$$\frac{1}{\Delta x^2}\{\sum_{j\in N_v(i)}[u_v(j,\rho)-u_v(i,\rho)]_+\rho_j - \sum_{j\in N_v(i)}[u_v(i,\rho)-u_v(j,\rho)]_+\rho_i\}$$

$$=\frac{1}{\Delta x^2}\{[u_v(x_i+e_v\Delta x,\rho)-u_v(x_i,\rho)]\int_{C_{i+e_v\Delta x}}\rho(t,x)dx$$

$$-[u_v(x_i,\rho)-u_v(x_i-e_v\Delta x,\rho)]\int_{C_i}\rho(t,x)dx\}$$

$$= \frac{1}{\Delta x^2}\{[\frac{\partial u_v(x_i,\rho)}{\partial x_v}\Delta x + \frac{1}{2}\frac{\partial u_v(x_i,\rho)}{\partial x_v}\Delta x^2]\int_{C_{i+e_v\Delta x}}\rho(t,x)dx$$

$$-[\frac{\partial u_v(x_i,\rho)}{\partial x_v}\Delta x - \frac{1}{2}\frac{\partial u_v(x_i,\rho)}{\partial x_v}\Delta x^2]\int_{C_{i+e_v\Delta x}}\rho(t,x)dx + O(\Delta x^3)\}$$

$$= \frac{1}{\Delta x}\frac{\partial u_v(x_i,\rho)}{\partial x_v}[\int_{C_{i+e_v\Delta x}}\rho(t,x)dx - \int_{C_i}\rho(t,x)dx]$$

$$+\frac{1}{2}\frac{\partial^2 u_v(x_i,\rho)}{\partial x_i^2}[\int_{C_{i+e_v\Delta x}}\rho(t,x)dx + \int_{C_i}\rho(t,x)dx] + O(\Delta x^3)$$

$$= \frac{\partial u_v(x_i,\rho)}{\partial x_v}\int_{C_i}\frac{\rho(t,x+e_v\Delta x)-\rho(t,x)}{\Delta x}dx$$

$$+\frac{\partial^2 u_v(x_i,\rho)}{\partial x_i^2}\int_{C_i}\frac{\rho(t,x+e_v\Delta x)+\rho(t,x)}{2}dx + O(\Delta x)$$

$$=\int_{C_i}[f_v(x,\rho)\frac{\partial\rho(t,x)}{\partial x_v}+\frac{\partial f_v(x,\rho)}{\partial x_v}\rho(t,x)]dx + O(\Delta x) \quad (Since\ \frac{\partial u_v(x,\rho)}{\partial x_v}=f_v(x,\rho))$$

$$=\int_{C_i}\nabla_{x_v}\cdot\big(\rho(t,x)\nabla_{x_v}f_v(x,\rho)\big)dx + O(\Delta x)\ .$$

$$(70)$$

Similarly, we can show the same results for other possible permutations, such as $u_v(x_i - e_v\Delta x,\rho) \geq u_v(x_i,\rho) \geq u_v(x_i + e_v\Delta x,\rho)$, $u_v(x_i,\rho) \geq u_v(x_v - e_v\Delta x,\rho) \geq u_v(x_i + e_v\Delta x,\rho)$ *etc.*

Therefore, the R.H.S. of (69) *becomes the sum of* $F_i(x)$

$$\frac{d\rho_i}{dt} - \frac{1}{\Delta x^2}\sum_{v=1}^{d}\{\sum_{j\in N_v(i)}[u_v(j,\rho)-u_v(i,\rho)]_+\rho_j - \sum_{j\in N_v(i)}[u_v(i,\rho)-u_v(j,\rho)]_+\rho_j\}$$

$$=\int_{C_i}\{\frac{\partial\rho(t,x)}{\partial t}-\sum_{v=1}^{d}\nabla_{x_v}\cdot\big(\rho(t,x)\nabla_{x_v}f_v(x,\rho)\big)\}dx + dO(\Delta x)$$

$$=dO(\Delta x),$$

which is the proposed first order discretization for PDE (68).

5.4   Examples

We illustrate the proposed semi discretization by some numerical experiments.

Example 29 *We consider a 2 dimensional nonlinear interaction-diffusion equation in granular gas [13, 93],*

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [\rho \nabla (W * \rho)] + \beta \Delta \rho,$$

where $W(x - y) = \|x - y\|^3$ and $\|\cdot\|$ is a 2 norm. This PDE has a unique stationary measure (Gibbs measure),

$$\rho^*(x) = \frac{1}{K} e^{-\frac{\int_{\mathbb{R}^2} W(x-y)\rho^*(y)dy}{\beta}}, \quad where \quad K = \int_{\mathbb{R}^2} e^{-\frac{\int_{\mathbb{R}^2} W(x-y)\rho^*(y)dy}{\beta}} dx.$$

To approximate the solution of the above PDE, we apply the semi-discretization (69)

$$\frac{d\rho_i}{dt} = \frac{1}{\Delta x^2} \{ \sum_{j \in N(i)} \rho_j (\sum_{i=1}^{n} w_{ij}\rho_i - \sum_{j=1}^{n} w_{ij}\rho_j + \beta \log \rho_j - \beta \log \rho_i)_+$$

$$- \sum_{j \in N(i)} \rho_i (\sum_{j=1}^{n} w_{ij}\rho_j - \sum_{i=1}^{n} w_{ij}\rho_i + \beta \log \rho_i - \beta \log \rho_j)_+ \}.$$

Let $\beta = 0.01$ and $\Delta x = 0.5$ on the lattice $[-10, 10] \times [-10, 10]$. We solve the above ODE system by Euler method with time step $\Delta t = 10^{-4}$.
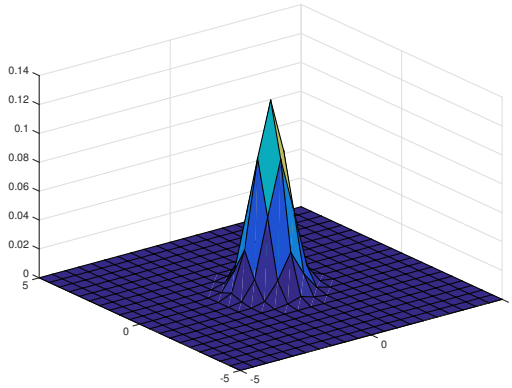
Figure 19: Stationary measure of interaction diffusion equation.
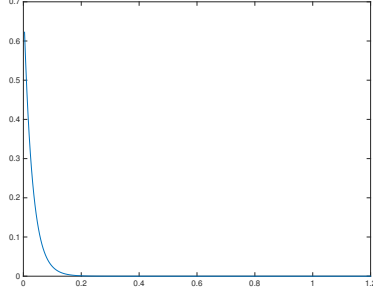
Figure 20

Theoretically, it is known that the Gibbs measure converges to a Delta measure supported at origin when $\beta \to 0$. In above figure, the behavior of Log-Laplacian reflects such result. Furthermore, as in Figure 20, Log-Laplacian reflects that the free energy decreases exponentially.

We illustrate the semi-discretization for general diffusion PDEs, which are not gradient flows.

Example 30 *We consider the Fokker-Planck equation*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \left(\rho \begin{pmatrix} x_2 \\ \alpha(1 - x_1^2) - x_2 \end{pmatrix} \right) = \beta \Delta_{x_2} \rho,$$

whose underlying state is the stochastic van der Pol oscillator

$$dx_1 = x_2 dt$$

$$dx_2 = [\alpha(1 - x_1^2)x_2 - x_1]dt + \sqrt{2\beta}dW_t.$$

We apply the semi-discretization (69) to solve this PDE:

$$\frac{d\rho_i}{dt} = \frac{1}{\Delta x^2} \Big\{ \sum_{j \in N_1(i)} \rho_j [u_{1j} - u_{1i}]_+ - \sum_{j \in N_1(i)} \rho_i [u_{1i} - u_{1j}]_+ $$
$$+ \sum_{j \in N_2(i)} \rho_j [u_{2j} - u_{2i}]_+ - \sum_{j \in N_2(i)} \rho_i [u_{2i} - u_{2j}]_+ \Big\},$$

where

$$u_{1j} = -x_1 x_2 |_{x=j}, \quad u_{2j} = -\alpha(1 - x_1^2)x_2 + \frac{1}{2}x_2^2 + \beta \log \rho_j |_{x=j}.$$

122

Let $\alpha = 1$, $\epsilon = 0.5$ with the square lattice on $[-10, 10] \times [-10, 10]$ with $\Delta x = 0.4082$. We apply Euler method with step size $\Delta t = 1.67 \times 10^{-4}$ to approximate the stationary measure.
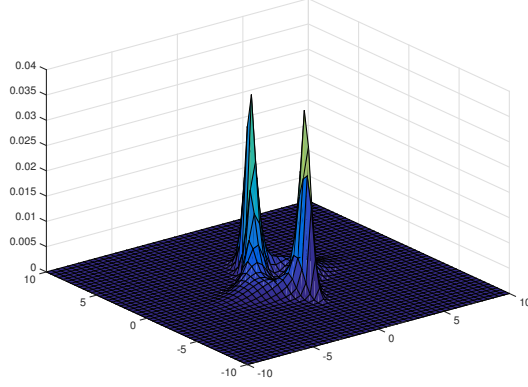
Figure 21: Stationary measure of stochastic van der Pol oscillator.

Similarly, we consider the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \left(\rho \begin{pmatrix} x_2 \\ -2\xi\omega x_2 + \omega x_1 - \omega^2 r x_1^3 \end{pmatrix} \right) = \beta \Delta_{x_2} \rho,$$

associated with the stochastic Duffing oscillator

$$dx_1 = x_2 dt$$

$$dx_2 = [-2\xi\omega x_2 + \omega x_1 - \omega^2 r x_1^3]dt + \epsilon dW_t.$$

Let $\xi = 0.2$, $\omega = 1$, $r = 0.1$, $\epsilon = 0.5$ with a square lattice on $[-10, 10] \times [-10, 10]$. The computed invariant measure is shown below.
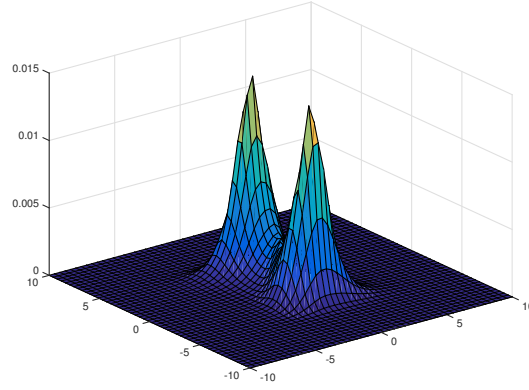
Figure 22: Stationary measure of stochastic Duffing oscillator.

5.5   Discussions

In this chapter, we have derived a semi-discretization scheme for a certain type of PDEs. Compared with traditional discretization methods, the new rule scheme following distinct properties:

- It brings the effect of Log Laplacian. The difference of log term coincides with the rule of Laplacian in PDE, from its "boundary repeller property" to the asymptotic convergence result;

- If the PDE is a gradient flow, the semi-discretization keeps the gradient flow structure;

- The graph in scheme naturally handles PDEs' underlying states and boundary conditions.

In addition, our semi-discretization can be applied to the other type of PDEs. For example, consider the nonlinear diffusion PDE in $\mathbb{R}^d$

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [\rho \nabla (V(x) + \beta \Delta(\rho^m))], \quad m > 1,$$

which is associated with the free energy

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^d} V(x)\rho(x)dx + \beta \int_{\mathbb{R}^d} \frac{1}{m-1}\rho(x)^m dx.$$

Thus, our discretization rule provides the following result.

Corollary 38 *The gradient flow of a discrete free energy*

$$\mathcal{F}(\rho) = \sum_{i=1}^{n} v_i \rho_i + \beta \sum_{i=1}^{n} \frac{1}{m-1} \rho_i^m, \quad m > 1,$$

on the metric manifold $(\mathcal{P}_o(G), W_{2;\mathcal{F}})$ *is*

$$\frac{d\rho_i}{dt} = \frac{1}{\Delta x^2} \{ \sum_{j \in N(i)} \rho_j [F_j(\rho) - F_i(\rho)]_+ - \sum_{j \in N(i)} \rho_i [F_i(\rho) - F_j(\rho)]_+ \}, \tag{71}$$

for any $i \in V$. *Here*

$$F_i(\rho) = v_i + \beta \frac{m}{m-1} \rho_i^{m-1}.$$

There exists a unique solution $\rho(t) : [0, \infty) \to \mathcal{P}(G)$ *to equation (71) with initial measure* $\rho^0 \in \mathcal{P}(G)$.

Proof 34 *Since* $F_i(\rho)$ *is a Lipschitz continuous function on a bounded manifold* $\mathcal{P}(G)$, *there exists a unique solution of* (71) *if* $\rho(t) \in \mathcal{P}(G)$. *So we only need to show that* $\rho(t) \in \mathcal{P}(G)$ *for any* $t \geq 0$. *Since* $\rho(0) \in \mathcal{P}(G)$, *it is only sufficient to show that the boundary of probability manifold* $\partial \mathcal{P}(G)$ *is a repeller.* **Claim**: *for any* $t \geq 0$, $i^* \in L$ *with* $\rho_{i^*}(t) = 0$, $\frac{d}{dt} \rho_{i^*}(t) \geq 0$. **Proof of claim**: *Since* $F(\rho)$ *is continuous on a bounded manifold, then* $F_i(\rho)$ *is bounded for all* $i \in L$. *Combining with the fact* $\rho_{i^*}(t) = 0$, *we have*

$$\frac{d}{dt} \rho_{i^*}(t) = \frac{1}{\Delta x^2} \{ \sum_{j \in N(i^*)} \rho_j [F_j(\rho) - F_{i^*}(\rho)]_+ - \sum_{j \in N(i^*)} \rho_{i^*} [F_{i^*}(\rho) - F_j(\rho)]_+ \}$$

$$= \frac{1}{\Delta x^2} \{ \sum_{j \in N(i^*)} \rho_j [F_j(\rho) - F_{i^*}(\rho)]_+ \} \geq 0 ,$$

.

We demonstrate the above result through a numerical example.

Example 31 *Consider the 2 dimensional nonlinear diffusion PDE in [23].*

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [\rho \nabla (V(x))] + \beta \Delta (\rho^m),$$

where $V(x) = \frac{x^2}{2}$ and $m = 2$. We solve this PDE by the semi-discretization (71)

$$\frac{d\rho_i}{dt} = \frac{1}{\Delta x^2}\{ \sum_{j \in N(i)} \rho_j[v_j - v_i + \frac{\beta m}{m-1}(\rho_j^{m-1} - \rho_i^{m-1})]_+$$

$$- \sum_{j \in N(i)} \rho_i[v_i - v_j + \frac{\beta m}{m-1}(\rho_i^{m-1} - \rho_j^{m-1})]_+\}$$

with $\beta = 0.01$ and $\Delta x = 0.2$ on a lattice $[-2, 2] \times [-2, 2]$. The solution of the semi discretization is approximated by Euler method with time step $\Delta t = 0.1$.
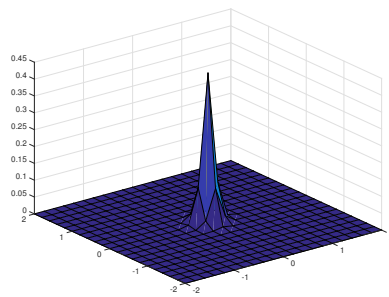
(a) Plot of $\rho(t, x)$ at time $t = 10$ (b) Free energy vs time
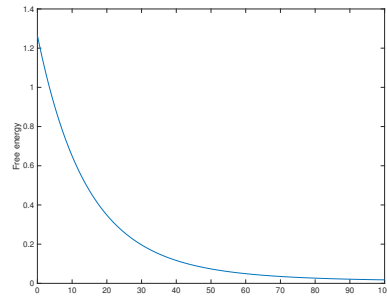
Figure 23: Stationary measure and convergence speed of Nonlinear diffusion equation

CHAPTER VI

PART 2: A NEW ALGORITHM FOR OPTIMAL CONTROL WITH CONSTRAINTS

In this chapter, we focus on the computation part of this thesis. We design a SDE based algorithm for optimal control with constraints.

Optimal control with constraints seeks to determine the input (control) to a dynamical system that optimizes a given performance functional (maximize profit, minimize cost, etc), while satisfying different kinds of constraints. Mathematically, the problem can usually be posed as

$$\min_{x,u} \int_{t_0}^{t_f} L(x(t), u(t), t)dt + \psi(x(t_f), t_f), \tag{72}$$

where the state variable $x(t) \in \mathbb{R}^n$, and the control $u(t) \in \mathbb{R}^r$ are subject to a dynamical system

$$\dot{x} = f(x(t), u(t), t), \quad t_0 \leq t \leq t_f,$$

$$x(t_0) = x_0, \quad M(t_f, x(t_f)) = 0,$$

with state (phase) and control constraints

$$\phi(x(t), t) \geq 0, \quad \varphi(u(t), t) \geq 0, \quad t_0 \leq t \leq t_f.$$

In literature, $x(t)$ is called the trajectory or path. $L \colon \mathbb{R}^n \times \mathbb{R}^r \times \mathbb{R}^+ \to \mathbb{R}$ is the Lagrangian; $\psi \colon \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}$ the terminal cost, and t_f the terminal time, which may be undetermined in some problems. $\phi \colon \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^p$ is the state constraint and $\varphi \colon \mathbb{R}^r \times \mathbb{R}^+ \to \mathbb{R}^q$ the control constraint. For technical simplicity, we assume that L, ϕ, φ, M are continuously differentiable with respect to x and t in this chapter.

Because many engineering problems can be formulated into the framework of optimal control (72) [57, 67], the optimal control theory has vast applications [5, 6, 20]. However, due to the complexity of those applications, few of them can be solved analytically. Thus numerical methods are often employed instead. Traditionally, the methods are divided into three categories, (1) state-space, (2) indirect, and (3) direct methods. *State-space* approaches apply the principle of dynamic programming, which states that each sub-arc of the optimal trajectory must be optimal. It leads to the well-known Hamilton-Jacobi-Bellman (HJB) equations, which are non-linear partial differential equations (PDEs) [12, 74]. *Indirect methods* employs the necessary condition of optimality known as Pontryagin's Maximum Principle [81]. This leads to a boundary value problem, which is then solved by numerical methods. Thus this approach is also referred to as "first optimize, then discretize". The boundary value problem is usually solved by shooting techniques or by collocation, for examples, neighboring extremal algorithm, gradient algorithm, quasi-linearization algorithm [9, 21, 64, 70], just to name a few. *Direct methods* take the "first discretize, then optimize" idea. They convert the original continuous infinite dimensional control problem into a finite dimensional optimization problem. This conversion is achieved by, for example, approximating the original control by piecewise constant controls. The resulting discrete problem becomes a large scale standard nonlinear programming problem (NLP) which can be solved by many well established algorithms such as Newton's method, Quasi-Newton methods [40, 45, 49, 75]. Direct methods are nowadays the most widespread and successfully used techniques.

Different from the existing methods, in this chapter, we design a new fast numerical method focusing on a special class of problem (72). That is the optimal trajectory exhibits certain structures, known as separability [29, 30, 31, 65]. Simply put, a path $\gamma \colon [t_0, t_f] \to \mathbb{R}^n$ is said to be separable, if there exists finite number of points, called

junctions

$$(\tilde{x}_0, , \tilde{x}_1, \cdots, \tilde{x}_{N+1}), \quad \tilde{x}_i = (t_i, x_i) \in \mathbb{R}^+ \times \mathbb{R}^n$$

such that γ can be represented as

$$\gamma_0(\tilde{x}_0, \tilde{x}_1) \cdot \gamma_c(\tilde{x}_1, \tilde{x}_2) \cdot \gamma_0(\tilde{x}_2, \tilde{x}_3) \cdot \gamma_c(\tilde{x}_3, \tilde{x}_4) \cdots \gamma_0(\tilde{x}_N, \tilde{x}_{N+1})$$

where $\gamma_0(\tilde{x}_i, \tilde{x}_{i+1})$ is the optimal trajectory connecting $\tilde{x}_i$ and $\tilde{x}_{i+1}$ with inactive constraints and $\gamma_c(\tilde{x}_i, \tilde{x}_{i+1})$ is the optimal trajectory connecting $\tilde{x}_i$ and $\tilde{x}_{i+1}$ with active constraints and $\gamma_0 \cdot \gamma_c$ is the concatenation of two trajectories.

The significance of being separable is that the determination of the entire path boils down to the determination of only a finite number of junctions and the determination of a finite number of optimal trajectories of smaller sizes, namely, γ_0 and γ_c, for which the constraints are either inactive or active on the entire segment. On the other hand, in many applications, γ_0 and γ_c can be computed either analytically or numerically by more efficient algorithms. In this way, the original infinite dimensional problem of finding the whole path is converted into a finite dimensional problem - determine a finite number of junctions, while the constraints can also be naturally handled as functions of junctions. Thus one gains a tremendous dimension reduction.

The resulting finite dimensional problem can be handled by many established algorithms, for example, the gradient descent method. In this case, each steady state of the gradient descent flow is a, possibly local, minimizer. It is evident from many applications that the total number of minimizers can often be very large. Therefore, it is highly desirable to design methods that are capable of obtaining the global optimal trajectory. In this chapter, we adopt a recently developed global optimization strategy, called intermittent diffusion (ID) [32], to do so. The idea is to add noise (diffusions) to the gradient flow intermittently. When the noise is turned off, one gets a pure gradient flow and it quickly converges to a local minimizer. On the other hand, when the noise is turned on, the perturbed flow becomes stochastic differential

equations (SDEs), and it has positive probability to jump out of the local traps and converges to other minimizers, including the global one. It can be shown that the local minimizers obtained will include the global one with probability arbitrarily close to 1 if appropriate perturbations are added. We call the method outlined above *Method of Evolving Junctions* (MEJ).

In the literature, the concept of junction has been introduced in the past [51], and used in Indirect methods [17, 18]. Most of them use junctions as shooting parameters to solve the Hamiltonian systems. For example, the one proposed in [17] uses a continuation method, also called homotopy method, together with the shooting method for the boundary value ODEs derived from maximum principle. This is different from how we use junctions, namely, we directly derive equations that govern the evolution of junctions to achieve the optimal control requirements.

Because MEJ is designed for separable problems and leverage the structure of the optimal solutions, it is able to overcome some well-known limitations of the aforementioned three general methods. Namely, HJB approach, which gives the global solution, can be computationally expensive and suffers from the notorious problem known as "curse of dimensionality". Indirect methods guarantee to find local optimal solutions, while carefully designed, if possible, initializations are needed when one wants to find the global optimal solutions. Direct methods require finer discretization (smaller time steps) if better accuracy is expected, and this leads to higher computational cost.

We arrange this chapter as follows. In section 2, we explain the idea of separability and give the algorithm for MEJ. In section 3, we use the new method to solve two linear quadratic optimal control problems. One is the test problem introduced in [63], the other is the robot path-planning problem. Through them, we demonstrate the advantages of MEJ.

6.1 Method of Evolving Junctions

In this section, we derive MEJ to solve the optimal control (72) by three steps. Firstly, we introduce the separable structure, and search the global minimizer from all trajectories determined by such a structure. This allows us to convert (72) into a finite dimensional optimization problem. Secondly, we apply the intermittent diffusion (ID) to find the global minimizer by solving initial value SDEs. In the third step, we present the criteria to add and remove junctions dynamically during the process. In the end of this section, we combine the three steps together and form an algorithm for MEJ.

In order to better explain our idea, we consider (72) only with state constraints, and the control constraints are omitted in the introduction of MEJ. However, with nominal modifications, the proposed method can be applied to problems with control constraints as well. In this chapter, the two presented numerical experiments contain both state and control constraints.

6.1.1 The Separable Structure

We start with the definition of separability.

Definition 39 *A path $x(t)$ is said to be separable if there exists a finite partition: $t_0 < t_1 < t_2 < \cdots < t_N < t_{N+1} = t_f$ such that $x(t)|_{[t_i, t_{i+1}]}$ alternates between the free space and on the boundary of the constraints $\phi = (\phi_k)_{k=1}^p \in \mathbb{R}^p$. In other words,*

$$
\hat{\phi}(x(t), t) \begin{cases} = 0, & t \in [t_i, t_{i+1}], \quad i \ odd; \\ > 0, & t \in (t_i, t_{i+1}), \quad i \ even, \end{cases}
$$

where $\hat{\phi}(x(t), t) = \min_{k \in \{1, \cdots, p\}} \phi_k(x(t), t).$

Remark 9 *To simplify our derivation, we consider the case where the optimal trajectory is in the free space at the beginning, which implies that i is even for free space and i is odd for constraints.*

The notion of separability has been given in [88] in which only trajectories consisting of three parts are considered ($N = 2$). In this chapter, we denote $\tilde{x}_i := (t_i, x(t_i))$ and call them **junctions**. Two junctions $\tilde{x}_i = (t_i, x_i)$, $\tilde{x}_{i+1} = (t_{i+1}, x_{i+1})$ define the optimal control in the free space as

$$J_0(\tilde{x}_i, \tilde{x}_{i+1}) := \min_{x,u} \int_{t_i}^{t_{i+1}} L(x(t), u(t), t)dt,$$

where $\dot{x} = f(x, u, t)$, $x(t_i) = x_i$, $x(t_{i+1}) = x_{i+1}$, $\hat{\phi}(x(t), t) > 0$, $t \in (t_i, t_{i+1})$, with optimal trajectory denoted as

$$\gamma_0(\tilde{x}_i, \tilde{x}_{i+1}) := \arg\min_{x} \int_{t_i}^{t_{i+1}} L(x(t), u(t), t)dt;$$

and the optimal control problem on the boundary of the constraints as (subscript "c" means constrained)

$$J_c(\tilde{x}_i, \tilde{x}_{i+1}) := \min_{x,u} \int_{t_i}^{t_{i+1}} L(x(t), u(t), t)dt,$$

where $\dot{x} = f(x, u, t)$, $x(t_i) = x_i$, $x(t_{i+1}) = x_{i+1}$, $\hat{\phi}(x(t), t) = 0$, $t \in [t_i, t_{i+1}]$, with optimal trajectory denoted as

$$\gamma_c(\tilde{x}_i, \tilde{x}_{i+1}) := \arg\min_{x} \int_{t_i}^{t_{i+1}} L(x(t), u(t), t)dt.$$

The separability of the optimal trajectory enables us to restrict our search of optimal trajectories in a subset H,

$$H := \{\gamma \ : \ \gamma \text{ is determined by finitely many junctions}\}.$$

More precisely, if $\gamma \in H$, there exists a finite sequence of junctions on the boundary of the constraints, $(\tilde{x}_0, \tilde{x}_1, \cdots, \tilde{x}_{N+1})$ such that γ can be represented as

$$\gamma_0(\tilde{x}_0, \tilde{x}_1) \cdot \gamma_c(\tilde{x}_1, \tilde{x}_2) \cdot \gamma_0(\tilde{x}_2, \tilde{x}_3) \cdot \gamma_c(\tilde{x}_3, \tilde{x}_4) \cdots \gamma_0(\tilde{x}_N, \tilde{x}_{N+1}). \tag{73}$$

Here $\gamma_0 \cdot \gamma_c$ is the concatenation of two trajectories.

As a result, if the trajectory of (72) is separable and determined by junctions, then the cost functional of (72) can be represented by junctions:

$$J(\tilde{x}_0, \tilde{x}_1, \cdots, \tilde{x}_{N+1}) := \sum_{1 \leq i \leq N, \ i \ \text{odd}} [J_0(\tilde{x}_{i-1}, \tilde{x}_i) + J_c(\tilde{x}_i, \tilde{x}_{i+1})].$$

Moreover, there is a hidden constraint. That is, the optimal trajectory in the free space connecting $\tilde{x}_i$ and $\tilde{x}_{i+1}$ must not violate the constraints. In other word, we require that for i even,

$$V(\tilde{x}_i, \tilde{x}_{i+1}) := \min_{t_i \leq t \leq t_{i+1}} \hat{\phi}(\gamma_0(\tilde{x}, \tilde{x}_{i+1})(t), t) = 0. \tag{74}$$

It ensures that the trajectory determined by junctions satisfies (72)'s constraints. Here we call $V(\tilde{x}_i, \tilde{x}_{i+1}) = 0$ as the visbility constraints.

For problems with separable structures, any optimal trajectories must be in H. As a result, in order to find an optimal trajectory, only the optimal junctions need to be computed. We gain a tremendous dimension reduction since the number of junctions is finite. In other words, problem (72) becomes

$$\min_{\tilde{x}_0, \cdots, \tilde{x}_{N+1}} J(\tilde{x}_0, \tilde{x}_1, \cdots, \tilde{x}_{N+1}), \tag{75}$$

subject to $V(\tilde{x}_i, \tilde{x}_{i+1}) = 0$, for i even.

Remark 10 *Here we require that (72) on free space or constraints can be solved easily, either by an analytical solution or other rapid numerical methods. Indeed, linear quadratic optimal control problems with proper constraints satisfies these criteria.*

We emphasize that solving optimization problem (75) has two challenges. One is that we intend to solve for the global minimizer. The other is that the dimension of optimization problem is unknown, since the number and index of junctions are unknown for the optimal trajectory. We apply the intermittent diffusion (ID) to conquer the first problem, and a new inserting and removing junctions from the system to treat the second challenge. They are presented in the next two subsections.

133

6.1.2 Intermittent Diffusion

ID is a global optimization strategy developed in [32]. It is to find minimizers of (75) by stochastic differential equations on a boundary manifold :

$$d\tilde{x} = P_{\tilde{x}}[-\nabla J(\tilde{x})d\theta + \sigma(\theta)dW(\theta)], \tag{76}$$

where $\tilde{x} = (\tilde{x}_0, \cdots, \tilde{x}_{N+1})$; $W(\theta)$ is the standard Brownian motion in $\mathbb{R}^n$; θ is an artificial time variable that is different from t, the time variable used in problem (72); $P_{\tilde{x}}$ is the orthogonal projection onto the tangent plane to $\tilde{x}$. In other words, if we denote a feasible direction set

$$\mathcal{F}(\tilde{x}) := \{\mathbf{q} \mid \nabla V(\tilde{x}_i, \tilde{x}_{i+1}) \cdot \mathbf{q} = 0, \text{ for } i \text{ even}, \ \|\mathbf{q}\| = 1\},$$

then $P_{\tilde{x}}(p)$, representing the projection of any vector p onto the feasible direction of $\mathcal{F}(\tilde{x})$, is defined by

$$-\frac{P_{\tilde{x}}(\mathbf{p})}{\|P_{\tilde{x}}(\mathbf{p})\|} := \arg\min_{\mathbf{q}\in\mathcal{F}(\tilde{x})} \mathbf{q}\cdot\mathbf{p}, \quad \|P_{\tilde{x}}(\mathbf{p})\| := \min_{\mathbf{q}\in\mathcal{F}(\tilde{x})} |\mathbf{q}\cdot\mathbf{p}|.$$

Here $\sigma(\theta)$ is a piecewise constant function, which is used to add the noise intermittently. More precisely, $\sigma(\theta) = \sum_{j=1}^{m} \sigma_j \chi_{[S_j, T_j]}(\theta)$, where $0 = S_1 < T_1 < S_2 < T_2 < \cdots < S_m < T_m < S_{m+1} = T$ and $\chi_{[S_j, T_j]}$ is the characteristic function of interval $[S_j, T_j]$. If $\sigma(\theta) \neq 0$ (76) is a well-defined SDE [56], whose solution has positive probability to escape the attraction of any local minimizers; If $\sigma(\theta) = 0$, we obtain the projected gradient flow, whose solution has the ability to visit a particular local minimizer. Here we denote

$$\nabla^c J(\tilde{x}) := P_{\tilde{x}}(\nabla J(\tilde{x})).$$

Remark 11 *Here the computation of $\nabla J(\tilde{x})$ and $\nabla V(\tilde{x})$ depends on (72) in free space or constraints. If $\gamma_0(\tilde{x}_i, \tilde{x}_{i+1})$, $\gamma_c(\tilde{x}_i, \tilde{x}_{i+1})$ have analytic solutions, e.g. example 2, then they can be found accordingly; If $\gamma_0(\tilde{x}_i, \tilde{x}_{i+1})$, $\gamma_c(\tilde{x}_i, \tilde{x}_{i+1})$ are not easy to obtain, e.g. example 1, they are approximated by finite differences.*

6.1.3 Handling Dimension Changes

To maintain separability, we introduce the following two operations in the process to add or remove junctions as needed.

Insert junctions. During the evolution of junctions according to (76), $\gamma_0(\tilde{x}_k, \tilde{x}_{k+1})$ may intersect with the interior of the constrained region. In other words, there may exist a time t such that $\hat{\phi}(t, \gamma_0(\tilde{x}_k, \tilde{x}_{k+1})(t)) < 0$. In order to maintain separability, we insert the intersection points into the set of junctions. Let $(\tilde{x}_0, \tilde{x}_1, \cdots, \tilde{x}_N, \tilde{x}_{N+1})$ be the set of junctions representing the current path and assume $\gamma_0(\tilde{x}_k, \tilde{x}_{k+1})$ intersects with $\hat{\phi} = 0$ at $\tilde{y}$ where $\hat{\phi}(\tilde{y}) = 0$ for the first time (without loss of generality, assume there is only one such intersection). We add $\tilde{y}$ as a new junction and the path becomes $(\tilde{x}_0, \cdots, \tilde{x}_k, \tilde{y}, \tilde{y}, \tilde{x}_{k+1}, \cdots, \tilde{x}_N, \tilde{x}_{N+1})$. It is easy to see that the cost of the new path remains the same

$$
J(\tilde{x}_0, \ldots, \tilde{x}_k, y, y, \tilde{x}_{k+1}, \cdots, \tilde{y}_{n+1}) - J(\tilde{x}_0, \ldots, \tilde{x}_{N+1})
$$
$$
= J_0(\tilde{x}_k, \tilde{y}) + J_c(\tilde{y}, \tilde{y}) + J_0(\tilde{y}, \tilde{x}_{k+1}) - J_0(\tilde{x}_k, \tilde{x}_{k+1})
$$
$$
= J_0(\tilde{x}_k, \tilde{y}) + J_0(\tilde{y}, \tilde{x}_{k+1}) - J_0(\tilde{x}_k, \tilde{x}_{k+1}) = 0.
$$

With the new set of junctions, we have another gradient flow for $\{\tilde{y}_k\}$ which is also expressed by (76). However, the number of equations is strictly larger. **Remove junctions.** Junctions need to be removed if doing so results in a path with smaller cost. This case happens when two junctions $\tilde{x}_k$ and $\tilde{x}_{k+1}$ on the boundary meet each other during the flow, i.e. $\tilde{x}_k = \tilde{x}_{k+1}$. By the triangle inequality, we have

$$
J_0(\tilde{x}_{k-1}, \tilde{x}_{k+2}) \leq J_0(\tilde{x}_{k-1}, \tilde{x}_k) + J_0(\tilde{x}_{k+1}, \tilde{x}_{k+2}),
$$

The original path $(\cdots, \tilde{x}_{k-1}, \tilde{x}_k, \tilde{x}_{k+1}, \tilde{x}_{k+2}, \cdots)$ can be shortened to obtained the path $(\cdots, \tilde{x}_{k-1}, \tilde{x}_{k+2}, \cdots)$. However, $\gamma_0(\tilde{x}_{k-1}, \tilde{x}_{k+2})$ may intersect with $\hat{\phi} = 0$. Hence, to maintain separability, as in the insertion case, we add the intersections into the set of junctions. It should be noted that unlike the process of adding junctions, removing junctions causes a jump in the gradient flow.

6.1.4 Algorithm

With all the components discussed above, we are ready to state our algorithm.

Method of Evolving Junctions

Input: Constraint ϕ and ψ,

starting and ending points x_0 and M,

running cost L, terminal time t_f, and ODE f,

number of intermittent diffusion intervals m.

Output: The optimal set γ_{opt} of junctions.

---

1. Initialization. Find the initial path $\gamma^{(0)} = (\tilde{x}_0, \cdots, \tilde{x}_{n+1})$;

2. Select duration of diffusion $\Delta T_l,\ l \le m$;

3. Select diffusion coefficients $\sigma_l,\ l \le m$;

4. **for** $l = 1 : m$

5. $\quad \gamma^{(l)} = \gamma^{(0)}$;

6. $\quad$ **for** $j = 1 : \Delta T_l$

7. $\qquad$ Find $\nabla^c J(\gamma^{(l)})$.

8. $\qquad$ Update $\gamma^{(l)}$ according to (76) with $\sigma(\theta) = \sigma_l$;

9. $\qquad$ Remove junctions from or add junctions to $\gamma^{(l)}$ when necessary;

10. $\quad$ **end**

11. $\quad$ **while** $\|\nabla^c J(\gamma^{(l)})\| > \epsilon$

12. $\qquad$ Update $\gamma^{(l)}$ according to (76) with $\sigma(\theta) = 0$;

13. $\quad$ **end**

14. **end**

15. Compare $J(\gamma^{(l)}), l \le m$ and set $\gamma_{opt} = \mathrm{argmin}_{l \le m} J(\gamma^{(l)})$;

Remark 12 *For step* 12, *problem* (72) *becomes a usual optimization problem with fixed dimensions, hence we can apply other efficient constrained optimization method to solve, such as Newton Method, quasi Newton Method and so on. An example in a shortest path problem has been studied in next section .*

6.2   Examples

In this section, we present two examples solved by MEJ.

6.2.1   Example 1: A standard linear quadratic control problem

Let us consider

$$\min_{x_1, x_2, u} \int_0^1 x_1(t)^2 + x_2(t)^2 + 0.005u(t)^2 dt,$$

subject to

$$\dot{x}_1(t) = x_2(t),$$

$$\dot{x}_2(t) = -x_2(t) + u(t), \quad t \in [0, 1],$$

$$x_1(0) = 0, \quad x_2(0) = -1,$$

with a state constraint

$$x_2(t) \le d(t), \quad t \in [0, 1],$$

and control constraints

$$-20 \le u(t) \le 20, \quad t \in [0, 1],$$

where

$$d(t) = 8(t - \frac{1}{2})^2 - \frac{1}{2}.$$

This is the linear quadratic problem considered in [63, 82].

First, through the following three steps, we reformulate this optimal control problem into a finite dimensional optimization problem following the idea of MEJ.

For convenience, we denote a junction $\tilde{x}_i = (t_i, x_1(t_i), x_2(t_i), u(t_i))$.

Step 1: The optimal trajectory in the free space can be solved analytically. Here free space means both state and control constraint are not active in a period (t_i, t_{i+1}). In this case, we just need to consider trajectory solving the following control problem:

$$J_0(\tilde{x}_i, \tilde{x}_{i+1}) = \min_{x,u} \int_{t_i}^{t_{i+1}} L(x_1(t), x_2(t), u(t))dt \tag{77}$$

subject to

$$\dot{x}_1(t) = x_2(t), \quad \dot{x}_2(t) = -x_2(t) + u(t), \quad t_i \leq t \leq t_{i+1},$$

and $x_1(t_i)$, $x_2(t_i)$, $x_2(t_{i+1})$ is fixed by $\tilde{x}_1$ and $\tilde{x}_2$. Here $x_1(t_{i+1})$ is not fixed since the state constraint is only for $x_2(t)$, while $x_1(t)$ is free.

We notice that (98) is an optimal control problem without a state constraint, and it can be solved by Pontryagin's maximum principle. Let us define Hamiltonian

$$H(\lambda, x_1, x_2, u) = L(x_1, x_2, u) + \lambda^T f(x, u),$$

where $x = (x_1, x_2)^T$, $\lambda = (\lambda_1, \lambda_2)^T$, $f(x, u) = (x_2, -x_2 + u)^T$. Then the optimality conditions become

$$\begin{cases} \frac{\partial H}{\partial x} = -\dot{\lambda} \\ \frac{\partial H}{\partial u} = 0 \\ \dot{x} = f(x, u) \end{cases} \tag{78}$$

with boundary condition $x_1(t_i)$, $x_2(t_i)$, $x_2(t_{i+1})$ are fixed by $\tilde{x}_i$, $\tilde{x}_{i+1}$ and $\lambda_1(t_{i+1}) = 0$. Moreover, since H is quadratic and f is linear, system (78) is reduced to a linear system of ODEs:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{u} \\ \dot{\lambda}_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 200 & 1 & 100 \\ -2 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ u \\ \lambda_1 \end{pmatrix}, \quad \begin{pmatrix} x_1(t_i) \\ x_2(t_i) \\ x_2(t_{i+1}) \\ \lambda_1(t_{i+1}) \end{pmatrix} = \begin{pmatrix} \tilde{x}_i|_{x_1} \\ \tilde{x}_i|_{x_2} \\ \tilde{x}_{i+1}|_{x_2} \\ 0 \end{pmatrix}. \tag{79}$$

Its solution $x^*(t), u^*(t)$ can be calculated analytically in term of $(\tilde{x}_i, \tilde{x}_{i+1})$:

$$(x^*(t), u^*(t), \lambda_1^*(t))^T = c_1 e^{10\sqrt{2}t} v_1 + c_2 e^{-10\sqrt{2}t} v_2 + c_3 e^t v_3 + c_4 e^{-t} v_4$$

where v_1, v_2, v_3, v_4 are eigenvectors corresponding to eigenvalues $10\sqrt{2}, -10\sqrt{2}, 1$ and -1 respectively, and c_1, c_2, c_3, c_4 are constants determined by $(\tilde{x}_i, \tilde{x}_{i+1})$. Hence, if we substitute $x^*(t), u^*(t)$ into the running cost, then

$$J_0(\tilde{x}_i, \tilde{x}_{i+1}) = \int_{t_i}^{t_{i+1}} L(x_1^*(t), x_2^*(t), u^*(t))dt \ ,$$

becomes a function of $\tilde{x}_i, \tilde{x}_{i+1}$.

Step 2: We consider the optimal trajectory with active constraints. In general, the active constraints can be divided into three types. In first case, both state and control constraints are active, i.e. $x_2(t) = d(t)$, $|u(t)| = 20$ for $t \in [t_i, t_{i+1}]$. In this example, this case is not possible since $|u(t)| = |\dot{d}(t) + d(t)| < 20$, for any $t \in [0, 1]$. In second case, the control constraint is active while the state constraint is not active, i.e. $u^*(t) = 20$ or -20 for $t \in [t_i, t_{i+1}]$. Since the control is known, the path is uniquely determined by ODE system

$$\dot{x}_1^*(t) = x_2^*(t), \quad \dot{x}_2^*(t) = -x_2^*(t) + u^*(t), \quad t_i \leq t \leq t_{i+1},$$

where are given junctions $\tilde{x}_i, \tilde{x}_{i+1}$. In third case, state constraint is active while control constraint is not active. I.e.

$$\dot{x}_1(t) = x_2(t), \quad \dot{x}_2(t) = -x_2(t) + u(t), \quad x_2(t) = d(t), \quad t_i \leq t \leq t_{i+1}.$$

From the state constraint

$$x_2^*(t) = d(t) = 8(t - \frac{1}{2})^2 - \frac{1}{2},$$

we can directly solve

$$u^*(t) = d(t) + \dot{d}(t) = 8(t - \frac{1}{2})^2 + 16(t - \frac{1}{2}) - \frac{1}{2}$$

and

$$
\begin{aligned}
x_1^*(t) &= x_1^*(t_i) + \int_{t_i}^{t} x_2^*(s)ds \\
&= x_1^*(t_i) + \frac{8}{3}[(t - \frac{1}{2})^3 - (t_i - \frac{1}{2})^3] - \frac{1}{2}(t - t_i)
\end{aligned}
$$

139

for $t \in (t_i, t_{i+1}]$.

For these cases, by substituting $x^*(t)$, $u^*(t)$ into the running cost, we obtain

$$J_c(\tilde{x}_i, \tilde{x}_{i+1}) = \int_{t_i}^{t_{i+1}} L(x_1^*(t), x_2^*(t), u^*(t))dt,$$

as a function of $\tilde{x}_i$ and $\tilde{x}_{i+1}$.

Step 3: The visibility function becomes

$$V(\tilde{x}_i, \tilde{x}_{i+1}) = \min_{t_i \leq t \leq t_{i+1}} d(t) - x_2^*(t),$$

where i is even and $x_2^*(t)$ is from (79). Moreover, we can obtain control constraints by junctions

$$|U(\tilde{x}_i, \tilde{x}_{i+1})| = \max_{t_i \leq t \leq t_{i+1}} |x_2^*(t) + \dot{x}_2^*(t)|.$$

Combining the above three steps, we achieve the finite dimensional optimization problem:

$$\min_{\tilde{x}_1, \cdots, \tilde{x}_n} \sum_{1 \leq i \leq n,\ i\ \text{odd}} [J_0(\tilde{x}_{i-1}, \tilde{x}_i) + J_c(\tilde{x}_i, \tilde{x}_{i+1})] \tag{80}$$

subject to

$$V(\tilde{x}_i, \tilde{x}_{i+1}) = 0, \quad i \text{ even}; \quad -20 \leq U(\tilde{x}_i, \tilde{x}_{i+1}) \leq 20 \quad \text{for all } i.$$

We give more details about the algorithm. Although there are four components in each junction $\tilde{x}_i = (t_i, x_1(t_i),\ x_2(t_i), u(t_i))$, the later three are functions of t_i. In other words, if we know t_i, we can compute the other three analytically. From the constraint, $x_2(t) = d(t)$, we know $x_2(t_i) = d(t_i)$. Substituting this to the equation $\dot{x}_2(t) = -x_2(t) + u(t)$, we obtain $u(t_i) = d(t_i) - \dot{d}(t_i)$. Then using $\dot{x}_1(t) = x_2(t)$, we obtain

$$x_1(t_i) = x_1(t_{i-1}) + \int_{t_{i-1}}^{t_i} d(t)dt$$

if i is even. When i is odd, equation (79) gives the value of $x_1(t_i)$ directly. Using this knowledge, we conclude if we know t_is, we know all the junctions. This enables us to write the optimization (80) as a problem depending only on t_i's.

140

Then we apply ID as given in (76) to this optimization problem:

$$dt_i = P_{\tilde{x}}(-\nabla_{t_i}\hat{J}(t_1,\cdots,t_n)d\theta + \sigma(\theta)dW(\theta)).$$

In our implementation, we evaluate the gradient by finite difference:

$$\frac{\partial \hat{J}(t_1,\cdots,t_n)}{\partial t_i} \approx \frac{\hat{J}(t_1,\cdots,t_i+h,\cdots,t_n) - \hat{J}(t_1,\cdots,t_i,\cdots,t_n)}{h}$$

where the step size $h = 10^{-9}$. While $\hat{V}(t_i,t_{i+1}) = \min_{t_i \le t \le t_{i+1}}(d(t) - x_2^*(t))$ is a one dimensional optimization problem, we compute it by the Newton's method. We stop the gradient flow at $\|\nabla^c \hat{J}(t_1,\cdots,t_n)\| \le 10^{-6}$.

Our experiment with $m = 4$ finds two minimizers. One of global minimizers is with objective function value 0.1721 as shown in Figure 24. The other shown in Figure 25 is a local minimizer with objective function value 0.1725. Both are smaller than the optimal objective value 0.1727 reported in [63, 82].

(a) $x_1(t)$

(b) $u(t)$

(c) $x_2(t)$. The red path is $d(t)$, the blue path is $x_2(t)$.
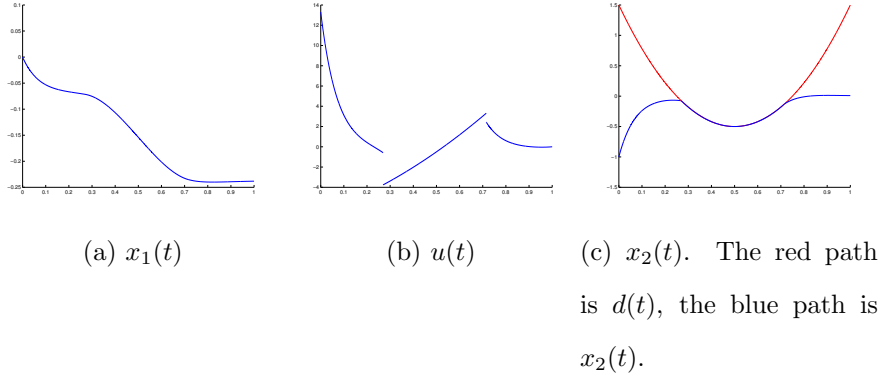
Figure 24: Linear quadric control: Global minimizer

141

(a) $x_1(t)$ (b) $u(t)$ (c) $x_2(t)$. The red path is $d(t)$, the blue path is $x_2(t)$.
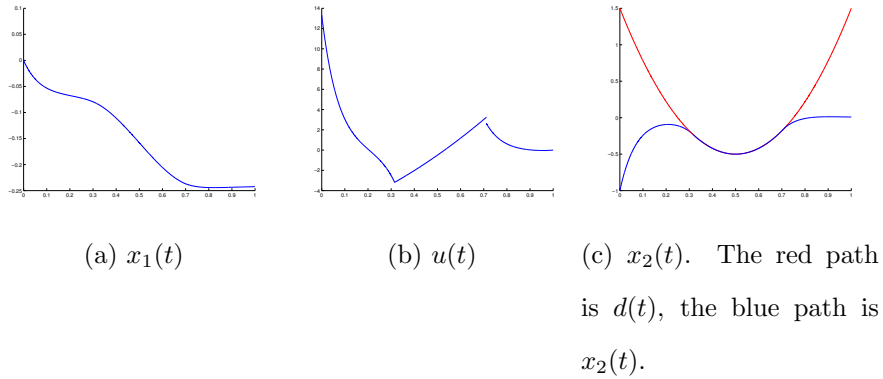
Figure 25: Linear quadric control I: Local minimizer

It is evident that the difference between the global and local minimizers is very small in terms of the objective function values, but the junctions in the control $u(t)$ are quite different and they are located at different positions. For example, the first junction for the global minimizer is around $t = 0.2658$, while the first junction for the local minimizer is around $t = 0.3156$. In addition, the control variable $u(t)$ at the first junction of the global minimizer is discontinuous, while its counterpart in the local minimizer is continuous in Figure 25.

Furthermore, in this example, there are two junctions, resulting a system of SDEs with 2 unknowns. This means that to compute the solutions, we only numerically solve an initial value problem for a system of SDEs with 2 equations by the simple Euler-Maruyama scheme. Compared the Direct methods, the dimension reduction in MEJ is significant.

We would like to point out that MEJ can be easily extended to handle more complicated situation for this type of linear quadratic optimal control problems. Here we give another example with the same objective functional, the same initial and terminal conditions, but replace the constraint by $d(t) = 8(t-\frac{1}{2})^2 - \frac{1}{2} + \frac{4}{5}\sin(5\pi t)$. As indicated in Figure 26, there are two humps in the constraints, leading to 4 junctions in MEJ. The dimension of the gradient flow is 4. We depict the identified a global

142

(a) $x_1(t)$　　(b) $u(t)$　　(c) $x_2(t)$. red path is $d(t)$, blue path is $x_2(t)$.
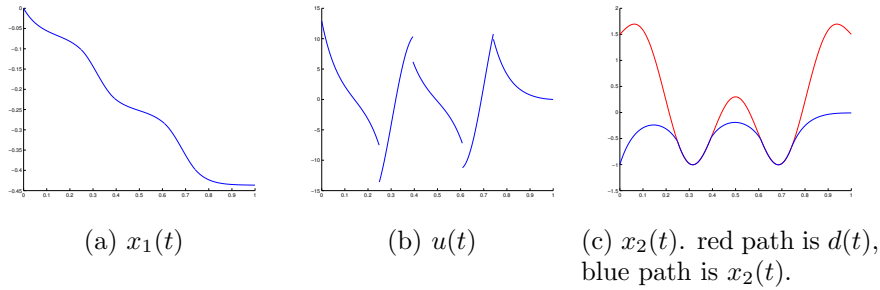
Figure 26: Global minimizer

optimal solution in Figure 26, and the objective functional values is 0.50825, where the tolerance is $\|\nabla^c J\| < 10^{-6}$. In contrast, the Direct method needs a special discretization to handle this situation, as reported in [82].

In summary, compared to the Direct methods in [63, 82], MEJ has the following advantages:

1. The dimension is changed fundamentally. All direct methods approximate, via discretization in t, the infinite dimensional Banach space by $\mathbb{R}^n$ with n large enough to meet the accuracy requirements, while our method leverage the separability, and only consider trajectories determined by a finite number of junctions.

2. MEJ can obtain the global minimizer and achieve desirable accuracy without suffering the restrictions on the discretization in t. This is not the case for the Direct methods. For instance, as we observed in Figures 24 and 25, the local minimizer is very similar to the global one. The Direct method needs to use very small discretization step size, meaning large n, to achieve enough accuracy to distinguish them. In MEJ, it only needs two junctions to compute both.

3. MEJ treats the constraints in a natural way, through the visibility function. The equations governing the evolution of the junctions are defined on the boundaries of the constraints. The Direct methods [63, 82] require all discretization points

satisfying the constraints and the dynamical systems.

6.2.2 Example 2: Path planning in dynamic environments

In this example, we use MEJ to find the minimal cost path in an environment where obstacles are moving obstacles. This example is motivated by practical problems in robotics, in which it is considered as a very challenging problem on its own. To the best of our knowledge, only a few studies have been devoted to the optimal solution for such a problem [47, 91]. Due to its complexity, it is not our intention to a give full description on how to apply MEJ to the general situation of this problem, rather, we refer readers to our recent study reported in [28] for the complete details.

In simple words, our goal is to find the optimal path for a robot in the plane moving from a starting point to a target location with minimal cost, such as fuel consumption, while avoiding collisions with several obstacles that also move in the environment. To be mathematically precise, we want to solve the following optimal control problem:

$$\min_{\gamma, v} \int_0^T L(t, \gamma, v) dt, \tag{81}$$

subject to

$$\dot{\gamma} = v, \quad t \in [0, T],$$

$$\gamma(0) = x, \quad \gamma(T) = y,$$

$$\gamma(t) \in \mathbb{R}_c^2(t),$$

$$||v(t)|| \leq v_m,$$

where $\gamma(t) \in \mathbb{R}^2$ is the parametrization of the robot moving path, v and t represent the velocity and time respectively, $\gamma(0), \gamma(T)$ are the initial and terminal points, constant v_m is the maximal speed that the robot can move, $|| \cdot ||$ is the 2-norm, and $\mathbb{R}_c^2(t)$ is the time dependent obstacle-free space defined as follows: Let $P_1(t), \cdots, P_N(t)$ be N time-dependent open subsets of $\mathbb{R}^2$ representing obstacles, which are moving at

constant speed,

$$P_k(t) = \{p + v_k t \mid p \in P_k\}, \quad v_k \text{ is a constant velocity of } P_k,$$

and $\mathbb{R}_c^2(t) = \mathbb{R}^2 \setminus (\cup_i^N P_i(t))$. We consider the Lagrangian

$$L(t, \gamma(t), \dot{\gamma}(t)) = \dot{\gamma}(t)^2 + c, \quad c \in \mathbb{R}^1 \text{ is a given constant.}$$

Hence the cost functional $J(\gamma) = \int_0^T \dot{\gamma}^2 dt + cT$ represents the robot's fuel consumption, meaning it costs more for fast speed while stalled (or slow) motion is also inefficient. It can be viewed as the weighted average of kinetic energy consumption and arrival time.

For convenience, we denote a junction $\tilde{x}_i = (t_i, x_i)$, where $x_i = \gamma(t_i)$.

Step 1: The optimal trajectory in the free space can be solved analytically.

Lemma 40 *The optimal path γ_i connecting the junction pair $(\tilde{x}_i, \tilde{x}_{i+1})$ with inactive constraint is a line with constant speed. I.e.*

$$\gamma_i(t) = \frac{x_{i+1} - x_i}{t_{i+1} - t_i}(t - t_i) + x_i.$$

Proof 35 *It is a classical problem in calculus of variation. Denote $L(t, x, v) = v^2 + c$, then the optimal path satisfies the Euler-Lagrange equation*

$$\nabla_x L(t, \gamma, \dot{\gamma}) - \frac{d}{dt} \nabla_v L(t, \gamma, \dot{\gamma}) = 0 \Rightarrow -2\frac{d}{dt}(\dot{\gamma}) = 2\ddot{\gamma} = 0,$$

meaning that the optimal solution is with zero acceleration.

If we substitute such optimal path into the running cost, we obtain

$$J_0(\tilde{x}_i, \tilde{x}_{i+1}) = \frac{(x_{i+1} - x_i)^2}{t_{i+1} - t_i} + c(t_{i+1} - t_i),$$

as a function of $\tilde{x}_i$, $\tilde{x}_{i+1}$.

Step 2: We consider the optimal trajectory with active constraints.

Lemma 41 *The optimal path γ_i connecting a junction pair $(\tilde{x}_i, \tilde{x}_{i+1})$ with active constraint is one of follows:*

(a) It is a line with maximal speed;

(b) It is a geodesic on moving obstacle with relative constant speed;

(c) It is a geodesic on moving obstacle with maximal speed.

Proof 36 *It contains three cases:*

(a) The speed constraint is active while the path constraint is not. I.e.

$$\|\dot{\gamma}(t)\| = v_m;$$

(b) The path constraint is active while the speed constraint is not. I.e. there exists an obstacle P_k, such that

$$\gamma(t) \in \partial P_k(t);$$

(c) Both path and speed constraints are active. I.e. there exists an obstacle P_k, such that

$$\gamma(t) \in \partial P_k(t) \ and \ \|\dot{\gamma}(t)\| = v_m.$$

Let's illustrate three cases separately. For case (a), it is not hard to show that the optimal path is a line with maximal speed $\|\dot{\gamma}\| = v_m$.

For case (b), the control problem forms

$$\min\{ \int_{t_i}^{t_{i+1}} (\dot{\gamma}^2(t) + c)dt \mid \gamma(t_i) = x_i, \ \gamma(t_{i+1}) = x_{i+1}, \ \phi_k(t, \gamma(t)) = 0\}. \tag{82}$$

Let's solve it explicitly. We parametrize the boundary of obstacle P_k by $\alpha(u)$, where $u \in [0, l_k]$ is an arc-length parameter and l_k is the obstacle's perimeter. Hence $\gamma(t)$ is represented by its relative position $u(t)$ on the obstacle

$$\gamma(t) = \alpha(u(t)) + v_k \cdot t.$$

Then (82) is equivalent to a new form

$$\min\left\{\int_{t_i}^{t_{i+1}} L_1(t, u, \dot{u})dt \mid u(t_i) = u_i, \ u(t_{i+1}) = u_{i+1}\right\},$$

where

$$L_1(t, u, \dot{u}) = \dot{u}(t)^2 + 2(\alpha_u(u(t)) \cdot v_k)\dot{u}(t) + v_k^2 + c, \tag{83}$$

and $x_i = \alpha(u_i) + v_k \cdot t_i$, $x_{i+1} = \alpha(u_{i+1}) + v_k \cdot t_{i+1}$.

Again, it is a standard calculus of variation problem. Notice that

$$\frac{\partial}{\partial u}L_1 = 2(\alpha_{uu} \cdot v_k)\dot{u}, \quad \frac{\partial}{\partial \dot{u}}L_1 = 2\dot{u} + 2(\alpha_u \cdot v_k),$$

and $\frac{d}{dt}L_{1\dot{u}} = 2\ddot{u} + 2(\alpha_{uu} \cdot v_k)\dot{u}$. Then from the Euler-Lagrange equation

$$\frac{\partial}{\partial u}L_1(t, u, \dot{u}) - \frac{d}{dt}\frac{\partial}{\partial \dot{u}}L_1(t, u, \dot{u}) = 0 \Rightarrow \ddot{u} = 0.$$

Hence the optimal path is with a relative constant speed.

Hence we can obtain γ_i. Since the obstacle is connected set in $\mathbb{R}^2$, there are two local minimizer paths, which is from either clockwise $u_+(t)$ direction or counterclockwise direction $u_-(t)$:

$$u_+(t) = u_i + \frac{u_{i+1} - u_i}{t_{i+1} - t_i}(t - t_i); \quad u_-(t) = u_i - l_k + \frac{(u_{i+1} - u_i + l_k)}{t_{i+1} - t_i}(t - t_i). \tag{84}$$

If we denote $\gamma_+(t)$, $\gamma_-(t)$ by $u_+(t)$, $u_-(t)$, the optimal path γ_i satisfies

$$\gamma_i(t) = \arg\min_{\gamma_+, \gamma_-}\{J(\gamma_+), J(\gamma_-)\}.$$

For case (c), the robot's path is uniquely determined by both path and speed constraints:

$$\phi_k(t, \gamma(t)) = 0, \quad \|\dot{\gamma}(t)\| = v_m.$$

Similarly in case (b), by the arc-length parametrization,

$$\dot{u}^2 + 2\alpha_u(u) \cdot v_k \dot{u} + v_k^2 = v_m^2.$$

147

There are two solutions depending on clockwise or counter-clockwise direction

$$\dot{u}_+ = -\alpha_u(u_+) \cdot v_k + \sqrt{\alpha_u(u_+)^2 + (v_m^2 - v_k^2)};$$

$$\dot{u}_- = -\alpha_u(u_-) \cdot v_k - \sqrt{\alpha_u(u_-)^2 + (v_m^2 - v_k^2)}.$$

Similarly as in case (b), we find

$$\gamma_i(t) = \arg \min_{\gamma_+,\gamma_-} \{J(\gamma_+), J(\gamma_-)\}.$$

If we substitute the optimal path into running cost, we obtain $J_c(\tilde{x}_i, \tilde{x}_{i+1})$ as a function of $\tilde{x}_i$ and $\tilde{x}_{i+1}$.

Step 3: We express the constraints by junctions.

Given two junctions $\tilde{x}_i$ and $\tilde{x}_{i+1}$, the visibility function $V(\tilde{x}_i, \tilde{x}_{i+1})$ determines whether the line connecting them with a constant velocity only intersects the moving obstacles at junctions. A point $\gamma(t)$ on the line, outside the obstacle P_k is according to:

$$\phi_k(\gamma(t), t) \geq 0,$$

where ϕ_k is the sign distance function,

$$\phi_k(t, y) = \begin{cases} \text{dist}(y, \partial P_k(t)), & \text{if } y \in P_k(t) ; \\ -\text{dist}(y, \partial P_k(t)), & \text{if } y \in \mathbb{R}^2 \setminus P_k(t), \end{cases}$$

with $\text{dist}(y, \partial P_k(t)) = \inf_{x \in \partial P_k(t)} \|x - y\|$.

Then the visibility function becomes,

$$V(\tilde{x}_i, \tilde{x}_{i+1}) = \min_{t_i \leq t \leq t_{i+1}} \phi_k(\gamma(t), t) = 0,$$

and the minimizer can only be achieved at the junction points $\tilde{x}_i$ or $\tilde{x}_{i+1}$.

Moreover, the control constraints can also be determined by junctions.

$$U(\tilde{x}_i, \tilde{x}_{i+1}) = \max_{t_i \leq t \leq t_{i+1}} \|\dot{\gamma}(t)\| \leq v_m, \quad \text{for all } i.$$

Combining all steps, the optimal control has been transformed to the following finite dimensional optimization problem:

$$\min_{\tilde{x}_1,\ldots,\tilde{x}_n} \sum_{1 \leq i \leq n,\ i \text{ odd}} [J_0(\tilde{x}_{i-1}, \tilde{x}_i) + J_c(\tilde{x}_i, \tilde{x}_{i+1})]$$

subject to

$$V(\tilde{x}_i, \tilde{x}_{i+1}) = 0, \quad \text{i even;} \quad U(\tilde{x}_i, \tilde{x}_{i+1}) \leq v_m, \quad \text{all } i.$$

We give more details about the algorithm by considering two cases.

Fixed terminal time At the beginning, let the terminal time be fixed at $T = 1$ and the running cost be $L = \dot{\gamma}^2$. The starting and ending points are $X = (-2, 0.5)$, $Y = (20, 0.5)$. The obstacles are all disks, with centers $(0, 0)$, $(4.5, 3)$, $(8, -3)$, $(10, 4)$, $(12, -3)$, $(15, -4)$ and radiuses 1, 1, 1, 1.2, 1, 1. They all move at constant velocities, which are $(3, 5)$, $(-2, -5)$, $(-2, 4.5)$, $(0, -5.5)$, $(1, 5.5)$ and $(1, 5.5)$.

By letting $m = 6$ in MEJ, the algorithm finds two minimizers. One is a global minimizer with cost 510.353, whose trajectory passes four moving obstacles, see Figure 27 or movie in https://youtu.be/ziq0GQZGVeE.

The other is a local minimizer with cost 535.273, whose trajectory passes five moving obstacles, see movie in https://youtu.be/AO3Cy5J1-Rg. Here we want to emphasize the speed of this algorithm. By doing simulations on a 2013 Macbook Air with CPU core i5, 1.8G HZ, RAM 4GB, the average time of finding one local minimizer is around 20 seconds.

Undetermined terminal time Secondly, we show that MEJ can work with a unknown terminal time T, which is also a variable in (72). Let the running cost be $L = \dot{\gamma}^2 + 200$. The starting and ending points are $X = (-2, 0.5)$, $Y = (10, 0.5)$. There are two obstacles, which are disks with centers $(0, 0)$, $(6.5, 3)$ and radiuses $1, 1$. They move at constant velocities $(5, 0)$, $(-5, 0)$. The local minimizer is with cost 353.16 and terminal time 0.895, see Figure 28 or movie in https://youtu.be/KDKLCW1bFYw.

(a)   (b)

(c)   (d)

(e)   (f)

(g)   (h)

(i)   (j)
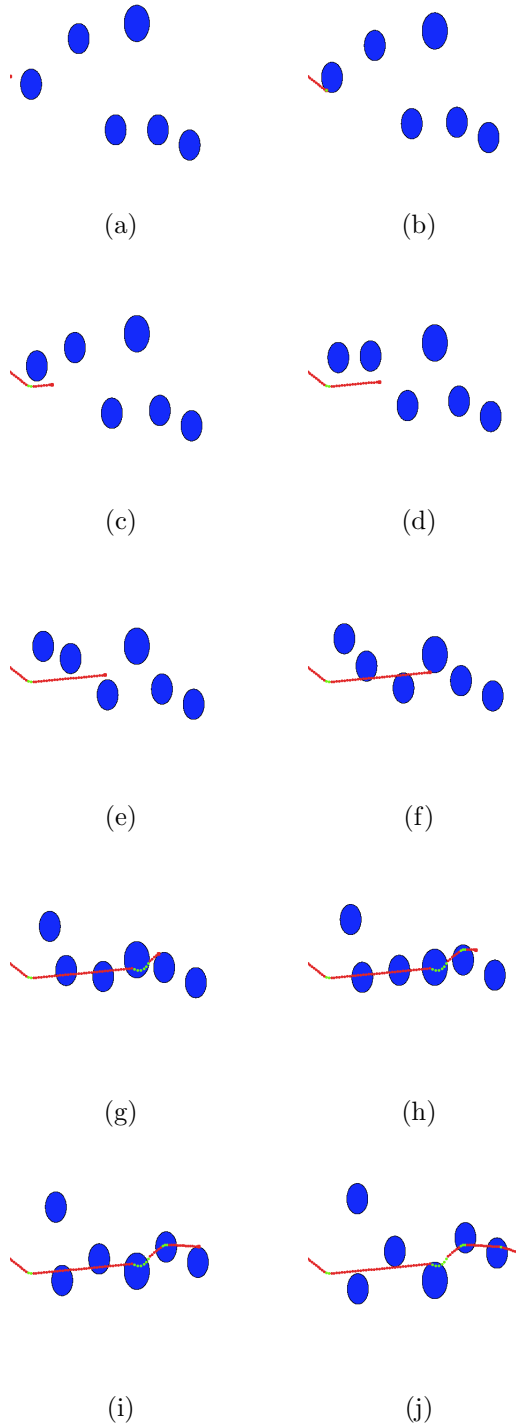
Figure 27: Optimal path in dynamic environments: Fixed terminal time: This is a snapshot of the global optimal path (red) for the drone while avoiding collisions with 6 moving obstacles (blue). The green part of the path indicates that the path travels along the moving obstacle boundary.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)
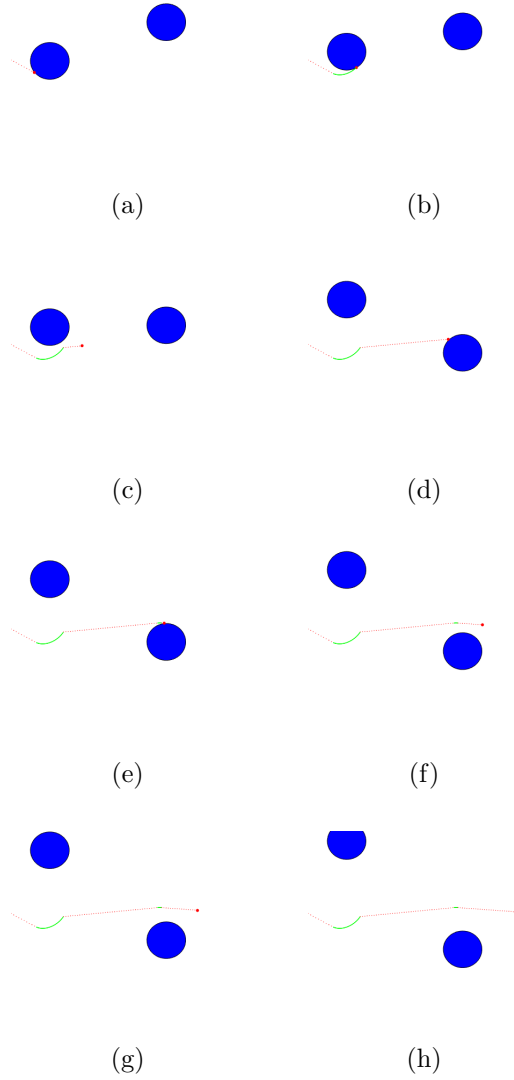
Figure 28: Optimal path in dynamic environments: Undetermined terminal time: This is a snapshot of the global optimal path (red) for the drone while avoiding collisions with 6 moving obstacles (blue). The green part of the path indicates that the path travels along the moving obstacle boundary.

Remark 13 *There are some additional interesting observations from MEJ. The optimal path in step 2 case (b) can be easily derived from geometry viewpoint. If we denote the robot's path by its relative position γ_r on ∂P_k, then*

$$\gamma(t) = \gamma_r(t) + \int_{t_i}^{t_{i+1}} v_k(t)dt, \quad t \in [t_i, t_{i+1}], \quad \gamma_r(t) \in \partial P_k.$$

Hence the running cost becomes

$$L(t, \gamma, \dot{\gamma}) := L_r(\dot{\gamma}_r) = [(\dot{\gamma}_r(t) + v_k)]^2 + c.$$

The the orignal optimal control problem forms

$$\inf\{\int_{t_i}^{t_{i+1}} L_r(\dot{\gamma}_r)dt \; : \; \gamma_r(t_i) = x_i, \; \gamma_r(t_{i+1}) + \int_{t_i}^{t_{i+1}} v_k dt = x_{i+1}\}.$$

By the Euler-Lagrange equation in geometry,

$$\frac{d}{dt}\nabla_{\dot{\gamma}_r} L_r(\dot{\gamma}_r) = \nabla_{\gamma_r} L_r(\dot{\gamma}_r) \Rightarrow 2\frac{D}{dt}\dot{\gamma}_r + \frac{D}{dt}v_k = 0.$$

Notice carefully that the left-hand side of the equation involves the time-derivative of a curve which is valued in tangent space. Hence $\frac{D}{dt}$ is a covariant derivative along the curve γ in ∂P_k. So if v_k is a constant, we have $\frac{D}{dt}\dot{\gamma}_r = 0$. It means that the relative path $\gamma_r(t)$ is a constant speed geodesic.

6.3   Acceleration technique

In this section, we further improve the MEJ presented in previous section. In short, we use an approximated Newton method to replace the gradient flow in MEJ, while retaining the overall MEJ framework including the SDEs to help the solution jump out of the traps of local minimizers. Such a replacement significantly reduces the computational time in finding the local minimizers.

In particular, we use a shortest path problem to illustrate the idea. Mathematically, the problem can be formulated as following. Let (X, d) be a length space, such as $\mathbb{R}^2$, where d is the distance defined on X, and $P_1, \cdots, P_N$ be N open

subsets of X, representing obstacles with boundaries $\{\partial P_k\}_{k=1}^{N}$. Given two points $x, y \in X_c = X \setminus \cup_{i=1}^{N} P_i$, we define the admissible set of paths connecting x and y to be the curves that have no intersection with all the interior of obstacles, i.e.

$$\mathcal{A}(x, y, X_c) = \{\, \gamma : [0, 1] \to X \mid \gamma(0) = x, \gamma(1) = y, \gamma \in X_c \},$$

where γ is absolutely continuous. For each admissible path, its length in Euclidean space is

$$J(\gamma(\theta)) = \int_0^1 |\dot{\gamma}(\theta)| \, \mathrm{d}\theta.$$

Then finding the shortest path can be posed as an optimization problem:

$$\gamma^* = \operatorname{argmin}_{\gamma \in \mathcal{A}(x,y,X_c)} J(\gamma). \tag{85}$$

Let's review MEJ. We start with a geometric structure, called separable, possessed by all shortest paths.

Definition 42 *A path* $\gamma \colon [0, 1] \to X_c$ *is* separable *if there exists a finite number of points* $\{x_1, x_2, \cdots, x_n\}$ *with* $x_i \in \partial P_{k_i}$, $k_i \leq N$, *such that* γ *concatenates line segments and partial curves on the boundaries of the obstacles, i.e.*

$$\gamma = \gamma_0(x, x_1) \cdot \gamma_c(x_1, x_2) \cdot \gamma_0(x_2, x_3) \cdot \gamma_c(x_3, x_4) \cdots \gamma_0(x_n, y), \tag{86}$$

where $\gamma_0(x_{i-1}, x_i)$ *is the line segment connecting* x_{i-1} *and* x_i *and* $\gamma_c(x_{i-1}, x_i)$ *is the geodesic on the boundary* ∂P_{k_i} *between the two points.*

A simple example is shown in Figure 29, and we call x_i a junction,

Theorem 43 *Let* ∂P_k *be a finite combination of convex and concave curves (surfaces). Then* γ^* *is separable. Moreover, each line segment* $\overline{x_{i-1}x_i}$ *is tangent to the obstacle* ∂P_{k_i}.
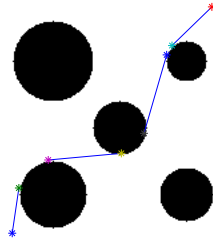
Figure 29: Each connecting point between a line segment and a boundary of the obstacles is a junction.

Figure 30: Each junction on a boundary is connected to the points before and after it by a straight line segment and an arc of the boundary.

Therefore the length of the shortest path is a function depending on the junctions $\{x_1, \ldots, x_n\}$,

$$J(x_1, \ldots, x_n) = \sum_{i=1}^{n} J(x_{i-1}, x_i),$$

where $J(x_{i-1}, x_i)$ represents the distance connecting (x_{i-1}, x_i):

$$J(x_{i-1}, x_i) = \begin{cases} \|x_{i-1} - x_i\|, & \text{if i is odd;} \\ \text{dist}_c(x_{i-1}, x_i), & \text{if i is even,} \end{cases}$$

in which $\|\cdot\|$ is the Euclidean norm.

Based on this theorem, MEJ restricts the search space to the set of all admissible paths with separable structures, a finite dimensional subset of $\mathcal{A}(x, y, X_c)$. More precisely, MEJ finds the shortest path by solving the following optimization problem,

$$\min_{x_1, \cdots, x_n} J(x_1, \cdots, x_n). \tag{87}$$

154

To find the global solution of (87), MEJ uses the intermittent diffusion (ID), a SDE based global optimization method developed in [32]. More precisely, it solves

$$d\hat{x} = -\nabla J(\hat{x})dt + \sigma(t)dW_t, \tag{88}$$

where $\hat{x} = \{x_1, \cdots, x_n\}$ represents the junctions, W_t the standard Brownian motion in $\mathbb{R}^n$, and $\sigma(t)$ a piecewise constant function

$$\sigma(t) = \sum_{j=1}^{m} \sigma_j \chi_{[S_j, T_j]}(t), \tag{89}$$

with $0 = S_1 < T_1 < \cdots < S_m < T_m < S_{m+1} = T$ and $\chi_{[S_j, T_j]}$ being the characteristic function of interval $[S_j, T_j]$.

If $\sigma(t) = 0$, the equation (88) becomes a gradient descent flow which converges to a local minimizer; if $\sigma(t) > 0$, the path has a certain (positive) probability, controlled by $\sigma(t)$, to jump out of the local traps, and therefore to reach the global solution.

6.3.1 The Newton-like algorithm

In this subsection, we present a new approximate Newton method in $\mathbb{R}^2$ shortest path problem, by finding the line segments tangent to the obstacles directly, to replace the gradient flow in (88) when $\sigma(t) = 0$.

In order to explain our method more clearly, we introduce an arc-length parameter θ to represent junctions. Let $x_i = x(\theta_i)$, $x_i^s = x(\theta_i^s)$, $x_i^c = x(\theta_i^c)$, where θ_i, θ_i^s, θ_i^c are arc-length parameters on the corresponding boundaries, and super index s indicates the junction connected to x_i by a straight line, c denotes the junction connected to x_i by a boundary arc, see Figure 31 for an illustration. With these notations, the length of the curve containing one straight line segment and the boundary arc $\gamma_0(x_i^s, x_i) \cdot \gamma_c(x_i, x_i^c)$ becomes

$$J_i(\theta) = \|x(\theta_i) - x(\theta_i^s)\| + d(\theta_i, \theta_i^c),$$

where $d(\theta_i, \theta_i^c) = \min\{d^+(\theta_i, \theta_i^c), d^-(\theta_i, \theta_i^c)\}$, with d^+, d^- representing the counterclockwise and clockwise distance on the obstacle boundary between $x(\theta_i^c)$ and $x(\theta_i)$

155

as illustrated in Figure 31.

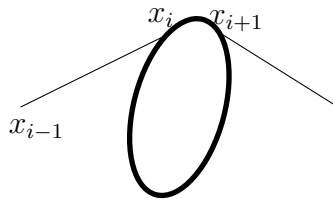(a)   (b)
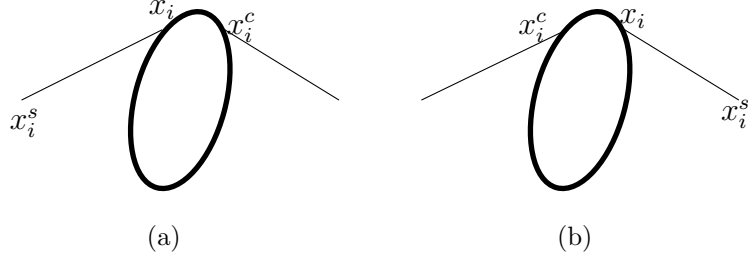
Figure 31: Two different scenarios for each junction on a boundary that is connected to the points before and after it by a straight line segment and an arc of the boundary.

Thus, the optimization problem (87) becomes

$$\min_{\theta_1,\cdots,\theta_n} J(\theta) = \frac{1}{2} \sum_{i=1}^{n} J_i(\theta),$$

where $\theta = (\theta_1, \cdots, \theta_n)$. And the intermittent diffusion (88) is

$$d\theta = -\nabla J(\theta)dt + \sigma(t)dW_t, \tag{90}$$

in which we have

$$\frac{\partial J}{\partial \theta_i} = \frac{x(\theta_i) - x(\theta_i^s)}{\|x(\theta_i) - x(\theta_i^s)\|} \cdot \dot{x}(\theta_i) + \text{sign}(d^+(\theta_i, \theta_i^c) - d^-(\theta_i, \theta_i^c)),$$

where $\dot{x}(\theta_i) = \frac{dx(\theta_i)}{d\theta_i}$.

Our main idea of this section is that instead of using the gradient flow to find local minimizer of (87), we apply the Newton method to solve $\nabla J(\theta) = 0$ directly. And this is equivalent to solving the tangent condition in Theorem 43 as stated in the next theorem. Here we denote $J_i^{(k)}(\theta) = \frac{\partial^k J_i}{\partial \theta_i^k}(\theta)$, $k = 2, 3$.

Theorem 44 *If θ^* is the local minimizer of (87), then the following statements are equivalent:*

(i)   *The line segment is tangent to the obstacle;*

156

(ii)   The second order derivative $J_i^{(2)}(\theta^*) = 0$, for $i = 1, \ldots, n$.

Moreover, the third order derivative satisfies

$$|J_i^{(3)}(\theta^*)| = |\kappa(\theta_i^*)|^2,$$

and

$$\frac{\partial J_i^{(2)}}{\partial \theta_j}(\theta^*) = 0,$$

for any $i, j = 1, \ldots, n$ and $i \neq j$.

Proof 37 *First, we show that solving $\nabla J(\theta) = 0$ implies the line connecting the junctions being tangent to the obstacles. Since*

$$\frac{\partial J}{\partial \theta_i} = g(\theta_i, \theta_i^s) + sign(d^+(\theta_i, \theta_i^c) - d^-(\theta_i, \theta_i^c)) = 0, \tag{91}$$

where

$$g(\theta_i, \theta_i^s) = \frac{x(\theta_i) - x(\theta_i^s)}{\|x(\theta_i) - x(\theta_i^s)\|} \cdot \dot{x}(\theta_i). \tag{92}$$

Hence $\nabla J(\theta) = 0$ is to solve $g(\theta_i, \theta_i^s)^2 = 1$ for each i. Moreover, since θ is the arc length parameter, $\dot{x}(\theta_i)$ is a unit vector, which implies that $g(\theta_i, \theta_i^s)$ is the inner product of two unit vectors. Then $g(\theta_i, \theta_i^s)^2 = 1$ means $x(\theta_i) - x(\theta_i^s)$ is parallel to tangent vector $\dot{x}(\theta_i)$, which implies tangent property.

To show the equivalence of (i) and (ii), we need to prove that (i) implies (ii): Without loss of generality, let us assume $sign(d^+(\theta_i, \theta_i^c) - d^-(\theta_i, \theta_i^c)) = -1$. Since $g(\theta_i, \theta_i^s) \leq 1$, solving (91) is equivalent to finding the maximizer θ^* of

$$\max_{\theta_i, \theta_i^s} g(\theta_i, \theta_i^s).$$

then it must satisfy

$$g_{\theta_i}(\theta_i^*, \theta_i^{s*}) = 0.$$

Since $J_i^{(2)}(\theta) = g_{\theta_i}(\theta_i, \theta_i^s)$, we have $J_i^{(2)}(\theta^*) = 0$.

(ii) implies (i): A direct computation gives the second order derivative of J:

$$J_i^{(2)}(\theta) = \frac{1 - g(\theta_i, \theta_i^s)^2 + (x(\theta_i) - x(\theta_i^s)) \cdot \ddot{x}(\theta_i)}{\|x(\theta_i) - x(\theta_i^s)\|}. \tag{93}$$

If $J_i^{(2)}(\theta^*) = 0$ for all $i = 1, \ldots, n$, we have

$$1 - g(\theta_i^*, \theta_i^{s*})^2 + (x(\theta_i^*) - x(\theta_i^{s*})) \cdot \ddot{x}(\theta_i^*) = 0.$$

Notice that $1 - g(\theta_i^*, \theta_i^{s*})^2 \geq 0$. Moreover, since $\overline{x(\theta_i^*)x(\theta_i^{s*})}$ first intersects obstacle P_{k_i} at point $x(\theta_i^*)$, which implies angle between vector $x(\theta_i^*) - x(\theta_i^{s*})$ and $\ddot{x}(\theta_i^*)$ is not larger than $\frac{\pi}{2}$, $(x(\theta_i^*) - x(\theta_i^{s*})) \cdot \ddot{x}(\theta_i^*) \geq 0$. Hence the solution satisfies $g(\theta_i^*, \theta_i^s)^2 = 1$, which implies the tangent property.
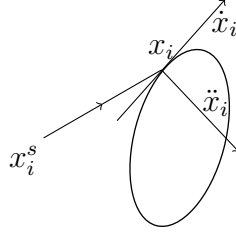
The third derivative is

$$J_i^{(3)}(\theta) = (\frac{1}{\|x(\theta_i) - x(\theta_i^s)\|})^{(1)} \cdot J_i^{(2)}(\theta) +$$
$$\frac{1}{\|x(\theta_i) - x(\theta_i^s)\|} \cdot [2\dot{x}(\theta_i) \cdot \ddot{x}(\theta_i) - 2g(\theta_i) \cdot J_i^{(2)}(\theta)$$
$$+ (x(\theta_i) - x(\theta_i^s)) \cdot \dddot{x}(\theta_i)].$$

Considering that θ_i is an arc-length parameter, we have $\dot{x}(\theta_i) \cdot \ddot{x}(\theta_i) = 0$. Combining it with $J_i^{(2)}(\theta^*) = 0$, we can show

$$J_i^{(3)}(\theta^*) = \frac{(x(\theta_i^*) - x(\theta_i^{*s})) \cdot \dddot{x}(\theta_i^*)}{\|x(\theta_i^*) - x(\theta_i^{*s})\|}.$$

By the tangent property, (37) can be formulated as

$$|J_i^{(3)}(\theta^*)| = |\dot{x}(\theta_i^*) \cdot \dddot{x}(\theta_i^*)|.$$

Since

$$F(\theta_i) = \int_0^{\theta_i} \dot{x}(u) \dddot{x}(u) du$$

$$= \dot{x}(0) \ddot{x}(0) - \int_0^{\theta_i} \ddot{x}(u)^2 du,$$

and $|\kappa(u)| = |\ddot{x}(u)|$, *then*

$$\dot{x}(\theta_i) \cdot \dddot{x}(\theta_i) = \frac{dF(\theta_i)}{d\theta_i} = -\kappa^2(\theta_i),$$

which implies $|J_i^{(3)}(\theta^*)| = \kappa^2(\theta_i^*)$.

In the end, we show that $\frac{\partial J_i^{(2)}}{\partial \theta_j}(\theta^*) = 0$ *for* $j \neq i$. *Since* $J_i^{(2)}(\theta)$ *depends on* θ_i^s *and* θ_i, *we only need to show* $\frac{\partial J_i^{(2)}}{\partial \theta_i^s}(\theta^*) = 0$. *By direct computations, we have*

$$\frac{\partial J_i^{(2)}}{\partial \theta_i^s}(\theta^*) = -\frac{2 g_{\theta_i^s}(\theta_i^*, \theta_i^{s*}) g(\theta_i^*, \theta_i^{s*})}{\|x(\theta_i^*) - x(\theta_i^{s*})\|}.$$

And since

$$g_{\theta_i^s}(\theta_i^*, \theta_i^{s*}) = 0,$$

then $\frac{\partial J^{(2)}}{\partial \theta_i^s}(\theta^*) = 0$, *which finishes the proof.*

Now, we are ready to present our method. We want to solve the tangency condition $\nabla J(\theta) = 0$ directly through the Newton method. By Theorem 44, it can be found that $\nabla J(\theta) = 0$ is a degenerate system, i.e. its Jacobian matrix becomes 0 at θ^*. Hence we can solve the system

$$J^{(2)}(\theta) = (J_1^{(2)}(\theta), \cdots, J_n^{(2)}(\theta)) = 0,$$

instead. We use an approximate Jacobian matrix of $J^{(2)}(\theta)$ given by

$$H(\theta) = \text{diag}(J_i^{(3)}(\theta)).$$

And it leads to the following iterations,

$$\theta^{k+1} = \theta^k - H^{-1}(\theta^k) J^{(2)}(\theta^k). \tag{94}$$

We must point out that in this iterative scheme, we consider obstacles with non-zero curvature boundary. Hence, H is an invertible diagonal matrix so that Newton method is valid. In fact, when boundaries are straight lines or curves with curvature close to 0, we can simply adjust the construction of $H(\theta)$ to continue the Newton step. For example, let $H(\theta) = \mathrm{diag}(J_i^{(3)}(\theta) + \lambda_i I)$, where λ_i is a selected scalar.

Remark 14 *The reason for being an "approximate" Newton method is that, only at the minimizer θ^*, the Jacobian matrix of $J^{(2)}(\theta)$ is exactly as $H(\theta)$. Otherwise, $H(\theta)$ is an approximation to the Jacobian.*

Remark 15 *In our implementation, we choose the parametrization direction, either clockwise or counter-clockwise, according to the initial condition θ^0. For instance, if $sign(d^+(\theta_i^0, \theta_i^{0c}) - d^-(\theta_i^0, \theta_i^{0c})) = 1$, we parametrize P_{k_i} clockwisely.*

Moreover, we prove the super-linear convergence rate of Newton-like algorithm by the following theorem. Here we denote $DJ^{(2)}(\theta)$ as the Jacobian matrix of $J^{(2)}(\theta)$.

Theorem 45 *Let $J^{(2)}(\theta): \mathbb{R}^n \to \mathbb{R}^n$ be smooth, and there is no zero-curvature point for all obstacles in $\mathbb{R}^2$, then there exists $\epsilon < 0$, such that if the iteration (94) starts at $\|\theta^0 - \theta^*\| < \epsilon$, θ^k converges to θ^* superlinearly.*

Proof 38 *We prove the theorem by two steps. Firstly, we use the fixed point theorem to show that there exists a sufficient small ϵ, such that θ^k converges to θ^*. In other words, consider a map $l: \mathbb{R}^n \to \mathbb{R}^n$,*

$$l(\theta) = \theta - H^{-1}(\theta)J^{(2)}(\theta).$$

We need to find a small neighbor of θ^*, such that $\sup_{\theta_B(\theta^*, \epsilon)} \|Dl(\theta)\| < 1$. To show this, we directly calculate

$$Dl(\theta) = I - H^{-1}(\theta)DJ^{(2)}(\theta) + H^{-2}(\theta)H'(\theta)J^{(2)}(\theta).$$

160

Substitute $H(\theta^*) = DJ^{(2)}(\theta^*)$ and $J^{(2)}(\theta^*) = 0$ into above equation, we have $Dl(\theta^*) = 0$. By the continuity of $Dl(\theta)$, we can show $\sup_{\theta_B(\theta^*,\epsilon)} \|Dl(\theta)\| < 1$. Hence we prove the convergence result.

Secondly, we show that the convergence rate is superlinear. To show this, let $e_k = \theta^* - \theta^k$, we need to show $\lim_{k\to\infty} \|e_{k+1}\|/\|e_k\| = 0$. Since θ_k converges to θ^*, we only consider the bounded region $B(\theta^*, \epsilon)$. On one hand, by the Taylor expansion of $J^2(\theta)$

$$0 = J^{(2)}(\theta^*) = J^{(2)}(\theta^k + e_k) = J^{(2)}(\theta^k) + DJ^{(2)}(\theta^k)e_k + O(\|e_k\|^2).$$

Hence

$$DJ^{(2)}(\theta^k)^{-1}J^{(2)}(\theta^k) = -e^k + O(\|e_k\|^2).$$

On the other hand, substitute e_{k+1}, e_k into equation (94)

$$
\begin{aligned}
e_{k+1} =& \theta^* - \theta^{k+1} = \theta^* - (\theta^k - H^{-1}(\theta^k)J^{(2)}(\theta^k)) \\
=& e_k + DJ^{(2)}(\theta^k)J^{(2)}(\theta^k) + [H^{-1}(\theta^k) - DJ^{(2)}(\theta^k)]J^{(2)}(\theta^k) \qquad (95) \\
=& [H^{-1}(\theta^k) - DJ^{(2)}(\theta^k)]J^{(2)}(\theta^k) + O(\|e_k\|^2).
\end{aligned}
$$

We need to consider $\|H^{-1}(\theta^k) - DJ^{(2)}(\theta^k)]J^{(2)}(\theta^k)\|$ in term of e_k. Since $DJ^{(2)}(\theta^*)$ is invertible and $J^{(2)}(\theta)$ is smooth, then $DJ^{(2)}(\theta)^{-1}$ exists and is a smooth function when $\theta \in B(\theta^*, \epsilon)$. Moreover we apply Taylor expansion of function $\|DJ^{(2)}(\theta)^{-1} - H^{-1}(\theta)\|$:

$$
\begin{aligned}
\|DJ^{(2)}(\theta^k)^{-1} - H^{-1}(\theta^k)\| \leq& \|DJ^{(2)}(\theta^*)^{-1} - H^{-1}(\theta^*)\| + C\|\theta^k - \theta^*\| \\
=& C\|\theta^k - \theta^*\| = C\|e_k\|,
\end{aligned}
$$

where

$$C = \sup_{\theta \in B(\theta^*,\epsilon)} \|DJ^{(2)}(\theta)^{-1} - H^{-1}(\theta)\|.$$

Combine all results into (95)

$$
\begin{aligned}
\|e_{k+1}\| \leq& \|DJ^{(2)}(\theta^k)^{-1} - H^{-1}(\theta^k)\|\|J^{(2)}(\theta^k)\| + O(\|e_k\|^2) \\
=& C\|e_k\|\|J^{(2)}(\theta^k)\| + O(\|e_k\|^2).
\end{aligned}
$$

Hence

$$\frac{\|e_{k+1}\|}{\|e_k\|} \leq C\|J^{(2)}(\theta^k)\| + O(\|e_k\|).$$

Since θ_k converges θ^*, $\lim_{k\to\infty} J^{(2)}(\theta^k) = 0$ and $\lim_{k\to\infty} \|e_k\| = 0$. Substitute them, we obtain $\lim_{k\to\infty} \|e_{k+1}\|/\|e_k\| = 0$, which finishes the proof.

With all the components discussed above, we are ready to state our algorithm.

MEJ with Newton-like acceleration

Input: Number of intermittent diffusion intervals m.

Output: The optimal set γ^* for the junctions.

---

1.   Initialization. Find the initial path $\gamma^{(0)} = (\theta_1, \cdots, \theta_n)$;

2.   Select the duration of diffusion ΔT_l, $l \leq m$;

3.   Select diffusion coefficients σ_l, $l \leq m$;

4.   **for** $l = 1 : m$

5. $\gamma^{(l)} = \gamma^{(0)}$;

6.   **for** $j = 1 : \Delta T_l$

7.   Find $\nabla J(\gamma^{(l)})$.

8.   Update $\gamma^{(l)}$ according to (90) with $\sigma(t) = \sigma_l$;

9.   Remove junctions from or add junctions to $\gamma^{(l)}$ when necessary;

10.   **end**

11.   **while** $\|\nabla J(\gamma^{(l)})\| > \epsilon$

12.   Update $\gamma^{(l)}$ according to (94) with $\sigma(t) = 0$;

13.   **end**

14.   **end**

15.   Compare $J(\gamma^{(l)})$, $l \leq m$ and set $\gamma_{opt} = \mathrm{argmin}_{l \leq m} J(\gamma^{(l)})$;

6.3.2 Numerical experiments

In this subsection, we use two numerical examples to show the effectiveness of new algorithm.

Example 1: In this case, the obstacles are 5 disks with centers $(1,1)$, $(1.5, 1.5)$, $(0.5, 0.5)$, $(1.5, 0.5)$, $(0.5, 1.5)$ and radius 0.2, 0.2, 0.3, 0.25, 0.15 respectively. The starting and ending points are $X = (1.8, 0.2)$, $Y = (0.1, 1.7)$. We take ID step $m = 20$ defined in formula (89). Figure 32 shows the four shortest paths found by the algorithm. They are local minimizers and the global minimizer is shown in (C).

(a) $L = 3.0167$ (b) $L = 2.5762$

(c) $L = 2.3575$ (d) $L = 2.5045$
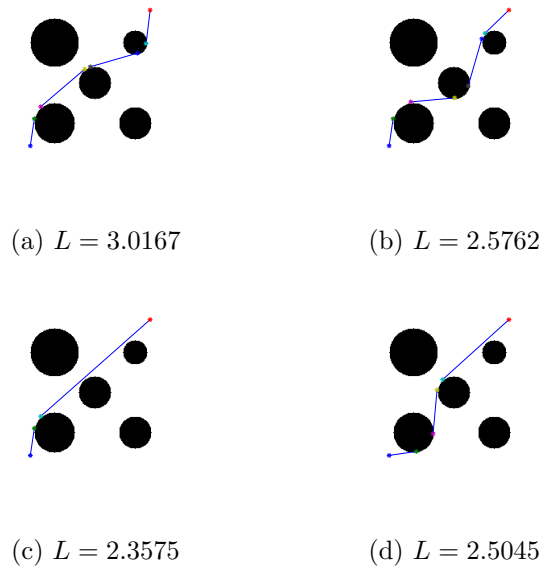
Figure 32: Shortest path problem: Multiple obstacles

It is worth to point out that by using a 2013 Macbook Air with CPU core i5, 1.8G HZ, RAM 4GB, our method needs only 1.75 seconds, while the method in [29] spends 485.777 seconds. Here two methods compute 10 local minimizers, where our method uses 0.175 seconds in average for each local minimizer and the gradient descent needs 48.58 seconds. This indicates that the computation time is reduced more than 200 times, which is due to the super-linear convergence result in Theorem 45. The reason for such an improvement is that it takes within 10 steps for the Newton
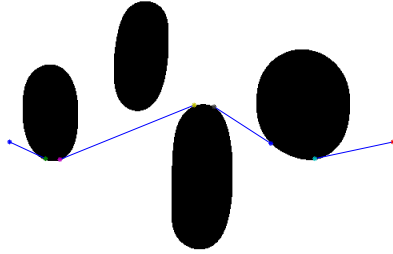
Figure 33: Shortest path problem: General obstacles

method to find a local minimizer while the gradient flow often takes much more steps. Moreover, it still has the advantage of stochastic method approaching the global minimizer. In other words, the larger m, the larger probability is to obtaining a global minimizer. To show that numerically, by computing 10 independent simulations when each individual simulation is with $m = 20$, among the 10 simulations, we find the global minimizer 5 times, which is Figure (C). While letting $m = 50$, we observe that Figure (C) happens 7 times among 10 simulations.

Example 2: Consider general obstacles same as in [29]. There are four obstacles with starting point $X = (0.5, 0.002)$ and ending point $Y = (0.5, 0.98)$, see Figure 33. Instead of having analytical parametrization as in Example 1, we obtain curvature, principle norm though level set method [78]. To compare with [29] for one local minimizer, we only take ID step $m = 1$ defined in formula (89), our method needs 3.719 seconds, while the method in [29] takes 215.840 seconds.

6.4 Differential games

In this section, we apply MEJ to a broader settings, differential games [19]. Here the "differential" refers that all players' behaviors are subject to differential equations. And the "game" means that each player faces his own optimal control problem, in

which the strategies are control variables and the payoffs (costs) are functionals[1].

In details, the differential game with N players is described as follows: Player v has his own cost functional to minimize. All players' strategies satisfies certain constraints. In mathematics, the game is represented as

$$\min_{x^v(t)} \ J_v(x^v(t), x^{-v}(t)) \quad \text{s.t.} \quad (x^v(t), x^{-v}(t)) \in S. \tag{96}$$

Here $x^{-v}(t) = (x^1(t), \cdots, x^{v-1}(t), x^{v+1}(t), \cdots, x^N(t))$ means all players' choices other than player v and S is the common strategy set.

In this sequel, we consider mainly a multiple robots' path-planning problem to illustrate.

Example 32 *Consider a two players' game. They design paths for two robots traveling from starting points A, B to destination positions A_1, B_1. Each player tries to minimize his own robots' gas consumption and does't want to see his own drone getting too close to the others. Let's denote $x^1(t)$, $x^2(t) \in \mathbb{R}^2$ as two robots' paths. Then the game means*

$$\min_{x^1(t)} \int_0^T \dot{x}^1(t)^2 dt \quad s.t \quad x^1(0) = A, \quad x^1(T) = A_1, \quad dist(x^1(t), x^2(t)) \geq \epsilon,$$

and

$$\min_{x^2(t)} \int_0^T \dot{x}^2(t)^2 dt \quad s.t \quad x^2(0) = B, \quad x^2(T) = B_1, \quad dist(x^1(t), x^2(t)) \geq \epsilon,$$

where T is a fixed terminal time, ϵ is the smallest allowed distance during two drones' travel and the quadric Lagrangian represents the fuel consumption of the robot.

One can give the definition of Nash equilibrium in differential games. Similarly to classical games, it describes a special status in which each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing his own strategy.

---

[1]In contrast, the classical game is "static" game, which doesn't contain the time variable in its description. In static game, each player faces his own optimization problem, in which the strategies are variables and payoffs (costs) are functions. See [44].

Definition 46 $x^*(t) = (x^{*1}(t), \cdots, x^{*N}(t))$ *is a NE if*

$$J_v(x^{*v}(t), x^{*-v}(t)) \leq J_v(x^v(t), x^{*-v}(t)), \quad \text{for any} \quad (x^v(t), x^{*-v}(t)) \in S.$$

People usually find NEs by each player's optimality conditions, which requires to solve multiple Pontryagin's maximal principles or a system of Hamilton-Jacobi-Bellman equations. In practice, these methods are numerically painful because of constraints. Instead of following traditional ways, we apply MEJ to find NEs. Similarly, we transfer the game from Banach space into games in finite dimensional spaces.

To simply the illustration, we apply MEJ in potential games. Potential game is a special type of game, which bridges the optimal control and differential game. It means that there exists a objective functional, named potential, such that the NE is the minimizer of potential. Example 32 is a potential game with potential $\int_0^T \sum_{v=1}^2 (\dot{x}^v(t))^2 dt$. In other words, the game means

$$\min_{x^1, x^2} \int_0^T \dot{x}^1(t)^2 + \dot{x}^2(t)^2 dt \tag{97}$$

subject to

$$x^1(0) = A, \quad x^1(T) = A_1, \quad x^2(0) = B, \quad x^2(T) = B_1, \quad \text{dist}(x^1(t), x^2(t)) \geq \epsilon.$$

We solve (97) by MEJ. To keep the presentation simple, we denote $x^1(t)$, $x^2(t)$ as $x(t)$, $y(t)$, where robots are called X, Y. Whenever two drones are with distance ϵ, the junctions are defined. For convenience, we denote junctions by $\tilde{x}_i = (t_i, x_i, y_i)$, where $x_i = x(t_i)$, $y_i = y(t_i)$.

Step one: We solve the optimal path with inactive constraints:

$$\min_{x,y} \int_{t_i}^{t_{i+1}} u(t)^2 + v(t)^2 dt \tag{98}$$

subject to

$$\dot{x} = u, \quad \dot{y} = v, \quad t \in [t_i, t_{i+1}],$$

$$x(t_i) = x_i, \quad y(t_i) = y_i, \quad x(t_{i+1}) = x_{i+1}, \quad y(t_{i+1}) = y_{i+1}.$$

166

It is easy to show that the optimal path for robot X is a constant velocity line connecting x_i, x_{i+1}, while the optimal path for robot Y is a constant velocity line connecting y_i, y_{i+1}. Hence we obtain the optimal cost functional

$$J(\tilde{x}_i, \tilde{x}_{i+1}) = \frac{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}{t_{i+1} - t_i}$$

as a function of $\tilde{x}_i$, $\tilde{x}_{i+1}$.

Step two: We solve the optimal path with active constraints:

$$\min_{x,y,u,v} \int_{t_i}^{t_{i+1}} u(t)^2 + v(t)^2 dt \tag{99}$$

subject to

$$\dot{x} = u, \quad \dot{y} = v, \quad t \in [t_i, t_{i+1}],$$

$$x(t_i) = x_i, \quad y(t_i) = y_i, \quad x(t_{i+1}) = x_{i+1}, \quad y(t_{i+1}) = y_{i+1},$$

$$\text{dist}(x(t), y(t)) = \epsilon, \quad t \in [t_i, t_{i+1}].$$

If we parameterize $(x(t), y(t))$ in a particular way, we will obtain an equivalent optimal control problem without constraints. Let's re-parameterize the trajectories $(x(t), y(t))$ by $(y(t), u(t))$, where $u(t)$ satisfies

$$x(t) = y(t) + \epsilon \begin{pmatrix} \cos u(t) \\ \sin u(t) \end{pmatrix},$$

which is the angle position of Robot X with respect to Robot Y. Since two robots travel in $\mathbb{R}^2$, we denote $y(t) = (y_1(t), y_2(t))$, $u(t) = (u_1(t), u_2(t))$. Hence the optimal control (99) forms

$$\min_{y_1,y_2,u,v_1,v_2,w} \int_{t_i}^{t_{i+1}} 2(v_1^2 + v_2^2) - 2\epsilon w(\sin u \cdot v_1 - \cos u \cdot v_2) + \epsilon^2 w^2 dt$$

subject to

$$\dot{y}_1 = v_1, \quad \dot{y}_2 = v_2, \quad \dot{u} = w, \quad t \in [t_i, t_{i+1}],$$

$$y(t_i) = y_i, \quad y(t_{i+1}) = y_{i+1},$$

$$u(t_i) = u_i, \quad u(t_{i+1}) = u_{i+1}.$$

167

We solve the above optimal control by optimality conditions. Denote the running cost

$$f(y_1, y_2, u, v_1, v_2, w) = 2(v_1^2 + v_2^2) - 2\epsilon w(\sin u \cdot v_1 - \cos u \cdot v_2) + \epsilon^2 w^2,$$

and Hamiltonian

$$H(y_1, y_2, u, v_1, v_2, w, \lambda_1, \lambda_2, \lambda) = f(y_1, y_2, u, v_1, v_2, w) + \lambda_1 v_1 + \lambda_2 v_2 + \lambda w.$$

By Pontryagin's maximal principle, we obtain

$$
\begin{cases}
H_{v_1} = 0 \\
H_{v_2} = 0 \\
H_w = 0 \\
\dot{\lambda}_1 = -H_{y_1} \\
\dot{\lambda}_2 = -H_{y_2} \\
\dot{\lambda} = -H_u
\end{cases}
\rightarrow
\begin{cases}
4v_1 - 2\epsilon w \sin u + \lambda_1 = 0 \\
4v_2 + 2\epsilon w \cos u + \lambda_2 = 0 \\
-2\epsilon v_1 + 2\epsilon \cos u \cdot v_2 + 2\epsilon^2 w + \lambda = 0 \\
\dot{\lambda}_1 = 0 \\
\dot{\lambda}_2 = 0 \\
\dot{\lambda} = 2\epsilon w(\cos u \cdot v_1 + \sin u \cdot v_2).
\end{cases}
$$

Interestingly, the above ODE system has an explicit solution. By solving the above system, we find

$$\dot{w} = \ddot{u} = 0.$$

It means the angle accretion of robot X relative to Y is 0. I.e. the optimal path forms

$$u(t) = u_i + \frac{u_{i+1} - u_i}{t_{i+1} - t_i}(t - t_i),$$

and

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = y_i + \begin{pmatrix} -\epsilon(\cos u(t) - \cos u_i) + c_1(t - t_i) \\ -\epsilon(\sin u(t) - \sin u_i) + c_2(t - t_i) \end{pmatrix},$$

where

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{y_{i+1} - y_i}{t_{i+1} - t_i} + \frac{\epsilon}{2}\begin{pmatrix} \frac{\cos u_{i+1} - \cos u_i}{t_{i+1} - t_i} \\ \frac{\sin u_{i+1} - \sin u_i}{t_{i+1} - t_i} \end{pmatrix}.$$

Hence the optimal cost functional becomes

$$J(\tilde{x}_i, \tilde{x}_{i+1}) = [\frac{\epsilon^2}{2}(\frac{u_{i+1} - u_i}{t_{i+1} - t_i})^2 + 2(c_1^2 + c_2^2)](t_{i+1} - t_i)$$

a function of $\tilde{x}_i$ and $\tilde{x}_{i+1}$.

Step three: We derive the visibility functions by junctions:

$$V(\tilde{x}_i, \tilde{x}_{i+1}) = \min_{t_i \leq t \leq t_{i+1}} \text{dist}(x^*(t), y^*(t)) - \epsilon,$$

where i is even and $x^*(t)$, $y^*(t)$ represent the optimal paths connecting junctions x_i, y_i and x_{i+1}, y_{i+1}, which are lines with constant velocities .

Combine all steps, we obtain a finite dimensional optimization:

$$\min_{x_1,\dots,x_n,y_1,\dots,y_n} \sum_{i=0}^{n} J(x_i, x_{i+1}, y_i, y_{i+1}), \quad \text{s.t.} \quad V(x_i, x_{i+1}, y_i, y_{i+1}) = 0, \quad i \text{ even.}$$

We demonstrate MEJ by numerical two examples. Firstly, we design a two player game. The game is in 2-dimensional environment, where the robot X travels from $A = (0,0)$ to $A_1 = (1,1)$ and robot Y travels from $B = (0,1)$ to $B_1 = (1,0)$. Let the terminal time be $T = 1$ and the safe distance be $\epsilon = 0.2$. We obtain a NE by MEJ, see Figure 34.
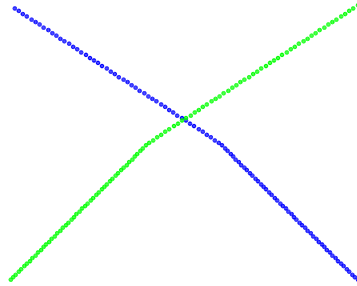
Figure 34: This is the snap short of two robots' game. Optimal paths for two robots represented by blue, green respectively form the Nash equilibrium.

Secondly, we consider a three robots' game. We design an environment with three robots traveling from starting points $A = (0, \sqrt{3})$, $B = (-1,0)$, $C = (1,0)$ to ending

169

points A_1, B_1, C_1, which are centers of $\overline{BC}$, $\overline{AC}$, $\overline{AB}$. Let the terminal time be $T = 1$ and the safe distance be $\epsilon = 0.2$. We solve a NE by MEJ, see Figure 35.
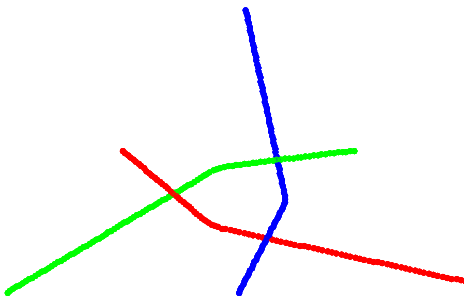
Figure 35: This is the snap short of three robots' game. Optimal paths for three robots represented by blue, red, green respectively form the Nash equilibrium.

6.5 Conclusions

In summary, we present MEJ for the separable optimal control problems with both state and control constraints. The method has following advantages compared to the existing methods:

1. Significant dimension reduction. Optimal control problems are in general considered as infinite dimensional problems in Banach spaces. By leveraging the separability structure of the optimal solution, MEJ reformulates the objectives and constraints in terms of junctions living on the boundaries of the constraints. In this way, MEJ restricts its search space to a finite dimensional subset of all feasible solutions without loss of any possible optimal solutions. This fundamentally changes the computation complexity and achieves significant dimensional reduction.

2. Fast and accurate. Since MEJ only needs to solve initial value SDEs, it can be more efficient than solving PDEs, boundary value ODEs and constrained

NLPs. So it has the potential to be very fast. Our experiments confirm this claim. Moreover, MEJ does not create additional accuracy restrictions from the discretization of the computed path.

3. MEJ has the ability to find the globally optimal trajectory as well as a series of locally optimal trajectories by the adoption of ID.

On the other hand, MEJ creates some theoretical questions that are very interesting on their own. For example, the SDEs solved by MEJ may change dimensions dynamically during its course, and the time and location of the change cannot be prescribed a priori. This is a question that has not been studied in mathematics at all. Compared to the existing methods for optimal control problems, MEJ requires customized reformulation to convert the original problem into a constrained optimization in terms of junctions. For certain problems, this may not be trivial tasks, since we require to solve the sub-optimal control problem by analytical solutions or fast numerical methods. Nevertheless, we demonstrated through examples that MEJ can be applied to several challenging problems including linear quadric problems with constraints, the optimal path planning with moving obstacles and a special differential game.

CHAPTER VII

PART 3: STOCHASTIC OSCILLATOR

7.1 Introduction

In this chapter, we consider the third part of this thesis, which is mainly about the analysis of stochastic oscillator in modeling.

It is well understood that many engineering and physical systems, like oscillators, can be modeled by deterministic dynamical systems having stable limit cycles (periodic orbits) as attractors.

A prototypical example is the van der Pol oscillator, that is governed by the second order differential equation:

$$\ddot{x} - \alpha(1 - x^2)\dot{x} + x \; = \; 0 \; . \tag{100}$$

It is well known that, for positive α, every solution of (100), except the origin, is attracted to the unique orbitally stable limit cycle, and that the strength of the damping, α, is intimately related to the rate at which trajectories approach this limit cycle.

However, in practice, noise is inevitable, and this motivates including random perturbation effects in the differential equations models. Among the many ways in which this has been done, we will focus on the case when the randomness takes the form of a forcing term. For example, when we add random noise to (100), we will obtain the following equation:

$$\ddot{x}_\epsilon - \alpha(1 - x_\epsilon^2)\dot{x}_\epsilon + x_\epsilon + \epsilon\xi = 0 \; , \tag{101}$$

where ξ represents the random perturbation, and ϵ is a small (positive) value. In (101), and hereafter, x_ϵ will denote the "solution" when (100) has been subject to

172

random perturbations as in (101). (For later reference, note that noise has been added to the original second order problem, prior to converting it into a first order system; see below).

A commonly used model of random perturbation ξ is white noise; i.e., $\xi = dW_t$, where W_t is the standard 1-dimensional Brownian motion. In this case, (101) becomes the classical stochastic van der Pol oscillator (weakly perturbed, for ϵ small). Other models of noise have been studied in [10, 11].

The presence of noise in a differential equation brings in several new challenges that require different approaches from those of deterministic dynamics. Of course, the key fact is that the dynamics will depend on the noise, not on the initial conditions. One of the most dramatic impacts of this fact is that (for any model of noise of which we are aware) the stable limit cycle gets destroyed. Finally, it is also worth realizing that noise causes changes in both phase and amplitude of the solutions. The impact on the phase is usually termed *phase noise*, or *time jitter* in the engineering literature [50], and considerable progress has been made, both in mathematics and engineering, toward understanding phase noise. For example, it is well appreciated that phase noise can become arbitrarily large even for perturbations that remain small [33]. Moreover, for white noise, a fundamentally important and striking result (see [8, 46]) states that –with probability arbitrarily close to 1– trajectories asymptotically escape from any neighborhood of the deterministic limit cycle!

However, in real life, things do not appear to be nearly as bad. We give three examples to support this statement. First, consider the circuits (oscillators) commonly used in cellular phones: these have a base frequency of around 1GHz, oscillating in excess of 10^9 times per second. While being subject to unavoidable random ambient disturbances, a cell phone oscillator typically works continuously for days, even

months or years, without experiencing any break down. Second, in laboratory stud-ies[1] on a cantilevered piezoelectric energy harvester, which is a electroelastic system converting ambient vibrations generated by stochastic perturbations into electricity through the direct piezoelectric effect, no breakdown caused by random perturbation was actually ever observed. Finally, the reports in [26] indicate that trajectories of a weakly perturbed van der Pol oscillator remain bounded and linger near the deter-ministic limit cycle. In fact, the results of this cited numerical study are consistent with our own numerical simulations of equation (101), with white noise perturbations, over long times; see Figure 36. Clearly, trajectories appear to remain in a tubular neighborhood of the deterministic limit cycle, and do not become arbitrarily large.
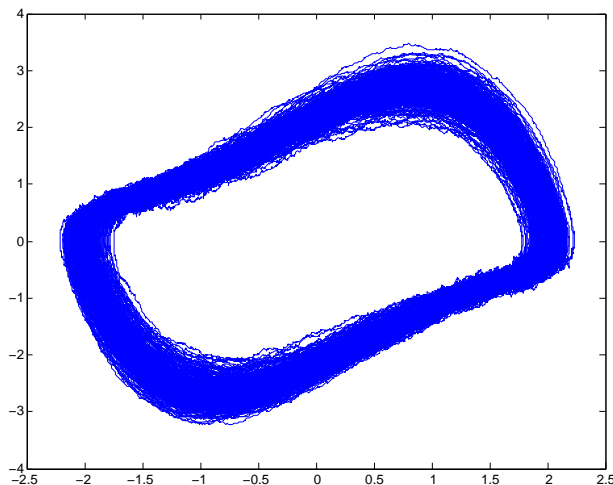
Figure 36: Long time behavior of (101) in numerical experiments.

This discrepancy between existing theoretical predictions and practical observa-tions is likely due to two factors: (i) the asymptotic nature of the theoretical results,

which typically require an extremely long time to be observable (if at all), and/or (ii) the inadequate modeling of the noise, meaning that practical random perturbations must have bounded strength (there is no noise perturbation with infinite energy), which is different from the white noise assumption commonly used in theoretical studies. [To explain the numerical results summarized in Figure 36, we note that –although we do not force any restriction on the random number generator used to mimic white noise– the pseudo random number generator used in our computation does not (and cannot) produce infinitely large perturbations.]

The above state of affairs provided us with the main motivation to carry out the present study. In particular, the above point (ii) is our key concern in this work.

We will focus on second order dissipative systems (oscillators) that posses an orbitally stable limit cycle surrounding a unique unstable equilibrium (at the origin), and we will study the impact of noise on these systems. Our main goals are: (1) to provide a new mathematical model for *realistic* random perturbations so that the trajectories of the stochastic oscillators resemble the phenomena observed in practice; and, (2) to study the behavior of solutions of these stochastic oscillators.

7.1.1 A new model of noise

Accounting for the possibility that standard white noise can generate infinitely large perturbations (albeit with arbitrarily small probability), while a realistic model of noise should never inject infinitely large energy into the system, here we propose a new model of noise that we believe serve as an appropriate model for random perturbations arising in practice.

Namely, we will require that the random perturbations ξ belong to the event set B defined as:

$$B \;=\; \{\, \omega \mid \sup_{|t-s|\leq T} |W_t(\omega) - W_s(\omega)| \leq M \,\} \,. \tag{102}$$

In (102), T and M are two given positive constants, t and s are any two instants of

time at most T-apart, and ω is the event of a Brownian path.

Note that B, a subset of all Brownian paths, is the collection of those Brownian motions that have bounded finite time increments. However, note that a path in B can still diverge to infinity as $t \to \infty$. Now, if one is interested in the finite time behavior of the system, then the probability of a Brownian path not in B can be made arbitrarily small by taking M large enough, because of Hölder continuity of the Brownian motion path. However, if infinite time is considered, we observe that B has measure zero in the set of all Brownian paths defined for $t \in [0, \infty)$. Nevertheless, this does not imply there are not sufficiently many paths in B for $t \in [0, \infty)$: in fact, B contains un-countably many paths for $t \in [0, \infty)$, maintaining key characteristics of Brownian motion, including the general order of continuity of $\frac{1}{2}$.

7.1.2 Our results

We shall show that selecting random perturbations ξ from B for perturbing a deterministic oscillator with attracting limit cycle, and whose right-hand-side satisfies a local Lipschitz condition[2], will give well defined solution trajectories that remain bounded for all times. With reference to (100) and (101), it is worth emphasizing that this does not mean that, for all t, $(x_\epsilon(t), y_\epsilon(t))^T$ will stay close to its deterministic counterpart $(x(t), y(t))^T$. In fact, the phase differences between stochastic and deterministic trajectories can become large in time. On the other hand, we will show that $(x_\epsilon(t), y_\epsilon(t))^T$ remains close to the deterministic limit cycle for all t, and we will further show several desirable properties of the stochastic trajectories relative to our new model of noise.

To witness, if we take a short segment transversal to the limit cycle (a "section"), we will show that the stochastic trajectories will return to this section, under appropriate conditions. As a consequence, we will set forth a proposal for defining the

---

[2]The local Lipschitz condition becomes a global Lipschitz condition if the solutions remain bounded.

Poincaré return map relative to the stochastic oscillators. This is very different from the scenario obtained when one uses standard white noise, in which case there is no guarantee that a trajectory will return to a given section.

In comparison to the Poincaré map for deterministic systems, our proposal of Poincaré map for the stochastic systems has some new features that have not been studied before. Namely, unlike the deterministic case, there is no longer just a first return point for a trajectory "going around the origin once." In fact, a solution path can (and does) intersect the given section repeatedly, and it could do so infinitely many times, while the trajectory goes around the origin just one time. As a consequence of this observation, our proposal will be to relate to each given section a *return interval* and an associated *distribution for the return points*; both return interval and distribution will depend on the section. An important outcome of the above proposal is that we will have at least three different Poincaré maps: (i) that associated to the first return points distribution, (ii) that associated to the average of the return points distribution, and (iii) that associated to the last return points distribution.

Finally, we will also investigate the evolution of the probability density function of the stochastic oscillator with noise in B. In the present case, the processes are no longer Markovian, because the random perturbations depend on their past in an interval of length T, and not only on their current values. This inhibits the possibility to write a standard Fokker-Planck equation (see below). What we shall show is that, under appropriate conditions, the probability density function can be given by rational functions depending on solutions of a pair of diffusive partial differential equations (PDEs) with vanishing boundary conditions on finite intervals.

The chapter is arranged as follows. In section 7.2, we consider a dissipative oscillator subject to random perturbations from B, and we show local (in time) boundedness of trajectories. In section 7.3, we introduce our proposal of stochastic Poincaré map, and show the main result of this paper, the global boundedness of solutions. Lastly,

in section 7.4, we study the evolution of the probability density function in terms of the solutions of some associated PDEs.

Notation: Throughout this work, the vector norm is always the 2-norm, which will be indicated simply as $\| \cdot \|$.

7.1.3   Relation to previous results

A lot of effort has been devoted to study the changes that solutions undergo under the effect of white noise perturbations. But, unfortunately, the existing results require modeling assumptions which make them inapplicable to our problem. We justify this claim below.

For a planar system of differential equations, the basic model considered is the stochastic differential equations (SDE)

$$dX = g(X)dt + A(X)dW_t \ , \tag{103}$$

where $X(t) = (x(t), y(t))^T$, the term W_t comprises two independent 1-dimensional Brownian motions, and the diffusion coefficient is such that the matrix AA^T is full rank. The latter property is often referred to as "uniform ellipticity." We refer to the excellent expositions in [7, 8, 46, 92], for details and further references. But, it is worth pointing out that the system(s) of interest to us, such as (101), do not fit into the model (103). This can be readily seen if we convert the second order equation (101) into a first order system, say

$$\begin{cases} dx_\epsilon = y_\epsilon dt, \\ dy_\epsilon = [\alpha(1 - x_\epsilon^2)y_\epsilon - x_\epsilon]dt + \epsilon dW_t \ . \end{cases} \tag{104}$$

It is very important to observe that random noise is only added to the second equation in (104). Mathematically, this is easily explained as having added the perturbation to the original second order equation (100), prior to converting it into first order system. But, a more intrinsic and deep reason is that x and y are related to the current and

voltage, respectively, which have a fixed relationship for a given circuit. So, it is not physically justified to add independent noise to both equations in (104).

7.2   *Local boundedness of solutions*

In this section, we introduce our model of stochastically perturbed system with noise set B, and show local (in time) boundedness of trajectories.

We consider a second order system

$$\ddot{x} = f(x, \dot{x}) , \tag{105}$$

where f is a smooth function of its arguments. We rewrite (105) as the first order system

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ f(x, y) \end{pmatrix} , \quad \text{or simply as} \quad \dot{X} = b(X), \quad X = \begin{pmatrix} x \\ y \end{pmatrix} , \tag{106}$$

and we will always work under the following **assumptions** on (106):

(i) the origin is the only equilibrium of (106), $b(0) = 0$, and it is an unstable focus;

(ii) the system possesses a globally orbitally stable limit cycle Γ, corresponding to a periodic solution of period $0 < T_0$;

(iii) for X: $\|X\| \le C$, where C is any (arbitrary, but finite) positive constant, the function b is smooth and locally Lipschitz, with Lipschitz constant L (usually, L will depend on C).

Finally, the solution of (106) with initial condition $X(0) = u$, will be written as $\phi^t(u)$.

Of course, as a consequence of the above assumptions, all orbits of (106) (except the origin) will approach Γ. In Section 7.3, we will quantify better the rate of approach to Γ. Finally, note that, in general, the function f may depend on a parameter α, as in (100), or even on several parameters, and it must be tacitly understood that the above assumptions must hold for all allowed values of the parameter(s); in particular,

the period T_0 (which usually will depend on the problem's parameters) must remain finite as the parameter(s) varies (vary).

Our specific interest in this work is in the following perturbation of (106):

$$\begin{cases} dx_\epsilon &= y_\epsilon dt \\ dy_\epsilon &= f(x_\epsilon, y_\epsilon)dt + \epsilon dW_t \ , \end{cases} \tag{107}$$

where ϵ is a (small) positive parameter, and W_t is a 1-dimensional Wiener process so that $\xi = dW_t$ is in the event set B given in (102). An initial condition of (107) is written as $X_\epsilon(0)$, and its solution as $X_\epsilon(t, \omega)$, or simply $X_\epsilon(t)$, if no confusion can arise.

As already remarked, there are good modeling reasons for considering the noise model given by the set B. The key reason, for us, has been to adopt a model of noise more in tune with what is typically observed in practice, whereby realistic ambient noise is bounded within a finite time interval (unlike, say, white noise). Indeed, on intervals of length T, noise realizations from B are locally bounded, which is meaningful since, in real world scenarios, energy is always bounded, and no perturbation can become unbounded in finite time. (Still, note that noise realizations from the set B can still become eventually unbounded, since the total increment is not constrained to remain bounded.)

As added benefit, restricting to the event set B, we will be able to show important mathematical properties of the model (107). Most notably, we will be able to propose a definition of Poincaré map, see Section 7.3. But, first, below we show that stochastic trajectories remain bounded in a finite time interval.

Theorem 47 *Let B be the set defined in (102), and let ω be any event from B. Then, for $\epsilon > 0$ sufficiently small, solutions of (107) are locally bounded:*

$$\sup_{0 \leq t \leq T} \|X_\epsilon(t, \omega)\| < \infty \ ,$$

where T is the interval width appearing in (102).

Proof 39 *We will argue by contradiction. So, suppose that $\sup_{0 \leq t \leq T} \|X_\epsilon(t, \omega)\| = \infty$.*

Let C_1 be the maximum of $\|\phi^t(u)\|$, for $0 \leq t \leq T$, along the deterministic limit cycle:

$$C_1 = \max_{0 \leq t \leq T} \|\phi^t(u)\| \ , \quad u \in \Gamma \ .$$

Then, there must exist a constant $C > 2C_1$, for which the stopping time

$$\tau(\omega) = \inf\{t : \|X_\epsilon(t, \omega)\| = C\} \ ,$$

must satisfy

$$\tau(\omega_0) \leq T \ ,$$

for some $\omega_0 \in B$. In other words, for such ω_0, we have

$$\sup_{0 \leq s \leq \tau(\omega_0)} \|X_\epsilon(s, \omega_0)\| \ = \ C \ . \tag{108}$$

Consider this event ω_0. From [85], we know that there exists a strong solution $X_\epsilon(t, \omega_0)$ up to time $\tau(\omega_0)$ (see [59] for the definition of strong solution). Let L be the local Lipschitz constant of b when $\|X\| \leq C$. Hence, for $t \leq \tau(\omega_0)$, we have

$$\|X_\epsilon(t, \omega_0) - X(t)\| = \| \int_0^t b(X_\epsilon(s, \omega_0)) - b(X(s))ds + \epsilon \begin{pmatrix} 0 \\ W_t(\omega_0) \end{pmatrix} \|,$$

$$\leq L \int_0^t \|X_\epsilon(s, \omega_0) - X(s)\|ds + \epsilon|W_t(\omega_0)| \ .$$

From Gronwall's Lemma, and since $\omega_0 \in B$, we obtain

$$\sup_{0 \leq s \leq \tau(\omega_0)} \|X_\epsilon(s) - X(s)\| \leq \epsilon e^{L\tau(\omega_0)} \sup_{0 \leq s \leq \tau(\omega_0)} |W_s(\omega_0)|$$

$$\leq \epsilon e^{L\tau(\omega_0)} M \ . \tag{109}$$

Also, since $\tau(\omega_0) \leq T$, using the triangular inequality in (109) we get

$$\sup_{0 \leq s \leq \tau(\omega_0)} \|X_\epsilon(s)\| \leq \sup_{0 \leq s \leq \tau(\omega_0)} \|X(s)\| + \epsilon e^{\tau(\omega_0)L} M$$

$$\leq C_1 + \epsilon e^{LT} M \ .$$

Therefore, if $\epsilon < \frac{C_1}{2e^{LT}M}$, then

$$\sup_{0 \leq s \leq \tau(\omega_0)} \|X_\epsilon(s)\| \leq \frac{3}{2}C_1 < C \ ,$$

contradicting equation (108).

Theorem 47 establishes closeness, for short time, between stochastic and deterministic solutions (see (109)), and it gives a lead on how to define the return map (Poincaré map) in stochastic systems. We do this next.

7.3 Stochastic Poincaré map and global boundedness

In this section, we introduce our proposal of stochastic Poincaré map, for (107). Then, using the notation resulting from the definition of Poincaré map, we will show our main theorem: global boundedness of solutions of (107).

7.3.1 Poincaré map

First, recall the definition of Poincaré map in the deterministic setting. We do this in the plane, since we are interested in the model (106), though of course the definition can be easily given in $\mathbb{R}^d$, $d > 2$.

Consider the general differential equation for $X \in \mathbb{R}^2$

$$\frac{dX}{dt} = b(X) \ , \quad X(0) = u \ , \tag{110}$$

where b is a locally Lipschitz smooth vector field. Let $\phi^t(u)$ be the flow associated to (110) and suppose that (110) has a periodic solution of period T_0, and let Γ be the orbit corresponding to this periodic solution. Therefore, $\phi^{t+T_0}(p) = \phi^t(p)$, $t \in \mathbb{R}$, $p \in \Gamma$.

The Poincaré map provides a useful tool for studying periodic orbits, whereby a periodic orbit becomes a fixed point of the Poincaré map.

Definition 48 *Let p be a point on Γ and S be a local cross section at p: a smooth 1-dimensional arc, intersecting Γ (only) at p, transversally. Given an open and connected neighborhood U of p, $U \subset S$, for every point $u \in U$, define the* first return time $\tau(u)$ *as*

$$\tau(u) \;=\; \inf\{t > 0 \mid \phi^t(u) \in S\} \;.$$

Then, the Poincaré map $P : U \to S$ is defined by

$$P(u) \;=\; \phi^{\tau(u)}(u) \;,\; u \in U \;.$$
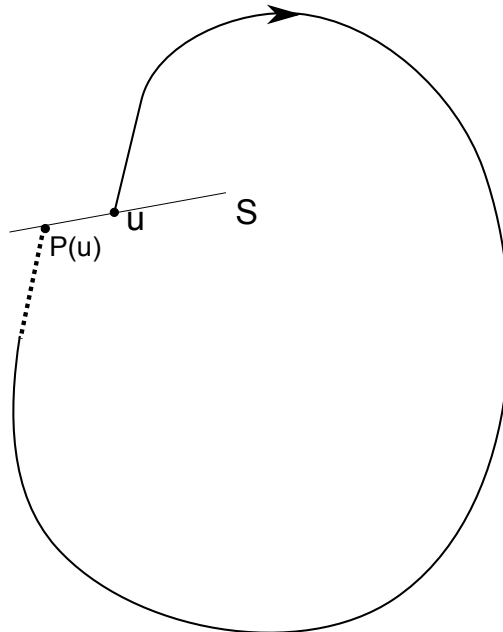

Figure 37: Poincaré map

Clearly $P(p) = p$ and $\tau(p) = T_0$.

With the help of the above definition, we can finally clarify assumption (ii), that we made in Section 7.2 relative to system (106). We say that Γ is attractive, with *rate* α_0, $0 < \alpha_0 < 1$, if:

$$\|P(u) - p\| \leq \alpha_0 \|u - p\| \;, \text{ for all } u \in U \;. \tag{111}$$

183

As before, in case (106) depends on parameters, it has to be understood that the inequality $\alpha_0 < 1$ must hold uniformly with respect to parameters variation.

7.3.2 Stochastic Poincaré map

Consider now the general SDE associated to (110):

$$dX_\epsilon(t) = b(X_\epsilon(t))dt + \epsilon \begin{pmatrix} 0 \\ dW_t(\omega_0) \end{pmatrix}, \ X_\epsilon(0) = u, \qquad (112)$$

where $X_\epsilon(\cdot)$, $u \in \mathbb{R}^2$ and W_t is a standard Wiener process in $\mathbb{R}^1$. In this case, it is known (see [59]) that a unique strong solution exists locally, that is for times before the stochastic solution blows up to infinity.

If we try to define a (stochastic) Poincaré map for (112), we face some intrinsic challenges.

1. With nonzero probability, the stochastic trajectory will not return "after one loop" to a given section, even though the unperturbed trajectory is periodic.

2. Even if the stochastic trajectory returns to a given section, the first return point doesn't represent all return points to that section. In fact, the trajectory can intersect a given section several times, even infinitely many times. [There is no monotonicity of motion with respect to a given section].

Selecting random perturbations from B in (102), and relative to the model (107), allows us to solve the above difficulties. In fact, the following two facts hold as a consequence of Theorem 47 and of (the proof of) Theorem 50 below.

(1) First, for all events in B, the stochastic trajectory of (107) will return to a given section.

(2) Second, although the stochastic trajectory may repeatedly enter and exit (or even stay for a while in) a certain section, it will have to leave such section within a finite time. See Figure 38 for an illustration of this fact.

184

Note that a stochastic path can (and does) intersect a given section repeatedly, and it could do so even infinitely many times, before leaving the section.

By virtue of points (1) and (2), and this last observation above, our **proposal** is to

Associate to a given section both a return interval and a distribution for the return points; both return interval and distribution will depend on the section.
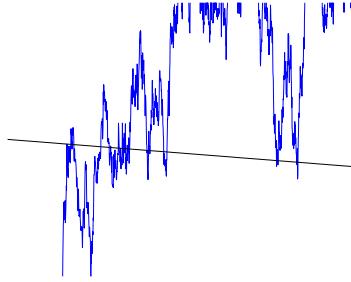
Figure 38: Return points of stochastic trajectory: the first return interval (solid segment)

Let us set forth our proposal more precisely.

Consider a section S and a neighborhood U of $p \in \Gamma$ as in Definition 48. For $\omega \in B$, and $u \in U$, let X_ϵ be the the stochastic trajectory of (107), such that $X_\epsilon(0) = u$.

To begin with, we introduce the first return time

$$\tau_\epsilon(u, \omega) = \inf\{t \mid X_\epsilon(t) \in S,\ t \in [\frac{1}{2}T_0, \frac{3}{2}T_0]\}\ , \tag{113}$$

and the last return time

$$\sigma_\epsilon(u, \omega) = \sup\{t \mid X_\epsilon(t) \in S,\ t \in [\frac{1}{2}T_0, \frac{3}{2}T_0]\}\ . \tag{114}$$

(Both of these values are well defined, for sufficiently small ϵ, because of Theorems 47 and 50.)

185

Then, we define the return "interval" $E(u, \omega)$ or simply E:

$$E = \{X_\epsilon(t) \mid X_\epsilon(t) \in S, \ t \in [\tau_\epsilon, \sigma_\epsilon]\} \ . \tag{115}$$

(Again, Theorems 47 and 50 ensure that E is well defined.)

Our proposal is to associate to E in (115) a " return distribution" function $P_{u,\omega}$, by which we can solve for the sample average of the set E. Since the distribution conveys all information on the return points, the Poincaré map (call it P_ϵ) should be constructed as a **point-to-distribution** map, for each stochastic path:

$$P_\epsilon : u \in U \to P_{u,\omega} \ .$$

At the same time, from the foregoing, it is natural to define three different point-to-point maps, all of which can be computed in a numerical simulation: first return map, last return map and average return map.

Definition 49 *Let the cross section S, neighborhood U, and Poincaré map P, be defined as in Definition 48, for the unperturbed system (106).*

Let $\omega \in B$, and X_ϵ be the solution of (107). Then, we define stochastic Poincaré maps P_ϵ, $P_\epsilon : U \to S$, as follows.

- *For $\epsilon = 0$, $P_0 = P$.*

- *For $\epsilon \neq 0$, and any $u \in U$, then we define:*

  - *the first return map*
    $$P_\epsilon(u, \omega) = X_\epsilon(\tau_\epsilon(u, \omega)) \ , \tag{116}$$
    where τ_ϵ is defined in (113);

  - *the last return map*
    $$P_\epsilon(u, \omega) = X_\epsilon(\sigma_\epsilon(u, \omega)) \ , \tag{117}$$
    where σ_ϵ is defined in (114);

186

– and the average return map

$$P_\epsilon(u, \omega) = \int_E y \, dP_{u,\omega}(y) \,, \tag{118}$$

where E is defined in (115) and $P_{u,\omega}$ is the return distribution function associated to E.

Remark 16 *In the recent work [53], the authors proposed a Poincaré map definition for the van der Pol oscillator, subject to standard white noise perturbation on ϵdW_t. Assuming that trajectories return to a given section (although there is a positive probability that they will not return), the authors further looked at the first return map and for sufficient small ϵ, argued that this map can be viewed as a Gaussian perturbation of the deterministic map. By comparison, restricting to noise from within B, we actually prove that trajectories always return (for ϵ sufficiently small) to a given section. Furthermore, our proposal of Poincare map takes into account all return points, and it gives a more detailed description relative to a given section, description which is not availably by a simple Gaussian process.*

7.3.3 Global boundedness

By exploiting the Poincaré map, we will show our main theorem, which we state next.

Theorem 50 *Consider the system (107), where ω is in B defined by (102). Assume that (111) holds, and that the assumptions of Lemmata 51 and 52 below hold (in particular, (123)). Then, for $\epsilon > 0$ sufficiently small, the stochastic trajectories of (107) are globally bounded:*

$$\sup_{t \geq 0} \|X_\epsilon(t, \omega)\| < \infty \,.$$

To prove this result, we will proceed according to the following steps.

1. We define the Poincaré section S as a line section, and show closeness between unperturbed and stochastic solutions during the first return time; see Lemma 51.

2. We construct the first return map P_ϵ (see Definition 49), and sharpen the result of Lemma 51 about closeness of the stochastic trajectory and the unperturbed Poincaré map for the first return to the given section; see Lemma 52.

3. Combining the above closeness results, and asymptotic stability of the deterministic limit cycle, we show that there exists a neighborhood (an interval) U_ϵ of $p \in \Gamma$, $U_\epsilon \subset S$, invariant under the stochastic Poincaré map: $P_\epsilon(U_\epsilon) \subset U_\epsilon$. This will complete the proof.

First of all, we define the Poincaré section. Through the polar representation of points in the plane, we take the section to be a line segment placed at a given angular value θ_0:

$$S \ = \ \{(x,y) \in \mathbb{R}^2 \mid x = r(\theta_0)\cos(\theta_0) \ , \ y = r(\theta_0)\sin(\theta_0)\} \ , \tag{119}$$

where $a(\theta_0) \leq r(\theta_0) \leq b(\theta_0)$, and a and b are chosen sufficiently close so that the line segment S intersects Γ transversally at just one point. With this, we can identify points of the stochastic trajectory that returns to this section:

$$\theta_\epsilon(t) = \theta_\epsilon(0) = \theta_0 \ .$$

Naturally, the neighborhood $U \subset S$ as in Definition 49, now becomes an open subinterval of S containing the intersection with Γ. This way of proceeding will be validated in Lemma 51.

To illustrate, in Figure 39, we show a typical stochastic trajectory of (101) starting from the section S, and traveling around the origin once before returning to S.

With this, we can be more specific about the meaning of first return time with respect to the section (cfr. with (116)):

$$\tau_\epsilon^1 = \inf\{t \mid \theta_\epsilon(t) = \theta_\epsilon(0) - 2\pi\} \ . \tag{120}$$

Without loss of generality, here below we will also make a simplification in the definition of B:

188
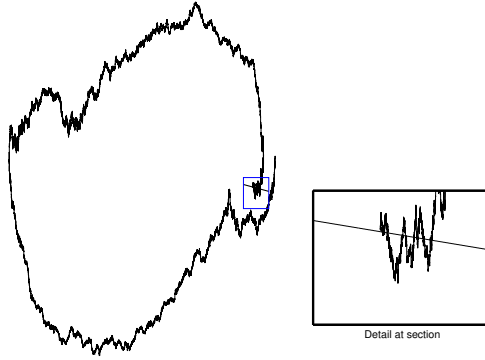
Figure 39: One realization for the stochastic van der Pol oscillator: on the section are all return points

"we require $T = 2T_0$ in (102), where $T_0 > 0$ is the period of the periodic solution of (106)."

This choice is legitimate, since we can always modify M to ensure that Theorem 47 holds for $T = 2T_0$. So, to reiterate, the event set B is henceforth given by

$$B = \{ \omega \mid \sup_{|t-s| \leq T} |W_t(\omega) - W_s(\omega)| \leq M , \text{ and } T = 2T_0 \} . \tag{121}$$

Finally, let L be the local Lipschitz constant of $b(X)$ in (106) for $\|X\| \leq R$, for R sufficiently large to enclose Γ.

We further consider the annular neighborhood of radius $\epsilon e^{LT} M$ around the periodic trajectory Γ; see Figure 40. Finally, we let times t_1 and t_2 be defined as follows (again, see Figure 40):

$$t_1 = \inf\{t > \frac{1}{2}T_0 \mid \|X(t) - X(T_0)\| = \epsilon e^{LT} M\} ,$$
$$t_2 = \sup\{t < \frac{3}{2}T_0 \mid \|X(t) - X(T_0)\| = \epsilon e^{LT} M\} . \tag{122}$$

Note that t_1, t_2 depend on ϵ, and that by continuity of the strong solution $X_\epsilon(t)$, $\tau_\epsilon^1 \in (t_1, t_2)$. Also, note that for $\epsilon > 0$ sufficiently small, with t_1, t_2 as in (122), then for $i = 1, 2$, we have:

$$\int_0^1 b\left(X(t_i + s(T_0 - t_i))\right) ds \neq 0 . \tag{123}$$

We are now ready to show closeness after the first return time.

Lemma 51 *Let B be as in (121), $\omega \in B$, and let L, T_0, and τ_ϵ^1 be as above. Then, for $\epsilon > 0$ sufficiently small, we have:*

$$\tau_\epsilon^1 < 2T_0 \ ,$$

and

$$\sup_{0 \le t \le \tau_\epsilon^1} \|X_\epsilon(t, \omega) - X(t)\| < C_0 \epsilon \ ,$$

where $C_0 = e^{LT} M$.

Proof 40 *Proceeding as in the derivation of (109), we have*

$$\sup_{0 \le s \le T} \|X_\epsilon(s) - X(s)\| \le \epsilon e^{LT} M \ . \tag{124}$$

Consider the deterministic system (106). For either $i = 1$ or $i = 2$, we have

$$X(T_0) - X(t_i) \;=\; \int_{t_i}^{T_0} b(X(s)) ds \;=\; \int_0^1 b\left(X(t_i + s(T_0 - t_i))\right) ds \; (T_0 - t_i) .$$

Since (123) holds, we can solve for $|T_0 - t_i|$ from this last equation:

$$|T_0 - t_i| = \frac{\|X(T_0) - X(t_i)\|}{\left\| \int_0^1 b\left(X(t_i + s(T_0 - t_i))\right) ds \right\|} \ . \tag{125}$$

Now, recall that $\tau_\epsilon^1 \in (t_1, t_2)$. Thus, by choosing ϵ:

$$\epsilon < \min_{i=1,2} \left[\frac{T_0 \left\| \int_0^1 b\left(X(t_i + s(T_0 - t_i))\right) ds \right\|}{4 e^{LT} M} \right] ,$$

and using this bound in (125), we get $\tau_\epsilon^1 < T = 2T_0$.

Hence, for $\omega \in B$, and for ϵ sufficiently small, inequality (124) holds up to time τ_ϵ^1.
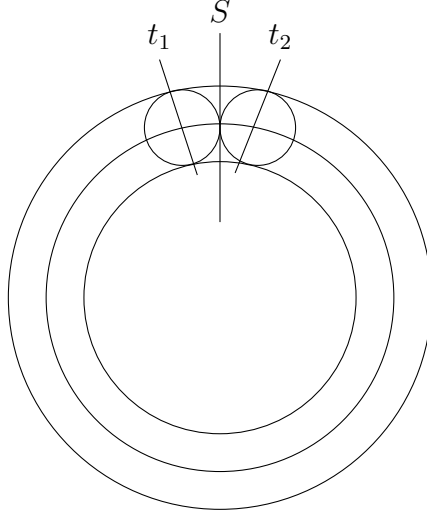
Figure 40: The middle circle represents Γ. The stochastic trajectories $X_\epsilon(\omega, t)$, $\omega \in B$, are inside the annulus.

Next, let P_ϵ denote the first return map as in (116), corresponding to the section S defined as in (119). We show the closeness on the section S.

Lemma 52 *With same notation and assumptions as in Lemma 51, let $X_\epsilon(\omega, t)$, be the solution of (107) for $0 \leq t \leq \tau_\epsilon^1$, with initial value $X_\epsilon(0)$ in U, a sufficiently small open interval of S.*

Then, for $\epsilon > 0$ sufficiently small,

$$\|X_\epsilon(\tau_\epsilon^1) - P_0(X_\epsilon(0))\| \leq 5C_0\epsilon.$$

Proof 41 *From Lemma 51, taking $t = \tau_\epsilon^1$, (124) becomes*

$$\|X_\epsilon(\tau_\epsilon^1) - \phi^{\tau_\epsilon^1}(X_\epsilon(0))\| \leq C_0\epsilon \ .$$

Also, we have both

$$\phi^{\tau_\epsilon^1}(X_\epsilon(0)), \ P_0(X_\epsilon(0)) \in \cup_{t \in [t_1, t_2]} \mathbf{B}(X(t), C_0\epsilon),$$

where t_1 and t_2 are defined in (122), and $\mathbf{B}(X(t), C_0\epsilon)$ are circles with center $X(t)$ and radius $C_0\epsilon$. Therefore, we have

$$\|X(\tau_\epsilon^1) - P_0(X_\epsilon(0))\| \leq 4C_0\epsilon \ ,$$

191

and so

$$\|X_\epsilon(\tau_\epsilon^1) - P_0(X_\epsilon(0))\| \leq \|X_\epsilon(\tau_\epsilon^1) - X(\tau_\epsilon^1)\| + \|X(\tau_\epsilon^1) - P_0(X_\epsilon(0))\| \leq 5C_0\epsilon .$$

Finally, we show global boundedness and complete the proof of Theorem 50, by using the stability of the deterministic limit cycle, as expressed by (111), and local boundedness of X_ϵ.

In the proof below, we will need to compare the stochastic and deterministic solutions. For this reason, we will use the following notations. For all $k = 1, 2, \ldots$, we write

$$\phi^{\Delta_\epsilon^k}\left(X_\epsilon(\tau_\epsilon^k)\right) \ , \text{ where } \Delta_\epsilon^k := \tau_\epsilon^{k+1} - \tau_\epsilon^k \ ,$$

for the solution of the deterministic equation (106) at time τ_ϵ^{k+1}, which started at time τ_ϵ^k with initial condition $X_\epsilon(\tau_\epsilon^k)$. Above, we have recursively defined the values τ_ϵ^k as the k-th "first return" times:

$$\tau_\epsilon^k = \inf\{t > \tau_\epsilon^{k-1} \mid \theta_\epsilon(t) = \theta_\epsilon(0) - 2\pi\} \ , \quad k = 1, 2, \ldots, \quad \tau_\epsilon^0 = 0 . \tag{126}$$

These values τ_ϵ^k will be shown to be well defined in the proof below.

Proof 42 (Proof of Theorem 50) *First, let us show that*

$$\sup_k \|X_\epsilon(\tau_\epsilon^k)\| < \infty .$$

The proof uses the fact that there exists an interval U_ϵ on the section S, such that

$$U_\epsilon = \{X \in S \mid \|X - p\| \leq R_0\} \ , \text{ such that } P_\epsilon(U_\epsilon) \subset U_\epsilon .$$

We show this last fact by induction, in the process showing that the times τ_ϵ^k's are well defined.

If $k = 1$, from Lemma 51, τ_ϵ^1 exists, in particular, from (122), we have $\tau_\epsilon^1 \in (\frac{1}{2}T_0, \frac{3}{2}T_0)$. From Lemma 52, we have

$$\|X_\epsilon(\tau_\epsilon^1) - P_0(X_\epsilon(0))\| \leq R_1\epsilon ,$$

where $R_1 = 5C_0$, and $C_0 = e^{LT}M$ with L the Lipschitz constant of b for X: $\|X\| \leq R$, with a sufficient large constant R to enclose the limit cycle Γ.

Also, since $X_\epsilon(0) \in U_\epsilon$, denoting with $p \in S$ the fixed point of P_0 and using (111), we have

$$\|X_\epsilon(\tau_\epsilon^1) - p\| \leq \|X_\epsilon(\tau_\epsilon^1) - P_0(X_\epsilon(0))\| + \|P_0(X_\epsilon(0)) - p\|$$

$$\leq R_1\epsilon + \alpha_0 R_0 \ .$$

So, if $\epsilon < \frac{(1-\alpha_0)R_0}{R_1}$ and a fortiori if $\epsilon < \frac{(1-\alpha_0)^2 R_0}{R_1}$, then $X_\epsilon(\tau_\epsilon^1) \in U_\epsilon$.

By induction, suppose that for $j = 1, \ldots, N$, the times τ_ϵ^j are well defined, that satisfy

$$\tau_\epsilon^{j-1} + \frac{1}{2}T_0 \leq \tau_\epsilon^j \leq \tau_\epsilon^{j-1} + \frac{3}{2}T_0 \, ,$$

and

$$\|X_\epsilon(\tau_\epsilon^N) - p\| \leq \sum_{k=0}^{N-1} \alpha_0^k R_1\epsilon + \alpha_0 R_0 \ .$$

Note that, since $0 < \alpha_0 < 1$, then $\sum_{k=0}^{N-1} \alpha_0^k \leq \sum_{k=0}^{\infty} \alpha_0^k = \frac{1}{1-\alpha_0}$. Therefore, since $\epsilon < \frac{(1-\alpha_0)^2 R_0}{R_1}$, we will obtain $X_\epsilon(\tau_\epsilon^N) \in U_\epsilon$.

Now, when $k = N+1$, we want to show that τ_ϵ^{N+1} exists and $X_\epsilon(\tau_\epsilon^{N+1}) \in U_\epsilon$. Let's consider the equation (106) with initial condition $X_\epsilon(\tau_\epsilon^N)$. By Gronwall inequality, we have

$$\sup_{\tau_\epsilon^N \leq s \leq \tau_\epsilon^N + T} \|X_\epsilon(s) - \phi^{\Delta_\epsilon^N}\left(X_\epsilon(\tau_\epsilon^N)\right)\| \leq C_0\epsilon \ ,$$

hence $\tau_\epsilon^N + \frac{1}{2}T_0 < \tau_\epsilon^{N+1} < \tau_\epsilon^N + \frac{3}{2}T_0$. Also, since

$$\|X_\epsilon(\tau_\epsilon^{N+1}) - \phi^{\Delta_\epsilon^N}\left(X_\epsilon(\tau_\epsilon^N)\right)\| \leq C_0\epsilon \ ,$$

similarly to the proof of Lemma 52, then

$$\|X_\epsilon(\tau_\epsilon^{N+1}) - P_0(X_\epsilon(\tau_\epsilon^N))\| \leq R_1\epsilon \ . \tag{127}$$

Using contractility of the Poincaré map as expressed by (111), and $X_\epsilon(\tau_\epsilon^N) \in U_\epsilon$, we have

$$\|P_0(X_\epsilon(\tau_\epsilon^N)) - p\| \leq \alpha_0 \|X_\epsilon(\tau_\epsilon^N) - p\| \ . \tag{128}$$

Combining inequalities (127) and (128), we obtain

$$\|X_\epsilon(\tau_\epsilon^{N+1}) - p\| \leq \|X_\epsilon(\tau_\epsilon^{N+1}) - P_0(X_\epsilon(\tau_\epsilon^N))\| + \|P_0(X_\epsilon(\tau_\epsilon^N)) - p\|$$

$$\leq R_1\epsilon + \alpha_0\|X_\epsilon(\tau_\epsilon^N) - p\| \leq \sum_{k=0}^{N} \alpha_0^k R_1\epsilon + \alpha_0 R_0 \ .$$

In particular, since $\epsilon < \frac{(1-\alpha_0)^2 R_0}{R_1}$, then $X(\tau_\epsilon^{N+1}) \in U_\epsilon$, and this completes the induction process.

Finally, since for all k, $\tau_\epsilon^{k-1} + \frac{1}{2}T_0 < \tau_\epsilon^k < \tau_\epsilon^{k-1} + \frac{3}{2}T_0$, then τ_ϵ^k must be in between two consecutive multiples of the period T_0. As a consequence of this, for any time t we can write $t = \tau_\epsilon^k + s$, for some k, and with $0 \leq s \leq T_0$. Using again Theorem 47, we then obtain

$$\sup_{t \geq 0} \|X_\epsilon(t)\| < \infty \ ,$$

which completes the proof.

Remark 17 *The main implication of Theorem 50 is that, although random perturbation in B will not be bounded for all times, the stochastic trajectories will remain within a tubular neighborhood of the deterministic limit cycle.*

Remark 18 *To illustrate the situation, consider a system (106) which is unambiguously representable in polar coordinates (for example, the van der Pol oscillator), use polar coordinates (r, θ) for the deterministic problem, and $(r_\epsilon(t), \theta_\epsilon(t))$ to model amplitude and phase in the stochastically perturbed version. Theorem 50 implies that -as long as the perturbation is selected from within the set B- the amplitude ρ_ϵ remains bounded:*

$$\sup_{t \geq 0, \ \omega \in B} |r_\epsilon(t) - r(t)| < \infty \ .$$

In turns, this helps explaining why we observe no catastrophic breakdown in cell-phone service, in agreement with practical experience.

At the same time, we must emphasize that the phase perturbation does become unbounded:

$$\sup_{t \geq 0, \ \omega \in B} |\theta_\epsilon(t) - \theta(t)| = \infty \ .$$

In turns, this helps explaining why we may (and do) lose cell phone connection during lengthy conversations; see details in [33].

To sum up, although perturbations occur in both amplitude and phase, there is a clear distinction among the two: in particular, the strong stability property of the deterministic limit cycle ensures that the amplitude perturbations remain bounded.

7.4 Connection with Fokker-Planck equations

In this section, we attempt deriving PDEs for the transition density function associated to the trajectories of (107), with $\omega \in B$. First, we review some known results and give needed notations.

7.4.1 Diffusion process and partial differential equations

For completeness, here we review the standard derivation of PDEs for diffusion processes; for details, see [59].

Consider a d-dimensional Markov family $\{X_t, \mathcal{F}_t\}$, which is a diffusion process with drift vector $b = (b_1, \ldots, b_d)$ and diffusion matrix $a = \{a_{ik}\}_{1 \leq i, k \leq d}$.

This means that for any $f \in C^2$, one has

$$\lim_{t \to 0} \frac{1}{t} [\mathrm{E}\, f(X_t | X_0 = x) - f(x)] = (Lf)(x), \ \forall x \in \mathbb{R}^d \ ,$$

where the infinitesimal operator L is given by

$$(Lf)(x) = \frac{1}{2} \sum_{i=1}^{d} \sum_{k=1}^{d} a_{ik}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_k} + \sum_{i=1}^{d} b_i(x) \frac{\partial f(x)}{\partial x_i} \ .$$

Suppose that the Markov family of X_t has a transition density function

$$P(X_t \in dy | X_0 = x) = \rho(t, x, y) dy; \ \forall x \in \mathbb{R}^d \ , \ t > 0 \ .$$

195

Then, $\rho(t, x, y)$ satisfies the forward Kolmogorov (Fokker-Planck) equation, for fixed $x \in \mathbb{R}^d$:

$$\rho_t(t, x, y) = L^* \rho(t, x, y); \ (t, y) \in (0, \infty) \times \mathbb{R}^d \ ,$$

and the backward Kolmogorov equation, for fixed $y \in \mathbb{R}^d$:

$$\rho_t(t, x, y) = L \rho(t, x, y); \ (t, x) \in (0, \infty) \times \mathbb{R}^d \ ,$$

where the adjoint operator L^* is given by

$$(L^* f)(x) = \frac{1}{2} \sum_{i=1}^{d} \sum_{k=1}^{d} \frac{\partial^2 (a_{ik}(x) f(x))}{\partial x_i \partial x_k} - \sum_{i=1}^{d} \frac{\partial (b_i(x) f(x))}{\partial x_i}, \ \forall x \in \mathbb{R}^d \ .$$

7.4.2 Killed diffusions

Let us also introduce the killed diffusion PDE, by considering the one dimensional diffusion process

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \ , \quad X_0 = x \ , \tag{129}$$

where b, σ are Lipschitz functions, and W_t is a standard Wiener process. Consider events set C:

$$C = \{\omega : \sup_{0 \le s \le t} |X_s| \le M_0\} \ .$$

Define the first exit time $\tau_C = \inf\{t : |X_t| = M_0\}$. The killed diffusion is defined as

$$X_t^C = \begin{cases} X_t, & \text{if } t < \tau_C \ ; \\ X_{\tau_C}, & \text{if } t \ge \tau_C \ . \end{cases} \tag{130}$$

(a) $X_{\tau_C} = M_0$ (b) $X_{\tau_C} = -M_0$
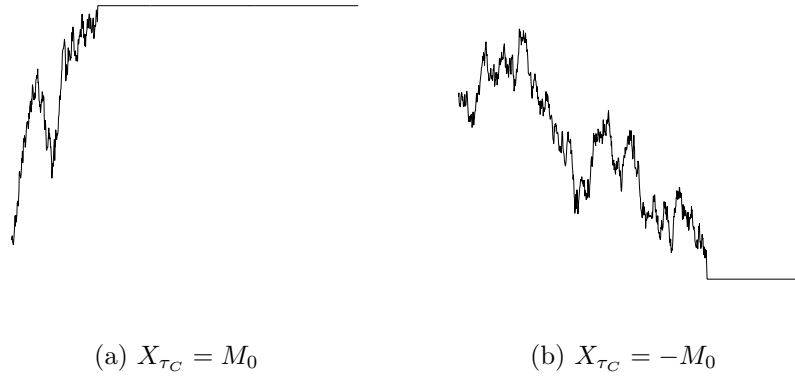
Figure 41: Killed diffusion

Consider the transition density function $\rho(t, x, y)$ of X_t^C:

$$\rho(t, x, y)dy = P(X_t^C \in dy | X_0 = x) \ .$$

In Lemma 53, we give the Fokker-Planck equation for the killed diffusion X_t^C, which is a PDE with vanishing boundary conditions on a finite interval. For the proof of Lemma 53, we refer to [52] and [90].

Lemma 53 *For fixed x, $\rho(t, x, y)$ is a weak solution of*

$$
\begin{cases}
\frac{\partial \rho}{\partial t} = -(b\rho)_y + \frac{1}{2}(\sigma^2 \rho)_{yy} \ , & |y| < M_0 \ , \\
\rho(t, x, y) = 0 \ , & |y| = M_0 \ , \\
\rho(0, x, y) = \delta_0(x - y) \ .
\end{cases}
$$

7.4.3 Derivation of PDEs

However, there are difficulties in following the above standard steps to derive the evolution conditioned on B, because:

- $X_\epsilon(t)$ is not a Markov process, since it depends both on values in the past and in the future; in fact, $X_\epsilon(t)$ depends on the full set of values in the time interval $(t - T, t + T)$.

197

Because of the above difficulty, we restrict to a subset of B which allows us to restart the process at certain times, and which is more amenable to analysis. As in (121), take $T = 2T_0$, where T_0 is the period of the deterministic limit cycle. Then, we consider the events set

$$B_0 = \{\omega \mid \sup_{0 \leq t \leq T_0} |W_{t+kT_0} - W_{kT_0}| \leq \frac{1}{2}M\} . \tag{131}$$

Clearly, B_0 in (131) is a subset of B in (121). With respect to B, B_0 has the advantage that, on each time interval of width T_0, the Wiener process increments are independent of that previous time interval of width T_0. Moreover, for first time interval, by introducing the absolute running maximum

$$M_t = \sup_{0 \leq s \leq t} |W_t| , \ t \leq T_0 ,$$

$(X_\epsilon(t), W_t, M_t)$ forms a Markov process, since condition B_0 is nothing but the restriction to those events for which M_t is bounded.

Motivated by the above, we will restrict to B_0. Then, on the first time interval, $(X_\epsilon(t), W_t, M_t)$ will be analyzed on separated subintervals $(0, t)$ and (t, T_0), where the first subinterval can be analyzed as a killed diffusion process and the second one can be analyzed by a standard PDE approach.

To be more precise, we will solve for the transition density function conditioned on events B_0:

$$\rho(t, X, X_\epsilon(0) \mid B_0)dX = P(X_\epsilon(t) \in dX \mid B_0, X_\epsilon(0)) , \tag{132}$$

where P represents probability function.

We divide our approach in three steps.

(i) From 0 to $t \leq T_0$, we introduce the new process $z_t = W_t$, and solve for the density function of $(X_\epsilon(t), z_t)$ at (X, z):

$$P(X_\epsilon(t) \in dX, z_t \in dz, M_t \leq \frac{1}{2}M \mid X_\epsilon(0)) . \tag{133}$$

198

As in killed diffusions, the corresponding equation is a PDE with vanishing boundary conditions.

(ii) For the remaining time from t to T_0, we will solve for the probability function

$$P(\sup_{t \le s \le T_0} |W_s| \le \frac{1}{2}M \mid W_t = z) .$$

By the Markov property of $(X_\epsilon(t), W_t, M_t)$, we will then form the transition density function on $(0, T_0)$.

(iii) Finally, for any time t, by the Markov property, we will derive the transition density function by connecting to the value obtained at the right-end point of the previous time interval.

We are now ready to give details of our approach. For our basic model (107), with dW_t from B_0 in (131), introduce the new process $z_t = z_0 + W_t$, so that equation (107) becomes

$$\begin{cases} dx_\epsilon = y_\epsilon dt , \\ dy_\epsilon = f(x_\epsilon, y_\epsilon)dt + \epsilon dW_t , \\ dz_t = dW_t . \end{cases} \tag{134}$$

For a test function $g(x, y, z) \in C^2(R^3)$, the infinitesimal operator corresponding to the process $(X_\epsilon(t), z_t)$ is

$$(Lg)(x, y, z) = yg_x + f(x, y)g_y + \frac{1}{2}(\epsilon^2 g_{yy} + 2\epsilon g_{yz} + g_{zz}) . \tag{135}$$

Now we begin the derivation on each time interval. Let

$$\tau_z = \inf\{t : |z_t| = \frac{1}{2}M\} , \quad M_t = \sup_{0 < s < t} |z_s| ,$$

and consider events up to N time intervals

$$B_0^N = \{\omega : \sup_{0 \le t \le T_0} |W_{t+kT_0} - W_{kT_0}| \le \frac{1}{2}M , \quad k \le N\} .$$

199

We first derive transition density for $X_\epsilon(t)$ conditioned on B_0^N:

$$u_N(t, X, X_\epsilon(0))dX = P(X_\epsilon(t) \in dX \mid X_\epsilon(0), B_0^N) , \quad \text{for } N = 1, 2, \ldots \quad (136)$$

As discussed above, this derivation goes through three steps.

Step one. We begin with transition density function on $(0, t)$, which plays a core role in this derivation. Since $(X_\epsilon(t), z_t)$ is also a diffusion process, condition transition density for $(X_\epsilon(t), z_t)$ with event $\{M_t \le \frac{1}{2}M\}$ is the same as a killed diffusion process, where we only cut off on the z_t part. To be more precisely, we define following killed diffusion by

$$(X_\epsilon^M(t), z_t^M) = \begin{cases} (X_\epsilon(t), z_t), & \text{if } t < \tau_z ; \\ (X_\epsilon(\tau_z), z_{\tau_z}), & \text{if } t \ge \tau_z . \end{cases} \quad (137)$$

The transition density of $(X_\epsilon^M(t), z_t^M)$ is the same as (133).

In details, consider B_0^1. For $\tau_z > t$ and fixed $X_\epsilon(0)$, we derive the transition density function u of process $(X_\epsilon(t), z_t)$ with events $\{M_t \le \frac{1}{2}M\}$:

$$\begin{aligned} u(t, X, z, X_\epsilon(0))dXdz &= P(X_\epsilon(t) \in dX, z_t \in dz, M_t \le \frac{1}{2}M \mid X_\epsilon(0), z_0 = 0) \\ &= P(X_\epsilon^M(t) \in dX, z_t^M \in dz \mid X_\epsilon(0), z_0 = 0) . \end{aligned}$$

For fixed $(X_\epsilon(0), z_0)$, we also denote $u(t, X, z, X_\epsilon) = u(t, X, z)$. The corresponding Fokker-Planck equation becomes

$$\begin{cases} \frac{\partial u}{\partial t} = L^* u, & (X, z) \in D , \\ u(t, X, z) = 0, \ |z| = \frac{1}{2}M , \\ u(0, X, z) = \delta_{(0,0,0)}(X - (x_\epsilon(0), y_\epsilon(0)), z - z_0) , \end{cases} \quad (138)$$

where

$$D = \mathbb{R} \times \mathbb{R} \times (-\frac{1}{2}M, \frac{1}{2}M) .$$

We delay justification of equation (138) until the end.

Step two. Here we discuss the event on (t, T_0). At time t, restricting to the events in B_0^1 is equivalent to requiring that the process z_t remains bounded up to time T_0.

Consider the probability of z_t remaining bounded until time t while starting at point z:

$$v(t, z) = P(\tau_z > t \mid z_0 = z) .$$

Here v represents probability of the events $\{\sup_{0 \leq s \leq t} |W_s + z| \leq \frac{1}{2}M\}$. Then (see [59]), $v(t, z)$ satisfies the following PDE:

$$\begin{cases} v_t = \frac{1}{2}v_{zz}, & -\frac{1}{2}M < z < \frac{1}{2}M , t > 0 , \\ v(t, \frac{1}{2}M) = v(t, -\frac{1}{2}M) = 0 , & t > 0 , \\ v(0, z) = 1, & -\frac{1}{2}M < z < \frac{1}{2}M . \end{cases} \tag{139}$$

Here, (139) can be solved. The remaining probability becomes

$$v(T_0 - t, z) = P(\sup_{t \leq s \leq T_0} |W_s| \leq \frac{1}{2}M \mid W_t = z) ,$$

and probability of B_0^1 is

$$v(T_0, 0) = P(B_0^1) .$$

Combining step one and two: Since $(X_\epsilon(t), z_t, M_t)$ forms a Markov process, we obtain the joint transition density function for $(X_\epsilon(t), z_t)$ with B_0^1:

$$\rho(t, X, z, X_\epsilon(0), B_0^1)dXdz = P(X_\epsilon(t) \in dX, z_t \in dz, B_0^1 \mid X_\epsilon(0), z_0 = 0) ,$$

which satisfies

$$\rho(t, X, z, X_\epsilon(0), B_0^1) = u(t, X, z)P(\sup_{t \leq s \leq T_0} |W_s| \leq \frac{1}{2}M \mid W_t = z) .$$

And the marginal density becomes

$$\rho(t, X, X_\epsilon(0), B_0^1) = \int_{-\frac{1}{2}M}^{\frac{1}{2}M} \rho(t, X, z, X_\epsilon(0), B_0^1)dz .$$

The derivation becomes as following: for the first time interval $(0, T_0)$, recall u_1 defined in (136) represents transition density function for $X_\epsilon(t)$ conditioned on B_0^1, which statisfies

$$u_1(t, X, X_\epsilon(0)) = \frac{\rho(t, X, X_\epsilon(0), B_0^1)}{P(B_0^1)}$$

$$= \frac{\int_{-\frac{1}{2}M}^{\frac{1}{2}M} u(t, X, z) v(T_0 - t, z) dz}{v(T_0, 0)} ,$$

where u satisfies equation (138) with $X_\epsilon(0)$, $z_0 = 0$ and v is the solution of equation (139).

Step three. "Refreshing". X_t under B_0 can be seen as a refreshed process at each time kT_0. And by applying the Markov property, we can derive general transition density function for $X_\epsilon(t)$.

Consider the events set B_0^{N+1}. We denote w as the transition density function for $(X_\epsilon(t), z_t)$ with events B_t:

$$w(t, X, z, X_\epsilon(0), z_0) dX dz = P(X_\epsilon(t) \in dX, z(t) \in dz, B_t \mid X_\epsilon(0), z_0) ,$$

where

$$B_t = \left\{ \omega : \sup_{\substack{0 < s < T_0 \\ s+kT_0 \le t}} |W_{s+kT_0} - W_{kT_0}| \le M \right\} .$$

Again, for fixed $\bar{X}_0$, we denote $w = w(t, X, z)$, and by Chapman−Kolmogorov equation

$$w(t, X, z) = \int_{\mathbb{R}^2} \dots \int_{\mathbb{R}^2} u(t, \bar{X}, \bar{X}_N) \prod_{i=0}^{N-1} u(T_0, \bar{X}_{i+1}, \bar{X}_i) d\bar{X}_1 d\bar{X}_2 \dots d\bar{X}_N ,$$

where we used the notation $\bar{X}_i = (X_i, z_i)$, and $\bar{X}_0 = (X_\epsilon(0), 0)$. Combining events from time t to $(N+1)T_0$, we have

$$u_{N+1}(t, X, X_\epsilon(0)) = \frac{\int_{-\frac{1}{2}M}^{\frac{1}{2}M} w(t, X, z) v((N+1)T_0 - t, z) dz}{P(B_0^{N+1})}$$

$$= \frac{\int_{-\frac{1}{2}M}^{\frac{1}{2}M} w(t, X, z) v((N+1)T_0 - t, z) dz}{v(T_0, 0)^{N+1}} .$$

Arbitrary t. From the independent increments property of B_0 for each time interval, we can derive the transition density function for $X_\epsilon(t)$ conditioned on B_0. Indeed, for any t, there exists $N = 0, 1, \ldots$, such that $t \in [NT_0, (N+1)T_0]$. Then, we simply have

$$\rho(t, X, X_\epsilon(0) \mid B_0) = u_{N+1}(t, X, X_\epsilon(0)) .$$

Finally, we justfy equation (138).

Proof of (138). The basic approach we use is standard; e.g., see [52] and [89].

The boundary conditions can be given as $u(0, x, y, z) = \delta_{(0,0,0)}(x - x_\epsilon(0), y - y_\epsilon(0), z - z(0))$ and $u(t, x, y, \pm \frac{1}{2}M) = 0$. Next, we follow the same steps used to derive the Fokker-Planck equation for the diffusion process.

To simplify notation, denote $Y = (x, y, z)$ and $Y_t = (x_\epsilon(t), y_\epsilon(t), z(t))$. Consider any test function $h(x, y, z) = h(Y) \in C^2$ with compact support. Then,

$$\int_D h(Y) \frac{\partial u(t, Y)}{\partial t} dY = \int_D h(Y) \lim_{\Delta t \to 0} \frac{u(t + \Delta t, Y) - u(t, Y)}{\Delta t} dY , \qquad (140)$$

where $D = \mathbb{R} \times \mathbb{R} \times (-\frac{1}{2}M, \frac{1}{2}M)$. Again, consider the process (Y_t, M_t). Since (Y_t, M_t) is a Markov process, if we denote its density function with $\bar{\rho}$, which is defined by

$$\bar{\rho}(t, Y, Y(0))dY = P(Y_t \in dY, M_t \leq \frac{1}{2}M | Y(0)) ,$$

Therefore the Chapman$-$Kolmogorov equation implies

$$u(t + \Delta t, Y) = \bar{\rho}(t + \Delta t, Y, Y(0))$$
$$= \int_D u(t, Z) \bar{\rho}(\Delta t, Y, Z) dZ .$$

Above, the last equality comes from the Markov property. Hence, equation (140) becomes

$$
\begin{aligned}
&\int_D h(Y) \frac{\partial u(t, Y)}{\partial t} dY \\
&= \int_D h(Y) \lim_{\Delta t \to 0} \frac{\int_D u(t, Z) \bar{\rho}(\Delta t, Y, Z) dZ - u(t, Y)}{\Delta t} dY \\
&= \lim_{\Delta t \to 0} \frac{\int_D \int_D h(Y) u(t, Z) \bar{\rho}(\Delta t, Y, Z) dZ dY - \int_D h(Y) u(t, Y) dY}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\int_D \int_D h(Y) u(t, Z) \bar{\rho}(\Delta t, Y, Z) dY dZ - \int_D h(Z) u(t, Z) dZ}{\Delta t} \\
&= \int_D \lim_{\Delta t \to 0} \frac{\int_D h(Y) \bar{\rho}(\Delta t, Y, Z) dY - h(Z)}{\Delta t} u(t, Z) dZ \ .
\end{aligned}
\tag{141}
$$

where the second and last equalities are justified by the dominated convergence theorem, and the third equality comes from Fubini's theorem. Since $\mathrm{E}(\mathbf{1}_{\{\tau_z \le \Delta t\}}) = o(\Delta t)$ and h is a bounded function, then

$$
\begin{aligned}
&\lim_{\Delta t \to 0} \frac{\int_D h(Y) \bar{\rho}(\Delta t, Y, Z) dY - h(Z)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\mathrm{E}\, h(Y_{\Delta t}) \mathbf{1}_{\{\tau_z > \Delta t\}} - h(Z)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \left[\frac{\mathrm{E}\, h(Y_{\Delta t}) - h(Z)}{\Delta t} - \frac{\mathrm{E}\, h(Y_{\Delta t}) \mathbf{1}_{\{\tau_z \le \Delta t\}}}{\Delta t} \right] = Lh(Z) \ ,
\end{aligned}
$$

where L is the infinitesimal operator defined by (135). Hence (141) becomes

$$
\int_D h(Y) \frac{\partial u(t, Y)}{\partial t} dy = \int_D Lh(Z) u(t, Z) dZ \ .
$$

Integrating by parts, using $u(t, Z) = 0$ on the boundary and letting $Z = Y$ on the right-hand-side, we then obtain

$$
\int_D h(Y) \frac{\partial u(t, Y)}{\partial t} dY = \int_D h(Y) L^* u(t, Y) dY \ ,
$$

which gives the equation (138).

7.5   Conclusions

In this work, motivated by practical observations (real world phenomena, laboratory experiments, and numerical simulations) on typical engineering circuitries, we reconsidered what model of noise is appropriate for the mathematical modeling of

stochastic perturbation of second order systems of differential equations that admit stable limit cycles. Whereas classical models consider stochastic DEs where perturbations come from standard Brownian motion paths, we restricted the class of allowed disturbances, to avoid pumping infinite energy into the system through the noise. In essence, our new model consists in selecting those Brownian paths that have bounded increments in finite time.

Of course, there are new challenges when one gives up familiar ground, such as white noise perturbations, and indeed we have encountered technical difficulties especially insofar as obtaining viable expression for the transition density function. However, by selecting the allowed perturbations from within our proposed event set, we were able to adopt many classical tools from dynamical systems, and show some interesting mathematical results, that further appear to be more in tune with practically observed circuitry behaviors.

Relative to our set of allowed stochastic perturbations, our main results have been the following.

(i) We proved global boundedness of the stochastic trajectories, and we showed that they remain (for small values of the parameter ϵ appearing in front of the perturbation term) in the neighborhood of the deterministic limit cycle.

(ii) We proposed, and ensured the existence, of stochastic Poincaré map(s) as a point-to-distribution map, and further introduced three point-to-point Poincaré maps: first, last, and average return maps.

(iii) We associated the study of transition densities to a pair of PDEs.

REFERENCES

[1] AKIN, E., *The geometry of population genetics*, vol. 280. Springer Science & Business Media, 1979.

[2] AMBROSIO, L., *Lecture notes on optimal transport problems.* Springer, 2003.

[3] AMBROSIO, L. and GANGBO, W., "Hamiltonian ODEs in the Wasserstein space of probability measures," *Communications on Pure and Applied Mathematics*, pp. 18–53, 2008.

[4] AMBROSIO, L., GIGLI, N., and SAVARÉ, G., *Gradient flows: in metric spaces and in the space of probability measures.* Springer Science & Business Media, 2006.

[5] ANCONA, F. and BRESSAN, A., "Patchy vector fields and asymptotic stabilization," *ESAIM: Control, Optimization and Calculus of Variations*, vol. 4, pp. 445–471, 1999.

[6] ANCONA, F. and BRESSAN, A., "Nearly time optimal stabilizing patchy feedbacks," in *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, vol. 24, pp. 279–310, Elsevier, 2007.

[7] ARNOLD., L., *Stochastic Differential Equations: Theory and Applications.* John Wiley & Sons, 1974.

[8] ARNOLD., L., *Random Dynamical Systems, 2nd Edition.* Springer-Verlag, Berlin, 2003.

[9] BALAKRISHNAN, A. V. and NEUSTADT, L. W., *Computing methods in optimization problems: proceedings.* Academic Press, 1964.

[10] BAXENDALE, P. H., "Lyapunov exponents and stability for the stochastic Duffing-van der Pol oscillator," in *IUTAM Symposium on Nonlinear Stochastic Dynamics*, pp. 125–135, Springer, 2003.

[11] BAXENDALE, P. H., "Stochastic averaging and asymptotic behavior of the stochastic Duffing–van der Pol equation," *Stochastic Processes and Their Applications*, vol. 113, no. 2, pp. 235–272, 2004.

[12] BELLMAN, R., "Dynamic programming and the smoothing problem," *Management Science*, vol. 3, no. 1, pp. 111–113, 1956.

[13] BENEDETTO, D., CAGLIOTI, E., CARRILLO, J. A., and PULVIRENTI, M., "A non-Maxwellian steady distribution for one-dimensional granular media," *Journal of statistical physics*, vol. 91, no. 5-6, pp. 979–990, 1998.

[14] BLANCHET, A. and CARLIER, G., "Optimal transport and Cournot-Nash equilibria," *arXiv preprint arXiv:1206.6571*, 2012.

[15] BLANCHET, A. and CARLIER, G., "From Nash to Cournot–Nash equilibria via the Monge–Kantorovich problem," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 372, no. 2028, p. 20130398, 2014.

[16] BOBKOV, S. G. and TETALI, P., "Modified logarithmic Sobolev inequalities in discrete settings," *Journal of Theoretical Probability*, vol. 19, no. 2, pp. 289–336, 2006.

[17] BONNANS, J. F. and HERMANT, A., "Stability and sensitivity analysis for optimal control problems with a first-order state constraint and application to continuation methods," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 14, no. 04, pp. 825–863, 2008.

[18] BONNARD, B., FAUBOURG, L., LAUNAY, G., and TRÉLAT, E., "Optimal control with state constraints and the space shuttle re-entry problem," *Journal of Dynamical and Control Systems*, vol. 9, no. 2, pp. 155–199, 2003.

[19] BRESSAN, A., "Noncooperative differential games. a tutorial," 2010.

[20] BRESSAN, A. and HONG, Y., "Optimal control problems on stratified domains,"

[21] BRYSON, A. E. and HO, Y.-C., *Applied optimal control: optimization, estimation, and control.* Taylor & Francis Group, 1975.

[22] CARDALIAGUET, P., DELARUE, F., LASRY, J.-M., and LIONS, P.-L., "The master equation and the convergence problem in Mean field games," *arXiv preprint arXiv:1509.02505*, 2015.

[23] CARRILLO, J. A., MCCANN, R. J., VILLANI, C., and OTHERS, "Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates," *Revista Matematica Iberoamericana*, vol. 19, no. 3, pp. 971–1018, 2003.

[24] CATTIAUX, P., GUILLIN, A., and MALRIEU, F., "Probabilistic approach for granular media equations in the non-uniformly convex case," *Probability theory and related fields*, vol. 140, no. 1-2, pp. 19–40, 2008.

[25] CHE, R., HUANG, W., LI, Y., and TETALI, P., "Convergence to global equilibrium for Fokker-Planck equations on a graph and Talagrand-type inequalities," *arXiv preprint arXiv:1409.0711*, 2014.

[26] CHEN, K., SUN, Z., YIN, S., WANG, Y., and WANG, X.-R., "Deformation of limit cycle under perturbations," *Superlattices and microstructures*, vol. 37, no. 3, pp. 185–191, 2005.

[27] Chow, S.-N., Huang, W., Li, Y., and Zhou, H., "Fokker–Planck equations for a free energy functional or Markov process on a graph," *Archive for Rational Mechanics and Analysis*, vol. 203, no. 3, pp. 969–1008, 2012.

[28] Chow, S.-N., Li, W., and Zhou, H., "Optimal path in dynamic environments by Method of evolving junctions," *In preparation.*

[29] Chow, S.-N., Lu, J., and Zhou, H., "Fast numerical methods based on SDEs for several problems related to the shortest path," *accepted to MAA special issue in honor of Stanley Osher on his 70th birthday.*

[30] Chow, S.-N., Lu, J., and Zhou, H., "Finding the shortest path by evolving junctions on obstacle boundaries (E-JOB): An initial value ODE's approach," *Applied and Computational Harmonic Analysis*, 2012.

[31] Chow, S.-N., Lu, J., and Zhou, H., "Shortest path amid 3-d polyhedral obstacles," *SIAM scientific computing*, 2012.

[32] Chow, S.-N., Yang, T.-S., and Zhou, H., "Global optimizations by intermittent diffusion," *International Journal of Bifurcation and Chaos*, 2013.

[33] Chow, S.-N. and Zhou, H.-M., "An analysis of phase noise and Fokker–Planck equations," *Journal of Differential Equations*, vol. 234, no. 2, pp. 391–411, 2007.

[34] Chow, S.-N., Dieci, L., Li, W., and Zhou, H., "A new numerical scheme for Fokker-Planck equations based optimal transport," *in preparation.*

[35] Chow, S.-N., Li, W., Lu, J., and Zhou, H., "Method of evolving junctions: A new approach to optimal control with constraints,," *submitted.*

[36] Chow, S.-N., Li, W., and Zhou, H., "Fokker-Planck equations on strategy graphs and population games," *in preparation.*

[37] Chow, S.-N., Li, W., and Zhou, H., "A Newton-like algorithm for the shortest path based on the method of evolving junctions," *accepted in Communications in mathematical sciences.*

[38] Chow, S.-N., Li, W., and Zhou, H., "Optimal transport on finite graphs," *in preparation.*

[39] Degond, P., Liu, J.-G., and Ringhofer, C., "Large-scale dynamics of Mean-field games driven by local Nash equilibria," *Journal of Nonlinear Science*, vol. 24, no. 1, pp. 93–115, 2014.

[40] Dennis, J. E. and Schnabel, R. B., *Numerical methods for unconstrained optimization and nonlinear equations*, vol. 16. Society for Industrial and Applied Mathematics, 1987.

[41] DIACONIS, P., SALOFF-COSTE, L., and OTHERS, "Logarithmic Sobolev inequalities for finite Markov chains," *The Annals of Applied Probability*, vol. 6, no. 3, pp. 695–750, 1996.

[42] DIECI, L., LI, W., and ZHOU, H., "A new model for realistic random perturbations of stochastic oscillators," *submitted.*

[43] ERBAR, M. and MAAS, J., "Ricci curvature of finite Markov chains via convexity of the entropy," *Archive for Rational Mechanics and Analysis*, vol. 206, no. 3, pp. 997–1038, 2012.

[44] FACCHINEI, F. and KANZOW, C., "Generalized Nash equilibrium problems," *4OR*, vol. 5, no. 3, pp. 173–210, 2007.

[45] FLETCHER, R., *Practical methods of optimization.* John Wiley & Sons, 2013.

[46] FREIDLIN, M. and WENTZELL, A. D., *Random perturbations of dynamical systems.* Springer, 2012.

[47] FUJIMURA, K. and SAMET, H., "Planning a time-minimal motion among moving obstacles," *Algorithmica*, vol. 10, no. 1, pp. 41–63, 1993.

[48] GANGBO, W., NGUYEN, T., TUDORASCU, A., and OTHERS, "Hamilton-Jacobi equations in the Wasserstein space," *Methods and Applications of Analysis*, pp. 155–184, 2008.

[49] GILL, P. E., MURRAY, W., and WRIGHT, M. H., *Practical optimization.* Academic press, 1981.

[50] HAJIMIRI, A., LIMOTYRAKIS, S., and LEE, T. H., "Jitter and phase noise in ring oscillators," *Solid-State Circuits, IEEE Journal of*, vol. 34, no. 6, pp. 790–804, 1999.

[51] HARTL, R. F., SETHI, S. P., and VICKSON, R. G., "A survey of the maximum principles for optimal control problems with state constraints," *SIAM review*, vol. 37, no. 2, pp. 181–218, 1995.

[52] HENKEL, H., *Range-based parameter estimation in diffusion models.* PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2010.

[53] HITCZENKO, P. and MEDVEDEV, G. S., "The Poincaré map of randomly perturbed periodic motion," *Journal of Nonlinear Science*, vol. 23, no. 5, pp. 835–861, 2013.

[54] HOFBAUER, J. and SIGMUND, K., *The theory of evolution and dynamical systems: mathematical aspects of selection.* Cambridge University Press Cambridge, 1988.

[55] HOFBAUER, J. and SIGMUND, K., "Evolutionary game dynamics," *Bulletin of the American Mathematical Society*, vol. 40, no. 4, pp. 479–519, 2003.

[56] HSU, E. P., *Stochastic analysis on manifolds*, vol. 38. American Mathematical Soc., 2002.

[57] IOFFE, A. D., TIKHOMIROV, V. M., and MAKOWSKI, K., *Theory of extremal problems*, vol. 6. Elsevier, 2009.

[58] JORDAN, R., KINDERLEHRER, D., and OTTO, F., "The variational formulation of the Fokker–Planck equation," *SIAM journal on mathematical analysis*, vol. 29, no. 1, pp. 1–17, 1998.

[59] KARATZAS, I. and SHREVE, S., *Brownian motion and stochastic calculus*, vol. 113. Springer Science & Business Media, 2012.

[60] KILLINGBACK, T. and DOEBELI, M., "Spatial evolutionary game theory: Hawks and Doves revisited," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 263, no. 1374, pp. 1135–1144, 1996.

[61] LASRY, J.-M. and LIONS, P.-L., "Mean field games," *Japanese Journal of Mathematics*, vol. 2, no. 1, pp. 229–260, 2007.

[62] LI, W., CHOW, S., MAGNUS, E., LU, J., and ZHOU, H., "Optimal path in dynamic environments," *in preparation*.

[63] LIU, Y., ITO, S., LEE, H., and TEO, K., "Semi-infinite programming approach to continuously-constrained linear-quadratic optimal control problems," *Journal of Optimization Theory and Applications*, vol. 108, no. 3, pp. 617–632, 2001.

[64] LIVNE, E., MINEAU, D., BRYSON, A., and DENHAM, W., "Optimal programming problems with inequality constraints. ii-solution by steepest-ascent," *AIAA Journal*, vol. 2, no. 1, pp. 25–34, 1964.

[65] LU, J., DIAZ-MERCADO, Y., EGERSTEDT, M., ZHOU, H., and CHOW, S.-N., "Shortest paths through 3-dimensional cluttered environments," *accepted to International Conference on robotics and automation*, 2014.

[66] MAAS, J., "Gradient flows of the entropy for finite Markov chains," *Journal of Functional Analysis*, vol. 261, no. 8, pp. 2250–2292, 2011.

[67] MALANOWSKI, K., "Stability and sensitivity of solutions to nonlinear optimal control problems," *Applied Mathematics and Optimization*, vol. 32, no. 2, pp. 111–141, 1995.

[68] MARKOWICH, P. and VILLANI, C., "On the trend to equilibrium for the Fokker-Planck equation: An interplay between physics and functional analysis," in *Physics and Functional Analysis, Matematica Contemporanea (SBM) 19*, Citeseer, 1999.

[69] MATSUI, A., "Best response dynamics and socially stable strategies," *Journal of Economic Theory*, vol. 57, no. 2, pp. 343–362, 1992.

[70] MC REYNOLDS, S. R. and BRYSON JR, A. E., "A successive sweep method for solving optimal programming problems.," tech. rep., DTIC Document, 1965.

[71] MIELKE, A., "Geodesic convexity of the relative entropy in reversible Markov chains," *Calculus of Variations and Partial Differential Equations*, vol. 48, no. 1-2, pp. 1–31, 2013.

[72] MONDERER, D. and SHAPLEY, L. S., "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[73] NASH, J. F. and OTHERS, "Equilibrium points in n-person games," *Proceedings of the national academy of sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[74] NAVASCA, C. and KRENER, A. J., "Patchy solutions of Hamilton-Jacobi-Bellman partial differential equations," in *Modeling, estimation and control*, pp. 251–270, Springer, 2007.

[75] NOCEDAL, J. and WRIGHT, S. J., *Numerical optimization.* Springer Science+ Business Media, 2006.

[76] NOWAK, M. A., *Evolutionary dynamics.* Harvard University Press, 2006.

[77] OLLIVIER, Y., "Ricci curvature of Markov chains on metric spaces," *Journal of Functional Analysis*, vol. 256, no. 3, pp. 810–864, 2009.

[78] OSHER, S. and FEDKIW, R. P., *Level set methods and dynamic implicit surfaces.* Springer Verlag, 2003.

[79] OTTO, F., "The geometry of dissipative evolution equations: the porous medium equation," 2001.

[80] OTTO, F. and VILLANI, C., "Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality," *Journal of Functional Analysis*, vol. 173, no. 2, pp. 361–400, 2000.

[81] PONTRIAGIN, L. S., *The mathematical theory of optimal processes*, vol. 4. CRC Press, 1962.

[82] POTSCHKA, A., *Handling path constraints in a direct multiple shooting method for optimal control problems.* PhD thesis, Diploma thesis, Universität Heidelberg, 2006.

[83] ROTHAUS, O., "Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities," *Journal of functional analysis*, vol. 42, no. 1, pp. 102–109, 1981.

[84] SANDHOLM, W. H., "Evolutionary game theory," in *Encyclopedia of Complexity and Systems Science*, pp. 3176–3205, Springer, 2009.

[85] SCHENK-HOPPÉ, K. R., "Deterministic and stochastic Duffing-van der Pol oscillators are non-explosive," *Zeitschrift für angewandte Mathematik und Physik ZAMP*, vol. 47, no. 5, pp. 740–759, 1996.

[86] SIGMUND, K. and NOWAK, M. A., "Evolutionary game theory," *Current Biology*, vol. 9, no. 14, pp. R503–R505, 1999.

[87] SMITH, M. J., "The stability of a dynamic model of traffic assignment-an application of a method of Lyapunov," *Transportation Science*, vol. 18, no. 3, pp. 245–252, 1984.

[88] SPEYER, J. L., MEHRA, R. K., and BRYSON JR, A. E., "The separate computation of arcs for optimal flight paths with state variable inequality constraints," *Advanced Problems and Methods for space flight optimization*, pp. 53–68, 1969.

[89] STROOCK, D. W., *Partial differential equations for probabilists.* No. 112.

[90] STROOCK, D. W. and VARADHAN, S. S., *Multidimensional diffusion processes.* Springer, 2007.

[91] VAN DEN BERG, J. and OVERMARS, M., "Planning time-minimal safe paths amidst unpredictably moving obstacles," *The International Journal of Robotics Research*, 2008.

[92] VARADHAN, S. R., "Large deviations and applications," in *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pp. 1–49, Springer, 1988.

[93] VILLANI, C., "A review of mathematical topics in collisional kinetic theory," *Handbook of mathematical fluid dynamics*, vol. 1, pp. 71–305, 2002.

[94] VILLANI, C., *Topics in optimal transportation.* No. 58, American Mathematical Soc., 2003.

[95] VILLANI, C., *Optimal transport: old and new*, vol. 338. Springer Science & Business Media, 2008.

[96] VON NEUMANN, J. and MORGENSTERN, O., *Theory of games and economic behavior (60th Anniversary Commemorative Edition).* Princeton university press, 2007.

[97] YANO, K., "Some remarks on tensor fields and curvature," *Annals of Mathematics*, pp. 328–347, 1952.

[98] YANO, K. and OTHERS, "Some integral formulas and their applications.," *The Michigan Mathematical Journal*, vol. 5, no. 1, pp. 63–73, 1958.

VITA

Wuchen Li was born in Tancheng, Linyi, Shandong, China on April 6 1988. At age 17, he went to Shandong university in Jinan, China, and graduated with a Bachelor of Science in mathematics in June 2009. From September 2009 to May 2011, he studied mathematics in Chinese academy of Science. In August of 2011, he came to Atlanta and joined the Georgia Institute of Technology to work with Professors Luca Dieci and Haomin Zhou in the School of Mathematics.