Rare Event

in

Stochastic Systems

CAO, Yu SID: 52640130

Department of Mathematics,

City University of Hong Kong

Supervisor: Dr Xiang Zhou

Date: March 27, 2015

Contents

List of Figures

List of Tables

Abstract

This report purports to study transition problem in stochastic dynamical systems under small noise (rare event), mainly by using transition path theory (TPT). TPT is a recently developed mathematical method to describe the transition mechanics in stochastic dynamical systems and it has shown to be an effective tool with wide application in biology, chemical physics and etc: dynamical bottleneck, the region that limits the transition rate, could be detected and studied. Representative dominant transition path, a concept built under the TPT, will be briefly introduced. In addition, most probable exit path will also be discussed as comparison, which is proposed and defined under the framework of large deviation principle (LDP) and which can also be understood by introducing Lagrange multiplier in stochastic mapping. What is more, numerical results on classical models by using TPT will be presented with illustration of how TPT could be applied to shed light on transition mechanics.

Keywords: Stochastic dynamical system, transition path theory, most probable exit path, representative dominant transition path, diffusion process, Markov chain.

1 Introduction

1.1 Background of transition problem

Transition (reaction) rate problem has always been challenging and important in many fields such as chemical physics, biology and engineering. Its main task is to calculate the transition rate for some complex systems and perhaps to understand the mechanics of chemical reactions, nucleation and et al.

It dated back to Hoff and Arrhenius's description on escape rate

$$k = \nu \exp\left(-\frac{1}{k_B T} E\right) = \nu \exp\left(-\beta E\right) \tag{1.1}$$

where ν is the prefactor, k_B is the Boltzmann constant, T is the temperature and E is the energy barrier, $\beta = \frac{1}{k_B T}$ [6]. At low temperature, the transition rate decays exponentially. The mean transition time

$$\bar{t} \propto \frac{1}{k} \propto \exp\left(\frac{1}{k_B T} E\right) \tag{1.2}$$

is exponentially increasing when $T \to 0$. The long separation of time between metastable states makes the observation of a complete transition almost impossible in real physical systems and makes the cost of direct numerical simulation (DNS) too expensive to perform. That is exactly where the term *rare event* comes from.

Many theories and methods have been developed, intending to compute the transition rate and to describe the transition mechanics, such as transition state theory (TST), transition path sampling (TPS). In the framework of large deviation, Freidlin-Wentzell theory has worked out the most probable exit path (MPEP) in diffusion process, which minimizes the action functional, in the sense of infinitesimally small noise; additionally, it has shown the exponentially increasing mean first exit time from the asymptotically stable position to the boundary of its basin of attraction [5, 12]. Such optimal escape

path could be computed by string method, if the drift term in the diffusion process is a conservative vector field [3]. MPEP has also been discussed for both continuous-time stochastic dynamical systems and random maps in [1].

Another important recently developed theory to deal with transition rate is transition path theory (TPT). Transition path theory focuses on studying mechanics of stochastic dynamical systems, such as diffusion process [4, 11], Markov jump processes [8]. Its main aim is to offer statistical description of transition in stochastic dynamical systems. Its application varies from protein folding problem [10], complex network [2], data analysis [8] and etc. Some examples in explaining applications of transition path theory can be found in, e.g. [9].

1.2 Transition problems in dynamical systems

For a given ergodic stochastic dynamical system $\{X_t\}_{t\in T}$ defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with state space S, where T is the parameter set for time (e.g. discrete-time Markov chain (DTMC) has $T = \mathbb{Z}$, Markov jump process (MJP) $T = \mathbb{R}$ and diffusion process $T = \mathbb{R}$), we prescribe two disjoint subsets A and B of state space S ($A, B \subset S$, $A \cap B \neq \emptyset$). Normally, set A is chosen to be the position or region with certain metastability and set B the position or region which is far away from set A. A could be regarded as the normal state and B the failure state. In the context of chemical reaction, A could represent the reactant state and B the product state. We wonder how the system transits from normal state to failure state, or say how the chemical reaction takes place. To make this problem more realistic, assume $A \cup B \subsetneq S$, which guarantees the existence of transient states in S.

The definition of reactive trajectory is given below:

Definition 1. *1. If $T = \mathbb{R}$, a function φ defined on a time interval $[a, b]$ is called a reactive trajectory of length $b - a$ iff $\varphi(a) \in A$, $\varphi(b) \in B$, and $\varphi(t) \notin A \cup B$ for all $a < t < b$.*

2. *If $T = \mathbb{Z}$, a function φ defined on a set of consecutive integers $\{n, n+1, \cdots, n+l\}$ is called a reactive trajectory of length l iff $\varphi(n) \in A$, $\varphi(n+l) \in B$, and $\varphi(n+i) \notin A \cup B$ for any integer $0 < i < l$.*

T might be chosen as other sets in theory, but those cases will not be considered in this report.

We ask: in the long run,

1. what is the transition rate k^{AB} for transition from set A to set B?
2. which position in set A is the most probable exit point in the transition and which position in set B is the most probable entering point?
3. how can we get more insight into the transition mechanics?
4. among all reactive trajectories, which one is the most probable one? Or, which path dominates in the transition?

Let $\mathscr{S}$ be the space of all possible reactive trajectories. Consider the case with $T = \mathbb{Z}$ and $S = \mathbb{N}$ (e.g. DTMC on countable state space). If there is a measure $\mathcal{P}$ defined on $\mathscr{S}$ which gives the time average of occurrence of a set of paths in $\mathscr{S}$, that is, for a set of paths $C \subset \mathscr{S}$,

$$\mathcal{P}(C) := \lim_{n\to\infty} \frac{\#\text{ of complete paths in } C \text{ within time } [-n, n] \cap T}{2n+1}, \tag{1.3}$$

then question 1 is to calculate $\mathcal{P}(\mathscr{S})$. Question 2 is to calculate

$$\arg\max_{x\in A} \mathbb{E}_\varphi[\mathbf{1}_x(\varphi(0))] \tag{1.4}$$

$$\arg\max_{y\in B} \mathbb{E}_\varphi[\mathbf{1}_y(\varphi(\text{length of } \varphi))] \tag{1.5}$$

where $\mathbf{1}_x(\cdot)$ is the indicator function (here and below), $\varphi(0)$ means the starting position of path φ, $\varphi(\text{length of } \varphi)$ means the ending point of path φ. Question 4 is to optimize

φ on space $\mathscr{S}$,

$$\max_{\varphi} \mathcal{P}(\varphi). \tag{1.6}$$

Question 3 might be more abstract, philosophical and problem-dependent; it depends on how we interpret the word *mechanics*.

In general, the above illustration may not be well-defined mathematically; it just gives the framework and some idea of what are the problems we are dealing with; from my understanding, it could be adapted to many specific problems with some modification. In addition, the above definition seems to be straightforward, but it may not be viable in practice; later on, it will be explained how to calculate these statistical properties in questions 1-3 by using transition path theory.

In this report, I will review the transition path theory for both diffusion process and Markov chains in section 2, together with representative dominant transition path, firstly proposed in [8]. In section 3, it will be demonstrated how transition path theory works by utilizing some classical models, some of which have been discussed and studied in [9]. In section 4, a short review on MPEP and its comparison with TPT will be provided. In the last section, I will summarize the transition problem and, in particular, transition path theory.

2 Transition Path Theory

2.1 Preliminary definitions

In this section, we will review the transition path theory (TPT) being applied into different stochastic dynamical systems; these different versions are basically following the same idea with slight modification. In the following, some shared concepts are introduced in a general form to avoid duplicative definition.

Definition 2. *Definition of hitting time to set $A \cup B$:*

1. *If $T = \mathbb{R}$, hitting times are defined as follow,*

$$t^{AB,+}(t) := \inf\{s \geq t : X_s \in A \cup B\},$$
$$t^{AB,-}(t) := \sup\{s \leq t : X_s \in A \cup B\}.$$

2. *If $T = \mathbb{Z}$, hitting times are defined as follow,*

$$t^{AB,+}(n) := \inf\{m \geq n : X_m \in A \cup B\},$$
$$t^{AB,-}(n) := \sup\{m \leq n : X_m \in A \cup B\}.$$

Definition 3. *Forward committor function $q^+(x) : S \to [0, 1]$ is defined as the probability of reaching set B before hitting set A,*

$$q^+(x) := \mathbb{P}_x[X_{t^{AB,+}(0)} \in B].$$

Backward committor function $q^-(x) : S \to [0, 1]$ is defined as the probability of coming from set A instead of set B,

$$q^-(x) := \mathbb{P}_x[X_{t^{AB,-}(0)} \in A].$$

It could be deduced immediately from definition that $q^+(x) = 1$ if $x \in B$ and $q^+(x) = 0$

if $x \in A$; $q^-(x) = 0$ if $x \in B$ and $q^-(x) = 1$ if $x \in A$. For the time-reversible process, that is to say, we cannot distinguish the forward process and the backward process based on the statistics that we can obtain,

$$\begin{aligned} q^+(x) &= \mathbb{P}_x[X_{t^{AB,+}(0)} \in B] \\ &= 1 - \mathbb{P}_x[X_{t^{AB,+}(0)} \in A] \\ &= 1 - \mathbb{P}_x[X_{t^{AB,-}(0)} \in A] \\ &= 1 - q^-(x). \end{aligned} \tag{2.1}$$

This is, however, not true in general.

Definition 4. *Forward operator $\mathcal{L}$ is defined as*

$$(\mathcal{L}f)(x) := \lim_{s \downarrow 0} \frac{\mathbb{E}_x[f(X_s)] - f(x)}{s}$$

for function f. In the discrete-time case, $\mathcal{L}$ is defined simply as

$$\mathcal{L} := P - I,$$

where P is the transition matrix and I is the identity matrix. Backward operator $\mathcal{L}^-$ is defined likewise but for the backward process $\{X_{-t}\}_{t \in T}$.

2.2 Diffusion process

In this subsection, the following stochastic differential equation is being considered

$$\dot{X}_t = b(X_t) + \sqrt{2D(X_t)} dW_t \tag{2.2}$$

where $X_t \in \mathbb{R}^d$, drift term $b : \mathbb{R}^d \to \mathbb{R}^d$, tensor $D : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$, W_t is the Weiner process (Brownian motion). Define $a := \sqrt{D}\sqrt{D}^T$, which is a symmetric and non-negative definite tensor. In the below, it will review transition path theory and

respond to questions 1 and 2. The main references in this subsection are [11, 4]. A and B are closed sets in this case.

By transition path theory, many statistical properties could be calculated, such as two committor functions, $q^+(x)$ and $q^-(x)$; the density function to describe the frequency of position x being visited in reactive trajectory, denoted by $m^R(x)$; probability density of position x in $S\backslash(A\cup B)$ conditional on it being in the reactive trajectory, denoted by $m^{AB}(x)$; probability current $J(x)$; probability flux $J^{AB}(x)$; transition rate k^{AB}; mean transition time $\bar{t}^{AB}$ and so on.

It should be reminded in advance that in diffusion process, infinitesimal generator for forward process is

$$\mathcal{L}f(x) = b(x)\cdot\nabla f(x) + a(x) : \nabla\nabla f(x), \tag{2.3}$$

while generator for time-reversed process is

$$\mathcal{L}^- f(x) = -b(x)\cdot\nabla f(x) + \frac{2}{m(x)}\mathrm{div}(a(x)m(x))\cdot\nabla f(x) + a(x) : \nabla\nabla f(x). \tag{2.4}$$

2.2.1 Collection of reactive trajectories

For a path ω in the probability space, we can find a sequence of time $(t_n^A, t_n^B)_{n\in\mathbb{Z}}$ such that $X_t(\omega)$ is a reactive trajectory iff $t \in [t_n^A, t_n^B]$. That is, we extract all the reactive trajectories from the sample path ω. Then define the collection of reactive trajectories as

$$R := \cup_{n\in\mathbb{Z}}(t_n^A, t_n^B). \tag{2.5}$$

Define

$$\mathbf{1}_R(t) := \mathbf{1}_A(X_{t^{AB,-}(t)})\mathbf{1}_B(X_{t^{AB,+}(t)}),$$

which is an indicator function to specify whether the system is in the progress of transition from set A to B at time t.

2.2.2 Probability density at equilibrium and probability current

We know that such system is ergodic, which means it has a unique invariant measure μ, whose probability density function is denoted by $m(x)$.

Let $p(x,t)$ be the probability density at position x and time t. From *Fokker-planck equation*, we have

$$\frac{\partial}{\partial t}p(x,t) = -\nabla \cdot (b(x)p(x,t) - \mathrm{div}(a(x)p(x,t))). \tag{2.6}$$

We might as well introduce

$$J(x,t) := b(x)p(x,t) - \mathrm{div}(a(x)p(x,t)), \tag{2.7}$$

which exactly depicts the probability flux at time t.

When the system is at equilibrium, $\frac{\partial}{\partial t}m(x) = 0$, which gives

$$0 = -\nabla \cdot (b(x)m(x) - \mathrm{div}(a(x)m(x))) \tag{2.8}$$

and

$$J(x) = b(x)m(x) - \mathrm{div}(a(x)m(x)). \tag{2.9}$$

When the system is time-reversible, the term $J(x)$ will vanish, that is, $J(x) = 0$.

2.2.3 Committor functions

Committor functions satisfy the *backward Kolmogorov equation*, which means,

$$\begin{cases} \mathcal{L}q^+(x) = 0 \\ q^+(x) \mid_{x\in A} = 0 \quad q^+(x) \mid_{x\in B} = 1 \end{cases} \tag{2.10}$$

and

$$\begin{cases} \mathcal{L}^- q^-(x) = 0 \\ q^-(x) \mid_{x\in A} = 1 \quad q^-(x) \mid_{x\in B} = 0 \end{cases} \tag{2.11}$$

Equation (2.1) is derived based on the understanding and intuition in physics, while equations (2.10) and (2.11) could offer more strict proof of equation (2.1): in time-reversible system, $\mathcal{L} = \mathcal{L}^-$, so q^+ and q^- satisfy the same equation except with different boundary conditions; therefore, $1 - q^+$ satisfies exactly the same equation as to q^-, which leads into the relation $q^+ = 1 - q^-$.

2.2.4 Probability density functions

The density function $m^R(x)$ is the density for measure μ^R which is defined on $S\backslash(A\cup B)$ and satisfies

$$\mu^R(C) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathbf{1}_C(X_t)\mathbf{1}_R(t)\ dt. \tag{2.12}$$

It should be remarked that $Z^{AB} := \mu^R(S \setminus (A \cup B)) < 1$, since it only includes the contribution from the reactive trajectories. By ergodicity and strong Markov property, we can deduce that

$$\mu^R(C) = \int_C q^+(x)q^-(x)m(x)\ dx,$$

for any measurable set $C \subset S \setminus (A \cup B)$, which means

$$m^R(x) = q^+(x)q^-(x)m(x). \tag{2.13}$$

Probability density of reactive trajectories $m^{AB}(x)$ is the normalized density of $m^R(x)$ on space $S \setminus (A \cup B)$,

$$m^{AB}(x) = m^R(x)/Z^{AB}, \tag{2.14}$$

where Z^{AB} can be rewritten as

$$\int_{S\setminus(A\cup B)} m^R(x)\ dx = \int_S m^R(x)\ dx,$$

since $m^R(x) = 0$ for $x \in A \cup B$.

2.2.5 Probability flux of reactive trajectory

We intend to find a vector filed $J^{AB}(x)$ which describes the probability flow of reactive trajectories from A to B, since $J(x)$ does not distinguish whether the probability flux is contributed from the reactive trajectories or not. Specifically, for any surface E (the space enclosed by E is denoted by E_a),

$$\int_E J^{AB}(x)\cdot\hat{n}(x)\ d\sigma(x) = \\ \lim_{s\to 0}\frac{1}{s}\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left[\mathbf{1}_{E_a}(X(t))\mathbf{1}_{S\setminus E_a}(X(t+s)) - \mathbf{1}_{S\setminus E_a}(X(t))\mathbf{1}_{E_a}(X(t+s))\right]\mathbf{1}_R(t)\ dt \tag{2.15}$$

It could be derived that (see [11] for proof)

$$J^{AB}(x) = q^-(x)q^+(x)J(x) + q^-(x)m(x)a(x)\nabla q^+(x) - q^+(x)m(x)a(x)\nabla q^-(x). \tag{2.16}$$

2.2.6 Transition rate and mean transition time

For any hyperplane $E \subset S \setminus (A \cup B)$,

$$k^{AB} = \int_E J^{AB}(x)\cdot\hat{n}(x)\ d\sigma(x). \tag{2.17}$$

Regardless of the surface chosen, this equation gives the same value for transition rate, since $\nabla \cdot J^{AB}(x) = 0$ for $x \notin A \cup B$. This is physically understandable, since the

transition rate should not depend on the diving surface chosen to observe the system.

The mean transition time

$$\begin{aligned}\bar{t}^{AB} &:= \lim_{T\to\infty} \frac{1}{N_T} \int_{-T}^{T} \mathbf{1}_{S\setminus(A\cup B)}(X(t))\mathbf{1}_R(t)\ dt \\ &= \lim_{T\to\infty} \frac{\frac{1}{2T}\int_{-T}^{T} \mathbf{1}_{S\setminus(A\cup B)}(X(t))\mathbf{1}_R(t)\ dt}{N_T/2T} \\ &= \frac{Z^{AB}}{k^{AB}}\end{aligned} \tag{2.18}$$

where N_T means the number of reactive trajectories within time period $[-T, T]$.

2.2.7 Most probable exit and entrance position

Vector field J^{AB} depicts the flux of reactive trajectories. In analogy, J^{AB} offers the velocity field of probability flows in space S, in the similar way as to water flow. Thus, it is reasonable to claim that $x_1^* := \arg\max_{x\in A} J^{AB}(x)\cdot\hat{n}(x)$ is the most probable exit position from set A and $x_2^* := \arg\max_{x\in B} J^{AB}(x)\cdot\hat{n}(x)$ is the most probable entering position into set B. Note that $\hat{n}(x)$ points outward of set A if $x\in\partial A$; it points inward to set B if $x\in\partial B$.

2.3 Markov chains

This subsection will review TPT in the context of discrete-time homogeneous Markov chain on countable state space, with reference from [7, 8]; TPT for continuous time Markov chain has similar forms. Without loss of generality, let $S = \mathbb{N}$.

As has been mentioned, the forward operator $\mathcal{L} = P - I$. Then the discrete probability distribution $\pi = (\pi_i)_{i\in S}$ satisfies

$$\pi L = 0. \tag{2.19}$$

The backward operator $\mathcal{L}^-$ has elements

$$\mathcal{L}^-_{ij} = \frac{\pi_j}{\pi_i}\mathcal{L}_{ji}. \tag{2.20}$$

2.3.1 Collection of reactive trajectories

Similarly, the collection of reactive trajectories

$$R := \cup_{n\in\mathbb{Z}}[t_n^A, t_n^B]$$

where $X_t(\omega)$ is a reactive trajectories iff $t_n^A \le t \le t_n^B$ for some integer n (of course, $t \in \mathbb{Z}$ herein). In the same way, define $\mathbf{1}_R(n) := \mathbf{1}_A(X_{t^{AB,-}(n)})\mathbf{1}_B(X_{t^{AB,+}(n+1)})$.

2.3.2 Committor functions

For forward committor function q^+, $q_i^+ := q^+(i) = 1$ if $i \in B$; $q_i^+ = 0$ if $i \in A$. For $i \notin A \cup B$,

$$\sum_{j\in S}\mathcal{L}_{ij}q_j^+ = 0. \tag{2.21}$$

As for backward committor function q^-, $q_i^- := q^-(i) = 1$ if $i \in A$; $q_i^- = 0$ if $i \in B$. For $i \notin (A \cup B)$,

$$\sum_{j\in S}\mathcal{L}^-_{ij}q_j^- = 0. \tag{2.22}$$

2.3.3 Probability distribution of reactive trajectory

The measure of reactive trajectories $m^R(\cdot) : S \to [0, 1]$ is a function which is defined as

$$m^R(i) := \lim_{T\to\infty}\frac{1}{2T+1}\sum_{n=-T}^{T}\mathbf{1}_{\{i\}}(X_n)\mathbf{1}_R(n), \qquad \forall i \in S \setminus (A \cup B). \tag{2.23}$$

By ergodicity,

$$m^R(i) = q_i^- \pi_i q_i^+. \tag{2.24}$$

$m^R(\cdot)$ is not a probability distribution, since usually

$$Z^{AB} := \sum_{i \in S} m^R(i) < 1. \tag{2.25}$$

The normalized probability distribution of reactive trajectories is denoted by $m^{AB}(\cdot)$,

$$m^{AB}(i) = m^R(i)/Z^{AB}, \qquad \forall i \in S \setminus (A \cup B). \tag{2.26}$$

2.3.4 Probability current

The probability current, which intends to sketch the contribution of transition between any two states to the ensemble behavior of global transition, is denoted by $J^{AB} : S \times S \to \mathbb{R}$, which is defined by

$$J_{ij}^{AB} = J^{AB}(i,j) := \lim_{T \to \infty} \frac{1}{2T+1} \sum_{n=-T}^{T} \mathbf{1}_{\{i\}}(X_n) \mathbf{1}_{\{j\}}(X_{n+1}) \mathbf{1}_R(n). \tag{2.27}$$

With the assumption that the DTMC is ergodic, we can deduce that, for $i \neq j$,

$$J_{ij}^{AB} = q_i^- \pi_i P_{ij} q_j^+ = q_i^- \pi_i L_{ij} q_j^+. \tag{2.28}$$

Since it is of little interest to include the contribution of jump from any state to itself, we can just set $J_{ii}^{AB} = 0$ by default.

The conservation law of probability current holds:

$$\sum_{j \in S} J_{ij}^{AB} = \text{ flow out of } i = \text{ flow into } i = \sum_{j \in S} J_{ji}^{AB} \qquad \forall i \in S \setminus (A \cup B). \tag{2.29}$$

2.3.5 Transition rate and mean transition time

Transition rate is the frequency of transition from A to B, defined by

$$\begin{aligned} k^{AB} &:= \lim_{T\to\infty} \frac{1}{2T+1} \sum_{n=-T}^{T} \mathbf{1}_A(X_n)\mathbf{1}_{S\setminus A}(X_{n+1})\mathbf{1}_R(n) \\ &= \sum_{i\in A, j\notin A} J_{ij}^{AB} \\ &= \sum_{i\in A, j\in S} J_{ij}^{AB}, \end{aligned} \tag{2.30}$$

since $J_{ij} = 0$ if $j \in A$.

The mean transition time

$$\begin{aligned} \bar{t}^{AB} &:= \lim_{T\to\infty} \frac{\sum_{n=-T}^{T} \mathbf{1}_R(n)}{N_T} \\ &= \lim_{T\to\infty} \frac{\frac{1}{2T+1}\sum_{n=-T}^{T} \mathbf{1}_R(n)}{\frac{1}{2T+1} N_T} \\ &= \frac{Z^{AB} + k^{AB}}{k^{AB}} \\ &= 1 + \frac{Z^{AB}}{k^{AB}}, \end{aligned} \tag{2.31}$$

where N_T means the number of transition within time period $[-T, T]$.

In lines 2-3 of equation (2.31),

$$\frac{1}{2T+1} \sum_{n=-T}^{T} \mathbf{1}_R(n) = \frac{1}{2T+1} \sum_{n=-T}^{T} (\mathbf{1}_R(n)\mathbf{1}_A(X_n) + \mathbf{1}_R(n)\mathbf{1}_{S\setminus(A\cup B)}(X_n)).$$

The second part gives Z^{AB} by letting $T \to \infty$. We know that

$$\mathbf{1}_R(n)\mathbf{1}_A(X_n) = \sum_{i\in A, j\in S} \mathbf{1}_{\{i\}}(X_n)\mathbf{1}_{\{j\}}(X_{n+1})\mathbf{1}_R(n),$$

so

$$\lim_{T\to\infty} \frac{1}{2T+1} \sum_{n=-T}^{T} \mathbf{1}_R(n)\mathbf{1}_A(X_n) = \sum_{i\in A, j\in S} J_{ij}^{AB} = k^{AB}.$$

Thus, we obtain

$$\bar{t}^{AB} = 1 + \frac{Z^{AB}}{k^{AB}}.$$

Note that in discrete-time case, some concepts may have different forms, due to the jumping process.

2.3.6 Most probable exit and entrance point

Equation (1.4) could be rewritten as

$$\begin{aligned} i^* &= \arg\max_{i\in A} \lim_{T\to\infty} \frac{1}{2T+1} \sum_{n=-T}^{T} \mathbf{1}_{\{i\}}(X_n)\mathbf{1}_R(n) \\ &= \arg\max_{i\in A} \sum_{j\in S} J_{ij}^{AB}. \end{aligned}$$

The term $\sum_{j\in S} J_{ij}^{AB}$ basically measures how much probability leaks from point $i \in A$ to set B. i^* is the most probable exit point in set A. Similarly, the most probable enter point into set B is

$$j^* = \arg\max_{j\in B} \sum_{i\in S} J_{ij}^{AB}.$$

2.4 Representative dominant transition path

In equation (2.29), $\{J_{ij}^{AB}\}_{(i,j)\in S\times S}$ generates a network flow, in which we might be interested in studying the structure of flows, that is, finding the representative path that describes where the flow is denser. In paper [8], representative dominant transition path has been proposed and studied for Markov jump process with countable state space; in this subsection, the main idea of representative dominant transition path will be reviewed.

The effective probability current

$$J_{ij}^{+} := \max(J_{ij}^{AB} - J_{ji}^{AB}, 0), \tag{2.32}$$

which depicts the net probability current from state i to state j. With J^{+}, a directed graph $G(V, E) = G\{J^{+}\}$ can be generated: the edges are the pairs which have non-negative effective probability current, namely, the set of edges $E = \{(e_1, e_2) : J_{e_1,e_2}^{+} > 0\}$; the set of nodes V is a set whose elements appear at least once in any pair (e_1, e_2) in E.

Transition path means simple transition path herein without distinguish. The reason of limiting the objects being studied to the simple path is that path with loops is more complicated; more importantly it does not represent the real probability flow. For a transition path, its capacity of carrying flows is confined by the minimal effective probability current along this path. Based on this observation, for a path $\varphi = (\varphi_i)_{0 \le i \le n}$, the *min-current* $c(\varphi)$ is defined as

$$c(\varphi) = \min_{e=(i,j)\in\varphi} J_{ij}^{+}, \tag{2.33}$$

and the *dynamical bottleneck* is the argument which achieves the min-current,

$$(b_1, b_2) = \arg \min_{e=(i,j)\in\varphi} J_{ij}^{+}. \tag{2.34}$$

Based on our argument before, it is apparent that the most desired path should have maximal min-current. Let W be the space of all transition paths and define

$$c_{\max} := \max_{\varphi \in W} c(\varphi), \tag{2.35}$$

then the set of paths which have $c_{\max}$ as min-current is

$$W_D := \{\varphi \in W : c(\varphi) = c_{\max}\}. \tag{2.36}$$

The element in W_D is referred as *dominant transition path*.

If we assume that we only have one edge which achieves c_{max}, namely, the global bottleneck is unique, denoted by (b_1, b_2), any path $\varphi = (\varphi_i)_{0\leq i\leq n}$ in W_D can be partitioned into two parts $\varphi_{\mathcal{L}} := (\varphi_j)_{0\leq j\leq n^*}$ and $\varphi_{\mathcal{R}} := (\varphi_j)_{n^*+1\leq j\leq n}$ if the position of node b_1 is n^* in this path. Assume we already know the most desired path $\varphi^*_{\mathcal{L}}$ in the set $G_{\mathcal{L}} := \{\varphi_{\mathcal{L}} : \varphi \in W_D\}$ and $\varphi^*_{\mathcal{R}}$ in the set $G_{\mathcal{R}} := \{\varphi_{\mathcal{R}} : \varphi \in W_D\}$, then easily we might conclude that the representative dominant transition path

$$\varphi^* = (\varphi^*_{\mathcal{L}}, \varphi^*_{\mathcal{R}}). \tag{2.37}$$

Then the question becomes how to obtain the φ^*_L and φ^*_R. Similarly, the new sets of dominant transition paths for sub-graphs $G_{\mathcal{L}}$ and $G_{\mathcal{R}}$ can be defined and then φ^*_L and φ^*_R can be defined and obtained in the same way as φ^*. The definition of representative dominant transition path is based on the algorithm paradigm called divide-and-conquer (D&C). The detailed algorithms and computational cost analysis could be found in [8]. Representative dominant transition path provides another perspective to understand the mechanics of transition in stochastic dynamical systems with countable state space. It serves as one possible response to question 3 about transition mechanics. The generalization of representative dominant transition path is possible but it will not be discussed here.

3 Applications

Transition path theory has wide applications as mentioned in the introduction section and this section serves as a short introduction and review of some classical models where TPT could be applied. We will focus on (1) time-reversible diffusion process with gradient drift term and isotropic diffusion tensor and (2) randomly generated Markov chain. Numerical results, some additional theoretical results and remarks will been provided in the following.

3.1 Diffusion process

In equation (2.2), let the diffusion tensor $\sqrt{D} = \sigma I$, then tensor $a = \sigma^2 I$; let drift term $b(x) = -\nabla V(x)$, where scalar function V is usually known as potential function. Hence, the general diffusion equation (2.2) is simplified into

$$\dot{X}_t = -\nabla V(X_t) + \sqrt{2}\sigma dW_t. \tag{3.1}$$

σ is related to the temperature in physical models: for instance, $\sigma^2 = k_B T r = r/\beta$ in Langevin equation where r is the friction coefficient.[1]

The invariant measure density $m(x)$ could be solved explicitly and it has the form of

$$m(x) = Z^{-1} \exp(-\frac{1}{\sigma^2} V(x)) \tag{3.2}$$

where $Z = \int_{\mathbb{R}^d} \exp(-\frac{1}{\sigma^2} V(x))$, the factor for normalization. Assume that $V(x)$ under consideration increases sharply when $|x| \to \infty$ so that $Z < \infty$. Such probability density function is sometimes known as *Boltzmann-Gibbs probability density*. It could be easily verified that $-bm + \mathrm{div}(am) = 0$ for this case, which means such diffusion process is

[1] In Langevin equation, the diffusion tensor, however, has slight different form, compared with the assumption here.

time-reversible. Therefore, $q^-(x) = 1 - q^+(x)$ for $x \in S$.

With the form of invariant measure density $m(x)$, the conclusion could be drawn that we could limit the numerical methods into certain space $\mathbf{S}$, and the space outside is negligible in its contribution to the whole system. The diffusion process is well-defined on the space $\mathbb{R}^d$, but boundary conditions need to be imposed in order to solve it numerically. Based on the observation on $m(x)$, Dirichlet boundary condition could be applied without problem if $\mathbf{S}$ is reasonably large enough. As for the committor functions, Neumann (reflective) boundary condition appears to be more natural, that is, $\partial_{\hat{n}} q|_{\partial S} = 0$ or $\nabla q \cdot \hat{n} = 0$ where q stands for both forward committor function and backward committor function, and $\hat{n}$ is the normal vector at the boundary $\partial\mathbf{S}$.

Focus is put on two classical models for potential V: one is double-well potential

$$V(x_1, x_2) = \frac{5}{2}(x_1^2 - 1)^2 + 5x_2^2$$

and another is triple-well potential

$$\begin{aligned} V(x_1, x_2) = {} & 3\exp\left(-x_1^2 - (x_2 - 1/3)^2\right) - 3\exp\left(-x_1^2 - (x_2 - 5/3)^2\right) \\ & - 5\exp\left(-(x_1 - 1)^2 - x_2^2\right) - 5\exp\left(-(x_1 + 1)^2 - x_2^2\right) \\ & + 0.2x_1^4 + 0.2(x_2 - 1/3)^4. \end{aligned}$$

These two examples have been used and illustrated in [9]. I will use them as examples here for easy illustration and reproduce the numerical results in that paper.

3.1.1 Double-well potential

In double-well potential system, there exists two local minima $(\pm 1, 0)$ separated by a saddle point $(0, 0)$. The space chosen for numerical approximation $\mathbf{S} = [-2, 2] \times [-1.5, 1.5]$ with grids 501×501. Set A and B are chosen as the neighborhood around

$(-1, 0)$ and $(1, 0)$ respectively with potential $V(x_1, x_2) < 0.4$. σ is chosen to be 1. The contour plot of potential $V(x_1, x_2)$ could be seen in figure 1.

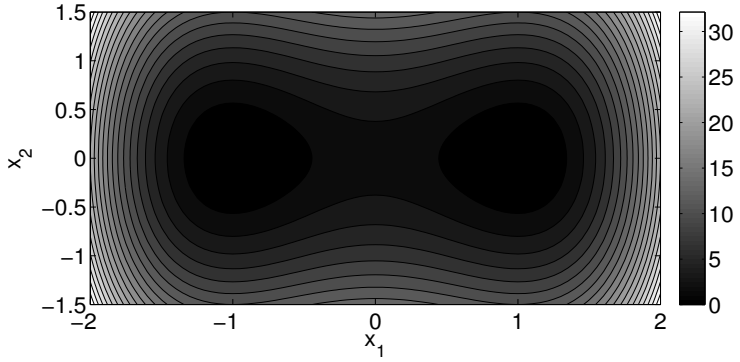
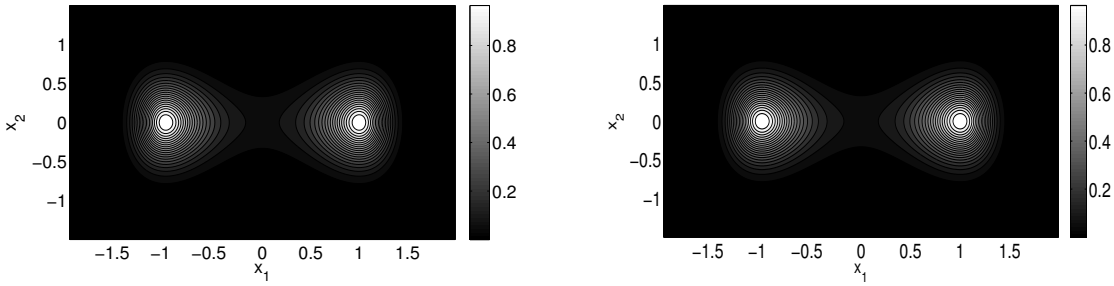


Figure 1: Potential $V(x_1, x_2)$.

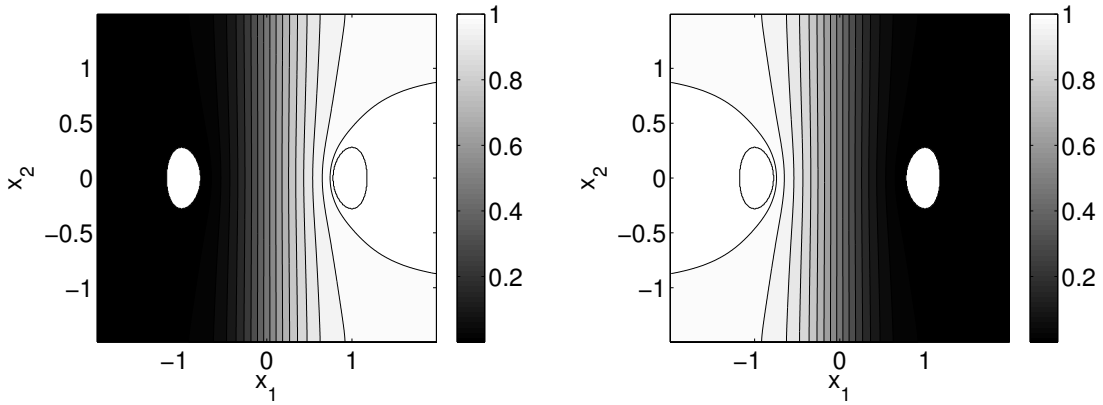
The probability density at equilibrium $m(x)$ is presented in figure 2 as well as the Boltzmann-Gibbs probability density. Figure 2 demonstrates in the numerical sense that equation (3.2) holds. The probability density is higher at around two local minima, which confirms the metastability of these two local minima in diffusion process. The forward committor function q^+ and backward committor function q^- have been visualized in figure 3; to facilitate the visualization of set boundaries, the value of committor functions at both set A and B is set to be 1. In figure 3a, the value of q^+ at the right half plane is higher than that in the left half plane, which means the particle at right half plane has higher probability to reach set B than that at the left half plane, and which is intuitively understandable. As for backward committor function q^-, the situation is totally reversed. By checking the numerical values of q^+ and q^-, the equation (2.1) holds with high accuracy.

Probability density of reactive trajectories $m^{AB}(x)$ has been computed and shown in figure 4. The region around the saddle point $(0, 0)$ has higher value in $m^{AB}(x)$, which means in the transition, the particle stays at this region more frequently than other areas; such region is sometimes referred as *dynamical bottleneck* or *transition state region*. In this case, $J(x_1, x_2)$ should approximately be zero; the result does not change too much either by using J calculated from equation (2.9) or by simply replacing J



(a) Invariant measure density $m(x)$ (b) Boltzmann-Gibbs probability density

Figure 2: Invariant measure density $m(x)$ and Boltzmann-Gibbs probability density.



(a) Forward committor function q^+. (b) Backward committor function q^-.

Figure 3: Committor functions

with zero. The vector field $J^{AB}(x_1, x_2)$ is shown in figure 5. As we might notice, $J^{AB}(x)$ points outward at the boundary of set A and points inward at the boundary of set B, since the effective reactive trajectories must start from set A and end at set B. The transition rate k^{AB} calculated from equation (2.17) with y-axis as diving surface is approximately 0.07764 and after enlarging space $\mathbf{S}$, transition rate k^{AB} does not change too much.

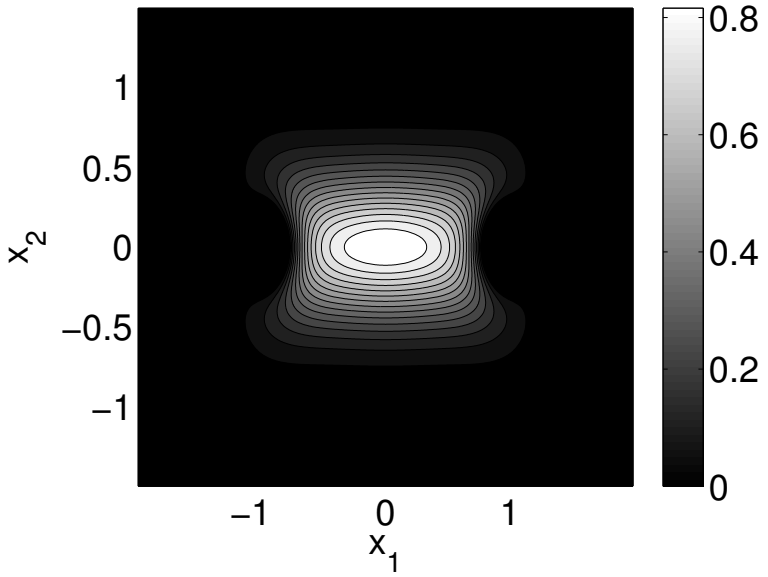


Figure 4: Probability density of reactive trajectories $m^{AB}(x)$.

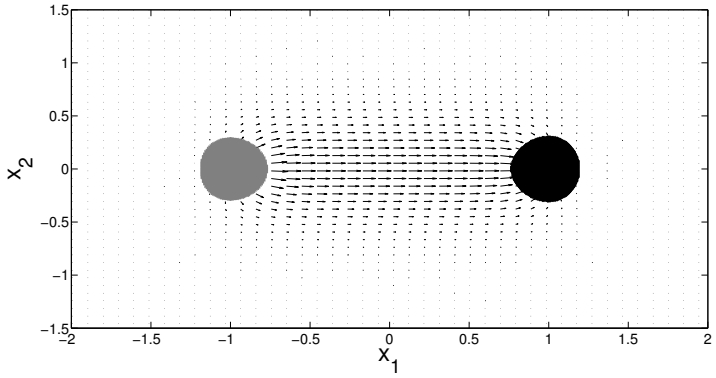


Figure 5: The vector field J^{AB}. Set A is colored in gray; set B is colored in black.

3.1.2 Triple-well potential

In double-well potential system, the dynamical bottleneck is trivial since for a particle from set A to set B, the cheapest way is to cross the saddle point, that is, (0, 0), and thus the region around that saddle point must be visited more frequently than other areas. In triple-well potential system, there are three local minima at around $(\pm 1.05, -0.04)$, $(0, 1.54)$, three saddle points at around $(\pm 0.6, 1.1)$, $(0, -0.3)$ and one local maximum at round $(0, 0.52)$. Let set A be the circle around the local minimum

$(-1.05, -0.04)$ with radius 0.2 and set B be the circle around the local minimum $(1.05, -0.04)$ with radius 0.2.

As we could notice in figure 6 for potential $V(x_1, x_2)$, there are two possible paths that particle in set A could choose to reach set B: one is to cross the saddle point $(0, -0.3)$ to reach set B directly and another is to cross the saddle points $(\pm 0.6, 1.1)$, between which it has to reach the neighborhood of local minimum $(0, 1.54)$. These two main pathways are prominent under different conditions. The potential at saddle point $(\pm 0.6, 1.1)$ is -1.65 and the potential at saddle point $(0, -0.3)$ is -1.38. From the large deviation principle (LDP), when the noise is infinitesimally small, the optimal path will be that that minimizes the potential barrier, which means, the most probable exit path crosses the saddle points $(\pm 0.6, 1.1)$ before it reaches set B.

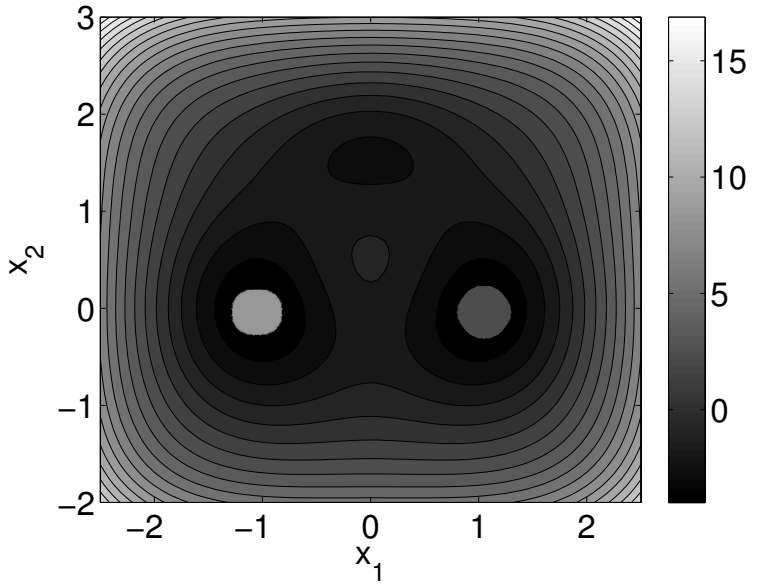


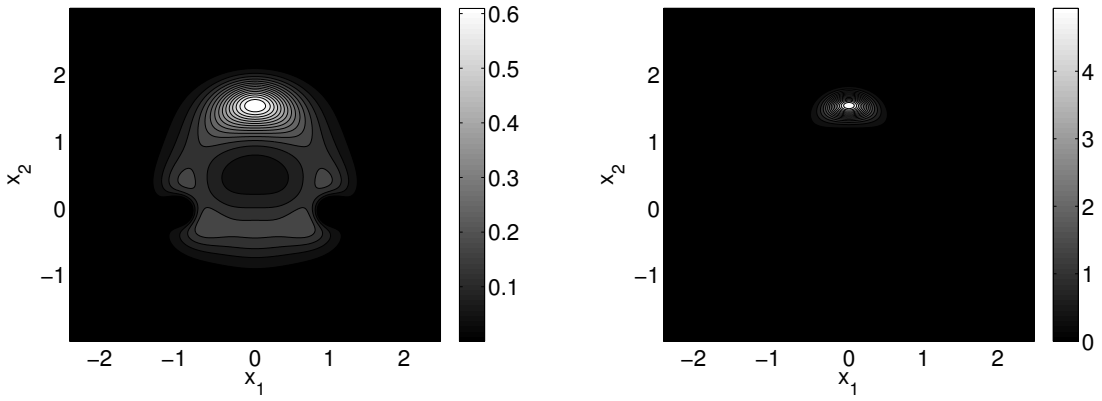
Figure 6: Potential $V(x_1, x_2)$. Set A is colored in light gray, set B in dark gray.

Here we choose $\sigma = 1/\sqrt{1.67}$ and $1/\sqrt{6.67}$ following paper [9] to verify this idea. The probability density of reactive trajectories $m^{AB}(x)$ is given in figure 7: at low noise, most probably, it will stays near the local minimum $(0, 1.54)$, which demonstrates that it must cross saddle points $(\pm 0.6, 1.1)$; when it is under higher noise, the probability density $m^{AB}(x)$ at the region near the saddle point $(0, -0.3)$ increases, compared with

the case with low noise, and there are two dynamical bottlenecks, one the region near saddle point $(0, -0.3)$ and another the region near local minimum $(0, 1.54)$.

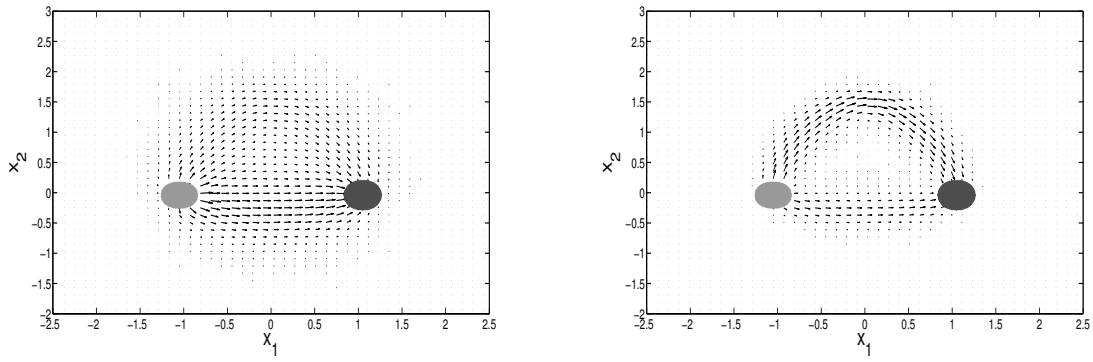
The probability flux $J^{AB}(x)$ for reactive trajectories in figure 8 further confirms this finding. When noise is relatively lower, the streamline that passes through saddle points $(\pm 0.6, 1.1)$ is dominating, while such dominance becomes weaker when noise increases. This could be explained by using physics: when noise is larger, that is, the temperature T for such system is larger, the particle has more kinetic energy to cross the energy barrier at saddle point $(0, -0.3)$ so that the chance for its hopping into set B through this saddle point contributes more to the whole system.

Transition rates k^{AB} have been computed for the sake of completeness: when $\sigma = 1/\sqrt{1.67}$, the transition rate is 1.88×10^{-2}. When $\sigma = 1/\sqrt{6.67}$, $k^{AB} = 8.81 \times 10^{-8}$. The decrement of transition rate when noise decreases is natural and trivial. These two values are close to the results presented in paper [9].



(a) $\sigma = 1/\sqrt{1.67}$. (b) $\sigma = 1/\sqrt{6.67}$.

Figure 7: Probability density of reactive trajectories $m^{AB}(x)$ under different noise σ.



(a) $\sigma = 1/\sqrt{1.67}$. (b) $\sigma = 1/\sqrt{6.67}$.

Figure 8: Probability flux $J^{AB}(x)$ under different noise σ.

3.2 Markov chain

A random generated Markov chain $\{X_n\}_{n\in\mathbb{Z}}$ with four-state space $S = [1, 2, 3, 4]$, is used as the simplest example to illustrate how transition path theory works. The one-step transition matrix for this Markov chain is

$$P = \begin{bmatrix} 0.4539 & 0.2906 & 0.1907 & 0.0648 \\ 0.3537 & 0.4600 & 0.0362 & 0.1501 \\ 0.5584 & 0.0211 & 0.3236 & 0.0968 \\ 0.1737 & 0.5098 & 0.1597 & 0.1568 \end{bmatrix}$$

Let set $A = \{1\}$ and set $B = \{4\}$.

3.2.1 Basic quantities

Many properties have been listed in table 1. Additionally, $Z^{AB} = 0.0839$ computed from equation (2.25) and $k^{AB} = 0.0723$ obtained from equation (2.30). The probability

current

$$J^{AB} = \begin{bmatrix} 0 & 0.0341 & 0.0118 & 0.0264 \\ 0 & 0 & 0.0012 & 0.0337 \\ 0 & 0.0008 & 0 & 0.0123 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

property	value
probability distribution at equilibrium $\pi(x)$	$\begin{bmatrix} 0.4075 & 0.3271 & 0.1578 & 0.1076 \end{bmatrix}$
forward committor function q^+	$\begin{bmatrix} 0 & 0.2881 & 0.1522 & 1 \end{bmatrix}$
backward committor function q^-	$\begin{bmatrix} 1 & 0.6856 & 0.8041 & 0 \end{bmatrix}$
measure of reactive trajectories $m^R(\cdot)$	$\begin{bmatrix} 0 & 0.0646 & 0.0193 & 0 \end{bmatrix}$
probability distribution of reactive trajectories $m^{AB}(\cdot)$	$\begin{bmatrix} 0 & 0.7700 & 0.2300 & 0 \end{bmatrix}$

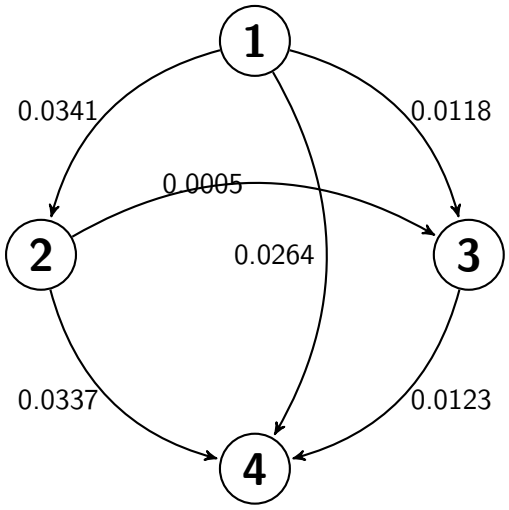
Table 1: Properties in Markov chain.

3.2.2 Representative dominant transition path

According to equation (2.32), the effective probability current J^+ is

$$J^+ = \begin{bmatrix} 0 & 0.0341 & 0.0118 & 0.0264 \\ 0 & 0 & 0.0005 & 0.0337 \\ 0 & 0 & 0 & 0.0123 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The diagram for effective probability current J^+ is



All the transition paths from 1 to 4 are

- $1 \to 2 \to 4$ with min-current 0.0337;
- $1 \to 2 \to 3 \to 4$ with min-current 0.0005;
- $1 \to 3 \to 4$ with min-current 0.0118;
- $1 \to 4$ with min-current 0.0264.

According to equation (2.35), $c_{\max} = 0.0337$ and therefore the first dynamical bottleneck is the edge $(b_1, b_2) = (2, 4)$ with weight $c_{\max} = 0.0337$. Then, let $\{2\}$ be the new set B and we need to find the representative dominant transition path in the graph with nodes $\{1, 2\}$, which leads to trivial solution $\varphi^*_{\mathcal{L}} = (1, 2)$. Finally, the overall representative dominant transition path for this Markov chain is $\varphi^* = (1, 2, 4)$, which means in the sense of statistics, most probably, the transition path will start from state 1, pass through state 2 and finally reach state 4.

4 Discussion on TPT and MPEP

The main idea of most probable exit path (MPEP) is to maximize the probability of path, under fixed starting point, ending point and the transition time T. This section will give a short review of MPEP to response to Question 4. The main reference for this section is [1].

For the stochastic mapping, $X_{n+1} = f(X_n) + \sigma\eta_n$, where η_n are iid standard normal random variables. We consider $X_n \in \mathbb{R}$ for convenience, that is, dim = 1. With fixed initial position x_0 at time 0 and ending position x_n at time n, for a particular path φ linking them, its probability

$$\mathcal{P}[\varphi] \propto \exp\left(-\frac{1}{\sigma^2} S[x_0, x_1, \cdots, x_n]\right), \tag{4.1}$$

where cost function

$$S[x_0, x_1, \cdots, x_n] = \frac{1}{2}\sum_{i=0}^{n-1}(x_{i+1} - f(x_i))^2.$$

The total probability of having a path φ connecting these two points is

$$P(x_n \mid x_0) = \int_{\varphi} \mathcal{P}[\varphi]\ d\{\varphi\} \tag{4.2}$$

When $\sigma \to 0$, the asymptotic approximation for $P(x_n \mid x_0)$ is

$$P(x_n \mid x_0) \to \gamma \exp\left(-\frac{1}{\sigma^2} S_{\min}\right), \tag{4.3}$$

where γ is the prefactor for correction. Equation (4.3) is guaranteed by the fact that when σ approaches 0, the term with $S > S_{\min}$ will vanish somehow.

Let $x_{i+1} - f(x_i) = \xi_i$ and if we introduce the Lagrange multiplier

$$\bar{S} = \frac{1}{2}\sum_{i=0}^{n-1} \xi_i^2 + \lambda_i[x_{i+1} - f(x_i) - \xi_i]. \tag{4.4}$$

By taking derivative with respect to ξ_i and x_{i+1}, we get $\xi_i - \lambda_i = 0$ and $\lambda_i - \lambda_{i+1} f'(x_{i+1}) = 0$. Then,

$$\begin{cases} x_{i+1} = f(x_i) + \lambda_i \\ \lambda_{i+1} = [f'(x_{i+1})]^{-1}\lambda_i \end{cases} \tag{4.5}$$

For a diffusion process $\dot{X}_t = f(X_t) + dW_t$, it could be approximated by

$$x_{n+1} = x_n + f(x_n)\Delta t + \sqrt{\Delta t} B(0,1), \tag{4.6}$$

where $B(0,1)$ is the standard Brownian motion. This approximation comes from the definition of Ito integral. Let $\lambda_i/\Delta t \to \lambda$, we could deduce that

$$\begin{cases} \dot{x} = f(x) + \lambda \\ \dot{\lambda} = -f'(x)\lambda \end{cases} \tag{4.7}$$

Then the analysis of most probable exit path turns into manifold analysis on a higher dimensional space (x, λ). In addition, if we treat λ as momentum, equation (4.7) forms the *Wentzell-Freidlin Hamiltonian* system with $H = \frac{1}{2}\lambda^2 + \lambda f$.

In summary, TPT and MPEP both intend to shed light on transition mechanics of some stochastic dynamical systems while their approaches are somewhat different. TPT works with finite noise cases and it takes the time period of observation to be infinity at first to make use of the ergodicity; MPEP works with zero-temperature cases, that is, infinitesimally small noise cases, and it makes the noise intensity approach zero before letting observation period T be infinity.

5 Conclusion

In summary, this report intends to answer four questions proposed at the beginning about transition problem under small noise in stochastic dynamical systems. In section 2, we have discussed the transition path theory in detail in both diffusion process and Markov chains. Many statistical properties have been derived, such as probability density at equilibrium, probability measure of reactive trajectories, probability flux of reactive trajectories and transition rate. The related numerical results have been presented in section 3 by using some classical models such as triple-well potential and finite state DTMC.

The main idea of TPT is to consider the ensemble behavior of reactive trajectories rather than focusing on a single reactive trajectory, because in real physical system, it seems that it is not realistic nor important to study single reactive trajectory. In the context of Markov chains, the representative dominant transition path is defined to depict the region where the reactive trajectories are denser. Representative dominant transition path and most probable exit path are both intending to sketch the transition mechanics, while they are applicable in different cases, one in the finite noise case, and another in the infinitesimally small noise case. When noise in TPT approaches zero in some stochastic dynamical systems, these two concepts are supposed to be consistent, which is, however, not guaranteed in general, as far as I know.

References

[1] S. Beri, R. Mannella, D. G. Luchinsky, A. N. Silchenko, and P. V. E. McClintock. Solution of the boundary value problem for optimal escape in continuous stochastic systems and maps. *Phys. Rev. E*, 72:036131, Sep 2005.

[2] Maria Cameron and Eric Vanden-Eijnden. Flows in complex networks: Theory, algorithms, and application to lennard–jones cluster rearrangement. *Journal of Statistical Physics*, 156(3):427–454, 2014.

[3] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *The Journal of Chemical Physics*, 126, 2007.

[4] Weinan E and Eric vanden Eijnden. Towards a theory of transition paths. *Journal of Statistical Physics*, 123(3):503–523, 2006.

[5] Mark I. Freidlin and Alexander D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag, 2012.

[6] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62:251–341, Apr 1990.

[7] Ling Lin. *Mathematical Study of Milestoning*. PhD thesis, New York University, September 2013.

[8] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Transition path theory for markov jump processes. *Multiscale Model. Simul.*, 7(3):1192–1219, January 2009.

[9] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *The Journal of Chemical Physics*, 125(8), 2006.

[10] Frank Noé, Christof Schütte, Eric Vanden-Eijnden, Lothar Reich, and Thomas R. Weikl. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 106(45):19011–19016, 2009.

[11] E. Vanden-Eijnden. Transition path theory. In Mauro Ferrario, Giovanni Ciccotti, and Kurt Binder, editors, *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology*, volume 1, pages 453–493. Springer, 2006.

[12] Xiang Zhou. *Study of Noise-induced Transition Pathways in Non-gradient systems Using Adaptive Minimum Action Method*. PhD thesis, Princeton University, September 2009.