Transient Behavior of Stochastic Dynamics

by

Shuo Liu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Applied Mathematics

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

This thesis concerns dynamical systems subject to small noise perturbations. Our purpose is to obtain a deep understanding of how small noise perturbations influence the original unperturbed dynamical system, especially over long but finite time intervals. We consider two special systems, the random linear recurrence equation which is a discrete system, and the underdamped Langevin equation which is a continuous system.

According to the classical perturbation theory, small perturbations do not impact much of the dynamics over a short time. When the time tends to infinity, in many situations, the flow can eventually merge with one of the steady states of the perturbed system. As far as we know, small perturbations gradually influence the dynamics in an accumulative way over time, resulting in the so-called transient behavior which describes the process of such a gradual change of dynamics from the unperturbed system to the steady state of the corresponding perturbed system.

We demonstrate this change by utilizing the cut-off phenomenon in the random linear recurrence system. We find the time window when the solution evolves from a deterministic value to the limiting distribution. The underdamped Langevin equation is essentially a slow-fast system admitting three time scales. The unperturbed Hamiltonian system plays a dominant role in the short time scale. In the intermediate time scale, the fast variables have already tended to their stationary distributions and the slow variables transition from their initial deterministic values to the marginal distribution of the slow variables with respect to the system's stationary distribution. Finally, in the long time scale, all variables are slaved by the stationary distribution.

Keywords: Perturbed Dynamical Systems, Cut-Off Phenomenon, Random Linear Recurrence Equation, Slow-Fast Dynamical Systems, Underdamped Langevin Equation, Multiscale Analysis, Stationary Distribution

Preface

I confirm that this is my original work, and the use of all material from other sources has been properly and fully acknowledged. Chapter 2 of this thesis has been published as Barrera, Gerardo, and Shuo Liu, "A Switch Convergence for a Small Perturbation of a Linear Recurrence Equation." *Brazilian Journal of Probability and Statistics*, 35.2 (2021), pp. 224-241.

Acknowledgements

First and foremost, I would like to thank my advisor and chair of my committee, Professor Yingfei Yi, for providing the opportunity to finish my research at the University of Alberta. His support and guidance helped me understand the importance of transient behavior in applied mathematics. Furthermore, his suggestions helped me overcome the research bottlenecks in time. I also could not have completed the Ph.D. program without my supervisory committee, who provided not only academic knowledge but also research experience. I sincerely appreciate every member of my supervisory committee, Professor Michael Yi Li, Professor Hao Wang, and Professor Zhongwei Shen. Meanwhile, I am grateful to Professor Yaozhong Hu and Professor Thomas Hillen who were arm's length examiners of my candidacy exam. Their farseeing suggestions broadened my understanding of transient behaviors.

I gratefully acknowledge the financial support provided by Professor Yingfei Yi, Professor Zhongwei Shen, and our department. Their support relieved my economic pressure and helped me devote myself entirely to my research. In addition, I am indebted to our Graduate Chair, Professor Feng Dai, and Graduate Program Administrator, Mr. Peng Wang. Their wise suggestions and efficient support have made the project process go smoothly.

I appreciate Dr. Barrera Gerardo, who is not only my coauthor of "A Switch Convergence for a Small Perturbation of a Linear Recurrence Equation" but also my torchbearer of stochastic analysis. I am grateful to Professor Yao Li, who shared his hybrid method of the simulation of stationary distribution. Having borrowed his core idea, I could generate the data-driven method to simulate the solution in different

time scales efficiently. I want to thank Professor Zhongwei Shen, who helped me to understand the big picture of the connection between the perturbation theory and the time limiting distribution. Meanwhile, he guided me to utilize the averaging principle to demonstrate the transient behavior in the underdamped Langevin equation.

Lastly, I would like to thank my parents who supported me in studying abroad. They encouraged me to hold on to my dream, enabling me to rediscover my direction and advance towards my goal in times of despair. I am not adept at expressing my feelings. I just want to say I LOVE you. I want to spend more time with you before it is too late. It is time for your son to take care of you.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation and objectives

There are many uncertainties which are introduced by the environment and objects themselves. To deal with those uncertainties, modellers add small noise perturbations to the deterministic systems, such as the predator-prey model [1], the population model [2] and the general circulation model [3]. There is much excellent research analyzing the existence and uniqueness of equilibria (stationary distributions) of corresponding stochastic systems [4]. In this work, we study when and how solutions transition from their initial values to equilibria.

In general, the deterministic system is called the unperturbed system. Its solution is denoted as X_t^0. Because of the lack of knowledge of the environment, we often use white noise to represent it without loss of generality. Then the perturbed system becomes a stochastic system. We use X_t^ϵ to denote its solution, where ϵ is the intensity of the perturbation.

There are two well-known results of the dynamical system with a small perturbation. One is that the perturbed system is mainly slaved by the unperturbed system when ϵ tends to zero in any finite time interval by the classical perturbation theory [5]. The expression is as follows:

$$\lim_{\epsilon \to 0} \mathbb{P}\left\{ \max_{t \in [0,T]} \left| X_t^\epsilon - X_t^0 \right| > \delta \right\} = 0,$$

for every $T > 0$ and $\delta > 0$. It implies that a small enough perturbation does not much affect the unperturbed system within the initial finite time. We name this convergence as the *perturbation limiting behavior*. The other significant result is that,

with any fixed ϵ, X_t^ϵ tends to one of the equilibria of the perturbed system when time tends to infinity. Let X_∞^ϵ be a random variable defined on a common probability space, whose distribution is the same as the equilibrium. With a suitable distance, we have

$$\lim_{t\to\infty} \mathbf{d}\left(X_t^\epsilon, X_\infty^\epsilon\right) = 0,$$

which is the classical limiting behavior.

The first result shows that there is not much difference between the unperturbed and perturbed systems at the beginning. The small perturbation has barely begun to affect the unperturbed system. We are interested in when the small perturbation starts to powerfully impact the system. The second result presents that the small noise perturbations ultimately affect the unperturbed systems when time tends to infinity. We want to know, rather than the infinite time, when X_t^ϵ and X_∞^ϵ are close enough.

After knowing the start time and end time of the influence introduced by the small perturbation, we plan to track where the solution is during the intermediate time, i.e., how the flow transforms from the almost deterministic solution X_t^0 to an equilibrium of the perturbed system X_∞^ϵ. The primary purpose of this work is to understand this type of transition named the *transient behavior*. In the past, the following three types of phenomena have been found in the first order systems.

i) Metastability

   The classic example to present the metastability behavior is one-dimensional diffusion in a double-well landscape at low temperature. The position of a particle is governed by the following stochastic differential equation:

$$\mathrm{d}x_t = -V'\left(x_t\right)\mathrm{d}t + \sqrt{2\epsilon}\mathrm{d}W_t,$$

   where W is a Brownian motion, ϵ is the intensity of the temperature, and V is a double-well potential as follows:

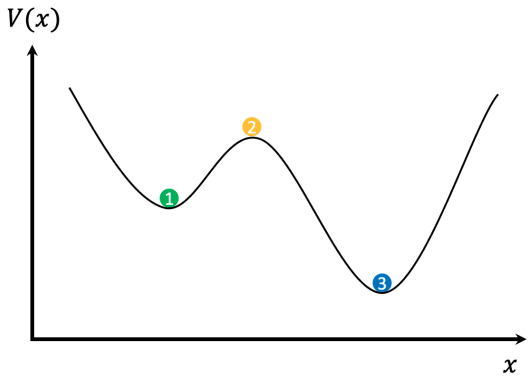


Figure 1.1: Double-well potential

In the unperturbed system, positions 1 and 3, wave troughs in Figure 1.1, are asymptotical stable states. All particles starting from the left basin converge to position 1, while those starting from the right converge to position 3. Because of the small noise perturbation, some particles inside a basin can be transitioned to the critical point, position 2, after an exponentially long time. Then, those particles enter one of the basins randomly and converge to the corresponding asymptotical stable states exponentially fast. We refer to [6] for details on the mathematical description of the transition from the metastable to the ultimate stationary distribution.

ii) Cut-off Phenomenon

The cut-off phenomenon was firstly introduced by Aldous and Diaconis to measure how many times the cards can be shuffled evenly [7]. The shuffle is modeled by a discrete Markov chain X_t^n, where X is the order of n cards after t random shuffles. Its stationary state is denoted as π^n. The total variation distance is utilized to measure the distance between X_t^n and π^n. There is an abrupt convergence of this distance in a specific cut-off window. The cut-off time is $\frac{3}{2}\log_2 n$ which has entered popular science as "7 shuffles are enough". This phenomenon can be drawn with a fixed parameter n as follows:

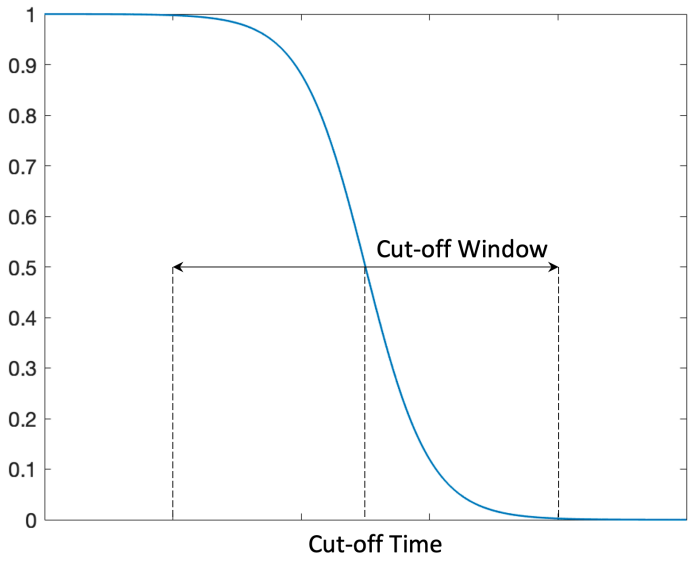


Figure 1.2: Cut-off phenomenon

Figure 1.2 is a schematic graph which demonstrates the abrupt drop of the distance between X_t^n and π^n in the cut-off window. We provide the rigorous definitions of the cut-off time and the cut-off window in Section 2.1.

iii) Quasi-stationary Distribution

The quasi-stationary distribution is applied to the population model [8]. We choose the Logistic Feller diffusion process as an example. The population's size is governed by the following stochastic differential equation:

$$\mathrm{d}Z_t^\epsilon = Z_t^\epsilon \left(r - cZ_t^\epsilon\right) \mathrm{d}t + \epsilon\sqrt{Z_t^\epsilon}\mathrm{d}W_t.$$

0 is the global absorbing state. Z_t reaches extinction in finite time almost surely. Let T_0 be the first hitting time at 0. It is necessary to observe the long time behavior conditioned on the time before T_0. The quasi-stationary distribution is defined as a measure α such that for all $t > 0$,

$$\mathbb{P}_\alpha \left(Z_t \in A | T_0 > t\right) = \alpha\left(A\right).$$

The existence and uniqueness of this quasi-stationary distribution have been proved in [8]. Thus, Z_t transitions to this quasi-stationary state after a relatively

short time. Then, it maintains the quasi-stationary distribution over a long time, and is extinct at the very end.

The above three phenomena describe the transient behavior from different angles, including microscopic particles and macroscopic distribution. They are all derived from the first order systems with noise perturbation. We are more interested in the transient behavior in the second order systems case. The second order systems can be converted to the first order systems. However, that transformation will introduce singularity in the noise party. Thus, there are not too many mathematical techniques developed. This work will discuss discrete and continuous second order cases separately.

1.2 Thesis outline

1.2.1 Discrete second order system with small noise perturbation.

We first analyze a random linear recurrence equation in Chapter 2. We initially focused on the Brownian oscillator, a classical second order system with a small perturbation as follows:

$$\ddot{x}_t + \gamma \dot{x}_t + \kappa x_t = \epsilon \dot{W}_t \quad \text{for any } t \geq 0.$$

The existence and uniqueness of the strong solution are demonstrated in [4]. We utilize the forward difference scheme to obtain a discretization which is

$$X^{\epsilon}_{t+2} = (2 - \gamma h) X^{\epsilon}_{t+1} - \left(1 - \gamma h + \kappa h^2\right) X^{\epsilon}_{t} + \epsilon h^{3/2} \xi_{t+2} \quad \text{for any } t \in \mathbb{N}_0,$$

where h is the step size of time and $\{\xi_t\}$ is white noise.

We prove that if $h < \min\{2/\gamma, \gamma/\kappa\}$, then the limiting random variable X^{ϵ}_{∞} exists. Because of the 2-string past, X^{ϵ}_t is not a Markov process anymore. Diaconis presented that even for the finite Markov chains, there are not always cut-off phenomena [9]. For our non-Markov process, we discuss the existence of the cut-off phenomenon with respect to the initial conditions and the eigenvalues of its characteristic polynomial in Section 2.3. When the eigenvalues are real, the cut-off phenomenon exists for almost

all initial conditions. When the argument of the eigenvalues is rational, we found a non-empty set of initial conditions such that the cut-off phenomenon exists. When the argument is irrational, we cannot find a suitable initial condition for the cut-off phenomenon.

The general linear recurrence equation with a small white noise has been applied in various models, such as Zellner's structural econometric model [10], the stock price's model [11] and the forecasting the unemployment rate [12]. We generalize our result in the pth order case as follows:

$$X^\epsilon_{t+p} = \phi_1 X^\epsilon_{t+p-1} + \phi_2 X^\epsilon_{t+p-2} + \cdots + \phi_p X^\epsilon_t + \epsilon\xi_{t+p} \quad \text{for any } t \in \mathbb{N}_0.$$

In Section 2.1, we state the main result of the cut-off phenomenon and its consequences. The solution X^ϵ_t abruptly merges with its limiting distribution in a cut-off window under suitable conditions. The cut-off time and window are

$$t^\epsilon = \mathcal{O}\left(\ln\left(\epsilon^{-1}\right)\right) \ \textit{and}\ w^\epsilon = \mathcal{O}(1).$$

We provide the proof of them in Section 2.2. For the sake of completeness, some results about the distribution of the random linear recurrence and its limiting behavior are demonstrated in Appendix A. Appendix B summarizes some properties about the total variation distance between Gaussian distributions. Moreover, Appendix D states some elementary limit behaviors.

1.2.2 Continuous second order system with small noise perturbation.

In Chapter 3, we analyze the underdamped Langevin equation,

$$\begin{aligned} \mathrm{d}x^\epsilon_t &= y^\epsilon_t \mathrm{d}t, \\ \mathrm{d}y^\epsilon_t &= \left(-\nabla V(x^\epsilon_t) - \epsilon y^\epsilon_t\right)\mathrm{d}t + \sqrt{2\epsilon\beta^{-1}}\mathrm{d}W_t, \end{aligned} \tag{1.1}$$

where V is the potential energy and W_t is a d-dimensional Brownian motion. The corresponding unperturbed system is a Hamiltonian equation,

$$\begin{aligned} \mathrm{d}x^0_t &= y^0_t \mathrm{d}t, \\ \mathrm{d}y^0_t &= -\nabla V(x^0_t)\mathrm{d}t. \end{aligned} \tag{1.2}$$

For consistency, we still use X_t^ϵ and X_t^0 to denote the solutions of (1.1) and (1.2), respectively. Since (1.1) has a unique invariant measure which does not depend on ϵ [13], let X_∞ be the random variable whose distribution is the same as the distribution of the invariant measure.

In order to understand the transient behavior precisely, in Section 3.1, we study the one-dimensional harmonic case where $V(x) = \frac{1}{2}kx^2$. The explicit solution can be obtained in this case. We generalize the perturbation limiting behavior from finite time to a short time scale, $t_\epsilon \ll \frac{1}{\epsilon}$. X_t^ϵ is still close to X_t^0 during this relatively long time. When $t_\epsilon = \frac{\theta}{\epsilon}$, we find that a part of the distribution of X_t^ϵ coincides with the distribution of X_∞. Thus, the transient behavior occurs in this critical time scale. In the long time scale, $t_\epsilon \gg \frac{1}{\epsilon}$, X_t^ϵ merges with X_∞ when ϵ tends to zero.

We study the d-dimensional non-harmonic case in Section 3.3. In the same long time scale, $t_\epsilon \gg \frac{1}{\epsilon}$, X_t^ϵ is close to X_∞ in the total variation distance. When $t_\epsilon \ll \frac{1}{\epsilon}$, we obtain the perturbation limiting behavior of the total energy,

$$E(x, y) \coloneqq \frac{1}{2}\|y\|_2^2 + V(x),$$

which is one of the conserved quantities of (1.2). In the short time scale, E_t^ϵ tends to $E_t^0 \equiv E_0$. Therefore, X_t^ϵ is far from X_∞ whose density function's support is $\mathbb{R}^{2d}$ [13].

During the critical time scale, we study the one-dimensional case in Section 3.4. We apply the averaging principle on E which is also a slow variable of (1.1). The averaged equation is

$$\mathrm{d}\overline{E}_t^c = \left(\frac{1}{2} - I(\overline{E}_t^c)\omega(\overline{E}_t^c)\right)\mathrm{d}t + \sqrt{I(\overline{E}_t^c)\omega(\overline{E}_t^c)}\mathrm{d}W_t, \tag{1.3}$$

where I is the action variable and ω is the frequency. (1.3) approximately describes the transient behavior of E with respect to θ when ϵ is small enough. We provide some general properties of (1.3) in Section 3.4.3.

The last objective is understanding the transient behavior in full coordinates, especially in the one-dimensional case. The numerical simulations are utilized to demonstrate the transition of the fast variable of (1.1). In order to simulate in a relatively long time, we develop an efficient data-driven method in Section 3.2, which is a hybrid method combining the Monte-Carlo simulation and the simulation of the Fokker-Planck equation of (1.1) [14, 15]. The data-driven method treats the rough

Monta-Carlo result as a dataset, which reduces the large number of iterations in the classical Monte-Carlo simulation. Then, it utilizes the numerical Fokker-Planck equation's scheme as a constraint to obtain the optimal solution.

In Section 3.2.3, we demonstrate the transition of the angle which is a fast variable of (1.1). It seems that the fast variable has started to transition in the short time scale. However, to date, we cannot provide a rigorous theorem to describe the transient behavior of the fast variable. For integrity, the Matlab code of our data-driven method and the simulation of (1.3) are presented in Appendix E.

Chapter 2

The cut-off phenomenon in a linear recurrence equation with a small perturbation

Linear recurrence equations have been widely used in several areas of applied mathematics and computer science. In applied science, they can be used to model the future of a process that depends linearly on a finite string, for instance: in population dynamics to model population size and structure [16–18]; in economics to model the interest rate, the amortization of a loan and price fluctuations [19–21]; in computer science for analysis of algorithms [22, 23] and in statistics for the autoregressive linear model [24, 25]. In theoretical mathematics, people apply it to find the coefficients of series solutions in differential equations, whose details are presented in Chapters 4–5 of [26]; to prove Hilbert's tenth problem over $\mathbb{Z}$ [27] and to provide expansions of some second order operators in approximation theory [28]. For a complete understanding of applications of the linear recurrence equation, we recommend the introduction of the monograph by [29] and the references therein.

We consider random dynamics that arise from a linear homogeneous recurrence equation with a control term given by independent and identically distributed (i.i.d. for short) random variables with Gaussian distribution. To be precise, given $p \in \mathbb{N}$, $\phi_1, \phi_2, \ldots, \phi_p \in \mathbb{R}$ with $\phi_p \neq 0$, we define the linear homogeneous recurrence of degree p as follows:

$$x_{t+p} = \phi_1 x_{t+p-1} + \phi_2 x_{t+p-2} + \cdots + \phi_p x_t \quad \text{for any } t \in \mathbb{N}_0, \tag{2.1}$$

where $\mathbb{N}_0$ denotes the set of non-negative integers. To single out a unique solution

of (2.1) one should assign initial conditions $x_0, x_1, \ldots, x_{p-1} \in \mathbb{R}$. Recurrence (2.1) is called a recurrence with p-history since it only depends on a p-number of earlier values.

We consider a small perturbation of (2.1) by adding Gaussian noise as follows: given $\epsilon > 0$ fixed, consider the random dynamics

$$X^{\epsilon}_{t+p} = \phi_1 X^{\epsilon}_{t+p-1} + \phi_2 X^{\epsilon}_{t+p-2} + \cdots + \phi_p X^{\epsilon}_{t} + \epsilon \xi_{t+p} \quad \text{for any } t \in \mathbb{N}_0, \tag{2.2}$$

with initial conditions $X^{\epsilon}_0 = x_0, \ldots, X^{\epsilon}_{p-1} = x_{p-1}$, and $(\xi_t : t \geq p)$ is a sequence of i.i.d. random variables with Gaussian distribution with zero mean and variance one. Since the random linear recurrence (2.2) depends on its p-string past, it is not Markovian. However, it is straightforward to convert as a linear first order random matrix recurrence, which is Markovian. Denote by $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space where the sequence $(\xi_t : t \geq p)$ is defined, then the random dynamics (2.2) can be defined as a stochastic process in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Notice that $\epsilon > 0$ is a parameter that controls the magnitude of the noise. When $\epsilon = 0$ the deterministic model (2.1) is recovered from the stochastic model (2.2). Since $(\xi_t : t \geq p)$ is a sequence of i.i.d. random variables with Gaussian distribution, the model (2.2) could be understood as a regularization of (2.1).

To our knowledge, this model was initially used by [30] ($p = 2$) to model the presence of random disturbances of a harmonic oscillator for investigating hidden periodicities and their relation to the observations of sunspots.

In this chapter we obtain a nearly-complete characterization of the convergence in the total variation distance between the distribution of X^{ϵ}_t and its limiting distribution as t increases. Under general conditions, we state in Section 2.1 when the intensity of the control ϵ is fixed. Then, the random linear recurrence goes to a limiting distribution in the total variation distance as time passes. We show that this convergence is abrupt in the following sense: the total variation distance between the distribution of the random linear recurrence and its limiting distribution drops abruptly over a negligible time (time window) around a threshold time (cut-off time) from near one to near zero. It means that if we run the random linear recurrence before a time window around the cut-off time, the process is not well mixed and after a time window around the cut-off time becomes well mixed. This fact is known as a

cut-off phenomenon in the context of stochastic processes.

Suppose that we model a system by a random process $(X_t^\epsilon : t \geq 0)$, where the parameter ϵ denotes the intensity of the noise and assume that X_∞^ϵ is its equilibrium. A natural question that arises is the following: *with a fixed ϵ and an error $\eta > 0$, how much time $\tau(\epsilon, \eta)$ do we need to run the model $(X_t^\epsilon : t \geq 0)$ in order to be close to its equilibrium X_∞^ϵ with an error at most η in a suitable distance?* The latter is known as a *mixing time* in the context of random processes. In general, it is hard to compute and/or estimate $\tau(\epsilon, \eta)$. The cut-off phenomenon provides a strong answer in a small regime ϵ. Roughly speaking, as ϵ goes to zero, it means that after a deterministic time $\tau^*(\epsilon)$, the system is "almost" in its equilibrium within any error η. We provide a precise definition in Section 2.1.

The cut-off phenomenon was extensively studied in the eighties to describe the phenomenon of abrupt convergence, which appears, for example, in models of cards' shuffling, Ehrenfests' urn, and random transpositions, see for instance [9]. In general, it is a challenging problem to prove that a specific model exhibits a cut-off phenomenon. It requires a complete understanding of the dynamics of the specific random process. For an introduction to this concept, we recommend Chapter 18 in [31] for discrete Markov chains in a finite state, [32] for discrete Markov chains with infinite countable state space and [33–35] for Stochastic Differential Equations in a continuous state space.

2.1 Main theorem

One of the most important problems in dynamical systems is the study of the limit behavior of their evolution for forward times. To the linear recurrence (2.1) we can associate a characteristic polynomial

$$f(\lambda) = \lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_p \quad \text{for any } \lambda \in \mathbb{C}. \tag{2.3}$$

From now on to the end of this chapter, we assume

(**H**) all the roots of (2.3) have modulus strictly less than one.

From (**H**) we can prove that for any string of initial values $x_0, \ldots, x_{p-1} \in \mathbb{R}$, x_t goes exponentially fast to zero as t goes to infinity (see Theorem 1 in [36] for more details).

In the stochastic model (2.2), (**H**) implies that the process $(X_t^\epsilon, t \in \mathbb{N}_0)$ is strongly ergodic, i.e., for any initial conditions $x_0, \ldots, x_{p-1}$, the random recurrence X_t^ϵ converges in the so-called total variation distance as t goes to infinity to a random variable X_∞^ϵ. For further details see Lemma 27 in Appendix A.

Given $m \in \mathbb{R}$ and $\sigma^2 \in (0, +\infty)$, denote by $\mathcal{N}(m, \sigma^2)$ the Gaussian distribution with mean m and variance σ^2. Later on, we see that for $t \geq p$, the random variable X_t^ϵ has distribution $\mathcal{N}(x_t, \epsilon^2\sigma_t^2)$, where x_t is given by (2.1) and $\sigma_t^2 \in (0, +\infty)$. Moreover, the random variable X_∞^ϵ has a distribution $\mathcal{N}(0, \epsilon^2\sigma_\infty^2)$ with $\sigma_\infty^2 \in (0, +\infty)$.

Since the random recurrence (2.2) is linear in the inputs which are independent Gaussian random variables, the distribution of X_t^ϵ (for $t \geq p$) and its limiting distribution X_∞^ϵ is also Gaussian. For details, see Lemma 26 and Lemma 27 in Appendix A. Then a natural way to measure its discrepancy is by the total variation distance. Given two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ on the measure space $(\Omega, \mathcal{F})$, the total variation distance between the probabilities $\mathbb{P}_1$ and $\mathbb{P}_2$ is given by

$$\mathbf{d}_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{F \in \mathcal{F}} |\mathbb{P}_1(F) - \mathbb{P}_2(F)|.$$

When X, Y are random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we write $\mathbf{d}_{\mathrm{TV}}(X, Y)$ instead of $\mathbf{d}_{\mathrm{TV}}(\mathbb{P}(X \in \cdot), \mathbb{P}(Y \in \cdot))$, where $\mathbb{P}(X \in \cdot)$ and $\mathbb{P}(Y \in \cdot)$ denote the distribution of X and Y under $\mathbb{P}$, respectively. Then we define

$$d^\epsilon(t) := \mathbf{d}_{\mathrm{TV}}\left(X_t^\epsilon, X_\infty^\epsilon\right) = \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}(x_t, \epsilon^2\sigma_t^2), \mathcal{N}(0, \epsilon^2\sigma_\infty^2)\right)$$

for any $t \geq p$. Notice that the above distance depends on the initial conditions $x_0, \ldots, x_{p-1} \in \mathbb{R}$. To make the notation more fluid, we avoid its dependence from our notation. For a complete understanding of the total variation distance between two arbitrary probabilities with densities, we recommend Section 3.3 in [37] and Section 2.2 in [38]. Nevertheless, for the sake of completeness, we provide an Appendix B that contains the properties and bounds for the total variation distance between Gaussian distributions that we used to prove Theorem 2, which is the main theorem of this chapter.

The goal is to study the so-called *cut-off phenomenon* in the total variation distance when ϵ goes to zero for the family of the stochastic processes

$$(X^\epsilon := (X_t^\epsilon : t \in \mathbb{N}_0) : \epsilon > 0)$$

for fixed initial conditions $x_0, \dots, x_{p-1}$.

In the sequel we introduce the formal definition of cut-off phenomenon. Recall that for any $z \in \mathbb{R}$, $\lfloor z \rfloor$ denotes the greatest integer less than or equal to z. Consider the family of stochastic processes $(X^\epsilon : \epsilon > 0)$. According to [39], the cut-off phenomenon can be expressed in three increasingly sharp levels as follows:

Definition 1 *The family $(X^\epsilon : \epsilon > 0)$ has*

i) cut-off at $(t^\epsilon : \epsilon > 0)$ with cut-off time t^ϵ if t^ϵ goes to infinity as ϵ goes to zero and

$$\lim_{\epsilon \to 0^+} d^\epsilon(\lfloor \delta t^\epsilon \rfloor) = \begin{cases} 1 & \text{if } 0 < \delta < 1, \\ 0 & \text{if } \delta > 1. \end{cases}$$

ii) window cut-off at $((t^\epsilon, w^\epsilon) : \epsilon > 0)$ with cut-off time t^ϵ and time cut-off w^ϵ if t^ϵ goes to infinity as ϵ goes to zero, $w^\epsilon = o(t^\epsilon)$ and

$$\lim_{b \to -\infty} \liminf_{\epsilon \to 0^+} d^\epsilon(\lfloor t^\epsilon + b w^\epsilon \rfloor) = 1$$

and

$$\lim_{b \to +\infty} \limsup_{\epsilon \to 0^+} d^\epsilon(\lfloor t^\epsilon + b w^\epsilon \rfloor) = 0.$$

iii) profile cut-off at $((t^\epsilon, w^\epsilon) : \epsilon > 0)$ with cut-off time t^ϵ, time cut-off w^ϵ and profile function $G : \mathbb{R} \to [0,1]$ if t^ϵ goes to infinity as ϵ goes to zero, $w^\epsilon = o(t^\epsilon)$,

$$\lim_{\epsilon \to 0^+} d^\epsilon(\lfloor t^\epsilon + b w^\epsilon \rfloor) =: G(b) \qquad \text{exists for any } b \in \mathbb{R}$$

together with $\lim_{b \to -\infty} G(b) = 1$ and $\lim_{b \to +\infty} G(b) = 0$.

Bearing all this in mind, we can analyze how this convergence happens, precisely the statement of the following theorem.

Theorem 2 (Main theorem) *Assume that (**H**) holds. For a given initial condition $x = (x_0, \dots, x_{p-1}) \in \mathbb{R}^p \setminus \{0_p\}$ assume that there exist $r = r(x) \in (0,1)$, $l = l(x) \in \{1, \dots, p\}$ and $v_t = v(t,x) \in \mathbb{R}$ such that*

i)

$$\lim_{t \to +\infty} \left| \frac{x_t}{t^{l-1} r^t} - v_t \right| = 0,$$

ii) $\sup_{t\geq 0} |v_t| < +\infty$,

iii) $\liminf_{t\to+\infty} |v_t| > 0$.

Then the family of random linear recurrences

$$(X^\epsilon := (X^\epsilon(t) : t \in \mathbb{N}_0) : \epsilon > 0)$$

has window cut-off as ϵ *goes to zero with cut-off time*

$$t^\epsilon = \frac{\ln(1/\epsilon)}{\ln(1/r)} + (l-1)\frac{\ln\left(\frac{\ln(1/\epsilon)}{\ln(1/r)}\right)}{\ln(1/r)}$$

and time window

$$w^\epsilon = C + o_\epsilon(1),$$

where C *is any positive constant and* $\lim_{\epsilon\to 0^+} o_\epsilon(1) = 0$. *In other words,*

$$\lim_{b\to-\infty} \liminf_{\epsilon\to 0^+} d^\epsilon(\lfloor t^\epsilon + bw^\epsilon \rfloor) = 1$$

and

$$\lim_{b\to+\infty} \limsup_{\epsilon\to 0^+} d^\epsilon(\lfloor t^\epsilon + bw^\epsilon \rfloor) = 0,$$

where $d^\epsilon(t) = \mathbf{d}_{\mathrm{TV}}\left(X_t^\epsilon, X_\infty^\epsilon\right)$ *for any* $t \geq p$.

Roughly speaking, the argument of the proof consists of fairly intricate calculations of the distributions of X_t^ϵ, $t \geq p$ and its limiting distribution X_∞^ϵ whose distributions are Gaussian. Then the cut-off phenomenon is proved from a refined analysis of their means and variances and "explicit calculations and bounds" for the total variation distance between Gaussian distributions. This analysis also provides a mild case where the cut-off phenomenon does not occur.

Remark 3 *Notice that* $\sup_{t\geq 0} |v_t| < +\infty$ *and* $\limsup_{t\to+\infty} |v_t| < +\infty$ *are actually equivalent. However,* $\liminf_{t\to+\infty} |v_t| > 0$ *does not always imply* $\inf_{t\geq 0} |v_t| > 0$.

Remark 4 *Roughly speaking, the number* r *corresponds to the absolute value of some roots of* (2.3) *and* l *is related to their multiplicities.*

Remark 5 *Under the conditions of Theorem 2, the total variation distance between the distribution of X_t^ϵ and its limiting distribution X_∞^ϵ changes abruptly from one to zero in a time window w^ϵ of constant order around the cut-off time t^ϵ of logarithmic order in ϵ.*

We introduce the definition of maximal set. We say that a set $\mathcal{A} \subset \mathbb{R}^p$ is a maximal set that satisfies the property **P** if and only if any set $\mathcal{B} \subset \mathbb{R}^d$ that satisfies the property **P** is a subset of $\mathcal{A}$.

In the case when all the roots of (2.3) are real numbers we will see in Lemma 9 that there exists a maximal set $\mathcal{C} \subset \mathbb{R}^p$ such that any initial condition $x := (x_0, \ldots, x_{p-1}) \in \mathcal{C}$ fulfills Condition i), Condition ii) and Condition iii) of Theorem 2. Moreover, $\mathcal{C}$ has full measure with respect to the Lebesgue measure on $\mathbb{R}^p$. If we only assume (**H**) and no further assumptions, we will see in Corollary 13 that Condition iii) of Theorem 2 may not hold. We conjecture that cut-off phenomenon does not hold when condition iii) fails.

Remark 6 *If* (**H**) *does not hold, then it is not hard to prove that the variance of X_t^ϵ tends to $+\infty$ as $t \to +\infty$. As a consequence, the random linear recurrence* (2.2) *does not converge in distribution. Therefore the cut-off phenomenon problem is not well-posed.*

2.2 Proof

Observe that for any $t \geq p$, X_t^ϵ has Gaussian distribution with mean x_t and variance $\sigma^2(t, \epsilon, x_0, \ldots, x_{p-1}) \in (0, +\infty)$. By Lemma 26 in Appendix A, under assumption (**H**), we obtain $\sigma^2(t, \epsilon, x_0, \ldots, x_{p-1}) = \epsilon^2 \sigma_t^2$, where $\sigma_t^2 \in [1, +\infty)$ and it does not depend on the initial conditions $x_0, x_1, \ldots, x_{p-1}$.

The following lemma asserts that the random dynamics (2.2) is strongly ergodic when (**H**) holds.

Lemma 7 *Assume that* (**H**) *holds. As t goes to infinity, X_t^ϵ converges in the total variation distance to a random variable X_∞^ϵ that has Gaussian distribution with zero mean and variance $\epsilon^2 \sigma_\infty^2 \in [\epsilon^2, +\infty)$.*

For the sake of brevity, the proof of the last lemma is given in Lemma 27 in Appendix A. Recall that

$$d^{\epsilon}(t) = \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}(x_t, \epsilon^2\sigma_t^2), \mathcal{N}(0, \epsilon^2\sigma_\infty^2)\right) \qquad \text{for any } t \geq p.$$

In order to analyze the cut-off phenomenon for the distance $d^{\epsilon}(t)$, for the convenience of computations we first study another distance, as the following lemma states.

Lemma 8 *For any $t \geq p$ we have*

$$|d^{\epsilon}(t) - D^{\epsilon}(t)| \leq R(t)$$

where

$$D^{\epsilon}(t) = \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{x_t}{\epsilon\sigma_\infty}, 1\right), \mathcal{N}(0,1)\right)$$

and

$$R(t) = \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(0, \sigma_t^2), \mathcal{N}(0, \sigma_\infty^2)).$$

Proof. Notice that the expressions $d^{\epsilon}(t)$ and $D^{\epsilon}(t)$ depend on the parameter ϵ and the initial conditions $x_0, x_1, \ldots, x_{p-1}$. Nevertheless, the term $R(t)$ does not depend on ϵ and on the initial conditions $x_0, x_1, \ldots, x_{p-1}$. Let $t \geq p$. By the triangle inequality we obtain

$$d^{\epsilon}(t) \leq \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}(x_t, \epsilon^2\sigma_t^2), \mathcal{N}(x_t, \epsilon^2\sigma_\infty^2)\right) + \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}(x_t, \epsilon^2\sigma_\infty^2), \mathcal{N}(0, \epsilon^2\sigma_\infty^2)\right).$$

By i) and ii) of Lemma 28 we have

$$d^{\epsilon}(t) \leq R(t) + D^{\epsilon}(t).$$

On the other hand, by ii) of Lemma 28 we notice

$$D^{\epsilon}(t) = \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}(x_t, \epsilon^2\sigma_\infty^2), \mathcal{N}(0, \epsilon^2\sigma_\infty^2)\right).$$

By the triangle inequality we obtain

$$D^{\epsilon}(t) \leq \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}(x_t, \epsilon^2\sigma_\infty^2), \mathcal{N}(x_t, \epsilon^2\sigma_t^2)\right) + \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}(x_t, \epsilon^2\sigma_t^2), \mathcal{N}(0, \epsilon^2\sigma_\infty^2)\right).$$

Again, by i) and ii) of Lemma 28 we have

$$D^{\epsilon}(t) \leq R(t) + d^{\epsilon}(t).$$

Gluing all pieces together we deduce

$$|d^\epsilon(t) - D^\epsilon(t)| \leq R(t) \quad \text{for any } t \geq p.$$

■

Now, we have all the tools to prove Theorem 2.

Proof of Theorem 2. By Lemma 7 and Lemma 31 we have

$$\lim_{t\to+\infty} R(t) = 0.$$

In order to analyze $D^\epsilon(t)$ we observe that

$$\frac{x_t}{\epsilon\sigma_\infty} = \frac{t^{l-1}r^t}{\epsilon\sigma_\infty}\left(\frac{x_t}{t^{l-1}r^t} - v_t\right) + \frac{t^{l-1}r^t}{\epsilon\sigma_\infty}v_t, \tag{2.4}$$

where $l \in \{1,\dots,p\}$, $r \in (0,1)$, and v_t are given by Condition i). By Lemma 34 in Appendix D we have

$$\lim_{\epsilon\to0^+} \frac{(t^\epsilon)^{l-1}r^{t^\epsilon}}{\epsilon} = 1.$$

For any $t \geq 0$, define $p_t = \frac{t^{l-1}r^t}{\epsilon\sigma_\infty}\left(\frac{x_t}{t^{l-1}r^t} - v_t\right)$ and $q_t = \frac{t^{l-1}r^t}{\epsilon\sigma_\infty}v_t$. Then for any $b \in \mathbb{R}$ we have

$$|p_{\lfloor t^\epsilon+bw^\epsilon\rfloor}| \leq \left(\frac{t^\epsilon+bw^\epsilon}{t^\epsilon}\right)^{l-1} \frac{(t^\epsilon)^{l-1}r^{t^\epsilon+bw^\epsilon-1}}{\epsilon\sigma_\infty} \times \left|\frac{x_{\lfloor t^\epsilon+bw^\epsilon\rfloor}}{(\lfloor t^\epsilon+bw^\epsilon\rfloor)^{l-1}r^{\lfloor t^\epsilon+bw^\epsilon\rfloor}} - v_{\lfloor t^\epsilon+bw^\epsilon\rfloor}\right|.$$

By Condition i) we have

$$\lim_{\epsilon\to0^+} p_{\lfloor t^\epsilon+bw^\epsilon\rfloor} = 0 \quad \text{for any } b \in \mathbb{R}. \tag{2.5}$$

Now, we analyze an upper bound for $|q_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|$. Notice that

$$|q_{\lfloor t^\epsilon+bw^\epsilon\rfloor}| \leq \left(\frac{t^\epsilon+bw^\epsilon}{t^\epsilon}\right)^{l-1} \frac{(t^\epsilon)^{l-1}r^{t^\epsilon+bw^\epsilon-1}}{\epsilon\sigma_\infty}M,$$

where $M = \sup_{t\geq0}|v_t|$. By Condition ii) we know $M < +\infty$. Then

$$\limsup_{\epsilon\to0^+} |q_{\lfloor t^\epsilon+bw^\epsilon\rfloor}| \leq \frac{Mr^{bC-1}}{\sigma_\infty} \quad \text{for any } b \in \mathbb{R}. \tag{2.6}$$

From equality (2.4), relation (2.5), inequality (2.6) and ii) of Lemma 33 we have

$$\limsup_{\epsilon\to0^+} \frac{|x_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|}{\epsilon\sigma_\infty} \leq \frac{Mr^{bC-1}}{\sigma_\infty} \quad \text{for any } b \in \mathbb{R}.$$

Using i) of Lemma 32 we have

$$\limsup_{\epsilon\to 0^+} \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{|x_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|}{\epsilon\sigma_\infty},1\right),\mathcal{N}(0,1)\right)\leq \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{Mr^{bC-1}}{\sigma_\infty},1\right),\mathcal{N}(0,1)\right)$$

for any $b\in\mathbb{R}$. Since $r\in(0,1)$, by Lemma 31 we have

$$\lim_{b\to+\infty}\limsup_{\epsilon\to 0^+} \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{|x_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|}{\epsilon\sigma_\infty},1\right),\mathcal{N}(0,1)\right)=0. \tag{2.7}$$

In order to analyze a lower bound for $|q_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|$, note

$$|q_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|\geq\left(\frac{t^\epsilon+bw^\epsilon-1}{t^\epsilon}\right)^{l-1}\frac{(t^\epsilon)^{l-1}r^{t^\epsilon+bw^\epsilon}}{\epsilon\sigma_\infty}|v_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|$$

for any $b\in\mathbb{R}$. By Condition iii) and iii) of Lemma 33 we have

$$\liminf_{\epsilon\to 0^+}|q_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|\geq\frac{r^{bC}}{\sigma_\infty}\liminf_{\epsilon\to 0^+}|v_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|\geq\frac{mr^{bC}}{\sigma_\infty}, \tag{2.8}$$

where $m=\liminf_{t\to+\infty}|v_t|\in(0,+\infty)$. From equality (2.4), relation (2.5), inequality (2.8) and ii) of Lemma 33 we have

$$\liminf_{\epsilon\to 0^+}\frac{|x_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|}{\epsilon\sigma_\infty}\geq\frac{mr^{bC}}{\sigma_\infty}\quad\text{ for any } b\in\mathbb{R}.$$

From ii) of Lemma 32 we have

$$\liminf_{\epsilon\to 0^+} \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{|x_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|}{\epsilon\sigma_\infty},1\right),\mathcal{N}(0,1)\right)\geq \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{r^{bC}}{\sigma_\infty}m,1\right),\mathcal{N}(0,1)\right)$$

for any $b\in\mathbb{R}$. Since $r\in(0,1)$, by iii) Lemma 29 we have

$$\lim_{b\to-\infty}\liminf_{\epsilon\to 0^+} \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{|x_{\lfloor t^\epsilon+bw^\epsilon\rfloor}|}{\epsilon\sigma_\infty},1\right),\mathcal{N}(0,1)\right)=1. \tag{2.9}$$

From (2.7) and (2.9) we have

$$\lim_{b\to+\infty}\limsup_{\epsilon\to 0^+} D^\epsilon(\lfloor t^\epsilon+bw^\epsilon\rfloor)=0 \text{ and } \lim_{b\to-\infty}\liminf_{\epsilon\to 0^+} D^\epsilon(\lfloor t^\epsilon+bw^\epsilon\rfloor)=1.$$

Recall that $\lim_{t\to+\infty} R(t)=0$. By Lemma 8 and i) of Lemma 33 we obtain

$$\limsup_{\epsilon\to 0^+} d^\epsilon(\lfloor t^\epsilon+bw^\epsilon\rfloor)\leq\limsup_{\epsilon\to 0^+} D^\epsilon(\lfloor t^\epsilon+bw^\epsilon\rfloor).$$

Now, sending $b \to +\infty$ we have

$$\lim_{b\to+\infty} \limsup_{\epsilon\to 0^+} d^\epsilon(\lfloor t^\epsilon + bw^\epsilon \rfloor) = 0.$$

Similarly, by Lemma 8 and ii) of Lemma 33 we obtain

$$\liminf_{\epsilon\to 0^+} D^\epsilon(\lfloor t^\epsilon + bw^\epsilon \rfloor) \leq \liminf_{\epsilon\to 0^+} d^\epsilon(\lfloor t^\epsilon + bw^\epsilon \rfloor).$$

Now, sending $b \to -\infty$ we have

$$\lim_{b\to-\infty} \liminf_{\epsilon\to 0^+} d^\epsilon(\lfloor t^\epsilon + bw^\epsilon \rfloor) = 1.$$

■

Now, we provide a precise estimate of the rate of the convergence to zero of (2.1). Let us recall some well-known facts about p-linear recurrences. By the celebrated Fundamental Theorem of Algebra we have at most p roots in the complex numbers for (2.3). Denote by $\lambda_1, \ldots, \lambda_q \in \mathbb{C}$ the different roots of (2.3) with multiplicity $m_1, \ldots, m_q$ respectively, where $1 \leq q \leq p$. Then

$$x_t = \sum_{j_1=1}^{m_1} c_{1,j_1} t^{j_1-1} \lambda_1^t + \sum_{j_2=1}^{m_2} c_{2,j_2} t^{j_2-1} \lambda_2^t + \ldots + \sum_{j_q=1}^{m_q} c_{q,j_q} t^{j_q-1} \lambda_q^t \tag{2.10}$$

for any $t \in \mathbb{N}_0$, where the coefficients $c_{1,1}, \ldots, c_{1,m_1}, \ldots, c_{q,1}, \ldots, c_{q,m_q}$ are uniquely obtained from the initial conditions $x_0, \ldots, x_{p-1}$ (see Theorem 1 in [36] for more details). Moreover, for any initial conditions $(x_0, \ldots, x_{p-1}) \in \mathbb{R}^p \setminus \{0_p\}$ we have

$$(c_{1,1}, \ldots, c_{1,m_1}, \ldots, c_{q,1}, \ldots, c_{q,m_q}) \in \mathbb{C}^p \setminus \{0_p\}.$$

Notice that the right-hand side of (2.10) may have complex numbers. When all the roots of (2.3) are real numbers we can establish the precise exponential behavior of x_t as t goes by.

Lemma 9 (Real roots) *Assume that all the roots of* (2.3) *are real numbers. Then there exists a non-empty maximal set* $\mathcal{C} \subset \mathbb{R}^p$ *that satisfies the following: for any* $x = (x_0, \ldots, x_{p-1}) \in \mathcal{C}$ *there exist* $r := r(x) > 0$, $l := l(x) \in \{1, \ldots, p\}$ *and* $v_t := v(t, x) \in \mathbb{R}$ *such that*

$$\lim_{t\to+\infty} \left| \frac{x_t}{t^{l-1} r^t} - v_t \right| = 0.$$

Moreover, we have $\sup_{t\geq 0} |v_t| < +\infty$ *and* $\liminf_{t\to+\infty} |v_t| > 0$.

Proof. Recall that the constants $c_{1,1}, \ldots, c_{1,m_1}, \ldots, c_{q,1}, \ldots, c_{q,m_q}$ in representation (2.10) depend on the initial conditions $x_0, x_1, \ldots, x_{p-1}$. In order to avoid technicalities, without loss of generality we assume that for each $1 \leq j \leq q$ there exists at least one $1 \leq k \leq m_j$ such that $c_{j,j_k} \neq 0$. If the last assumption is not true for some $1 \leq j \leq q$, then the root λ_j does not appear in representation (2.10) for an specific initial conditions $x_0, x_1, \ldots, x_{p-1}$, then we can remove it from representation (2.10) and apply the method described below.

Denote by $r = \max\limits_{1 \leq j \leq q} |\lambda_j| > 0$. Since all the roots of (2.3) are real numbers, after multiplicity at most two roots of (2.3) have the same absolute value. The function $\text{sign}(\cdot)$ is defined over the domain $\mathbb{R} \setminus \{0\}$ by $\text{sign}(x) = {}^{x}\!/_{|x|}$. Only one of the following cases can occur.

i) There exists a unique $1 \leq j \leq q$ such that $|\lambda_j| = r$. Let

$$l = \max\{1 \leq s \leq m_j : c_{j,s} \neq 0\}.$$

Then

$$\lim_{t \to +\infty} \left| \frac{x_t}{t^{l-1} r^t} - c_{j,l} (\text{sign}(\lambda_j))^t \right| = 0.$$

In this case $\mathcal{C} = \mathbb{R}^p \setminus \{0_p\}$.

ii) There exist $1 \leq j < k \leq q$ such that $|\lambda_j| = |\lambda_k| = r$. Without loss of generality, we can assume $0 < \lambda_k = -\lambda_j$. Let

$$l_j = \max\{1 \leq s \leq m_j : c_{j,s} \neq 0\}$$

and

$$l_k = \max\{1 \leq s \leq m_k : c_{k,s} \neq 0\}.$$

If $l_j < l_k$ or $l_k < l_j$ then by taking $l = \max\{l_j, l_k\}$ we have

$$\lim_{t \to +\infty} \left| \frac{x_t}{t^{l-1} r^t} - c_{\star,l} (\text{sign}(\lambda_\star))^t \right| = 0,$$

where $\star = j$ if $l_j = l$ and $\star = k$ if $l_k = l$. In this case $\mathcal{C} = \mathbb{R}^p \setminus \{0_p\}$. If $l_j = l_k$ then by taking $l = l_j$, $v_t = (-1)^t c_{j,l} + c_{k,l}$ we have

$$\lim_{t \to +\infty} \left| \frac{x_t}{t^{l-1} r^t} - v_t \right| = 0.$$

Notice that $\sup_{t\geq 0}|v_t| < +\infty$. By taking

$$\mathcal{C} = \{(x_0, \ldots, x_{p-1}) \in \mathbb{R}^p : -c_{j,l} + c_{k,l} \neq 0 \text{ and } c_{j,l} + c_{k,l} \neq 0\}$$

we have $\liminf_{t\to+\infty}|v_t| > 0$.

■

Remark 10 *From the proof of Lemma 9, we can state precisely $\mathcal{C}$. Moreover, $\mathcal{C}$ has full measure with respect to the Lebesgue measure on $\mathbb{R}^p$.*

Rather than the real roots case, the following lemma provides a fine estimate about the behavior of (2.1) as t increases in general setting.

Lemma 11 (General case) *For any $x = (x_0, \ldots, x_{p-1}) \in \mathbb{R}^p \setminus \{0_p\}$ there exist $r := r(x) > 0$, $l := l(x) \in \{1, \ldots, p\}$ and $v_t := v(t, x) \in \mathbb{R}$ such that*

$$\lim_{t\to+\infty}\left|\frac{x_t}{t^{l-1}r^t} - v_t\right| = 0,$$

where

$$v_t = \sum_{j=1}^{m}\left(\alpha_j\cos(2\pi\theta_j t) + \beta_j\sin(2\pi\theta_j t)\right)$$

with $(\alpha_j, \beta_j) := (\alpha_j(x), \beta_j(x)) \in \mathbb{R}^2 \setminus \{(0,0)\}$, $m := m(x) \in \{1, \ldots, p\}$, and $\theta_j := \theta(x) \in [0, 1)$ for any $j \in \{1, \ldots, m\}$. Moreover, $\sup_{t\geq 0}|v_t| < +\infty$.

Proof. From (2.10) we have for any $t \in \mathbb{N}_0$

$$x_t = \sum_{j_1=1}^{m_1} c_{1,j_1} t^{j_1-1}\lambda_1^t + \sum_{j_2=1}^{m_2} c_{2,j_2} t^{j_2-1}\lambda_2^t + \ldots + \sum_{j_q=1}^{m_q} c_{q,j_q} t^{j_q-1}\lambda_q^t.$$

Without loss of generality we assume that for any $k \in \{1, \ldots, q\}$ there exists $j \in \{1, \ldots, m_k\}$ such that $c_{k,j} \neq 0$. Let $l_k := \max\{1 \leq j \leq m_k : c_{k,j} \neq 0\}$. Then x_t can be rewritten as

$$x_t = \sum_{j_1=1}^{l_1} c_{1,j_1} t^{j_1-1}\lambda_1^t + \sum_{j_2=1}^{l_2} c_{2,j_2} t^{j_2-1}\lambda_2^t + \ldots + \sum_{j_q=1}^{l_q} c_{q,j_q} t^{j_q-1}\lambda_q^t,$$

where $c_{k,l_k} \neq 0$ for each k. For each k let $r_k := \|\lambda_k\|$ be its complex modulus. Without loss of generality we assume:

i) $r_1 \leq \cdots \leq r_q$,

ii) there exists an integer $\tilde{h}$ such that $r_{\tilde{h}} = \cdots = r_q$,

iii) $l_{\tilde{h}} \leq \cdots \leq l_q$,

iv) there exists an integer $h \geq \tilde{h}$ such that $l_h = \cdots = l_q$.

Let $r := r_q$ and $l := l_q$. By taking $v_t = r^{-t}(c_{h,l}\lambda_h^t + \cdots + c_{q,l}\lambda_q^t)$ we have

$$\lim_{t\to+\infty} \left| \frac{x_t}{t^{l-1}r^t} - v_t \right| = 0,$$

where $\lambda_h, \ldots, \lambda_q$ have the same modulus r, but they have different arguments $\theta_j \in [0, 1)$. Then

$$v_t = \sum_{j=h}^{q} \left(\alpha_j \cos(2\pi\theta_j t) + \beta_j \sin(2\pi\theta_j t)\right).$$

Since $c_{k,l_k} \neq 0$ for each $h \leq k \leq q$, α_j and β_j are not both zero for any $h \leq j \leq q$. After relabelling we have the desired result. ■

Remark 12 *Under no further conditions on Lemma 11, we cannot guarantee that* $\liminf_{t\to+\infty} |v_t| > 0$. *For instance, the following corollary provides sufficient conditions for which* $\liminf_{t\to+\infty} |v_t| = 0$.

Following [40], the numbers $\vartheta_1, \ldots, \vartheta_m$ are rationally independent if

$$k_1\vartheta_1 + \ldots + k_m\vartheta_m \notin \mathbb{Z} \quad \text{for any } (k_1, \ldots, k_m) \in \mathbb{Z}^m \setminus \{0_m\}.$$

Corollary 13 *Assume that* $\theta_1, \ldots, \theta_m$ *are rationally independent. Then* $\liminf_{t\to+\infty} |v_t| = 0$.

Proof. For any $j \in \{1, \ldots, m\}$ notice that $d_j := \sqrt{\alpha_j^2 + \beta_j^2} > 0$, and let $\cos(\gamma_j) = {}^{\alpha_j}/_{d_j}$ and $\sin(\gamma_j) = {}^{\beta_j}/_{d_j}$. Then v_t can be rewritten as

$$v_t = \sum_{j=1}^{m} d_j \cos(2\pi\theta_j t - \gamma_j).$$

Let $\gamma = -(\frac{\gamma_1}{2\pi}, \ldots, \frac{\gamma_m}{2\pi})$ be in the m-dimensional torus $(\mathbb{R}/\mathbb{Z})^m$. Then the set $\{(\gamma + (\theta_1 t, \ldots, \theta_m t)) \in (\mathbb{R}/\mathbb{Z})^m, t \in \mathbb{N}\}$ is dense in $(\mathbb{R}/\mathbb{Z})^m$ (see Corollary 4.2.3 in [40] for more detials). Consequently, $\liminf_{t\to+\infty} |v_t| = 0$. ■

2.3 Example

In this section, we consider the celebrated Brownian oscillator

$$m\ddot{x}_t + \gamma\dot{x}_t + \kappa x_t = \epsilon\dot{W}_t \quad \text{for any } t \geq 0, \tag{2.11}$$

where x_t denotes the position at time t of the holding mass m with respect to its equilibrium position, $\gamma > 0$ denotes the damping constant, $\kappa > 0$ denotes the restoration constant (Hooke's constant) and $(W_t : t \geq 0)$ is a Brownian motion. For each initial displacement from the equilibrium position $x_0 = u$ and initial velocity $\dot{x}_0 = v$ we have a unique solution of (2.11). For further details, see Chapter 8 in [4].

Without loss of generality, we can assume that the mass m is one. Using the classical forward difference approximation with the step size $h > 0$ (fixed), we obtain

$$\frac{1}{h^2}(x_{(n+2)h} - 2x_{(n+1)h} + x_{nh}) + \frac{\gamma}{h}(x_{(n+1)h} - x_{nh}) + \kappa x_{nh} = \frac{\epsilon}{h}(B_{(n+3)h} - B_{(n+2)h})$$

for any $n \in \mathbb{N}_0$ with the initial condition $x_0 = u$ and $x_h = x_0 + \dot{x}_0 h = u + vh$. For consistency, let $X_t = x_{th}$ for any $t \in \mathbb{N}_0$. The latter can be rewritten as

$$X^\epsilon_{t+2} = (2 - \gamma h)\, X^\epsilon_{t+1} - \left(1 - \gamma h + \kappa h^2\right) X^\epsilon_t + \epsilon h(B_{(t+3)h} - B_{(t+2)h}) \tag{2.12}$$

for any $t \in \mathbb{N}_0$. Notice that the sequence $(B_{(t+3)h} - B_{(t+2)h} : t \in \mathbb{N}_0)$ are i.i.d. random variables with Gaussian distribution with zero mean and variance h. Therefore

$$X^\epsilon_{t+2} = (2 - \gamma h)\, X^\epsilon_{t+1} - \left(1 - \gamma h + \kappa h^2\right) X^\epsilon_t + \epsilon h^{3/2}\xi_{t+2} \quad \text{for any } t \in \mathbb{N}_0,$$

where $(\xi_{t+2} : t \in \mathbb{N}_0)$ is a sequence of i.i.d. random variables with standard Gaussian distribution. This is exactly a linear recurrence of degree 2 with control sequence $(\epsilon h^{3/2}\xi_{t+2} : t \in \mathbb{N}_0)$, and its characteristic polynomial is given by

$$\lambda^2 + (\gamma h - 2)\lambda + (1 - \gamma h + \kappa h^2). \tag{2.13}$$

To fulfill assumption (**H**) we deduce the following conditions.

i) If $\gamma^2 - 4\kappa > 0$, then polynomial (2.13) has two distinct real roots. In this case a sufficient condition to verify (**H**) is $h \in (0, 2/\gamma)$.

ii) If $\gamma^2 - 4\kappa = 0$, then polynomial (2.13) has two repeated real roots. In this case (**H**) is equivalent to $h \in (0, \gamma/\kappa)$.

iii) If $\gamma^2 - 4\kappa < 0$, then polynomial (2.13) has two complex conjugate roots. In this case (**H**) is equivalent to $h \in (0, \gamma/\kappa)$.

In other words, there exists $h^* \in (0,1)$ such that for each $h \in (0, h^*)$ the characteristic polynomial (2.13) satisfies assumption (**H**). From here to the end of this section, we assume that $h \in (0, h^*)$.

Now, we compute r, l, v_t and $\mathcal{C}$ which appear in Lemma 9. Let λ_1 and λ_2 be roots of (2.13). Denote $r_1 = \|\lambda_1\|$ and $r_2 = \|\lambda_2\|$. Recall the function $\text{sign}(\cdot)$ is defined over the domain $\mathbb{R} \setminus \{0\}$ by $\text{sign}(x) = x/|x|$. We assume that $(x_0, x_1) \neq (0,0)$. We analyze as far as possible when the conditions of Theorem 2 are fulfilled for the model (2.12).

i) **Real roots with different absolute values.** λ_1 and λ_2 are real and $r_1 \neq r_2$. In this case,

$$x_t = c_1\lambda_1^t + c_2\lambda_2^t \quad \text{for any } t \in \mathbb{N}_0,$$

where c_1 and c_2 are real constants determined by initial conditions x_0, x_1. Since $(x_0, x_1) \neq (0,0)$, we have $(c_1, c_2) \neq (0,0)$. Without loss of generality assume that $r_1 > r_2$.

i.1) If $c_1 \neq 0$ then

$$\lim_{t\to+\infty} \left| \frac{x_t}{r_1^t} - c_1(\text{sign}(\lambda_1))^t \right| = 0.$$

i.2) If $c_1 = 0$ then $c_2 \neq 0$. Therefore

$$\lim_{t\to+\infty} \left| \frac{x_t}{r_2^t} - c_2(\text{sign}(\lambda_2))^t \right| = 0.$$

Consequently, $\mathcal{C} = \mathbb{R}^2 \setminus \{(0,0)\}$.

ii) **Real roots with the same absolute value.** λ_1 and λ_2 are real and $r := r_1 = r_2$.

ii.1) If $\lambda_1 = \lambda_2 = r\text{sign}(\lambda_1)$ then

$$x_t = c_1 r^t(\text{sign}(\lambda_1))^t + c_2 t r^t(\text{sign}(\lambda_1))^t \quad \text{for any } t \in \mathbb{N}_0,$$

where c_1 and c_2 are real constants determined by initial conditions x_0, x_1. Since $(x_0, x_1) \neq (0,0)$, we have $(c_1, c_2) \neq (0,0)$. The following cases are analyzed.

ii.1.1) If $c_2 \neq 0$ then

$$\lim_{t\to+\infty} \left| \frac{x_t}{tr^t} - c_2(\text{sign}(\lambda_1))^t \right| = 0.$$

ii.1.2) If $c_2 = 0$ then $c_1 \neq 0$. Therefore

$$\lim_{t\to+\infty} \left| \frac{x_t}{r^t} - c_1(\text{sign}(\lambda_1))^t \right| = 0.$$

Consequently, $\mathcal{C} = \mathbb{R}^2 \setminus \{(0,0)\}$.

ii.2) If $\lambda_1 \neq \lambda_2$ then

$$x_t = c_1 r^t + c_2(-r)^t \quad \text{for any } t \in \mathbb{N}_0,$$

where c_1 and c_2 are real constants determined by initial conditions x_0, x_1. Therefore

$$\lim_{t\to+\infty} \left| \frac{x_t}{r^t} - (c_1 + c_2(-1)^t) \right| = 0.$$

Consequently,

$$\begin{aligned} \mathcal{C} =& \{(x_0, x_1) \in \mathbb{R}^2 : c_1 + c_2 \neq 0 \text{ and } c_1 - c_2 \neq 0\} \\ =& \{(x_0, x_1) \in \mathbb{R}^2 : x_0 \neq 0 \text{ and } x_1 \neq 0\}. \end{aligned}$$

iii) **Complex conjugate roots.** Since the coefficients of the characteristic polynomial are real, if λ is a root of the polynomial, then conjugate $\overline{\lambda}$ is also a root. We can assume that $\lambda_1 = re^{i2\pi\theta}$ and $\lambda_2 = re^{-i2\pi\theta}$ with $r \in (0,1)$ and $\theta \in (0,1) \setminus \{1/2\}$. In this setting

$$x_t = c_1 r^t \cos(2\pi\theta t) + c_2 r^t \sin(2\pi\theta t) \quad \text{for any } t \in \mathbb{N}_0,$$

where c_1 and c_2 are real constants determined by initial conditions x_0, x_1. Thus

$$\lim_{t\to+\infty} \left| \frac{x_t}{r^t} - (c_1 \cos(2\pi\theta t) + c_2 \sin(2\pi\theta t)) \right| = 0.$$

Since $(x_0, x_1) \neq (0,0)$, we have $(c_1, c_2) \neq (0,0)$. Let $c = \sqrt{c_1^2 + c_2^2}$ and define γ satisfying $\cos(\gamma) = c_1/c$ and $\sin(\gamma) = c_2/c$. Consequently,

$$v_t := c_1 \cos(2\pi\theta t) + c_2 \sin(2\pi\theta t) = c \cos(2\pi\theta t - \gamma) \quad \text{for any } t \in \mathbb{N}_0.$$

Observe that γ depends on the initial conditions x_0 and x_1. Let us analyze under which conditions on x_0 and x_1 we have $\liminf_{t\to+\infty} |v_t| > 0$.

iii.1) If θ is a rational number then the sequence $(\cos(2\pi\theta t-\gamma), t\in\mathbb{N}_0)$ takes finite number of values. Notice that there exists $t_0\in\mathbb{N}_0$ such that $2\pi\theta t_0-\gamma=\pi/2+k\pi$ for some $k\in\mathbb{Z}$, if and only if $\cos(2\pi\theta t_0-\gamma)=0$. Therefore, $\liminf_{t\to+\infty}|v_t|>0$ if and only if

$$\mathcal{C}=\{(x_0,x_1)\in\mathbb{R}^2: 2\pi\theta t-\gamma\neq\frac{\pi}{2}+k\pi \quad \text{for any } t\in\mathbb{N}_0,\ k\in\mathbb{Z}\}.$$

iii.2) If θ is an irrational number, then by Corollary 4.2.3 in [40] the set $\{(\theta t-\gamma/2\pi)\in\mathbb{R}/\mathbb{Z}: t\in\mathbb{N}_0\}$ is dense in the circle $\mathbb{R}/\mathbb{Z}$. Consequently, the set

$$\{\cos(2\pi\theta t-\gamma): t\in\mathbb{N}_0\} \quad \text{is dense in } [-1,1].$$

Therefore, for any γ we have $\liminf_{t\to+\infty}|v_t|=0$, which implies $\mathcal{C}=\emptyset$.

2.4 Summary

To our knowledge, there is no general theory about the existence of the cut-off phenomenon in the discrete system with a small noise perturbation. Even for the finite Markov chains, Diaconis presented that drunkard's walk does not show the cut-off phenomenon. The meticulous analysis is necessary for the random linear recurrence (2.2), which depends on its p-string past.

Like the famous shuffle model, when the intensity of the small noise perturbation ϵ is fixed, under the assumptions in Theorem 2, we provide the large enough iteration number after which X_t^ϵ is close enough to X_∞^ϵ, which is $t^\epsilon=\mathcal{O}\left(\ln\left(\epsilon^{-1}\right)\right)$. Meanwhile, this cut-off time can be obtained for the discretization of the pth order system with a small noise perturbation. We only need a sufficient condition, which is

$$\liminf_{t\to+\infty}|v_t(x)|>0, \text{ where } \lim_{t\to+\infty}\left|\frac{x_t}{t^{l-1}r^t}-v_t\right|=0$$

and r is the spectral radius of its characteristic equation. Thus, in the pth order system with a small noise perturbation, we can find an appropriate truncation time to improve iterative efficiency when we simulate its stationary distribution.

Chapter 3

The evolution of the Langevin equation through multiple time scales

The Langevin equation reveals the motion of a Brownian particle (position x) in a fluid with much smaller particles. The following shows the equation under Newton's second law,

$$m\ddot{x}_t = -\nabla V(x_t) - \gamma\dot{x}_t + \sqrt{2D}\dot{W}_t, \quad \text{in } \mathbb{R}^d \text{ for } t \geq 0 \tag{3.1}$$

where m is the mass of the Brownian particle, $V(x)$ is the potential energy function, $\gamma\dot{x}$ is the frictional force and $\sqrt{2D}\dot{W}_t$ is the thermal noise. The Brownian motion W_t introduced by Einstein in 1905 describes stochastic fluctuations. Moreover, because of the fluctuation-dissipation theorem [41, 42], the coefficients in (3.1) satisfy

$$D = \gamma\kappa_B T_{abs},$$

where κ_B is Boltzmann's constant and T_{abs} is the absolute temperature.

By denoting $\beta^{-1} = \kappa_B T_{abs}$ and setting $m = 1$, (3.1) can be written as

$$\ddot{x}_t = -\nabla V(x_t) - \gamma\dot{x}_t + \sqrt{2\gamma\beta^{-1}}\dot{W}_t, \tag{3.2}$$

where $\gamma > 0$ and $\beta > 0$. Introducing the momentum $y_t = \dot{x}_t$, we can rewrite the Langevin equation (3.2) in the phase space representation,

$$\begin{cases} \mathrm{d}x_t = y_t \mathrm{d}t, \\ \mathrm{d}y_t = (-\nabla V(x_t) - \gamma y_t)\,\mathrm{d}t + \sqrt{2\gamma\beta^{-1}}\mathrm{d}W_t. \end{cases} \tag{3.3}$$

Also, let $(x_0, y_0)^T \in \mathbb{R}^{2d}$ be the initial condition. In order to represent the long time behavior of system (3.3), we focus on its *invariant measure* whose definition is as follows:

Definition 14 *Let X_t be a Markov process with Markov semigroup P_t. A probability measure μ is invariant under X_t if and only if*

$$P_t^* \mu = \mu,$$

where P_t^* is the corresponding adjoint semigroup.

Therefore, when the initial condition X_0 is distributed according to μ, the distribution measure of X_t is μ for each t.

When $V(x)$ is a smooth *confining function* [13],

$$\lim_{|x| \to +\infty} V(x) = +\infty \text{ and } e^{-\beta V(x)} \in L^1\left(\mathbb{R}^d\right) \text{ for } \beta \in \mathbb{R}^+,$$

(3.3) has a unique invariant measure whose distribution is called Gibbs distribution. The density of the measure is

$$\rho_\beta(x, y) = Z^{-1} e^{-\beta H(x,y)}, \tag{3.4}$$

where the corresponding *Hamiltonian* is

$$H(x, y) = \frac{1}{2} y^2 + V(x)$$

and the normalization factor is

$$Z = \int_{\mathbb{R}^{2d}} e^{-\beta H(x,y)} \mathrm{d}x\mathrm{d}y,$$

(see Proposition 6.1 in [13] for more details). For a fixed damping coefficient γ, the solution of (3.3) converges to its invariant measure in distribution. We call this type of convergence as the *time limiting behavior*.

The perturbation theory, as another perspective, can be applied to analyze the limiting behavior of the Langevin equation in terms of the parameter γ on a finite time interval $[0, T]$. This convergence type is called the *perturbation limiting behavior*. In general, two limit regimes are established by γ, which are the overdamped case ($\gamma \gg 1$) and the underdamped case ($\gamma \ll 1$).

In the overdamped case, the damping coefficient tends to infinity. In order to characterize the perturbation limiting behavior, we choose the time scale $\tau = \gamma t$. Let $x^\gamma(t) = x(\gamma t)$. Then the Langevin equation (3.2) can be written as

$$\gamma^{-2}\ddot{x}_t^\gamma = -\nabla V(x_t^\gamma) - \dot{x}_t^\gamma + \sqrt{2\beta^{-1}}\dot{W}_t.$$

Roughly speaking, when $\gamma \to \infty$, the second order term vanishes, resulting in the following Smoluchowski stochastic differential

$$\dot{x}_t^S = -\nabla V(x_t^S) + \sqrt{2\beta^{-1}}\dot{W}_t.$$

In fact, one can prove rigorously that when $\mathbb{E}|\dot{x}_t^\gamma|^2 \leq C$

$$\left(\mathbb{E}\left|x_t^\gamma - x_t^S\right|^2\right)^{\frac{1}{2}} \leq C\gamma^{-1},$$

where x_τ^γ and x_τ^S have the same initial conditions (see Section 6.5.1 in [13] for more details). In literature x_τ^S is called *the Smoluchowski limit* of x_τ^γ.

In the underdamped case, the damping coefficient tends to zero. Set $\epsilon = \gamma$ to emphasize that the parameter is small and to distinguish it from the overdamped case. Then (3.3) can be written as

$$\begin{cases} \mathrm{d}x_t^\epsilon = y_t^\epsilon \mathrm{d}t, \\ \mathrm{d}y_t^\epsilon = \left(-\nabla V(x_t^\epsilon) - \epsilon y_t^\epsilon\right)\mathrm{d}t + \sqrt{2\epsilon\beta^{-1}}\mathrm{d}W_t. \end{cases} \tag{3.5}$$

The corresponding Hamiltonian system is

$$\begin{cases} \mathrm{d}x_t^0 = y_t^0 \mathrm{d}t, \\ \mathrm{d}y_t^0 = -\nabla V(x_t^0)\mathrm{d}t. \end{cases}$$

By [5, Theorem 1.3], for any $T > 0$ and $\delta > 0$

$$\lim_{\epsilon \to 0} \mathbb{P}\left\{ \max_{t\in[0,T]} \left| \begin{pmatrix} x_t^\epsilon \\ y_t^\epsilon \end{pmatrix} - \begin{pmatrix} x_t^0 \\ y_t^0 \end{pmatrix} \right| > \delta \right\} = 0. \tag{3.6}$$

The solution of the corresponding Hamiltonian system is called *the Freidlin-Wentzell limit* of the solution of (3.5).The perturbation limiting behavior describes that in a fixed time interval $[0, T]$, the perturbed dynamical system tends to the unperturbed

one when the perturbation tends to zero. The Smoluchowski limit is an Ornstein-Uhlenbeck process, whose unique invariant measure has the density

$$\rho_\beta(x) = Z^{-1} e^{-\beta V(x)}, \text{where } Z = \int_{\mathbb{R}^d} e^{-\beta V(x)} \mathrm{d}x. \tag{3.7}$$

On one hand, we first let $\gamma \to \infty$ in (3.3) to obtain its Smoluchowski limit, and then set $t \to \infty$ to find the invariant measure (3.7). On the other hand, when we first let $t \to \infty$, the solution of (3.3) will tend to the invariant measure (3.4). Although (3.4) does not depend on the perturbation parameter γ, we can set $\gamma \to \infty$. Then the position variable x_t will still tend to the invariant measure (3.7), which is the marginal distribution of (3.4) on x. Thus, the two limits in terms of γ and t commute.

However, the above commutative property does not take place in the underdamped case. On one hand, we first let $\epsilon \to 0$ to obtain its Freidlin-Wentzell limit, which only has periodic solutions. On the other hand, similarly, the time limit (3.4) is reached first. Since (3.4) is independent of ϵ, when set $\epsilon \to 0$, the solution of (3.3) will tend to the invariant measure. Thus the time limiting process and the perturbation limiting process cannot commutate with each other in the underdamped case. Thus we need to establish a proper connection between those two limiting process. The following question arises.

- Are there appropriate time scales to describe the evolution from the perturbation limiting behavior to the time limiting behavior?

We can describe the underdamped Langevin dynamics more precisely if we have a positive answer to the above question. When the solution is close to the Hamiltonian system, the perturbation limiting behavior is only analyzed in finite time. In comparison, the time limiting behavior indicates that the solution will tend to its invariant measure after a long time. We want to find a critical time scale t_ϵ^c in between. Any time duration shorter than t_ϵ^c is called the short time scale t_ϵ^s. Up to t_ϵ^s the solution is restricted around a Hamiltonian. We denote t_ϵ^l as a long time scale, which is longer than t_ϵ^c. After t_ϵ^l the solution will eventually tend to the invariant measure (3.4). Thus we expect the following results.

Expected Results

i) There exists a short time scale t_ϵ^s; such that when $t_\epsilon \ll t_\epsilon^s$,

$$\lim_{\epsilon\to 0} \mathrm{d}_1\left(\left(x_{t_\epsilon}^\epsilon, y_{t_\epsilon}^\epsilon\right)^T, \left(x_{t_\epsilon}^0, y_{t_\epsilon}^0\right)^T\right) = 0.$$

ii) There exists a long time scale t_ϵ^l; such that when $t_\epsilon \gg t_\epsilon^l$,

$$\lim_{\epsilon\to 0} \mathrm{d}_2\left(\left(x_{t_\epsilon}^\epsilon, y_{t_\epsilon}^\epsilon\right)^T, \left(x_\infty, y_\infty\right)^T\right) = 0,$$

where $(x_\infty, y_\infty)^T$ is a random variable with the density function ρ_β.

Notice that we need to find suitable distances, d_1 and d_2, in order to describe the limiting behaviors. For instance, in the short time scale, $\left(x_{t_\epsilon}^0, y_{t_\epsilon}^0\right)^T$ is a point on the deterministic Hamiltonian, and consequently the total variation distance between the point and any distribution supported on $\mathbb{R}^{2d}$ is equal to 1. Thus, the total variation distance should not be chosen as d_1.

3.1 Analysis of the one-dimensional harmonic oscillator with external noise

We start with a case, when $d = 1$ and $V(x) = \frac{1}{2}kx^2$ ($V'(x) = kx$ is a linear function) where $k > 0$. This potential V is called the Harmonic Potential. Without loss of generality, we set $\beta = 2$.

3.1.1 The explicit solution

(3.5) can be written as

$$\mathrm{d}\begin{pmatrix} x_t^\epsilon \\ y_t^\epsilon \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k & -\epsilon \end{pmatrix}\begin{pmatrix} x_t^\epsilon \\ y_t^\epsilon \end{pmatrix}\mathrm{d}t + \begin{pmatrix} 0 \\ \sqrt{\epsilon} \end{pmatrix}\mathrm{d}W_t, \tag{3.8}$$

where W_t is a one-dimensional Brownian motion. Note that this is a linear system. The corresponding Hamiltonian system is

$$\mathrm{d}\begin{pmatrix} x_t^0 \\ y_t^0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix}\begin{pmatrix} x_t^0 \\ y_t^0 \end{pmatrix}\mathrm{d}t.$$

Set $x_0^\epsilon = x_0^0 = x_0$ and $y_0^\epsilon = y_0^0 = y_0$, such that two systems have the same initial conditions. Let

$$A \coloneqq \begin{pmatrix} 0 & 1 \\ -k & -\epsilon \end{pmatrix}, \; B \coloneqq \begin{pmatrix} 0 \\ \sqrt{\epsilon} \end{pmatrix}, \; X_t^\epsilon \coloneqq \begin{pmatrix} x_t^\epsilon \\ y_t^\epsilon \end{pmatrix}, \text{ and } X_t^0 \coloneqq \begin{pmatrix} x_t^0 \\ y_t^0 \end{pmatrix}.$$

By [43, (6.6)], we have

$$X_t^\epsilon = e^{At}\left[X_0^\epsilon + \int_0^t e^{-As} B \, \mathrm{d}W_s\right]. \tag{3.9}$$

Let m_t^ϵ and Σ_t^ϵ be the mean and covariance matrix of X_t^ϵ respectively. Since X_t^ϵ is the solution of a linear stochastic differential system, it is a Gaussian process and $X_t^\epsilon \overset{\mathcal{D}}{=} \mathcal{N}(m_t^\epsilon, \Sigma_t^\epsilon)$.

Since $\epsilon^2 \ll 4k$, the eigenvalues of A are

$$\lambda_1 = -\frac{\epsilon}{2} + \frac{\sqrt{4k-\epsilon^2}}{2} i \text{ and } \lambda_2 = -\frac{\epsilon}{2} - \frac{\sqrt{4k-\epsilon^2}}{2} i,$$

which are complex conjugates. Set $b \coloneqq \frac{1}{2}\sqrt{4k-\epsilon^2}$. We compute mean and covariance matrix of X_t^ϵ as follows:

$$\begin{aligned} m_t^\epsilon &= e^{At} X_0^\epsilon \\ &= \left[e^{-\frac{\epsilon}{2}t}\cos(bt)\,\mathbf{I} + \frac{1}{b} e^{-\frac{\epsilon}{2}t}\sin(bt)\left(A + \frac{\epsilon}{2}\mathbf{I}\right)\right]\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\ &= e^{-\frac{\epsilon}{2}t}\begin{pmatrix} x_0\cos(bt) + \frac{1}{b}\left(\frac{\epsilon}{2}x_0 + y_0\right)\sin(bt) \\ y_0\cos(bt) - \frac{1}{b}\left(kx_0 + \frac{\epsilon}{2}y_0\right)\sin(bt) \end{pmatrix}, \end{aligned} \tag{3.10}$$

where $\mathbf{I}$ denotes the 2×2 identity matrix. By [43, (6.14)],

$$\begin{aligned} \Sigma_t^\epsilon &= e^{At}\left[\int_0^t e^{-As} B \left(e^{-As}B\right)^T \mathrm{d}s\right] e^{A^T t} \\ &= \epsilon \int_0^t e^{A(t-s)} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} e^{A^T(t-s)} \, \mathrm{d}s \\ &= \epsilon \int_0^t e^{As} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} e^{A^T s} \, \mathrm{d}s =: \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{pmatrix}, \end{aligned} \tag{3.11}$$

where

$$\begin{aligned}
\sigma_{1,1} &= \epsilon \int_0^t e^{-\epsilon s} \frac{1}{b^2} \sin^2(bs)\, \mathrm{d}s \\
&= \frac{e^{-\epsilon t}}{4b^2+\epsilon^2} \left[2\left(e^{\epsilon t}-1\right) - \epsilon \frac{\sin(2bt)}{b} - \epsilon^2 \frac{1-\cos(2bt)}{2b^2} \right]; \\
\sigma_{1,2} = \sigma_{2,1} &= \epsilon \int_0^t e^{-\epsilon s} \left[\frac{1}{b} \sin(bs)\cos(bs) - \frac{\epsilon}{2b^2} \sin^2(bs) \right] \mathrm{d}s \\
&= \frac{\epsilon e^{-\epsilon t}}{4b^2+\epsilon^2} \left[1-\cos(2bt) + \epsilon^2 \frac{1-\cos(2bt)}{4b^2} \right]; \\
\sigma_{2,2} &= \epsilon \int_0^t e^{-\epsilon s} \left[\cos^2(bs) - \frac{\epsilon}{b} \sin(bs)\cos(bs) + \frac{\epsilon^2}{4b^2} \sin^2(bs) \right] \mathrm{d}s \\
&= \frac{e^{-\epsilon t}}{4b^2+\epsilon^2} \Bigg[2b^2\left(e^{\epsilon t}-1\right) + \epsilon b \sin(2bt) + \epsilon^2 \frac{\cos(2bt)+e^{\epsilon t}-2}{2} \\
&\qquad\qquad + \epsilon^3 \frac{\sin(2bt)}{4b} + \epsilon^4 \frac{1-\cos(2bt)}{8b^2} \Bigg].
\end{aligned}$$

Since X_0^ϵ is deterministic ($\Sigma_0^\epsilon = \mathbf{0}$), Σ_t^ϵ is independent of $(x_0, y_0)^T$.

On one hand, for any fixed ϵ, when $t \to \infty$ we have

$$m_t^\epsilon \to \mathbf{0} \text{ and } \Sigma_t^\epsilon \to \begin{pmatrix} \frac{1}{2k} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} =: \Sigma.$$

Let X_∞^ϵ be the random variable whose distribution is $\mathcal{N}(\mathbf{0}, \Sigma)$. Then X_∞^ϵ is the limit of X_t^ϵ in distribution. We known that the density of the unique invariant measure of (3.8) is

$$\rho_2(x,y) = Z^{-1} e^{-2(\frac{1}{2}y^2 + \frac{1}{2}kx^2)} = Z^{-1} e^{-y^2 - kx^2}, \tag{3.12}$$

which is exactly $\mathcal{N}(\mathbf{0}, \Sigma)$. Since X_∞^ϵ does not depend on ϵ, we denote X_∞ as X_∞^ϵ for simplicity.

On the other hand, for any fixed t, we let $\epsilon \to 0$ to find

$$m_t^\epsilon \to \begin{pmatrix} x_0 \cos\left(\sqrt{k}t\right) + \frac{1}{\sqrt{k}} y_0 \sin\left(\sqrt{k}t\right) \\ y_0 \cos\left(\sqrt{k}t\right) - \sqrt{k} x_0 \sin\left(\sqrt{k}t\right) \end{pmatrix} = X_t^0 \text{ and } \Sigma_t^\epsilon \to \mathbf{0}. \tag{3.13}$$

Thus the mean of X_t^ϵ converges to the solution of the corresponding Hamiltonian solution for any t. Meanwhile, there is no variation of the limit of X_t^ϵ. Roughly speaking, X_t^ϵ tends to the Hamiltonian solution X_t^0 as ϵ tends to zero, which is the perturbation limiting behavior.

3.1.2 Multi-scale behaviors

We first analyze the mean m_t^ϵ. Notice that $e^{-\frac{\epsilon}{2}t}$ controls the value of m_t^ϵ in (3.10). Thus we use its exponent to classify different time scales. There three time scales as follows:

1) Let $t_\epsilon \ll \frac{1}{\epsilon}$. When $\epsilon \to 0$, we have

$$\epsilon t_\epsilon \to 0,\ b = \frac{1}{2}\sqrt{4k-\epsilon^2} \to \sqrt{k} \text{ and } bt_\epsilon - \sqrt{k}t_\epsilon \to 0,$$

and hence,

$$\begin{aligned} m_{t_\epsilon}^\epsilon &= e^{-\frac{\epsilon}{2}t_\epsilon}\begin{pmatrix} x_0\cos(bt_\epsilon) + \frac{1}{b}\left(\frac{\epsilon}{2}x_0 + y_0\right)\sin(bt_\epsilon) \\ y_0\cos(bt_\epsilon) - \frac{1}{b}\left(kx_0 + \frac{\epsilon}{2}y_0\right)\sin(bt_\epsilon)\end{pmatrix} \\ &\to \begin{pmatrix} x_0\cos(\sqrt{k}t_\epsilon) + \frac{1}{\sqrt{k}}y_0\sin(\sqrt{k}t_\epsilon) \\ y_0\cos(\sqrt{k}t_\epsilon) - \sqrt{k}x_0\sin(\sqrt{k}t_\epsilon)\end{pmatrix} = X_{t_\epsilon}^0, \end{aligned}$$

which is equal to the solution of the Hamiltonian system at time t_ϵ.

2) Let $t_\epsilon = \frac{\theta}{\epsilon}$. When $\epsilon \to 0$,

$$\epsilon t_\epsilon \to \theta \text{ and } bt_\epsilon - \sqrt{kt_\epsilon{}^2 - \frac{\theta^2}{4}} \to 0,$$

resulting in

$$m_{t_\epsilon}^\epsilon \to e^{-\frac{\theta}{2}}\begin{pmatrix} x_0\cos\left(\sqrt{kt_\epsilon{}^2 - \frac{\theta^2}{4}}\right) + \frac{1}{\sqrt{k}}y_0\sin\left(\sqrt{kt_\epsilon{}^2 - \frac{\theta^2}{4}}\right) \\ y_0\cos\left(\sqrt{kt_\epsilon{}^2 - \frac{\theta^2}{4}}\right) - \sqrt{k}x_0\sin\left(\sqrt{kt_\epsilon{}^2 - \frac{\theta^2}{4}}\right)\end{pmatrix}.$$

Although the limit is not exactly equal to $e^{-\frac{\theta}{2}}X_{t_\epsilon}^0$, it is on the same Hamiltonian with the initial condition $e^{-\frac{\theta}{2}}X_0^0$. Thus, we define a rotation map on the Hamiltonian orbit $\mathrm{Rot}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ by setting

$$\mathrm{Rot}_\theta\left(X^0(t)\right) = X^0\left(\sqrt{t^2 - \frac{\theta^2}{4k}}\right), \tag{3.14}$$

when $t \geq \frac{\theta}{2\sqrt{k}}$.

3) Let $t_\epsilon \gg \frac{1}{\epsilon}$. Then $\epsilon t_\epsilon \to \infty$. Since the exponential multiplier tends to zero and the vector multiplier is bounded, we have $m_{t_\epsilon}^\epsilon \to \mathbf{0}$. Then the mean of $X_{t_\epsilon}^\epsilon$ has already tended to the mean of its invariant measure.

Next, we use the same time scales to analyze the covariance matrix Σ_t^ϵ.

1) Let $t_\epsilon \ll \frac{1}{\epsilon}$. Then we have

$$\begin{aligned}\sigma_{1,1}(t_\epsilon) &= \frac{e^{-\epsilon t_\epsilon}}{4b^2+\epsilon^2}\left[2\left(e^{\epsilon t_\epsilon}-1\right)-\epsilon\frac{\sin(2bt_\epsilon)}{b}-\epsilon^2\frac{1-\cos(2bt_\epsilon)}{2b^2}\right]\\ &\to \frac{2e^{-\epsilon t_\epsilon}\left(e^{\epsilon t_\epsilon}-1\right)}{4k}\to 0,\\ \sigma_{1,2}(t_\epsilon) &= \sigma_{2,1}(t_\epsilon)\to 0 \text{ and } \sigma_{2,2}(t_\epsilon)\to 0,\end{aligned}$$

as ϵ tends to zero. Since $\Sigma_{t_\epsilon}^\epsilon \to \mathbf{0}$, $X_{t_\epsilon}^\epsilon$ only depends on its mean value.

2) Let $t_\epsilon = \frac{\theta}{\epsilon}$. When $\epsilon \to 0$,

$$\begin{aligned}\sigma_{1,1}(t_\epsilon) &= \frac{e^{-\epsilon t_\epsilon}}{4b^2+\epsilon^2}\left[2\left(e^{\epsilon t_\epsilon}-1\right)-\epsilon\frac{\sin(2bt_\epsilon)}{b}-\epsilon^2\frac{1-\cos(2bt_\epsilon)}{2b^2}\right]\\ &\to \frac{2e^{-\epsilon t_\epsilon}\left(e^{\epsilon t_\epsilon}-1\right)}{4k}\to \frac{1}{2k}\left(1-e^{-\theta}\right),\\ \sigma_{1,2}(t_\epsilon) &= \sigma_{2,1}(t_\epsilon)\to 0 \text{ and } \sigma_{2,2}(t_\epsilon)\to \frac{1}{2}\left(1-e^{-\theta}\right),\end{aligned}$$

Therefore $\Sigma_{t_\epsilon}^\epsilon \to \Sigma\cdot\left(1-e^{-\theta}\right)$, where Σ is the covariance matrix of X_∞.

3) Let $t_\epsilon \gg \frac{1}{\epsilon}$. Then $\epsilon t_\epsilon \to \infty$. Letting $\epsilon \to 0$ we obtain

$$\begin{aligned}\sigma_{1,1}(t_\epsilon) &= \frac{e^{-\epsilon t_\epsilon}}{4b^2+\epsilon^2}\left[2\left(e^{\epsilon t_\epsilon}-1\right)-\epsilon\frac{\sin(2bt_\epsilon)}{b}-\epsilon^2\frac{1-\cos(2bt_\epsilon)}{2b^2}\right]\\ &\to \frac{2e^{-\epsilon t_\epsilon}\left(e^{\epsilon t_\epsilon}-1\right)}{4k}\to \frac{1}{2k},\\ \sigma_{1,2}(t_\epsilon) &= \sigma_{2,1}(t_\epsilon)\to 0 \text{ and } \sigma_{2,2}(t_\epsilon)\to \frac{1}{2},\end{aligned}$$

i.e., $\Sigma_{t_\epsilon}^\epsilon \to \Sigma$. In this time scale the covariance matrix of $X_{t_\epsilon}^\epsilon$ has already tended to the covariance matrix of its invariant measure.

3.1.3 Core Observations

In summary, the multi-scale behaviors of the mean and covariance matrix lead to the following results:

$$\begin{cases} m_{t_\epsilon}^\epsilon - X_{t_\epsilon}^0 \xrightarrow{\epsilon\to 0} \mathbf{0}, & \Sigma_{t_\epsilon}^\epsilon \xrightarrow{\epsilon\to 0} \mathbf{0} & t_\epsilon \ll \frac{1}{\epsilon} \quad (3.15)\\ m_{t_\epsilon}^\epsilon - \mathrm{Rot}_\theta(X_{t_\epsilon}^0)\cdot e^{-\frac{\theta}{2}} \xrightarrow{\epsilon\to 0} \mathbf{0}, & \Sigma_{t_\epsilon}^\epsilon \xrightarrow{\epsilon\to 0} \Sigma\cdot\left(1-e^{-\theta}\right) & t_\epsilon = \frac{\theta}{\epsilon} \quad (3.16)\\ m_{t_\epsilon}^\epsilon \xrightarrow{\epsilon\to 0} \mathbf{0}, & \Sigma_{t_\epsilon}^\epsilon \xrightarrow{\epsilon\to 0} \Sigma & t_\epsilon \gg \frac{1}{\epsilon} \quad (3.17)\end{cases}$$

where $\mathrm{Rot}_\theta(\cdot)$ is the rotation map given by (3.14).

When $t_\epsilon \ll \frac{1}{\epsilon}$, the solution $X^\epsilon_{t_\epsilon}$ as a Gaussian process tends to the corresponding Hamiltonian solution $X^0_{t_\epsilon}$. This time scale is named the short time scale, which is denoted as t^s_ϵ. When $t_\epsilon \gg \frac{1}{\epsilon}$, $X^\epsilon_{t_\epsilon}$ has already tended to its stationary measure. Thus we call this time scale the long time scale, denoted as t^l_ϵ.

When $t_\epsilon = \frac{\theta}{\epsilon}$, $X^\epsilon_{t_\epsilon}$ is transient from the Hamilton solution to the invariant measure. Roughly speaking, the total energy has already decreased $1 - e^{-\theta}$ times. Meanwhile the solution starts to demonstrate a certain degree of variation which is $1 - e^{-\theta}$ times as large as the variation of the invariant measure.

3.1.4 The main result

In this section, we determine a proper distance to measure the difference among the solution X^ϵ_t, the corresponding Hamiltonian solution X^0_t, and the invariant measure X_∞. To do so, we need to check whether the covariance matrix of X^ϵ_t is singular.

Lemma 15 *The distribution of X^ϵ_t is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$ for $t > 0$.*

Proof. By [43, Proposition 6.4], for the non-degeneracy, we need to fulfill the controllability of the pair of functions (A, B). Since A and B are constant matrices, by [43, Proposition 6.5], we only need to verify that the controllability matrix $C = (B, AB)$ has full rank. Clearly,

$$\mathrm{rank}(C) = \mathrm{rank}\left(\begin{pmatrix} 0 & \sqrt{\epsilon} \\ \sqrt{\epsilon} & -\epsilon\sqrt{\epsilon} \end{pmatrix}\right) = 2, \text{when } \epsilon \neq 0.$$

■

Since the distributions of X^ϵ_t and X_∞ are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$, we use the total variation distance to measure their discrepancy. The following theorem describes how far away the solution X^ϵ_t is from its invariant measure X_∞ in the different time scales by using the total variation distance.

Theorem 16

i) If $\epsilon t_\epsilon \ll 1$, *then*

$$\lim_{\epsilon \to 0} \mathbf{d}_{\mathrm{TV}} \left(X^\epsilon_{t_\epsilon}, X_\infty \right) = 1.$$

ii) If $\epsilon t_\epsilon = \theta$ *where* $\theta \in \mathbb{R}^+$, *then there exists a constant* $c(x_0, y_0, \theta) \in (0,1)$ *such that*

$$\lim_{\epsilon \to 0} \mathbf{d}_{\mathrm{TV}} \left(X^\epsilon_{t_\epsilon}, X_\infty \right) = c.$$

iii) If $\epsilon t_\epsilon \gg 1$, *then*

$$\lim_{\epsilon \to 0} \mathbf{d}_{\mathrm{TV}} \left(X^\epsilon_{t_\epsilon}, X_\infty \right) = 0.$$

Proof.

i) In (3.15) we know that when $\epsilon \to 0$,

$$\Sigma^\epsilon_{t_\epsilon} \to \mathbf{0}.$$

Therefore, we have by ii) of Lemma 29,

$$\lim_{\epsilon \to 0} \mathbf{d}_{\mathrm{TV}} \left(X^\epsilon_{t_\epsilon}, X_\infty \right) = \lim_{\epsilon \to 0} \mathbf{d}_{\mathrm{TV}} \left(\mathcal{N} \left(m^\epsilon_{t_\epsilon}, \Sigma^\epsilon_{t_\epsilon} \right), \mathcal{N} \left(\mathbf{0}, \Sigma \right) \right) = 1.$$

ii) Using iii) of Lemma 31 we can first standardize the stationary measure to obtain

$$\begin{aligned} \mathbf{d}_{\mathrm{TV}} \left(X^\epsilon_{t_\epsilon}, X_\infty \right) &= \mathbf{d}_{\mathrm{TV}} \left(\mathcal{N} \left(m^\epsilon_{t_\epsilon}, \Sigma^\epsilon_{t_\epsilon} \right), \mathcal{N} \left(\mathbf{0}, \Sigma \right) \right) \\ &= \mathbf{d}_{\mathrm{TV}} \left(\mathcal{N} \left(\Sigma^{-\frac{1}{2}} m^\epsilon_{t_\epsilon}, \Sigma^{-\frac{1}{2}} \Sigma^\epsilon_{t_\epsilon} \Sigma^{-\frac{1}{2}} \right), \mathcal{N} \left(\mathbf{0}, \mathbf{I} \right) \right). \end{aligned} \tag{3.18}$$

(3.16) then yields

$$\epsilon t_\epsilon = \theta \text{ and } \Sigma^{-\frac{1}{2}} \Sigma^\epsilon_{t_\epsilon} \Sigma^{-\frac{1}{2}} \to \left(1 - e^{-\theta}\right) \mathbf{I},$$

as $\epsilon \to 0$. It remains to deal with the mean of the first component in (3.18).

We note that

$$
\begin{aligned}
&\lim_{\epsilon\to 0}\left|\left|\Sigma^{-\frac{1}{2}}m^{\epsilon}_{t_\epsilon}\right|\right|_2^2\\
=&\lim_{\epsilon\to 0}e^{-\epsilon t_\epsilon}\left|\left|\begin{pmatrix}\sqrt{2k} & 0\\ 0 & \sqrt{2}\end{pmatrix}\begin{pmatrix}x_0\cos\left(bt_\epsilon\right)+\frac{1}{b}\left(\frac{\epsilon}{2}x_0+y_0\right)\sin\left(bt_\epsilon\right)\\ y_0\cos\left(bt_\epsilon\right)-\frac{1}{b}\left(kx_0+\frac{\epsilon}{2}y_0\right)\sin\left(bt_\epsilon\right)\end{pmatrix}\right|\right|_2^2\\
=&\lim_{\epsilon\to 0}e^{-\epsilon t_\epsilon}\left[2k\left(x_0\cos\left(bt_\epsilon\right)+\frac{1}{b}\left(\frac{\epsilon}{2}x_0+y_0\right)\sin\left(bt_\epsilon\right)\right)^2\right.\\
&\qquad\left.+2\left(y_0\cos\left(bt_\epsilon\right)-\frac{1}{b}\left(kx_0+\frac{\epsilon}{2}y_0\right)\sin\left(bt_\epsilon\right)\right)^2\right]\\
=&\lim_{\epsilon\to 0}e^{-\epsilon t_\epsilon}\left[2kx_0^2\cos^2\left(bt_\epsilon\right)+\frac{2k}{b^2}y_0^2\sin^2\left(bt_\epsilon\right)+2y_0^2\cos^2\left(bt_\epsilon\right)\right.\\
&\qquad\left.+\frac{2k^2}{b^2}x_0^2\sin\left(bt_\epsilon\right)+\mathcal{O}(\epsilon)\right]\\
=&e^{-\theta}\left[2kx_0^2\left(\cos^2\left(\sqrt{k{t_\epsilon}^2-4^{-1}\theta}\right)+\sin^2\left(\sqrt{k{t_\epsilon}^2-4^{-1}\theta}\right)\right)\right.\\
&\qquad\left.+2y_0^2\left(\cos^2\left(\sqrt{k{t_\epsilon}^2-4^{-1}\theta}\right)+\sin^2\left(\sqrt{k{t_\epsilon}^2-4^{-1}\theta}\right)\right)\right]\\
=&2e^{-\theta}\left(kx_0^2+y_0^2\right).
\end{aligned}
$$

Let $K_\theta \coloneqq \sqrt{2}e^{-\frac{\theta}{2}}\sqrt{kx_0^2+y_0^2} > 0$. Applying i) of Lemma 29 and Lemma 32, we obtain the limit of the mean and the covariance respectively. Since $\theta > 0$, we have

$$
\begin{aligned}
&\lim_{\epsilon\to 0}\mathbf{d}_{\mathrm{TV}}(X^{\epsilon}_{t_\epsilon},X_\infty)\\
=&\mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\begin{pmatrix}K_\theta\\ 0\end{pmatrix},\left(1-e^{-\theta}\right)\mathbf{I}\right),\mathcal{N}\left(\mathbf{0},\mathbf{I}\right)\right)\in(0,1).
\end{aligned}
$$

iii) In (3.17) we obtain that

$$
m^{\epsilon}_{t_\epsilon}\to\mathbf{0}\text{ and }\Sigma^{\epsilon}_{t_\epsilon}\to\Sigma,
$$

as $\epsilon\to 0$. Therefore, according to i) of Lemma 29, the limiting distance

$$
\lim_{\epsilon\to 0}\mathbf{d}_{\mathrm{TV}}(X^{\epsilon}_{t_\epsilon},X_\infty)=\lim_{\epsilon\to 0}\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m^{\epsilon}_{t_\epsilon},\Sigma^{\epsilon}_{t_\epsilon}),\mathcal{N}(\mathbf{0},\Sigma))=0.
$$

■

Before we move on, let us remark that Theorem 16's part iii) gives an estimate about the long time scale t_ϵ^l where $t_\epsilon^l \gg \epsilon^{-1}$. In this time scale, the distribution of $X_{t_\epsilon}^\epsilon$ eventually tends to its invariant measure when $\epsilon \to 0$. In part ii), a portion of the mass of the distribution of $X_{t_\epsilon}^\epsilon$ has been transferred to its invariant measure. We denote this scenario as a transient state. We call t_ϵ^c the critical time scale when $\epsilon t_\epsilon^c = \theta$. Part i) simply says that when $t_\epsilon \ll \epsilon^{-1}$ the solution is far from the invariant measure.

Recall in the short time scale t_ϵ^s, we have that $X_{t_\epsilon^s}^\epsilon$ tends to $X_{t_\epsilon^s}^0$. This limit is a degenerate Gaussian distribution and the support of X_∞ is $\mathbb{R}^2$. Consequently,

$$\lim_{\epsilon \to 0} \mathbf{d}_{\mathrm{TV}}(X_{t_\epsilon^s}^\epsilon, X_\infty) = 1.$$

Because of part i) our short time scale should be no longer than $t_\epsilon \ll \epsilon^{-1}$. In order to verify whether $t_\epsilon \ll \epsilon^{-1}$ is the exact short time scale, it is necessary to compare the solution X_t^ϵ with the Hamilton solution X_t^0. However, X_t^0 is a point mass for each t and the support of X_t^ϵ is $\mathbb{R}^2$. Similarly,

$$\mathbf{d}_{\mathrm{TV}}(X_t^\epsilon, X_t^0) \equiv 1 \text{ for any } t \in \mathbb{R}^+.$$

Therefore the total variation distance does not properly describe the discrepancy between X_t^ϵ and X_t^0. We start to use the type of the law of large numbers in order to describe the convergence in the short time scale as follows:

Theorem 17 (Law of Large Numbers) *Let $\epsilon t_\epsilon \ll 1$. Then*

$$\lim_{\epsilon \to 0} \mathbb{P}\left(\left|X_{t_\epsilon}^\epsilon - X_{t_\epsilon}^0\right| > \delta\right) = 0 \quad \textit{for every } \delta > 0.$$

Proof. By Chebyshev's inequality and the triangle inequality, for every $\delta > 0$

$$\begin{aligned}\mathbb{P}\left(\left|X_{t_\epsilon}^\epsilon - X_{t_\epsilon}^0\right| > \delta\right) &\le \frac{\mathbb{E}\left|X_{t_\epsilon}^\epsilon - X_{t_\epsilon}^0\right|^2}{\delta^2} \\ &\le \frac{2}{\delta^2}\mathbb{E}\left|X_{t_\epsilon}^\epsilon - m_{t_\epsilon}^\epsilon\right|^2 + \frac{2}{\delta^2}\left(m_{t_\epsilon}^\epsilon - X_{t_\epsilon}^0\right)^2,\end{aligned}$$

where $m_{t_\epsilon}^\epsilon$ is the mean of $X_{t_\epsilon}^\epsilon$. According to (3.15), the first term

$$\mathbb{E}\left|X_{t_\epsilon}^\epsilon - m_{t_\epsilon}^\epsilon\right|^2 = \Sigma_{t_\epsilon}^\epsilon \to \mathbf{0}, \text{ as } \epsilon \to 0.$$

Similarly, when $\epsilon \to 0$ the second term

$$\left(m^{\epsilon}_{t_\epsilon} - X^0_{t_\epsilon}\right)^2 \to \mathbf{0}$$

Combining these two limits, we have the theorem proved. ■

Theorem 17 asserts that when $\epsilon t_\epsilon \ll 1$ the solutions X^ϵ_t and X^0_t can be arbitrarily close to each other in probability one. Therefore we set this time scale as the short time scale t^s_ϵ. Applying the explicit solution, we obtain a stronger result in the spirit of the central limit theorem.

Since the Harmonic potential is smooth, we can follow the expansion in [44, (1.6)] to obtain

$$X^\epsilon_t = X^{(0)}_t + \sqrt{\epsilon} X^{(1)}_t + R^\epsilon_2(t), \tag{3.19}$$

where R^ϵ_2 is the remainder which is controlled by ϵ [5, Theorem 2.2]. Expanding both sides of (3.8) in powers of $\sqrt{\epsilon}$, we compare the equal power terms of ϵ on both sides to obtain the following differential equations:

$$\begin{aligned}
&\mathrm{d}\begin{pmatrix} x^{(0)}_t \\ y^{(0)}_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix} \begin{pmatrix} x^{(0)}_t \\ y^{(0)}_t \end{pmatrix} \mathrm{d}t && \begin{pmatrix} x^{(0)}_0 \\ y^{(0)}_0 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}. \\
&\mathrm{d}\begin{pmatrix} x^{(1)}_t \\ y^{(1)}_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix} \begin{pmatrix} x^{(1)}_t \\ y^{(1)}_t \end{pmatrix} \mathrm{d}t + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mathrm{d}W_t && \begin{pmatrix} x^{(1)}_0 \\ y^{(1)}_0 \end{pmatrix} = \mathbf{0}.
\end{aligned}$$

In the above, $X^{(0)}_t$ is exactly the solution of the corresponding Hamiltonian system X^0_t. $X^{(1)}_t$ is a Markov Gaussian process, whose mean and covariance matrix are denoted by $m^{(1)}_t$ and $\Sigma^{(1)}_t$ respectively. Similar to (3.10) and (3.11) we obtain

$$m^{(1)}_t = \mathbf{0} \text{ and } \Sigma^{(1)}_t = \begin{pmatrix} \frac{t}{2k} - \frac{\sin\left(2\sqrt{k}t\right)}{4k\sqrt{k}} & \frac{\sin^2\left(\sqrt{k}t\right)}{2k} \\ \frac{\sin^2\left(\sqrt{k}t\right)}{2k} & \frac{t}{2} + \frac{\sin\left(2\sqrt{k}t\right)}{4\sqrt{k}} \end{pmatrix}.$$

We are now ready to obtain a result of central limit theorem type as follows:

Theorem 18 (Central Limit Theorem) *Let $\epsilon t_\epsilon \ll 1$. Then*

$$\lim_{\epsilon \to 0} \mathbf{d}_{\mathrm{TV}} \left(\frac{X^\epsilon_{t_\epsilon} - X^0_{t_\epsilon}}{\sqrt{\epsilon}}, X^{(1)}_{t_\epsilon} \right) = 0.$$

Proof. $X^\epsilon_{t_\epsilon}$ and $X^{(1)}_{t_\epsilon}$ are random variables with Gaussian distribution. Therefore, by ii) of Lemma (31) we have

$$\begin{aligned}
\mathbf{d}_{\mathrm{TV}}\left(\frac{X^\epsilon_{t_\epsilon}-X^0_{t_\epsilon}}{\sqrt{\epsilon}}, X^{(1)}_{t_\epsilon}\right) &= \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{m^\epsilon_{t_\epsilon}-m^0_{t_\epsilon}}{\sqrt{\epsilon}}, \frac{\Sigma^\epsilon_{t_\epsilon}}{\epsilon}\right), \mathcal{N}\left(\mathbf{0}, \Sigma^{(1)}_{t_\epsilon}\right)\right) \\
&= \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{m^\epsilon_{t_\epsilon}-m^0_{t_\epsilon}}{\sqrt{\epsilon t_\epsilon}}, \frac{\Sigma^\epsilon_{t_\epsilon}}{\epsilon t_\epsilon}\right), \mathcal{N}\left(\mathbf{0}, \frac{\Sigma^{(1)}_{t_\epsilon}}{t_\epsilon}\right)\right) \\
&\le \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\frac{m^\epsilon_{t_\epsilon}-m^0_{t_\epsilon}}{\sqrt{\epsilon t_\epsilon}}, \frac{\Sigma^\epsilon_{t_\epsilon}}{\epsilon t_\epsilon}\right), \mathcal{N}(\mathbf{0}, \Sigma)\right) \\
&\quad + \mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(\mathbf{0}, \frac{\Sigma^{(1)}_{t_\epsilon}}{t_\epsilon}\right), \mathcal{N}(\mathbf{0}, \Sigma)\right) =: C_1 + C_2.
\end{aligned}$$

Supposing $t_\epsilon \to \infty$ as $\epsilon \to 0$, we have ${t_\epsilon}^{-1}\Sigma^{(1)}_{t_\epsilon} \to \Sigma$. Applying i) of Lemma 29, we have $C_2 \to 0$ as $\epsilon \to 0$. For C_1 we consider the mean and the variance separately. When $\epsilon t_\epsilon \ll 1$,

$$\lim_{\epsilon\to 0}\frac{\Sigma^\epsilon_{t_\epsilon}}{\epsilon t_\epsilon} = \lim_{\epsilon\to 0}\begin{pmatrix} (\epsilon t_\epsilon)^{-1}\sigma_{1,1} & (\epsilon t_\epsilon)^{-1}\sigma_{1,2} \\ (\epsilon t_\epsilon)^{-1}\sigma_{2,1} & (\epsilon t_\epsilon)^{-1}\sigma_{2,2}\end{pmatrix}, \tag{3.20}$$

where $\sigma_{i,j}$'s are defined as in (3.11). Then

$$\begin{aligned}
\lim_{\epsilon\to 0}\frac{\sigma_{1,1}}{\epsilon t_\epsilon} &= \lim_{\epsilon\to 0}\frac{e^{-\epsilon t_\epsilon}}{\epsilon t_\epsilon\left(4b^2+\epsilon^2\right)}\left[2\left(e^{\epsilon t_\epsilon}-1\right) - \epsilon\frac{\sin(2bt_\epsilon)}{b} - \epsilon^2\frac{1-\cos(2bt_\epsilon)}{2b^2}\right] \\
&= \lim_{\epsilon\to 0}\frac{1}{2k}\frac{1-e^{-\epsilon t_\epsilon}}{\epsilon t_\epsilon} + \frac{e^{-\epsilon t_\epsilon}}{t_\epsilon}B_1(\epsilon) + \frac{\epsilon e^{-\epsilon t_\epsilon}}{t_\epsilon}B_2(\epsilon) = \frac{1}{2k}, \text{ where}
\end{aligned}$$

$$\begin{aligned}
B_1(\epsilon) &= -\frac{\sin(2bt_\epsilon)}{4b^3+\epsilon^2 b}, \text{ and} \\
B_2(\epsilon) &= -\frac{1-\cos(2bt_\epsilon)}{8b^4+2\epsilon^2 b^2}
\end{aligned}$$

are bounded functions. Similarly,

$$\lim_{\epsilon\to 0}\frac{\sigma_{1,2}}{\epsilon t_\epsilon} = \lim_{\epsilon\to 0}\frac{\sigma_{2,1}}{\epsilon t_\epsilon} \to 0 \quad \text{and} \quad \lim_{\epsilon\to 0}\frac{\sigma_{2,2}}{\epsilon t_\epsilon} \to \frac{1}{2}.$$

Thus the limit in (3.20) is Σ.

We now estimate the mean. Since

$$\begin{aligned}
&m^\epsilon_{t_\epsilon} - m^0_{t_\epsilon} \\
&= \begin{pmatrix} e^{-\frac{\epsilon}{2}t_\epsilon}\left[x_0\cos(bt_\epsilon) + \frac{1}{b}\left(\frac{\epsilon}{2}x_0 + y_0\right)\sin(bt_\epsilon)\right] - \left[x_0\cos\left(\sqrt{k}t_\epsilon\right) + \frac{1}{\sqrt{k}}y_0\sin\left(\sqrt{k}t_\epsilon\right)\right] \\ e^{-\frac{\epsilon}{2}t_\epsilon}\left[y_0\cos(bt_\epsilon) - \frac{1}{b}\left(kx_0 + \frac{\epsilon}{2}y_0\right)\sin(bt_\epsilon)\right] - \left[y_0\cos\left(\sqrt{k}t_\epsilon\right) - \sqrt{k}x_0\sin\left(\sqrt{k}t_\epsilon\right)\right]\end{pmatrix}
\end{aligned}$$

has a complicated expression in the L^2 norm, we only estimate its L^1 norm for the sake of simplicity.

$$\left|\left|\frac{m^\epsilon_{t_\epsilon} - m^0_{t_\epsilon}}{\sqrt{\epsilon t_\epsilon}}\right|\right|_1 \leq I_1 + I_2 + I_3 + I_4 + \mathcal{O}(\epsilon),$$

$$\begin{aligned}
\text{where } \ I_1(\epsilon) &= \frac{1}{\sqrt{\epsilon t_\epsilon}}\left|e^{-\frac{\epsilon}{2}t_\epsilon}x_0\cos(bt_\epsilon) - x_0\cos\left(\sqrt{k}t_\epsilon\right)\right|,\\
I_2(\epsilon) &= \frac{1}{\sqrt{\epsilon t_\epsilon}}\left|\frac{1}{b}e^{-\frac{\epsilon}{2}t_\epsilon}y_0\sin(bt_\epsilon) - \frac{1}{\sqrt{k}}y_0\sin\left(\sqrt{k}t_\epsilon\right)\right|,\\
I_3(\epsilon) &= \frac{1}{\sqrt{\epsilon t_\epsilon}}\left|e^{-\frac{\epsilon}{2}t_\epsilon}y_0\cos(bt_\epsilon) - y_0\cos\left(\sqrt{k}t_\epsilon\right)\right|, \text{ and}\\
I_4(\epsilon) &= \frac{1}{\sqrt{\epsilon t_\epsilon}}\left|\frac{k}{b}e^{-\frac{\epsilon}{2}t_\epsilon}x_0\sin(bt_\epsilon) - \sqrt{k}x_0\sin\left(\sqrt{k}t_\epsilon\right)\right|.
\end{aligned}$$

The structures of these four terms are similar. For I_1 we have

$$\begin{aligned}
I_1(\epsilon) &\leq \frac{1}{\sqrt{\epsilon t_\epsilon}}\left|e^{-\frac{\epsilon}{2}t_\epsilon}x_0\cos(bt_\epsilon) - e^{-\frac{\epsilon}{2}t_\epsilon}x_0\cos\left(\sqrt{k}t_\epsilon\right)\right|\\
&\quad + \frac{1}{\sqrt{\epsilon t_\epsilon}}\left|e^{-\frac{\epsilon}{2}t_\epsilon}x_0\cos\left(\sqrt{k}t_\epsilon\right) - x_0\cos\left(\sqrt{k}t_\epsilon\right)\right|\\
&= |x_0|\left|\cos(bt_\epsilon) - \cos\left(\sqrt{k}t_\epsilon\right)\right|\frac{e^{-\frac{\epsilon}{2}t_\epsilon}}{\sqrt{\epsilon t_\epsilon}} + \left|x_0\cos\left(\sqrt{k}t_\epsilon\right)\right|\frac{1-e^{-\frac{\epsilon}{2}t_\epsilon}}{\sqrt{\epsilon t_\epsilon}}\\
&=: D_1(\epsilon) + D_2(\epsilon).
\end{aligned}$$

When $\epsilon t_\epsilon \ll 1$, we let ϵ tend to 0 to obtain

$$\begin{aligned}
D_1(\epsilon) &\leq |x_0|\left|\cos(bt_\epsilon) - \cos\left(\sqrt{k}t_\epsilon\right)\right|\frac{e^{-\frac{\epsilon}{2}t_\epsilon}}{\sqrt{\epsilon t_\epsilon}} = |x_0|\,|\sin(\theta)|\,\frac{e^{-\frac{\epsilon}{2}t_\epsilon}\left(\sqrt{k}t_\epsilon - bt_\epsilon\right)}{\sqrt{\epsilon t_\epsilon}}\\
&\leq |x_0|\,\frac{e^{-\frac{\epsilon}{2}t_\epsilon}t_\epsilon}{\sqrt{\epsilon t_\epsilon}}\left(\sqrt{k} - \frac{1}{2}\sqrt{4k-\epsilon^2}\right)\\
&= \frac{|x_0|}{8}e^{-\frac{\epsilon}{2}t_\epsilon}t_\epsilon{}^{\frac{1}{2}}\epsilon^{-\frac{1}{2}}\left(\epsilon^2 + \mathcal{O}\left(\epsilon^2\right)\right) \to 0 \ \text{ and}\\
D_2(\epsilon) &\leq |x_0|\left(\frac{\sqrt{\epsilon t_\epsilon}}{2} + \mathcal{O}\left(\sqrt{\epsilon t_\epsilon}\right)\right) \to 0.
\end{aligned}$$

So, $\lim\limits_{\epsilon\to 0} I_1(\epsilon) = 0$. Similarly, we obtain that $\lim\limits_{\epsilon\to 0} I_2(\epsilon) = \lim\limits_{\epsilon\to 0} I_3(\epsilon) = \lim\limits_{\epsilon\to 0} I_4(\epsilon) = 0$. Thus,

$$\lim_{\epsilon\to 0}\left|\left|\frac{m^\epsilon_{t_\epsilon} - m^0_{t_\epsilon}}{\sqrt{\epsilon t_\epsilon}}\right|\right|_2 \leq \lim_{\epsilon\to 0}\sqrt{2}\left|\left|\frac{m^\epsilon_{t_\epsilon} - m^0_{t_\epsilon}}{\sqrt{\epsilon t_\epsilon}}\right|\right|_\infty \leq \lim_{\epsilon\to 0}\sqrt{2}\left|\left|\frac{m^\epsilon_{t_\epsilon} - m^0_{t_\epsilon}}{\sqrt{\epsilon t_\epsilon}}\right|\right|_1 = 0.$$

The theorem now follows by applying i) of Lemma 29.

Following the same arguments, we can still prove the theorem if t_ϵ is bounded when $\epsilon \to 0$. ■

Theorem 18 asserts that under the normalization of the order $\sqrt{\epsilon}$ the difference between X_t^ϵ and X_t^0, as a random process, is close to the first order approximation $X_t^{(1)}$. The variance of $X_t^{(1)}$ plays a crucial role in demonstrating the variation of the solution.

In conclusionIn, for the one-dimensional harmonic case, there exists a short time scale $t_\epsilon^s \ll \frac{1}{\epsilon}$ over which the corresponding Hamiltonian solution captures the solution X_t^ϵ. Meanwhile, there exists a long time scale $t_\epsilon^l \gg \frac{1}{\epsilon}$ when the solution is close enough to its invariant measure. $t_\epsilon^c = \mathcal{O}\left(\frac{1}{\epsilon}\right)$ naturally becomes the critical time scale when the solution evolves from the Hamiltonian solution to the invariant measure. Because of the rotation (3.14) it is hard to construct a transition system in $\mathbb{R}^2$ from the Hamiltonian solution to the invariant measure. We will use the averaging system to approximate the above transient behavior in Section 3.4.

3.2 Simulation of the one-dimensional Langevin equation

In the last section, we explore an exquisite transient behavior of the one-dimensional Langevin equation under the harmonic potential. In this section, we will visualize the short and long time behaviors using numerical methods developed in [45]. In order to simulate a relatively long time behavior, we need to use a cost-effective algorithm. Then we can apply this algorithm to the non-harmonic potential case and intuitively determine the short and long time scales.

3.2.1 Algorithm of the one-dimensional Langevin equation

In order to simulate (3.8), there are two classical approaches: the Monte–Carlo simulation and the simulation of the Fokker-Planck equations. [14] and [15] are elaborate surveys of main techniques.

We utilize the Monte–Carlo simulation to observe the transient phenomenon. Then, using the first order stochastic Taylor series, we have the following Euler-Maruyama

scheme:

$$x_{n+1} = x_n + y_n \Delta t,$$
$$y_{n+1} = y_n - k x_n \Delta t - \epsilon y_n \Delta t + \sqrt{\epsilon} \Delta W_n,$$

where ΔW_n is a Gaussian random variable with expected value 0 and variance Δt. Given an initial condition, an application of the above scheme provides a sample path. We denote $\left(x_n^{(i)}, y_n^{(i)}\right)$ as the ith sample path where n denotes the particle at time $n\Delta t$. Then we will use the Monte-Carlo simulation to approximate the probability density function $u_n(x, y)$.

Because of the symmetric property of the Hamiltonian differential equations, we set a symmetric domain $[-L, L]^2$. We construct the regular grid with the even space steps r. Denote M as the total number of steps in each direction. Let $u_n^*(i, j)$ be the numerical approximation of the density at $(-L + ir, -L + jr)$, where $i, j = 0, \cdots, M$. In our discrete domain, we naturally consider the density of the element

$$O_{i,j} \coloneqq \left[-L + ir - \frac{r}{2}\,,\, -L + ir + \frac{r}{2}\right) \times \left[-L + jr - \frac{r}{2}\,,\, -L + jr + \frac{r}{2}\right).$$

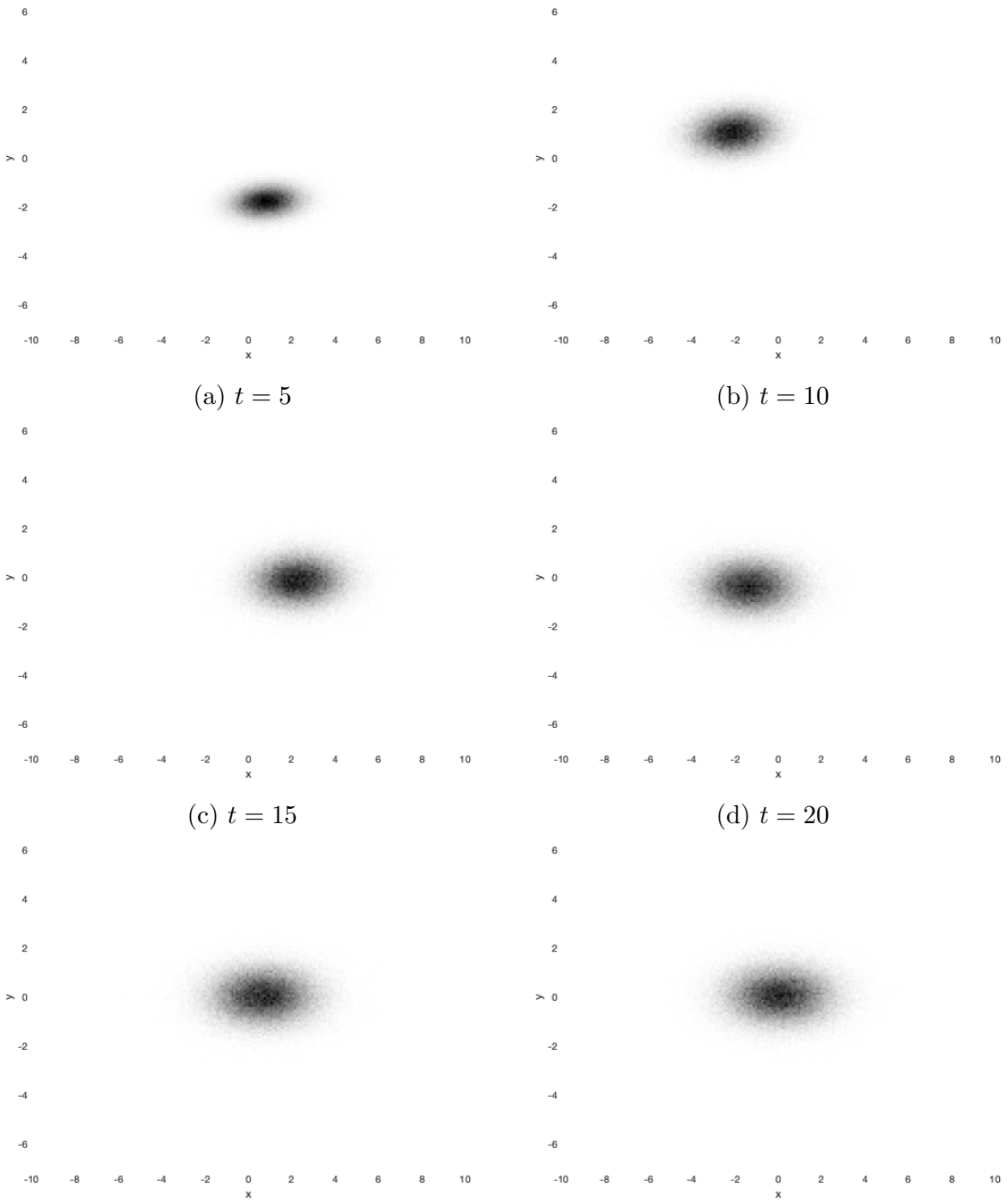
Let N be the total number of sample paths. The Monte-Carlo simulation can be implemented by

$$u_n^N(i, j) = \frac{1}{Nr^2} \sum_{k=1}^{\infty} \mathbf{1}_{O_{i,j}} \left(x_n^{(k)}, y_n^{(k)}\right).$$

When N tends to infinity, by the Law of Large Numbers we have

$$u_n^N(i, j) \to u_n \left(-L + ir, -L + jr\right).$$

We use MATLAB to implement the Monte-Carlo scheme. The code is in Appendix E.2. We fix the initial condition at $(2, 2)$ with the parameters $k = 0.25$ and $\epsilon = 0.1$. Figure 3.1 demonstrates the updates of the probability density function with respect to time.



(a) $t = 5$ (b) $t = 10$

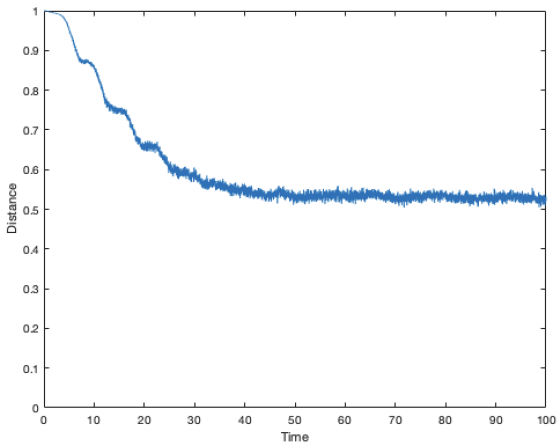
(c) $t = 15$ (d) $t = 20$

(e) $t = 40$ (f) $t = 100$

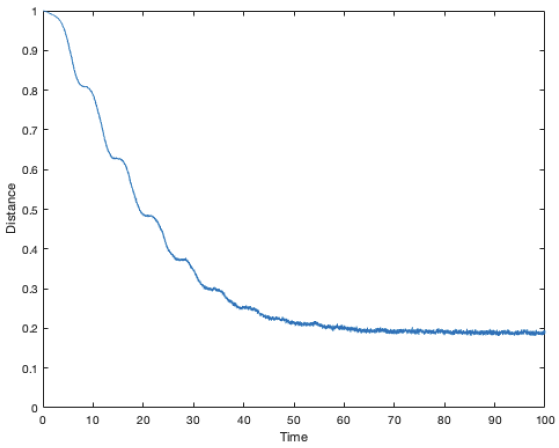
Figure 3.1: Time evolution with $k = 0.25$ and $\epsilon = 0.1$

As for the first observation, when the time passes 30, the distribution becomes stable, which gives rise to the time limiting behavior or the stationary distribution. Recall that $\mathcal{N}(\mathbf{0}, \Sigma)$ is the stationary distribution. We can compare it with our time evolution results in the total variation distance. The comparison code is provided

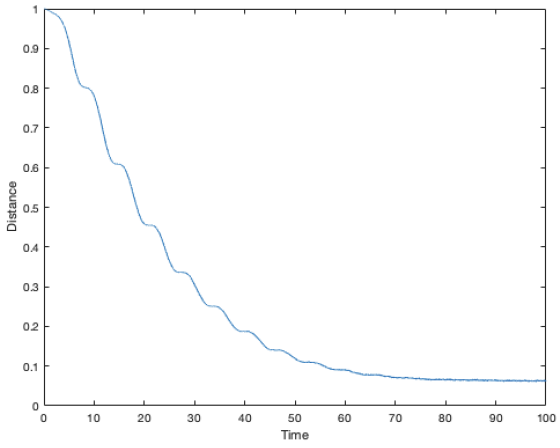
in Appendix E.3. Figure 3.2c indicates the convergent phenomenon in a non-linear shape. However, the simulated distance between the solution and the stationary distribution does not converge to zero exactly. At least there is a lower bound larger than 0.05 in that figure.



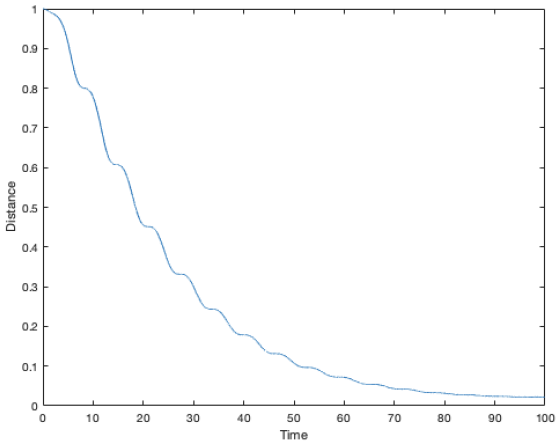
(a) Extremely Lack ($N = 10^3$)



(b) Lack ($N = 10^4$)



(c) Regular ($N = 10^5$)



(d) Abundant ($N = 10^6$)

Figure 3.2: Convergence of the difference

On one hand, the error increases tremendously when we only gather a lack of sample paths in Figure 3.2a and 3.2b. Especially when we gather a significant lack of samples, the limit of the total variation distance almost everywhere exceeds 0.5 with a visible variation. On the other hand, in Figure 3.2d, when we have abundant samples (10^6), the convergence curve seems the least noisy. The limiting value seems close to zero in the abundant case. Simultaneously, the variation of that curve is the

smallest among the four cases.

In conclusion, such an inaccuracy is mainly produced by the lack of the number of samples in the Monte-Carlo simulation. Of course, one could try to generate a large number of samples, but this will cost massive iterations. Even in Figure 3.2d, the terminal time is a mere 100; it takes more than an hour to simulate the densities based on 10^6 samples. Since our problem applies the multi-scale analysis, we need a more cost-effective algorithm, especially in the long time scale ($t_\epsilon \gg \epsilon^{-1}$). For instance, in the time scale $t_\epsilon = \epsilon^{-1.5}$, we need to observe the densities when ϵ tends to zero. If $\epsilon = 10^{-4}$, the terminal time will be 10^6.

Application of the data-driven method

We are inclined to use the data-driven method to keep the balance between the computational cost and the accuracy. Roughly speaking, the data-driven method, as an approach in between, takes the Monte–Carlo's data as a reference, then searches for the optimal solution subjecting to the Fokker-Planck equation. This novel idea was first introduced by Li [46], in which he mainly focuses on the simulation of the steady state of small noise SDE. Since for every single time we can collect Monte-Carlo data as a proper reference, we will generalize it to the simulation of the densities of X_t^ϵ at every moment. We demonstrate this technique in the harmonic potential problem as follows:

By [13] we deduce the corresponding Fokker-Planck equation of (3.8) as follows:

$$\begin{aligned} u_t &= -\frac{\partial}{\partial x}\left[yu\right] - \frac{\partial}{\partial y}\left[\left(-kx - \epsilon y\right)u\right] + \frac{1}{2}\epsilon u_{yy}, \\ &= \frac{\epsilon}{2}u_{yy} - yu_x + \left(kx + \epsilon y\right)u_y + \epsilon u, \end{aligned}$$

with an initial condition

$$u(x, y, 0) = \delta\left(x_0, y_0\right).$$

The traditional finite difference approach is to design a large enough space domain $[-L, L]^2$. Gridding the area into $M \times M$ pieces, we obtain $(M+1) \times (M+1)$ nodes in total. The space step is denoted by r as before. Discretize the time interval $[0, T]$ into N pieces, and let Δt be the time step. The discretized value is

$$u_{i,j}^n \coloneqq u(-L + ir, -L + jr, n\Delta t),$$

where $i, j = 0, \cdots, M$ and $n = 0, \cdots, N$. The second-order finite difference discretizations of the differential operator in space are

$$\begin{aligned}
\delta_y^2 u_{i,j} &= \frac{1}{r^2}\left[u_{i,j+1} - 2u_{i,j} + u_{i,j-1}\right], \\
\delta_x u_{i,j} &= \frac{1}{2r}\left[u_{i+1,j} - u_{i-1,j}\right], \\
\delta_y u_{i,j} &= \frac{1}{2r}\left[u_{i,j+1} - u_{i,j-1}\right].
\end{aligned}$$

We apply the Crank-Nicolson scheme in time. Then

$$\frac{u_{i,j}^{n+1} - u_{i,j}^{n}}{\Delta t} = \frac{1}{2}\left[\mathcal{D}u_{i,j}^{n+1} - \mathcal{D}u_{i,j}^{n}\right],$$

where the linear operator

$$\mathcal{D}u_{i,j} \coloneqq \frac{\epsilon}{2}\delta_y^2 u_{i,j} - y\delta_x u_{i,j} + (kx + \epsilon y)\delta_y u_{i,j} + \epsilon u_{i,j}.$$

Rearrange the discretized value in a vector as follows:

$$\mathbf{u} \coloneqq \left[u_{0,0}, \ldots, u_{M,0}, u_{0,1}, \ldots, u_{M,1}, \ldots, u_{0,M}, \ldots, u_{M,M}\right]^T.$$

As a variable in a linear system, $\mathbf{u}$ has $(M+1) \times (M+1)$ components. The Crank-Nicolson scheme demonstrates the recurrence relationships among the nodes in the inner part of the space domain. The total number of that relationships is $(M-1) \times (M-1)$. Since the space domain is large enough, a convectional way is to implement the free boundary condition, which introduces $4M$ relationships on the boundary nodes. In general, we can write down the recurrence relationships as

$$T_1 \mathbf{u}^{n+1} = T_2 \mathbf{u}^{n},$$

where T_1 is a nonsingular $(M+1) \times (M+1)$ matrix. Since $\mathbf{u}$ is the solution of the density, we have an extra equation which is

$$\mathbf{1}^T \mathbf{u}^{n+1} = r^{-2}. \tag{3.21}$$

We call it the unit-one condition. Because of that extra condition, the system becomes an overdetermined problem. Denote

$$T \coloneqq \begin{pmatrix} T_1 \\ \mathbf{1}^T \end{pmatrix} \text{ and } \mathbf{b} \coloneqq \begin{pmatrix} T_2 \mathbf{u}^n \\ r^{-2} \end{pmatrix}.$$

The last step of the traditional technique is an optimization problem as follows:

$$\min ||T\mathbf{u} - \mathbf{b}||_2 .$$

The traditional method is valuable, especially when we have preliminary estimates for Fokker-Planck's solution. However, if a portion of masses escapes from the space domain, it is hard to count this portion with this technique. Therefore, selecting the space domain becomes the key to make the above method working.

In order to overcome the above difficulty, the data-driven method makes use of the Monte-Carlo simulation. Even though the simulation provides rough and noisy data, it can be used as a preliminary estimate. While it does not seem easy to fix the boundaries and the boundary conditions in the traditional method, the data-driven method avoids all of them. What we have left is only the PDE mechanism given by the Crank-Nicolson scheme. Those $(M-1) \times (M-1)$ equations become constraint conditions. We denote it as

$$A\mathbf{u}^{n+1} = B\mathbf{u}^n.$$

Because of the lack of equations, the above linear system has infinite solutions. We want to find the closest one to the Monte-Carlo reference. With a similar definition, we can rewrite our Monte-Carlo result as a vector $\mathbf{v}^n$. Denote

$$\mathbf{b} \coloneqq B\mathbf{u}^n.$$

Then the optimization problem is

$$\begin{aligned} \min \quad & \left|\left|\mathbf{u}^{n+1} - \mathbf{v}^{n+1}\right|\right|_2 \\ \text{subject to} \quad & A\mathbf{u}^{n+1} = \mathbf{b}. \end{aligned}$$

For simplicity of the simulation, by Theorem 3.1 in [46], we can define the correction term $\mathbf{x}^{n+1} \coloneqq \mathbf{u}^{n+1} - \mathbf{v}^{n+1}$. The above optimization problem can be rewritten as

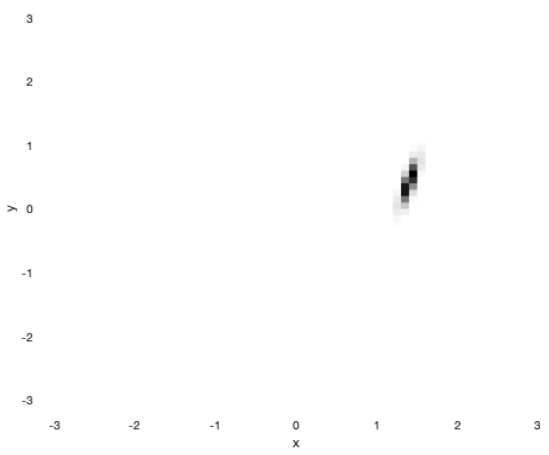
$$\begin{aligned} \min \quad & \left|\left|\mathbf{x}^{n+1}\right|\right|_2 \\ \text{subject to} \quad & A\mathbf{x}^{n+1} = \mathbf{d}, \end{aligned} \tag{3.22}$$

where $\mathbf{d} \coloneqq \mathbf{b} - A\mathbf{v}^{n+1}$.

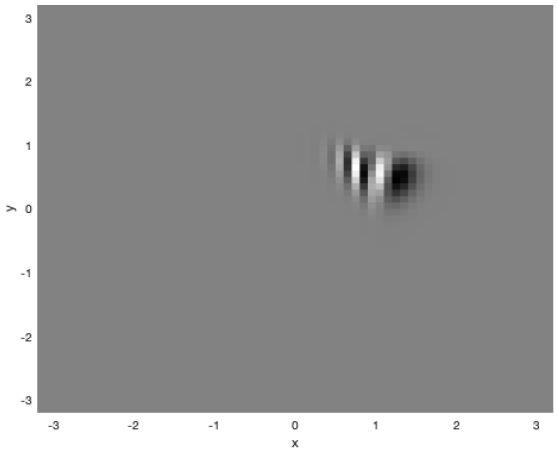
We note that even the unit-one condition (3.21) is eliminated. Thus the optimization is much more dependent on the reference data from the Monte-Carlo simulation.

Even if a portion of masses is transported out of the observing space domain, the Monte-Carlo reference will preserve this information and update the optimization's solution $\mathbf{u}^{n+1}$. This allows us to visualize the distribution locally rather than having to establish a sufficiently large domain that covers almost all samples. The code of the Crank-Nicolson scheme with the above optimization is provided in Appendix E.4.

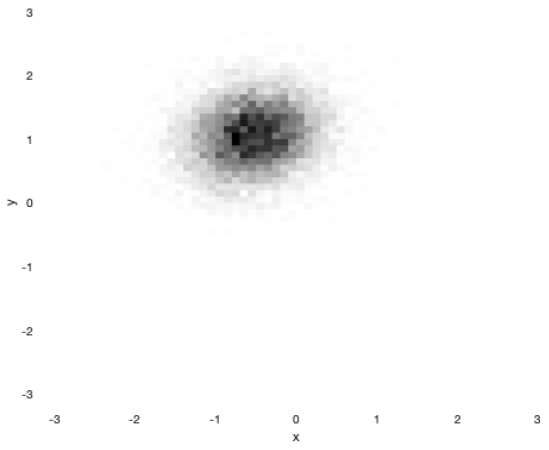
We utilize this scheme in our SDE system with a determinate initial condition. This type of initial condition becomes a delta function in the Fokker-Planck equation. In the following example we set $x_0 = y_0 = 1$, $k = 1$ and $\epsilon = 0.1$.



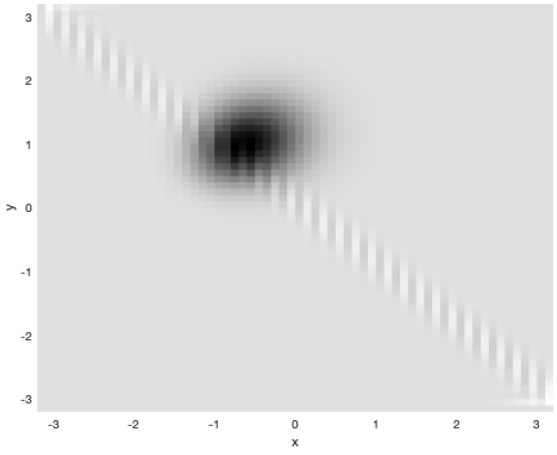
(a) $t = 0.5$



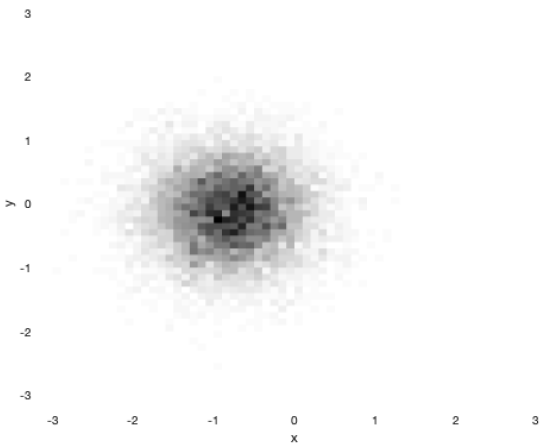
(b) $t = 0.5$



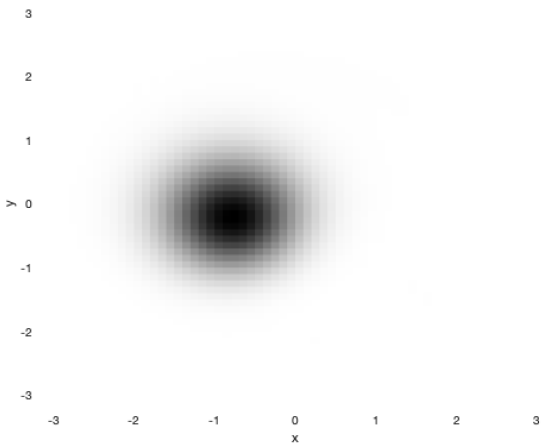
(c) $t = 5$



(d) $t = 5$



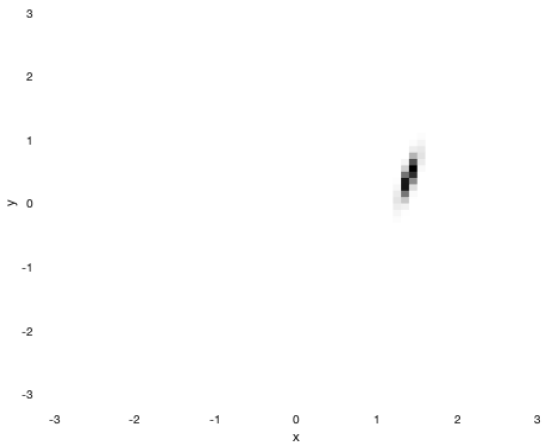
(e) $t = 10$



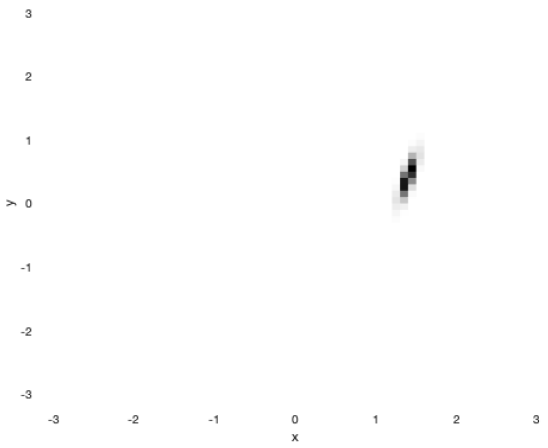
(f) $t = 10$

The figures on the left are the reference data provided by the Monte-Carlo simulation. We list the corresponding data-driven results on the right. There is an enormous difference in the initial time period. At the beginning, $t = 0.5$, and the exact solution of the density should be close to a delta function as the Monte-Carlo result. However, the data-driven result is fully supported in our compact domain, and obtains negative values in a non-empty set. When $t = 5$, the main portion is concentrated on our expected Gaussian distribution. Nevertheless, there is still a visible error that some masses escape from our expectation: a Gaussian distribution. Eventually, when $t = 10$, the data-driven method selects the correct density function.

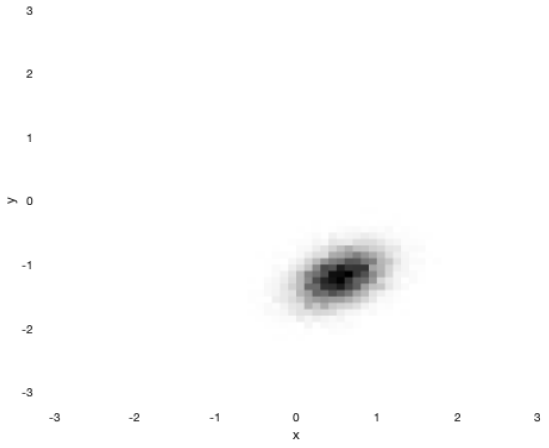
The cause of this phenomenon should be the delta function. When we input a delta function as the reference in the data-driven method, there should be more than one optimal solution satisfying the Crank-Nicolson scheme. It will lead to difficulty in optimization.We set up a threshold, *Density_Max_Threshold* in Appendix E.4. When the maximum of the reference density is larger than the threshold, this density is regarded as a delta-type of density. In this situation, we stop the optimization and directly copy the reference as our simulation. We operate the optimization only when the maximum reference density is less than the threshold. With this threshold, we have the following simulation.



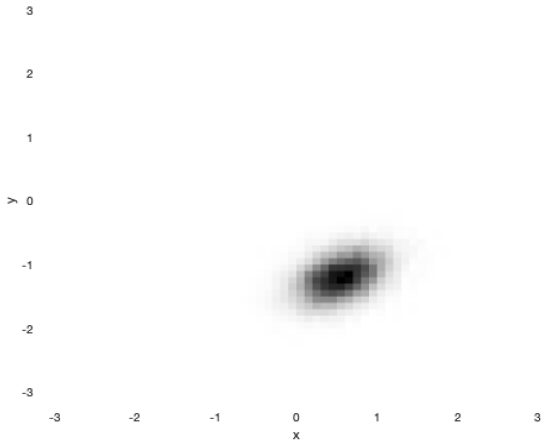
(a) $t = 0.5$



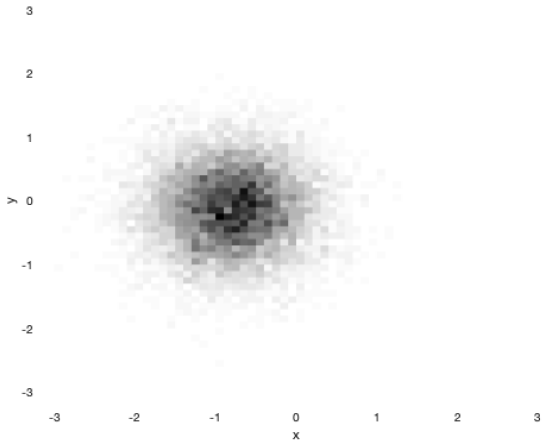
(b) $t = 0.5$



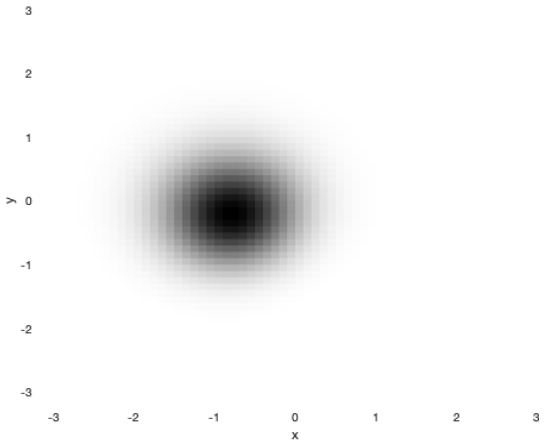
(c) $t = 2$



(d) $t = 2$



(e) $t = 10$



(f) $t = 10$

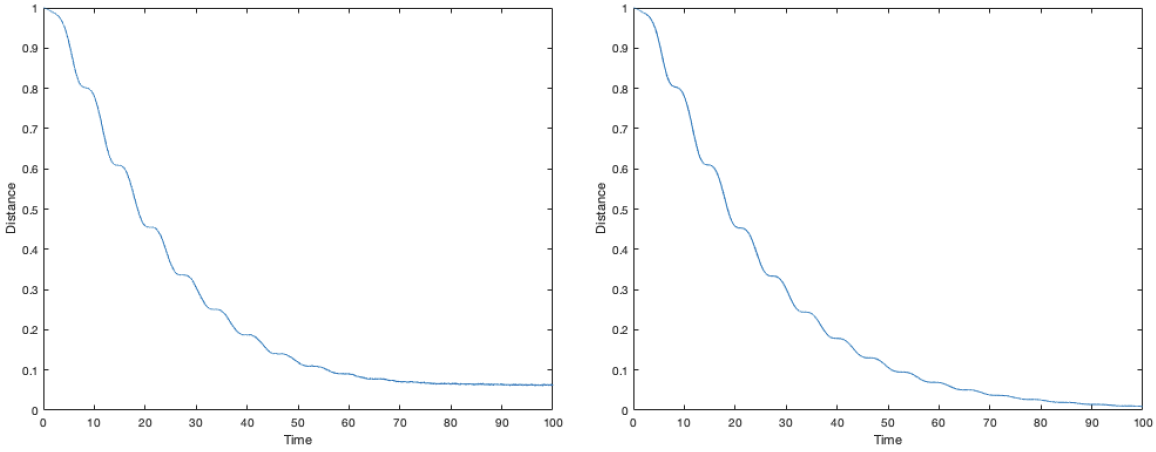
Monte-Carlo's reference is demonstrated on the left, while the results of the data-driven method with the threshold are on the right. When $t = 0.5$ we borrow Monte-

Carlo's result as our simulation. We set the threshold,

$$Density\_Max\_Threshold = 2. \tag{3.23}$$

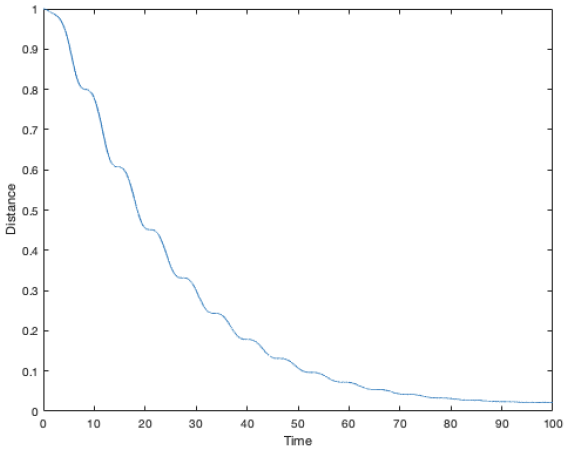
Through the simulation, the threshold succeeds when $t = 1.95$. After passing the threshold, the data-driven method operates smoothly and successfully selects the expected solution. Thus, we avoid this error from the source.

We compare the data-driven results with the stationary distribution among different samples' scales. In Figure 3.6 we select the sufficient large scales which are 10^5 and 10^6. We know there is an improvement in the Monte-Carlo method when we collect large samples from the comparison between Figure 3.6a and 3.6c. However, when we treat Monte-Carlo result as reference data and implement the optimization 3.22, the feedback of the simulations of both cases is almost the same. This phenomenon is demonstrated by Figure 3.6b and 3.6d. Thus the convergence curve is stable with respect to the samples' scales. As a fair deduction, if we generate samples more than our large scale, for instance, $N = 10^{10}$, its data-driven result should keep the same shape as the results shown in Figure 3.6b and 3.6d.

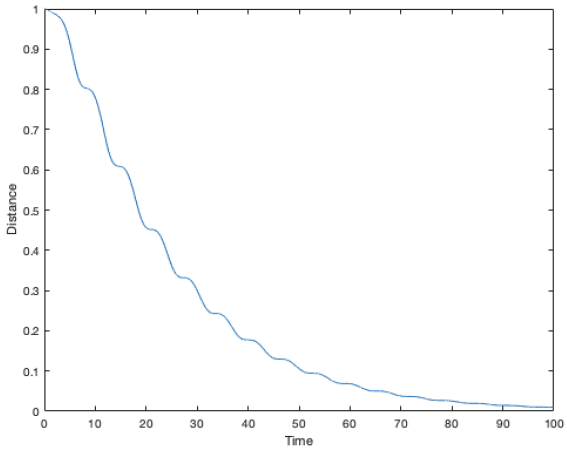


(a) Regular ($N = 10^5$) (b) Data-Driven ($N = 10^5$)

Figure 3.6a: Enough samples (first form)



(c) Abundant ($N = 10^5$)

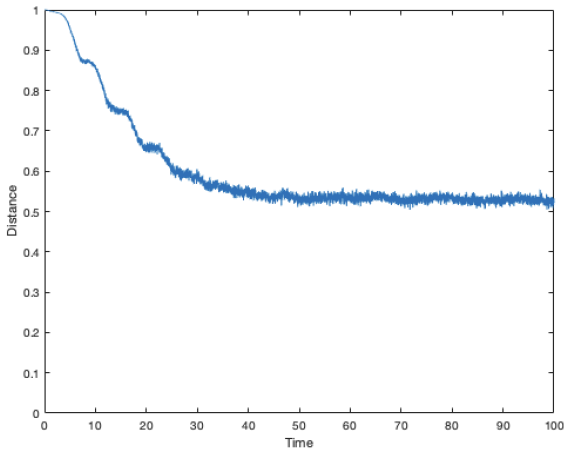


(d) Data-Driven ($N = 10^5$)

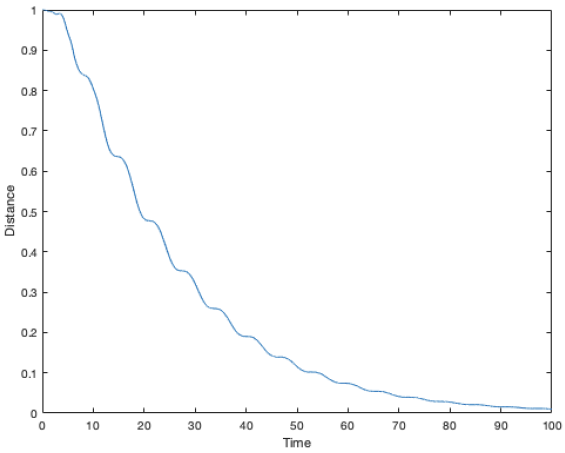
Figure 3.6b: Enough samples (second form)

Recall that there is a significant error when we generate a lack of samples in Monte-Carlo. We give such noisy data a chance to be the reference, since it saves computational expense.

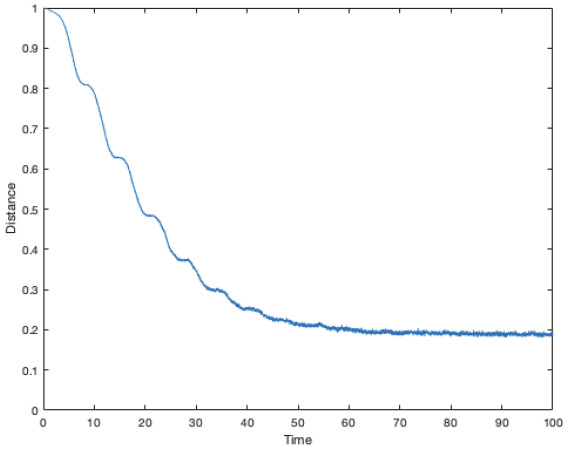
Figure 3.7d demonstrates the data-driven result with lack of samples' Monte-Carlo ($N = 10^4$). This data-driven result is consistent with the abundant samples' Monte-Carlo ($N = 10^6$) in Figure 3.6d intuitively. Without considering the numerical PDE part, we save 99 percent of computing time but obtain a similar result. When we choose the worst Monte-Carlo reference ($N = 10^3$ and present in Figure 3.7a), the corresponding data-driven result unexpectedly stays the same as the abundant samples' Monte-Carlo. However, it saves 99.9 percent of computing time. In Figure 3.7b there is only a small difference in the initial time period. During that time, the threshold does not succeed. Thus, this difference is produced by the lack of information of Monte-Carlo part.



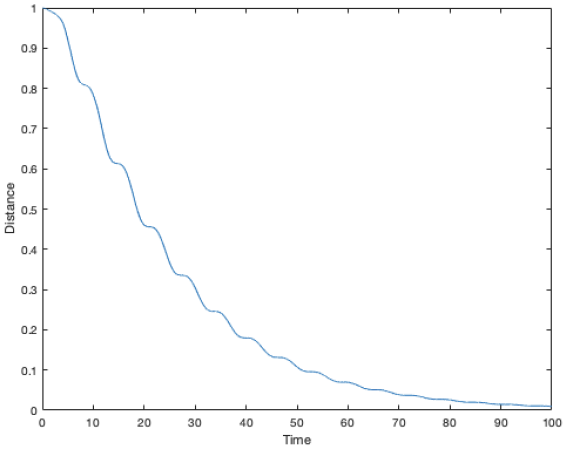
(a) Extremely Lack ($N = 10^3$)



(b) Data-Driven ($N = 10^3$)



(c) Lack ($N = 10^4$)



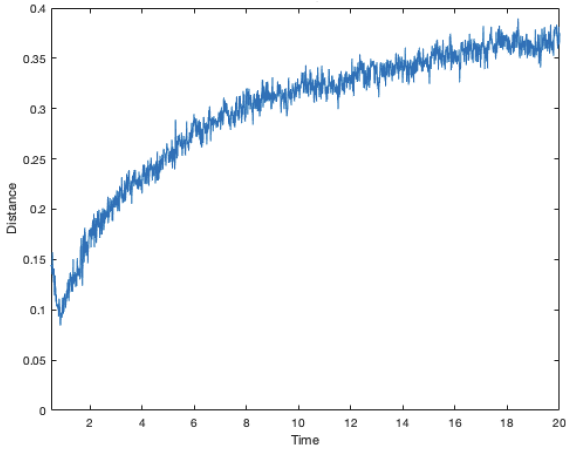
(d) Data-Driven ($N = 10^4$)

Figure 3.7: Lack of samples

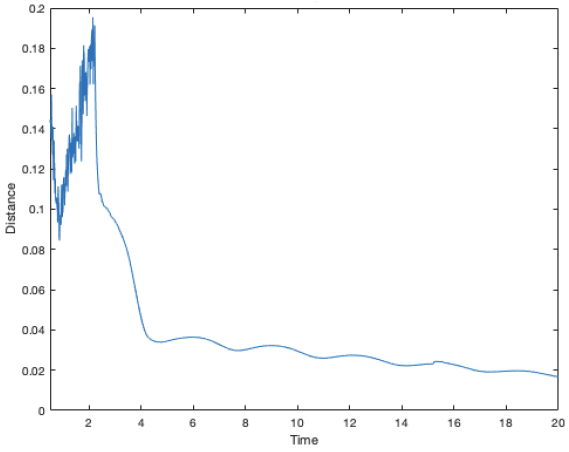
Therefore, Monte-Carlo's data preserves more information beyond direct observation, a noisy step function. With the strong constraint conditions provided by the scheme of the Fokker-Planck equations, only a few samples will recover the density function. The data-driven method is stable in simulating the densities, especially for a long time.

We analyze the accuracy heuristically. We need to compare the simulation given by the data-driven method with the exact solution (3.9). The simulation code of that exact solution is demonstrated in Appendix E.1. We still use the total variation distance to measure the difference. It will not be accurate when discretizing a delta function, even in the exact solution. Any tiny error causes the total variation to be

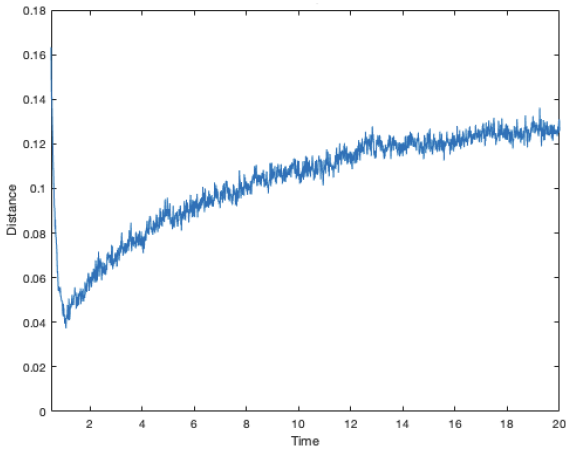
equal to 1. Therefore the starting time of the following error estimates is 0.5 rather than 0. We will discuss the initial time period in Section 3.2.2, which is called short time scale in that section. Keep the same initial setting. Let 20 be the end of the time. We make the comparison among different scales in the number of samples.



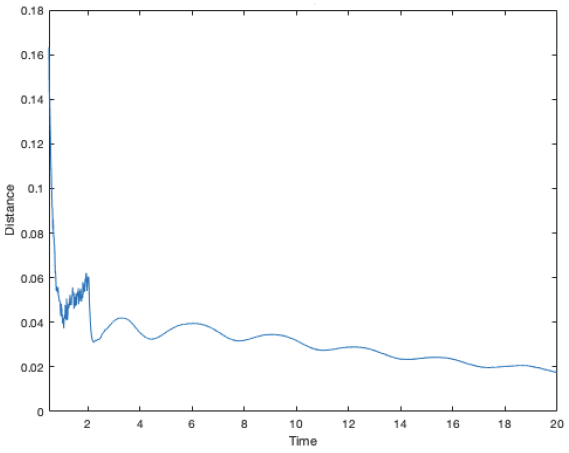
(a) Extremely Lack ($N = 10^3$)



(b) Data-Driven ($N = 10^3$)

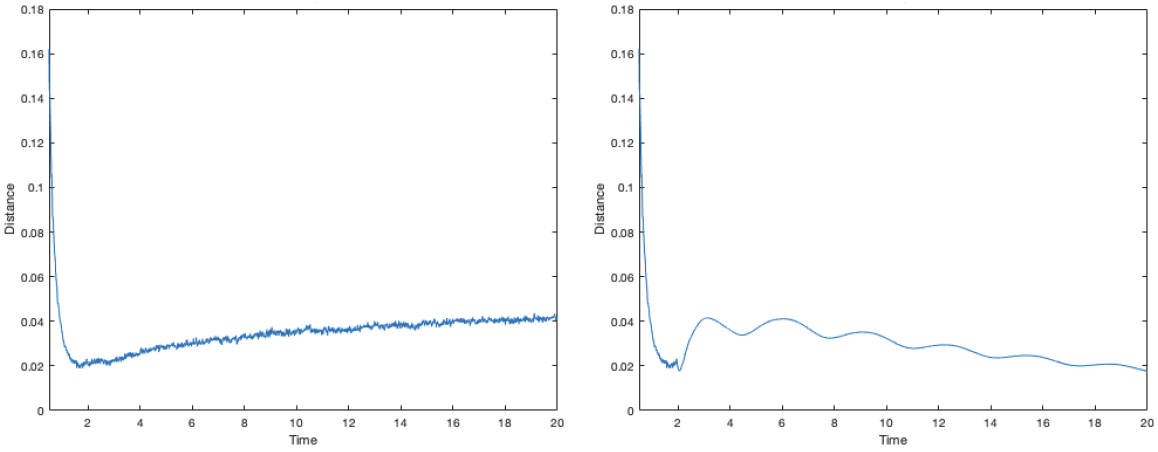


(c) Lack ($N = 10^4$)



(d) Data-Driven ($N = 10^4$)

Figure 3.8a: Error analysis (first form)



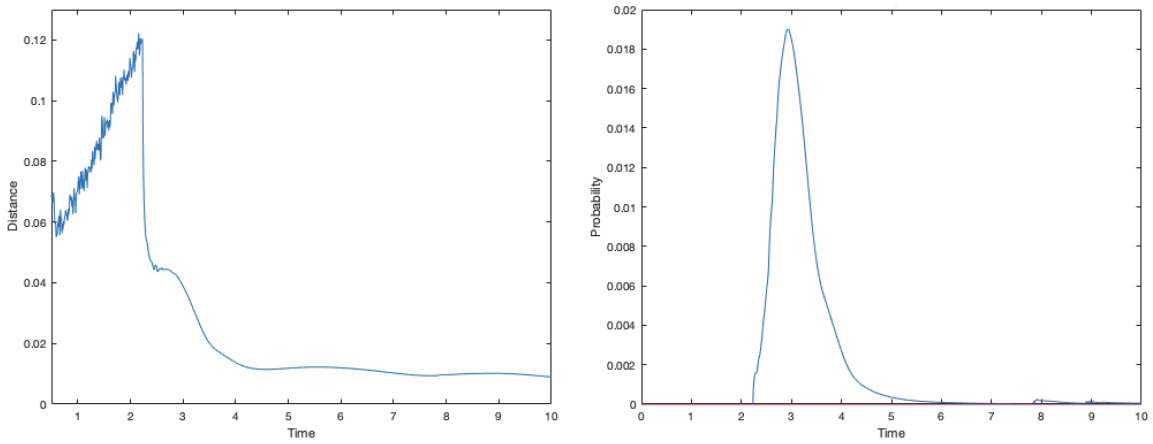
(e) Regular ($N = 10^5$) (f) Data-Driven ($N = 10^5$)

Figure 3.8b: Error analysis (second form)

When $N = 10^5$ the threshold of time is 1.96. In Figure 3.8f the error is the same as Monte-Carlo reference before that threshold. After that time, the error curve becomes smooth because the simulation result is an optimal solution for a PDE. A low level of error captures the curve. The overall decline should be expected to continue because the simulation and the exact solution tend to have a stationary distribution. In Figure 3.8e, when the time crosses the threshold, the error fluctuates around 0.04, which is mainly caused by the noise of the Monte-Carlo method.

In general, compared with Monte-Carlo references, the data-driven method provides more accurate results indeed. However, enriching the sample paths only decreases the errors in the initial time period, when only the Monte-Carlo scheme is in the running status and the optimal scheme is not invoked. When the time passes the threshold, the optimal scheme is in the running status as well. Because of the mechanism given by the Fokker-Planck equation, there is almost no difference between $N = 10^4$ and $N = 10^5$. Thus 10^4 samples are sufficient as reference data.

It is not hard to find that there are always some densities simulated by the data-driven method that have negative values, especially at the beginning of utilizing the data-driven method. In Figure 3.9b we accumulate the negative part of densities at each time. There is a noticeable portion right after the threshold time.



(a) Data-Driven ($N = 10^4$) (b) Masses in the Negative Domain

Figure 3.9: Negative density problem

That phenomenon does not lead by the Crank-Nicolson scheme. For the same data reference, there should be several optimal solutions. Unfortunately, the minimum norm least-squares solution, provided by *lsqminnorm* in MATLAB, is separate from the exact solution. Therefore, we use the *lsqnonneg* in MATLAB to extract the non-negative solution.

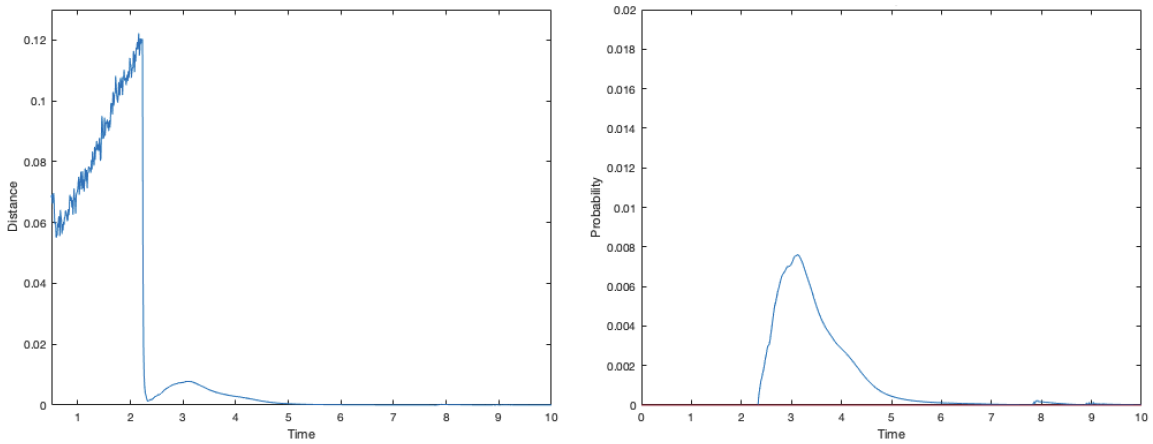


Figure 3.10: 10 steps non-negative optimization

Since the operand of *lsqnonneg* is large, after the threshold, we attempt 10 steps of the non-negative optimization. The negative densities arise again when we start to solve the minimum norm least-squares solution after the 10 steps. However, it prevents more than half of the masses from being transported to negative compared with Figure 3.9b. More importantly, as time passes, it is far more accurate than

before. The total variation distance in Figure 3.9a maintains 0.01, while with our new technique it almost decreases to 0 after $t = 5$.

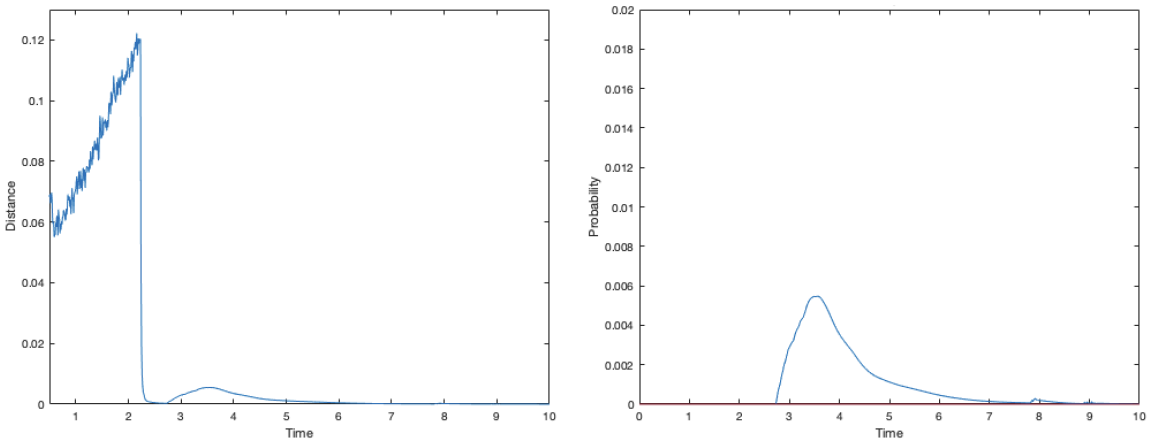
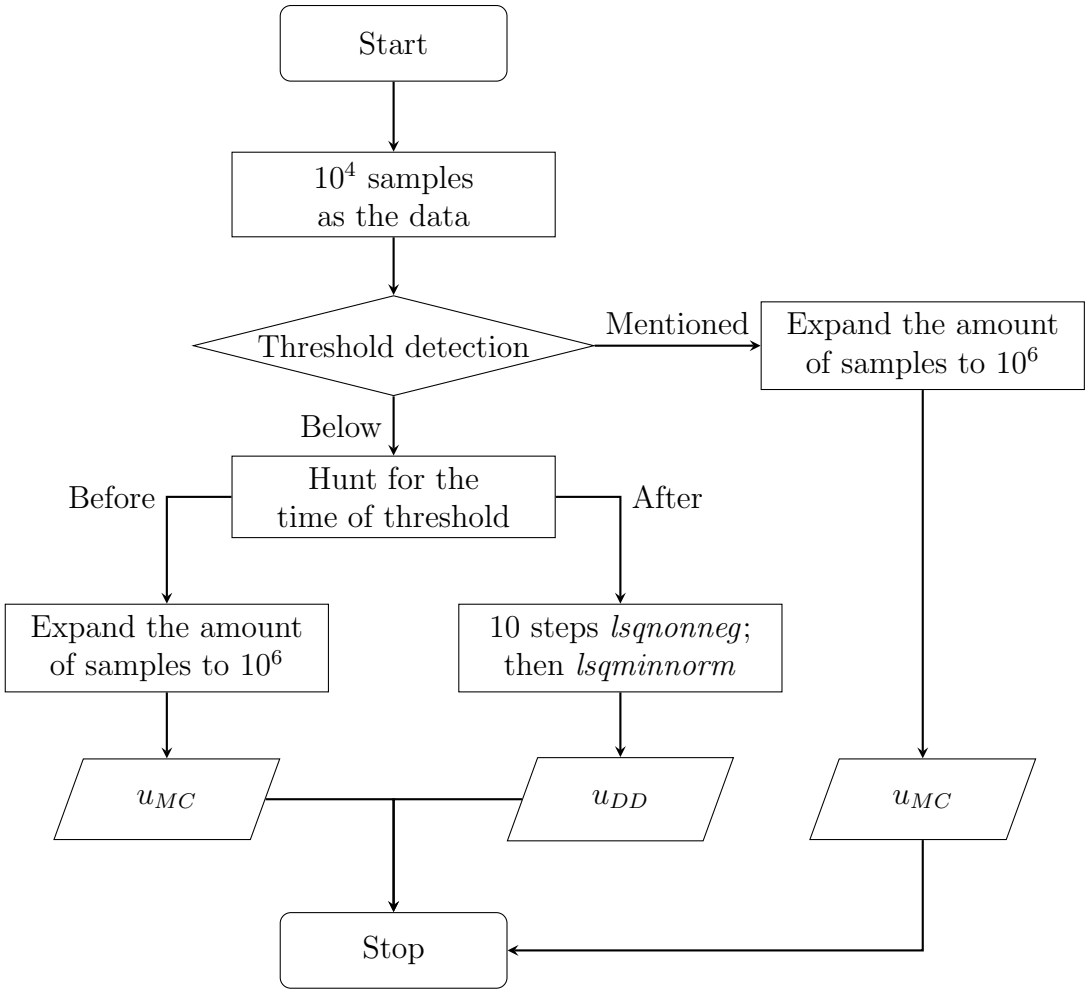


Figure 3.11: 50 steps non-negative optimization

When increasing the number of steps to 50, we should obtain a better result naturally. However, compared with the right ones in Figure 3.10 and Figure 3.11, the result is quite different from what we expected. After 50 steps of non-negative optimization, the negative mass's error is almost identical to the 10 steps' case. Nevertheless, that process costs 3343 seconds which is more than 10 times expending of the 10 steps' case. The full of non-negative least-squares optimizations without considering the computing power should be prime. However, in our multiscale analysis, we need to consider the efficiency of long time simlulation. Therefore, we suggest using the 10 steps' non-negative least-squares optimizations. The technological process can be generated as follows:



In the threshold detection, we will search for the maximum value of the densities of the last 10 steps. If any values in the last 10 steps exceeds the threshold, then the Monte-Carlo still provides a delta-like data reference. Thus we will not operate our data-driven method. Instead, we increase the number of samples to sufficiently large. Let the Monte-Carlo result u_{MC} represent the numerical solution. In general, the short time scheme always follows this flow.

On the contrary, when all of those density values are below the threshold, we can at least accomplish 10 steps non-negative optimization. Therefore, we apply our data-driven scheme. Combine the Monte-Carlo result u_{MC} in the front and the data-driven result u_{DD} in the back as our numerical solution.

3.2.2 Multiscale simulation in the harmonic potential

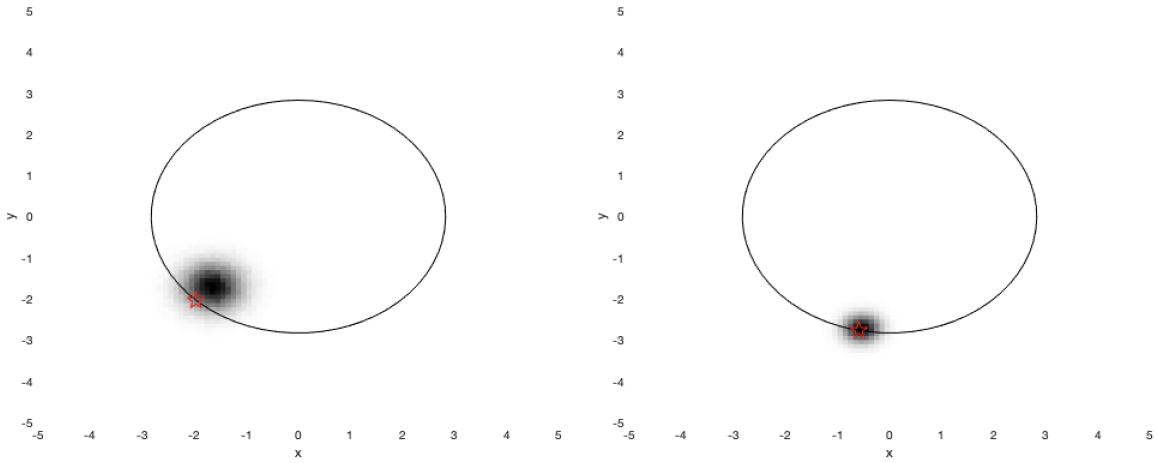
So far, we have enough tools to implement the multiscale analysis numerically. Let us recall our theoretical results first. Theorem 17 and Theorem 18 demonstrate that, in the short time scale ($t_\epsilon \ll \epsilon^{-1}$), the solution of the Langevin equation $X^\epsilon_{t_\epsilon}$ is close to the solution of the Hamiltonian system at the same time $X^0_{t_\epsilon}$ with small ϵ. In the long time scale ($t_\epsilon \gg \epsilon^{-1}$), Theorem 16 ensures $X^\epsilon_{t_\epsilon}$ eventually tends to its stationary distribution when ϵ tends to zero.

Numerical simulation of the short time scale

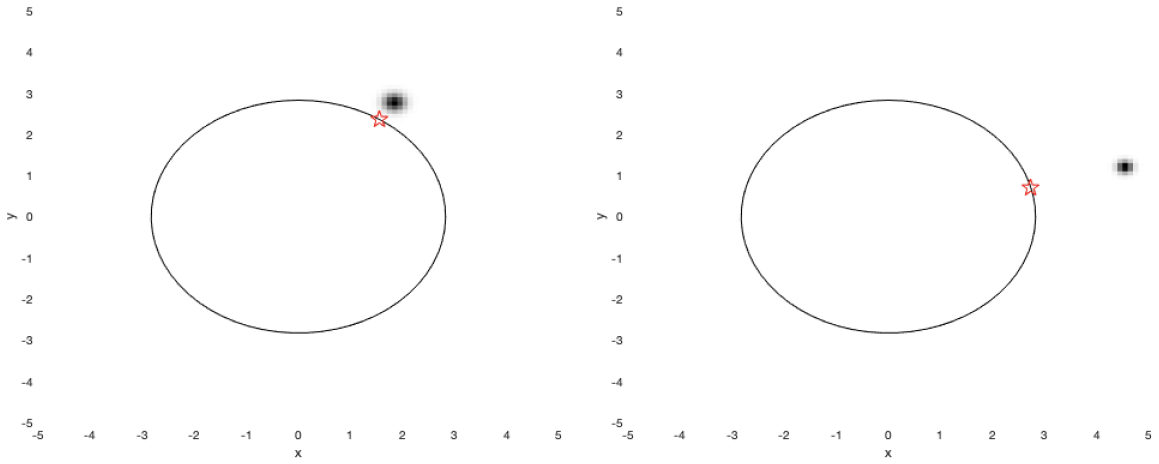
The limiting distribution theoretically satisfies a delta function at $X^0_{t_\epsilon}$. When ϵ is small, t_ϵ is not long enough to escape the front time interval of delta-like densities. Thus our data-drive method is reduced to the Monte-Carlo method. The accuracy depends on both the scales of samples and the iteration scheme.

Theoretically, we need to compare the distribution of $X^\epsilon_{t_\epsilon}$ with a single point on the Hamiltonian orbit $X^0_{t_\epsilon}$. Since $X^\epsilon_{t_\epsilon}$ is a random variable with 2 dimensional Gaussian distribution, the first two moments are sufficient to represent it. By Theorem 17 and Theorem 18, the mean will tend to $X^0_{t_\epsilon}$ while the variance will tend to zero when ϵ tends to zero, where $X^0_{t_\epsilon}$ is on the Hamiltonian orbit with the same initial condition as $X^\epsilon_{t_\epsilon}$.

We determinate the end time point $t_\epsilon = \epsilon^{-0.5}$ as a short time scale. The initial settings are $k = 1$, $x^\epsilon_0 = 2$ and $y^\epsilon_0 = 2$. Operating the code in Appendix E.6 we obtain the densities of $X^\epsilon_{t_\epsilon}$ where ϵ changes from 10^{-1} to 10^{-4}. We select the Euler-Maruyama iteration in our Monte-Carlo simulation. The result is demonstrated in Figure 3.12. The black curve is the Hamiltonian orbit. The red star denotes the corresponding Hamiltonian solution X^0_t at t_ϵ. When ϵ is small, the distribution of X^ϵ_t exactly tends to a delta-like distribution as we expect. However, there is a visible difference between the mean of that distribution and $X^\epsilon_{t_\epsilon}$ in Figure 3.12d. Intuitively, the differences will diverge when ϵ tends to zero. It seems that the algorithm properly maintains the value on the angle direction but accumulates errors in the energy level direction.



(a) $\epsilon = 0.1$ (b) $\epsilon = 0.01$



(c) $\epsilon = 0.001$ (d) $\epsilon = 0.0001$

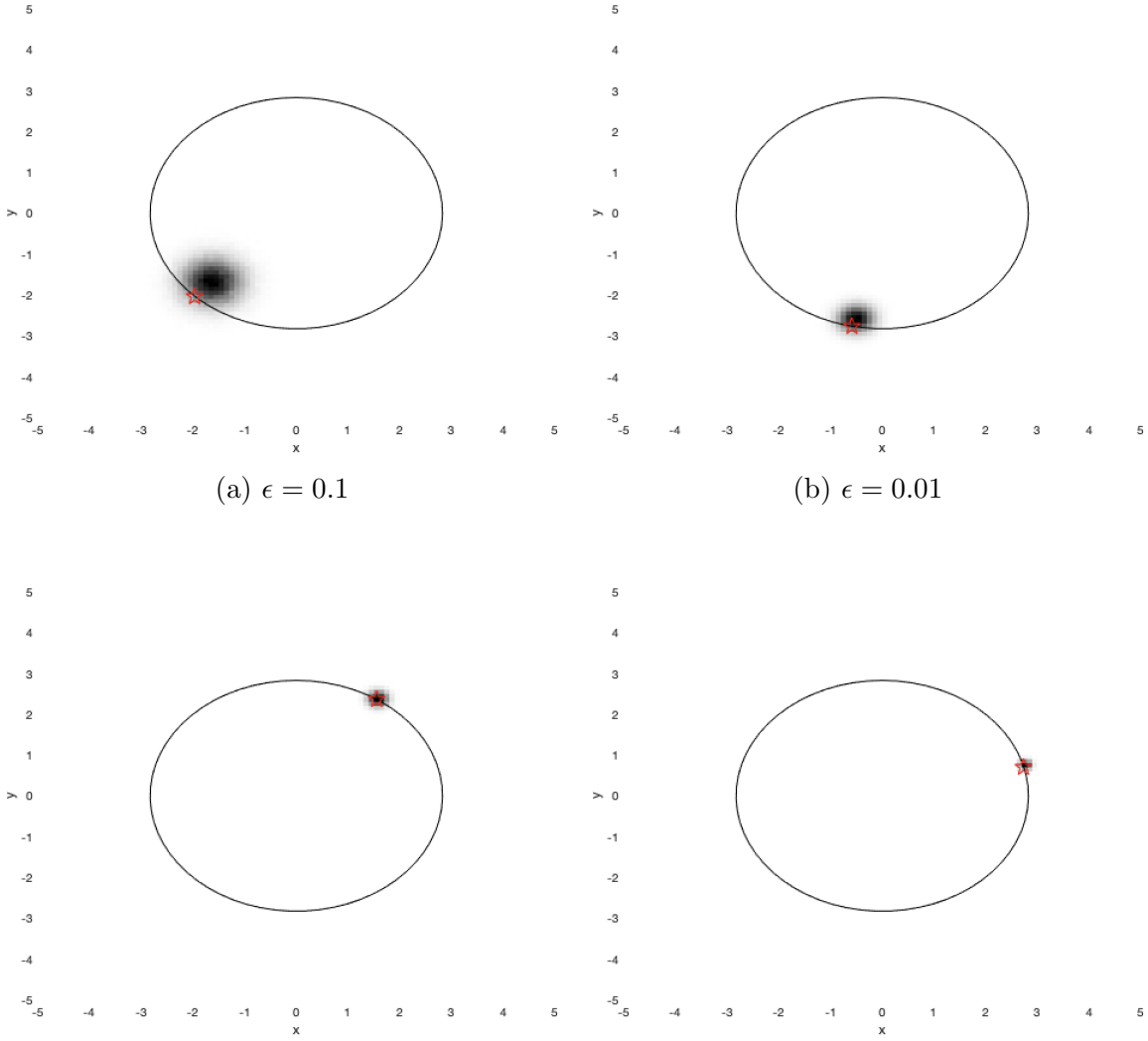
Figure 3.12: Evolution of densities in the short time scale

The iteration scheme provides this mistake. Even for the deterministic Hamiltonian system, both the Euler iteration and the implicit Euler iteration will destroy the Hamiltonian orbit. In order to keep the conservative property of a Hamiltonian system, people need to use a conservative algorithm. The conservative algorithm of the deterministic system is provided in Appendix C. Unfortunately, in the Euler-Maruyama iteration, the deterministic part is exact the Euler iteration. In order to eliminate this error, we need to utilize a stochastic version of the conservative algorithm.

By Section 2.3 of [47], the stochastic version of the leapfrog method is

$$\begin{aligned} x_{hs} &= x_n - y_n \frac{\Delta t}{2}, \\ y_{n+1} &= y_n - k x_{hs} \Delta t - \epsilon y_n + \sqrt{\epsilon} \Delta W_n, \\ x_{n+1} &= x_{hs} + y_{n+1} \frac{\Delta t}{2}, \end{aligned}$$

which is a conservative algorithm. We change the iteration scheme in Appendix E.6 to the leapfrog method above. Keep the same initial settings above. The evolution is presented in Figure 3.13



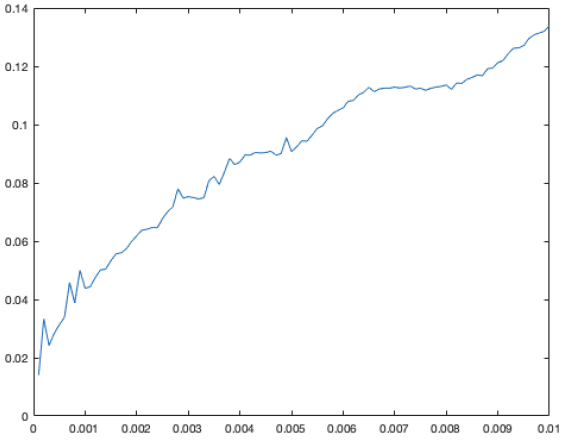
(a) $\epsilon = 0.1$ (b) $\epsilon = 0.01$

(c) $\epsilon = 0.001$ (d) $\epsilon = 0.0001$

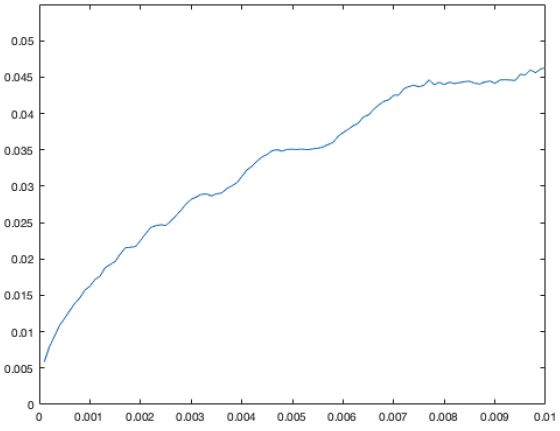
Figure 3.13: Evolution of densities in the short time scale with the leapfrog scheme

Intuitively our numerical result of $X^{\epsilon}_{t_\epsilon}$ gets closer to the Hamiltonian solution X^0_t

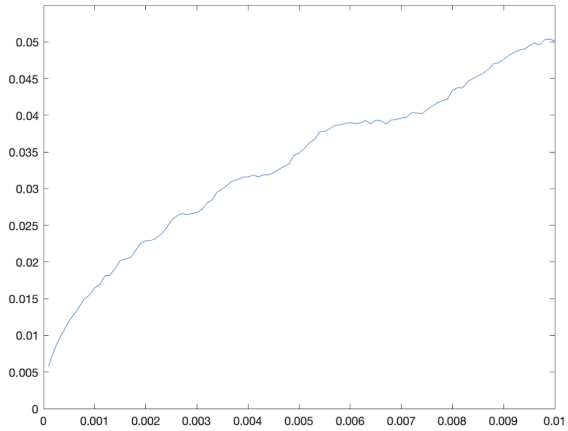
at $t = t_\epsilon$. Since X_t^ϵ is a Gaussian process in the harmonic case, the mean and variance can entirely determine it. In order to evaluate the difference between $X_{t_\epsilon}^\epsilon$ and $X_{t_\epsilon}^0$, we can directly simulate the mean and variance of $X_{t_\epsilon}^\epsilon$. Then compare the mean with $X_{t_\epsilon}^0$. Simultaneously, compare the variance with zero. In Appendix E.6 variable *diff_2norm* accumulates the differences between $\mathbb{E}X_{t_\epsilon}^\epsilon$ and $X_{t_\epsilon}^0$ in the second norm. Meanwhile use *sigma_x* and *sigma_y* to collect the variances of marginal distributions of $X_{t_\epsilon}^\epsilon$ in the x-direction and y-direction respectively.



(a) $\left|\left|\mathbb{E}X_{t_\epsilon}^\epsilon - X_{t_\epsilon}^0\right|\right|_2$ depends on ϵ



(b) Variance in x depends on ϵ



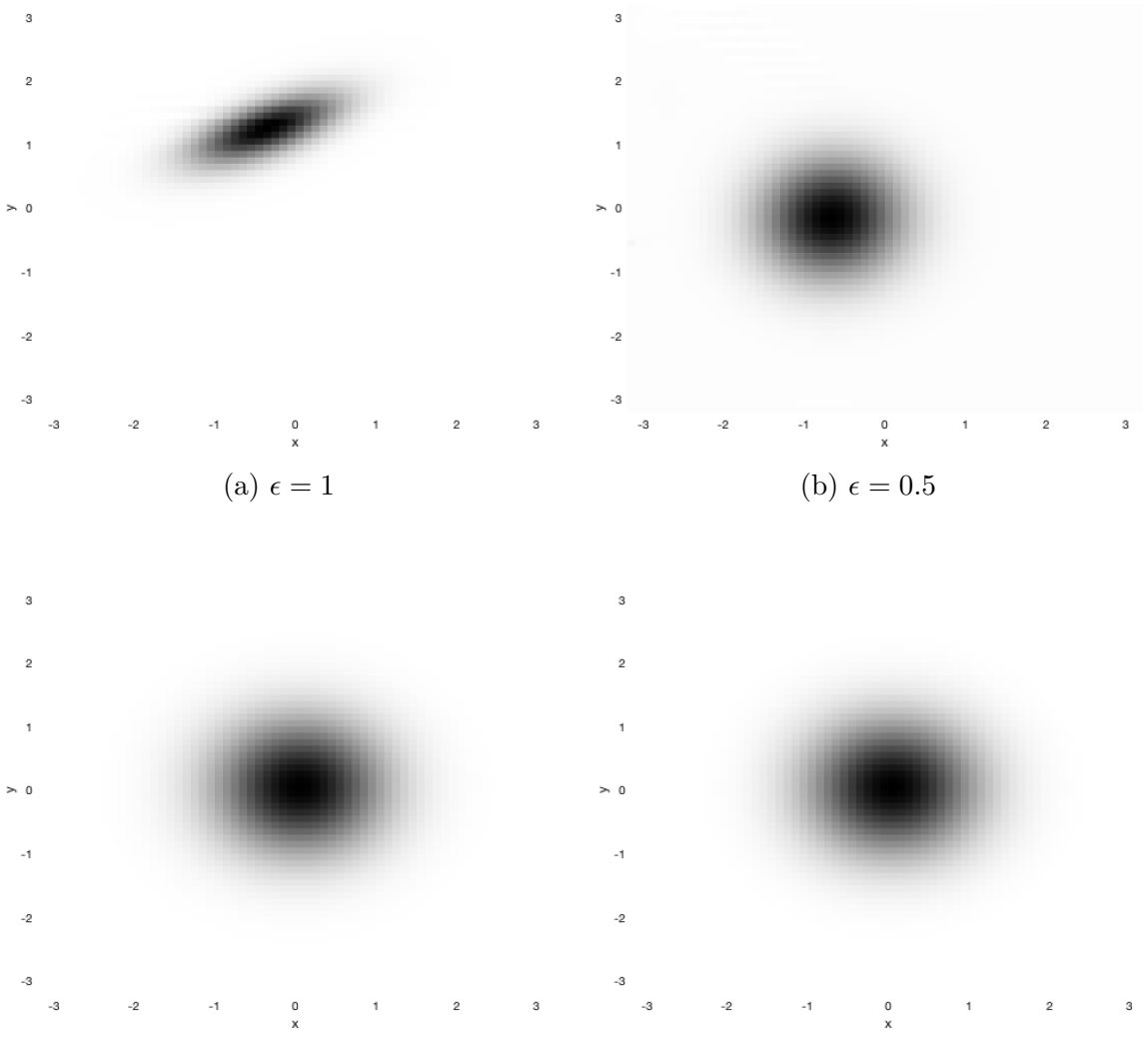
(c) Variance in y depends on ϵ

Figure 3.14: Convergency of the mean difference and the variances

With the leapfrog scheme, we successfully visualize our Theorem 17 and Theorem 18 in Figure 3.14. In the short time scale $X_{t_\epsilon}^\epsilon$ indeed tends to the point $X_{t_\epsilon}^0$. We will utilize a similar scheme in the nonlinear case to estimate the short time scale before giving a theoretical proof.

Numerical simulation of the long time scale

We follow our data-driven method introduced in the last section. With each fixed ϵ we can simulate the densities of X_t^ϵ evolving with time. We set the end time point as t_ϵ, where $t_\epsilon \gg \epsilon^{-1}$. In the following simulation we determinate $t_\epsilon = \epsilon^{-1.5}$. For convenience of the simulation, in system (3.8) we make the initial settings where $k = 1$, $x_0^\epsilon = 1$ and $y_0^\epsilon = 1$. Let ϵ change from 1 to 10^{-2}. We obtain a sequence of densities of $X_{t_\epsilon}^\epsilon$ with respect to ϵ. The code of this simulation is in Appendix E.5.



(a) $\epsilon = 1$ (b) $\epsilon = 0.5$

(c) $\epsilon = 0.01$ (d) Stationary Distribution

Figure 3.15: Evolution of densities in the long time scale

The last figure in Figure 3.15 presents the corresponding stationary distribution. (3.12) is the explicit form of it. Intuitively when $\epsilon = 0.01$, the distribution of $X_{t_\epsilon}^\epsilon$ is

almost the same as the stationary distribution on the right. We use the total variation distance in Theorem 16 to measure the above difference. Therefore we compare the sequence of simulations with the stationary distribution in the total variation distance.

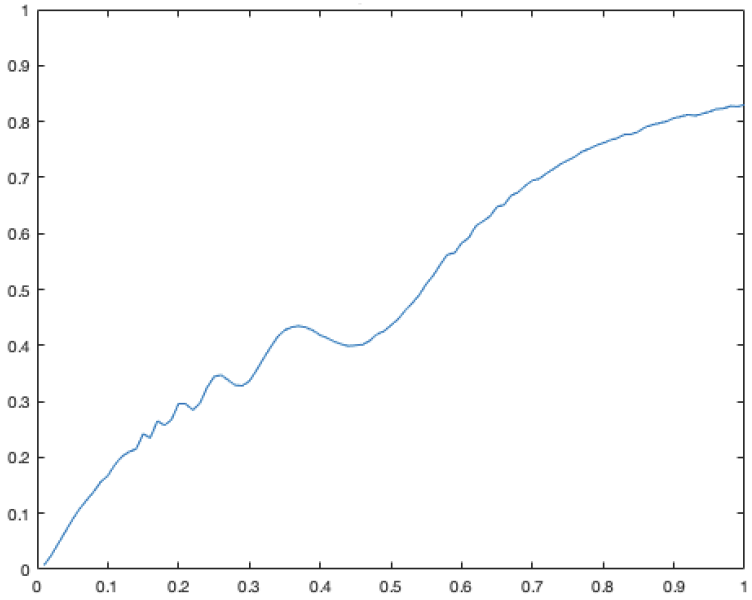


Figure 3.16: $\mathbf{d}_{\mathrm{TV}}\left(X^{\epsilon}_{t_\epsilon}, X_\infty\right)$ depends on ϵ

When ϵ tends to zero, the total variation distance is precisely zero. With this data, in the future, we will estimate the convergency rate as a function depending on the small variable ϵ.

3.2.3 Multiscale simulation in the non-harmonic potential case

In the non-harmonic case, the only difference is the derivative of the potential term,

$$\begin{cases} \mathrm{d}x^{\epsilon}_t = y^{\epsilon}_t \mathrm{d}t, \\ \mathrm{d}y^{\epsilon}_t = \left(-V'(x^{\epsilon}_t) - \epsilon y^{\epsilon}_t\right) \mathrm{d}t + \sqrt{2\epsilon\beta^{-1}}\mathrm{d}W_t. \end{cases} \tag{3.24}$$

Rather than being kx in the harmonic case (3.8), it is a general nonlinear function. Therefore the corresponding Hamiltonian orbit will be modified by the structure of the potential.

But the forms of the damping $-\epsilon y \mathrm{d}t$ and noise $\sqrt{2\epsilon\beta^{-1}}\mathrm{d}W_t$ do not change. Roughly speaking, the damping term will lead the solution to its stable equilibrium, and the

noise term will introduce the variation of the solution. In the harmonic case, we know those two terms do not take effect in the short time scale. The Hamiltonian system dominates the solution. In the long time scale, they are fully functional in the system, such that the solution tends to its stationary distribution, which is the equilibrium of the stochastic differential equation. We suppose the above two terms provide the multiscale phenomenon. Thus the short time scale and long time scale of the system (3.24) should be the same as the harmonic case, which are $t_\epsilon^s \ll \epsilon^{-1}$ and $t_\epsilon^l \gg \epsilon^{-1}$ respectively.

In this section, we utilize a single well potential,

$$V(x) = \frac{1}{2}x^2 + \frac{1}{4}x^4,$$

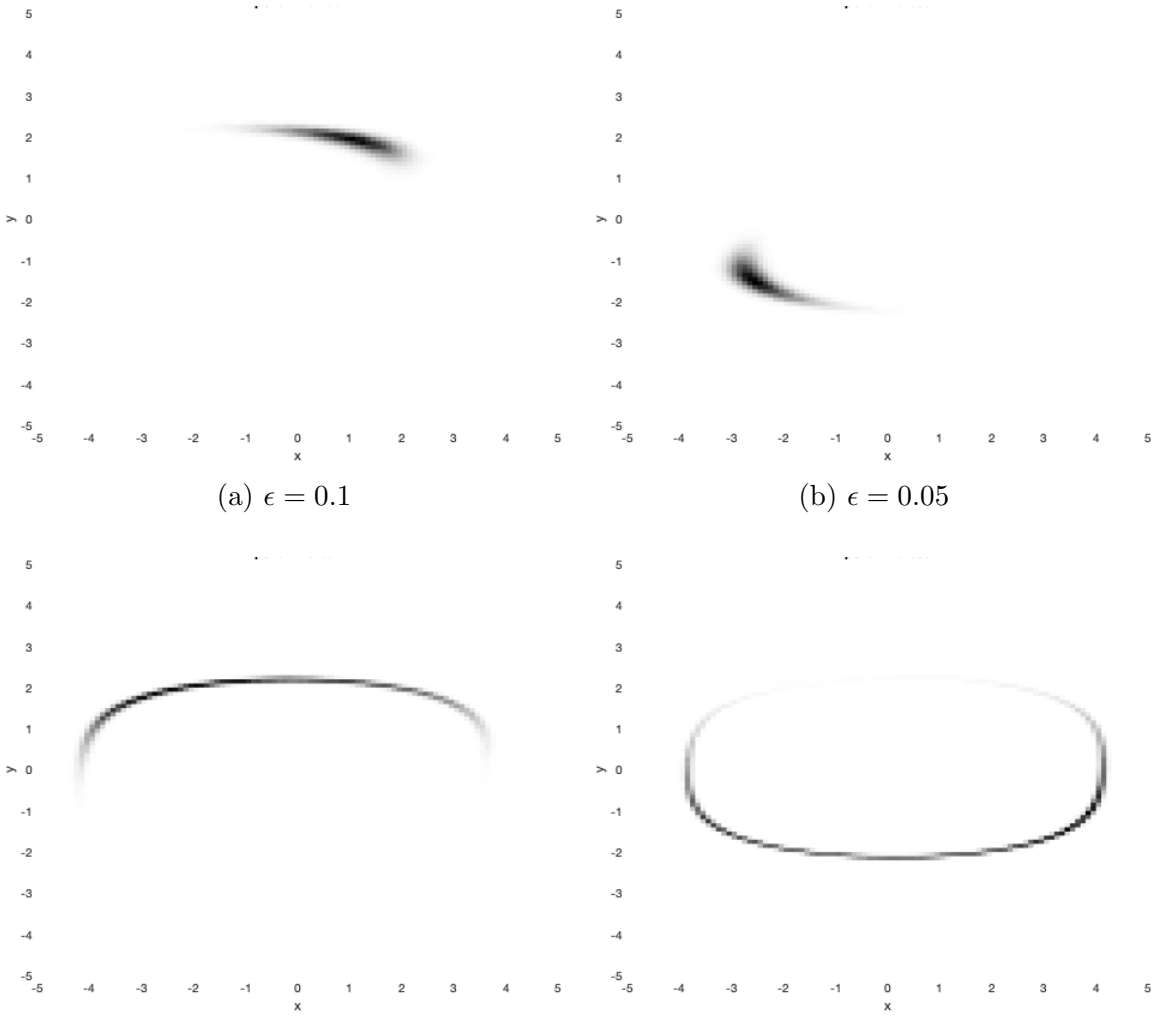
as a toy model of the non-harmonic potential, and let $\beta = 2$. Like the harmonic case, we will still select different time scales to observe the short time and long time behaviors.

Numerical simulation of the short time scale

In the short time scale, the distribution of the limiting result is degenerate. We directly apply the Monte-Carlo method as what we do in the harmonic case. The leapfrog method is chosen as the iteration algorithm to preserve the Hamiltonian dynamics. The initial settings are $x_0^\epsilon = 1$ and $y_0^\epsilon = 1$. Let the time scale $t_\epsilon = \epsilon^{-0.5}$ where ϵ changes from 10^{-2} to 10^{-5}. The result is demonstrated in Figure 3.17. The blue curve is the corresponding Hamiltonian orbit.

As the first observation, when ϵ tends to zero, the simulation result gets closer to the Hamiltonian orbit. However, it does not tend to have a delta-like distribution as in the harmonic case. In this specific time scale, the limiting distribution seems to be supported on the entire curve with a certain degree of variation around it. We must understand whether the Langevin dynamics or numerical discretization provides this phenomenon.

Two reasons lead us to believe that numerical discretization introduces this variation. First, the simulation will present this singular phenomenon even if the real limiting distribution is a delta distribution at $X_{t_\epsilon}^0$. Our initial condition is a delta function at $(x_0^\epsilon, y_0^\epsilon)$. In order to discretize it, we have to approximate it by a uniform



(a) $\epsilon = 0.1$ (b) $\epsilon = 0.05$

(c) $\epsilon = 0.001$ (d) $\epsilon = 0.0001$

Figure 3.17: Evolution of densities in the short time scale

distribution at the grid covering the initial condition. In the harmonic case, this approximation introduces a nonnegligible error at the front period of the time, which is demonstrated in Figure 3.8e. In the non-harmonic case, this type of error cannot be avoided in a short time scale.

Consider the Langevin dynamics in the energy and angle variables. The time derivative of the angle variable is the frequency along a fixed Hamiltonian orbit. The error which is introduced by the initial condition influences both the energy and the angle. In the harmonic case, the frequencies are the same on all energy levels. Then the errors of the energy and the angle are independent. Thus the ranges of errors in both directions are similar, and the limiting distribution is a delta-like distribution

which is demonstrated in Figure 3.13. In the non-harmonic case, the frequency is a function that depends on energy. A slight variance in the energy causes a noticeable change in the angle. Even if there is only a tiny energy error, the nonlinear dynamics will multiply the error in the angle direction. This should be one of the reasons for us to observe the mixing in the angle in Figure 3.17d.

We will prove in the short time scale, $t_\epsilon^s \ll \epsilon^{-1}$, the energy is close to the initial energy almost surely. Thus the visible variance of the energy in Figure 3.17d should also be introduced by the numerical discretization.

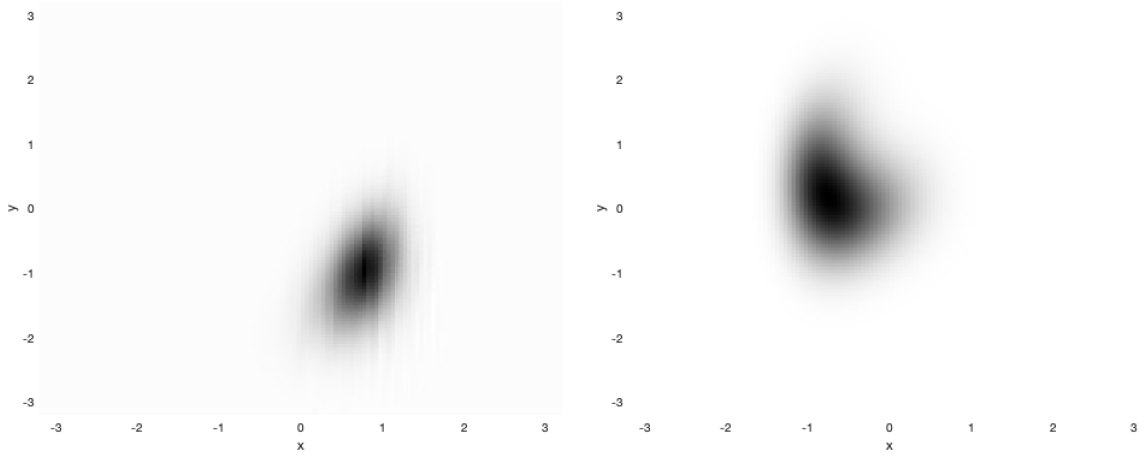
In conclusion, this error can be relieved by utilizing finer gridding, collecting more samples, and selecting smaller ϵ. However, each method will cause exponential growth in the computation cost. Those methods also do not overcome the essential part of the error. In the non-harmonic case, any energy error leads to the mixing in the angle. Thus, how we can visualize the short time behavior in the non-harmonic case is an interesting and unsolved problem.

Numerical simulation of the long time scale

In this section, we will compare the simulation result of X_t^ϵ with its equilibrium in the total variation distance. By (3.4), the stationary distribution is

$$\rho_2(x, y) = Z^{-1} e^{-y^2 - x^2 - \frac{1}{2}x^4},$$

where Z is the normalization factor. We select the time scale $t_\epsilon = \epsilon^{-1.5}$ as an example of the long time scale. Let $x_0^\epsilon = 1$, $y_0^\epsilon = 1$ and ϵ change from 1 to 10^{-2}. Then the sequence of densities of $X_{t_\epsilon}^\epsilon$ presents in Figure 3.18. In this specific time scale, the distributions of $X_{t_\epsilon}^\epsilon$ tend to its stationary distribution. Figure 3.19 demonstrates the difference's evolution. When ϵ tends to zero, the difference tends to zero.



(a) $\epsilon = 1$ (b) $\epsilon = 0.5$

(c) $\epsilon = 0.01$ (d) Stationary Distribution

Figure 3.18: Evolution of densities in the long time scale

When we determine the time scale $t_\epsilon = \epsilon^{-1.3}$ or $t_\epsilon = \epsilon^{-1.7}$, the densities still converge to the stationary distribution when ϵ tends to zero. Therefore, the above verified time scales are long time scales. Of course, we cannot obtain all of the long time scales by the enumeration method. However, after enumerating more cases, we can conclude that the density always tends to the stationary distribution when the time passes $t_\epsilon = \epsilon^{-1}$. We suppose that the long time scale is $t_\epsilon^l \gg \epsilon^{-1}$.

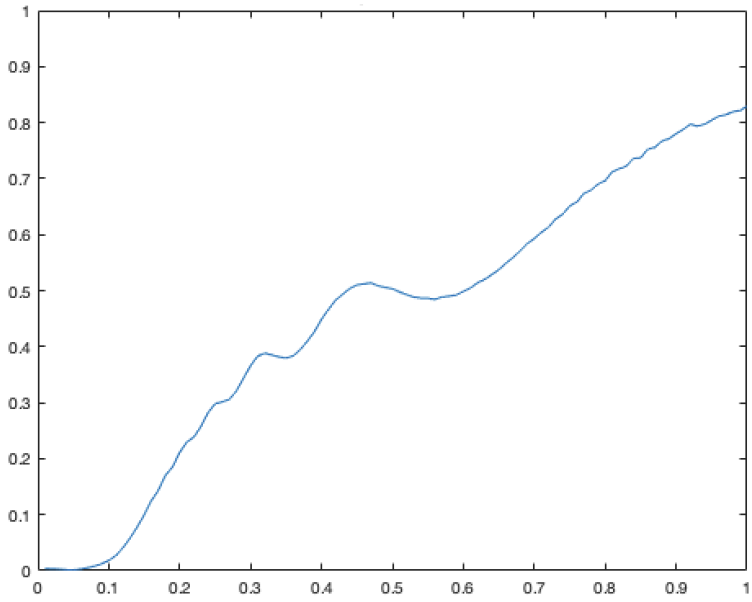


Figure 3.19: $\mathbf{d}_{\mathrm{TV}}\left(X^{\epsilon}_{t_\epsilon}, X_\infty\right)$ depends on ϵ

3.3 Analysis of the non-harmonic oscillator with external noise

This section provides some theoretical results regarding the short and long time behaviors of the non-harmonic Langevin equation (3.5). The simulation in the previous section provides us with an estimate of the time scales in the one-dimensional case. We will generalize the short and long time scales to d-dimensional Langevin equation.

3.3.1 Short time behavior

We aim to obtain results similar to Theorem 17 and Theorem 18 in the short time scale. In the nonlinear case, it appears to be challenging to track all of the coordinates of $X^{\epsilon}_{t_\epsilon}$. The simulation in Section 3.2.3 presents the complexity of the angle variable in the one-dimensional case. Thus we start to analyze the total energy

$$E(x, y) = \frac{1}{2}||y||_2^2 + V(x),$$

where the first term is the kinetic energy and the second term is the potential energy.

The total energy is one of the conserved quantities in the Hamiltonian system.

Directly differentiating the total energy of X_t^0, we have

$$\mathrm{d}E\left(x_t^0, y_t^0\right) = \nabla E \cdot \begin{pmatrix} \mathrm{d}x_t^0 \\ \mathrm{d}y_t^0 \end{pmatrix} = \nabla V(x_t^0) \cdot y_t^0 \mathrm{d}t + y_t^0 \cdot \left(-\nabla V(x_t^0)\right) = 0.$$

Therefore $E\left(x_t^0, y_t^0\right) \equiv E(x_0, y_0)$ for any t. Denote the value as E_0 for convenience in the following theorem.

The Hamiltonian orbit should capture the solution X_t^ϵ in a proper short time scale. Then the energy of X_t^ϵ is more or less the same as that at the initial energy E_0. The simulation in Section 3.2.3 did not entirely visualize the short time behaviors because of the sensitive angle variable. However, it illustrates that when $t_\epsilon \ll \epsilon^{-1}$ the energy does not start to dissipate, and the majority distribution of X_t^ϵ stays around the Hamiltonian orbit. Therefore we suppose the short time scale is $t_\epsilon \ll \epsilon^{-1}$.

Theorem 19 (Law of Large Numbers on the total energy) *Let $\epsilon t_\epsilon \ll 1$. Then*

$$\lim_{\epsilon \to 0} \mathbb{P}\left(\sup_{0 \le t \le t_\epsilon} \left| E\left(x_t^\epsilon, y_t^\epsilon\right) - E_0 \right| > \delta \right) = 0 \quad \textit{for every } \delta > 0.$$

Proof. Applying the Itô formula on $E(x_t^\epsilon, y_t^\epsilon)$, we obtain

$$\mathrm{d}E\left(x_t^\epsilon, y_t^\epsilon\right) = \epsilon\left(\beta^{-1} d - ||y_t^\epsilon||_2^2\right)\mathrm{d}t + \sqrt{2\epsilon\beta^{-1}} y_t^\epsilon \cdot \mathrm{d}W_t. \tag{3.25}$$

Since $V(x) \ge 0$, we obtain

$$\frac{1}{2}||y_t^\epsilon||_2^2 \le E\left(x_t^\epsilon, y_t^\epsilon\right) = E_0 + \beta^{-1} d\epsilon t - \epsilon \int_0^t ||y_s^\epsilon||_2^2 \mathrm{d}s + \sqrt{2\epsilon\beta^{-1}} \int_0^t y_s^\epsilon \cdot \mathrm{d}W_s.$$

Taking the expectation in the above yields

$$\mathbb{E}||y_t^\epsilon||_2^2 \le 2E_0 + 2\beta^{-1} d\epsilon t. \tag{3.26}$$

Then, the integral form of (3.25) can be estimated as follows:

$$\left|E\left(x_t^\epsilon, y_t^\epsilon\right) - E_0\right| \le \beta^{-1} d\epsilon t + \epsilon \int_0^t ||y_s^\epsilon||_2^2 \mathrm{d}s + \sqrt{2\epsilon\beta^{-1}} \left| \int_0^t y_s^\epsilon \cdot \mathrm{d}W_s \right|.$$

Therefore,

$$\sup_{0 \le t \le t_\epsilon} \left|E\left(x_t^\epsilon, y_t^\epsilon\right) - E_0\right| \le \beta^{-1} d\epsilon t_\epsilon + \epsilon \int_0^{t_\epsilon} ||y_s^\epsilon||_2^2 \mathrm{d}s + \sqrt{2\epsilon\beta^{-1}} \sup_{0 \le t \le t_\epsilon} \left| \int_0^t y_s^\epsilon \cdot \mathrm{d}W_s \right|.$$

Since $\epsilon t_\epsilon \to 0$ as $\epsilon \to 0$, there exists a $\epsilon_0(\delta)$ such that when $\epsilon < \epsilon_0$ we have $2\beta^{-1}d\epsilon t_\epsilon < \delta$. It follows that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le t_\epsilon}\left|E\left(x_t^\epsilon, y_t^\epsilon\right) - E_0\right| > \delta\right) \\
&\le \mathbb{P}\left(\epsilon\int_0^{t_\epsilon}||y_s^\epsilon||_2^2\mathrm{d}s + \sqrt{2\epsilon\beta^{-1}}\sup_{0\le t\le t_\epsilon}\left|\int_0^t y_s^\epsilon\cdot\mathrm{d}W_s\right| > \frac{\delta}{2}\right) \\
&\le \mathbb{P}\left(\epsilon\int_0^{t_\epsilon}||y_s^\epsilon||_2^2\mathrm{d}s > \frac{\delta}{4}\right) + \mathbb{P}\left(\sqrt{2\epsilon\beta^{-1}}\sup_{0\le t\le t_\epsilon}\left|\int_0^t y_s^\epsilon\cdot\mathrm{d}W_s\right| > \frac{\delta}{4}\right). \qquad (3.27)
\end{aligned}$$

By using the Fubini's theorem and (3.26), we have

$$\mathbb{E}\left[\epsilon\int_0^t||y_s^\epsilon||_2^2\mathrm{d}s\right] = \epsilon\int_0^t\mathbb{E}||y_s^\epsilon||_2^2\mathrm{d}s \le 2E_0\epsilon t + \beta^{-1}d\epsilon^2t^2.$$

Applying the Markov's inequality to the first term of (3.27), we have

$$\mathbb{P}\left(\epsilon\int_0^{t_\epsilon}||y_s^\epsilon||_2^2\mathrm{d}s > \frac{\delta}{4}\right) \le \frac{4}{\delta}\,\mathbb{E}\left[\epsilon\int_0^{t_\epsilon}||y_s^\epsilon||_2^2\mathrm{d}s\right] \le \frac{8E_0}{\delta}\epsilon t_\epsilon + \frac{4\beta^{-1}d}{\delta}\epsilon^2{t_\epsilon}^2,$$

whose upper bound tends to zero when $\epsilon \to 0$.

By using the Cauchy–Schwarz inequality and the multidimensional Itô isometry, we have

$$\begin{aligned}
\mathbb{E}\left[\sqrt{2\epsilon\beta^{-1}}\left|\int_0^t y_s^\epsilon\cdot\mathrm{d}W_s\right|\right] &= \sqrt{2\epsilon\beta^{-1}}\,\mathbb{E}\left[\left|\int_0^t y_s^\epsilon\cdot\mathrm{d}W_s\right|\mathbf{1}_{\left\{\left|\int_0^t y_s^\epsilon\cdot\mathrm{d}W_s\right|\ge 0\right\}}\right] \\
&\le \sqrt{2\epsilon\beta^{-1}}\,\mathbb{E}\left[\left(\int_0^t y_s^\epsilon\cdot\mathrm{d}W_s\right)^2\right]^{\frac{1}{2}}\mathbb{P}\left(\left|\int_0^t y_s^\epsilon\cdot\mathrm{d}W_s\right|\ge 0\right)^{\frac{1}{2}} \\
&= \sqrt{2\epsilon\beta^{-1}}\,\mathbb{E}\left[\int_0^t||y_s^\epsilon||_2^2\mathrm{d}s\right]^{\frac{1}{2}} \le \left(4\beta^{-1}E_0\epsilon t + 2\beta^{-2}d\epsilon^2t^2\right)^{\frac{1}{2}}.
\end{aligned}$$

Applying the Doob's martingale inequality on the second term of (3.27), we have

$$\begin{aligned}
\mathbb{P}\left(\sqrt{2\epsilon\beta^{-1}}\sup_{0\le t\le t_\epsilon}\left|\int_0^t y_s^\epsilon\cdot\mathrm{d}W_s\right| > \frac{\delta}{4}\right) &\le \frac{4}{\delta}\,\mathbb{E}\left[\sqrt{2\epsilon\beta^{-1}}\left|\int_0^{t_\epsilon} y_s^\epsilon\cdot\mathrm{d}W_s\right|\right] \\
&\le \frac{4}{\delta}\left(4\beta^{-1}E_0\epsilon t_\epsilon + 2\beta^{-2}d\epsilon^2{t_\epsilon}^2\right)^{\frac{1}{2}},
\end{aligned}$$

which tends to zero when $\epsilon \to 0$. ■

This theorem indicates that, in the short time scale t_ϵ, the energy of X_t^ϵ is close to E_0 for any $t \le t_\epsilon$ almost surely. The result validates in $[0, t_\epsilon]$. At the endpoint t_ϵ, the energy of $X_{t_\epsilon}^\epsilon$ is obviously close to E_0. Thus, we obtain the following corollary.

Corollary 20 (Law of Large Numbers on the total energy) *Let* $\epsilon t_\epsilon \ll 1$. *Then*

$$\lim_{\epsilon\to 0}\mathbb{P}\left(\left|E\left(x^\epsilon_{t_\epsilon}, y^\epsilon_{t_\epsilon}\right) - E_0\right| > \delta\right) = 0 \quad \textit{for every } \delta > 0.$$

Notice that Corollary 20 is a generalization of Theorem 17 in the non-harmonic d-dimensional case. The difference is that we investigate all coordinates in Theorem 17, but only one slow variable in Corollary 20, which is the total energy E. The total energy remain close to its initial value in the short time scale. It is a piece of partial evidence that $t_\epsilon \ll \frac{1}{\epsilon}$ is a proper short time scale.

3.3.2 Long time behavior

In the harmonic case, $X^\epsilon_{t_\epsilon}$ eventually tends to its equilibrium state X_∞ which is the unique invariant measure of the Langevin equation. In general, the same holds for the non-harmonic cases. In order to find the proper time scale in the non-harmonic case, we exam the convergence rate in detail.

The generator of the underdamped Langevin equation (3.5) is

$$\mathcal{L} = [y\cdot\nabla_x - \nabla_x V(x)\cdot\nabla_y] + \epsilon\left[-y\cdot\nabla_y + \beta^{-1}\Delta_y\right] =: \mathcal{L}_{ham} + \epsilon\mathcal{L}_{FD},$$

where $\mathcal{L}_{ham}$ is the Liouville operator of the corresponding Hamiltonian system and $\mathcal{L}_{FD}$ is the fluctuation-dissipation part. The density function $\rho(x,y,t)$ of the distribution of X^ϵ_t satisfies the Fokker–Planck equation

$$\begin{aligned}\partial_t\rho &= -y\cdot\nabla_x\rho + \nabla_x V\cdot\nabla_y\rho + \epsilon\left(\nabla_y\cdot(y\rho) + \beta^{-1}\Delta_y\rho\right),\\ \rho(x,y,0) &= \rho_0(x,y),\end{aligned}$$

where ρ_0 is the initial density. Let ρ_β denote the density of the stationary distribution. Let $h(x,y,t)$ satisfy $\rho(x,y,t) = h(x,y,t)\rho_\beta(x,y)$. It is straight forward to check that

$$\begin{aligned}\partial_t h &= -\mathcal{L}_{ham}h + \epsilon\mathcal{L}_{FD}h =: \mathcal{L}_{kin}h,\\ h(x,y,0) &= \rho_\beta^{-1}(x,y)\rho_0(x,y).\end{aligned}$$

The equation of h is named as the kinetic Fokker-Planck equation. Write $\rho(x,y,t) = f(x,-y,t)\rho_\beta(x,y)$. By direct differentiation we have

$$\begin{aligned}\partial_t f &= \mathcal{L}f,\\ f(x,y,0) &= \rho_\beta^{-1}(x,-y)\rho_0(x,-y).\end{aligned}$$

Thus f satisfies the backward Kolmogorov equation. ρ, h and f have similar time evolutions. Since $\mathcal{L}$ is not self-adjoint, the Poincaré inequality is not applicable. Therefore one cannot use the classical techniques of the convergence of the Smoluchowski equation, whose details are demonstrated in [13, Section 4.5] and [48, Section 2].

In this degenerate case, one uses hypocoercivity to control the convergence. Villani demonstrates the convergence of h in the following $\mathcal{H}^1$ norm

$$||h||^2_{\mathcal{H}^1(\rho_\beta)} \coloneqq \int_{\mathbb{R}^n\times\mathbb{R}^n} \left(|\nabla_x h(x,y)|^2 + |\nabla_y h(x,y)|^2\right) \rho_\beta(x,y)dxdy.$$

The main reason to chose this norm is to satisfy the Gronwall inequality [49, Section 3]. He provides the first assumption of convexity.

$$\left|\nabla^2 V\right| \leq c\left(1 + |\nabla V|\right),$$

which avoids the potential V to growth faster than exponential function. The second assumption he made is a modified Poincaré inequality

$$\int \left(|\nabla_x h|^2 + |\nabla_y h|^2\right) \rho_\beta dxdy \geq \kappa \left[\int h^2 \rho_\beta dxdy - \int h\rho_\beta dxdy\right]^2.$$

With above two assmptions he proves the exponential decay in the $\mathcal{H}^1$ norm [50]. In [49, Section 3] he provides the $L^2(\rho_\beta)$ version

$$\left|\left|h_t - \int h_0\rho_\beta dxdy\right|\right|_{L^2(\rho_\beta)} \leq Ce^{-\lambda t}\left|\left|h_0 - \int h_0\rho_\beta dxdy\right|\right|_{L^2(\rho_\beta)},$$

$$\text{where}\quad ||h||^2_{L^2(\rho_\beta)} \coloneqq \int_{\mathbb{R}^n\times\mathbb{R}^n} |h(x,y)|^2 \rho_\beta(x,y)dxdy.$$

We can write $h_t = \rho_t/\rho_\beta$, then

$$\left|\left|h_t - \int h_0\rho_\beta dxdy\right|\right|^2_{L^2(\rho_\beta)} = ||\rho_t - \rho_\beta||^2_{L^2\left(\rho_\beta^{-1}\right)} = \chi^2\left(\rho_t, \rho_\beta\right), \tag{3.28}$$

where the last term is χ^2-divergence. By [51, Lemma 2.7] we have

$$\mathbf{d}_{\mathrm{TV}}\left(\rho_t, \rho_\beta\right) \leq \sqrt{\chi^2\left(\rho_t, \rho_\beta\right)}. \tag{3.29}$$

Thus rather than L^2 norm, we obtain an exponential convergence of the densities in the total variation distance. However, the convergence rate λ is not analyzed precisely in those pioneering results.

There are other studies considering this exponential convergence with the different norms or in the different spaces. With the bounded condition of the Hessian of $V(x)$, Baudoin proves the exponential convergence of the densities in Wasserstein distance via the Bakry-Émery machinery [52]. In [53], CAO, LU, and WANG estimate the convergence rate precisely in a modified Banach space which includes a finite time interval, $I = (0, T)$. With a fixed Poincaré constant they estimate that the convergence rate is $\mathcal{O}(\epsilon)$, when $\epsilon \to 0$. Because of the framework of the technique, this result validates in a finite time.

Grothaus and Stilgenbauer provide a rigorous derivation of the convergence rate as a function of the damping coefficient ϵ [54]. Rather than the kinetic Fokker-Planck equation, they directly analyze the backward Kolmogorov equation. The two assumptions are similar to what Villani assumes. The difference is that the Poincaré inequality is in the L^2 norm. Their result is

$$\left|\left| f_t - \int f_0 \rho_\beta dxdy \right|\right|_{L^2(\rho_\beta)} \leq \nu_1 e^{-\nu_2 t} \left|\left| f_0 - \int f_0 \rho_\beta dxdy \right|\right|_{L^2(\rho_\beta)}, \tag{3.30}$$

where $\nu_1 \in (0, \infty)$,

$$\nu_2 = \frac{\nu_1 - 1}{\nu_1} \frac{\epsilon}{n_1 + n_2\epsilon + n_3\epsilon^2},$$

and positive constants $n_i, i = 1, \ldots, 3$ only depend on $V(x)$ and β. By (3.28) and (3.29) we can deduce the total variation version of (3.30). Therefore we have a similar result as part iii) of Theorem 16.

Corollary 21 *Assume the probability measure $\rho_\beta dxdy$ satisfies a Poincaré inequality*

$$||\nabla\phi||^2_{L^2(\rho_\beta)} \geq \Lambda \left|\left| \phi - \int_{\mathbb{R}^n \times \mathbb{R}^n} \phi \rho_\beta dxdy \right|\right|_{L^2(\rho_\beta)},$$

for $\Lambda \in (0, \infty)$ and all $\phi \in C_c^\infty(\mathbb{R}^{2n})$. Also assume that there exists a constant c such that

$$\left|\nabla^2 V(x)\right| \leq c\,(1 + |\nabla V(x)|) \text{ for all } x \in \mathbb{R}^n.$$

If $\epsilon t_\epsilon \gg 1$, then

$$\lim_{\epsilon \to 0} \mathbf{d}_{\mathrm{TV}}\left(X^\epsilon_{t_\epsilon}, X_\infty\right) = 0.$$

When ϵ tends to zero, we have $\epsilon\nu_2$ tends to infinity. Then the right hand side of the inequality (3.30) tends to zero, which is the result we want to prove. As expected, the solution fully merges with its stationary measure in the long time scale.

3.4 Transient behavior of the one-dimensional Langevin equation

We have shown that, in the short time scale ($t_\epsilon \ll \epsilon^{-1}$), X_t^ϵ is captured by the initial Hamiltonian orbit. In the long time scale ($t_\epsilon \gg \epsilon^{-1}$), X_t^ϵ eventually tends to its stationary measure. Then the solution undergoes a transition from a deterministic curve to the stationary measure during the critical time scale $t_\epsilon = \mathcal{O}(\epsilon^{-1})$. In this section, we will demonstrate how the solution evolves from the Hamiltonian orbit to its stationary measure in the critical time scale $\left(t_\epsilon = \frac{\theta}{\epsilon}\right)$.

In the one-dimensional harmonic case, for each fixed θ the mean of the total energy decreases $1 - e^{-\theta}$ times from the initial total energy. Meanwhile, the variance X_t^ϵ increases $1 - e^{-\theta}$ times the variance of its stationary measure. When θ tends to zero, both factors are equal to zero, which means the initial Hamiltonian orbit precisely captures the solution. When θ tends to infinity, both factors are equal to one, which means the solution reaches its stationary measure. Therefore, the variation of θ demonstrates the transient behavior. However, because of the rotation map (3.14) we cannot find a proper system to describe the behavior with respect to θ in the critical time scale.

In the one-dimensional non-harmonic case, from the simulation, we know the angle variable becomes even more sensitive than the harmonic case. We cannot describe the transient behavior in the full dimensions. Fortunately, the Langevin equation is a slow-fast system. In the critical time scale, the energy as a slow variable plays a more important role than the angle variable. We will analyze the transition of energy in detail.

3.4.1 Construct the averaging system of the energy

We first construct the slow-fast structure of the Langevin equation. By changing variables into the position and the energy space, (x, E). The energy is defined as

$$E(x, y) = \frac{1}{2}y^2 + V(x).$$

Applying Itô's formula, we find

$$\mathrm{d}E\left(x_t^\epsilon, y_t^\epsilon\right) = \epsilon\left(\beta^{-1} - y_t^{\epsilon 2}\right)\mathrm{d}t + \sqrt{2\epsilon\beta^{-1}}y_t^\epsilon \mathrm{d}W_t.$$

Without loss of generality, we set $\beta = 2$. We drop the superscript ϵ to simplify the notation. Since we focus on the distribution, the above equation can be rewritten as

$$\begin{aligned}
\mathrm{d}E_t &= \epsilon\left(\frac{1}{2} - y_t^2\right)\mathrm{d}t + \sqrt{\epsilon y_t^2}\mathrm{d}W_t, \\
&= \epsilon\left[\frac{1}{2} - 2\left(E_t - V(x_t)\right)\right]\mathrm{d}t + \sqrt{2\epsilon\left(E_t - V(x_t)\right)}\mathrm{d}W_t, \\
&=: \epsilon b\left(x_t, E_t\right)\mathrm{d}t + \sqrt{\epsilon}\sigma\left(x_t, E_t\right)\mathrm{d}W_t.
\end{aligned}$$

For the position variable

$$\mathrm{d}x_t = y_t\mathrm{d}t = \begin{cases} \sqrt{2\left(E_t - V(x_t)\right)}\mathrm{d}t, & y_t \geq 0 \\ -\sqrt{2\left(E_t - V(x_t)\right)}\mathrm{d}t, & y_t < 0 \end{cases} =: f\left(x_t, E_t\right)\mathrm{d}t.$$

Compared with x, E is the relatively slow variable. In order to apply the averaging principle to our SDEs [55], we need to change the time scale to the critical time scale. Let $E_t^c = E_{\frac{t}{\epsilon}}$ and $x_t^c = x_{\frac{t}{\epsilon}}$. Then the SDEs are

$$\begin{aligned}
\mathrm{d}E_t^c &= b\left(x_t^c, E_t^c\right)\mathrm{d}t + \sigma\left(x_t^c, E_t^c\right)\mathrm{d}W_t, \\
\mathrm{d}x_t^c &= \frac{1}{\epsilon}f\left(x_t^c, E_t^c\right)\mathrm{d}t.
\end{aligned} \tag{3.31}$$

We drop the superscript and subscript. The generator of (3.31) is

$$\mathcal{L} = \frac{1}{\epsilon}f(x,E)\frac{\partial}{\partial x} + b(x,E)\frac{\partial}{\partial E} + \frac{1}{2}\sigma^2(x,E)\frac{\partial^2}{\partial E^2} =: \frac{1}{\epsilon}\mathcal{L}_0 + \mathcal{L}_1.$$

Let u^ϵ be the solution of the backward Kolmogorov equation

$$\frac{\partial u^\epsilon}{\partial t} = \left(\frac{1}{\epsilon}\mathcal{L}_0 + \mathcal{L}_1\right)u^\epsilon.$$

We seek a series expansion of u^ϵ

$$u^\epsilon = u_0 + \epsilon u_1 + \mathcal{O}\left(\epsilon^2\right).$$

Then in different scales of ϵ we have

$$\mathcal{O}\left(\epsilon^{-1}\right) \quad \mathcal{L}_0 u_0 = 0, \tag{3.32}$$

$$\mathcal{O}\left(1\right) \quad \mathcal{L}_0 u_1 = -\mathcal{L}_1 u_0 + \frac{\partial u_0}{\partial t}. \tag{3.33}$$

(3.32) implies that u_0 is in the null space of $\mathcal{L}_0$. Thus u_0 is independent of x, which means $u_0 = u_0(E,t)$. Let $\rho^\infty(x;E)$, as a density function, be the element in the null space of $\mathcal{L}_0^*$. Fixing E, we obtain

$$\mathcal{L}_0^*\rho^\infty(x;E) = 0 \tag{3.34}$$

$$\text{and} \quad \oint_{C(E)} \rho^\infty(x;E)\mathrm{d}x = 1, \tag{3.35}$$

where the integral is taken over one complete period on

$$C(E) = \left\{ (x,y) \;\middle|\; \frac{1}{2}y^2 + V(x) = E \right\}.$$

The Fredholm alternative of (3.33) indicates

$$-\mathcal{L}_1 u_0 + \frac{\partial u_0}{\partial t} \perp \mathrm{Null}\{\mathcal{L}_0^*\},$$

which implies

$$\oint_{C(E)} \rho^\infty(x;E) \left(\frac{\partial u_0}{\partial t} - b(x,E)\frac{\partial u_0}{\partial E} - \frac{1}{2}\sigma^2(x,E)\frac{\partial^2 u_0}{\partial E^2} \right) \mathrm{d}x = 0.$$

Since ρ^∞ is a density function, the equation can be simplified as

$$\frac{\partial u_0}{\partial t} = \oint_{C(E)} b(x,E)\rho^\infty(x;E)\mathrm{d}x \frac{\partial u_0}{\partial E} + \frac{1}{2}\oint_{C(E)} \sigma^2(x,E)\rho^\infty(x;E)\mathrm{d}x \frac{\partial^2 u_0}{\partial E^2},$$

which is a backward Kolmogorov equation of a one-dimenstional averaged equation

$$\mathrm{d}\overline{E} = \bar{b}(\overline{E})\mathrm{d}t + \bar{\sigma}(\overline{E})\mathrm{d}W_t, \tag{3.36}$$

where

$$\bar{b}(E) = \oint_{C(E)} b(x,E)\rho^\infty(x;E)\mathrm{d}x,$$

$$\bar{\sigma}(E) = \left[\oint_{C(E)} \sigma^2(x,E)\rho^\infty(x;E)\mathrm{d}x \right]^{\frac{1}{2}}.$$

In order to obtain $\bar{b}$ and $\bar{\sigma}$ precisely, we need to find the density of the steady state of the fast variable, which is $\rho^\infty(x;E)$. According to (3.34), the steady state satisfies

$$\mathcal{L}_0^*\rho^\infty(x;E) = -\frac{\partial}{\partial x}\left(f(x,E)\rho^\infty(x;E)\right) = 0.$$

Thus there exists a function $h(E)$, which only depends on E, such that

$$\rho^{\infty}(x;E) = \frac{h(E)}{f(x,E)} = \frac{h(E)}{y}.$$

Since $\rho^{\infty}(x;E)$ is a density function, applying (3.35) we have

$$\oint_{C(E)} \rho^{\infty}(x;E)\mathrm{d}x = \oint_{C(E)} \frac{h(E)}{y}\mathrm{d}x = 1.$$

Therefore,

$$h(E) = \frac{1}{\langle y^{-1}\rangle}, \text{ where } \langle y\rangle = \oint_{C(E)} y\mathrm{d}x.$$

We plug ρ^{∞} back into (3.36) to obtain

$$\bar{b}(E) = \frac{1}{2} - \frac{\langle y\rangle}{\langle y^{-1}\rangle} \text{ and } \bar{\sigma}(E) = \sqrt{\frac{\langle y\rangle}{\langle y^{-1}\rangle}}.$$

We introduce the classical action variable I and the frequency ω to simplify the expression of the averaged equation.

$$\begin{aligned} I(E) &= \frac{1}{2\pi}\oint_{C(E)} y\mathrm{d}x, \\ \omega(E) &= 2\pi\left(\oint_{C(E)} \mathrm{d}t\right)^{-1} = 2\pi\left(\oint_{C(E)} \frac{\mathrm{d}x}{y}\right)^{-1}. \end{aligned}$$

Therefore the averaged equation can be rewritten as

$$\mathrm{d}\overline{E} = \left(\frac{1}{2} - I(\overline{E})\omega(\overline{E})\right)\mathrm{d}t + \sqrt{I(\overline{E})\omega(\overline{E})}\mathrm{d}W_t. \tag{3.37}$$

Changing the time scale back to the original one, the averaged equation reads

$$\mathrm{d}\overline{E}^{\epsilon}_t = \epsilon\left(\frac{1}{2} - I(\overline{E}^{\epsilon}_t)\omega(\overline{E}^{\epsilon}_t)\right)\mathrm{d}t + \sqrt{\epsilon I(\overline{E}^{\epsilon}_t)\omega(\overline{E}^{\epsilon}_t)}\mathrm{d}W_t.$$

Applying the averaging principle [57], let E^{ϵ}_t be the energy at $(x^{\epsilon}_t, y^{\epsilon}_t)$, then

Theorem 22 (Averaging Principle of Total Energy)

i) The marginal distribution of E^c_t converges to the distribution of $\overline{E}^c_t$ as $\epsilon \to 0$ for any $t \in [0,T]$.

ii) The marginal distribution of E^{ϵ}_t converges to the distribution of $\overline{E}^{\epsilon}_t$ for any $t \in \left[0, \frac{T}{\epsilon}\right]$.

In Section 3.3 we knew that, for the d-dimensional non-harmonic case, E_t^ϵ tends to E_0 in the short time scale. In the long time scale, E_t^ϵ tends to E_∞ which is the energy of the stationary measure. We specifically consider that $d = 1$ in this section. There exists $\theta \in \mathbb{R}^+$ such that $\epsilon t_\epsilon = \theta$ in the critical time scale. By part ii) of Theorem 22, E_t^ϵ weakly converges, as $\epsilon \to 0$, to $\overline{E}_t^\epsilon$ when $t = \frac{\theta}{\epsilon}$. Let $E(\theta)$ and $\overline{E}(\theta)$ be the above two limits respectively. Then $\overline{E}(\theta)$ is a proper approximation of $E(\theta)$ which is our target system. Meanwhile, by definition

$$\lim_{\epsilon \to 0} \overline{E}^\epsilon_{\frac{\theta}{\epsilon}} = \overline{E}(\theta).$$

Then $\overline{E}(\theta)$ is the solution of (3.37) at time θ. Therefore, in the critical time scale, the averaged equation (3.37) demonstrates the evolution from the energy of the initial Hamiltonian to the energy of the stationary measure.

3.4.2 The averaged system of the harmonic oscillator with external noise

In this section, we consider the Harmonic Potential, $V(x) = \frac{1}{2}kx^2$. By Green's theorem, we have

$$I(E) = \frac{1}{2\pi} \oint_{C(E)} y \mathrm{d}x = \frac{1}{2\pi} \iint_{D(E)} \mathrm{d}x\mathrm{d}y = \frac{E}{\sqrt{k}},$$

where $D(E)$ is the region in $\mathbb{R}^2$ bounded by $C(E)$. By directly differentiation, we have

$$\begin{aligned} \frac{\mathrm{d}I}{\mathrm{d}E} &= \frac{1}{2\pi} \frac{\mathrm{d}}{\mathrm{d}E} \oint_{C(E)} \sqrt{2E - kx^2} \mathrm{d}x = \frac{1}{2\pi} \oint_{C(E)} \frac{\mathrm{d}x}{\sqrt{2E - kx^2}} \\ &= \frac{1}{2\pi} \oint_{C(E)} \frac{\mathrm{d}x}{\dot{x}} = \frac{1}{2\pi} \oint_{C(E)} \mathrm{d}t = \frac{T}{2\pi} = \frac{1}{\omega(E)}, \\ \text{then} \quad I(E)\omega(E) &= \frac{E}{\sqrt{k}} \left(\frac{\mathrm{d}I}{\mathrm{d}E} \right)^{-1} = E. \end{aligned}$$

Therefore, in the harmonic case, the averaged equation (3.37) is

$$\mathrm{d}\overline{E} = \left(\frac{1}{2} - \overline{E} \right) \mathrm{d}t + \sqrt{\overline{E}} \mathrm{d}W_t; \ \overline{E}_0 = E_0.$$

which is a Cox–Ingersoll–Ross (CIR) model.

[58] provided the corresponding moment generating function which is

$$\mathbb{E}\left(e^{u\overline{E}_\theta} \right) = \left[1 - \frac{1 - e^{-\theta}}{2} u \right]^{-1} \times \exp\left\{ \frac{E_0 e^{-\theta} u}{1 - \frac{1 - e^{-\theta}}{2} u} \right\}. \tag{3.38}$$

When $\theta = 0$, we have $\mathbb{E}\left(e^{u\overline{E}_0}\right) = e^{uE_0}$. This is to say that the averaged process starts from the energy of the initial Hamiltonian as we construct. When θ tends to infinity, we have

$$\lim_{\theta\to\infty} \mathbb{E}\left(e^{u\overline{E}_\theta}\right) = \frac{2}{2-u}.$$

Thus the stationary distribution of $\overline{E}$ is an exponential distribution whose rate equals 2. We will discuss this limit when comparing the averaged process with the original process in the critical time scale.

In the harmonic case, we know the explicit means and variances of the Gaussian process X_t^ϵ. In Section 3.1.2 we compute the limiting state in the critical time scale. When $\epsilon t_\epsilon = \theta$, as $\epsilon \to 0$, the means and variances of $X_{t_\epsilon}^\epsilon$ respectively tend to

$$m_\theta = e^{-\frac{\theta}{2}} \begin{pmatrix} x_0 \cos\left(\sqrt{k{t_\epsilon}^2 - \frac{\theta^2}{4}}\right) + \frac{1}{\sqrt{k}} y_0 \sin\left(\sqrt{k{t_\epsilon}^2 - \frac{\theta^2}{4}}\right) \\ y_0 \cos\left(\sqrt{k{t_\epsilon}^2 - \frac{\theta^2}{4}}\right) - \sqrt{k} x_0 \sin\left(\sqrt{k{t_\epsilon}^2 - \frac{\theta^2}{4}}\right) \end{pmatrix} =: \begin{pmatrix} m_1 \\ m_2 \end{pmatrix},$$

$$\Sigma_\theta = \begin{pmatrix} \frac{1}{2k}\left(1 - e^{-\theta}\right), & 0 \\ 0, & \frac{1}{2}\left(1 - e^{-\theta}\right) \end{pmatrix} =: \begin{pmatrix} \sigma_1^2, & 0 \\ 0, & \sigma_2^2 \end{pmatrix}.$$

Then the moment generating function of the energy of $X_{t_\epsilon}^\epsilon$ is

$$\begin{aligned} \mathbb{E}\left(e^{uE_\theta}\right) &= \iint_{\mathbb{R}^2} e^{u\left(\frac{1}{2}y^2 + \frac{1}{2}kx^2\right)} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(x-m_1)^2}{\sigma_1^2} + \frac{(x-m_2)^2}{\sigma_2^2}\right]} \mathrm{d}x\mathrm{d}y \\ &= \frac{1}{\sqrt{1 - k\sigma_1^2 u}} e^{\frac{km_1^2\sigma_1^2 u}{2\sigma_1^2\left(1 - k\sigma_1^2 u\right)}} \frac{1}{\sqrt{1 - \sigma_2^2 u}} e^{\frac{m_2^2\sigma_2^2 u}{2\sigma_2^2\left(1 - \sigma_2^2 u\right)}} \\ &= \left[1 - \frac{1 - e^{-\theta}}{2} u\right]^{-1} \times \exp\left\{\frac{\left[\frac{k}{2}m_1^2 + \frac{1}{2}m_2^2\right] u}{1 - \frac{1-e^{-\theta}}{2} u}\right\} \\ &= \left[1 - \frac{1 - e^{-\theta}}{2} u\right]^{-1} \times \exp\left\{\frac{E_0 e^{-\theta} u}{1 - \frac{1-e^{-\theta}}{2} u}\right\}, \end{aligned}$$

which is exactly the moment generating function of $\overline{E}_\theta$ in (3.38). This result implies $E\left(X_{t_\epsilon}^\epsilon\right)$ converges to $\overline{E}_\theta$ in distribution with fixed θ, which is given by the averaging principle.

With a similar computation, the moment generating function of the energy of the stationary measure is

$$\mathbb{E}\left(e^{uE_\infty}\right) = \frac{2}{2-u} = \lim_{\theta\to\infty} \mathbb{E}\left(e^{u\overline{E}_\theta}\right).$$

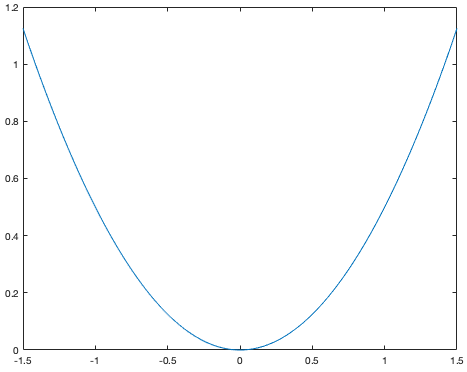
The stationary distribution of the averaged equation (3.37) happens to be the distribution of the energy of X_∞. Therefore, at least in the harmonic case, the averaging principle can be generated to include the stationary distribution. Moreover, the averaged equation demonstrates the entire transition of the energy. It starts from the energy of the initial Hamiltonian and converges to the energy of the stationary state of X_t^ϵ.

3.4.3 The averaged system of the non-harmonic oscillator with external noise

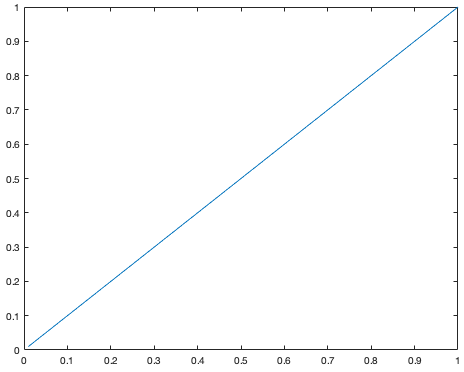
In the averaged equation (3.37), the term, $F(E) := I(E)\omega(E)$, plays a critical role. In the harmonic case $F(E) = E$. Since it is a linear function, (3.37) becomes a CIR model whose density is well-known explicitly for each θ. In the non-harmonic case, we try to find a suitable type of potential such that $F(E)$ is still a linear function.

Since it is impossible to have the explicit form of $F(E)$ with general potentials, we start to simulate $F(E)$ among the following different types of potentials, and to collect the general properties from simulations. We select the code of the double-well potential as an example in Appendix E.7.

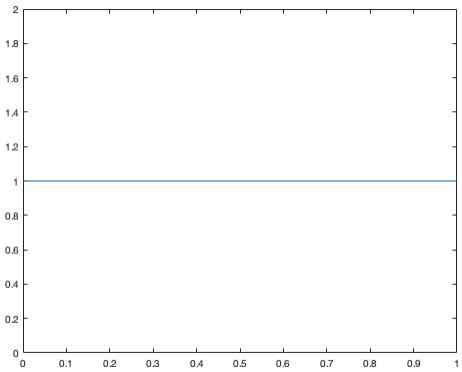
i) Harmonic potential: $V(x) = \frac{1}{2}x^2$



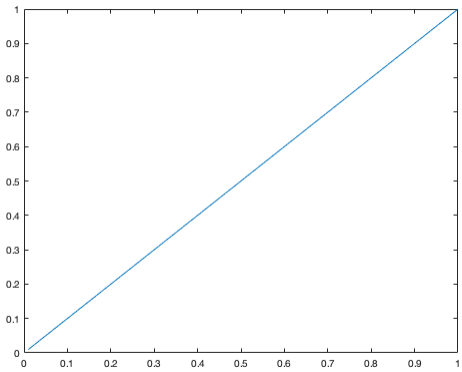
(a) $V(x)$



(b) $I(E)$

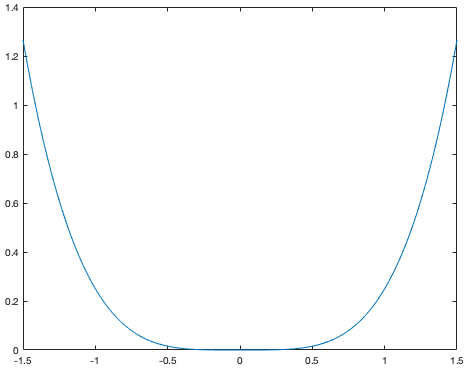


(c) $\omega(E)$

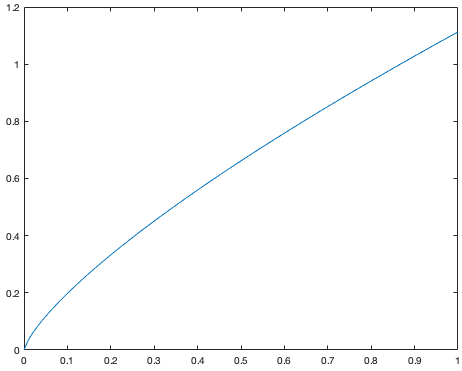


(d) $F(E)$

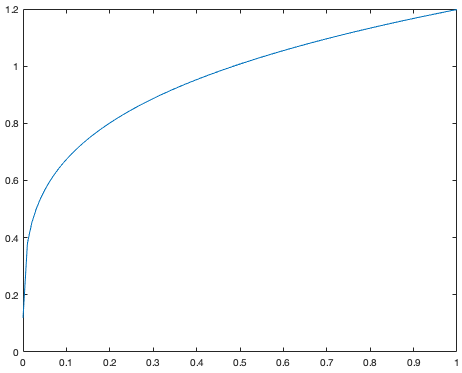
ii) Single-well potential: $V(x) = \frac{1}{4}x^4$



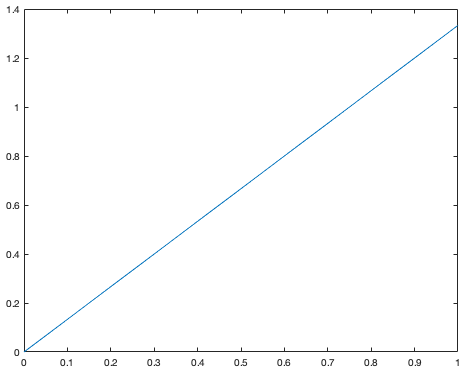
(a) $V(x)$



(b) $I(E)$

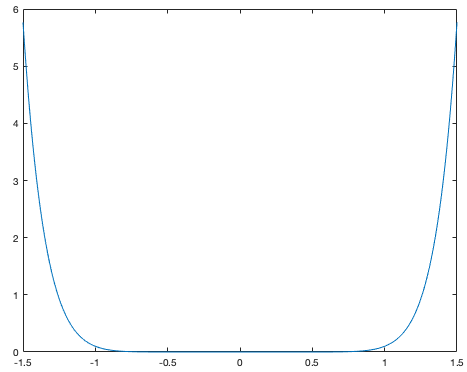


(c) $\omega(E)$

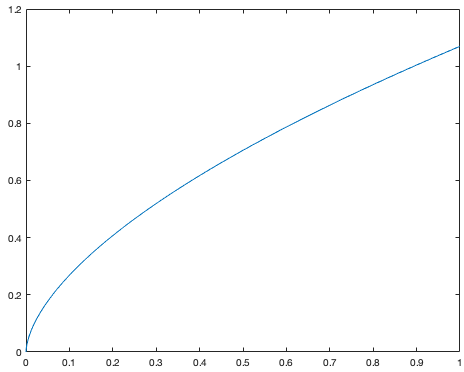


(d) $F(E)$

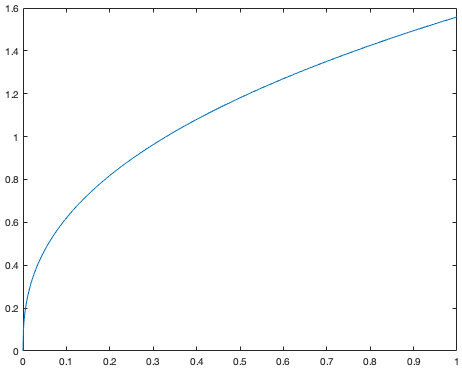
iii) Single-well potential: $V(x) = \frac{1}{10}x^{10}$



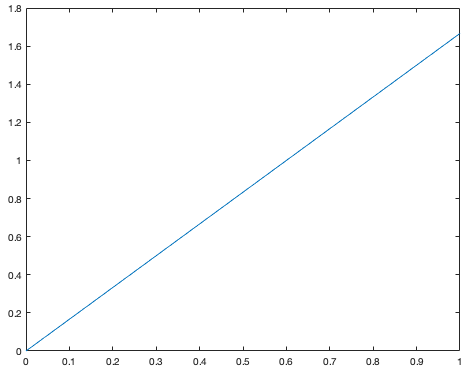
(a) $V(x)$



(b) $I(E)$

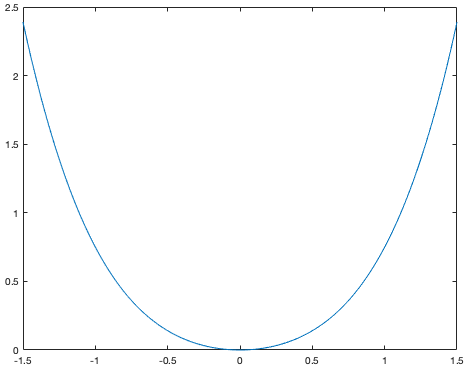


(c) $\omega(E)$

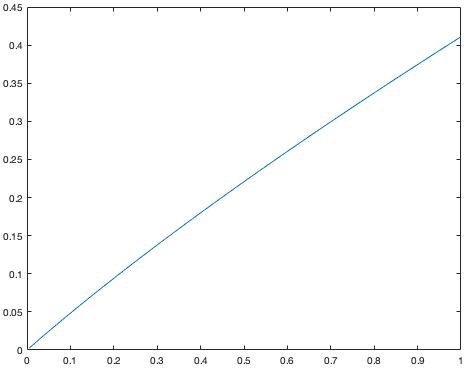


(d) $F(E)$

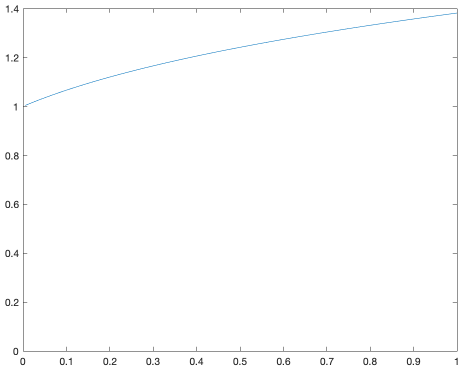
iv) Single-well potential: $V(x) = \frac{1}{2}x^2 + \frac{1}{4}x^4$



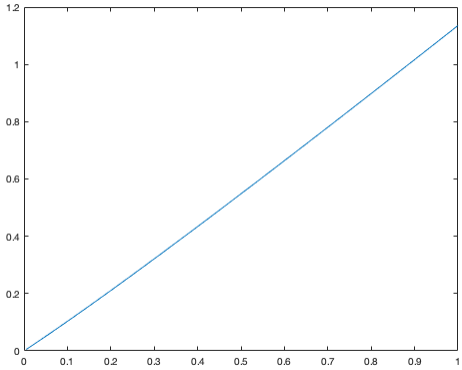
(a) $V(x)$



(b) $I(E)$

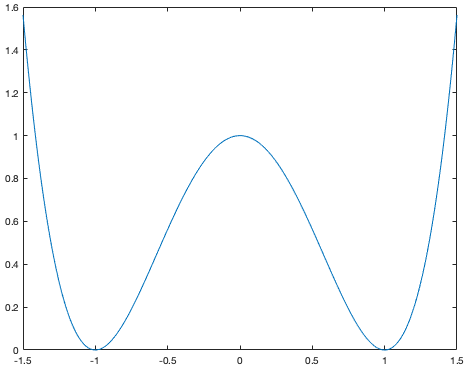


(c) $\omega(E)$

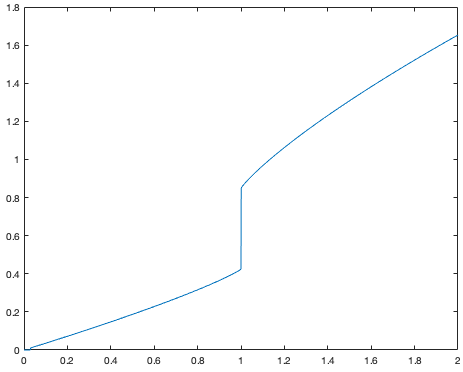


(d) $F(E)$

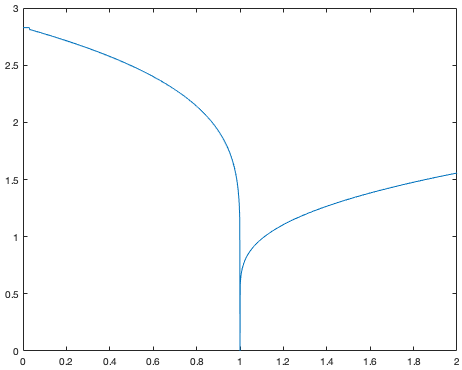
v) Double-well potential: $V(x) = 1 - 2x^2 + x^4$



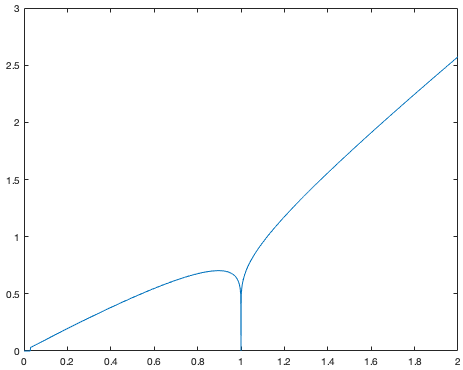
(a) $V(x)$



(b) $I(E)$



(c) $\omega(E)$



(d) $F(E)$

The function $F(E)$ is almost linear with different slopes in the first four cases, which are single-well potentials. We use Maple to compute $F(E)$ among those cases. In the first three cases, $F(E)$ are linear functions. But the symbolic computation does not provide a linear result when $V(x) = \frac{1}{2}x^2 + \frac{1}{4}x^4$. Therefore, we consider $V(x) = \frac{1}{2n}x^{2n}$ first.

Proposition 23 *If $V(x) = \frac{1}{2n}x^{2n}$, where $n \in \mathbb{Z}^+$. Then $F(E) = \frac{2n}{n+1}E$. Furthermore, the averaged equation is a CIR model.*

Proof. The intersection points are $x^+ = (2nE)^{\frac{1}{2n}}$ and $x^- = -(2nE)^{\frac{1}{2n}}$. Since $V(x)$ is an even function we have

$$F(E) = \frac{\int_0^{x^+} (2E - 2V(x))^{\frac{1}{2}} \, \mathrm{d}x}{\int_0^{x^+} (2E - 2V(x))^{-\frac{1}{2}} \, \mathrm{d}x} = 2 \, \frac{\int_0^{x^+} (E - V(x))^{\frac{1}{2}} \, \mathrm{d}x}{\int_0^{x^+} (E - V(x))^{-\frac{1}{2}} \, \mathrm{d}x} =: 2 \, \frac{p(E)}{q(E)}.$$

By changing of variables and integration by part, we have

$$\begin{aligned} p(E) &= \int_0^{(2nE)^{\frac{1}{2n}}} \left(E - \frac{1}{2n}x^{2n}\right)^{\frac{1}{2}} \mathrm{d}x = 2^{\frac{1}{2n}} n^{\frac{1}{2n}-1} \int_0^{\sqrt{E}} t^2 \left(E - t^2\right)^{\frac{1}{2n}-1} \mathrm{d}t \\ &= (2n)^{\frac{1}{2n}} \int_0^{\sqrt{E}} \left(E - t^2\right)^{\frac{1}{2n}} \mathrm{d}t = (2n)^{\frac{1}{2n}} \, E^{\frac{1}{2n}+\frac{1}{2}} \int_0^1 (1 - s^2)^{\frac{1}{2n}} \mathrm{d}s. \end{aligned}$$

By the similarly changing of variables, we have

$$\begin{aligned} q(E) &= \int_0^{(2nE)^{\frac{1}{2n}}} \left(E - \frac{1}{2n}x^{2n}\right)^{-\frac{1}{2}} \mathrm{d}x = 2^{\frac{1}{2n}} n^{\frac{1}{2n}-1} \int_0^{\sqrt{E}} \left(E - t^2\right)^{\frac{1}{2n}-1} \mathrm{d}t \\ &= 2^{\frac{1}{2n}} n^{\frac{1}{2n}-1} E^{\frac{1}{2n}-\frac{1}{2}} \int_0^1 (1 - s^2)^{\frac{1}{2n}-1} \mathrm{d}s. \end{aligned}$$

Therefore,

$$\begin{aligned} F(E) &= 2n \frac{\int_0^1 (1 - s^2)^{\frac{1}{2n}} \mathrm{d}s}{\int_0^1 (1 - s^2)^{\frac{1}{2n}-1} \mathrm{d}s} E \\ &= 2n \frac{\Gamma\left(\frac{1}{2} + \frac{1}{2n}\right) \Gamma\left(1 + \frac{1}{2n}\right)}{\Gamma\left(\frac{1}{2n}\right) \Gamma\left(\frac{3}{2} + \frac{1}{2n}\right)} E = \frac{2n}{n+1} E. \end{aligned}$$

■

For this particular potential, we trace the solution E_t and derive the following proposition.

Proposition 24 *Assume* $V(x) = \frac{1}{2n}x^{2n}$, *where* $n \in \mathbb{Z}^+$, *and* $E_0 > 0$. *If* $n = 1$, *the process of* E_t *is a strictly positive process. If* $n > 1$, *the boundary* 0 *is accessible and instantaneously reflecting.*

Proof. Since the solution E_t is a CIR process, the behavior of E_t can be estimated by the *Feller condition* (see [59] for more details). We need to modify the equation back to the classical form

$$\mathrm{d}r_t = k\left(\theta - r_t\right)\mathrm{d}t + \sigma\sqrt{r_t}\mathrm{d}W_t,$$

and compare $2k\theta$ to σ^2.

If $n = 1$, we obtain $F(E) = E$. Then $k\theta = \frac{1}{2}$, $k = 1$, and $\sigma = 1$. Therefore,

$$2k\theta = 1 \leq 1 = \sigma^2,$$

i.e., the Feller condition holds. Thus, the solution stays strictly positive for any positive initial value.

If $n > 1$, we obtain $F(E) = \frac{2n}{n+1}E$. Then $k\theta = \frac{1}{2}$, $k = \frac{2n}{n+1}$, and $\sigma = \sqrt{\frac{2n}{n+1}}$. Therefore,

$$0 \leq 2k\theta = 1 < \frac{2n}{n+1} = \sigma^2,$$

i.e., the Feller condition does not hold. Thus, the solution can hit 0. Since $0 \leq 2k\theta$, the origin strongly reflecting. ■

For more general potential $V(x)$, since $F(E)$ is a nonlinear function, the solution of the averaged equation (3.37) is not a CIR process. The corresponding moment generating function is hard to obtain. Using the Laplace transform, Feller presents the positivity preserving of the CIR process ($F(E)$ is a linear function) [60]. The following theorem provides the non-negativity preserving in the general potential case.

Proposition 25 *The solution of the averaged equation* (3.37) *is non-negativity preserving.*

Proof. The averaging equation can be rewritten as

$$\mathrm{d}E_t = \left(\frac{1}{2} - F(E_t)\right)\mathrm{d}t + \sqrt{F(E_t)}\mathrm{d}W_t. \tag{3.39}$$

Since $\frac{1}{2} - F(E) > -F(E)$, we make a comparison between (3.39) and

$$\mathrm{d}\widetilde{E_t} = -F\left(\widetilde{E_t}\right)\mathrm{d}t + \sqrt{F\left(\widetilde{E_t}\right)}\mathrm{d}W_t, \tag{3.40}$$

with the same initial condition. Because of the comparison principle, $E_t \leq \widetilde{E_t}$ a.e.. Suppose that 0 is an absorbing state for $\widetilde{E_t}$. For any $E_0 = \widetilde{E_0} \geq 0$, we have $\widetilde{E_t} \geq 0$ a.e.. Therefore, $E_t \leq 0$ a.e..

For the absorbing state, we need to prove $F(0) = 0$. Recall that

$$F(E) = I(E)\omega(E) = \frac{\oint_{C(E)} y\mathrm{d}x}{\oint_{C(E)} y^{-1}\mathrm{d}x},$$

which is a continuous function. When $E \to 0$, we treat $V(x)$ as a single-well potential locally. Therefore,

$$I(E) = \frac{1}{\pi}\int_{x_-}^{x_+} (2E - 2V(s))^{\frac{1}{2}}\,\mathrm{d}s$$

where x_+ and x_- are the x-coordinates of the intersection points with $V(x)$ and E, and they are in the positive and negative domains separately. Since the integrand is bounded and limits of integral tend to 0 as $E \to 0$, we obtain

$$\lim_{E\to 0} I(E) = 0.$$

Since $V(0) = V'(0) = 0$, $V(x)$ can be approximated by $\frac{V''(0)}{2}x^2$ when x is small enough. Therefore,

$$\begin{aligned}\omega(E) &= \frac{\pi}{\int_{x_-}^{x_+} (2E - V(s))^{-\frac{1}{2}}\,\mathrm{d}s} \\ &\approx \frac{\pi}{2}\frac{1}{\int_0^{\sqrt{\frac{2E}{V''(0)}}} (2E - V''(0)s^2)^{-\frac{1}{2}}\,\mathrm{d}s} = \sqrt{V''(0)}.\end{aligned}$$

Consequently, we obtain

$$\lim_{E\to 0}\omega(E) = \sqrt{V''(0)}, \text{ and } F(0) = \lim_{E\to 0} F(E) = 0.$$

■

We start to simulate the averaged equation and summarize the general properties. The corresponding Fokker–Planck equation is

$$\rho_\theta = \frac{\partial^2}{\partial E^2}\left[\frac{1}{2}F(E)\rho\right] - \frac{\partial}{\partial E}\left[\left(\frac{1}{2} - F(E)\right)\rho\right].$$

Directly applying the Euler-Maruyama scheme to the averaged equation, we have

$$E_{n+1} = E_n + \left[\frac{1}{2} - F\left(E_n\right)\right]\Delta t + \sqrt{F\left(E_n\right)}\Delta W_n.$$

Because ΔW_n is a Gaussian random variable, the iteration enters the domain such that $F\left(E_n\right)$ is negative with non-zero probability. Then square root of $F\left(E_n\right)$ is not well-defined. The algorithms have been analyzed for the classical CIR process during the last few years. The full truncation technique of the simulation of a CIR model is proposed in [59], and [61]. It overcomes the square root's problem by replacing the negative value with zero. This idea will be applied in our simulations.

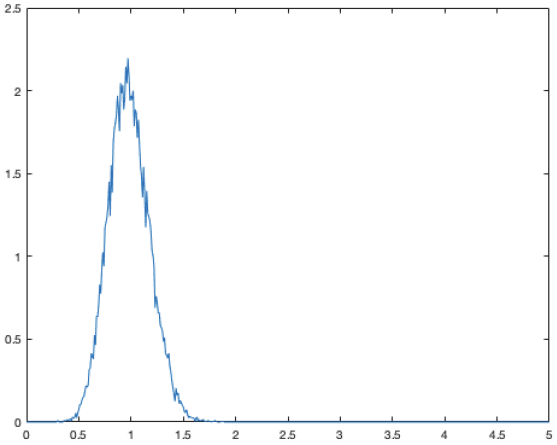
The logarithmic Euler-Maruyama is another ingenious scheme to simulate the strictly positive process [62]. It generates a new variable by logarithmic transformation. Then the domains of variables change from $\mathbb{R}^+$ to $\mathbb{R}$. We known that the Feller condition is rarely satisfied when $V(x) = \frac{1}{2n}x^{2n}$. Then the process hits the origin with positive probability when $n > 1$. Simultaneously, the new variable tends to negative infinity. Therefore, it is hard to apply the logarithmic Euler-Maruyama in the averaged equation.

We introduce the generalized full truncation scheme as follows:

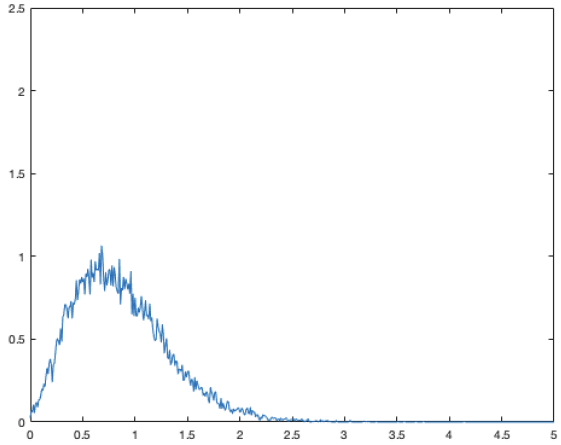
$$E_{n+1} = E_n + \left[\frac{1}{2} - F^+\left(E_n\right)\right]\Delta t + \sqrt{F^+\left(E_n\right)}\Delta W_n,$$

where $F^+ = \max(F, 0)$. For each potential energy mentioned above, we can simulate the process directly. Meanwhile, we collect the properties of their means and variances. We continuously select the code of the double-well potential as an example in Appendix E.8.

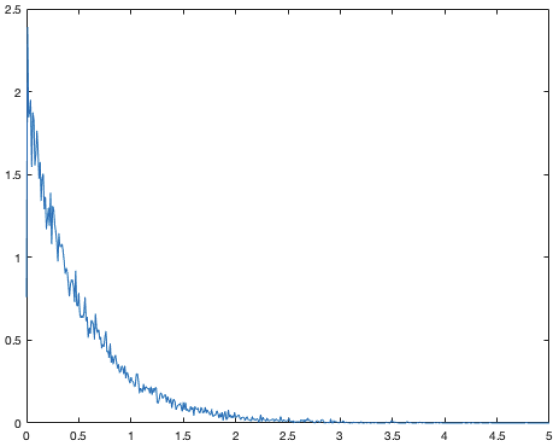
i) Harmonic potential: $V(x) = \frac{1}{2}x^2$



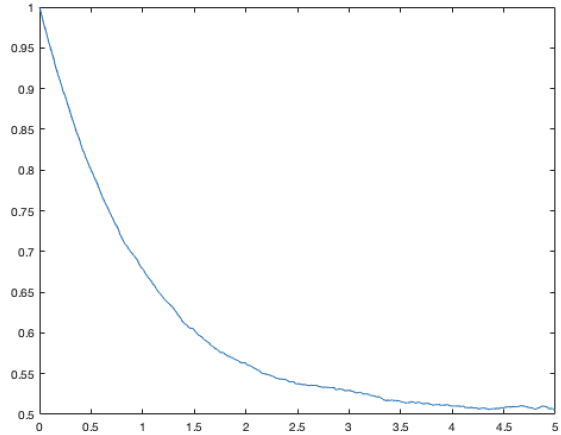
(a) $t = 0.04$



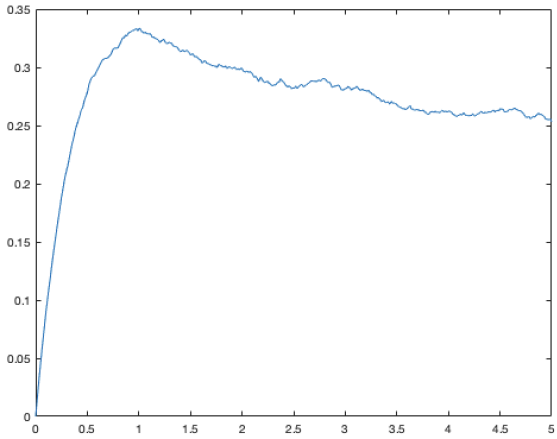
(b) $t = 0.3$



(c) $t = 5$

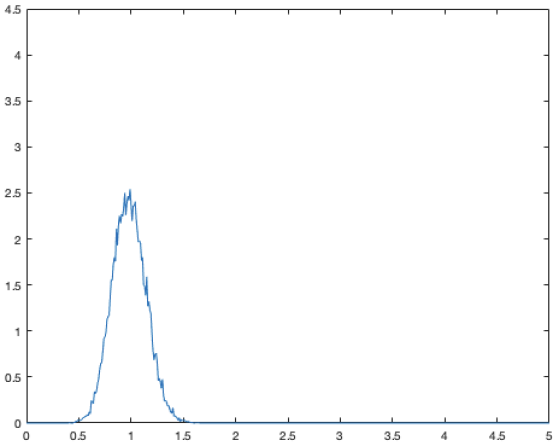


(d) Mean

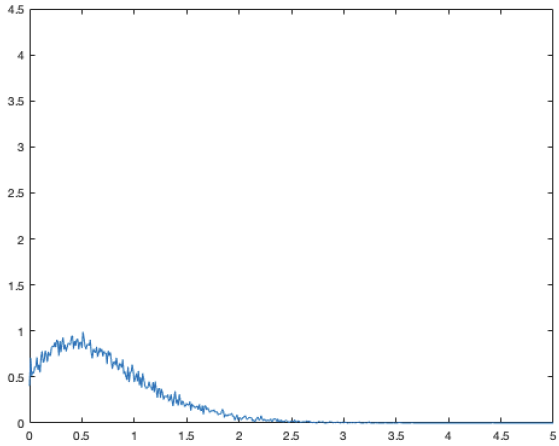


(e) Variance

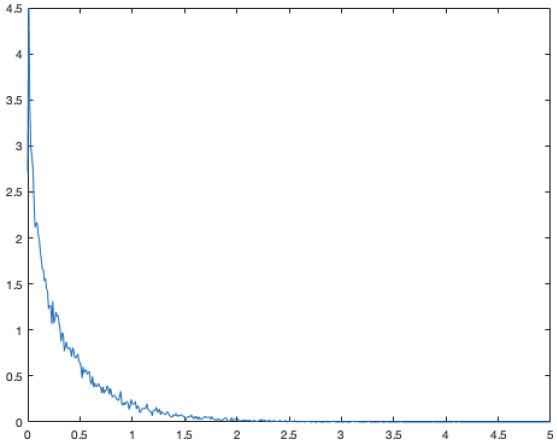
ii) Single-well potential: $V(x) = \frac{1}{4}x^4$



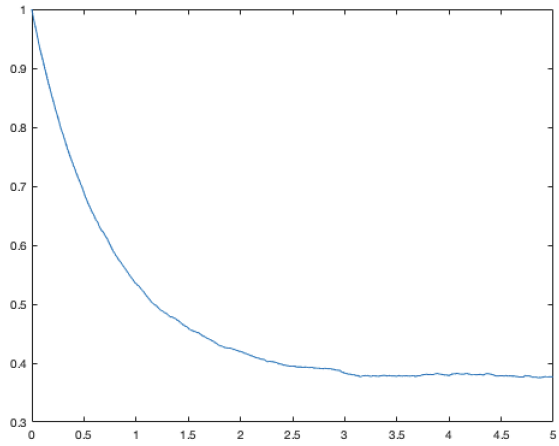
(a) $t = 0.02$



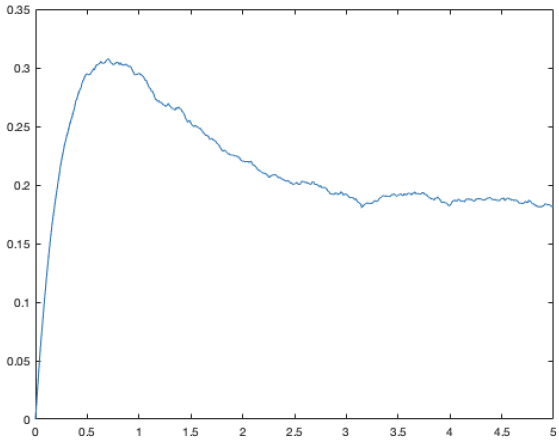
(b) $t = 0.4$



(c) $t = 5$

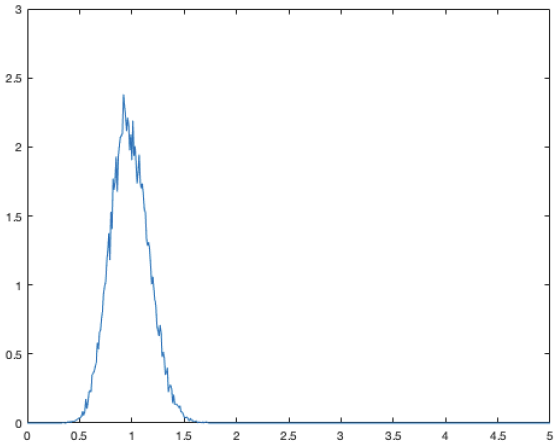


(d) Mean

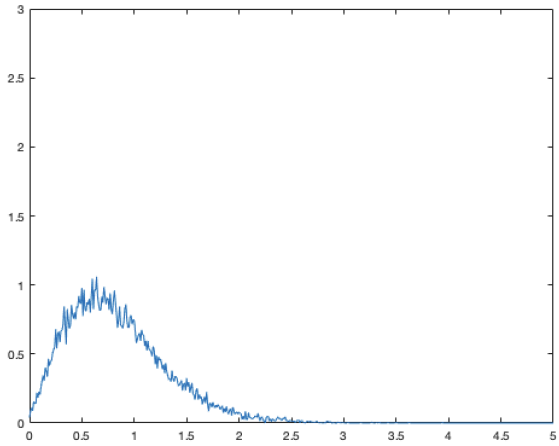


(e) Variance

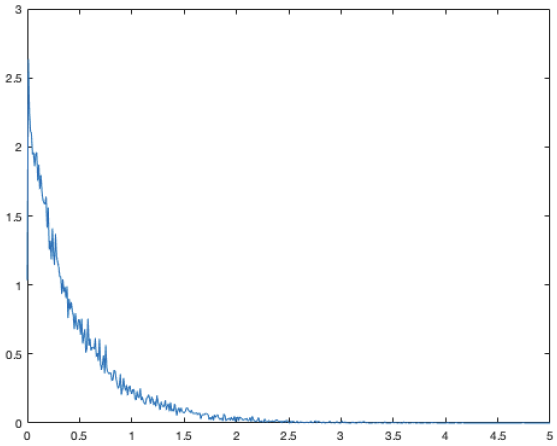
iii) Single-well potential: $V(x) = \frac{1}{2}x^2 + \frac{1}{4}x^4$



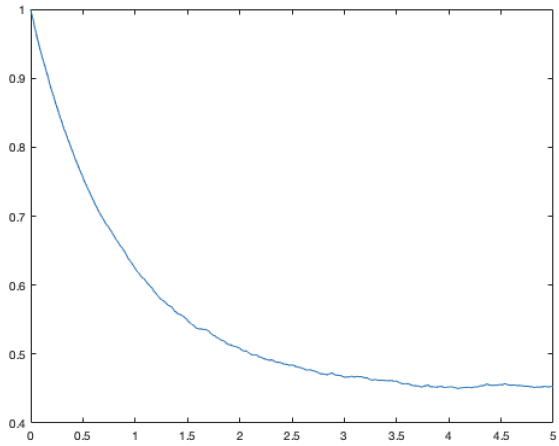
(a) $t = 0.03$



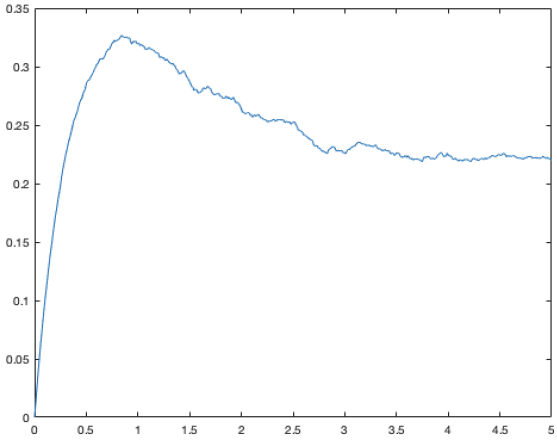
(b) $t = 0.3$



(c) $t = 5$

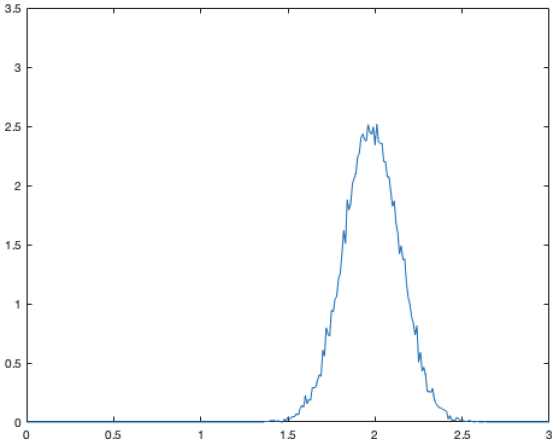


(d) Mean

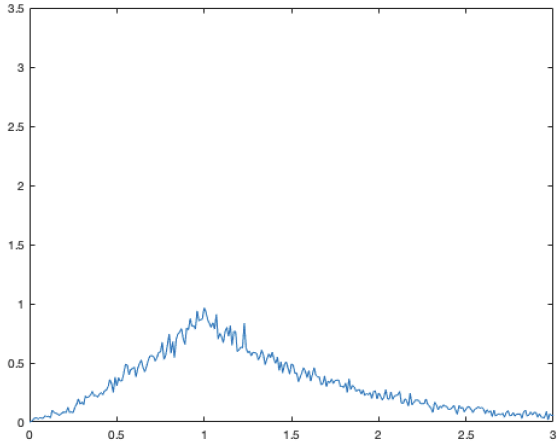


(e) Variance

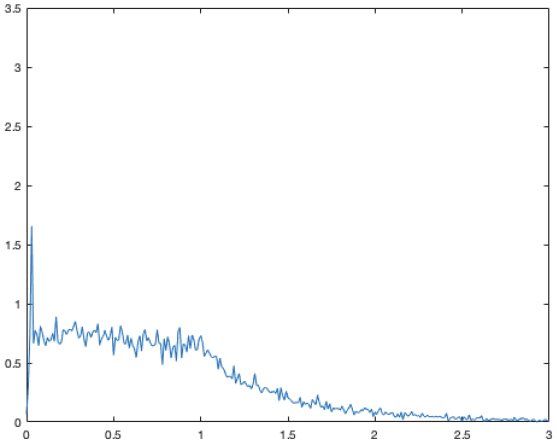
iv) Double-well potential: $V(x) = 1 - 2x^2 + x^4$ and $E_0 = 2$



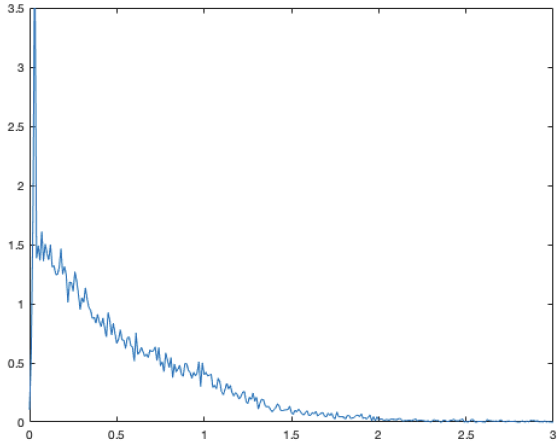
(a) $t = 0.01$



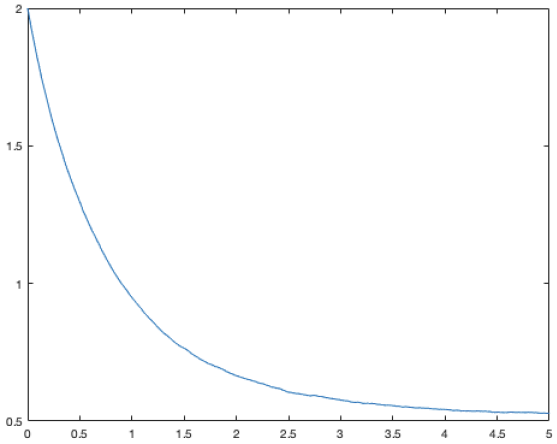
(b) $t = 0.5$



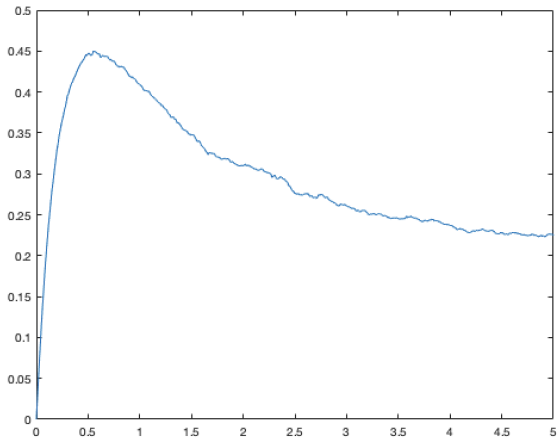
(c) $t = 1.4$



(d) $t = 5$



(e) Mean



(f) Variance

For different potential energies, we first show the change in the density of total energy with respect to time. The tendencies of their means and variances are presented in the last two graphs. In the single-well cases, the evolution of densities of E_t is similar, which is demonstrated in cases i), ii) and iii). Intuitively, the density moves from a unimodal distribution to its stationary distribution, which looks like an exponential distribution.

In our symmetric double-well case shown in case iv), we know the critical potential energy is $V(0) = 1$. When $V > 1$, the local behavior is the same as a single-well potential. When $V < 1$, the particle stays in one of the small wells. Because of the potential's symmetry, the wells' selection can be ignored. In the evolution simulation, a barrier at $V = 1$ decelerates the transition of the densities. Meanwhile, analyzing the stationary distribution, this barrier collects more densities inside the domain, $0 < E < 1$. Therefore, the stationary distribution looks like a piecewise exponential distribution. Different potential energies lead to different transient behaviors of the total energies. We want to find common ground for the evolution of E_t among different potentials.

We record the characteristics of the evolutions by mean and variance. For all the different potential energy functions, the trend of the mean of E_t remains consistent. The same goes for its variance. In general, the mean of the total energy monotonously converges to the mean of its stationary distribution. The variation of the variance generates a unimodal function. It starts from zero and converges to the variance of the stationary distribution. The incipient zero is caused by our initial total energy, which is a one-point distribution. To date, we do not find the essential mechanism inducing the peak. We need to pay more effort in this generalized CIR process (3.37) to comprehend the transition from the short to the long time limit.

3.5 Summary

In this chapter, we analyze the underdamped Langevin equation (3.5). It is well-known that there is perturbation limiting behavior when ϵ tends to zero in a compact time interval. The perturbation limit is a solution on the corresponding initial Hamiltonian without damping and noise. For any fixed ϵ, there is a time limiting behavior.

When t tends to infinity, the solution tends to its invariant measure (3.4). We attempt to construct the connection between those limits in the d-dimensional space.

The analysis starts from the one-dimensional harmonic case. The underdamped Langevin equation has an explicit solution which is a Gaussian process. We obtain the following three time scales by discussing the means and variances of the solution precisely.

i) The short time scale is $t_\epsilon^s \ll \frac{1}{\epsilon}$. In this time scale, when ϵ tends to zero, the solution tends to the perturbation limit. This result increases the time interval of the perturbation theorem, which is a compact interval.

ii) The critical time scale is $t_\epsilon^c = \frac{\theta}{\epsilon}$. When ϵ tends to zero, the limiting distribution is in the transition from the initial Hamiltonian to the invariant measure. Both damping and noise influence the solution. The degree of that influence only depends on θ.

iii) The long time scale is $t_\epsilon^l \gg \frac{1}{\epsilon}$. Let ϵ tend to zero. Then, the solution eventually merges with the invariant measure.

Theorem 16, 17 and 18 present different behaviors in the different time scales.

In the d-dimensional non-harmonic case, we analyze the total energy E in the short time scale $t_\epsilon^s \ll \frac{1}{\epsilon}$, which is one of the conserved quantities of the unperturbed Hamiltonian equation. By applying the conservation property and Itô's formula, we estimate E_t^ϵ tends to $E_t^0 \equiv E_0$ in the short time scale when ϵ tends to zero in Theorem 19. Therefore, the small perturbations have not started to affect E, and there is no transient behavior of E until t_ϵ^s. In the long time scale $t_\epsilon^l \gg \frac{1}{\epsilon}$, by applying the classical hypocoercivity technique and a refined estimation of the convergence rate, we conclude the solution X_t^ϵ converges to its invariant measure in the total variation distance when ϵ tends to zero in Corollary 21. Thus, the transition of E should still occur in the critical time scale $t_\epsilon^c = \frac{\theta}{\epsilon}$.

In the critical time scale, we only analyze the one-dimensional case. Having applied the averaging principle to our slow-fast systems, we deduce the averaged equation (3.37) of the averaged total energy $\overline{E}$. The averaging principle demonstrates that E converges to the distribution of $\overline{E}$ as $\epsilon \to 0$, which is presented in Theorem 22. Thus,

we utilize (3.37) to approximate the evolution of the total energy from E_0 to its limit. (3.37) is a generalized CIR process. There is no previous work in this non-negative square root process, whose non-negativity preserving is proved in Proposition 25. In order to gather more characteristics of (3.37), we generalize the full truncation scheme, which simulates the classical CIR process. Under a variety of potentials, we observe the common monotone decrease of the mean and the common trend of the variance.

After understanding the transient behavior of E, we focus on the remaining fast variable in the one-dimensional case. Naturally, we choose the angle as the fast variable. Since we need to simulate (3.5) in a relatively long time, we develop a data-driven method. We apply a rough Monte-Carlo simulation (10^4 samples) to generate a dataset as a reference. Then, we borrow the equations in the numerical Fokker-Planck equation's scheme as constraints. Finally, we search for the closest solution to the reference. Compared with the classical Monte-Carlo simulation (10^6 samples), our algorithm gives a smoother and more accurate solution while reducing the number of sample paths. Compared with the classical numerical Fokker-Planck equation's scheme, we avoid iterating on a large space and provide a local numerical solution of the density function.

By using the data-driven method, we simulate the full coordinates of (3.5) in Section 3.2.3. Due to the numerical results, especially in the short time scale, we speculate that the fast variable starts to transition in a time scale shorter than t_ϵ^c, and finishes the transition right before t_ϵ^c. We will provide a rigorous theorem in the future.

Chapter 4

Conclusions and future works

In this work, we study the transient behavior in random linear recurrence equation (2.2) and the underdamped Langevin equation (3.5). Both of them are subject to small noise perturbations. We aim to understand how the small perturbations influence the original deterministic systems. Roughly speaking, the classical perturbation theorem asserts that, in any finite time interval, when the small parameter ϵ tends to zero, the unperturbed dynamics play a dominant role in the movements of particles. We want to introduce various time scales t_ϵ to learn more refined dynamics over long time intervals under small perturbations.

The linear recurrence equation with additive white noise is a classical model in applied science. In Section 2.3 we provide an example of a stochastic differential equation simulated by the forward difference approximation. That explicit scheme can be simplified as a random linear recurrence equation of degree 2. When (**H**) holds, there always exists a limit of the random linear recurrence equation. We analyze the cut-off phenomenon which describes the convergence of the Gaussian process generated by the random linear recurrence equation. The distances between random variables in the process, X_t, and its limit, X_∞, abruptly drop from one to zero around the cut-off time, t_ϵ. Thus, before implementing the recurrence precisely, we can estimate the time when the process is close enough to its equilibrium by the structure of the equation.

In the underdamped Langevin equation, we still focus on the transient process from the initial distribution to the stationary distribution. We distinguish three time scales in Section 3.1 in the one-dimensional harmonic oscillator. In the short time scale, t_ϵ^s,

particles are mainly governed by the unperturbed system, which is the corresponding Hamiltonian system. In the critical time scale, t_ϵ^c, the solution starts to make a transition from the initial Hamiltonian orbit to its stationary distribution. In the long time scale, t_ϵ^l, the solution merges with the stationary distribution eventually.

Because of the structure of the Langevin equation, we suppose the different time scales still exist in the d-dimensional non-harmonic oscillator, which is presented by the $2d$-dimensional non-linear Langevin system. By the hypocoercivity technique, we know t_ϵ^l is still a long time scale for the d-dimensional non-harmonic oscillator. During the long time scale the solution is close enough to its stationary distribution (Section 3.3.2). For the other two time scales, we have the following intermediate results.

Within the short time scale t_ϵ^s we find that the total energy remains very close to its initial value (Theorem 19), which is the same phenomenon as that for the one-dimensional harmonic oscillator. In order to make an understanding at the dynamics level of the Langevin system in t_ϵ^s, we plan to analyze the other $d-1$ slow variables first. The observed quantities that need to be verified are d slow variables. For a future study, we aim to prove that those d variables remain close to their respective initial values in this time scale.

Within the critical time scale t_ϵ^c, by using the averaging principle for the stochastic slow-fast system, we know that the averaged equations are useful in understanding the transition of slow variables from the initial Hamiltonian to their stationary distribution. Having chosen the total energy as the slow variable, we verify this in the one-dimensional harmonic case (Section 3.4.2). In the one-dimensional non-harmonic case, we derive the averaged equation of the total energy and use it to describe certain behaviors of that transition (Section 3.4.1 and 3.4.3). We will consider the d-dimensional non-harmonic case using averaging techniques in future studies.

We also have conducted multiscale simulations. Since the simulations is in a relatively long time, when ϵ is small, we introduce an efficient scheme in Section 3.2. We call it the data-driven method which is a hybrid with the Monte-Carlo method and the simulation of the corresponding Fokker-Planck equation. With the data-driven method, we nicely capture the three time scales for the one-dimensional case. The simulation not only provides visualizations of our results in the different time scales,

but also leads us to phenomena which need some rigorous theoretical proofs.

For instance, in the short time scale, we observe that the total energy is close to its initial value. Then we prove it in Theorem 19. But the angle, as a fast variable, is not precise even in simulation. The averaging principle is proved up to the critical time scale. We find the limit of the averaged equation is close to the stationary distribution of the corresponding slow variable in the one-dimensional case. In future studies, we might extend the time interval of the averaging principle. Thus, the d-dimensional simulation needs to be implemented.

For the d-dimensional oscillator, the Monte-Carlo method can be directly generalized to this $2d$-dimensional system. However, since the corresponding Fokker-Planck equation has $2d$ spatial variables, the classical finite difference method is not competent for simulating this equation. Fortunately, we find there is an innovative approach for the simulation of the high dimensional PDEs, which applies neural network [63]. In the future, we will utilize this approach as the second part of our data-driven method, and simulate those slow variables as data evidence of different time scales.

After developing the d-dimensional data-driven scheme and advanced theoretical proofs' techniques, we will investigate the different time scales in the general slow-fast system. Meanwhile, we will analyze how the averaged equations demonstrate the transient behaviors of slow variables from their initial values to corresponding stationary distributions.

Bibliography

[1] A. Das and G. Samanta, "Influence of environmental noises on a prey–predator species with predator-dependent carrying capacity in alpine meadow ecosystem", *Mathematics and Computers in Simulation*, vol. 190, pp. 1294–1316, 2021.

[2] T. Reichenbach, M. Mobilia, and E. Frey, "Noise and correlations in a spatial population model with cyclic competition", *Physical review letters*, vol. 99, no. 23, p. 238 105, 2007.

[3] J.-W. Lee, S.-W. Yeh, and H.-S. Jo, "Weather noise leading to el niño diversity in an ocean general circulation model", *Climate Dynamics*, vol. 52, no. 12, pp. 7235–7247, 2019.

[4] X. Mao, *Stochastic differential equations and applications*. Elsevier, 2007.

[5] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*. Springer science & business media, 2012.

[6] E. Olivieri and M. E. Vares, *Large deviations and metastability*. Cambridge University Press, 2005.

[7] D. Aldous and P. Diaconis, "Shuffling cards and stopping times", *The American Mathematical Monthly*, vol. 93, no. 5, pp. 333–348, 1986.

[8] S. Méléard and D. Villemonais, "Quasi-stationary distributions and population processes", *Probability Surveys*, vol. 9, pp. 340–410, 2012.

[9] P. Diaconis, "The cutoff phenomenon in finite markov chains", *Proceedings of the National Academy of Sciences*, vol. 93, no. 4, pp. 1659–1664, 1996.

[10] M. Boumans, "Economics, strategies in social sciences", *Encyclopedia of Social Measurement*, pp. 751–760, 2005.

[11] O. C. Ibe, *Fundamentals of Applied Probability and Random Processes*, Second Edition. Academic Press, 2014, pp. 345–368.

[12] A. F. Siegel and M. R. Wagner, *Practical Business Statistics (Eighth Edition)*, Eighth Edition. Academic Press, 2022, pp. 445–482.

[13] G. A. Pavliotis, *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*. Springer, 2014, vol. 60.

[14] D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations", *SIAM review*, vol. 43, no. 3, pp. 525–546, 2001.

[15] S. J. Malham and A. Wiese, "An introduction to sde simulation", *arXiv preprint arXiv:1004.0646*, 2010.

[16] B. Allen and M. A. Nowak, "Games among relatives revisited", *Journal of theoretical biology*, vol. 378, pp. 103–116, 2015.

[17] F. Dubeau, "The rabbit problem revisited", *Fibonacci Quart*, vol. 31, no. 3, pp. 268–274, 1993.

[18] S. Smale and R. F. Williams, "The qualitative analysis of a difference equation of population growth", *Journal of Mathematical Biology*, vol. 3, no. 1, pp. 1–4, 1976.

[19] C. Ferguson, "Learning, expectations, and the cobweb model", *Zeitschrift für Nationalökonomie*, vol. 20, no. 3-4, pp. 297–315, 1960.

[20] M. J. Flannery and C. M. James, "The effect of interest rate changes on the common stock returns of financial institutions", *The journal of Finance*, vol. 39, no. 4, pp. 1141–1153, 1984.

[21] H. Klee, "Some novel applications of a continuous system simulation language", *Computers & Industrial Engineering*, vol. 11, no. 1-4, pp. 385–389, 1986.

[22] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2009.

[23] J. Ouaknine and J. Worrell, "On the positivity problem for simple linear recurrence sequences", in *International Colloquium on Automata, Languages, and Programming*, Springer, 2014, pp. 318–329.

[24] H. Akaike, "Fitting autoregressive models for prediction", *Annals of the institute of Statistical Mathematics*, vol. 21, no. 1, pp. 243–247, 1969.

[25] R. Dahlhaus, "Fitting time series models to nonstationary processes", *The annals of Statistics*, vol. 25, no. 1, pp. 1–37, 1997.

[26] E. A. Coddington and N. Levinson, *Theory of ordinary differential equations*. Tata McGraw-Hill Education, 1955.

[27] I. Matiiasevich, *Hilbert's tenth problem*.

[28] R. Spigler and M. Vianello, "Liouville-green approximations for a class of linear oscillatory difference equations of the second order", *Journal of computational and applied mathematics*, vol. 41, no. 1-2, pp. 105–116, 1992.

[29] G. Everest, A. J. Van Der Poorten, I. Shparlinski, T. Ward, *et al.*, *Recurrence sequences*. American Mathematical Society Providence, RI, 2003, vol. 104.

[30] G. U. Yule, "Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer's sunspot numbers", *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 226, no. 636-646, pp. 267–298, 1927.

[31] D. A. Levin and Y. Peres, *Markov chains and mixing times*. American Mathematical Soc., 2017, vol. 107.

[32] S. Martínez and B. Ycart, "Decay rates and cutoff for convergence and hitting times of markov chains with countably infinite state space", *Advances in Applied Probability*, vol. 33, no. 1, pp. 188–205, 2001.

[33] G. Barrera and M. Jara, "Abrupt convergence for stochastic small perturbations of one dimensional dynamical systems", *Journal of Statistical Physics*, vol. 163, no. 1, pp. 113–138, 2016.

[34] G. Barrera and M. Jara, "Thermalisation for small random perturbations of dynamical systems", *The Annals of Applied Probability*, vol. 30, no. 3, pp. 1164–1208, 2020.

[35] G. Barrera, "Abrupt convergence for a family of ornstein–uhlenbeck processes", *Brazilian Journal of Probability and Statistics*, vol. 32, no. 1, pp. 188–199, 2018.

[36] G. S. Lueker, "Some techniques for solving recurrences", *ACM Computing Surveys (CSUR)*, vol. 12, no. 4, pp. 419–436, 1980.

[37] R.-D. Reiss, *Approximate distributions of order statistics: with applications to nonparametric statistics*. Springer science & business media, 2012.

[38] A. DasGupta, *Asymptotic theory of statistics and probability*. Springer Science & Business Media, 2008.

[39] J. Barrera and B. Ycart, "Bounds for left and right window cutoffs", *arXiv preprint arXiv:1310.0726*, 2013.

[40] M. Viana and K. Oliveira, *Foundations of ergodic theory*. Cambridge University Press, 2016.

[41] A. Einstein, "Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen", *Annalen der physik*, vol. 4, 1905.

[42] R. Kubo, "Brownian motion and nonequilibrium statistical mechanics", *Science*, vol. 233, no. 4761, pp. 330–334, 1986.

[43] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*. Springer Science & Business Media, 2012, vol. 113.

[44] M. I. Freidlin, "Random perturbations of dynamical systems: Large deviations and averaging", *Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo*, vol. 1, no. 2/3, pp. 183–216, 1994.

[45] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of markov chain monte carlo*. CRC press, 2011.

[46] Y. Li, "A data-driven method for the steady state of randomly perturbed dynamics", *arXiv preprint arXiv:1805.04099*, 2018.

[47] K. Burrage, I. Lenane, and G. Lythe, "Numerical methods for second-order stochastic differential equations", *SIAM journal on scientific computing*, vol. 29, no. 1, pp. 245–264, 2007.

[48] P. A. Markowich and C. Villani, "On the trend to equilibrium for the fokker-planck equation: An interplay between physics and functional analysis", *Mat. Contemp*, vol. 19, pp. 1–29, 2000.

[49] C. Villani, "Hypocoercive diffusion operators", in *International Congress of Mathematicians*, vol. 3, 2006, pp. 473–498.

[50] C. Villani, "Hypocoercivity", *arXiv preprint math/0609050*, 2006.

[51] A. B. Tsybakov, *Introduction to nonparametric estimation.* 2009.

[52] F. Baudoin, "Wasserstein contraction properties for hypoelliptic diffusions", *arXiv preprint arXiv:1602.04177*, 2016.

[53] Y. Cao, J. Lu, and L. Wang, "On explicit L^2-convergence rate estimate for underdamped langevin dynamics", *arXiv preprint arXiv:1908.04746*, 2019.

[54] M. Grothaus and P. Stilgenbauer, "Hilbert space hypocoercivity for the langevin dynamics revisited", *arXiv preprint arXiv:1608.07889*, 2016.

[55] G. Pavliotis and A. Stuart, *Multiscale methods: averaging and homogenization*. Springer Science & Business Media, 2008.

[56] M. Freidlin and A. Wentzell, "Some recent results on averaging principle", in *Topics in Stochastic Analysis and Nonparametric Estimation*, Springer, 2008, pp. 1–19.

[57] M. Freidlin and A. Wentzell, "Averaging principle for stochastic perturbations of multifrequency systems", *Stochastics and Dynamics*, vol. 3, no. 03, pp. 393–408, 2003.

[58] L. Overbeck and T. Ryden, "Estimation in the cox-ingersoll-ross model", *Econometric Theory*, vol. 13, no. 3, pp. 430–461, 1997.

[59] L. B. Andersen, P. Jäckel, and C. Kahl, "Simulation of square-root processes", *Encyclopedia of Quantitative Finance*, pp. 1642–1649, 2010.

[60] W. Feller, "Two singular diffusion problems", *Annals of mathematics*, pp. 173–182, 1951.

[61] A. Cozma and C. Reisinger, "Strong order 1/2 convergence of full truncation euler approximations to the cox–ingersoll–ross process", *IMA journal of numerical analysis*, vol. 40, no. 1, pp. 358–376, 2020.

[62] Y. Yi, Y. Hu, and J. Zhao, "Positivity preserving logarithmic euler-maruyama type scheme for stochastic differential equations", *Communications in Nonlinear Science and Numerical Simulation*, vol. 101, p. 105 895, 2021.

[63] Y. Khoo, J. Lu, and L. Ying, "Solving parametric pde problems with artificial neural networks", *European Journal of Applied Mathematics*, vol. 32, no. 3, pp. 421–435, 2021.

[64] S. Elaydi, *An introduction to difference equations*. Springer Science & Business Media, 2005.

[65] J. D. Hamilton, *Time series analysis*. Princeton university press, 2020.

[66] R. M. Neal *et al.*, "Mcmc using hamiltonian dynamics", *Handbook of markov chain monte carlo*, vol. 2, no. 11, p. 2, 2011.

[67] B. Leimkuhler and S. Reich, *Simulating hamiltonian dynamics*. Cambridge university press, 2004.

Appendix A: Variance representation of $X_t^{(\epsilon)}$

Since $(\xi_t : t \geq 0)$ is a sequence of i.i.d. random variables with standard Gaussian distribution, it is not hard to see that for any $t \geq p$ the random variable $X_t^{(\epsilon)}$ has Gaussian distribution, whose expectation is x_t. The next lemma provides a representation of its variance under assumption (**H**).

Now, for the sake of intuitive reasoning and in a conscious abuse of notation we introduce the following notation. For each $s \in \mathbb{N}_0$ denote by $\sum k_j = s$ the set

$$\left\{ (k_1, \ldots, k_p) \in \mathbb{N}_0^p : \sum_{j=1}^{p} k_j = s \right\}$$

and denote by $\sum_{\sum k_j = s}$ the sum of $\sum_{(k_1,\ldots,k_p) \in \sum k_j = s}$.

Lemma 26 *Assume that* (**H**) *holds. For any $t \geq p$, $X_t^{(\epsilon)}$ has Gaussian distribution with mean x_t and variance $\epsilon^2 \sigma_t^2$, where*

$$\sigma_t^2 = 1 + \left(\sum_{\sum k_j = 1} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \right)^2 + \cdots + \left(\sum_{\sum k_j = t-p} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \right)^2$$

and $\lambda_1, \ldots, \lambda_p$ are the roots of (2.3).

Proof. By the superposition principle, the solution of the non-homogeneous linear recurrence (2.2) can be written as the general solution of the homogeneous linear recurrence (2.1) plus a particular solution of the non-homogeneous linear recurrence (2.2) as follows:

$$X_t^{(\epsilon)} = x_t^{\text{gen}} + X_t^{(\text{par},\epsilon)} \quad \text{for any } t \in \mathbb{N}_0,$$

where $X_t^{(\text{par},\epsilon)}$ solves the non-homogeneous linear recurrence (2.2), x_t^{gen} solves the homogeneous linear recurrence (2.1) but possible both solutions do not fit the pre-

scribed initial conditions. The initial conditions are fitting after adding themselves (see Section 2.4 in [64] for more details).

To find a particular solution, we introduce the *Lag operator* $\mathbb{L}$ which acts as follows: $x_{t-1} = \mathbb{L} \circ x_t$. Its inverse, $\mathbb{L}^{-1}$, is defined as $\mathbb{L}^{-1} \circ x_t = x_{t+1}$. For more details about the Lag operator we recommend Chapter 2 in [65]. Notice that the random linear recurrence (2.2) can be rewritten as

$$(\mathbb{L}^{-p} - \phi_1 \mathbb{L}^{-p+1} - \cdots - \phi_p) \circ X_t^{(\mathrm{par},\epsilon)} = \epsilon \mathbb{L}^{-p} \circ \xi_t.$$

Then

$$(1 - \lambda_1 \mathbb{L})(1 - \lambda_2 \mathbb{L}) \cdots (1 - \lambda_p \mathbb{L}) \circ X_t^{(\mathrm{par},\epsilon)} = \epsilon \xi_t,$$

where $\lambda_1, \ldots, \lambda_p$ are the roots of (2.3). Since the modulus of the roots of (2.3) are strictly less than one, we have

$$X_t^{(\mathrm{par},\epsilon)} = (1 + \lambda_1 \mathbb{L} + \lambda_1^2 \mathbb{L}^2 + \cdots) \cdots (1 + \lambda_p \mathbb{L} + \lambda_p^2 \mathbb{L}^2 + \cdots) \circ \epsilon \xi_t$$

for any $t \geq p$. Since ξ_t is only defined for $t \geq p$, we have

$$X_t^{(\mathrm{par},\epsilon)} = \left(1 + \sum_{\sum k_i = 1} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \mathbb{L} + \cdots + \sum_{\sum k_i = t-p} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \mathbb{L}^{t-p} \right) \circ \epsilon \xi_t.$$

Consequently,

$$X_t^{(\epsilon)} = x_t^{\mathrm{gen}} + \epsilon \left(\xi_t + \sum_{\sum k_i = 1} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \xi_{t-1} + \cdots + \sum_{\sum k_i = t-p} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \xi_p \right) \tag{A.1}$$

for $t \geq p$, where x_t^{gen} satisfies (2.1). After fitting the initial conditions, we see that $(x_t^{\mathrm{gen}} : t \in \mathbb{N}_0)$ is the solution of (2.1) with initial data $x_0, \ldots, x_{p-1}$. Therefore $x_t^{\mathrm{gen}} = x_t$ for any $t \in \mathbb{N}_0$. Since $(\xi_t : t \geq p)$ are i.i.d. Gaussian random variables with zero mean and unit variance, $X_t^{(\epsilon)}$ is a Gaussian distribution for any $t \geq p$. Therefore it is characterized by its mean and variance. Since the expectation of $X_t^{(\epsilon)}$ is x_t, we only need to compute its variance. From (A.1) for any $t \geq p$ we have

$$\begin{aligned} &\mathrm{Var}\left(X_t^{(\epsilon)}\right) \\ &\quad = \epsilon^2 \left(1 + \left(\sum_{\sum k_j = 1} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \right)^2 + \cdots + \left(\sum_{\sum k_j = t-p} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \right)^2 \right). \end{aligned}$$

■

Lemma 27 *Assume that* (**H**) *holds. As t goes to infinity, $X_t^{(\epsilon)}$ converges in the total variation distance to a random variable $X_\infty^{(\epsilon)}$ that has Gaussian distribution with zero mean and variance $\epsilon^2\sigma_\infty^2 \in [\epsilon^2, +\infty)$.*

Proof. From Lemma 26 we have that for any $t \geq p$, $X_t^{(\epsilon)}$ has mean x_t which is the solution of (2.1) and variance $\epsilon^2\sigma_t^2$ where

$$\sigma_t^2 = 1 + \left(\sum_{\sum k_j = 1} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \right)^2 + \cdots + \left(\sum_{\sum k_j = t-p} \lambda_1^{k_1} \cdots \lambda_p^{k_p} \right)^2 .$$

Since all the roots of (2.3) have modulus strictly less than one, relation (2.10) yields that x_t converges to zero when t goes to infinity. By a simple counting argument one can see that

$$\text{Card}\left(\sum k_j = s \right) \leq (s+1)^p \quad \text{for any } s \in \mathbb{N}_0,$$

where Card denotes the cardinality of the given set. Then for any $t \geq p$

$$\begin{aligned}
\sigma_t^2 &= 1 + \left(\sum_{\sum k_j = 1} \lambda_1^{k_1} \dots \lambda_p^{k_p} \right)^2 + \cdots + \left(\sum_{\sum k_j = t-p} \lambda_1^{k_1} \dots \lambda_p^{k_p} \right)^2 \\
&\leq 1 + (2^p\kappa)^2 + \cdots + ((t-p+1)^p \kappa^{t-p})^2 \\
&= \sum_{j=0}^{t-p} (j+1)^{2p}\kappa^{2j} \leq \sum_{j=0}^{\infty} (j+1)^{2p}\kappa^{2j} < +\infty,
\end{aligned}$$

where $\kappa = \max\limits_{1 \leq j \leq n} |\lambda_j| < 1$. Since $1 \leq \sigma_t^2 \leq \sigma_{t+1}^2 \leq \sum\limits_{j=0}^{\infty} (j+1)^{2p}\kappa^{2j} < +\infty$ for any $t \geq p$, we deduce $\lim\limits_{t \to +\infty} \sigma_t^2$ exists. Denote by σ_∞^2 its value. Observe that $\sigma_\infty^2 \in [1, +\infty)$. It follows from Lemma 31 that $X_t^{(\epsilon)}$ converges in the total variation distance to $X_\infty^{(\epsilon)}$ as t goes to infinity, which has Gaussian distribution with zero mean and variance $\epsilon^2\sigma_\infty^2$.

■

Appendix B: Total variation distance between Gaussian distributions

In this section we provide some useful properties for the total variation distance between Gaussian distributions. Recall that $\mathcal{N}(m, \sigma^2)$ denotes the Gaussian distribution with mean $m \in \mathbb{R}$ and variance $\sigma^2 \in (0, +\infty)$. A straightforward computation leads

$$\mathbf{d}_{\mathrm{TV}}\left(\mathcal{N}\left(m_1, \sigma_1^2\right), \mathcal{N}\left(m_2, \sigma_2^2\right)\right) = \frac{1}{2} \int_{\mathbb{R}} \left| \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}} - \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-m_2)^2}{2\sigma_2^2}} \right| \mathrm{d}x$$

for any $m_1, m_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in (0, +\infty)$. For details see Lemma 3.3.1 in [37].

Lemma 28 *Let $m_1, m_2 \in \mathbb{R}$ and $\sigma_1^2, \sigma_2^2 \in (0, +\infty)$. Then*

i) $\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_1, \sigma_1^2), \mathcal{N}(m_2, \sigma_2^2)) = \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_1 - m_2, \sigma_1^2), \mathcal{N}(0, \sigma_2^2))$.

ii) $\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(cm_1, c^2\sigma_1^2), \mathcal{N}(cm_2, c^2\sigma_2^2)) = \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_1, \sigma_1^2), \mathcal{N}(m_2, \sigma_2^2))$ *for any* $c \neq 0$.

Proof. The proofs of i) and ii) proceed from the Change of Variable Theorem. ■

Lemma 29

i) For any $m \in \mathbb{R}$ and $\sigma^2 \in (0, +\infty)$ we have

$$\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m, \sigma^2), \mathcal{N}(0, \sigma^2)) = \frac{2}{\sqrt{2\pi}} \int_0^{\frac{|m|}{2\sigma}} e^{-\frac{x^2}{2}} \mathrm{d}x \leq \frac{|m|}{\sigma\sqrt{2\pi}}.$$

ii) For any $m_1, m_2 \in \mathbb{R}$ and $\sigma^2 \in (0, +\infty)$ such that $|m_1| \leq |m_2| < +\infty$ we have

$$\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_1, \sigma^2), \mathcal{N}(0, \sigma^2)) \leq \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_2, \sigma^2), \mathcal{N}(0, \sigma^2)).$$

iii) If $\lim_{t\to+\infty}|m_t| = +\infty$ *and* $\sigma^2 \in (0,+\infty)$ *then*

$$\lim_{t\to+\infty} \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_t,\sigma^2),\mathcal{N}(0,\sigma^2)) = 1.$$

Proof. Notice that ii) and iii) follow immediately from i). Therefore we only prove i). From ii) of Lemma 28 we can assume that $m \geq 0$. Observe that

$$\begin{aligned}
&\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m,\sigma^2),\mathcal{N}(0,\sigma^2)) \\
&= \frac{1}{2\sqrt{2\pi}\sigma}\int_{-\infty}^{\frac{m}{2}}\left(e^{-\frac{x^2}{2\sigma^2}} - e^{-\frac{(x-m)^2}{2\sigma^2}}\right)\mathrm{d}x + \frac{1}{2\sqrt{2\pi}\sigma}\int_{\frac{m}{2}}^{+\infty}\left(e^{-\frac{(x-m)^2}{2\sigma^2}} - e^{-\frac{x^2}{2\sigma^2}}\right)\mathrm{d}x \\
&= \frac{2}{\sqrt{2\pi}\sigma}\int_{0}^{\frac{m}{2}} e^{-\frac{x^2}{2\sigma^2}}\mathrm{d}x.
\end{aligned}$$

The latter easily implies the result. ■

Lemma 30 *For any* $\sigma^2 \in (0,1)\cup(1,+\infty)$ *we have*

$$\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(0,\sigma^2),\mathcal{N}(0,1)) = \frac{2}{\sqrt{2\pi}}\int_{\min\left\{x(\sigma),\frac{x(\sigma)}{\sigma}\right\}}^{\max\left\{x(\sigma),\frac{x(\sigma)}{\sigma}\right\}} e^{-\frac{x^2}{2}}\mathrm{d}x \leq \frac{2}{\sqrt{2\pi}}x(\sigma)\left|{}^1\!/_\sigma - 1\right|,$$

where $x(\sigma) = \sigma\left(\frac{\ln(\sigma^2)}{\sigma^2-1}\right)^{1/2}$. *Moreover, we have* $\lim_{\sigma^2\to 1} x(\sigma) = 1$.

Proof. In this case a formula for $\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(0,\sigma^2),\mathcal{N}(0,1))$ can be computed explicitly as we did in the proof of i) of Lemma 29. Indeed, if $\sigma^2 \in (0,1)$ observe that

$$\begin{aligned}
\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(0,\sigma^2),\mathcal{N}(0,1)) &= \frac{1}{2\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left|\frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}} - e^{-\frac{x^2}{2}}\right|\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\int_{0}^{+\infty}\left|\frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}} - e^{-\frac{x^2}{2}}\right|\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\left[\int_{0}^{x(\sigma)}\left(\frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}} - e^{-\frac{x^2}{2}}\right)\mathrm{d}x + \int_{x(\sigma)}^{+\infty}\left(e^{-\frac{x^2}{2}} - \frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}}\right)\mathrm{d}x\right] \\
&= \frac{2}{\sqrt{2\pi}}\int_{0}^{x(\sigma)}\left(\frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}} - e^{-\frac{x^2}{2}}\right)\mathrm{d}x \\
&= \frac{2}{\sqrt{2\pi}}\int_{x(\sigma)}^{\frac{x(\sigma)}{\sigma}} e^{-\frac{x^2}{2}}\mathrm{d}x \leq \frac{2}{\sqrt{2\pi}}x(\sigma)({}^1\!/_\sigma - 1).
\end{aligned}$$

On the other hand, if $\sigma^2 \in (1, +\infty)$ one can also deduce that

$$\mathbf{d}_{\mathrm{TV}}(\mathcal{N}(0,\sigma^2),\mathcal{N}(0,1)) = \frac{2}{\sqrt{2\pi}} \int_{\frac{x(\sigma)}{\sigma}}^{x(\sigma)} e^{-\frac{x^2}{2}} \mathrm{d}x \leq \frac{2}{\sqrt{2\pi}} x(\sigma)(1 - 1/\sigma).$$

The second part of the lemma is a direct computation. ■

Lemma 31 *If* $\lim_{t\to+\infty} m_t = m \in \mathbb{R}$ *and* $\lim_{t\to+\infty} \sigma_t^2 = \sigma^2 \in (0, +\infty)$*, then*

$$\lim_{t\to+\infty} \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_t,\sigma_t^2),\mathcal{N}(m,\sigma^2)) = 0.$$

Proof. The proof follows from the triangle inequality together with i) of Lemma 28, i) of Lemma 29 and Lemma 30. ■

Lemma 32 *Let* $\sigma^2 \in (0, +\infty)$.

i) If $\limsup_{t\to+\infty} |m_t| \leq C_0 \in [0, +\infty)$ *then*

$$\limsup_{t\to+\infty} \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_t,\sigma^2),\mathcal{N}(0,\sigma^2)) \leq \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(C_0,\sigma^2),\mathcal{N}(0,\sigma^2)).$$

ii) If $\liminf_{t\to+\infty} |m_t| \geq C_1 \in [0, +\infty)$ *then*

$$\liminf_{t\to+\infty} \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_t,\sigma^2),\mathcal{N}(0,\sigma^2)) \geq \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(C_1,\sigma^2),\mathcal{N}(0,\sigma^2)).$$

Proof.

i) Let $L := \limsup_{t\to+\infty} \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_t,\sigma^2),\mathcal{N}(0,\sigma^2))$. *Then there exists a subsequence* $(t_n : n \in \mathbb{N})$ *such that* $\lim_{n\to+\infty} t_n = +\infty$ *and*

$$\lim_{n\to+\infty} \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_{t_n},\sigma^2),\mathcal{N}(0,\sigma^2)) = L.$$

Since $\limsup_{t\to+\infty} |m_t| \leq C_0$*, we have* $\limsup_{n\to+\infty} |m_{t_n}| \leq C_0$. *Then there exists a subsequence* $(t_{n_k} : k \in \mathbb{N})$ *of* $(t_n : n \in \mathbb{N})$ *such that* $\lim_{k\to+\infty} t_{n_k} = +\infty$ *and* $\lim_{k\to+\infty} |m_{t_{n_k}}|$ *exists. We define* $C := \lim_{k\to+\infty} |m_{t_{n_k}}|$ *and notice that* $0 \leq C \leq C_0$. *From Lemma 31 we obtain*

$$\lim_{k\to+\infty} \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_{t_{n_k}},\sigma^2),\mathcal{N}(0,\sigma^2)) = \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(C,\sigma^2),\mathcal{N}(0,\sigma^2)).$$

Notice that $\lim_{k\to+\infty} \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(m_{t_{n_k}},\sigma^2),\mathcal{N}(0,\sigma^2)) = L$*, then by item ii) of Lemma 29 we deduce*

$$L = \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(C,\sigma^2),\mathcal{N}(0,\sigma^2)) \leq \mathbf{d}_{\mathrm{TV}}(\mathcal{N}(C_0,\sigma^2),\mathcal{N}(0,\sigma^2)).$$

ii) The proof of ii) follows from similar arguments as we did in i). We left the details to the interested reader.

■

Appendix C: The Conservative Algorithm of The Hamiltonian Dynamics

In order to show this nonconservative phenomenon contributed by different algorithms. We consider the most straightforward Hamilton system.

$$\begin{cases} \mathrm{d}x = y\mathrm{d}t, \\ \mathrm{d}y = -x\mathrm{d}t. \end{cases}$$

The corresponding Hamiltonian is

$$H(x, y) = \frac{1}{2}x^2 + \frac{1}{2}y^2.$$

One can easily check

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\partial H}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial H}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} = 0,$$

which is the conservation of that Hamiltonian. When we fix an initial condition, the flow will be captured by its own Hamiltonian orbit. Mathematically we have the following identity,

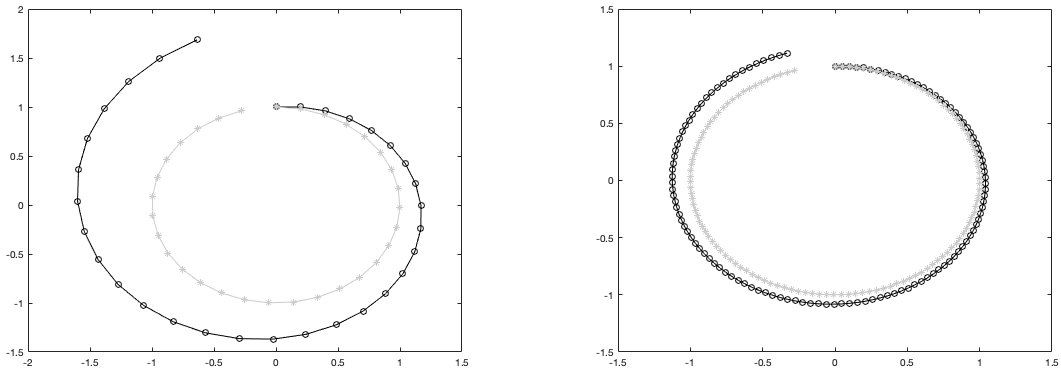
$$H(x_t, y_t) = H(x_0, y_0) \text{ for any } t.$$

When discretizing the time variable, we denote Δt as the step size. By the Euler method, the iteration can be written as

$$x_{n+1} = x_n + y_n\Delta t,$$
$$y_{n+1} = y_n - x_n\Delta t.$$

Let $x(0) = 0$ and $y(0) = 1$ be the initial condition. We known the exact solution is $x(t) = \sin(t)$ and $y(t) = \cos(t)$. For the conciseness of the figure, set time starts from

0 to 6. Figure C.1 exhibits that the accumulation of errors destroys the conservation law, where the iterative solution is marked black, and the exact solution is marked grey.



(a) $dt = 0.2$ (b) $dt = 0.05$

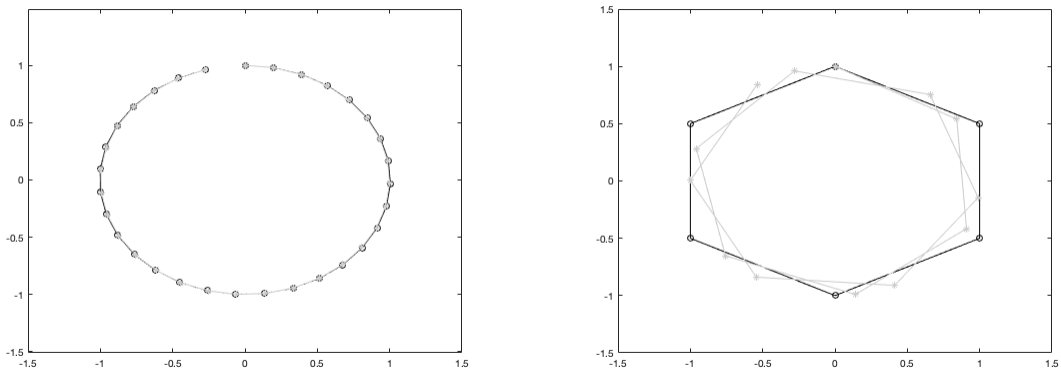
Figure C.1: Divergency of the Euler method

In figure C.1a, there is a distinct divergence from the exact solution when $dt = 0.2$. By refining the time steps to 0.05 we can alleviate the above divergency. However, with time going on, this divergence can not be ignored, especially when we observe the long time behaviors.

The $leapfrog$ method plays a vital role in preserving conservation. By introducing the half-step point, the iteration becomes

$$\begin{aligned} y_{hs} &= y_n - x_n \frac{\Delta t}{2}, \\ x_{n+1} &= x_n + y_{hs}\Delta t, \\ y_{n+1} &= y_{hs} - x_{n+1}\frac{\Delta t}{2}. \end{aligned}$$

We take the same initial condition as before.



(a) $dt = 0.2, t_{\max} = 6$ (b) $dt = 1, t_{\max} = 12$

Figure C.2: Divergency of the Leapfrog method

In figure C.2a, comparing with the Euler method(Figure C.1a), the Leapfrog method preserves the exact solution. Even when we select a tremendous time step, $dt = 1$, in figure C.2b, the iterated solution will almost maintain the same energy level as the exact solution. However, it leads to some errors in the angle variable. This type of error might cause some misunderstandings about the short time scale of the simulation. We will discuss it more in Section **??**. For more details of the Leapfrog method and other high order algorithms see [66] and [67].

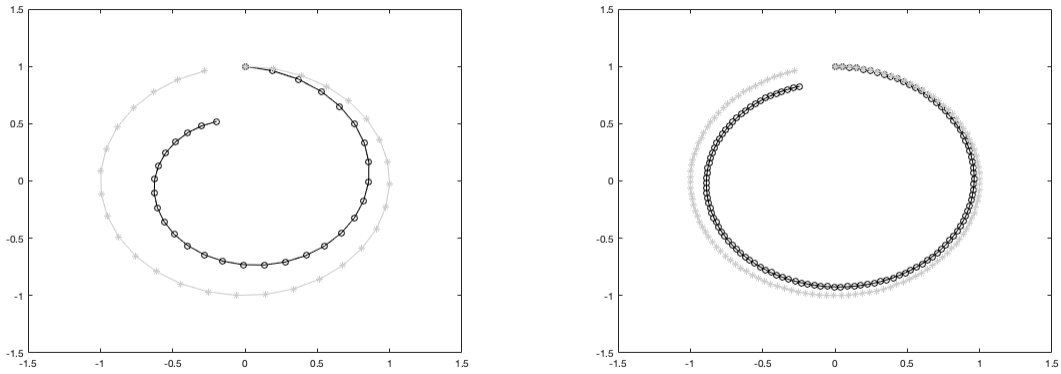
In general, implicit methods always provide more stability than explicit methods. However, the implicit methods cannot preserve the conservation as well. We will demonstrate the implicit Euler method at the end. The iteration is

$$x_{n+1} = x_n + y_{n+1}\Delta t,$$
$$y_{n+1} = y_n - x_{n+1}\Delta t.$$

As a toy model we can simplify it as an explicit expression,

$$x_{n+1} = \frac{x_n + y_n\Delta t}{1 + \Delta t^2},$$
$$y_{n+1} = y_n - x_{n+1}\Delta t.$$

The simulation result is as follows:



(a) $dt = 0.2$ (b) $dt = 0.05$

Figure C.3: Divergency of the implicit Euler method

Roughly speaking, it is a reverse result compared with the explicit Euler method. Rather than diverge to infinity, the iterated solution will converge to zero when the time is large enough. This convergency cannot be eliminated no matter how good the time step is. In figure C.3b we select $dt = 0.05$. People can generate a smaller dt but a larger $t_{\max}$ in order to understand this convergency better.

In conclusion, the above two types of Euler methods essentially destroy the conservative property. Therefore, the Leapfrog method should be our priority when we want to analyze the energy level in the conservative field.

Appendix D: Other mathematical Tools

In this section we state some elementary tools that we used all along this work. We state them here for the sake of completeness.

Lemma 33 *Let* $(a_\epsilon : \epsilon > 0)$ *and* $(b_\epsilon : \epsilon > 0)$ *be functions of real numbers. Assume that* $\lim_{\epsilon\to 0^+} b_\epsilon = b \in \mathbb{R}$. *Then*

i) $\limsup_{\epsilon\to 0^+}(a_\epsilon + b_\epsilon) = \limsup_{\epsilon\to 0^+} a_\epsilon + b.$

ii) $\liminf_{\epsilon\to 0^+}(a_\epsilon + b_\epsilon) = \liminf_{\epsilon\to 0^+} a_\epsilon + b.$

iii) $\liminf_{\epsilon\to 0^+}(a_\epsilon b_\epsilon) = b \liminf_{\epsilon\to 0^+} a_\epsilon \quad$ *when* $b > 0$.

Proof. The proofs proceed by definition of limit superior and limit inferior using subsequences. ■

Lemma 34 *For any* $\alpha \in \mathbb{R}$ *and* $r \in (0,1)$ *we have*

$$\lim_{\epsilon\to 0^+} \frac{(t^{(\epsilon)})^\alpha r^{t^{(\epsilon)}}}{\epsilon} = 1,$$

where $t^{(\epsilon)} = \frac{\ln(1/\epsilon)}{\ln(1/r)} + \alpha \frac{\ln\left(\frac{\ln(1/\epsilon)}{\ln(1/r)}\right)}{\ln(1/r)}$.

Proof. Note that $t^{(\epsilon)} = \log_r(\epsilon) - \alpha \log_r(\log_r(\epsilon))$, where $\log_r(\cdot)$ denotes the base–r logarithm function. A straightforward computation shows

$$\lim_{\epsilon\to 0^+} \frac{(t^{(\epsilon)})^\alpha r^{t^{(\epsilon)}}}{\epsilon} = \lim_{\epsilon\to 0^+} \left(1 - \alpha \frac{\log_r(\log_r(\epsilon))}{\log_r(\epsilon)}\right)^\alpha = 1.$$

■

Appendix E: Matlab Code

E.1 Exact solution

```
u_exact_1 = zeros(x_size, y_size, t_size);
u_exact = zeros(x_size, y_size, t_size);
Mean = zeros(2, 1);
Variance = zeros(2, 2);
b = 1 / 2 * sqrt(4 * k - ep^2);

%Initial distribution
x_initial = 1 + floor((x_0 - x_min + dh/2)/dh);
y_initial = 1 + floor((y_0 - y_min + dh/2)/dh);
if x_initial > 0 && x_initial <= x_size && y_initial > 0 && y_initial <= x_size
    u_exact_1(x_initial, y_initial, 1) = u_exact_1(x_initial, y_initial, 1) + 1/ dh^2;
    u_exact(x_initial, y_initial, 1) = u_exact(x_initial, y_initial, 1) + 1/ dh^2;
end

for kk = 2 : t_size
    Mean(1) = exp(-ep / 2 * t_coordinate(kk)) * (x_0 * cos(b * t_coordinate(kk))
    + 1 / b * (ep / 2 * x_0 + y_0) * sin(b * t_coordinate(kk)));

    Mean(2) = exp(-ep / 2 * t_coordinate(kk)) * (y_0 * cos(b * t_coordinate(kk))
    - 1 / b * (k * x_0 + ep / 2 * y_0) * sin(b * t_coordinate(kk)));

    Variance(1,1) = exp(-ep * t_coordinate(kk)) / (4 * b^2 + ep^2) * (2 *
    (exp(ep * t_coordinate(kk)) - 1) - ep * sin(2 * b * t_coordinate(kk)) / b
    - ep^2 * (1 - cos(2 * b * t_coordinate(kk))) / 2 / b^2);

    Variance(1,2) = ep * exp(-ep * t_coordinate(kk)) / (4 * b^2 + ep^2) *
    (1 - cos(2*b*t_coordinate(kk)) + ep^2 * (1 - cos(2*b*t_coordinate(kk)))/4/b^2);

    Variance(2,1) = Variance(1,2);

    Variance(2,2) = exp(-ep * t_coordinate(kk)) / (4 * b^2 + ep^2) * (2 *
    b^2 * (exp(ep * t_coordinate(kk)) - 1) + ep * b * sin(2 * b * t_coordinate(kk))
    + ep^2 * (cos(2 * b * t_coordinate(kk)) + exp(ep * t_coordinate(kk)) - 2) / 2
    + ep^3*sin(2*b*t_coordinate(kk))/4/b+ep^4*(1-cos(2*b*t_coordinate(kk)))/8/b^2);

    VV = det(Variance);
    for i = 1 : x_size
        for j = 1 : y_size
            u_exact_1(i, j, kk) = exp(-1 / 2 * ([x_coordinate(i) - Mean(1);
            y_coordinate(j) - Mean(2)].' * inv(Variance) * [x_coordinate(i)
            - Mean(1); y_coordinate(j) - Mean(2)]));
        end
    end
    S = sum( u_exact_1(:, :, kk) , 'all' );
    u_exact(:, :, kk) = u_exact_1(:, :, kk) / (S * dh^2);
end
```

E.2 Time evolution

```
%Fix Parameters
k = 0.25;
ep = 0.1;
x_0 = 2;
y_0 = 2;


%Space grid
x_min = -10;
x_max = 10;
y_min = -7;
y_max = 7;
dh = 0.1;
x_size = (x_max - x_min) / dh + 1;
y_size = (y_max - y_min) / dh + 1;
x_coordinate = linspace(x_min, x_max, x_size);
y_coordinate = linspace(y_min, y_max, y_size);
[x_graph,y_graph] = meshgrid(x_coordinate, y_coordinate);

%Time grid
t_min = 0;
t_max = 100;
dt = 0.01;
t_size = (t_max - t_min) / dt + 1;
t_coordinate = linspace(t_min, t_max, t_size);

u_mc = zeros(x_size, y_size, t_size);

%Initial distribution
x_initial = 1 + floor((x_0 - x_min + dh/2)/dh);
y_initial = 1 + floor((y_0 - y_min + dh/2)/dh);

if x_initial>0 && x_initial<=x_size && y_initial>0 && y_initial<=x_size
    u_mc(x_initial, y_initial, 1) = u_mc(x_initial, y_initial, 1)+1/ dh^2;
end


f = @(t,x,y) y;
g = @(t,x,y) - k * x - ep * y;

numsp = 10^4;
x_search = 0;
y_search = 0;

for i = 1 : numsp
    %MC Value Init
    x_value = x_0 ;
    y_value = y_0 ;

    for j = 2 : t_size
        x_value = x_value + dt * f(t_coordinate(j-1), x_value, y_value);
        y_value = y_value + dt * g(t_coordinate(j-1), x_value, y_value)
        + sqrt(ep * dt) * randn;
        x_search = 1 + floor((x_value - x_min + dh/2)/dh);
        y_search = 1 + floor((y_value - y_min + dh/2)/dh);

        if x_search>0 && x_search<=x_size && y_search>0 && y_search<=y_size
            u_mc(x_search, y_search, j) = u_mc(x_search, y_search, j)
            + 1/(numsp * dh^2);
        else
        disp(['Iteration:'num2str(i) ' ' 'Time:'num2str(t_coordinate(j))])
        end
    end
end
```

E.3 Comparison

```
%Define the stationary distribution
u_stationary_1 = zeros(x_size, y_size);
u_stationary = zeros(x_size, y_size);

for i = 1 : x_size
    for j = 1 : y_size
        u_stationary_1(i, j)
        = exp(-(y_coordinate(j))^2 - k*(x_coordinate(i))^2
        - k/2*(x_coordinate(i))^2);
    end
end

S = sum( u_stationary_1 , 'all' );
u_stationary = u_stationary_1 / (S * dh^2);

%Comparison
TV_MC = zeros(1, t_size);
TV_Hybrid = zeros(1, t_size);
for ii = 1 : t_size
    for i = 1 : x_size
        for j = 1 : y_size
            if u_stationary(i, j) > u_mc(i, j, ii)
                TV_MC(ii)=TV_MC(ii)+(u_stationary(i,j)-u_mc(i,j,ii))*dh^2;
            end
            if u_stationary(i, j) > u_hybrid(i, j, ii)
                TV_Hybrid(ii) = TV_Hybrid(ii) + (u_stationary(i, j)
                - u_hybrid(i, j, ii)) * dh^2;
            end
        end
    end
end
```

E.4 Crank-Nicolson scheme with the optimization

```
%Data-driven method
u_hybrid = zeros(x_size, y_size, t_size);
u_hybrid_col = zeros(x_size * y_size, t_size);
u_mc_col = zeros(x_size * y_size, t_size);
u_hybrid(:, :, 1) = u_mc(:, :, 1);


%Change MC result to a column
for kt = 1 : t_size
    flag = 1;
    for kx = 1 : x_size
        for ky = 1 : y_size
            u_mc_col(flag, kt) = u_mc(kx, ky, kt);
            flag = flag+1;
        end
    end
end


%matrix A B b for PDE
A = zeros((x_size - 2) * (y_size - 2) + 1, x_size * y_size);
A((x_size - 2) * (y_size - 2) + 1, :) = ones(1, x_size * y_size);
B = zeros((x_size - 2) * (y_size - 2) + 1, x_size * y_size);
B((x_size - 2) * (y_size - 2) + 1, :) = ones(1, x_size * y_size);
b = zeros((x_size - 2) * (y_size - 2) + 1, 1);
b(end) = dh^(-2);
```

```
for i = 2 : (x_size - 1)
    for j = 2 : (y_size - 1)
        A((i - 2) * (y_size - 2) + j - 1, (i - 2) * y_size + j)
        = -y_coordinate(j)/(4 * dh);
        A((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j - 1)
        = k*x_coordinate(i)/(4*dh)+ep*y_coordinate(j)/(4*dh)-ep/(4*dh^2);
        A((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j)
        = 1/dt - ep/2 + ep/(2 * dh^2);
        A((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j + 1)
        = -k*x_coordinate(i)/(4*dh)-ep*y_coordinate(j)/(4*dh)-ep/(4*dh^2);
        A((i - 2) * (y_size - 2) + j - 1, i * y_size + j)
        = y_coordinate(j)/(4 * dh);

        B((i - 2) * (y_size - 2) + j - 1, (i - 2) * y_size + j)
        = y_coordinate(j)/(4 * dh);
        B((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j - 1)
        = -k*x_coordinate(i)/(4*dh)-ep*y_coordinate(j)/(4*dh)+ep/(4*dh^2);
        B((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j)
        = 1/dt + ep/2 - ep/(2 * dh^2);
        B((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j + 1)
        = k*x_coordinate(i)/(4*dh)+ep*y_coordinate(j)/(4*dh)+ep/(4*dh^2);
        B((i - 2) * (y_size - 2) + j - 1, i * y_size + j)
        = -y_coordinate(j)/(4 * dh);
    end
end
A = sparse(A);
B = sparse(B);


%Initial Setting
%Find a step to start by Threshold
Density_Max_Threshold = 2;
Density_Max_t = 0;
Hybrid_Start = 0;

for Hybrid_Start = 1 : t_size
    Density_Max_t = max(u_mc_col(:, Hybrid_Start));
    if Density_Max_t > Density_Max_Threshold
        u_hybrid(:, :, Hybrid_Start) = u_mc(:, :, Hybrid_Start) ;
    else
        break;
    end
end

t_coordinate(Hybrid_Start)
u_hybrid_col(:, Hybrid_Start - 1) = u_mc_col(:, Hybrid_Start - 1);


%Optimization part
for l = Hybrid_Start : t_size
    b = B * u_hybrid_col(:, l - 1);
    d = b - A * u_mc_col(:, l);
    dx = lsqminnorm(A, d);
    u_hybrid_col(:, l) = u_mc_col(:, l) + dx;
end


%Change the hybrid column back to matrix
for kt = Hybrid_Start : t_size
    flag = 1;
    for kx = 1 : x_size
        for ky = 1 : y_size
            u_hybrid(kx, ky, kt) = u_hybrid_col(flag, kt);
            flag = flag + 1;
        end
    end
end
```

E.5 Long time simulation

```
clear all
close all

%Fix Parameters
k = 1;
x_0 = 1;
y_0 = 1;
t_min = 0;


%%Space %In here we just use the uniform dh(which can be change by dhx and dhy~)
x_min = -3.2;
x_max = 3.2;
y_min = -3.2;
y_max = 3.2;
dh = 0.1;
x_size = (x_max - x_min) / dh + 1;
y_size = (y_max - y_min) / dh + 1;
x_coordinate = linspace(x_min, x_max, x_size);
y_coordinate = linspace(y_min, y_max, y_size);
[x_graph,y_graph] = meshgrid(x_coordinate, y_coordinate);


%modify epsilon and t^epsilon
ep_v = fliplr([[10^-2 : 10^-2 : 1]]);
t_max_v = ep_v.^(-1.5);
ep_size = numel(ep_v);
u_mc_1 = zeros(x_size, y_size, ep_size);


%Initial distribution
u_mc_i = zeros(x_size, y_size);
x_initial = 1 + floor((x_0 - x_min + dh/2)/dh);
y_initial = 1 + floor((y_0 - y_min + dh/2)/dh);

if x_initial > 0 && x_initial <= x_size && y_initial > 0 && y_initial <= x_size
    u_mc_i(x_initial, y_initial) = u_mc_i(x_initial, y_initial) + 1/ dh^2;
end


for kep = 1 : ep_size

    ep = ep_v(kep);

    %Vector Field
    f = @(t,x,y) y;
    g = @(t,x,y) - k * x - ep * y;

    t_max = round(t_max_v(kep)*100)/100;
    dt = 0.01;
    t_size = floor((t_max - t_min) / dt + 1);
    t_coordinate = linspace(t_min, t_max, t_size);

    u_mc = zeros(x_size, y_size, t_size);
    u_mc(:, :, 1) = u_mc_i;


    %MCMC set up
    numsp = 10^4;
    x_search = 0;
    y_search = 0;

    %MCMC Main Part
    for i = 1 : numsp
        %MC Value Init
```

```
        x_value = x_0 ;
        y_value = y_0 ;

        for j = 2 : t_size
            x_value = x_value + dt*f(t_coordinate(j-1), x_value, y_value);
            y_value = y_value + dt * g(t_coordinate(j-1), x_value, y_value)
            + sqrt(ep * dt) * randn;

            x_search = 1 + floor((x_value - x_min + dh/2)/dh);
            y_search = 1 + floor((y_value - y_min + dh/2)/dh);

            if x_search>0 && x_search<=x_size && y_search>0 && y_search<=y_size
                u_mc(x_search, y_search, j) = u_mc(x_search, y_search, j)
                + 1/(numsp * dh^2); %Density function
            else
            disp(['Iteration:' num2str(i) ' ' 'Time:' num2str(t_coordinate(j))])
            end
        end
    end

    %PDE method
    u_hybrid = zeros(x_size, y_size, t_size);
    u_hybrid_col = zeros(x_size * y_size, t_size);
    u_mc_col = zeros(x_size * y_size, t_size);

    u_hybrid(:, :, 1) = u_mc(:, :, 1);

    %Change MC result to a column
    for kt = 1 : t_size
        flag = 1;
        for kx = 1 : x_size
            for ky = 1 : y_size
                u_mc_col(flag, kt) = u_mc(kx, ky, kt);
                flag = flag+1;
            end
        end
    end

    %matrix A B b for PDE
    A = zeros((x_size - 2) * (y_size - 2) + 1, x_size * y_size);
    A((x_size - 2) * (y_size - 2) + 1, :) = ones(1, x_size * y_size);

    B = zeros((x_size - 2) * (y_size - 2) + 1, x_size * y_size);
    B((x_size - 2) * (y_size - 2) + 1, :) = ones(1, x_size * y_size);

    b = zeros((x_size - 2) * (y_size - 2) + 1, 1);
    b(end) = dh^(-2);

    for i = 2 : (x_size - 1)
        for j = 2 : (y_size - 1)
            A((i - 2) * (y_size - 2) + j - 1, (i - 2) * y_size + j)
            = -y_coordinate(j)/(4 * dh);
            A((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j - 1)
            = k*x_coordinate(i)/(4*dh) + ep*y_coordinate(j)/(4*dh) - ep/(4*dh^2);
            A((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j)
            = 1/dt - ep/2 + ep/(2 * dh^2);
            A((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j + 1)
            = -k*x_coordinate(i)/(4*dh) - ep*y_coordinate(j)/(4*dh) - ep/(4*dh^2);
            A((i - 2) * (y_size - 2) + j - 1, i * y_size + j)
            = y_coordinate(j)/(4 * dh);

            B((i - 2) * (y_size - 2) + j - 1, (i - 2) * y_size + j)
            = y_coordinate(j)/(4 * dh);
            B((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j - 1)
            = -k*x_coordinate(i)/(4*dh) - ep*y_coordinate(j)/(4*dh) + ep/(4*dh^2);
            B((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j)
            = 1/dt + ep/2 - ep/(2 * dh^2);
            B((i - 2) * (y_size - 2) + j - 1, (i - 1) * y_size + j + 1)
```

```
            = k*x_coordinate(i)/(4*dh) + ep*y_coordinate(j)/(4*dh) + ep/(4*dh^2);
            B((i - 2) * (y_size - 2) + j - 1, i * y_size + j)
            = -y_coordinate(j)/(4 * dh);
        end
    end

    A = sparse(A);
    B = sparse(B);

    %Initial Setting
    %Find a step to start
    Density_Max_Threshold = 2;
    Density_Max_t = 0;
    Hybrid_Start = 0;

    for Hybrid_Start = 1 : t_size
        Density_Max_t = max(u_mc_col(:, Hybrid_Start));
        if Density_Max_t > Density_Max_Threshold
            u_hybrid(:, :, Hybrid_Start) = u_mc(:, :, Hybrid_Start) ;
        else
            break;
        end
    end

    t_coordinate(Hybrid_Start)
    u_hybrid_col(:, Hybrid_Start - 1) = u_mc_col(:, Hybrid_Start - 1);


    %Optimization part
    for l = Hybrid_Start : t_size
        b = B * u_hybrid_col(:, l - 1);
        d = b - A * u_mc_col(:, l);
        dx = lsqminnorm(A, d);
        u_hybrid_col(:, l) = u_mc_col(:, l) + dx;
    end

    %Change the hybrid column back to matrix
    for kt = Hybrid_Start : t_size
        flag = 1;
        for kx = 1 : x_size
            for ky = 1 : y_size
                u_hybrid(kx, ky, kt) = u_hybrid_col(flag, kt);
                flag = flag + 1;
            end
        end
    end

    u_mc_1(:, :, kep) = u_hybrid(:, :, end);
end
```

E.6 Short time simulation

```
%Fix Parameters
k = 1;
x_0 = 2;
y_0 = 2;
t_min = 0;
dt = 0.01;
numsp = 10^5;

%%Space %In here we just use the uniform dh
x_min = -5;
x_max = 5;
y_min = -5;
```

```
y_max = 5;
dh = 0.1;
x_size = (x_max - x_min) / dh + 1;
y_size = (y_max - y_min) / dh + 1;
x_coordinate = linspace(x_min, x_max, x_size);
y_coordinate = linspace(y_min, y_max, y_size);
[x_graph,y_graph] = meshgrid(x_coordinate, y_coordinate);


%modify epsilon and t^epsilon

ep_v = fliplr([10^-4 : 10^-3 : 10^-1 + 10^-4]);
t_max_v = ep_v.^(-0.5); %Short time scale

ep_size = numel(ep_v);
u_mc_1 = zeros(x_size, y_size, ep_size);

%Only linear muti Gaussian Mean
x_mean = zeros(ep_size,1);
y_mean = zeros(ep_size,1);
x_exact = zeros(ep_size,1);
y_exact = zeros(ep_size,1);

diff_2norm = zeros(ep_size,1);

for kep = 1 : ep_size

    ep = ep_v(kep);

    %Vector Field k x^2 case
    f = @(t,x,y) y;
    g = @(t,x,y)  - k * x - ep * y;

    t_max = round(t_max_v(kep)*100)/100;
    t_size = floor((t_max - t_min) / dt + 1);
    t_coordinate = linspace(t_min, t_max, t_size);

    u_mc = zeros(x_size, y_size);

    %MCMC set up
    x_search = 0;
    y_search = 0;

    %MCMC Main Part
    for i = 1 : numsp
        %MC Value Init
        x_value = x_0 ;
        y_value = y_0 ;
        x_his = 0;

        for j = 2 : t_size
            x_his = x_value;
            x_value = x_value + dt * f(t_coordinate(j-1), x_value, y_value);
            y_value = y_value + dt * g(t_coordinate(j-1), x_his, y_value)
            + sqrt(ep * dt) * randn;
        end

        x_search = 1 + floor((x_value - x_min + dh/2)/dh);
        y_search = 1 + floor((y_value - y_min + dh/2)/dh);

        if x_search>0 && x_search<=x_size && y_search>0 && y_search<=y_size
            u_mc(x_search, y_search) = u_mc(x_search, y_search) + 1/(numsp*dh^2);
            x_mean(kep) = x_mean(kep) + x_coordinate(x_search) * 1/numsp;
            y_mean(kep) = y_mean(kep) + y_coordinate(y_search) * 1/numsp;
        else
            disp(['Iteration:' num2str(i) ' ' 'Time: 'num2str(t_coordinate(j))])
        end
```

```
    end

    x_exact(kep) = x_0 * cos(t_max) + y_0 * sin(t_max);
    y_exact(kep) = y_0 * cos(t_max) - x_0 * sin(t_max);

    diff_2norm(kep)=((x_mean(kep)-x_exact(kep))^2+(y_mean(kep)-y_exact(kep))^2)^(1/2);

    u_mc_1(:, :, kep) = u_mc(:, :);
end
```

E.7 $F(E)$ of the double-well potential

```
function [F] = Double_F_energy(EA)

[x_p, x_n] = Double_terminal_point(EA); %use newton method find the properbound

if EA >= 0.03
    if (EA <= 0.9998) || (EA > 1)
        integrand_numerator = @(s) (2 * EA - 2 + 4.*s.^2 - 2.*s.^4).^(1/2);
        integrand_denominator = @(s)(2 * EA - 2 + 4.*s.^2 - 2.*s.^4).^(-1/2);
        numerator = integral(integrand_numerator, x_n, x_p);
        denominator = integral(integrand_denominator, x_n, x_p);
        F = numerator / denominator;
    else
        F = 0;
    end
else
        F = 0;
end

end
```

```
function [upper, lower] = Double_terminal_point(EE)

if EE >= 0.03
    if EE > 1  %untrapped orbits
        v = @(x) 1 - 2 * x^2 + x^4 - EE;
        upper = fzero(v,2);
        lower = - upper;
    elseif EE <= 0.9998
        v = @(x) 1 - 2 * x^2 + x^4 - EE;
        upper = fzero(v,1.5);
        k = fzero(v,0.1);
        if k > 0
            lower = k;
        else
            lower = fzero(v,0.9);
        end
    else
        v = @(x) 1 - 2 * x^2 + x^4 - EE;
        upper = fzero(v,2);
        lower = 0;
    end
else
    upper = 1;
    lower = 1;
end

end
```

E.8 Process of the double-well potential

```
clear all
close all

%Fix Parameters and initial condition
k = 1;
x_0 = 2;

%Space
t_min = 0;
t_max = 5;
dt = 0.01;
t_size = floor((t_max - t_min) / dt + 1);
t_coordinate = linspace(t_min, t_max, t_size);

x_min = 0;
x_max = 8;
dh = 0.01;
x_size = (x_max - x_min) / dh + 1;
x_coordinate = linspace(x_min, x_max, x_size);

%Density function set up
u_x = zeros(x_size, t_size);

%Initial distribution
u_x_i = zeros(x_size, 1);
x_initial = 1 + floor((x_0 - x_min + dh/2)/dh);
if x_initial > 0 && x_initial <= x_size
    u_x_i(x_initial) = u_x_i(x_initial) + 1/ dh;
end
u_x(:, 1) = u_x_i;

%MCMC set up
numsp = 2*10^4;
x_search = 0;

%MCMC Main Part as reference
for i = 1 : numsp

    %MC Value Init every sample start from the same IC
    x_value = x_0;
    for j = 2 : t_size
        F = Double_F_energy(x_value);
        M = max(0,F);
        x_value = x_value + (1/2 - M) * dt +  sqrt(M) * sqrt(dt) * randn;
        x_value = max(0,x_value);
        if imag(x_value) ~= 0
            break
        end
        x_search = 1 + floor((x_value - x_min + dh/2)/dh);
        if x_search > 0 && x_search <= x_size
            u_x(x_search, j) = u_x(x_search, j) + 1/(numsp * dh);
        else
            disp(['Iteration: ' num2str(i) '    ' 'Time: ' num2str(t_coordinate(j))])
        end
    end
end


%% plot density of energy

filename = 'Mid_Energy_AveSys_Double.gif';

fmat = moviein(t_size);

for i = 1 : t_size
```

```
    plot(x_coordinate, u_x(:,i))
    xlim([0 3])
    ylim([0 3.5])
    xlabel('Total energy');
    ylabel('Density');
    title(['Time = ' num2str(t_coordinate(i))]);

    fmat(:, i) = getframe;
    f = getframe(gcf);
    imind = frame2im(f);
    [imind, cm] = rgb2ind(imind, 256);
    if i == 1
        imwrite(imind, cm, filename, 'gif', 'Loopcount', inf, 'DelayTime', 0.1);
    else
        imwrite(imind, cm, filename, 'gif', 'WriteMode', 'append', 'DelayTime', 0.1);
    end
end
```