Large Deviations for Markov Jump Processes and Hamiltonian Trajectories

by

Daniël Matthijs Driessen

To obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday August 27, 2015 at 14:00 PM.

Student number:	1526324	
Project duration:	January 1, 2015 – July 9, 2015	
Thesis committee:	Prof. dr. F. H. J. Redig,	TU Delft, supervisor
	Dr. ir. L. E. Meester,	TU Delft
	Dr. J. L. A. Dubbeldam,	TU Delft

TU Delft Delft University of Technology

Acknowledgements

First of all I would like to express my appreciation to my supervisor Prof. Dr. Frank Redig for giving me the opportunity doing this research in the field of Large Deviation Theory. The meetings together were always very helpful and motivating and he made me feel comfortable to ask any question that was on my mind. Writing this Thesis I gained a lot of knowledge and I enjoyed doing it.
I would also like to thank Msc. R.C. Kraaij for taking the time helping me when I suddenly dropped by his office because I had a question.
A special thanks to my beloved girlfriend Elisabet for her love and support. She always cheered me up and motivated me when I got stuck.
I am very grateful to my family. My parents always trusted me in the choices I made for studying and gave me all the opportunities I needed. I appreciate it very much that I can always count on the love and support of my family.

Abstract

This master thesis is concerned with Large Deviation Theory in combination with Lagrangian and Hamiltonian dynamics. In particular, the Large Deviation behaviour of the empirical distribution of n independent two-state continuous-time Markov processes is studied. We start by looking at the general theory of Large Deviations in both the finite dimensional case as well as for infinite dimensional stochastic processes. After this, the connection is shown of Large Deviation Theory with Lagrangian and Hamiltonian dynamics. The Hamiltonian of the empirical distribution of n independent two-state Markov processes is derived and using this, via the Hamilton equations, the dynamics of this process are derived. That is, the most likely path that is taken when in time T we force the path to start in state a and end in state b. The main goal of this thesis is to find out more about the dynamics of this process (behaviour of the trajectories) and to derive an explicit equation for the so-called Action integral. We want to compute the Action for the general case. That is, the case in which the rate going from state one to state two can differ from the rate going from state two to state one. Once the Action integral is computed we look at the asymptotic behaviour of this Action integral. This is important as it says something about the probability of some trajectories occurring for T an extreme. We look at the asymptotics for both $T \to 0$ and the $T \to \infty$.

Contents

General Introduction

The great mathematician Jacob Bernoulli showed around 1689 that them mean of independent and identically distributed random variables converges to the expected value. Sometimes however, the unexpected happens. For finite sample size n, the mean can substantially deviate from the expectation, of course with small probability. This is where the theory of Large Deviations comes in. Simply put, Large Deviation Theory deals with such rare events. It is an important and active field in probability theory with applications in many different areas. Examples of such deviations of the mean from the expectation in a very general context are queuing theory, financial mathematics, thermodynamics, statistical mechanics and biology. The earliest work in this area dates back to Laplace (1749-1829) and Cramér (1893-1985) but the ideas and concepts got coherently introduced and formally defined only in 1966 in a paper by the Indian American mathematician Varadhan.

In this thesis Large Deviation Theory is combined with Lagrangian and Hamiltonian dynamics. As will be shown, there is a nice connection between the two areas. Lagrangian and Hamiltonian dynamics describe the motions and interactions of a system of particles. They are another way of looking at the well known equation of motion, derived by Newton namely $F = ma$. This equation is known as the second law of mechanics. The theory will be useful, in connection with Large Deviation Theory, when we look at the Large Deviation behaviour of stochastic processes. Stochastic processes represent the evolution of a random variable over time, which in the continuous case leads to continuous random paths. The stochastic process that is of main interest in this report is the two-state continuous-time Markov process. More specifically, we look at the measure of the ratio of particles in state 1 and 2 respectively. An important distinction that is made is between the symmetric and asymmetric case. In the symmetric case, the rate of jumping from state 1 to state 2 is the same as vice versa while in the asymmetric case these rates differ.

An important part is to find, in the symmetric as well as the asymmetric case, the path that will be taken (the most probable path) given some distribution a of particles at time $t = 0$ and distribution b at time $t = T$. To illustrate this with an example, we can consider a chemical solution with a large number of particles n. The particles pass (approximately) independently of each other from state 1 to state 2 and vice versa with rates γ_1 and γ_2 respectively. Now, taking T big enough, the particles will eventually be distributed according to the equilibrium values. That is, $\frac{\gamma_2}{\gamma_1+\gamma_2}$ will be the fraction in state 1 and $\frac{\gamma_1}{\gamma_1+\gamma_2}$ will be the fraction in state 2. However the goal is to find, if we assume we start with some distribution a and end with a distribution b, the most probable path among all of these atypical paths from a to b. We will see that this path is the path that minimizes the so-called action (the action will be formally defined in chapter 2) over all paths starting from a and going to b. Another important part of this report is to find an explicit expression of the action and the asymptotic behaviour of this action for T being very small or very large.

In the first chapter of this report the basics of Large Deviation Theory are treated, focussing on those parts that will be used in later chapters. The first part of the chapter deals with Large Deviations in finite dimensions. Here Cramer's Theorem is stated and proved. Cramer's Theorem is important as it tells something about Large Deviation behaviour of empirical means of independently and identically distributed random vectors. Later it is shown that there is a more general condition for the LDP to hold, this is the content of the Gärtner-Ellis Theorem. Here the independence of the random variables is not longer required. After this we go on with large Deviation Theory for whole random processes evolving over time. Mogulskii's theorem and Schilder's theorem (following very quickly once Mogulskii's theorem is proved) are stated and proved. Mogulskii's theorem relies on the finite dimensional result and the fact that using projective limits (Dawson-Gärtner theorem) this can lifted to the infinite-dimensional case. Projective limits, and in particular the Dawson-Gärtner theorem, are treated in the last section of chapter one.

In the first section of chapter two the Lagrangian and Hamiltonian formalism is treated. We only look at the special case when energy is a conserved quantity. We will prove the Principle of Least Action which states that the path taken by a particle is the path that minimizes the so-called Action. Later in this chapter the connection between Lagrangian and Hamiltonian dynamics on the one hand and Large Deviation theory on the other hand is explained. In the last part of this chapter we use this theory to derive the so-called Hamiltonian of the two-state continuous-time Markov process. This will be very useful in the last chapter.

In the last chapter of this report we use the theory of the previous two chapters and use it on the two-state continuous-time Markov process. With the derived Hamiltonian the dynamics of the process, that is the evolution of the process over time given some boundary conditions, are derived. We will see that this goes a lot easier in the symmetric case compared with the asymmetric case. Having derived the equations that govern the dynamics of the process explicitly, graphs can be drawn of the trajectories. The behaviour of the trajectories are very interesting and depend highly on the amount of time the system has to evolve. After this the Action Integral is calculated and the asymptotic behaviour is examined.

Chapter 1

Basics of Large Deviation Theory

1.1 Introduction

Large Deviation Theory is concerned with the asymptotic behaviour of so called rare events of sequences of probability measures. In order to illustrate the basic idea, let $X_1, X_2, ..., X_n$ be an i.i.d. sequence of random variables on a probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ where $\mathcal{B}(\mathbb{R})$ is the Borel sigma-field on $\mathbb{R}$. Let $\mu := EX_1$ and $\sigma^2 := Var X_1$. Furthermore let S_n denote the partial sums, i.e. $S_n = X_1 + ... + X_n$. We know, from the two fundamental theorems in probability theory, that for this sequence,

Strong Law of Large Numbers (SLLN)

$$\frac{1}{n} S_n \xrightarrow[n\to\infty]{} \mu \qquad \mathbb{P}\, almost\ surely \tag{1.1}$$

Central Limit Theorem (CLT)

$$\frac{1}{\sigma\sqrt{n}}(S_n - \mu n) \xrightarrow[n\to\infty]{} Z \qquad \mathbb{P}\, almost\ surely \tag{1.2}$$

From the SLLN we see that the empirical average $\frac{1}{n}S_n$ converges to μ as $n \to \infty$. The CLT shows the probabilty the partial sums differ from μn and that this goes in distribution to the Gaussian distribution. We see from (1.2) that deviations of the size $\sqrt{n}$ are "normal". Now we can look what happens if the deviations are of order n, i.e. the event $\{S_n \geq \alpha n\}$, where $\alpha := \mu + l$ and $l > 0$. As n tends to infinity, the probability of this event goes to zero of course (the distribution of $\frac{S_n}{n}$ converges to the degenerate distribution at μ), but the goal is to specify the rate of this. It can be expected that the probability decays exponentially. After all, if we assume we start with Gaussian distributed random variables we get,

$$\mathbb{P}(X_1 + ... + X_n \geq (\mu + l)n) = \mathbb{P}(\frac{S_n}{n} \geq \alpha) = \frac{\sqrt{n}}{\sqrt{\pi}} \int_\alpha^\infty e^{-\frac{nx^2}{2}} dx = e^{-\frac{n\alpha^2}{2} + o(n)} \simeq e^{-nI(\alpha)} \tag{1.3}$$

Where $I(\alpha) = \frac{\alpha^2}{2}$. By I we denote a rate function, which is explained below. By the symbol $\simeq$ we mean that the two sequences are exponentially equivalent. Two sequences of positive numbers (α_n) and (β_n) are exponentially equivalent iff[1],

$$\lim_{n\to\infty} \frac{1}{n}(\log \alpha_n - \log \beta_n) = 0.$$

This form of $I(\alpha)$ above reflects the fact that the distribution of the X_i was Gaussian to begin with. We will see in the following that, under some condition on the distribution of X_1, in general it holds that the decay is exponential in n and we have,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \approx \alpha\right) \simeq e^{-nI(\alpha)}. \tag{1.4}$$

[1] In definition 3 we define exponentially equivalence, this however is for probability measures

Now we will start explaining the basics of Large Deviation Theory and give some of the most important theorems. For the theorems in this chapter we follow for a large part the book "Large Deviations Techniques and Aplications" by Dembo and Zeitouni (later to be referred to as [1]). First we will look at the finite dimensional case. It is defined what is meant if a sequence of probability measures satisfies the Large Deviation Principle (LDP). Furthermore Cramer's Theorem and the Gärtner-Ellis Theorem are explained. After this, the large deviation behavior of sequences of paths of random processes are studied. This leads to the theorems of Mogulskii and Schilder. The last section deals with projective limits and the Dawsnon-Gärtner Theorem. This can be used to generalise an LDP from one topological space to another topological space and is used to proof Mogulskii's Theorem.

1.2 The basics of LDT in finite dimensions

Let $\{\mu_\varepsilon\}$ be a family of probability measures on a probability space $(\mathcal{X},\mathcal{B})$. Here $\mathcal{X}$ is a topological space so open and closed subsets of $\mathcal{X}$ are well defined. Now large deviation theory, and in particular the large deviation principle (LDP), characterizes the limiting behaviour of the family of probability measures $\{\mu_\varepsilon\}$ as $\varepsilon \to 0$. For the definition of the large deviation principle we need to define what is called a rate function.

Definition 1 (Rate Function)**.** *A rate function I is a lower semicontinuous mapping $I : \mathcal{X} \to [0,\infty]$ such that for all $\alpha \in [0,\infty)$, the level set $\psi_I(\alpha) := \{x : I(x) \le \alpha\}$ is a closed subset of $\mathcal{X}$. A rate function is called a good rate function if all level sets as defined above are compact subsets of $\mathcal{X}$. We define the effective domain $\mathcal{D}_I$ of I as the set of points in $\mathcal{X}$ of finite rate, i.e. $\mathcal{D}_I := \{x : I(x) < \infty\}$. When no confusion occurs we refer to $\mathcal{D}_I$ as the domain of I.*

With this definition we can now define what is meant by a set of probability measures satisfying the large deviation principle. For a set Γ, $\bar{\Gamma}$ denotes the closure of Γ, Γ° the interior of Γ, and Γ^c the complement of Γ. The infimum of a function over the empty set is defined as ∞.

Definition 2 (Large deviation principle)**.** *$\{\mu_\varepsilon\}$ satisfies the large deviation principle with a rate function I if, for all $\Gamma \in \mathcal{B}$,*

$$-\inf_{x\in\Gamma^\circ} I(x) \le \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\Gamma) \le \limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\Gamma) \le -\inf_{x\in\bar{\Gamma}} I(x) \tag{1.5}$$

When, in the following, we say that μ_ε satisfies the LDP we mean that the above holds for some rate function I. It is easy to see that if μ_ε satisfies the LDP and $\Gamma \in \mathcal{B}$ is such that

$$\inf_{x\in\Gamma^\circ} I(x) = \inf_{x\in\bar{\Gamma}} I(x) := I_\Gamma$$

then:

$$\lim_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\Gamma) = -I_\Gamma$$

A set Γ that satisfies the above is called an I continuity set. Note that since $\mu_\varepsilon(\mathcal{X}) = 1$ for all ε, it follows from the LDP that $\inf_{x\in\mathcal{X}} I(x) = 0$ for the upper bound to hold (note I is a non-negative function). Now when I is a good rate function it follows that there exists at least one point x for which $I(x) = 0$, so good rate functions attain their minimum. Furthermore note that when $\inf_{x\in\bar{\Gamma}} I(x) = 0$ the upper bound trivially holds and that when $\inf_{x\in\Gamma^\circ} I(x) = \infty$ the lower bound trivially holds. We can use this to reformulate the LDP given above into an equivalent statement, namely,

(a) (*Upper bound*) For every $\alpha < \infty$ and every measurable set Γ with $\bar{\Gamma} \subset \psi_I(\alpha)^c$,

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\Gamma) \le -\alpha. \tag{1.6}$$

(b) (*Lower bound*) For any $x \in \mathcal{D}_I$ and every measurable Γ with $x \in \Gamma^\circ$,

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\Gamma) \ge -I(x). \tag{1.7}$$

Cramér's Theorem
Assume we have n d-dimensional random i.i.d. vectors $X_1, X_2, ..., X_n$ distributed according to the probability law μ and with the logarithmic moment generating function defined as:

$$\Lambda(\lambda) := \log E[e^{\langle \lambda, X_1 \rangle}] \tag{1.8}$$

Cramér's theorem specifies the LDP for the laws μ_n of the empirical mean $\hat{S}_n := \frac{1}{n}\sum_{j=1}^{n} X_j$ of these i.i.d. d-dimensional random vectors $X_1, X_2, ..., X_n$. in particular it shows that the rate function satisfying this LDP is given by the Fenchel-Legendre transform of the logarithmic moment generating function $\Lambda(\lambda)$. This Fenchel-Legendre transform of $\Lambda(\lambda)$ is given by:

$$\Lambda^\star(x) := \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, x\rangle - \Lambda(\lambda)\}. \tag{1.9}$$

The proof will be given for the one dimensional case (d=1). For the general case we refer to [1], p.26. First we state the one-dimensional Cramér theorem.

Theorem 1 (Cramér's Theorem)**.** *Let $X_1, X_2, ..., X_n$ be a set of i.i.d. random vectors in $\mathbb{R}$. The sequence of measures $\{\mu_n\}$ (measures of the empirical mean of n vectors as defined above) satisfies the LDP with convex rate function $\Lambda^\star(.)$, namely:*
(a) For any closed set $F \subset \mathbb{R}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} \Lambda^\star(x). \tag{1.10}$$

(b) For any open set $G \subset \mathbb{R}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_{x\in G} \Lambda^\star(x). \tag{1.11}$$

In order to get a feeling for this theorem we use Cramer's theorem on a collection of i.i.d. Gaussians. In the introduction of this chapter we have seen what Large Deviation behaviour to expect.

Example 1 (Cramér's Theorem applied on Gaussians)**.** *Before we will prove the above theorem, first let us go back to the situation where we start with standard normally distributed random variables $X_1, X_2, ..., X_n$. Let μ_n be the distribution of the mean of n such random variables. We can look again at the event that $\mathbb{P}(X_1 + ... + X_n \geq (\mu + l)n) = \mathbb{P}(\frac{S_n}{n} \geq \alpha) = \mu_n[\alpha, \infty)$, like in the introduction, but now using Cramér's Theorem. We have in this case,*

$$\Lambda(\lambda) = \log E\, e^{\lambda X_1} = \log e^{\frac{\lambda^2}{2}} = \frac{\lambda^2}{2}.$$

So we get,

$$\Lambda^\star(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \frac{\lambda^2}{2}\} = x^2 - \frac{1}{2}x^2 = \frac{1}{2}x^2.$$

Where the last equality can be seen by noting that if we take the derivative with respect to λ and equating to zero we get $\lambda = x$. So Cramér's Theorem tells us now that ,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} \frac{1}{2}x^2 = -\frac{\alpha^2}{2}$$

Where $F = [\alpha, \infty)$. On the other hand, from the lower bound (1.11) from Theorem 1 it follows that,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_{x\in G} \frac{1}{2}x^2 = -\frac{\alpha^2}{2}$$

Where $G = (\alpha, \infty)$. Combining the two bounds it is clear that,

$$\mu_n(\alpha, \infty) \simeq e^{-\frac{n\alpha^2}{2}}$$

Which is the same as we found in the introduction.

To prove the stated theorem above we first need the following lemma which states some usefull properties of $\Lambda^\star(.)$ and $\Lambda(.)$.

Lemma 1. *(a) Λ is a convex function and $\Lambda^\star$ is a convex rate function.*
(b) If $\mathcal{D}_\Lambda = \{0\}$, then $\Lambda^\star$ is identically zero. If $\Lambda(\lambda) < \infty$ for some $\lambda > 0$, then $\bar{x} < \infty$ (possibly $\bar{x} = -\infty$), and for all $x \geq \bar{x}$, $\Lambda^\star(x)$ is a nondecreasing function given by:

$$\Lambda^\star(x) = \sup_{\lambda \geq 0}[\lambda x - \Lambda(\lambda)] \tag{1.12}$$

Similarly, if $\Lambda(\lambda) < \infty$ for some $\lambda < 0$, then $\bar{x} > -\infty$ (possibly $\bar{x} = \infty$), nd for all $x \leq \bar{x}$, $\Lambda^\star(x)$ is a nonincreasing function given by:

$$\Lambda^\star(x) = \sup_{\lambda \leq 0}[\lambda x - \Lambda(\lambda)] \tag{1.13}$$

When $\bar{x}$ is finite, $\Lambda^\star(\bar{x}) = 0$, and always.

$$\inf_{x \in \mathbb{R}} \Lambda^\star(x) = 0. \tag{1.14}$$

(c) $\Lambda(.)$ is differentiable in $\mathcal{D}^\circ_\Lambda$ with

$$\Lambda^{'}(\eta) = \frac{1}{M(\eta)} E[X_1 e^{\eta X_1}] \tag{1.15}$$

and

$$\Lambda^{'}(\eta) = y \implies \Lambda^\star(y) = ny - \Lambda(\eta). \tag{1.16}$$

Proof of lemma 1. (a) To show Λ is convex we can use Hölder's inequality, namely for any $\theta \in [0, 1]$:

$$\Lambda(\theta\lambda_1 + (1-\theta)\lambda_2) = \log E[(e^{\lambda_1 X_1})^\theta (e^{\lambda_2 X_1})^{(1-\theta)}] \leq \log\{E[e^{\lambda_1 X_1}]^\theta E[e^{\lambda_2 X_1}]^{(1-\theta)}\} = \theta\Lambda(\lambda_1) + (1-\theta)\Lambda(\lambda_2)$$

The convexity of $\Lambda^\star$ follows from its definition, namely $\Lambda^\star$ is a supremum of linear functions:

$$\begin{aligned}\theta\Lambda^\star(x_1) + (1-\theta)\Lambda^\star(x_2) &= \sup_{\lambda \in \mathbb{R}}\{\theta\lambda x_1 - \theta\Lambda(\lambda)\} + \sup_{\lambda \in \mathbb{R}}\{(1-\theta)\lambda x_2 - (1-\theta)\Lambda(\lambda)\} \\ &\geq \sup_{\lambda \in \mathbb{R}}\{(\theta x_1 + (1-\theta)x_2)\lambda - \Lambda(\lambda)\} = \Lambda^\star(\theta x_1 + (1-\theta)x_2).\end{aligned}$$

To show that $\Lambda^\star$ is a rate function note that $\Lambda(0) = 0$ and because $\Lambda^\star(x)$ is a supremum over all λ we clearly have $\Lambda^\star(x) \geq 0x - \Lambda(0) = 0$. To show lowersemicontinuity fix a sequence $\{x_n\}$ converging to x. Now, for every $\lambda \in \mathbb{R}$,

$$\liminf_{x_n \to x} \Lambda^\star(x_n) \geq \liminf_{x_n \to x}[\lambda x_n - \Lambda(\lambda)] = \lambda x - \Lambda(\lambda)$$

Thus in particular,

$$\liminf_{x_n \to x} \Lambda^\star(x_n) \geq \sup_{\lambda \in \mathbb{R}}[\lambda x - \Lambda(\lambda)] = \Lambda^\star(x)$$

(b) If $\mathcal{D}_\Lambda = \{0\}$, then $\Lambda^\star(x) = -\Lambda(0) = 0$ for all $x \in \mathbb{R}$. If $\Lambda(\lambda) = \log M(\lambda) < \infty$ for some $\lambda > 0$, then using Jensen's inequality we see that $\infty > M(\lambda)/\lambda = E\, e^{\lambda X_1}/\lambda \geq e^{E[\lambda X_1]}/\lambda > E[\lambda X_1]/\lambda = E[X_1] = \bar{x}$ (possibly $\bar{x} = -\infty$). Furthermore, for all $\lambda \in \mathbb{R}$, by using Jensen's inequality for concave functions we see,

$$\Lambda(\lambda) = \log E[e^{\lambda X_1}] \geq E[\log e^{\lambda X_1}] = \lambda\bar{x}$$

If $\bar{x} = -\infty$, then $\Lambda(\lambda) = \infty$ for λ negative, and (1.12) clearly holds in this situation. When $\bar{x}$ is finite, it follows from the preceding inequality that,

$$\Lambda^\star(\bar{x}) = \sup_{\lambda \in \mathbb{R}}\{\lambda\bar{x} - \Lambda(\lambda)\} \leq \sup_{\lambda \in \mathbb{R}}\{\lambda\bar{x} - \lambda\bar{x}\} = 0$$

From which it follows, because $\Lambda^\star \geq 0$, that $\Lambda^\star(\bar{x}) = 0$. In this case, for every $x \geq \bar{x}$ and every $\lambda < 0$,

$$\lambda x - \Lambda(\lambda) \leq \lambda\bar{x} - \Lambda(\lambda) \leq \Lambda^\star(\bar{x}) = 0$$

So we see (1.12) holds. Note that from (1.12) it follows that $\Lambda^\star(x)$ is a nondecreasing function of x $\forall x \in (\bar{x}, \infty)$. When $\Lambda(\lambda) < \infty$ for some $\lambda < 0$ we can proof (1.13) and the fact that $\Lambda^\star(x)$ is a nonincreasing function of x $\forall x \in (-\infty, \bar{x})$ by using the same arguments as above. It remains to show that $\inf_{x\in\mathbb{R}} \Lambda^\star(x) = 0$. This is already shown for the situation when $\mathcal{D}_\Lambda = \{0\}$ and when $\bar{x}$ is finite. Now assume $\bar{x} = -\infty$ while $\Lambda(\lambda) < \infty$ for some $\lambda > 0$. Then by Chebycheff's inequality and (1.12),

$$\log \mu([x, \infty]) = \log E[I_{X_1 - x \geq 0}] \leq \inf_{\lambda \geq 0} \log E[e^{\lambda(X_1 - x)}] = -\sup_{\lambda \geq 0}\{\lambda x - \Lambda(\lambda)\} = -\Lambda^\star(x).$$

Hence,

$$\lim_{x\to-\infty} \Lambda^\star(x) \leq \lim_{x\to-\infty} \{-\log \mu([x, \infty])\} = 0.$$

And we see that (1.14) also holds in this case. The validity of (1.14) in the situation that $\bar{x} = \infty$ while $\Lambda(\lambda) < \infty$ for some $\lambda < 0$, can be proved in the same way.
(c) The identity (1.15) follows by interchanging the order of differentiation and integration. Namely, assuming this is allowed we see,

$$\frac{d}{d\eta}\Lambda(\eta) = \frac{1}{M(\eta)}\frac{d}{d\eta}E[e^{\eta X_1}] = \frac{1}{M(\eta)}E[\frac{d}{d\eta}e^{\eta X_1}] = \frac{1}{M(\eta)}E[X_1 e^{\eta X_1}]$$

To justify the interchanging we use the dominated convergence theorem. Note that $f_\varepsilon(x) := (e^{(\eta+\varepsilon)x} - e^{\eta x})/\varepsilon$ converges pointwise to $\frac{d}{d\eta}e^{\eta x}$ for $\varepsilon \to 0$. Furthermore $\forall \varepsilon \in (-\delta, \delta)$,

$$|f_\varepsilon(x)| \leq \frac{e^{\eta x}(e^{\delta|x|} - 1)}{\delta} =: h(x)$$

Note that $h(x)$ is integrable as we assume we are in $\mathcal{D}^\circ_\Lambda$ (so $\Lambda(\lambda) < \infty$) and we can choose a $\delta > 0$ small enough.

To show (1.16), let $\Lambda'(\eta) = y$ and consider the function $g(\lambda) := \lambda y - \Lambda(\lambda)$. We showed that $\Lambda(\lambda)$ is a convex function, and since $g(\lambda)$ is a straight line minus this convex function, $g(\lambda)$ is concave. Note that $g'(\eta) = y - \Lambda'(\eta) = 0$ from which it follows that,

$$g(\eta) = \sup_{\lambda\in\mathbb{R}}\{\lambda y - \Lambda(\lambda)\} = \Lambda^\star(y)$$

□

Proof of Cramër's Theorem. (a) Let F be a non-empty closed set. Note that (1.10) trivially holds when $I_F = inf_{x\in F}\Lambda^\star(x) = 0$. Assume that $I_F > 0$. From the lemma 1 it follows that $\bar{x}$ exists, possibly as an extended real number. An application of Chebycheff's inequality yields $\forall x, \lambda \geq 0$

$$\begin{aligned} \mu_n([x, \infty]) &= E[I_{\hat{S}_n \geq 0}] \leq E[e^{n\lambda(\hat{S}_n - x)}] \\ &= e^{-n\lambda x}\prod_{i=1}^{n} E[e^{\lambda X_1}] = e^{-n[\lambda x - \Lambda(\lambda)]}. \end{aligned} \tag{1.17}$$

So, in particular $\mu_n([x, \infty]) \leq e^{-n\sup_{\lambda\geq 0}[\lambda x - \Lambda(\lambda)]}$. Therefore, if $\bar{x} < \infty$, then by (1.12), for every $x > \bar{x}$,

$$\mu_n([x, \infty)) \leq e^{-n\Lambda^\star(x)}. \tag{1.18}$$

By a similar argument, if $\bar{x} > -\infty$ then for every $x < \bar{x}$,

$$\mu_n((-\infty, x]) \leq e^{-n\Lambda^\star(x)}. \tag{1.19}$$

Now when $\bar{x}$ is finite, $\Lambda^\star(\bar{x}) = 0$, and because we assumed $I_F > 0$ it must be the case that $\bar{x} \in F^c$ (F^c is open as F is closed). Let (x_-, x_+) be the union of all the open intervals $(a, b) \in F^c$ that contain $\bar{x}$. Note that $x_- < x_+$ with either x_- or x_+ finite since both F^c and F are non empty. If x_- is finite, then

$x_- \in F$, and thus $\Lambda^\star(x_-) \geq \inf_{x\in F}\Lambda^\star(x) = I_F$. Likewise when x_+ is finite, $\Lambda^\star(x_+) \geq I_F$. Applying (1.19) for $x = x_+$ and (1.18) for $x = x_-$ we get:

$$\mu_n(F) \leq \mu_n((-\infty, x_-]) + \mu_n([x_+, \infty)) \leq 2e^{-nI_F}$$

Taking the natural logarithm of the above, dividing by n and taking the limit as $n \to \infty$ gives the desired upper bound as stated in (1.10).
Now suppose that $\bar{x} = -\infty$. Then, we see from part (b) of the above lemma that $\Lambda^\star$ is nondecreasing and $\lim_{x\to-\infty\Lambda^\star(x)} = 0$. Because we assumed $I_F > 0$ we must have $x_- = -\infty$ and $x_+ = inf\{x : x \in F\}$. Again we have that $x_+ \in F$ and consequently $\Lambda^\star(x_+) \geq I_F$. Moreover, $F \subset [x, \infty)$ and thus we can use (1.18) to get:

$$\mu_n(F) \leq \mu_n([x_+, \infty)) \leq e^{-nI_F}$$

Taking the natural logarithm of the above, dividing by n and taking the limit as $n \to \infty$ gives the desired upper bound as stated in (1.10).The case where $\bar{x} = \infty$ can be handled analogously. This shows that the upper bound holds.

(b) To prove the lower bound (1.11) we prove that for every $\delta > 0$ and every marginal law $\mu \in M_1(\mathbb{R})$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(-\delta, \delta) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = -\Lambda^\star(0). \tag{1.20}$$

Now we look at the transformation $Y = X + x$ which results in $\Lambda_Y(\lambda) = \Lambda(\lambda) - \lambda x$, and $\Lambda_Y^\star(.) = \sup_{\lambda\in\mathbb{R}}[\lambda(.+x) - \Lambda(\lambda)] = \Lambda^\star(.+x)$. Applying inequality (1.20) on Y we find, with ν the marginal law of Y, the following

$$\liminf_{n\to\infty} \frac{1}{n} \log \nu_n(-\delta, \delta) \geq -\Lambda_Y^\star(0).$$

And since $\nu_n(-\delta, \delta) = \mu_n(x-\delta, x+\delta)$ and $\Lambda_Y^\star(0) = \Lambda^\star(x)$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(x-\delta, x+\delta) \geq -\Lambda^\star(x). \tag{1.21}$$

Which holds for every $\delta > 0$ and every x. Now if we have an arbitrary open set G we can take the supremum over all $x \in G$ and all δ such that $(x-\delta, x+\delta) \subset G$, we get,

$$\begin{aligned}\sup_{\substack{x\in G\\ \delta s.t.\,(x-\delta,x+\delta)\subset G}} \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(x-\delta, x+\delta) &= \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \\ \geq \sup_{\substack{x\in G\\ \delta s.t.\,(x-\delta,x+\delta)\subset G}} -\Lambda^\star(x) &= \sup_{x\in G} -\Lambda^\star(x) = -\inf_{x\in G}\Lambda^\star(x).\end{aligned}$$

Which shows us we get the lower bound (1.11). It remains to prove (1.20). First suppose that $\mu((-\infty, 0)) > 0$, $\mu((0, \infty)) > 0$. From this it follows that $\Lambda(\lambda) \to \infty$ as $|\lambda| \to \infty$. Secondly, suppose that μ is supported on a bounded subset of $\mathbb{R}$ from which it follows that $\Lambda(.)$ is finite everywhere. Now it follows from part (c) of lemma 1 that $\Lambda(.)$ is a continuous differentiable function and hence there exists a finite η such that$\Lambda(\eta) = \inf_{\lambda\in\mathbb{R}}\Lambda(\lambda)$ and $\Lambda^{'}(\eta) = 0$.
Define a new probability measure $\tilde{\mu}$ in terms of μ via

$$\frac{d\tilde{\mu}}{d\mu}(x) = e^{\eta x - \Lambda\eta} \tag{1.22}$$

Note that this is indeed a probability measure because

$$\int_{\mathbb{R}} d\tilde{\mu} = \frac{1}{M(\eta)} \int_{\mathbb{R}} e^{\eta x} d\mu = \frac{M(\eta)}{M(\eta)} = 1.$$

This change of measure with the optimal η is called "Cramer's trick". Now let $\tilde{\mu}_n$ be the law governing $\tilde{S}_n$ when X_i are i.i.d. random variables of law $\tilde{\mu}$. Note that for every $\varepsilon > 0$,

$$\begin{aligned} \mu_n((-\varepsilon,\varepsilon)) &= \int_{|\sum_{i=1}^n x_i|<n\varepsilon} \mu(dx_1)...\mu(dx_n) \\ &\geq e^{-n\varepsilon|\eta|} \int_{|\sum_{i=1}^n x_i|<n\varepsilon} exp(\eta \sum_{i=1}^n x_i)\mu(dx_1)...\mu(dx_n) \\ &= e^{-n\varepsilon|\eta|} e^{n\Lambda(\eta)} \tilde{\mu}_n((-\varepsilon,\varepsilon)). \end{aligned} \tag{1.23}$$

By (1.15) and the choice of η,

$$E_{\tilde{\mu}}[X_1] = \frac{1}{M(\eta)} \int_{\mathbb{R}} x e^{\eta x} d\mu = \Lambda^{'}(\eta) = 0$$

Hence, by the law of large numbers,

$$\lim_{n\to\infty} \tilde{\mu}_n((-\varepsilon,\varepsilon)) = 1.$$

Now, from (1.23), taking the logarithm, dividing by n and the limit as n goes to infinity, that for every $0 < \varepsilon < \delta$,

$$\liminf_{n\to\infty} \frac{1}{n}\mu_n((-\delta,\delta)) \geq \liminf_{n\to\infty} \frac{1}{n}\mu_n((-\varepsilon,\varepsilon)) \geq \Lambda(\eta) - \varepsilon|\eta|,$$

Taking the limit $\varepsilon \to 0$ and noting that, using (1.16) of lemma 1 from which it follows that $\Lambda(\eta) = -\Lambda^\star(0)$, gives us (1.20).

Now suppose that μ is of unbounded support and still $\mu((-\infty,0)) > 0$, $\mu((0,\infty)) > 0$. There exists an $M < \infty$ big enough such that $\mu((-M,0)) > 0$ as well as $\mu((0,M)) > 0$. Let

$$\Lambda_M(\lambda) = \log \int_{-M}^{M} e^{\lambda x} d\mu.$$

Let ν denote the law of X_1 conditioned on $\{|X_1| \leq M\}$, and ν_n be the law of the corresponding $\hat{S}_n$ conditioned on $\{|X_i| \leq M, i = 1,...,n\}$. Then, for all n and every $\delta > 0$,

$$\mu_n((-\delta,\delta)) \geq \nu_n((-\delta,\delta))\mu([-M,M])^n. \tag{1.24}$$

Now the preceding proof holds for ν_n and so (1.20) holds for ν_n. The logarithmic moment generating function associated with ν is

$$\log E[e^{\lambda X_1}|\ |X_1| \leq M] = \log \frac{E[e^{\lambda X_1} I_{X_1 \leq M}]}{\mu([-M,M])} = \log \frac{\int_{-M}^{M} e^{\lambda x} d\mu}{\mu([-M,M])} = \Lambda_M(\lambda) - \log(\mu([-M,M])). \tag{1.25}$$

Now,

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n}\mu_n((-\delta,\delta)) &\geq \log\mu([-M,M]) + \liminf_{n\to\infty} \frac{1}{n}\nu_n((-\delta,\delta)) \\ &\geq \log\mu([-M,M]) + \inf_{\lambda\in\mathbb{R}} [\Lambda_M(\lambda) - \log(\mu([-M,M]))] \\ &= \inf_{\lambda\in\mathbb{R}} \Lambda_M(\lambda) \end{aligned}$$

Where the first inequality follows from (1.24) and the second inequality follows from combining (1.25) with the result (1.20) for ν_n. Now let $I_M = -\inf_{\lambda\in\mathbb{R}} \Lambda_M(\lambda)$ and $I^\star = \limsup_{M\to\infty} I_M$. It follows that

$$\liminf_{n\to\infty} \frac{1}{n}\mu_n((-\delta,\delta)) \geq -I^\star. \tag{1.26}$$

Note that $\Lambda_M(.)$ and thus so is $-I_M$. Moreover, $-I_M \leq \Lambda_M(0) \leq \Lambda(0) = 0$ and hence $-I^\star \leq 0$. Now since $-I_M$ is finite for M large enough, $-I^\star > -\infty$. From this, it follows that the level sets $\{\lambda : \Lambda_M(\lambda) \leq -I^\star\}$

are non-empty, compact sets that are nested with respect to M. So there is a point, say λ_0, in the intersection of the above sets. By Lebegue's monotone convergence theorem, $\Lambda(\lambda_0) = \lim_{M\to\infty} \lambda_M(\lambda_0) \leq -I^\star$. So using (1.26) we get

$$\liminf_{n\to\infty} \frac{1}{n}\mu_n((-\delta,\delta)) \geq -I^\star \geq \Lambda(\lambda_0) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda).$$

Showing (1.20) for μ of unbounded support.

Now in the case that either $\mu((-\infty,0)) = 0$ or $\mu((0,\infty)) = 0$, then $\Lambda(.)$ is a monotone function with $\inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = \log\mu(\{0\})$. In this case (1.20) follows from

$$\mu_n((-\delta,\delta)) \geq \mu_n(\{0\}) = \mu(\{0\})^n.$$

□

There is a multivariate version of Cramér's theorem dealing with the large deviations of the empirical means of i.i.d. random vectors in $\mathbb{R}^d$. For the prove see [1], p. 36. Cramér's theorem is limited to the case where all stochastic variables are i.i.d.. However, there is an extension to the case where the random variables are not i.i.d.. Consider a sequence of random vectors $Z_n \in \mathbb{R}^d$, where Z_n posseses the law μ_n, and logarithmic moment generating function

$$\Lambda_n(\lambda) := \log E[e^{\langle\lambda, Z_n\rangle}]. \tag{1.27}$$

Assumption 1. *We make the following assumption about the logarithmic moment generating functions. For each $\lambda \in \mathbb{R}^d$, the logarithmic moment generating function, defined as the limit*

$$\Lambda(\lambda) := \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda)$$

exists as an extended real number. Further, the origin belongs to the interior of $\mathcal{D}_\Lambda := \{\lambda \in \mathbb{R}^d : \Lambda(\lambda) < \infty\}$.

Note that we can see the sequence of vectors Z_n as the empirical mean of n vectors in the sequence of vectors $\{X_n\}$ (and μ_n as the corresponding law of the Z_n). Now when the X_n are i.i.d., μ_n is the law of the empirical mean $\hat{S}_n$ of the i.i.d. random vectors $X_i \in \mathbb{R}^d$. We see, in this situation, that for every $n \in \mathbb{Z}_+$,

$$\begin{aligned}\frac{1}{n}\Lambda_n(n\lambda) = \frac{1}{n}\log E[e^{n\langle\lambda, Z_n\rangle}] = \frac{1}{n}\log E[e^{n\langle\lambda, \frac{X_1+..+X_n}{n}\rangle}] = \frac{1}{n}\log E[e^{n\langle\lambda, X_1\rangle}] \\ = \frac{1}{n}\log E[e^{\langle\lambda, X_1\rangle}]^n = \log E[e^{\langle\lambda, X_1\rangle} =: \Lambda(\lambda)\end{aligned}$$

and assumption 1 holds when $0 \in \mathcal{D}^0_\Lambda$. Now, let $\Lambda^\star$ be the Fenchel-Legendre transform of $\Lambda(.)$ with $\mathcal{D}_{\Lambda^\star} = \{x \in \mathbb{R}^d : \Lambda^\star(x) < \infty\}$. The following theorem states the LDP for the measures μ_n of the random vectors $Z_n \in \mathbb{R}^d$ (that not necessarily are i.i.d.)

Theorem 2 (Gärtner-Ellis). *Let assumption 1 hold. (a) For any closed set F,*

$$\limsup_{n\to\infty} \frac{1}{n}\log\mu_n(F) \leq -\inf_{x\in F} \Lambda^\star(x). \tag{1.28}$$

(b) For any open set G,

$$\liminf_{n\to\infty} \frac{1}{n}\log\mu_n(G) \geq -\inf_{x\in G\cap\mathcal{F}} \Lambda^\star(x), \tag{1.29}$$

where $\mathcal{F}$ is the set of exposed points[2] of $\Lambda^\star$ whose exposing hyperplane belongs to $\mathcal{D}^\circ_\Lambda$.

(c) If Λ is an essentially smooth[3], lower semicontinuous function, then the LDP holds with the good rate function $\Lambda^\star(.)$.

[2]for the definition of exposed points and exposing hyperplane see [1] p.44
[3]for the definition of essentially smooth see [1] p.44

The proof is left out of this paper but the above shows the key properties for the logarithmic moment generating function for an LDP to hold. Namely, the properties stated in assumption 1 together with the property that the logarithmic moment generating function is an essentially smooth, lower semicontinuous function. Without the latter assumption still the upper bound, stated in theorem 2 part (a), and a weaker lower bound, stated in theorem 2 part (b), hold.

1.3 Sample Path Large Deviations

In the above we have seen finite dimensional LDP's for the empirical means of a sequence of random variables. Often, we are interested in rare events that depends on **a whole random process** that evolves over time. For instance, the probability that the path of a random process is an element of some set. Again when looking at the average of a sequence of random paths, this average will converge to some deterministic path. We can look what happens if the path is different from this deterministic path, in particular the rate the probability of such a path goes to zero as the number of paths n in the sequence increases.

1.3.1 Introduction and Mogulskii's Theorem

We can now look at a sequence of n random processes $\{Z_n(t)\}$, starting at $t = 0$ and ending at $t = T$. Given some conditions on the sequence of random processes, this sequence converges to a deterministic path for $n \to \infty$. Now we can look what happens to the probability if the path of this random process is an element of a set of "a-typical" paths, say Γ. Of course, for $n \to \infty$ this probability goes to zero. But we are interested in the rate of the decay. We will see below, that under suitable conditions, we find again an exponential decay in n of the following type,

$$\mathbb{P}(Z_n(t) \in \Gamma) \simeq e^{-n \inf_{\gamma_s \in \Gamma} \int_0^T L(\gamma_s, \dot{\gamma}_s) ds}. \tag{1.30}$$

Here, $L(\gamma_s, \dot{\gamma}_s)$ can be seen as the cost of a particular path on time s. The γ_s's are the different possible paths. Here, the connection with Langrangian dynamics becomes clear as we can see $\int_0^T L(\gamma_s, \dot{\gamma}_s) ds$ as the action of the path $\{\gamma(s)\}$. Again we look at a sequence $X_1, X_2, .. \in \mathbb{R}^d$ of i.i.d. random vectors with $\Lambda(\lambda) := \log E(e^{\langle \lambda, X_1 \rangle}) < \infty \; \forall \lambda \in \mathbb{R}^d$. As we have seen, Cramér's theorem allows the analysis of the large deviations of the empirical mean $\frac{1}{n} \sum_{i=1}^{n} X_i$ of the sequence of the i.i.d. random vectors. Now we will consider the large deviations joint behaviour of a family of random variables indexed by t.

Define

$$Z_n(t) := \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} X_i, \; 0 \leq t \leq 1, \tag{1.31}$$

and let μ_n be the law of $Z_n(.)$ in $L_\infty[0, 1]$. Note that, taking the limit of $n \to \infty$the random process converges to a deterministic path. Namely the path starting at zero and going linearly to the expected value of X_1 which is attained for $t = 1$.

Throughout, $|x| := \sqrt{\langle x, x \rangle}$ denotes the Euclidean norm in $\mathbb{R}^d$, $||f||$ denotes the supremum norm on $L_\infty[0, 1]$ and $\Lambda^\star(x) := sup_{\lambda \in \mathbb{R}^d}[\langle \lambda, x \rangle - \Lambda(\lambda)]$ denotes the Fenchel-Legendre transform of $\Lambda(.)$. Furthermore, the following definitions will be used,

Definition 3 (Exponentially equivalent). *Let $(\mathcal{Y}, d)$ be a metric space and $\{\mu_\varepsilon\}$ and $\{\tilde{\mu}_\varepsilon\}$ sequences of probability measures on $\mathcal{Y}$. The probability measures $\{\mu_\varepsilon\}$ and $\{\tilde{\mu}_\varepsilon\}$ are exponentially equivalent if there exist probability spaces $\{\Omega, \mathcal{B}_\varepsilon, P_\varepsilon\}$ and two families of $\mathcal{Y}$-valued random variables $\{Z_\varepsilon\}$ and $\{\tilde{Z}_\varepsilon\}$ with joint probability laws $\{P_\varepsilon\}$ and marginals $\{\mu_\varepsilon\}$ and $\{\tilde{\mu}_\varepsilon\}$, respectively, such that the following is satisfied:*

For each $\delta > 0$, the set $\{\omega : (\tilde{Z}_\varepsilon, Z_\varepsilon) \in \Gamma_\delta\}$ is $\mathcal{B}_\varepsilon$ measurable, and

$$\limsup_{\varepsilon \to 0} \varepsilon \log P_\varepsilon(\Gamma_\delta) = -\infty, \tag{1.32}$$

where

$$\Gamma_\delta := \{(\tilde{y}, y) : d(\tilde{y}, y) > \delta\} \subset \mathcal{Y} \times \mathcal{Y} \tag{1.33}$$

Definition 4 (Exponentially Tight)**.** *Suppose that all the compact subsets of $\mathcal{X}$ belong to $\mathcal{B}$. A family of probability measures $\{\mu_\varepsilon\}$ on $\mathcal{X}$ is exponentially tight if for every $\alpha < \infty$, there exist a compact set $\mathcal{K}_\alpha \subset \mathcal{X}$ such that*

$$\limsup_{\varepsilon \to 0} \varepsilon \mu_\varepsilon(\mathcal{K}_\alpha^c) < \alpha \tag{1.34}$$

Definition 3 is important as it allows to pass an LDP from one sequence of probability measures to another sequence of probability measures. This is a consequence of the following theorem[4],

Theorem 3. *If an LDP with a good rate function $I(.)$ holds for the probability measures $\{\mu_\varepsilon\}$, which are exponentially equivalent to $\{\tilde{\mu}_\varepsilon\}$, then the same LDP holds for $\{\tilde{\mu}_\varepsilon\}$.*

Definition 4 is important because of the following corrolary[5],

Corollary 1. *Let $\{\mu_\varepsilon\}$ be an exponentially tight family of probability measures on $\mathcal{X}$ equipped with the topology τ_1. If $\{\mu_\varepsilon\}$ satisfies an LDP with respect to a Hausdorff topology τ_2 on $\mathcal{X}$ that is coarser than τ_1, then the same LDP holds with respect to the topology τ_1.*

Below the theorem is stated that specifies the Large Deviation Principle for the defined $Z_n(t)$. In particular, it specifies the rate function $I(\phi)$ with which this LDP holds.

Theorem 4 (Mogulskii)**.** *The measures μ_n satisfy in $L_\infty[0, 1]$ the LDP with the good rate function*

$$I(\phi) = \begin{cases} \int_0^1 \Lambda^\star(\dot{\phi}(t)dt & \text{if} \quad \phi \in \mathcal{AC}, \phi(0) = 0 \\ \infty & \text{otherwise}, \end{cases} \tag{1.35}$$

Where $\mathcal{AC}$ denotes the space of absolutely continuous functions, i.e.,

$$\mathcal{AC} := \{\phi \in C([0, 1]) : \sum_{l=1}^{k} |t_l - s_l| \to 0, s_l < t_l \leq s_{l+1} < t_{l+1} \Rightarrow \sum_{l=1}^{k} |\phi(t_l) - \phi(s_l)| \to 0\}.$$

Note that $\phi : [0, 1] \to \mathbb{R}^d$ absolutely continuous implies that ϕ is differentiable almost everywhere, in particular, ϕ is the integral of a function in $L_1([0, 1])$ (Fundamental Theorem of Calculus for Lebesgue Integrals).

The proof of Mogulskii's theorem has the following structure. First we will show that there exists a continuous stochastic process $\tilde{Z}_n(.)$ in $L_\infty([0, 1])$ whose measures are exponentially equivalent (see definition 3) with the measures of $Z_n(.)$ in $L_\infty([0, 1])$. Next, $\mathcal{X}$ is defined, consisting of all the maps $f : [0, 1] \to \mathbb{R}^d$ mapping $t = 0$ to the origin. These maps $f : [0, 1] \to \mathbb{R}^d$ can be identified with paths of $\tilde{Z}_n(.)$. Now it can be shown that the measures of $\tilde{Z}_n(.)$ satisfy the LDP with good rate function (1.35) on $\mathcal{X}$ equiped with the topology of pointwise convergence on $[0, 1]$. This is done by first proving the LDP for all finite projections p_j from the functions f to ordered j-dimensional vectors (specified in lemma 6). For this finite-dimensional LDP we use the Gärtner-Ellis Theorem and the Contraction Principle. Now using projective limits via consistency we show that the LDP also holds in the infinite-dimensional space $\mathcal{X}$ (follows from the Dawson-Gärtner Theorem). The LDP then also holds for these measures on $C_0([0, 1])$ with the topology induced by $\mathcal{X}$, i.e. the pointwise convergence topology (This topology can be identified with τ_2 in corollary 1). This topology can then be strengthened in this space to the supremum norm topology (via corollary 1, the supremum norm topology can be identified with τ_2), using that the measures are exponentially tight in $C_0([0, 1])$ equipped with the supremum norm topology. Subsequently, it can be shown that the same LDP holds on $L_\infty([0, 1])$. And by the exponentially equivalentness of the measures for $\tilde{Z}_n(.)$ and $Z_n(.)$ we can show the LDP holds for μ_n. For the proof of Theorem 4 we need the following four lemma's.

[4]The proof of theorem 3 can be found in [1], p.130

[5]The proof of corollary 1 can be found in [1], p.129

Lemma 2. *Let ξ be a measurable subset of $\mathcal{X}$ such that $\mu_\varepsilon(\xi) = 1$ for all $\varepsilon > 0$. Suppose that ξ is equipped with the topology induced by $\mathcal{X}$.*
(a) If ξ is a closed subset of $\mathcal{X}$ and $\{\mu_\varepsilon\}$ satisfies the LDP in ξ with rate function I, then $\{\mu_\varepsilon\}$ satisfies the LDP in $\mathcal{X}$ with rate function $I^{'}$ such that $I^{'} = I$ on ξ and $I^{'} = \infty$ on ξ^c.
(b) If $\{\mu_\varepsilon\}$ satisfies the LDP in $\mathcal{X}$ with rate function I and $D_I \subset \xi$, then the same LDP holds in ξ. In particular, if ξ is a closed subset of $\mathcal{X}$, then $D_I \subset \xi$ and hence the LDP holds in ξ.

Lemma 3. *Let $\tilde{\mu}_n$ denote the law of $\tilde{Z}_n(.)$ in $L_\infty[(0,1])$, where*

$$\tilde{Z}_n(t) := Z_n(t) + (t - \frac{[nt]}{n})X_{[nt]+1} \tag{1.36}$$

is the polynomial approximation of $Z_n(t)$. Then the probability measures μ_n and $\tilde{\mu}_n$ are exponentially equivalent in $L_\infty[(0,1])$.

Lemma 4. *Let $\mathcal{X}$ consist of all the maps from $[0,1]$ to $\mathbb{R}^d$ such that $t = 0$ is mapped to the origin, and equip $\mathcal{X}$ with the topology of pointwise convergence on $[0,1]$. Then the probability measures $\tilde{\mu}_n$ of lemma 3 (defined on $\mathcal{X}$ by the natural embedding) satisfy the LDP in this Hausdorff topological space with the good rate function $I(.)$ as stated in (1.35).*

Lemma 5. *The probability measures $\tilde{\mu}_n$ are exponentially tight in the space $C_0([0,1])$ of all continuous functions $f : [0,1] \to \mathbb{R}^d$ such that $f(0) = 0$, equipped with the supremum norm topology.*

We will now proof the above lemma's.

Proof of Lemma 2. First of all, in the topology induced on ξ by $\mathcal{X}$ the open sets in ξ are of the form $\mathcal{G} \cap \xi$ with $\mathcal{G} \subseteq \mathcal{X}$ open. Similarly, the closed sets in the topology are the sets of the form $\mathcal{F} \cap \xi$ with $\mathcal{F} \subseteq \mathcal{X}$ closed. Furthermore, for all $\Gamma \in \mathcal{B}$ we have $\mu_\varepsilon(\Gamma) = \mu_\varepsilon(\Gamma \cap \xi)$.
(a)Suppose that an LDP holds in ξ, which is a closed subset of $\mathcal{X}$. Extend the rate function I to be a lower semicontinuous function on $\mathcal{X}$ by setting $I(x) = \infty$ for any $x \in \xi^c$. We see that $inf_{x \in \Gamma)} I(x) = inf_{x \in \Gamma \cap \xi} I(x)$ for any $\Gamma \subset \mathcal{X}$ and the large deviations lower and upper bound holds.
(b) Suppose that an LDP holds in $\mathcal{X}$. If ξ is closed, then $\mathcal{D}_I \subset \xi$ by the large deviations lower bound because $\mu_\varepsilon(\xi^c) = 0$ for all $\varepsilon > 0$ and ξ^c is open. Thus for the lower bound to hold we need $I(x) = \infty \quad \forall x \in \xi^c$. Now, $\mathcal{D}_I \subset \xi$ impies that $inf_{x \in \Gamma)} I(x) = inf_{x \in \Gamma \cap \xi} I(x)$ for any $\Gamma \subset \mathcal{X}$ and the large deviation lower and upper bounds hold for all measurable subsets of ξ. Furthermore, since the level sets $\psi_I(\alpha)$ are closed subsets of ξ, the rate function I remains lower semicontinuous when restricted to ξ. □

proof of Lemma 3. The sets $\{\omega : ||\tilde{Z}_n - Z_n|| > \eta\}$ are clearly measurable. Note that at times t where $[nt]$ in $\mathbb{Z}$, the processes are equal. At the rest of the times $Z_n(t)$ stays the same while $\tilde{Z}_n(t)$ moves $(t - \frac{[nt]}{n})X_{[nt]+1}$. From this we see $|\tilde{Z}_n(t) - Z_n(t)| \le |X_{|nt|+1}|/n$. Thus, for any $\eta > 0$ and any $\lambda > 0$,

$$P(||\tilde{Z}_n - Z_n|| > \eta) \le nP(|X_1| > n\eta) \le nE(e^{\lambda|X_1|})e^{-\lambda n\eta}.$$

Taking logarithms of the above and dividing by n, it now follows, since $\mathcal{D}_\Lambda = \mathbb{R}^d$, by considering $n \to \infty$ and later $\lambda \to \infty$ that for any $\eta, \lambda > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log P(||\tilde{Z}_n - Z_n|| > \eta) = -\infty.$$

Therefore, the probability measures μ_n and $\tilde{\mu}_n$ are exponentially equivalent (see definition 3). □

The proof of lemma 4 relies on an important theorem, the Dawson-Gärtner Theorem, which we will state and prove at the end of the chapter. First, we need to show a finite dimensional LDP. Using the Dawson-Gärtner Theorem we can use this finite dimensional result to lift the finite dimensional LDP to an infinite dimensional LDP for the measures $\tilde{\mu}_n$. We will start with proving the following finite dimensional result.

Lemma 6. *Let J denote the collection of all ordered finite subsets of $(0,1]$. For any $j = \{0 < t_1 < t_2 < ... < t_{|j|} \leq 1\} \in J$ and any $f : [0,1] \to \mathbb{R}^d$, let $p_j(f)$ denote the vector $(f(t_1), f(t_2), ..., f(t_{|j|})) \in (\mathbb{R}^d)^{|j|}$. Then the sequence of laws $\{\mu_n \circ p_j^{-1}\}$ satisfies the LDP in $(\mathbb{R}^d)^{|j|}$ with the good rate function*

$$I_j(\boldsymbol{z}) = \sum_{l=1}^{|j|} (t_l - t_{l-1}) \Lambda^\star \left(\frac{z_l - z_{l-1}}{t_l - t_{l-1}}\right), \tag{1.37}$$

where $\boldsymbol{z} = (z_1, ..., z_{|j|})$ and $t_0 = 0$, $z_0 = 0$.

proof of lemma 6. Fix $j \in J$ and observe that $\mu_n \circ p_j^{-1}$ is the law of the random vector

$$Z_n^j := (Z_n(t_1), Z_n(t_2), ..., Z_n(t_{|j|}).$$

Let

$$Y_n^j := (Z_n(t_1), Z_n(t_2) - Z_n(t_1), ..., Z_n(t_{|j|} - Z_n(t_{|j-1|}).$$

Now the LDP for Y_n^j follows from the Gärtner-Ellis theorem (theorem 2). Let $\underline{\lambda} := (\lambda_1, ..., \lambda_{|j|})$. Since by the independence of the X_i we have

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log E[e^{n\langle \underline{\lambda}, Y_n^j\rangle}] &= \lim_{n\to\infty} \frac{1}{n} \log E(e^{n\lambda_1 Z_n(t_1)}) E(e^{n\lambda_2 (Z_n(t_2) - Z_n(t_1))}) ... E(e^{n\lambda_{|j|}(Z_n(t_{|j|}) - Z_n(t_{|j|-1}))}) \\ &= \lim_{n\to\infty} \frac{1}{n} \log E(e^{n\lambda_1 X_1})^{[nt_1]} E(e^{n\lambda_2 X_1})^{[nt_2]-[nt_1]} ... E(e^{n\lambda_{|j|} X_1})^{[nt_{|j|}]-[nt_{|j|-1}]} \\ &= \lim_{n\to\infty} \sum_{l=1}^{|j|} \frac{1}{n} ([nt_l] - [nt_{l-1}]) \Lambda(\lambda_l) = \sum_{l=1}^{|j|} (t_l - t_{l-1}) \Lambda(\lambda_l) =: \Lambda_j(\underline{\lambda}). \end{aligned}$$

By the Gärtner-Ellis theorem, the rate function of Y_n^j is given by the Fenchel-Legendre transform of the finite and differentiable function $\Lambda_j(\underline{\lambda})$. Thus we have,

$$\begin{aligned} \Lambda_j^\star(\mathbf{y}) = \sup_{\underline{\lambda} \in (\mathbb{R}^d)^{|j|}} \{\langle \underline{\lambda}, \mathbf{y}\rangle - \Lambda_j(\underline{\lambda})\} &= \sup_{\underline{\lambda} \in (\mathbb{R}^d)^{|j|}} \{\sum_{l=1}^{|j|} (\langle \lambda_l, y_l\rangle - (t_l - t_{l-1}) \Lambda(\lambda_l))\} \\ &= \sum_{l=1}^{|j|} (t_l - t_{l-1}) \sup_{\lambda_l \in \mathbb{R}^d} \{\langle \lambda_l, \frac{y_l}{t_l - t_{l-1}}\rangle - \Lambda(\lambda_l)\} \\ &= \sum_{l=1}^{|j|} (t_l - t_{l-1}) \Lambda^\star \left(\frac{y_l}{t_l - t_{l-1}}\right). \end{aligned}$$

Since the map $Y_n^j \mapsto Z_n^j$ of $(\mathbb{R}^d)^{|j|}$ onto itself is continuous and one to one, the specified LDP in (1.37) for Z_n^j follows directly by the contraction principle from an LDP for Y_n^j. □

The probability measures $\{\mu_n \circ p_j^{-1}\}$ and $\{\tilde{\mu}_n \circ p_j^{-1}\}$ are exponentially equivalent in $(\mathbb{R}^d)^{|j|}$ which follows by lemma 3. As an immediate consequence we have the following corrolary.

Corollary 2. *For any $j \in J$, $\{\tilde{\mu}_n \circ p_j^{-1}\}$ satisfies the LDP in $(\mathbb{R}^d)^{|j|}$ with the good rate function (1.37) (which follows from Theorem 3).*

We will now proof lemma 4.

Proof of Lemma 4. A partial order by inclusions is defined on J as follows. For $i, j \in J$, $i = \{s_1, ..., s_{|i|}\} \leq j = \{t_1, ..., t_{|j|}\}$ iff for any l, $s_l = t_{q(l)}$ for some $q(l)$. In other words, $i \leq j$ iff the partition of i is included in the partition of j. Then, for $i \leq j \in J$, the projection

$$p_{ij} : (\mathbb{R}^d)^{|j|} \to (\mathbb{R}^d)^{|i|}$$

is defined in the natural way. Let $\tilde{\mathcal{X}}$ denote the projective limit of $\{\mathcal{Y}_j = (\mathbb{R}^d)^{|j|}\}_{j \in J}$ with respect to the projections p_{ij} i.e., $\tilde{\mathcal{X}} = \underleftarrow{lim} \mathcal{Y}_j$. We can identify $\tilde{\mathcal{X}}$ with the space $\mathcal{X}$. Indeed each $f \in \mathcal{X}$ corresponds

to $(p_j(f))_{j\in J}$ which belongs to $\tilde{\mathcal{X}}$ since $p_i(f) = p_{ij}((p_j(f)))$ for $i \leq j \in J$. In the reverse direction, each point $\mathbf{x} = (x_j)_{j\in J}$ of $\tilde{\mathcal{X}}$ may be identified with the map $f : [0, 1] \to \mathbb{R}^d$, where $f(t) = x_{(t)}$ for $t > 0$ and $f(0) = 0$. Furthermore, the projective topology on $\tilde{\mathcal{X}}$ coincides with the pointwise convergence topology of $\mathcal{X}$, and p_j as defined in lemma 6 are the canonical projections for $\tilde{\mathcal{X}}$. The LDP for $\{\tilde{\mu}_n\}$ in the Hausdorff topological space $\mathcal{X}$ follows by applying the Dawson-Gärtner theorem in conjunction with corollary 2. (Note that $(\mathbb{R}^d)^{|j|}$ are Hausdorff spaces and I_j are good rate functions.)
The rate function governing this LDP is

$$I_X(f) = \sup_{\substack{0=t_0<t_1<t_2<...<t_k\leq 1\\ k\in\mathbb{N}}} \sum_{l=1}^{k}(t_l - t_{l-1})\Lambda^\star(\frac{f(t_l)-f(t_{l-1})}{t_l - t_{l-1}}). \tag{1.38}$$

Since $\Lambda^\star$ is nonnegative, without loss of generality, assume now that $t_k = 1$. It remains to be shown that $I_X(.) = I(.)$. First we show $I(\phi) \geq I_X(\phi)$. Note that we can see

$$\frac{1}{t_i - t_{i-1}}\int_{t_{i-1}}^{t_i} \Lambda^\star(\dot{\phi}(t))dt$$

as the expectation of the stochast $\Lambda^\star(\dot{\phi}(t))$ with uniform density. Now the convexity of $\Lambda^\star$ implies by Jensen's inequality

$$\frac{1}{t_i - t_{i-1}}\int_{t_{i-1}}^{t_i} \Lambda^\star(\dot{\phi}(t))dt \geq \Lambda^\star(\frac{1}{t_i - t_{i-1}}\int_{t_{i-1}}^{i} \dot{\phi}(t)dt) = \Lambda^\star(\frac{\phi(t_i)-\phi(t_{i-1})}{t_i - t_{i-1}})$$

Now taking the sum for $l = 1$ to k such that $t_k = 1$ and taking the supremum over all partitions we get $I(\phi) \geq I_X(\phi)$. For the reversed inequality, consider $\phi \in \mathcal{AC}$. Let $g(t) := d\phi(t)/dt \in L_1[0, 1]$ and, for $k \geq 1$, define

$$g^k(t) := k\int_{([kt]-1)/k}^{[kt]/k} g(s)ds \quad t \in [0, 1), \quad g^k(1) = k\int_{1-1/k}^{1} g(s)ds.$$

Now we can look at (1.38) taking only values of $t_i = i/k$ (where we consider k points of time with $t_k = 1$). So we have $t_l - t_{l-1} = 1/k$. Observe that we have

$$\begin{aligned} I_X(\phi) \geq \liminf_{k\to\infty}\sum_{l=1}^{k}\frac{1}{k}\Lambda^\star(k[\phi(\frac{l}{k}) - \phi(\frac{l-1}{k})]) &= \liminf_{k\to\infty}\sum_{l=1}^{k}\frac{1}{k}\Lambda^\star(g^k(\frac{l}{k})) \\ &= \liminf_{k\to\infty}\int_0^1 \Lambda^\star(g^k(t))dt. \end{aligned} \tag{1.39}$$

By Lebesgue's theorem we have $\lim_{k\to\infty} g^k(t) = g(t)$ almost everywhere in $[0, 1]$. Hence, by Fatou's lemma and the lower semicontinuity of $\Lambda^\star(.)$,

$$\begin{aligned} \liminf_{k\to\infty}\int_0^1 \Lambda^\star(g^k(t))dt &\geq \int_0^1 \liminf_{k\to\infty}\Lambda^\star(g^k(t))dt \\ &\geq \int_0^1 \Lambda^\star(g(t))dt = I(\phi). \end{aligned} \tag{1.40}$$

From which we find, combining (1.39) and (1.40), $I_X(\phi) \geq I(\phi)$.

Finally, suppose that $\phi \in \mathcal{X}$ and $\phi \notin \mathcal{AC}$. We need to show $I_X = \infty$. In this situation there exist a $\delta > 0$ and $\{s_1^n < t_1^n \leq ... \leq s_{k_n}^n < t_{k_n}^n\}$ such that $\sum_{l=1}^{k_n}(t_l^n - s_l^n) \to 0$, while $\sum_{l=1}^{k_n}|\phi(t_l^n) - \phi(s_l^n)| \geq \delta$. Note that since $\Lambda^\star$ is nonnegative,

$$I_X(\phi) = \sup_{\substack{0<t_1<t_2<\ldots<t_k \\ \lambda_1,\ldots,\lambda_k\in\mathbb{R}^d}} \sum_{l=1}^{k} [\langle\lambda_l, \phi(t_l)-\phi(t_{l-1})\rangle - (t_l-t_{l-1})\Lambda(\lambda_l)]$$

$$\geq \sup_{\substack{0\leq s_1<t_1\leq s_2<t_2\leq\ldots\leq s_k<t_k \\ \lambda_1,\ldots,\lambda_k\in\mathbb{R}^d}} \sum_{l=1}^{k} [\langle\lambda_l, \phi(t_l)-\phi(s_l)\rangle - (t_l-s_l)\Lambda(\lambda_l)].$$

Hence, for $t_l = t_l^n$, $s_l = s_l^n$, and λ_l proportional to $\phi(t_l)-\phi(s_l)$ and with $|\lambda_l| = \rho$, the following bound is obtained:

$$I_X(\phi) \geq \limsup_{n\to\infty}\{\rho\sum_{l=1}^{k_n}|\phi(t_l^n)-\phi(s_l^n)| - [\sup_{|\lambda|=\rho}\Lambda(\lambda)]\sum_{l=1}^{k_n}(t_l^n-s_l^n)\} \geq \rho\delta.$$

(Recall that $\Lambda(.)$ is continuous everywhere.) The arbitrariness of ρ implies that in this situation $I_X(\phi) = \infty$, completing the proof of lemma 4. □

It remains to prove lemma 5. The proof relies on the following one-dimensional result.

Lemma 7. *Let X be a real valued random variable distributed according to the law ν. Then $E[e^{\delta\Lambda_\nu^\star(X)}] < \infty$ for all $\delta < 1$.*

The proof can be found in [1], lemma 5.1.14 on p.181.

Proof of Lemma 5. To show the exponential tightness of $\tilde{\mu}_n$ in $C_0([0, 1])$ equipped with the supremum norm topology, denote X_1^j the jth component of X_1, define

$$\Lambda_j(\lambda) := \log(E[\exp(\lambda X_1^j)]),$$

with $\Lambda_j^\star(.)$ the Fenchel-Legendre transform of $\Lambda_j(.)$. Fix $\alpha > 0$ and

$$K_\alpha^j := \{f \in \mathcal{AC} : f(0) = 0, \int_0^1 \Lambda_j^\star(\dot{f}_j(\theta))d\theta \leq \alpha\},$$

where $f_j(.)$ is the jth component of $f : [0, 1] \to \mathbb{R}^d$. Now let $K_\alpha := \cap_{j=1}^d K_\alpha^j$. Note that $d\tilde{Z}_n(t)/dt = X_{[nt]+1}$ for almost all $t \in [0, 1)$. Thus,

$$\tilde{\mu}_n(K_\alpha^c) \leq d\max_{j=1}^{d} P(\frac{1}{n}\sum_{i=1}^{n}\Lambda_j^\star(X_i^j) > \alpha).$$

Since $\{X_i\}_{i=1}^n$ are independent, it now follows by Chebycheff's inequality that for any $\delta > 0$,

$$\frac{1}{n}\log\tilde{\mu}_n(K_\alpha^c) \leq -\delta\alpha + \frac{1}{n}\log d + \max_{j=1}^{d}\log E[e^{\delta\Lambda_j^\star(X_1^j)}].$$

It follows by considering $\delta = \frac{1}{2}$ and $\alpha \to \infty$ that, because we know $E[e^{\delta\Lambda_j^\star(X_1^j)}] < \infty$, that $\lim_{\alpha\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\tilde{\mu}_n(K_\alpha^c) = -\infty$. So, for every $\alpha < \infty$ there exists a set $K_\alpha \subset \mathcal{X}$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\log\tilde{\mu}_n(K_\alpha^c) < \alpha$$

It remains to show K_α is compact. By the Arzeli-Ascoli theorem this follows if K_α is a bounded set of equicontinuous functions. For the equicontinuity, notice that if $f \in K_\alpha$, then the continuous function f is differentiable almost everywhere, and for all $0 \leq s < t \leq 1$ and $j = 1, 2, ..., d$,

$$\Lambda_j^\star(\frac{f_j(t)-f_j(s)}{t-s}) \leq \frac{1}{t-s}\int_s^t \Lambda_j^\star(\dot{f_j}(\theta))d\theta \leq \frac{\alpha}{t-s}.$$

Since $\Lambda_j^\star(x) \geq M|x| - \{\Lambda_j(M) \vee \Lambda_j(-M)\}$ for all $M > 0$, it follows that for all $(t-s) \leq \delta$,

$$|f_j(t) - f_j(s)| \leq \frac{1}{M}(\alpha + \delta\{\Lambda_j(M) \vee \Lambda_j(-M)\}). \tag{1.41}$$

Since $\Lambda_j(.)$ is continuous on $\mathbb{R}$, there exist $M_j = M_j(\delta)$ such that $\Lambda_j(M_j) \leq 1/\delta$ and $\Lambda_j(-M_j) \leq 1/\delta$, and $lim_{\delta\to 0} M_j(\delta) = \infty$. Hence, $\varepsilon(\delta) := max_{j=1,\ldots,d}(\alpha+1)/M_j(\delta)$ is a uniform modulus of continuity for the set K_α. Finally, K_α is bounded by regarding equation (1.41) and taking for instance $s = 0$ and $\delta = 1$. □

Having proved the lemma's 2-5 we can now prove the theorem which we set out to prove in the beginning, namely Mogulskii's theorem.

Proof of Mogulskii's Theorem. From lemma 4 we know that the $\{\tilde{\mu}_n\}$ satisfies the LDP in $\mathcal{X}$. The domain $\mathcal{D}_I \subset C_0([0,1])$, and $\tilde{\mu}_n(C_0([0,1])) = 1$ for all n. It follows now from lemma 2 that the LDP for $\{\tilde{\mu}_n\}$ also holds in the space $C_0([0,1])$ equipped with the topology induced by $\mathcal{X}$. This topology is the pointwise convergence topology. It is generated by the sets $V_{t,x,\delta} := \{g \in C_0([0,1]) : |g(t) - x| < \delta\}$ with $t \in (0,1]$, $x \in \mathbb{R}^d$ and $\delta > 0$. Each $V_{t,x,\delta}$ is an open set under the supremum norm, i.e. the supremum norm topology is a stronger topology than the pointwise convergence topology. Now the exponential tightness of the $\{\tilde{\mu}_n\}$ in $C_0([0,1])$ equipped with the supremum norm topology (by lemma 5) together with corollary 1 allows the strengthening of the LDP on $C_0([0,1])$ equipped with the supremum norm topology. Because $C_0([0,1])$ is a closed subset of $L_\infty([0,1])$ this same LDP holds also in $L_\infty([0,1])$ by using again lemma 2 (now in the opposite direction). The LDP for $\{\mu_n\}$ in the metric space $L_\infty([0,1])$ follows from the one for $\{\tilde{\mu}_n\}$ by using the exponential equivalentness of the two sets of measures and using theorem 3. □

Mogulskii's Theorem can be extended to the laws ν_ε of

$$Y_\varepsilon(t) = \varepsilon \sum_{i=1}^{[\frac{t}{\varepsilon}]} X_i, \quad 0 \leq t \leq 1. \tag{1.42}$$

Now $Z_n(t)$ and $\mu_n(t)$ correspond to the special case where $\varepsilon = n^{-1}$. The extension can be proved by showing that the measures of $\mu_n(t)$ and ν_ε are exponentially equivalent and using theorem 3. For the full proof see [1], p.183.

1.3.2 Large Deviations of Brownian Motion Paths

A well-known stochastic process is of course the standard Brownian Motion. For this process we can look for an LDP as well. We consider a sequence of Brownian motions scaled with parameter ε in order to get a converging sequence when $\varepsilon \to 0$. Let $\{w_t, t \in [0,1]\}$ denote a standard Brownian motion in $\mathbb{R}^d$. Now the scaled process is defined as follows:

$$w_\varepsilon(t) = \sqrt{\varepsilon} w_t \tag{1.43}$$

where we let v_ε be the probability measures induced by $w_\varepsilon(t)$ on $C_0([0,1])$. It can be shown that an exponential equivalent process to $w_\varepsilon(t)$ is the process $Y_\varepsilon(t)$ in (1.42) for a particular choice of the X_i. But first we state the theorem below for the LDP of the process stated in (1.43).

Let $H_1 := \{\int_0^1 f(s)ds : f \in L_2([0,1])\}$ denote the space of all absolutely continuous functions with square integrable derivative equipped with the norm $||g||_{H_1} = [\int_0^1 |\dot{g}(s)|^2 ds]^{\frac{1}{2}}$. The following theorem holds.

Theorem 5 (Schilder). $\{\nu_\varepsilon\}$ *satisfies, in* $C_0([0,1])$, *an LDP with the good rate function*

$$I_w(\phi) = \begin{cases} \frac{1}{2}\int_0^1 |\dot{\phi}(t)|^2 dt, & \phi \in H_1 \\ \infty & otherwise. \end{cases} \tag{1.44}$$

Proof. Define the process

$$\hat{w}_\varepsilon(t) := w_\varepsilon(\varepsilon[\frac{t}{\varepsilon}])$$

Note that if we let $\{X_i\}$ be a collection of i.i.d. $\mathcal{N}(0, 1)$ random variables, we have for $t \in [0, 1]$ and $\varepsilon > 0$,

$$\hat{w}_\varepsilon(t) = \sqrt{\varepsilon}w(\varepsilon[\frac{t}{\varepsilon}]) = \varepsilon\mathcal{N}(0, [\frac{t}{\varepsilon}]) = \varepsilon\sum_{i=1}^{[\frac{t}{\varepsilon}]} X_i.$$

So this is merely the process $Y_\varepsilon(.)$ defined in (1.42) where the X_i are standard normally distributed. So from the above we know now that the probability measures corresponding to $\hat{w}_\varepsilon(.)$ satisfy the LDP in $L_\infty([0, 1])$ with the good rate function $I(.)$ stated in Theorem 4. We have, for the standard normal variables considered here,

$$\Lambda(\lambda) = \log E[e^{\langle\lambda, X_1\rangle}] = \frac{1}{2}|\lambda|^2,$$

from which it follows that, for each $x \in \mathbb{R}^d$

$$\Lambda^\star(x) = \sup_{\lambda\in\mathbb{R}^d}\{\langle\lambda, x\rangle - \frac{1}{2}|\lambda|^2\}$$

This supremum is easy to compute by looking at each component separately . In this case we want to find $\sup_{\lambda\in\mathbb{R}}\{\lambda x - \frac{1}{2}\lambda^2\}$. Taking the derivative with respect to λ and equating to zero gives us $\lambda = x$. As this holds for every component we can write now,

$$\Lambda^\star(x) = \frac{1}{2}|x|^2.$$

Hence, from this it follows that $\mathcal{D}_I = H_1$ (square integrable derivative), and the rate function $I(.)$ of Theorem 4 specializes to $I_w(.)$ of (1.44). It remains to show the exponential equivalentness of the measures of $\hat{w}_\varepsilon$ and w_ε as then we have the LDP for w_ε by Theorem 3 in $L_\infty([0, 1])$. Now this can be restricted to $C_0([0, 1])$ by lemma 2 as $w_\varepsilon(.) \in C_0([0, 1])$ with probability one.

For the exponential equivalence of the measures we look at $\mathbb{P}(||w_\varepsilon - \hat{w}_\varepsilon|| \geq \delta)$, $\delta > 0$. Observe that for $t \in \{k\varepsilon : k \in \mathbb{N}\}$ the two processes are equal. For this we only have to consider $t \in [0, \varepsilon]$. Furthermore, $\hat{w}_\varepsilon$ stays constant in this interval of time and starts at the same point as w_ε. As $t \leq 1$ we have at most $([1/\varepsilon] + 1)$ intervals to consider, and hence we can write for any $\delta > 0$,

$$\mathbb{P}(||w_\varepsilon - \hat{w}_\varepsilon|| \geq \delta) \leq ([1/\varepsilon] + 1)\mathbb{P}(\sup_{0\leq t\leq\varepsilon}|w_\varepsilon(t)| \geq \delta)$$

Next we make use of an inequality we will state here but leave the proof to the reader (it also can be found in [1], p.185). We have that for any integer d and any $\tau, \varepsilon, \delta > 0$,

$$\mathbb{P}(\sup_{0\leq t\leq\tau}|w_\varepsilon(t)| \geq \delta) \leq 4de^{-\delta^2/2d\tau\varepsilon}. \tag{1.45}$$

Using the above inequality (1.45) we find the following,

$$\mathbb{P}(||w_\varepsilon - \hat{w}_\varepsilon|| \geq \delta) \leq 4d(\varepsilon^{-1} + 1)e^{-\delta^2/2d\varepsilon^2}$$

Taking the logarithm, multiplying by ε and taking the limit supremum for $\varepsilon \to 0$ gives us now,

$$\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{P}(||w_\varepsilon - \hat{w}_\varepsilon|| \geq \delta) = -\infty.$$

This concludes the proof of the Schilder Theorem. □

1.4 Projective limits and the Dawson Gärtner Theorem

Projective limits in the context of Large Deviations are used to lift a collection of LDPs in "small" finite dimensional spaces into the LDP of some "large" space $\mathcal{X}$. The idea is to identify $\mathcal{X}$ with the projective limit of a family of spaces $\{\mathcal{Y}_j\}_{j\in J}$. Now it seems likely that the LDP will hold for a given family $\{\mu_\varepsilon\}$ of probability measures on the space $\mathcal{X}$ if the LDP holds for any projection of μ_ε to $\{\mathcal{Y}_j\}_{j\in J}$. We will see this is indeed the case and it is stated in the Dawson Gärtner Theorem. The Dawson Gärtner Theorem is also proved below.

We will first define the notion of a projective limit. Let $(J,\leq)$ be a partially ordered, possibly uncountable, right filtered set (the latter means that for any $i,j\in J$ there exist a $k\in J$ such that both $i\leq k$ and $j\leq k$). A projective system $(\mathcal{Y}_j, p_{ij})_{i\leq j\in J}$ consists of Hausdorff topological spaces $\{\mathcal{Y}_j\}_{j\in J}$ and continuous maps $p_{ij}:\mathcal{Y}_j\to\mathcal{Y}_i$ such that for $i\leq j\leq k$, $p_{ik}=p_{ij}\circ p_{jk}$. The projective limit of this system, denoted by $\mathcal{X}=\underleftarrow{lim}\mathcal{Y}_j$, is the subset of the topological space $\mathcal{Y}=\prod_{j\in J}\mathcal{Y}_j$ consisting of all the elements $\mathbf{x}=(y_j)_{j\in J}$ for which $y_i=p_{ij}(y_j)$ whenever $i\leq j$ equipped with the topology induced by $\mathcal{Y}$. We already encountered projective limits in the proof of lemma 4 and lemma 6. In this case we wrote the space $\mathbb{X}$ of all functions $f:[0,1]\to\mathbb{R}^d$ with $f(0)=0$ equipped with the pointwise convergence topology as the projective limit of $\{(\mathbb{R}^d)^{|j|}\}_{j\in J}$. With $(\mathbb{R}^d)^{|j|}$ we can imagine that we take $|j|$ time values from a fucntion mapping to $\mathbb{R}^d$. As stated in the proof of lemma 4 we can define a partial order and the projections p_{ij} in a natural way such that the projective limit of $\{(\mathbb{R}^d)^{|j|}\}_{j\in J}$ exactly yields the space $\mathcal{X}$.

Now we will state the Contraction Princicple. This Theorem shows that the LDP is preserved under a continuous mapping $f:\mathcal{X}\to\mathcal{Y}$ between two Hausdorf topological spaces $\mathcal{X}$ and $\mathcal{Y}$. This Theorem is used in the proof of Lemma 6, furthermore it is used to proof the Dawson Gärtner Theorem below.

Theorem 6 (Contraction Principle). *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorf topological spaces and $f:\mathcal{X}\to\mathcal{Y}$ a continuous function. Consider a good rate function $I:\mathcal{X}\to[0,\infty]$.*
(a) For each $y\in\mathcal{Y}$, define

$$I'(y):=\inf\{I(x):x\in\mathcal{X},y=f(x)\}. \tag{1.46}$$

Then I' is a good rate function on $\mathcal{Y}$, where as usual the infimum over the empty set is taken as ∞.
(b) If I controls the LDP associated with a family of probability measures $\{\mu_\varepsilon\}$ on $\mathcal{X}$, then I' controls the LDP associated with the family of probability measures $\{\mu_\varepsilon\circ f^{-1}\}$ on $\mathcal{Y}$.

Part (b) of the Contraction Principle can be used in the proof of the Dawson-Gärtner Theorem which we will state and prove now. The proof of the Contraction principle can be found in [1], p.126.

Theorem 7 (Dawson-Gärtner Theorem). *Let μ_ε be a family of probability measures on $\mathcal{X}$, such that for any $j\in J$ the Borel probability measures on $\mu_\varepsilon\circ p_j^{-1}$ on $\mathcal{Y}_j$ satisfy the LDP with good rate function $I_j(.)$. Then $\{\mu_\varepsilon\}$ satisfies the LDP with good rate function*

$$I(\mathbf{x})=\sup_{j\in J}\{I_j(p_j(\mathbf{x}))\},\quad \mathbf{x}\in\mathcal{X}. \tag{1.47}$$

Proof of Dawson-Gärtner Theorem. As $I_j(.)$ is nonnegative, clearly $I(\mathbf{x})$ is nonnegative. For any $\alpha\in[0,\infty)$ and $j\in J$, let $\psi_{I_j}(\alpha)$ denote the compact level set of I_j, i.e. $\psi_{I_j}(\alpha):=\{y_j:I_j(y_j)\leq\alpha\}$. Note that for any $i\leq j\in J$, $p_{ij}:\mathcal{Y}_j\to\mathcal{Y}_i$, is a continuous map and $\mu_\varepsilon\circ p_i^{-1}=(\mu_i\circ p_j^{-1})\circ p_{ij}^{-1}$. Now we can, with the help of the Contraction Principle, relate the rate functions I_i and I_j. namely, by the Contraction Principle, the rate function I_i in a point y_i is the infimum of the rate function I_j taken over all the points y_j that map to y_i, i.e. $I_i(y_i)=inf_{y_j\in p_{ij}^{-1}(y_i)}I_j(y_j)$. Or, equivalently, we can write $\psi_{I_i}(\alpha)=p_{ij}(\psi_{I_j}(\alpha))$. Therefore,

$$\psi_j(\alpha)=\mathcal{X}\cap\prod_{j\in J}\psi_{I_j}(\alpha)=\underleftarrow{\lim}\psi_{I_j}(\alpha). \tag{1.48}$$

And $I(x)$ is a good rate function as the projective limit of compact subsets of $\mathcal{Y}_j$, $j \in J$, is a compact subset of $\mathcal{X}$ (by Tychonoff's Theorem).

Now that we know $I(x)$ is a good rate function we will show the bounds hold. First, to show the lower bound, it suffices to show that for every measurable set $\mathbf{A} \subset \mathcal{X}$ and each $\mathbf{x} \in \mathbf{A}^\circ$, there exists a $j \in J$ such that

$$\liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(\mathbf{A}) \geq -I_j(p_j(\mathbf{x})).$$

Because then surely ,

$$\liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(\mathbf{A}) \geq -\sup_{j \in J}\{I_j(p_j(\mathbf{x}))\} = I(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbf{A}^\circ.$$

Since the collection $\{p_j^{-1}(U_j) : U_j \subset \mathcal{Y}_j \textit{ is open}\}$ is a base of the topology of $\mathcal{X}$, there exists some $j \in J$ and an open set $U_j \in \mathcal{Y}_j$ such that $\mathbf{X} \in p_j^{-1}(U_j) \subset \mathbf{A}^\circ$. Thus, by using the large deviations lower bound for $\{\mu_\varepsilon \circ p_j^{-1}\}$ (second inequality below),

$$\begin{aligned}\liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(\mathbf{A}) &\geq \liminf_{\varepsilon \to 0} \varepsilon \log(\mu_\varepsilon \circ p_j^{-1}(U_j)) \\ &\geq -\inf_{y \in U_j} I_j(y) \geq -I_j(p_j(\mathbf{x})),\end{aligned}$$

as desired.

Considering the large deviations upper bound, fix a measurable set $\mathbf{A} \subset \mathcal{X}$ and let $\mathbf{A}_j := p_j(\overline{\mathbf{A}})$. Then $\mathbf{A}_i = p_{ij}(\mathbf{A}_j)$ for any $i \leq j$, implying that $p_{ij}(\overline{\mathbf{A}}_j) \subseteq \overline{\mathbf{A}}_i$ (since the p_ij are continuous). Hence, $\overline{\mathbf{A}} \subseteq \underleftarrow{lim}\ \overline{\mathbf{A}}_j$. To prove the inverse inclusion, assume $\mathbf{x}$ in not an element of $\overline{\mathbf{A}}$, i.e. $\mathbf{x} \in (\overline{\mathbf{A}})^c$. We need to show that now also $p_j(\mathbf{x}) \notin \overline{\mathbf{A}}_j$ for some $j \in J$. Since $(\overline{\mathbf{A}})^c$ is an open subset of $\mathcal{X}$, there exists some $j \in J$ and an open set $U_j \subseteq \mathcal{Y}_j$ such that $\mathbf{x} \in p_j{-}1(U_j) \subseteq (\overline{\mathbf{A}})^c$. Hence, for this value of j, $p_j(\mathbf{x}) \in U_j \subseteq \overline{\mathbf{A}}_j^c$, implying that $p_j(\mathbf{x}) \notin \overline{\mathbf{A}}_j$.
From the above two inclusions it follows $\overline{\mathbf{A}} = \underleftarrow{lim}\,\overline{\mathbf{A}}_j$.

Combining this identity with (1.48), it follows that for every $\alpha < \infty$,

$$\overline{\mathbf{A}} \cap \psi_I(\alpha) = \underleftarrow{lim}(\overline{\mathbf{A}}_j \cap \psi_{I_j}(\alpha)).$$

Now fix $\alpha < inf_{\mathbf{x} \in \overline{\mathbf{A}}} I(\mathbf{x})$, for which $\overline{\mathbf{A}} \cap \psi_I(\alpha) = \emptyset$. Now it also must hold that $\overline{\mathbf{A}}_j \cap \psi_{I_j}(\alpha) = \emptyset$ for some $j \in J$ (by Theorem B4 in [1], p.346). Therefore, as $A \subseteq p_j^{-1}(\overline{\mathbf{A}}_j)$, by the LDP upper bound associated with the Borel measures $\{\mu_\varepsilon \circ p_j^{-1}\}$,

$$\limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(\mathbf{A}) \leq \limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon \circ p_j^{-1}(\overline{\mathbf{A}}_j) \leq -\inf_{\mathbf{x} \in \overline{\mathbf{A}}_j} I_j(\mathbf{x}) \leq -\alpha.$$

This inequality holds for every measurable $\mathbf{A}$ and $\alpha < \infty$ such that $\overline{\mathbf{A}} \cap \psi_I(\alpha) = \emptyset$. Note that $\overline{\mathbf{A}} \cap \psi_I(\alpha) = \emptyset$ means that $\overline{\mathbf{A}} \subseteq (\psi_I(\alpha))^c$. We see from (1.6) that this yields the LDP upper bound for $\{\mu_\varepsilon\}$. Which proves the Dawson-Gärtner Theorem. □

1.5 Short Summary and Conclusion

We have seen that for many different measures often a Large Deviation Principle (LDP) exists. This principle describes bounds using a so called rate function for the measures μ_n or more generally μ_ε if $n \to \infty$ or $\varepsilon \to 0$ respectively. In the case of finite dimensions we have looked at the LDP of the empirical mean of independently an identically distributed random variables. Here it was seen that the rate function was equal to the Fenchel-Legendre transform of the logarithmic moment generating function of the stochastic variables. Later, with Gärtner-Ellis theorem, it was shown that it is not needed for the variables to be i.i.d. but that whenever a less restrictive condition (assumption 1) for the measures holds the above LDP holds. After this the large deviation behaviour of stochastic processes was analysed with as main result the Theorem of Mogulski. This theorem deals with the process,

$$Z_n(t) = \frac{1}{n}\sum_{i=1}^{|nt|} X_i, \; 0 \leq t \leq 1$$

Where the X_i are again assumed to be i.i.d.. In this case rate functions are functionals that have a function ϕ as argument. Mogulski's theorem states that he rate function $I(\phi)$ that describes the LDP for the measures corresponding to the process of Z_n is the integral with respect to time over the Fenchel-Legendre transform of the logarithmic moment generating function of X_i evaluated at the time derivative of $\phi(t)$. The proof of Mogulski relies on the finite dimensional result using projective limits. Namely, first the LDP can be derived for the sequence of laws $\{\tilde{\mu}_n \circ p_j^{-1}\}$, where p_j projects any function $f : [0, 1] \to \mathbb{R}^d$ to a finite vector of elements of this function evaluated at a finitely ordered subset of time points. Subsequently the LDP can be shown to hold for $\tilde{\mu}_n$ on the space of all functions $f : [0, 1] \to \mathbb{R}^d$ by applying the Dawson-Gärtner theorem. The Mogulski Theorem is later used for the proof of the Schilder theorem which deals with the LDP for (scaled) Brownian motion paths. It is shown that a process, exponentially equivalent to the scaled Brownian motion process, is a special case of the process where Mogulski Theorem deals with.

In the chapter below the connection is shown between Lagrangian and Hamiltonian dynamics and Large Deviation Theory. We have seen that the rate function for stochastic processes has an integral form. As it turns out, this rate function can be identified with the so called Action integral in Lagrangian dynamics. In large deviation theory we looked at the infimum of this rate function over all functions in some set we are looking at (for instance all paths starting at a point a end ending in a point b). This can be identified with the Principle of least action in the context of Lagrangian dynamics as will be discussed below.

Chapter 2

Langrangian and Hamiltonian dynamics and the connection with LDT

2.1 Introduction

Langrangian and Hamilton dynamics are different ways to look at the famous equation of motion given by Newton's second law $F = ma$ and do mechanics. As we know, Mechanics deals with the motion of particles. A big advantage of the Lagrangian and Hamiltonian formalism is that the equations of motion hold in any coordinate system (not only in inertial frames of reference). As was briefly mentioned, there is a nice connection between Lagrangian and Hamiltonian dynamics and Large Deviation Theory. This will be used in our final chapter to analyse the Large Deviation behaviour of a particular stochastic process. This process is given by the empirical distribution of n independent copies of a continuous-time Markov chain with two states. The first part of this chapter follows to some extent the book Classical Dynamics by David Tong, later referred to by [3]. First a very brief introduction in Newtonian mechanics is given in order to define the basic quantities used in the rest of this chapter and to show how the Lagrangian and Hamiltonian equations arise from Newton's equation. After this the Lagrangian and Hamiltonian formalism respectively are explained. In section 2.3 the connection between Lagrangian and Hamiltonian dynamics on the one hand and Large Deviation Theory on the other hand is explained and clarified using examples with the Brownian motion process.

2.2 Newtonian Mechanics, a very brief introduction

Lets look at a system of N particles. Particle i has mass m_i and position $\mathbf{r}_i$. Newton's Law now reads:

$$\mathbf{F}_i = \dot{\mathbf{p}}_i. \tag{2.1}$$

Where $\mathbf{p}_i = m\dot{\mathbf{r}}_i$ is the momentum of the ith particle. We assume that m doesn't depend on t. The total kinetic energy of this system of particles is $T = \frac{1}{2}\sum_i m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i$. We can write the time derivative of T as follows $\frac{dT}{dt} = \sum_i m_i \dot{\mathbf{r}}_i \ddot{\mathbf{r}}_i = \sum_i (\mathbf{F}_i^{ext} + \sum_{j\neq i} \mathbf{F}_{ij}) \cdot \dot{\mathbf{r}}_i$[1]. The change in total kinetic energy in the system of particles between times t_1 and t_2 is given by,

$$T(t_2) - T(t_1) = \int_{t_1}^{t_2} \frac{dT}{dt}\, dt = \int_{t_1}^{t_2} \sum_i (\mathbf{F}_i^{ext} + \sum_{j\neq i} \mathbf{F}_{ij}) \cdot \dot{\mathbf{r}}_i\, dt = \sum_i \int_{r_1}^{r_2} \mathbf{F}_i^{ext} \cdot d\mathbf{r} + \sum_i \sum_{j\neq i} \int_{r_1}^{r_2} \mathbf{F}_{ij} \cdot d\mathbf{r} \tag{2.2}$$

We will, in this chapter, assume we deal with conservative forces. We show that then necessarily the total energy of the system is conserved. Note that for conservative forces we can write,

[1] Here we split up the force in an external force $\mathbf{F}_i^{ext}$ and forces of the particles in the system that are acting upon each other

$$\begin{aligned}\mathbf{F}_i^{ext} &= -\nabla_i V_i(\mathbf{r}_1, ..., \mathbf{r}_N)\\ \mathbf{F}_{ij} &= -\nabla_i V_{ij}(\mathbf{r}_1, ..., \mathbf{r}_N)\end{aligned} \tag{2.3}$$

for some potential V_i and internal potentials V_{ij}. To get Newton's third law $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$ to hold togehter with the requirement that these forces are parallel with the vector $(\mathbf{r}_i - \mathbf{r}_j)$, the internal potentials should satisfy $V_{ij} = V_{ji}$ and $V_{ij} = V_{ij}(|\mathbf{r}_i - \mathbf{r}_j|)$. Where the latter means that the internal potential V_{ij} only depends on the distance between the ith and the jth particle. Furthermore we restrict $V_i(\mathbf{r}_1, ..., \mathbf{r}_N) = V_i(\mathbf{r}_i)$, which means that the external force on the ith particle does not depend on the positions of the other particles. Using this and filling (2.3) in (2,2), we get,

$$\begin{aligned}T(t_2) - T(t_1) = \sum_i \int_{r_1}^{r_2} -\nabla_i V_i(\mathbf{r}_i) \cdot d\mathbf{r} + \sum_i \sum_{j \neq i} \int_{r_1}^{r_2} -\nabla_i V_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) \cdot d\mathbf{r} = \\ -\sum_i (V_i(t_2) - V_i(t_1)) - \sum_i \sum_{j \neq i} (V_{ij}(t_2) - V_{ij}(t_1)) = \\ -(\sum_i V_i(t_2) + \sum_i \sum_{j \neq i} V_{ij}(t_2)) + (\sum_i V_i(t_1) + \sum_i \sum_{j \neq i} V_{ij}(t_1)) = -V(t_2) + V(t_1)\end{aligned}$$

From which it follows that,

$$V(t_1) + T(t_1) = V(t_2) + T(t_2) \equiv E \tag{2.4}$$

Where E is the energy of the system, which we see here is a conserved quantity. This is basic Newtonian mechanics. When all forces are known the paths can be calculated. In the next section we will explain Langrangian mechanics and the Principle of least action. Here the vector notation is changed in favour of a more general coordinate system.

2.3 Langrangian and Hamiltonian Mechanics

2.3.1 The Lagrangian Formalism

As in the above, let us assume we have a system of N particles with coordinates r_i for the ith particle. We can rewrite the positions of these N particles by taking $3N$ coordinates x_i with $i \in \{1, 2, .., 3N\}$. Now Newton's equation read,

$$\dot{p}_i = -\frac{\partial V}{\partial x_i} \tag{2.5}$$

Where $p_i = m_i \dot{x}_i$. This can be associated with a 3N dimensional space known as configuration space C. One point in C specifies all positions of the N particles and a path in C corresponds with the dynamics of this system of particles. Now we define the Langrangian $L(x_i, \dot{x}_i)$ of this system of particles, which is a function of the positions x_i and velocities $\dot{x}_i$.

Definition 5 (Langrangian)**.**

$$L(x_i, \dot{x}_i) = T(\dot{x}_i) - V(x_i) \tag{2.6}$$

Where as in the above $T(\dot{x}_i)$ is the kinetic energy of x_i and $V(x_i)$ the potential energy of x_i. Langrangian mechanics now makes use of the so called Principle of Least Action. This is equivalent to Newton's equations but another way of looking at this. Namely, we will see that the path taken by the system of particles is the one that minimizes the action S (as defined below). The action can be seen as a cost calculated for each path, the path with the lowest cost is the path that will be taken by the system[2].

[2] It does not necessarily have to be the minimum cost, but an extremum. In practice it often is a minimum hence the name Principle of Least Action

Definition 6 (Action)**.** *Let t_i and t_f be the initial and final time respectively. Let $x_i(t_i) := x_i^{initial}$ be the initial position and $x_i t_f := x_i^{final}$ the final position. The Action of the system in this time period for the paths $x^A(t)$ is given by,*

$$S[x_i(t)] = \int_{t_i}^{t_f} L(x_i(t), \dot{x}_i(t))dt. \tag{2.7}$$

Theorem 8 (Principle of Least Action)**.** *The actual path taken by x_i is an extremum of S.*

Proof. Consider the path $x_i(t)$ starting at $x_i^{initial}$ and ending at x_i^{final}. We consider now the slightly varied path, $x_i(t) \to x_i(t) + \delta x_i(t)$. But we fix the endpoints so that $\delta x_i(t_i) = \delta x_i(t_f) = 0$. Now the change in action is,

$$\begin{aligned}
\delta S = \delta[\int_{t_i}^{t_f} L dt] &= \int_{t_i}^{t_f} \delta L dt \\
&= \int_{t_i}^{t_f} \frac{\partial L}{\partial x_i}\delta x_i + \frac{\partial L}{\partial \dot{x}_i}\delta \dot{x}_i dt \\
&= \int_{t_i}^{t_f} (\frac{\partial L}{\partial x_i} - \frac{d}{dt}(\frac{\partial L}{\partial \dot{x}_i}))\delta x_i dt + [\frac{\partial L}{\partial \dot{x}_i})\delta x_i]_{t_i}^{t_f} \\
&= \int_{t_i}^{t_f} (\frac{\partial L}{\partial x_i} - \frac{d}{dt}(\frac{\partial L}{\partial \dot{x}_i}))\delta x_i dt
\end{aligned}$$

Where for the next to last equality we used integration by parts and for the last equality the fact that $\delta x_i(t_i) = \delta x_i(t_f) = 0$. Now if S would be at an extremum we would have $\delta S = 0$. This holds if and only if

$$\frac{\partial L}{\partial x_i} - \frac{d}{dt}(\frac{\partial L}{\partial \dot{x}_i}) = 0. \tag{2.8}$$

Now note that $\frac{\partial L}{\partial x_i} = -\frac{\partial V}{\partial x_i}$ and $\frac{\partial L}{\partial \dot{x}_i} = p_i$. So (2.8) is equivalent with

$$\frac{d}{dt}p_i = -\frac{\partial V}{\partial x_i}$$

And this is exactly Newton's equation. So we see that the action is an extremum if and only if Newton's equation hold and the proof is complete. □

Now the proof holds of course for all paths $\{x_i : i = 1, .., 3N\}$. The equations holding for all $i \in \{1, ..., 3N\}$ in (2.8) are called the Euler-Lagrange equations. These equations hold in any coordinate system which follows directly from the action principle as this is a statement about paths and not about coordinates. This means the Euler-Lagrange equations also hold in non-inertial reference frames (for instance rotating frames of reference) as opposed to Newton's equations.

2.3.2 The Hamiltonian Formalism

Note that with the Euler-Lagrange equations we have $3N$ second order differential equations (for every general coordinate x^A an equation). The Hamiltonian approach transforms this in a problem where $6N$ first order differential equations have to be solved. In the following, we assume without loss of generality we deal with N general coordinates. Recall that we have for the generalised momenta,

$$p_i = \frac{\partial L}{\partial \dot{x}_i} \qquad i \in \{1, 2, .., N\}. \tag{2.9}$$

Now if we rewrite the Euler-Lagrange equation with these generalised momenta we get,

$$\dot{p}_i = \frac{\partial L}{\partial x_i} \qquad i \in \{1, 2, .., N\}. \tag{2.10}$$

The idea is to remove $\dot{x}_i$ in favour of p_i. Note that the pair $\{x_i, p_i\}_{i\in\{1,..,N\}}$ determines the state of the system. That is, the positions of the particles at moment t but also the future evolution of the system. The pair $\{x_i, p_i\}_{i\in\{1,..,N\}}$ is a point in 6N-dimensional space, the so called phase space. Note that paths in this space can never cross as one point in this space also determines the future evolution.

The Legendre Transform

We will define the Hamiltonian as the Legendre transform of the Lagrangian with respect to $\dot{x}_i$. Consider an artbitrary two-dimensional function $f(x,y)$. Define now $u(x,y) = \frac{\partial F}{\partial x}$. The Legendre transform $\mathcal{L}_x[f(x,y)] := g(u,y)$ of $f(x,y)$ with respect to x is now given by,

$$\mathcal{L}_x[f(x,y)] := g(u,y) = ux(u,y) - f(x(u,y),y). \tag{2.11}$$

The geometrical interpretation of the Legendre transform is that for fixed y we take for every value u the maximal distance between the curves ux and $f(x,y)$. This is easily seen by noting that maximizing this distance we have,

$$\frac{d}{dx}(ux - f(x)) = 0 \qquad \Longrightarrow \qquad u = \frac{\partial f}{\partial x}.$$

In order to be able to have an inverse of the Legendre transform we need that the function on which the Legendre transform is applied is convex. From the above it follows that we can also define the legendre transform $\mathcal{L}_x[f(x,y)]$ as,

$$\mathcal{L}_x[f(x,y)] := \sup_x[ux - f(x,y)] \tag{2.12}$$

Such that it is not necessary to explicitly calculate the partial derivative $\frac{\partial f}{\partial x} = u$ and inverting this to write x as a function of u and y. Note that the two definitions (2.11) and (2.12) are the same when the supremum in (2.12) is a maximum.

Taking the Legendre transform doesn't lead to losing information. We can get the original function $f(x,y)$ back from the Legendre transform $g(u,y)$. The way to do this is to take again the Legendre transform of the Legendre transformed function. Note that:

$$\frac{\partial g}{\partial u} = u\frac{\partial x}{\partial u} + x(u,y) - \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} = x(u,y) \tag{2.13}$$

Where we used the chain rule and the fact that $u = \frac{\partial f}{\partial x}$. Furthermore, we have:

$$\frac{\partial g}{\partial y} = u\frac{\partial x}{\partial y} - \frac{\partial f}{\partial x}\frac{\partial x}{\partial y} - \frac{\partial f}{\partial y} = -\frac{\partial f}{\partial y} \tag{2.14}$$

Now note that taking the Legendre transform of $g(u,y)$ with respect to u gives:

$$\mathcal{L}_x^2[f(x,y)] = \mathcal{L}_u[g(u,y)] = \frac{\partial g}{\partial u}u(x,y) - g(u(x,y),y) = xu(x,y) - [xu(x,y) - f(x,y)] = f(x,y) \tag{2.15}$$

Which shows that applying the Legendre transform twice gives back the original function.

The Hamiltonian and the Hamilton equations

Now we will define the Hamiltonian to be the Legendre transform of the Lagrangian. Applying the Legendre transform on the Lagrangian with respect to the $\dot{x}_i$ variables, we get,

$$H(x_i,p_i,t) = \sum_i \frac{\partial L(x_i,\dot{x}_i,t)}{\partial \dot{x}_i}\dot{x}_i - L(x_i,\dot{x}_i,t) = \sum_i p_i\dot{x}_i - L(x_i,\dot{x}_i,t) \tag{2.16}$$

We remove $\dot{x}_i$ from the above equation in favour of p_i. Note that we can always write $\dot{x}_i$ as a function of x_i, p_i and t because $p_i = \frac{\partial L}{\partial \dot{x}_i} = p_i(x_i,\dot{x}_i,t)$ which we can invert to write $\dot{x}_i$ as a function of x_i, p_i and t. Now looking at the variation of H we get,

$$\begin{aligned} dH &= (\dot{x}_i dp_i + p_i d\dot{x}_i) - (\frac{\partial L}{\partial x_i}dx_i + \frac{\partial L}{\partial \dot{x}_i}d\dot{x}_i + \frac{\partial L}{\partial t}dt) \\ &= \dot{x}_i dp_i - \frac{\partial L}{\partial x_i} - \frac{\partial L}{\partial t}dt. \end{aligned} \tag{2.17}$$

On the other hand we can write the variation of H as,

$$dH = \frac{\partial H}{\partial x_i}dx_i + \frac{\partial H}{\partial p}dp + \frac{\partial H}{\partial t}dt. \tag{2.18}$$

Equating terms gives the following,

$$\begin{aligned} \dot{p}_i &= \frac{\partial L}{\partial x_i} = -\frac{\partial H}{\partial x_i} \\ \dot{x}_i &= \frac{\partial H}{\partial p_i} \\ -\frac{\partial L}{\partial t} &= \frac{\partial H}{\partial t} \end{aligned} \tag{2.19}$$

Where we used (2.10) above. These equations are the Hamilton equations. Going from the Euler-Lagrange to the Hamilton equations we have replaced n second order differential equations for x_i by $2n$ first order differential equations for x_i and p_i. Note that from (2.15) it follows that taking the legendre transform of the Hamiltonian $H(x_i, p, t)$ with respect to p we get back the Lagrangrian. The Hamiltonian is associated with the energy E of the system and is assumed constant over time (conservation of energy). Indeed, we see that using (2.19),

$$\frac{\partial E}{\partial t} = \frac{\partial}{\partial t}H(x_i, p, t) = \sum_i \frac{\partial H}{\partial x_i}\dot{x}_i + \frac{\partial H}{\partial p_i}\dot{p}_i = \sum_i -\dot{p}_i\dot{x}_i + \dot{x}_i\dot{p}_i = 0 \tag{2.20}$$

So in the remaining of this chapter we can write the Hamiltonian $H(x_i, p)$ without time dependence.

Maupertuis principle

We use the last part of this paragraph to write the action defined in (2.7) in another way, that will be usefull later on, using the Hamiltonian. We look at a one-dimensional system (but it is easily generalised to n dimensions) with the positiion and velocity indicated at time t by γ_t and t $\dot{\gamma}_t$ respectively. Now we can write the Hamiltonian as follows,

$$H(\gamma_t, p) = p\dot{\gamma}_t - L(\gamma_t, \dot{\gamma}_t, t). \tag{2.21}$$

Integrating between the starting time $t_o = 0$ and $t_f = T$ we get,

$$\int_0^T H(\gamma_t, p)dt = \int_0^T p\dot{\gamma}_t dt - \int_0^T L(\gamma_t, \dot{\gamma}_t, t)dt. \tag{2.22}$$

Note that we have seen in (2.20) that $H(\gamma_t, p, t) := E$ is constant over time. Furthermore, we recognize $\int_0^T L(\gamma_t, \dot{\gamma}_t, t)dt$ as the action $S_T[\gamma_t]$. So we can write equation (2.22) as follows:

$$ET = \int_0^T p\dot{\gamma}_t dt - S_T[\gamma_t]. \tag{2.23}$$

Rearranging terms, we get,

$$S_T[\gamma_t] = \int_0^T p\dot{\gamma}_t dt - ET = \int_0^T p d\gamma_t - ET \tag{2.24}$$

Now we see that when the energy is conserved, minimizing $\int_0^T p\dot{\gamma}_t dt$ is equivalent with minimizing the Action. This is known as Maupertuis principle.

2.4 The connection with Large Deviation Theory

In chapter 1 we have seen that for sample path large deviations, we can write (in a somewhat suggestive notation) the probability of a random process $\{Z_n(t)\}$, for which the LDP holds with rate function I, between times $t_i = 0$ and $t_f = 1$ following a path γ_t in the set Γ as,

$$\mathbb{P}(Z_n(t) \in \Gamma) \simeq e^{-n \inf_{\gamma_t \in \Gamma} \int_0^1 L(\gamma_s, \dot{\gamma}_s) ds}. \tag{2.25}$$

So the rate function $I(\gamma_t)$ is equal to the integral $\int_0^1 L(\gamma_s, \dot{\gamma}_s) ds$. Note that the probability is determined by the path that minimizes this integral. Minimizing this integral gives namely the path with the highest probability. This corresponds exactly with minimizing the Action in order to find the path that is taken by the system. So we can also see the integral $\int_0^1 L(\gamma_s, \dot{\gamma}_s) ds$ as the Action. Now lets assume we want to find the path γ_t that is most likely given that $\gamma_{t_i} = A$ and $\gamma_{t_f} = B$. Let Γ contain all possible paths between A and B. We want to find,

$$\inf_{\gamma_t \in \Gamma} \int_0^1 L(\gamma_s, \dot{\gamma}_s) ds. \tag{2.26}$$

So (2.26) shows we are looking at an extremum of the action $S[\gamma_t]$. Lagrangian and Hamiltonian dynamics tell us now that this path has to satisfy the Euler-Langrange equations and the Hamiltonian equations respectively. This can be used to find the path γ_t.

In the special case of Mogulskii's theorem we have,

$$\mathbb{P}(Z_n(t) \in \Gamma) \simeq e^{-n \inf_{\gamma_t \in \Gamma} \int_0^1 \Lambda^\star(\dot{\gamma}_s) ds}, \tag{2.27}$$

For the process $Z_n(t)$ defined in (1.31). In this case, when we want to find the most likely path between A and B we look for,

$$\inf_{\gamma_t \in \Gamma} \int_0^1 \Lambda^\star(\dot{\gamma}_s)) ds. \tag{2.28}$$

Which we again can see as the extremum of the action $S[\gamma_t] = \int_0^1 \Lambda^\star(\dot{\gamma}_s) ds$ with the Lagrangian $\Lambda^\star(\dot{\gamma}_s)$ (which in this case only depends on $\dot{\gamma}_s$). In the below we will give an example where we calculate the most likely Brownian path between the points A and B. This example is illustrative as all values can be computed explicitly.

Example 1 (Brownian Motion)**.** *We look at the process $w_\varepsilon(t)$ as defined in (1.43). Let Γ be the sets of all paths γ_t of $w_\varepsilon(t)$ such that $\gamma_0 = A$ and $\gamma_1 = B$. Now we know from Schilder's Theorem (Theorem 5) that,*

$$\mathbb{P}(w_\varepsilon(t) \in \Gamma) \simeq e^{-n \inf_{\gamma_t \in \Gamma} \frac{1}{2} \int_0^1 |\dot{\gamma}_s|^2 ds} \tag{2.29}$$

So we want to find,

$$\inf_{\gamma_t \in \Gamma} \frac{1}{2} \int_0^1 |\dot{\gamma}_s|^2 ds. \tag{2.30}$$

So we have $L(\gamma_s, \dot{\gamma}_s) = \frac{1}{2}\dot{\gamma}_s^2$. We can find the extremum of the action by solving the Euler-lagrange equations (2.8). We find:

$$0 = \frac{\partial L}{\partial \gamma_t} = \frac{d}{dt}(\frac{\partial L}{\partial \dot{\gamma}_t}) = \ddot{\gamma}_t$$

This means the time derivative of the optimal path $\gamma_t^\star$ is a constant. And we have $\gamma_t^\star = C_1 t + C_2$.

To find the values for C_1 and C_2 we can use the conditions that $\gamma_{t_i}^\star = A$ and $\gamma_{t_f}^\star = B$. Let $t_i = 0$ and $t_f = T$ where T is some positive constant. Filling in the beginpoint we find that $C_2 = A$. Filling in the endpoint we must *have that* $C_1 = \frac{B-A}{T}$*. So the optimal traject for a Brownian motion between the points A and B, i.e. the traject with the lowest cost, is given by,*

$$\gamma_t^\star = A + \frac{B-A}{T}t. \tag{2.31}$$

A second way to calculate this optimal trajectory is by using the Hamiltonian equations. Note that $\frac{\partial L}{\partial \dot\gamma_t} = \dot\gamma_t$*, so by (2.9) we have* $p_t = \dot\gamma_t$ *and we can write using (2.21):*

$$H(\gamma_t, p_t) = p_t^2 - L(\gamma_t, \dot\gamma_t) = p_t^2 - \frac{1}{2}\dot\gamma_t^2 = p_t^2 - \frac{1}{2}p_t^2 = \frac{1}{2}p_t^2. \tag{2.32}$$

From the Hamilton equations in (2,19) we find now that,

$$\dot p_t = -\frac{\partial H}{\partial \gamma_t} = 0$$
$$\dot\gamma_t = \frac{\partial H}{\partial p_t} = p_t$$

From this we find that $p_t = p$ *is a constant and that* $\gamma_t = pt + C_1$*. Calculating the two constants p and* C_1 *again we of course find the same values, namely* $C_1 = A$ *and* $p = \frac{B-A}{T}$*, leading to the same optimal trajectory* $\gamma_t^\star$*.*

The Hamiltonian for stochastic processes

We have seen that the Hamiltonian is the Legendre transform of the Lagrangian. We will now show a way to compute the Hamiltonian directly for a stochastic process $X_n(t)$. This is often a good strategy to find the Lagrangian of a stochastic process and with this the action integral. We have seen in example one that we could derive the Lagrangian directly for the Brownian motion process using Schilder's Theorem. For general processes however such an easy derivation does not exist and the best way for computing the Lagrangian is by taking the legendre transform of the Hamiltonian. This Hamiltonian can be found by the method shown below. For this we first state the following lemma.

Lemma 8 (Varadhan)**.** *Let the measures* $\{\mu_n\}$ *satisfy the LDP on* $\mathcal{X}$ *with rate function I. Let* $F : \mathcal{X} \to \mathbb{R}$ *be a continuous function that is bounded from above. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} e^{nF(x)} d\mu_n(x) = \sup_{x\in\mathcal{X}} [F(x) - I(x)]. \tag{2.33}$$

The proof of this lemma can be found in [2], p.32. Now let $X_n(t)$ be a stochastic process for which the LDP holds with rate function $I(.)$, and we assume $X_n(0) = x$. The Hamiltonian $H(x_t, p_t)$ of this process can be calculated as follows:

$$H(x, p) = \lim_{n\to\infty} \lim_{T\to 0} \frac{1}{nT} \log E_x[e^{n(X_n(T) - X_n(0))p}] \tag{2.34}$$

Where $E_x[.] := E[.|X_n(0) = x]$. This can be seen as follows. Take $F(\gamma) = (\gamma_T - \gamma_0)p$. By Varadhan's lemma we have that,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \log E_x[e^{n(X_n(T)-X_n(0))p}] &= \sup_\gamma [(\gamma_T - \gamma_0)p - I(\gamma)] \\ &= \sup_\gamma \left[(\gamma_T - \gamma_0)p - \int_0^T L(\gamma_s, \dot\gamma_s) ds\right].\end{aligned}$$

Note that for small T this is approximately equal to,

$$\sup_\gamma \left[(\gamma_T - \gamma_0)p - \int_0^T L(\gamma_s, \dot\gamma_s) ds\right] \approx \sup_{\dot\gamma} [T\dot\gamma_0 p - TL(\gamma_0, \dot\gamma_0)]$$

Dividing by T and taking the limit of $T \to 0$ we get,

$$\begin{aligned}
&\lim_{n\to\infty}\lim_{T\to 0}\frac{1}{nT}\log E_x[e^{n(X_n(T)-X_n(0))p}] \\
&= \lim_{T\to 0}\frac{1}{T}\sup_{\dot\gamma}[T\dot\gamma_0 p - TL(\gamma_0,\dot\gamma_0] \\
&= \lim_{T\to 0}\sup_{\dot\gamma}[\dot\gamma_0 p - L(\gamma_0,\dot\gamma_0] \\
&= \sup_{\dot x}[p\dot x - L(x,\dot x)]
\end{aligned} \tag{2.35}$$

Where in the last equality we replaced $\dot\gamma_0$ by $\dot x$. So we see, as we have at the right part in the above equation the Legendre transform of the Lagrangian with respect to $\dot x$ (i.e. the Hamiltonian) that equation (2.34) holds. To come back to the example of the Brownian motion above. Another way we could compute the Hamiltonian is by using (2.34). We will do this in the following example.

Example 2 (Brownian Motion revisited)**.** *We have again the process $w_\varepsilon(t) = \frac{1}{\sqrt{\varepsilon}}w(t)$, where $w(t)$ is a standard Brownian motion. We will now directly compute the Hamiltonian by using (2.34). We have,*

$$H(x,p) = \lim_{\varepsilon\to 0}\lim_{T\to 0}\frac{1}{T\varepsilon}\log E_x[e^{\varepsilon w_\varepsilon(T)p}] = \lim_{\varepsilon\to 0}\lim_{T\to 0}\frac{1}{T\varepsilon}\log E_x[e^{\sqrt{\varepsilon}w(T)p}] \tag{2.36}$$

We used that $w_\varepsilon(0) = 0$. Now, using the fact that $\sqrt{\varepsilon}w(T)$ is a normally distributed random variable with expectation 0 and variance εT together with the fact that the moment generating function (MGF) of a normally distributed random variable X with mean μ and standard deviation σ is given by:

$$MGF := E_x e^{tX} = e^{\mu T + \frac{1}{2}\sigma^2 t^2},$$

we have,

$$\begin{aligned}
\lim_{\varepsilon\to 0}\lim_{T\to 0}\frac{1}{T\varepsilon}\log E_x[e^{\sqrt{\varepsilon}w(T)p^2}] &= \lim_{\varepsilon\to 0}\lim_{T\to 0}\frac{1}{T\varepsilon}\log e^{\frac{1}{2}\varepsilon T p^2} \\
&= \lim_{\varepsilon\to 0}\lim_{T\to 0}\frac{1}{T\varepsilon}\frac{1}{2}\varepsilon T p^2 = \frac{1}{2}p^2.
\end{aligned} \tag{2.37}$$

So taking (2.36) and (2.37) together we have,

$$H(x,p) = \frac{1}{2}p^2 \tag{2.38}$$

And from this the Lagrangian can be obtained by noting that it is the Legendre transform of the Hamiltonian with respect to p. So we have,

$$L(\gamma_s,\dot\gamma_s) = \sup_p[p\dot\gamma_s - H(x,p)] \tag{2.39}$$

To calculate this we fill out $H(x,p)$, take the derivative with respect to p and equating to zero.

$$\frac{\partial}{\partial p}[p\dot\gamma_s - \frac{1}{2}p^2] = 0 \qquad \Longrightarrow \qquad p = \dot\gamma_s$$

This leads to:

$$L(\gamma_s,\dot\gamma, ys) = \frac{1}{2}\dot\gamma_s^2 \tag{2.40}$$

So, as expected, we find the same Lagrangian and Hamiltonian as in example 1.

Two-state continuous-time Markov processes

In the rest of this report we will look at the large deviation behaviour of a particular type of stochastic process, namely the continuous-time Markov process with two states. Markov processes are characterized by the property that, given the present state, the future is independent of the past. In other words, let $\{X(t)\}$ be a continuous-time stochastic process taking the value 1 or 2. The process $\{X(t)\}$ is a continuous time two-state Markov chain if for all $s, t \geq 0$, $0 \leq u < s$ and $i, j, x(u) \in \{1, 2\}$,

$$P(X(t+s) = j|X(s) = i, X(u) = x(u)) = P(X(t+s) = j|X(s) = i)$$

Furthermore, we assume the Markov chain is time-homogeneous, i.e. $P(X(t+s) = j|X(s) = i)$ is independent of s. Now let T_i, $i \in \{1, 2\}$ be the time that the process stays in state i before making a transition. From the Markov property it now follows that,

$$P(T_i > s+t|T_i > s) = P(T_i > t).$$

From this we see the random variable T_i is memoryless and thus exponentially distributed. We can now define $\{X(t)\}$ as a process going from state 1 to state 2 in exponential time with rate $\lambda_1 = \gamma$ and from state 2 to state 1 with rate $\lambda_2 = 1$. Note that without loss of generality we can assume that $\lambda_2 = 1$ as for the behaviour of the process only the relative rates matter. We are interested in the large deviation behaviour of the trajectory of,

$$(\mu_1^n, \mu_2^n) = (\frac{1}{n}\sum_{i=1}^{n} I(X_i(t) = 1), \frac{1}{n}\sum_{i=1}^{n} I(X_i(t) = 2)). \tag{2.41}$$

Here I is the indicator function. We see from this that $\mu_2^n = 1 - \mu_1^n$. For $n \to \infty$ the above trajectory goes to a deterministic constant path, namely the path governed by the kolmogorov forward equations,

$$\begin{aligned} \dot{\mu}_1 &= -\gamma\mu_1 + \mu_2 \\ \dot{\mu}_2 &= \gamma\mu_1 - \mu_2 \end{aligned} \tag{2.42}$$

The above equations are very intuitive as we know that we go with rate λ from state 1 to state 2 and with rate 1 from state 2 to state 1. Defining $x := \mu_1 - \mu_2$ we can convert the above two first order differential equations to one first order differential equation for x. Note that we easily retrieve the values of μ_1 and μ_2 from x as we know that $1 = \mu_1 + \mu_2$ and $x = \mu_1 - \mu_2$. This gives $\mu_1 = \frac{1+x}{2}$ and $\mu_2 = \frac{1-x}{2}$. The differential equation for x becomes,

$$\dot{x} = \dot{\mu}_1 - \dot{\mu}_2 = -2\gamma\mu_1 + 2\mu_2 = -(\gamma+1)x + (1-\gamma) \tag{2.43}$$

Solving this equation and using the boundary condition that $x_0 = a$, we get,

$$x_t = (a - \frac{1-\gamma}{1+\gamma})e^{-(\gamma+1)t} + \frac{1-\gamma}{1+\gamma} \tag{2.44}$$

This solution is the so-called typical path of x. Note that in the symmetric case ($\gamma = 1$), the typical path is given by $x_t = ae^{-2t}$. We are interested in the Large Deviation behaviour if x follows atypical paths. We see that we are dealing just like in the case of Cramer's Theorem with the mean of a stochastic variable $I(X_i(t) = 1)$. However in this case the process is not continuous as $I(X_i(t) = 1)$ fluctuates between 0 and 1. We would like to be able to calculate the optimal paths for $\boldsymbol{\mu}_n := (\mu_1^n, \mu_2^n)$ given that we start in some point $\boldsymbol{x}_0$ and end in some point $\boldsymbol{x}_T$. That is, we would like to calculate the path that has the highest probability of all paths between $\boldsymbol{x}_0$ and $\boldsymbol{x}_T$. For this, first we need to compute the Hamiltonian of the process that jumps with rate γ from state $\boldsymbol{x} := (\frac{i}{n}, \frac{n-i}{n})$ for some $i \leq n$ to state $\boldsymbol{y}_1 := (\frac{i-1}{n}, \frac{n-i+1}{n})$ and with rate 1 from state $\boldsymbol{x}$ to state $\boldsymbol{y}_2 := (\frac{i+1}{n}, \frac{n-i-1}{n})$. We do this using (2.34). So we know, the Hamiltonian is given by,

$$H(\boldsymbol{x}, \boldsymbol{p}) = \lim_{n\to\infty} \lim_{T\to 0} \frac{1}{nT} \log e^{-n\boldsymbol{x}\cdot\boldsymbol{p}} E_{\boldsymbol{x}_0}[e^{n\boldsymbol{\mu}_n\cdot\boldsymbol{p}}] \tag{2.45}$$

Now, taking T small enough we may assume at most one jump happens as the probability that multiple jumps will happen goes to zero much more rapidly for $T \to 0$. So we may write,

$$e^{-n\boldsymbol{x}\cdot\boldsymbol{p}}E_{\boldsymbol{x}_0}[e^{n\boldsymbol{\mu}_n\cdot\boldsymbol{p}}] = e^{-n\boldsymbol{x}\cdot\boldsymbol{p}}[e^{n\boldsymbol{x}\cdot\boldsymbol{p}}(1 - P_{\boldsymbol{x},\boldsymbol{y}_1} - P_{\boldsymbol{x},\boldsymbol{y}_2}]) + e^{n\boldsymbol{y}_1\cdot\boldsymbol{p}}P_{\boldsymbol{x},\boldsymbol{y}_1} + e^{n\boldsymbol{y}_2\cdot\boldsymbol{p}}P_{\boldsymbol{x},\boldsymbol{y}_2}]. \tag{2.46}$$

Now note that $P_{\boldsymbol{x},\boldsymbol{y}_1}$, the probability that a jump occurs from state $\boldsymbol{x}$ to state $\boldsymbol{y}_1$ (i.e. the first jump goes from state 1 to state 2), scales up to a constant with $\gamma\frac{i}{n}T$. The probability that a jump occurs from state $\boldsymbol{x}$ to state $\boldsymbol{y}_2$ (i.e. the first jump goes from state 2 to state 1) scales with $(1-\frac{i}{n})T$. Note that $\mu_1 = \frac{i}{n}$ and $\mu_2 = (1-\frac{i}{n})$. As we will later take the logarithm, the scaling constants will not matter, and we can fill out the unscaled probabilities in (2.46) to obtain,

$$\begin{aligned} e^{-n\boldsymbol{x}\cdot\boldsymbol{p}}E_{\boldsymbol{x}_0}[e^{n\boldsymbol{\mu}_n\cdot\boldsymbol{p}}] &= e^{-n\boldsymbol{x}\cdot\boldsymbol{p}}[e^{n\boldsymbol{x}\cdot\boldsymbol{p}}(1-(\gamma\mu_1+\mu_2)T) + e^{n\boldsymbol{y}_1\cdot\boldsymbol{p}}\gamma\mu_1 T + e^{n\boldsymbol{y}_2\cdot\boldsymbol{p}}\mu_2 T] \\ &= [1 + T\Big(\gamma\mu_1(e^{n\boldsymbol{p}\cdot(\boldsymbol{y}_1-\boldsymbol{x})}-1) + \mu_2(e^{n\boldsymbol{p}\cdot(\boldsymbol{y}_2-\boldsymbol{x})}-1)\Big)]. \end{aligned} \tag{2.47}$$

Taking the logarithm and using the fact that $\log(1+Tc) \approx Tc + o(T)$, we get,

$$\log e^{-n\boldsymbol{x}\cdot\boldsymbol{p}}E_{\boldsymbol{x}_0}[e^{n\boldsymbol{\mu}_n\cdot\boldsymbol{p}}] = T(\gamma\mu_1(e^{n\boldsymbol{p}\cdot(\boldsymbol{y}_1-\boldsymbol{x})}-1) + \mu_2(e^{n\boldsymbol{p}\cdot(\boldsymbol{y}_2-\boldsymbol{x})}-1)). \tag{2.48}$$

Note that the inner product $\boldsymbol{p}\cdot(\boldsymbol{y}_1-\boldsymbol{x})$ is equal to $\boldsymbol{p}\cdot(-\frac{1}{n},\frac{1}{n}) = \frac{p_2-p_1}{n}$. And the inner product $\boldsymbol{p}\cdot(\boldsymbol{y}_2-\boldsymbol{x})$ is equal to $\boldsymbol{p}\cdot(\frac{1}{n},-\frac{1}{n}) = \frac{p_1-p_2}{n}$. Using this, dividing by T and taking the limit for $T \to 0$ and $n \to \infty$ we obtain:

$$H(\boldsymbol{x},\boldsymbol{p}) = \gamma\mu_1(e^{p_2-p_1}-1) + \mu_2(e^{p_1-p_2}-1). \tag{2.49}$$

The above Hamiltonian will be used in the next chapter together with the derived Hamilton equations to determine the optimal paths of the two-state Markov jump process described above. For a more rigorous derivation of the above Hamiltonian we refer to [8].

2.5 Conclusion

We have seen in this chapter that the Principle of Least Action governs the path that is taken by the system as this is equivalent with Newton's equation of motion. The Action of a certain path is the Lagrangian of this path integrated over the time. The Principle of Least Action tells us that the path that is taken is the path that is an extremum of the Action. This is exactly where the connection with large deviation theory comes in. Here we also see that we look for a path that minimizes the rate function as only this path has influence on the value of the measure on some set of (unlikely) paths as we take the limit of $n \to \infty$. In other words, the probability of some set of (unlikely) paths is governed by the most likely path in the set. So, the Rate Function can be identified with the Action Integral. The Euler-Lagrange equations arising from the Principle of Least Action gives a system of second order differential equations. In the Hamiltonian formalism this is converted using a Legendre transform into a system of first order differential equations (but doubling the number of equations). Via a Legendre transform we can move back and forth between the Lagrangian and Hamiltonian. In section 2.4 a useful technique was shown to compute the Hamiltonian for stochastic processes. This is later used to compute the Hamiltonian of the two-state continuous-time Markov process This is the process we will analyse in the next chapter. Using the Hamiltonian and Hamilton equations we can derive the optimal paths (most likely path of all the paths in the set). Furthermore the Action Integral will be computed and the aymptotic behaviour of this Action Integral is analysed.

Chapter 3

Research

3.1 Introduction

In this chapter we will go on with the large deviations of the optimal trajectories for the empirical distribution of n independent 2-state Markov chains between starting time $t = 0$ and ending time $t = T$, i.e. the optimal trajectories of,

$$(\mu_1^n(t), \mu_2^n(t) = (\frac{1}{n}\sum_{i=1}^{n} I(X_i(t) = 1), \frac{1}{n}\sum_{i=1}^{n} I(X_i(t) = 2)) \qquad 0 \leq t \leq T. \tag{3.1}$$

Where, the $\{X_i(t), i = 1, .., n\}$ are independently distributed continuous time two-state Markov processes flipping from state 1 to state 2 with rate λ and from state 2 to state 1 with rate 1. As we have seen, in section 3 of the previous chapter (2.49), the Hamiltonian of the above process (Feng Kurtz Hamiltonian) is given by,

$$H = \gamma\mu_1(e^{p_2-p_1} - 1) + \mu_2(e^{p_1-p_1} - 2) \tag{3.2}$$

We are interested in computing the optimal paths, i.e. the paths with the lowest action, for the process defined in (3.1) between a starting point a and an ending point b. In the Brownian motion case we have seen that this is just a straight line between starting and end point. The interpretation of this is that the costs get higher when the trajectory moves away from a towards b (which makes sense as standard Brownian motion "wants" to stay around the same value, in this case a) and rapid movements are penalized relatively much. Hence, the lowest penalty (cost) is obtained when the trajectory is a straight line. This optimal trajectory is very intuitive. For the process (3.1) above it is more difficult to see how an optimal trajectory between two points would look like.

In this chapter we go on with the paper "Hamiltonian Dynamics and optimal transport for probability measures on two-point sets", by Frank Redig, in the following referred to as [4]. We will start by stating the Hamilton equations which will give 4 first order differential equations (two equations for μ_i and two equations for p_i). After this we will solve these equations for the symmetric case $\gamma = 1$ and later look at the general case. Furthermore, asymptotic behaviour of the action is examined.

3.2 Dynamics of the jump process

3.2.1 The Hamilton equations

The Hamilton equations corresponding to the Hamiltonian stated in (3.2) are given by,

$$\begin{aligned} \dot{p}_i &= -\frac{\partial H}{\partial \mu_i} \\ \dot{\mu}_i &= \frac{\partial H}{\partial p_i} \end{aligned} \tag{3.3}$$

Defining $u_1 = e^{p_1}$, $u_2 = e^{p_2}$ we obtain using the Hamilton equations in (3.3),

$$\begin{aligned}
\dot{u}_1 &= u_1\dot{p}_1 = u_1(-\gamma(e^{p_2-p_1}-1)) = -\gamma u_1(\frac{u_2}{u_1}-1) = \gamma(u_1-u_2)\\
\dot{u}_2 &= u_2\dot{p}_2 = u_2-(e^{p_1-p_2}-1) = (u_2-u_1)\\
\dot{\mu}_1 &= -\gamma\mu_1 e^{p_2-p_1}+\mu_2 e^{p_1-p_2} = -\gamma\mu_1\frac{u_2}{u_1}+\mu_2\frac{u_1}{u_2}\\
\dot{\mu}_2 &= \gamma\mu_1 e^{p_2-p_1}-\mu_2 e^{p_1-p_2} = \gamma\mu_1\frac{u_2}{u_1}-\mu_2\frac{u_1}{u_2}
\end{aligned} \tag{3.4}$$

As in chapter 2, we define $x := \mu_1 - \mu_2$. Furthermore we define $\xi := \frac{u_1}{u_2}$. Remember that from x we can always go back to μ_1 and μ_2 via $\mu_1 = \frac{1+x}{2}$ and $\mu_2 = \frac{1-x}{2}$. Now we can rewrite the Hamiltonian as,

$$H = \frac{1+x}{2}\gamma(\xi^{-1}-1)+\frac{1-x}{2}(\xi-1) \tag{3.5}$$

Note that if we take the derivative with respect to time in the above equation as well as taking this derivative in equation (3.2) we get zero. This means that energy is conserved and we can legitimately use the equations (3.4) stated in chapter 2 to describe the motion of the above system. Now using (3.4) we get the following two first order differential equations for x and ξ,

$$\begin{aligned}
\dot{x} &= \dot{\mu}_1-\dot{\mu}_2 = -\gamma\xi^{-1}+\xi+x(-\gamma\xi^{-1}-\xi)\\
\dot{\xi} &= \frac{u_2\dot{u}_1-u_1\dot{u}_2}{u_2^2} = -\gamma+(\gamma-1)\xi+\xi^2
\end{aligned} \tag{3.6}$$

3.2.2 The Symmetric Case

In the symmetric situation γ is equal to one. In this case the Hamilton equations reduce to,

$$\begin{aligned}
\dot{x} &= -\xi^{-1}+\xi+x(-\xi^{-1}-\xi)\\
\dot{\xi} &= -1+\xi^2
\end{aligned} \tag{3.7}$$

Taking the time derivative in the first equation of (3.7) we get rid of ξ, indeed,

$$\begin{aligned}
\frac{d^2x(t)}{dt^2}v &= \dot{\xi}+\xi^{-2}\dot{\xi}+\dot{x}(-\xi^{-1}-\xi)+x(\xi^{-2}\dot{\xi}-\dot{\xi})\\
&= \dot{\xi}+\xi^{-2}\dot{\xi}+(-\xi^{-1}+\xi+x(-\xi^{-1}-\xi))(-\xi^{-1}-\xi)+x(\xi^{-2}(-1+\xi^2)+1-\xi^2)\\
&= \xi^2-1+\xi^{-2}(\xi^2-1)+(\xi^{-2}-\xi^2)+x(\xi^2+\xi^{-2}+2-\xi^2-\xi^{-2}+1+1)\\
&= 4x
\end{aligned} \tag{3.8}$$

The solution of the above differential equation is easily found. Let's take $x = e^{rt}$. We get $r^2x - 4x = 0$, from which it follows that $r = 2$ or $r = -2$. The general solution is thus given by,

$$x(t) = c_1e^{2t}+c_2e^{-2t} \tag{3.9}$$

Note that we can also write the solution in terms of hyperbolic sines. Furthermore, the solutions of the above equation are time reversible. That is, if $\sinh(2t)$ is a solution then also $\sinh(2(T-t))$. So the unique optimal trajectory starting from $x(0) = a$ and arriving at $x(T) = b$ is given by,

$$x(t) = b\frac{\sinh(2t)}{\sinh(2T)}+a\frac{\sinh(2(T-t))}{\sinh(2T)} \tag{3.10}$$

Note that for $T << 1$ and thus also $t << 1$ we can use the linear approximation of $\sinh(2T)$ and we get,

$$x(t) \approx b\frac{t}{T}+a\frac{T-t}{T} = a+\frac{t}{T}(b-a) \tag{3.11}$$

So $x(t)$ goes for $T << 1$ approximately linearly from a to b. We will later plot the trajectories of x and see that for small T this is indeed the case.
From the first equation in (3.7) we can solve for ξ, we see

$$(1-x)\xi - \dot{x} - (1+x)\xi^{-1} = 0$$

Multiplying both sides of the above equation with ξ we get a quadratic equation in ξ with the following solution,

$$\xi = \frac{\dot{x} + \sqrt{\dot{x}^2 + 4(1-x^2)}}{2(1-x)} \tag{3.12}$$

Note that ξ is always positive and that $\sqrt{\dot{x}^2 + 4(1-x^2)}$ is also positive and bigger than $\dot{x}$ (because $x \in [-1, 1]$). So only one solution holds for ξ and that is the one with the plus sign above. Again rewriting the equation in (3.7) we get

$$\xi^{-1} = \frac{-\dot{x} + \xi(1-x)}{1+x}$$

Filling out (3.12) in the above equation, we find

$$\xi^{-1} = -\frac{\dot{x} - \sqrt{\dot{x}^2 + 4(1-x^2)}}{2(1+x)} \tag{3.13}$$

Using (3.5) together with (3.12) and (3.13) we find the following relationship between the conserved energy E (the Hamiltonian), the position x and the velocity $\dot{x}$

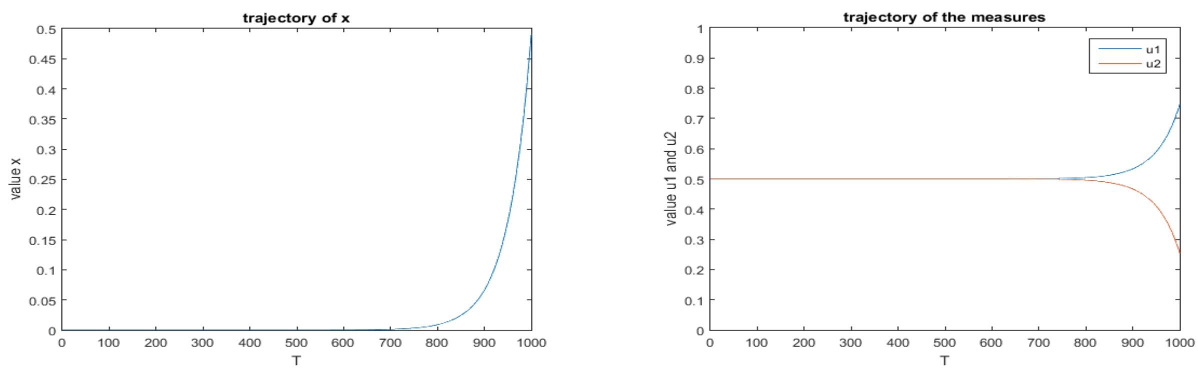
$$2E + 2 = \sqrt{\dot{x}^2 + 4(1-x^2)} \tag{3.14}$$

So we see that the above quantity E, the energy, is a constant of motion along Hamiltonian trajectories. The relationship between ξ and the energy is given by

$$\xi = \frac{\dot{x} + 2E + 2}{2(1-x)} \tag{3.15}$$

Graphing the trajectories

Now that we have the equation for x and we know the relationship between x and the measures μ_1 and μ_2 we can draw the trajectories for some values of a and b. Note that in the symmetric case, the typical behaviour is that $x = ae^{-2t}$. So starting at $a = 0$, typically we will stay at zero. Now let us see what happens if we begin at $x = 0$ but go to $x = 0.5$. We include both the graph for the trajectory of x as well as the graphs for the trajectories of the measures μ_1 and μ_2.

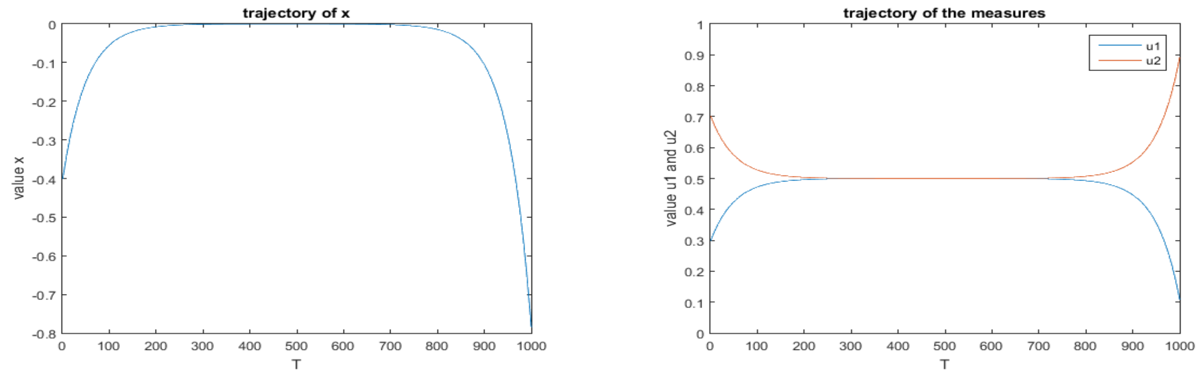


(a) Trajectory of x (b) Trajectory of the measures

Figure 3.1: Trajectories for a=0, b=0.5 and T=10

We see in the above figure that x stays a long time at the typical value 0 till a certain point were it goes exponentially fast to the value b were it should be at $t = T$. So, as we can see, for the values of the measures it means that both are starting at the equilibrium value 0.5 and stay there for about three quarters of the time after which they run off to $\mu_1 = 0.75$ and $\mu_2 = 0.25$.

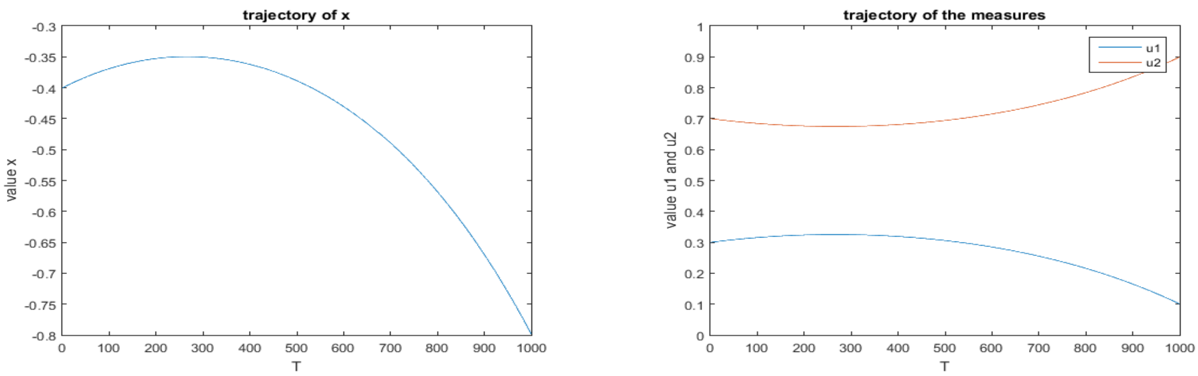
Now we will look what happens if we start at $\mu_1 = 0.3$ and $\mu_2 = 0.7$ ($x = -0.4$) and we go to $\mu_1 = 0.1$ and $\mu_2 = 0.9$ ($x = -0.8$). The reason these values were chosen is because in this case we start already under the typical value of x and the end value is even lower. It is interesting to see what the trajectory of x will look like. It may stay around the starting value and later decrease further to the ending value (like we saw in the picture above), it can also directly start descending towards the ending value. Another possibility is that it first increases to the typical value $x = 0$ and later decreases again.



(a) Trajectory of x (b) Trajectory of the measures

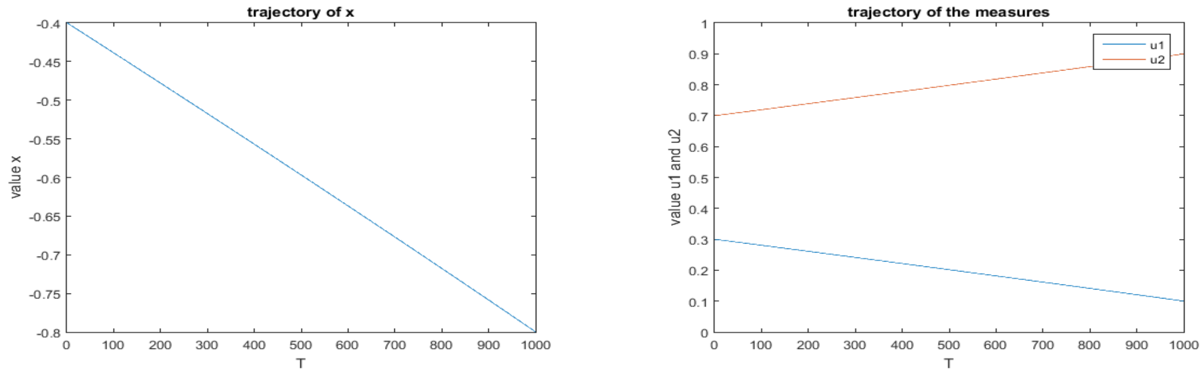
Figure 3.2: Trajectories for a=-0.4, b=-0.8 and T=10

We see that the latter is the case. The trajectory of x first goes back from its starting value $x = -0.4$ to its typical value $x = 0$ and then decreases to the ending value -0.8. Now let us look what happens if x has less time to go from a to b. We take instead of ending time $T = 10$ the ending times $T = 1$ and the more extreme ending time $T = 0.1$. We get the following results,



(a) Trajectory of x (b) Trajectory of the measures

Figure 3.3: Trajectories for a=-0.4, b=-0.8 and T=1



(a) Trajectory of x

(b) Trajectory of the measures

Figure 3.4: Trajectories for a=-0.4, b=-0.8 and T=0.1

We see that in figure 3.3, the time is too short to 'push' the trajectory to the equilibrium value. The trajectory of x increases however a little bit but then starts to decrease to its ending value $x = -0.8$ before hitting equilibrium. In figure 3.4 we see that the descend starts immediately and is roughly linear as we would expect from (3.11).

In a later section we will also look at the action integral of the trajectories of the symmetric case. We will now first look at the dynamics of x in the asymmetric case.

3.2.3 The Asymmetric Case

In this section we will look at the asymmetric situation. In this case the rate of going from state 1 to state 2 is different than the rate of going from state 2 to state 1. For this reason the typical value of x will not be zero anymore. To derive the trajectories of x we go back to the following differential equations for x and ξ with $\gamma \neq 1$,

$$\begin{aligned} \dot{x} &= -\gamma\xi^{-1} + \xi + x(-\gamma\xi^{-1} - \xi) \\ \dot{\xi} &= -\gamma + (\gamma - 1)\xi + \xi^2 \end{aligned} \tag{3.16}$$

Furthermore we can see using (3.5) the expression below for the energy E holds. Note that the Hamiltonian and thus the expression below is a constant of motion.

$$\begin{aligned} 2E &= x(\gamma\xi^{-1} - \xi) + \gamma\xi^{-1} + \xi + (1-\gamma)x - (1+\gamma) \\ &= (1-x)(\xi - 1) + (1+x)\gamma(\xi^{-1} - 1) \end{aligned} \tag{3.17}$$

Again, as in the symmetric case, we take the time derivative of the first equation of (3.16). In the symmetric case this led to the elimination of ξ. As we will see now, this will not happen in the asymmetric case. We get,

$$\begin{aligned} \ddot{x} &= (1-x)\dot{\xi} - \dot{x}\xi + (1+x)\gamma\frac{\dot{\xi}}{\xi^2} - \dot{x}\frac{\gamma}{\xi} \\ &= (1-x)\dot{\xi} - ((1-x)\xi - (1+x)\gamma\xi^{-1})\xi + (1+x)\gamma\frac{\dot{\xi}}{\xi^2} - ((1-x)\xi - (1+x)\gamma\xi^{-1})\frac{\gamma}{\xi} \\ &= (1-x)(\dot{\xi} - \xi^2 - \gamma) + (1+x)(\gamma + \gamma\frac{\dot{\xi}}{\xi^2} + \frac{\gamma^2}{\xi^2}) \end{aligned} \tag{3.18}$$

Concentrating on the terms multiplied by $(1-x)$ and $(1+x)$ respectively seperately, and filling out $\dot{\xi}$, we see for the factor of $(1-x)$,

$$\begin{aligned} \dot{\xi} - \xi^2 - \gamma &= -\gamma + (\gamma - 1)\xi + \xi^2 - \xi^2 - \gamma \\ &= (\gamma - 1)\xi - 2\gamma \end{aligned}$$

And for the factor of $(1+x)$ we get,

$$\begin{aligned}\gamma+\gamma\frac{\dot{\xi}}{\xi^2}+\frac{\gamma^2}{\xi^2} &= \gamma+\gamma\frac{-\gamma+(\gamma-1)\xi+\xi^2}{\xi^2}+\frac{\gamma^2}{\xi^2}\\ &= \gamma(\frac{\xi^2}{\xi^2}+\frac{-\gamma+(\gamma-1)\xi+\xi^2}{\xi^2}+\frac{\gamma}{\xi^2})\\ &= (\gamma-1)\gamma\xi^{-1}+2\gamma\end{aligned}$$

So we obtain the following expression for $\ddot{x}$

$$\ddot{x} = (1-x)((\gamma-1)\xi-2\gamma)+(1+x)((\gamma-1)\gamma\xi^{-1}+2\gamma) \tag{3.19}$$

Now we isolate the terms containing $(\gamma-1)$ and rewrite this in terms of the energy $2E$.

$$\begin{aligned}(\gamma-1)((1-x)\xi+(1+x)\gamma\xi^{-1}) &= (\gamma-1)((1-x)(\xi-1)+(1+x)\gamma(\xi^{-1}-1)+(1-x)+\gamma(1+x))\\ &= (\gamma-1)(2E+(\gamma-1)+(1-x)+\gamma(1+x))\\ &= (\gamma-1)(2E+(1+\gamma))+x(\gamma-1)^2\end{aligned}$$

Note that all the terms with ξ dropped out of the equation (but of course the energy is dependent of ξ). We can now write the secornd order differential equation for x in terms of the energy and the rate γ as follows,

$$\begin{aligned}\ddot{x} &= (\gamma-1)(2E+(1+\gamma))+x(\gamma-1)^2-2\gamma(1-x)+2\gamma(1+x)\\ &= (\gamma-1)(2E+1+\gamma)+x(\gamma+1)^2\end{aligned} \tag{3.20}$$

Solving the above second-order differential equation with boundaries $x(0)=a$ and $x(T)=b$ we find

$$x(t) = C_1e^{\delta t}+C_2e^{-\delta t}+C_3 \tag{3.21}$$

with,

$$\begin{aligned}C_1 &= \frac{(b-C_3)-(a-C_3)e^{-\delta T}}{2\sinh(\delta T)}\\ C_2 &= \frac{(a-C_3)e^{\delta T}-(b-C_3)}{2\sinh(\delta T)}\\ C_3 &= -\frac{(\delta-2)(2E+\delta)}{\delta^2}\\ \delta &= 1+\gamma\end{aligned} \tag{3.22}$$

We can rewrite this in terms of all hyperbolic sines as follows,

$$x(t) = (b-C_3)\frac{\sinh(\delta t)}{\sinh\delta T}+(a-C_3)\frac{\sinh(\delta(T-t))}{\sinh\delta T}+C_3 \tag{3.23}$$

When $T << 1$ and thus also $t << 1$ we can write $\sinh(\delta t)\approx\delta t$ and we see that the dynamics of x are approximately equal to,

$$x(t) \approx (b-C_3)\frac{\delta t}{\delta T}+(a-C_3)\frac{\delta T-\delta t}{\delta T}+C_3 = a+\frac{t}{T}(b-a) \tag{3.24}$$

Note that the constants C_3 and δ drop out of the equation. So again we get the same linear behaviour for $T << 1$ as in the symmetric case

.

In the following we will write $C := C_3$. Notice that in this case our solution is dependent on the energy (via the constant C) and this still has to be determined. In the symmetric case we didn't have this problem as the second-order differential equation for x did not depend on ξ anymore. In order to find the value of the energy we look at time $t=0$ (note that the energy is a constant). At this point we have, filling out that $x(0)=a$

$$\begin{aligned} \dot{x}_0 &= -\gamma\xi_0^{-1} + \xi_0 + a(-\gamma\xi_0^{-1} - \xi_0) \\ 2E &= a(\gamma\xi_0^{-1} - \xi_0) + \gamma\xi_0^{-1} + \xi_0 + (1-\gamma)a - (1+\gamma) \end{aligned} \tag{3.25}$$

Note that we can combine the two equations in (3.25) to write the energy as a function of $\dot{x}_0$ and by doing this ξ_0 is removed from the equation. We get,

$$2E = \sqrt{\dot{x}_0^2 + 4\gamma(1-a^2)} - (1+\gamma-(1-\gamma)a). \tag{3.26}$$

Note that in general the following expression is constant over time (doing the same as above but now not filled out $t=0$),

$$2E = \sqrt{\dot{x}_t^2 + 4\gamma(1-x_t^2)} - (1+\gamma-(1-\gamma)x_t). \tag{3.27}$$

The typical speed $\dot{x}_0$ at time zero with $x_0 = a$ and γ unequal to one is easily calculated. Note that in this case $\dot{\mu}_1(0) = -\gamma\mu_1(0) + \mu_2(0) = -\gamma\frac{1+a}{2} + \frac{1-a}{2}$ and $\dot{\mu}_2(0) = \gamma\mu_1(0) - \mu_2(0) = \gamma\frac{1+a}{2} - \frac{1-a}{2}$. So we have $\dot{x}_0 = \dot{\mu}_1(0) - \dot{\mu}_2(0) = -\gamma(1+a) + (1-a)$. Filling this out in the equation (3.26) for the speed we see that the energy becomes zero as it should. Furthermore, note that the typical speed at time zero is equal to zero if we fill out for a the equilibrium value of $a = \frac{2-\delta}{\delta}$. Which of course makes sense as in this case the x will not have to move to come on its equilibrium value.

taking the first equation of (3.16) we can solve for ξ again. First multiplying with ξ and rewriting the equation we get,

$$(1-x)\xi^2 - \dot{x}\xi - \gamma(1+x) = 0 \tag{3.28}$$

and thus,

$$\xi_t = \frac{\dot{x}_t + \sqrt{\dot{x}_t^2 + 4\gamma(1-x_t^2)}}{2(1-x_t)}. \tag{3.29}$$

Using (3.27) we find the following relation between ξ, $\dot{x}$ and the energy E

$$\xi_t = \frac{\dot{x}_t + 2E + (1+\gamma) - (1-\gamma)x_t}{2(1-x_t)}. \tag{3.30}$$

Still we want to solve for $2E$ in order to get an explicit formula for the motion x. To do this, note that we can use (3.23) to find an additional relation between $\dot{x}_0$ and the energy E. We see differentiating (3.23) and filling out $t=0$,

$$\dot{x}_0 = \delta(b-C)\frac{1}{\sinh\delta T} - \delta(a-C)\frac{\cosh(\delta T)}{\sinh\delta T}.$$

Solving for C we get,

$$C = \frac{\sinh(\delta T)}{\delta(\cosh(\delta T)-1)}\dot{x}_0 + \frac{a\cosh\delta T - b}{\cosh(\delta T)-1}$$

Now filling out C which depends on the $2E$ and solving for $2E$ leads to the following equation between $\dot{x}_0$ and E.

$$\begin{aligned} 2E &= A\dot{x}_0 + B \\ A &= \frac{\delta}{2-\delta}\frac{\sinh(\delta T)}{\cosh(\delta T)-1} \\ B &= \frac{\delta^2}{2-\delta}\frac{a\cosh(\delta T)-b}{\cosh(\delta T)-1} - \delta \end{aligned} \tag{3.31}$$

We can equate now the first expression of (3.31) and the expression in (3.26) and solve for $\dot{x}_0$. Filling out this value of $\dot{x}_0$ in the first expression of (3.31) gives the value of 2E which is needed to finally get an explicit equation for the trajectory x. Defining $D = B + \delta - (2-\delta)a$, we get equating (3.26) and (3.31),

$$\begin{aligned}&\sqrt{\dot{x}_0^2 + 4\gamma(1-a^2)} = A\dot{x}_0 + D\\ \Longleftrightarrow\ &\dot{x}_0^2 + 4\gamma(1-a^2) = A^2\dot{x}_0^2 + 2AD\dot{x}_0 + D^2\\ \Longleftrightarrow\ &(A^2-1)\dot{x}_0^2 + 2AD\dot{x}_0 - 4\gamma(1-a^2) + D^2 = 0\end{aligned}$$

Solving this quadratic equation we get,

$$\dot{x}_0 = \frac{-AD + \sqrt{4\gamma(1-a^2)(A^2-1) + D^2}}{(A^2-1)}. \tag{3.32}$$

Now filling out the values for A and D, we find after some algebra,

$$\begin{aligned}\dot{x}_0 =& \frac{a\delta\sinh(\delta T) - \frac{\delta^3}{(2-\delta)^2}\coth(\frac{1}{2}\delta T)(a\cosh(\delta T) - b)}{(\frac{\delta^2}{(2-\delta)^2} - 1)\cosh(\delta T) + (\frac{\delta^2}{(2-\delta)^2} + 1)} + \frac{(2-\delta)^2}{\delta^2\coth^2(\frac{1}{2}\delta T) - (2-\delta)^2}\cdot\Bigg[\frac{\delta^4}{(2-\delta)^2}\Big(\frac{-b + a\cosh(\delta T)}{\cosh(\delta T) - 1}\Big)^2\\ &- 2\delta^2 a\cdot\Big(\frac{-b + a\cosh(\delta T)}{\cosh(\delta T) - 1}\Big) + (2-\delta)^2a^2 + 4(\delta-1)(1-a^2)\Big(\frac{\delta^2}{(2-\delta)^2}\coth^2(\frac{1}{2}\delta T) - 1\Big)\Bigg]^{\frac{1}{2}}\end{aligned} \tag{3.33}$$

With this we can use (3.31) to solve for $2E$ and using this we can determine the constant C in (3.23) with which we can calculate the trajectory of x. We find after some algebra that,

$$\begin{aligned}2E = A\dot{x}_0 + B =& \frac{\frac{\delta^2}{(2-\delta)}a(\cosh(\delta T) + 1) - \frac{\delta^4}{(2-\delta)^3}\coth^2(\frac{1}{2}\delta T)(a\cosh(\delta T) - b)}{(\frac{\delta^2}{(2-\delta)^2} - 1)\cosh(\delta T) + (\frac{\delta^2}{(2-\delta)^2} + 1)}\\ &+ \frac{\delta(2-\delta)\sinh(\delta T)}{4(\delta-1)\cosh(\delta T) + \delta^2 + (2-\delta)^2}\cdot\Bigg[\frac{\delta^4}{(2-\delta)^2}\Big(\frac{-b + a\cosh(\delta T)}{\cosh(\delta T) - 1}\Big)^2 - 2\delta^2 a\cdot\Big(\frac{-b + a\cosh(\delta T)}{\cosh(\delta T) - 1}\Big)\\ &+ (2-\delta)^2a^2 + 4(\delta-1)(1-a^2)\Big(\frac{\delta^2}{(2-\delta)^2}\coth^2(\frac{1}{2}\delta T) - 1\Big)\Bigg]^{\frac{1}{2}} + \frac{\delta^2}{(2-\delta)}\frac{a\cosh(\delta T) - b}{\cosh(\delta T) - 1} - \delta\end{aligned}$$

Now with this it is not hard to calculate the constant C which is given by,

$$\begin{aligned}C =& \frac{(2-\delta)^2a(\cosh(\delta T) + 1) - \delta^2\coth^2(\frac{1}{2}\delta T)(a\cosh(\delta T) - b)}{4(\delta-1)\cosh(\delta T) + (\delta^2 + (2-\delta)^2)}\\ &+ \frac{(2-\delta)\sinh(\delta T)}{4\delta(\delta-1)\cosh(\delta T) + \delta^3 + \delta(2-\delta)^2}\cdot\Bigg[\delta^4\Big(\frac{-b + a\cosh(\delta T)}{\cosh(\delta T) - 1}\Big)^2 - 2\delta^2(2-\delta)^2a\cdot\Big(\frac{-b + a\cosh(\delta T)}{\cosh(\delta T) - 1}\Big)\\ &+ (2-\delta)^4a^2 + 4(\delta-1)(1-a^2)\Big(\delta^2\coth^2(\frac{1}{2}\delta T) - (2-\delta)^2\Big)\Bigg]^{\frac{1}{2}} + \frac{a\cosh(\delta T) - b}{\cosh(\delta T) - 1}\end{aligned} \tag{3.34}$$

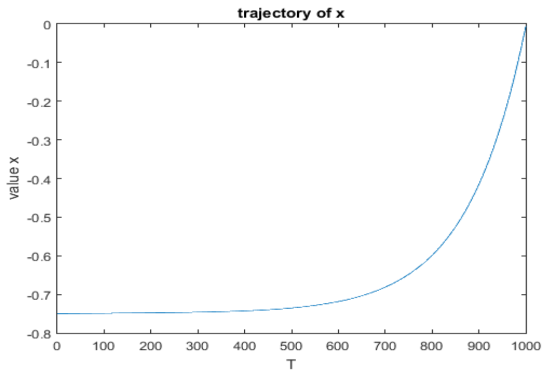
Note that when we are in the symmetric case $\gamma = 1$, and thus $\delta = 2$, a lot of the terms in the above equation immediately drop out and we get,

$$\begin{aligned}C = -\frac{4\coth^2(T)(a\cosh(2T) - b)}{4(\cosh(2T) + 1)} + \frac{a\cosh(2T) - b}{\cosh(2T) - 1} &= -\frac{\frac{\cosh(2T)+1}{\cosh(2T)-1}(a\cosh(2T) - b)}{\cosh(2T) + 1} + \frac{a\cosh(2T) - b}{\cosh(2T) - 1}\\ &= -\frac{a\cosh(2T) - b}{\cosh(2T) - 1} + \frac{a\cosh(2T) - b}{\cosh(2T) - 1} = 0\end{aligned}$$

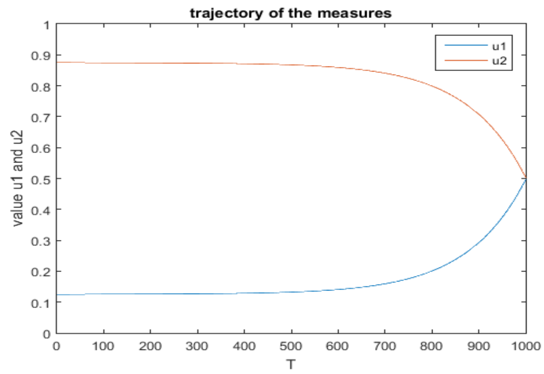
like we should get.

Graphing the Trajectories

Now that the equation of motion of x is explicit we can look at the behaviour of the trajectories of both x and the measures μ_1 and μ_2. In the asymmetric case the bigger the value of γ the faster particles jump from state 1 to state 2 (relative to the particles jumping from state 2 to state 1). The system typically converges to the distribution $\mu_1 = \frac{1}{\gamma+1}$ and $\mu_2 = \frac{\gamma}{\gamma+1}$. The first graphs plotted start on this typical equilibrium values for $\gamma = 7$ (so $x_0 = -\frac{6}{8}$) and the ending value is $x(T) = 0$.



(a) Trajectory of x

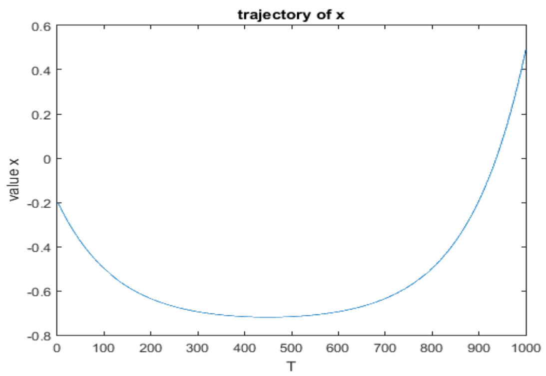


(b) Trajectory of the measures

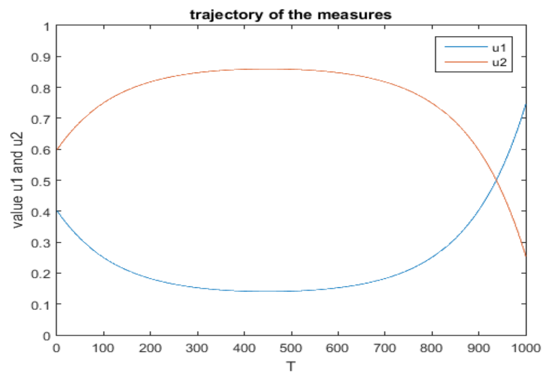
Figure 3.5: Trajectories for $a = -\frac{6}{8}$, $b = 0$, $T = 1$ and $\gamma = 7$

We see that x and the measures stay very close to the equilibrium value and later, after three quarters of the time has passed, the measures are going towards their destined value b.

Below, in figure 3.6 the graphs are shown for again $\gamma = 7$ but now starting above the equilibrium value for x and ending even higher above this equilibrium value. It can be seen that also in this asymmetric case the trajectory first goes towards the equilibrium value, and thus moves away from the destined value. On the last moment it moves rapidly to the destined value.



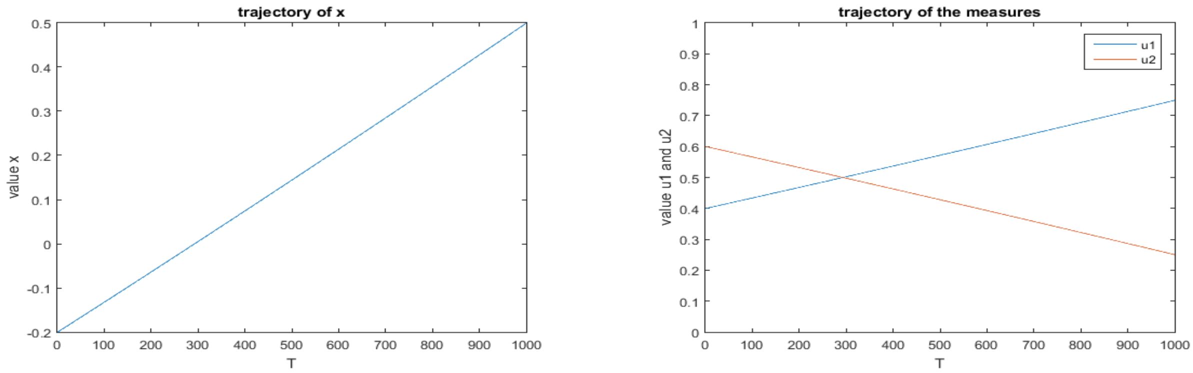
(a) Trajectory of x



(b) Trajectory of the measures

Figure 3.6: Trajectories for $a = -0.2$, $b = 0.5$, $T = 1$ and $\gamma = 7$

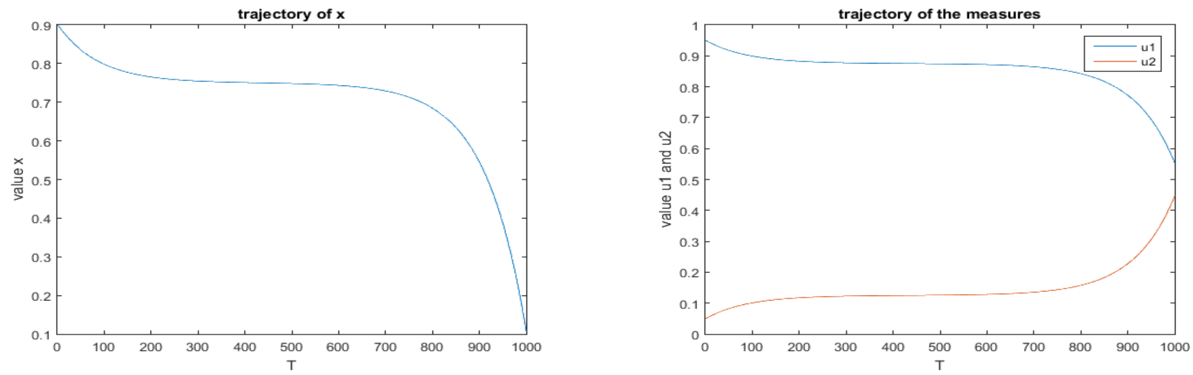
Now we will draw the graphs again for $a = -0.2$, $b = 0.5$ and $\gamma = 7$ but now letting T be very small, namely for $T = 0.01$. From (3.24) we expect the trajectories now to be approximately linear. We see below that this is indeed the case,



(a) Trajectory of x (b) Trajectory of the measures

Figure 3.7: Trajectories for $a=-0.2$, $b=0.5$, $T=0.01$ and $\gamma=7$

The last two graphs shown below are for the case that $\gamma=\frac{1}{7}$. So the roles are reversed between the measures μ_1 and μ_2. In this case the equilibrium value for x is $\frac{6}{8}$. In the graph we look at the situation that we start above the equilibrium value, namely $a=0.9$ and go below the equilibrium value towards $b=0.1$. In this case the trajectory of x descends first to the equilibrium value and stays close to this value until it descends rapidly towards its destined value.



(a) Trajectory of x (b) Trajectory of the measures

Figure 3.8: Trajectories for $a=0.9$, $b=0.1$, $T=10$ and $\gamma=\frac{1}{7}$

3.2.4 The trajectories for large T

We have seen that in the case T is very small, the trajectories are approximately linear between starting point a and ending point b. Furthermore we have seen that when T is big enough, the trajectory first goes from starting point a to the equilibrium value, stays here a while and subsequently goes to the ending value b. We suspect that Comparing the figures of the last two sections we see that the bigger the value of T is,the longer the trajectory stays close to the equilibrium value. In other words, the time period in which the trajectory goes up to its ending value seems to get relatively shorter for increasing T. We want to find the behaviour of this, in particular for $T\to\infty$.

Looking at (3.21) we see that for $T>>1$ and $t>>1$, we can write,

$$x(t)\approx(b-C)\frac{e^{\delta t}}{e^{\delta T}}+C. \tag{3.35}$$

Now we know from the pictures that $x(t)$ will go for t big enough but smaller than T to its equilibrium value $\frac{2-\delta}{\delta}$. This suggests that for large T, $C\to\frac{2-\delta}{\delta}$. This will be confirmed in paragraph 3.3.2. We see in this case from (3.35) that for $1<<t<<T$ that $x(t)$ is approximately equal to its equilibrium value. For t increasing, the trajectory will move a tiny bit towards its ending value b. But we want to know

when it properly starts moving towards b. Too define what we mean by this, fix a number $r \in [0, 1]$. Now we say that the trajectory has started moving towards b iff we have moved a distance $r|b - C|$ in the direction of b. We see that for this to hold, we need to have that,

$$r = \frac{e^{\delta t}}{e^{\delta T}}. \tag{3.36}$$

We denote the time at which we start moving towards b with t^*. From (3.36) it follows that,

$$t^* = T - \frac{1}{\delta}\log(r^{-1}) \tag{3.37}$$

The relative period of moving towards b is thus given by.

$$\frac{T - t^*}{T} = \frac{\frac{1}{\delta}\log(r^{-1})}{T} \tag{3.38}$$

For T very big we see that this period gets arbitrarily small. Furthermore, we see that each time we double the ending time T, the relative period of moving towards b is halved. In order to check this behaviour numerically, we looked at the case $a = 0$, $b = 0.8$ and $\gamma = 1$. We plotted x for 6 different values of T, each time doubling T. After this we computed the point of intersection of the trajectory with the line $k = 0.08$. Note that this corresponds with taking $r = 0.1$. The following table was obtained.

T	$\frac{T-t^*}{T}$
5	0.229
10	0.114
20	0.057
40	0.028
80	0.013
160	0.006

Table 3.1: Relative times of moving towards b

We see in the above table that it confirms our calculation above. For every time we double the ending time T we see in the table that we more or less halve the relative time of moving towards b. In the next section we will look at the Action and compute an explicit value for the Action. Furthermore we will look what happens in the case $T \to 0$ and $T \to \infty$.

3.3 The Action Integral and its Asymptotics for $T \to 0$ and $T \to \infty$

Recall from chapter 2 that the Action integral is the Langrangian of the system integrated over time. By the Principle of Least Action it is known that the optimal trajectory is the trajectory that minimizes this action integral. Although we already calculated the optimal paths for the symmetric and asymmetric case in the previous section we would like to see the behaviour of the Action Integral. In particular the asymptotics for $T \to 0$ and $T \to \infty$. Note that we find from the example for Brownian motion (paragraph 2.3, example 1, p.28) that the Action integral for the optimal path is easily calculated and is given by,

$$S_T(a, b) = \int_0^T \frac{1}{2}\dot{\gamma}_s^2 ds = \int_0^T \frac{1}{2}\Big(\frac{(a-b)}{T}\Big)^2 ds = \frac{(a-b)^2}{2T}$$

The asymptotics here are easily seen and intuitive. The action is zero for $a = b$ as in this case the path is typical. Otherwise, for $T \to 0$ the Action goes to infinity. This is intuitive because the shorter the time the process has to go from a to b, the steeper has to be the ascend or descend and the least likely is the path. On the other hand for $T \to \infty$ the Action goes to zero because the ascend or descend comes arbitrarily close to zero, so the path from a to b comes arbitrarily close to the typical path. In the case of the two-state Markov process we expect the same but we want to check this and we also want to find at which rates the Action integral goes to zero or infinity respectively. The action of the optimal trajectory γ_s starting at $\gamma_0 = a$ and arriving at $\gamma_T = b$ is given by,

$$\begin{aligned} S_T(a,b) &= \int_0^T L(\gamma_s, \dot{\gamma}_s) ds \\ &= \int_0^T p_1(s) d\mu_1(s) + \int_0^T p_2(s) d\mu_2(s) - ET \\ &= \frac{1}{2}\int_0^T p_1(s) - p_2(s) dx(s) - ET \\ &= \frac{1}{2}\int_0^T \log(\xi(s)) dx(s) - ET \end{aligned} \tag{3.39}$$

Where the second equality follows directly from Maupertuis's Principle. The third equality follows by remembering that $\mu_1 = \frac{1+x}{2}$ and $\mu_2 = \frac{1-x}{2}$ and thus $d\mu_1 = \frac{1}{2}dx$ and $\mu_2 = -\frac{1}{2}dx$. The fourth equality follows directly by noting that $\xi = \frac{e^{p_1}}{e^{p_2}}$.

We state now directly the value of the Action $S_T(a,b)$, i.e. the solution of the above equation (3.39) as the calculation is too elaborate to include in the main body of the text. For the calculation we refer to appendix A at the end of this report. We find the following explicit expression of the Action $S_T(a,b)$, only dependent on a, b, γ, T, s and the explicit constant C,

$$\begin{aligned} &2S_T(a,b) + 2ET = \\ &\quad \frac{s}{2\gamma}\left(\tan^{-1}\left(\frac{1}{s}\left[\frac{2\gamma(b-C)e^{\delta T} - 2\gamma(a-C)}{\sinh \delta T} + \frac{4\gamma C}{1-\gamma}\right]\right) - \tan^{-1}\left(\frac{1}{s}\left[\frac{2\gamma(b-C) - 2\gamma(a-C)e^{-\delta T}}{\sinh \delta T} + \frac{4\gamma C}{1-\gamma}\right]\right)\right) \\ &\quad - \frac{s}{2}\left(\tan^{-1}\left(\frac{1}{s}\left[\frac{2(b-C)e^{-\delta T} - 2(a-C)}{\sinh \delta T} + \frac{4\gamma C}{1-\gamma}\right]\right) - \tan^{-1}\left(\frac{1}{s}\left[\frac{2(b-C) - 2(a-C)e^{\delta T}}{\sinh \delta T} + \frac{4\gamma C}{1-\gamma}\right]\right)\right) \\ &\quad + \left[\frac{C}{2-\delta} + \frac{(b-C)e^{\delta T} - (a-C)}{2\sinh(\delta T)}\right] \log\left(\frac{(b-C)(\gamma e^{2\delta T} + 1) - (a-C)e^{\delta T}\delta}{\sinh(\delta T)} + \frac{4e^{\delta T}\gamma C}{2-\delta}\right) \\ &\quad + \left[-\frac{C\gamma}{2-\delta} + \frac{(a-C) - (b-C)e^{-\delta T}}{2\sinh(\delta T)}\right] \log\left(\frac{(b-C)(e^{-2\delta T} + \gamma) - (a-C)e^{-\delta T}\delta}{\sinh(\delta T)} + \frac{4e^{-\delta T}\gamma C}{2-\delta}\right) \\ &\quad - a\log\left(\frac{(b-C)\delta - (a-C)(\gamma e^{-\delta T} + e^{\delta T})}{\sinh(\delta T)} + \frac{4\gamma C}{(2-\delta)}\right) + (a-b)\log(2) - \delta T(b-C)\tanh^{-1}(\delta T) \\ &\quad + \delta T(a-C)\sinh^{-1}(\delta T) + (1-b)\log(1-b) - (1-a)\log(1-a). \end{aligned} \tag{3.40}$$

3.3.1 Small Time Asymptotics

We can look at the small time asymtotics of the above formula, so for $T \to 0$. For this we first have to look what happens to the constant C for $T \to 0$. Note that in the above formula the term $\frac{C}{2-\delta}$ appears a lot. Determining the small time asymtotics for the case $\delta = 2$ has to be handled separately. In the previous section it has been shown that C converges to zero for $\delta \to 2$. However, it also needs to be determined what happens to $\frac{C}{2-\delta}$ for $\delta \to 2$. First we look at what happens to C for $T \to 0$.

We will use the following notation. Two function $f(T)$ and $g(T)$ are asymptotically equivalent if and only if $\lim_{T\to 0} |\frac{f(T)}{g(T)}| = 1$. In this case we write $f \approx g$. From equation (3.34) we get,

$$
\begin{aligned}
C =& \frac{(2-\delta)^2 a(\cosh(\delta T)+1) - \delta^2\coth^2(\frac{1}{2}\delta T)(a\cosh(\delta T)-b)}{4(\delta-1)\cosh(\delta T)+(\delta^2+(2-\delta)^2)} \\
&+ \frac{(2-\delta)\sinh(\delta T)}{4\delta(\delta-1)\cosh(\delta T)+\delta^3+\delta(2-\delta)^2}\cdot\Bigg[\delta^4\Big(\frac{-b+a\cosh(\delta T)}{\cosh(\delta T)-1}\Big)^2 - 2\delta^2(2-\delta)^2 a\cdot\Big(\frac{-b+a\cosh(\delta T)}{\cosh(\delta T)-1}\Big) \\
&+ (2-\delta)^4 a^2 + 4(\delta-1)(1-a^2)\Big(\delta^2\coth^2(\frac{1}{2}\delta T) - (2-\delta)^2\Big)\Bigg]^{\frac{1}{2}} + \frac{a\cosh(\delta T)-b}{\cosh(\delta T)-1} \\
\approx& -\delta^2\frac{\cosh(\delta T)+1}{4(\delta-1)\cosh(\delta T)+(\delta^2+(2-\delta)^2)}\cdot\frac{a\cosh(\delta T)-b}{\cosh(\delta T)-1} + \frac{a\cosh(\delta T)-b}{\cosh(\delta T)-1} \\
&+ \frac{(2-\delta)\sinh(\delta T)}{2\delta^3}\sqrt{\delta^4\Big(\frac{-b+a\cosh(\delta T)}{\cosh(\delta T)-1}\Big)^2} \\
\approx& \frac{(2-\delta)(b-a)}{\delta^3 T}
\end{aligned} \tag{3.41}
$$

Where we assumed $b > a$. For $a > b$ the same expression holds only a and b switch places. To obtain the second to last equality we used the fact that in the square root the highest order term is the first term in the square root. Furthermore, it was used that,

$$\coth^2(\frac{1}{2}\delta T) = \frac{\cosh(\delta T)+1}{\cosh(\delta T)-1}.$$

For the last equality, note that

$$\lim_{T\to 0}\frac{\cosh(\delta T)+1}{4(\delta-1)\cosh(\delta T)+(\delta^2+(2-\delta)^2)} = \frac{1}{\delta^2}$$

Note that when $\delta = 2$, $C = 0$ but the fraction $\frac{C}{2-\delta} \approx \frac{b-a}{\delta^3 T}$ for $T \to 0$. Now that we have obtained the asymptotic behaviour of C we can look at what happens to $S_T(a,b)+ET$. First of all we can see from (A.4) in the appendices that for $T \to 0$,

$$s \approx \sqrt{\frac{4(\gamma C^2 + C^2)}{T^2}} \approx \frac{2(2-\delta)(b-a)}{\delta^2 T^2}$$

Now if we look at the first part of (3.40) we see that,

$$\tan^{-1}\Big(\frac{1}{s}\Big[\frac{2\gamma(b-C)e^{\delta T} - 2\gamma(a-C)}{\sinh\delta T} + \frac{4\gamma C}{1-\gamma}\Big]\Big) - \tan^{-1}\Big(\frac{1}{s}\Big[\frac{2\gamma(b-C) - 2\gamma(a-C)e^{-\delta T}}{\sinh\delta T} + \frac{4\gamma C}{1-\gamma}\Big]\Big) \approx cT^2$$

as $e^{\delta T}$ and $e^{-\delta T}$ both converge to 1. In line 1 of equation (3.40) this is multiplied with $\frac{s}{2\gamma}$ and from this we see that line 1 is of order 1. Exactly the same argument shows that also line 2 of equation (3.40) is of order 1.

Now going to the third line of equation (3.40) we get,

$$
\begin{aligned}
&\Big[\frac{C}{2-\delta} + \frac{(b-C)e^{\delta T} - (a-C)}{2\sinh(\delta T)}\Big]\log\Big(\frac{(b-C)(\gamma e^{2\delta T}+1) - (a-C)e^{\delta T}\delta}{\sinh(\delta T)} + \frac{4e^{\delta T}\gamma C}{2-\delta}\Big) \\
&\approx \Big[\frac{C}{2-\delta} + \frac{(b-C)e^{\delta T} - (a-C)}{2\sinh(\delta T)}\Big]\cdot\log(\frac{(b-C)\delta - (a-C)\delta}{T} + \frac{4\gamma(b-a)}{\delta^3 T}) \\
&\approx \Big[\frac{C}{2-\delta} + \frac{(b-C)e^{\delta T} - (a-C)}{2\sinh(\delta T)}\Big]\cdot\log(T^{-1})
\end{aligned}
$$

For the fourth line we can do the same. The logarithm converges the same and we can combine terms. In this way we write for the third and fourth line,

$$\begin{aligned}&\Big[\frac{C}{2-\delta}+\frac{(b-C)e^{\delta T}-(a-C)}{2\sinh(\delta T)}-\frac{C\gamma}{2-\delta}+\frac{(a-C)-(b-C)e^{-\delta T}}{2\sinh(\delta T)}\Big]\cdot\log(T^{-1})\\&=\Big[C+\frac{(b-C)2\sinh(\delta T)}{2\sinh(\delta T)}\Big]\cdot\log(T^{-1})\\&=b\log(T^{-1})\end{aligned}$$

The logarithm on the fifth line has the same asymptotic behaviour as the logarithm above. The rest on lines five and six in (3.40) are of order one or higher. Combining all the lines for $T \to 0$, and noting that $ET \to 0$ for $T \to 0$, we find the following small time asymptotics for the action,

$$S_T(a,b) \approx \frac{|b-a|}{2}\log(T^{-1}) \tag{3.42}$$

As follows from the above these small time asymptotics follow for all γ, thus for the symmetric as well as the asymmetric case. The asymptotic behaviour for small T found above corresponds with the results in the paper [4], p.5. Here the small time asymptotics are found by noting that for small $T, b > a$,

$$S_T(a,b) \approx \xi_0(b-a) - ET \tag{3.43}$$

Now, as ξ_0 only depends on $\dot{x}_0$ and the energy E, the asymptotics can be found by filling out the asymptotics of these two quantities in the formula for ξ_0.

3.3.2 Large Time Asymptotics

Now that we have seen what happens for $T \to 0$ we also want to know what happens if the time T goes to infinity. In other words, the amount of time the system has to go from state a to state b becomes arbitrary big. First, we look again at the asymptotic behaviour of C for $T \to \infty$.

We will use similar notation as before. Two function $f(T)$ and $g(T)$ are asymptotically equivalent if and only if $\lim_{T\to\infty}|\frac{f(T)}{g(T)}| = 1$. Again, in this case, we will write $f \approx g$. From equation (3.34) we get,

$$\begin{aligned}C=&\frac{(2-\delta)^2a(\cosh(\delta T)+1)-\delta^2\coth^2(\frac{1}{2}\delta T)(a\cosh(\delta T)-b)}{4(\delta-1)\cosh(\delta T)+(\delta^2+(2-\delta)^2)}\\&+\frac{(2-\delta)\sinh(\delta T)}{4\delta(\delta-1)\cosh(\delta T)+\delta^3+\delta(2-\delta)^2}\cdot\Bigg[\delta^4\Big(\frac{-b+a\cosh(\delta T)}{\cosh(\delta T)-1}\Big)^2-2\delta^2(2-\delta)^2a\cdot\Big(\frac{-b+a\cosh(\delta T)}{\cosh(\delta T)-1}\Big)\\&+(2-\delta)^4a^2+4(\delta-1)(1-a^2)\Big(\delta^2\coth^2(\frac{1}{2}\delta T)-(2-\delta)^2\Big)\Bigg]^{\frac{1}{2}}+\frac{a\cosh(\delta T)-b}{\cosh(\delta T)-1}\\&\approx -a+\frac{2-\delta}{4(\delta-1)}\sqrt{\delta^4a^2-2\delta^2(2-\delta)^2a^2+(2-\delta)^4a^2+16(\delta-1)^2(1-a^2)}+a\\&=\frac{2-\delta}{4(\delta-1)}\sqrt{16\frac{(\delta-1)^2}{\delta^2}}\\&=\frac{2-\delta}{\delta}\end{aligned} \tag{3.44}$$

Note that this limit of C for $T \to \infty$ is exactly equal to the equilibrium value of x. Looking at the trajectory of x in (3.23) in section 3.2.3 we see that this value of C is what to expect. Namely, for t far away from zero but not close yet to the final time T both terms $\frac{\sinh(\delta t)}{\sinh(\delta T)}$ and $\frac{\sinh(\delta(T-t))}{\sinh(\delta T)}$ are approximately zero and we have for this t that $x(t) \approx C$. In other words this confirms that x moves to its equilibrium value. Now we can look at what happens to s in (A.4) for $T \to \infty$. It is easily seen that the first fraction in the square root goes to zero. For the second fraction we get a constant such that s for $T \to \infty$ is given by,

$$s = \sqrt{-16\gamma^2\left(\frac{C}{2-\delta}\right)^2} = i \cdot D_3 \tag{3.45}$$

From this we see that s goes to an imaginary number for $T \to \infty$ (D_3 is defined in appendix A, (A.2), p.53). Now we use the fact that we can rewrite an inverse tangens as a natural logarithm using the following equation[1],

$$tan^{-1}(z) = \frac{1}{2 \cdot i}\log\left(\frac{1+i\cdot z}{1-i\cdot z}\right). \tag{3.46}$$

Combining (3.45) and (3.46) and looking at the asymptotic behaviour, we can write the first line of (3.40) as,

$$\begin{aligned}
&\frac{s}{2\gamma}\left(\tan^{-1}\left(\frac{1}{s}\left[\frac{2\gamma(b-C)e^{\delta T}-2\gamma(a-C)}{\sinh\delta T}+\frac{4\gamma C}{1-\gamma}\right]\right)-\tan^{-1}\left(\frac{1}{s}\left[\frac{2\gamma(b-C)-2\gamma(a-C)e^{-\delta T}}{\sinh\delta T}+\frac{4\gamma C}{1-\gamma}\right]\right)\right)\\
&\approx \frac{D_3}{4\gamma}\left[\log\left(\frac{2+\frac{(2-\delta)(b-C)}{C}}{-\frac{(2-\delta)(b-C)}{C}}\right)-\log\left(\frac{2+\frac{(2-\delta)(b-C)e^{-\delta T}}{C}}{-\frac{(2-\delta)(b-C)e^{-\delta T}}{C}}\right)\right]\\
&\approx \frac{1}{\delta}\left[\log\left(\frac{2+\delta(b-C)}{-\delta(b-C)}\right)-\log\left(\frac{2+\delta(b-C)e^{-\delta T}}{-\delta(b-C)e^{-\delta T}}\right)\right]\\
&= \frac{1}{\delta}\log\left(\frac{\big(2+\delta(b-C)\big)e^{-\delta T}}{2+\delta(b-C)e^{-\delta T}}\right).
\end{aligned}$$

The same can be done for the second line of (3.40). We get,

$$\begin{aligned}
&-\frac{s}{2}\left(\tan^{-1}\left(\frac{1}{s}\left[\frac{-2(b-C)e^{-\delta T}+2(a-C)}{\sinh\delta T}+\frac{4\gamma C}{1-\gamma}\right]\right)-\tan^{-1}\left(\frac{1}{s}\left[\frac{-2(b-C)+2(a-C)e^{\delta T}}{\sinh\delta T}+\frac{4\gamma C}{1-\gamma}\right]\right)\right)\\
&\approx -\frac{D_3}{4}\left[\log\left(\frac{2+\frac{(2-\delta)(a-C)e^{-\delta T}}{\gamma C}}{-\frac{(2-\delta)(a-C)e^{-\delta T}}{\gamma C}}\right)-\log\left(\frac{2+\frac{(2-\delta)(a-C)}{\gamma C}}{-\frac{(2-\delta)(a-C)}{\gamma C}}\right)\right]\\
&\approx -\frac{\gamma}{\delta}\left[\log\left(\frac{2+\frac{\delta(a-C)e^{-\delta T}}{\gamma}}{-\frac{\delta(a-C)e^{-\delta T}}{\gamma}}\right)-\log\left(\frac{2+\frac{\delta(a-C)}{\gamma}}{-\frac{\delta(a-C)}{\gamma}}\right)\right]\\
&= -\frac{\gamma}{\delta}\log\left(\frac{2+\frac{\delta(a-C)e^{-\delta T}}{\gamma}}{(2+\frac{\delta(a-C)}{\gamma})e^{-\delta T}}\right).
\end{aligned}$$

Now going to the third line of (3.40) we have,

$$\begin{aligned}
&\left[\frac{C}{2-\delta}+\frac{(b-C)e^{\delta T}-(a-C)}{2\sinh(\delta T)}\right]\log\left(\frac{(b-C)(\gamma e^{2\delta T}+1)-(a-C)e^{\delta T}\delta}{\sinh(\delta T)}+\frac{4e^{\delta T}\gamma C}{2-\delta}\right)\\
&\approx \left[\frac{C}{2-\delta}+(b-C)\right]\lim_{T\to\infty}\log\left((\gamma(b-C)+\frac{2\gamma C}{2-\delta})2e^{\delta T}\right)\\
&= \left[\frac{C}{2-\delta}+(b-C)\right]\left(\log\left(\gamma(b-C)+\frac{2\gamma C}{2-\delta}\right)+\log(2)+\delta T\right).
\end{aligned}$$

Where we split the logarithm in three parts for reasons that become clear later on. We see in the above that this part diverges for $T \to \infty$. Again this can be done for the fourth line, resulting in,

[1]See as a reference for instance http://mathworld.wolfram.com/NaturalLogarithm.html

$$\Big[-\frac{C\gamma}{2-\delta}+\frac{(a-C)-(b-C)e^{-\delta T}}{2\sinh(\delta T)}\Big]\log\Big(\frac{(b-C)(e^{-2\delta T}+\gamma)-(a-C)e^{-\delta T}\delta}{\sinh(\delta T)}+\frac{4e^{-\delta T}\gamma C}{2-\delta}\Big)$$
$$\approx -\frac{C\gamma}{2-\delta}\left(\log\Big(\frac{4\gamma C}{2-\delta}\Big)-\delta T\right).$$

And going to the first part of the fifth line we get,

$$-a\log\Big(\frac{(b-C)\delta-(a-C)(\gamma e^{-\delta T}+e^{\delta T})}{\sinh(\delta T)}+\frac{4\gamma C}{(2-\delta)}\Big)$$
$$\approx -a\left(\log\Big(\frac{2\gamma C}{2-\delta}-\gamma(a-C)\Big)+\log(2)\right).$$

Finally, for the rest on line five and line six of (3.40) we get,

$$\Big[(a-b)\log(2)-\delta T(b-C)\tanh^{-1}(\delta T)+\delta T(a-C)\sinh^{-1}(\delta T)+(1-b)\log(1-b)-(1-a)\log(1-a)\Big]$$
$$\approx (a-b)\log(2)-\delta T(b-C)+(1-b)\log(1-b)-(1-a)\log(1-a)$$

Now we can combine everything to get the asymptotic behaviour of $S_T(a,b)$. Note that we have to add ET to (3.40) first. Doing this we get the following expression,

$$\begin{aligned}
S_T(a,b) \approx{}& \frac{1}{2}\Bigg[\frac{1}{\delta}\log\left(\frac{(2+\delta(b-C))e^{-\delta T}}{2+\delta(b-C)e^{-\delta T}}\right)-\frac{\gamma}{\delta}\log\left(\frac{2+\frac{\delta(a-C)e^{-\delta T}}{\gamma}}{(2+\frac{\delta(a-C)}{\gamma})e^{-\delta T}}\right)\\
&+\Big[\frac{C}{2-\delta}+(b-C)\Big]\left(\log\Big(\gamma(b-C)+\frac{2\gamma C}{2-\delta}\Big)+\log(2)+\delta T\right)-\frac{C\gamma}{2-\delta}\left(\log\Big(\frac{4\gamma C}{2-\delta}\Big)-\delta T\right)\\
&-a\left(\log\Big(\frac{2\gamma C}{2-\delta}-\gamma(a-C)\Big)+\log(2)\right)+(a-b)\log(2)-\delta T(b-C)+(1-b)\log(1-b)\\
&-(1-a)\log(1-a)-2ET\Bigg]\\
\approx{}& \frac{1}{2}\Big(\frac{-\delta}{\delta}-\frac{\gamma\delta}{\delta}+\frac{C\delta}{2-\delta}+(b-C)\delta+\frac{C\gamma\delta}{2-\delta}-\delta(b-C)-2E\Big)T\\
={}& \frac{1}{2}\Big(-1-\gamma+\frac{\delta^2 C}{2-\delta}-2E\Big)T\\
={}& \frac{1}{2}\Big(-\delta+2E+\delta-2E\Big)T\\
={}& 0+O(1)
\end{aligned} \tag{3.47}$$

This is what we would expect for the Action. Because we have seen in the previous section that the system goes to the equilibrium value if $T >> 1$. Furthermore, we have seen that the relative period of moving towards b goes to zero for $T \to \infty$. Now if we look at the velocity of the trajectory at $t = 0$ for $T \to \infty$, we get,

$$\begin{aligned}
\dot{x}_0 ={}& \frac{a\delta\sinh(\delta T)-\frac{\delta^3}{(2-\delta)^2}\coth(\frac{1}{2}\delta T)(a\cosh(\delta T)-b)}{(\frac{\delta^2}{(2-\delta)^2}-1)\cosh(\delta T)+(\frac{\delta^2}{(2-\delta)^2}+1)}\\
&+\frac{(2-\delta)^2}{\delta^2\coth^2(\frac{1}{2}\delta T)-(2-\delta)^2}\cdot\Bigg[\frac{\delta^4}{(2-\delta)^2}\Big(\frac{-b+a\cosh(\delta T)}{\cosh(\delta T)-1}\Big)^2-2\delta^2 a\cdot\Big(\frac{-b+a\cosh(\delta T)}{\cosh(\delta T)-1}\Big)\\
&+(2-\delta)^2a^2+4(\delta-1)(1-a^2)\Big(\frac{\delta^2}{(2-\delta)^2}\coth^2(\frac{1}{2}\delta T)-1\Big)\Bigg]^{\frac{1}{2}}
\end{aligned} \tag{3.48}$$

$$\begin{aligned}
&\approx \frac{a\delta - \frac{\delta^3}{(2-\delta)^2}a}{\frac{\delta^2}{(2-\delta)^2}-1} + \frac{(2-\delta)^2}{4(\delta-1)}\sqrt{a^2\frac{\delta^4}{(2-\delta)^2} - 2\delta^2a^2 + (2-\delta)^2a^2 + 16\frac{(\delta-1)^2}{(2-\delta)^2}(1-a^2)} \\
&= -\delta a + \frac{(2-\delta)^2}{4(\delta-1)}\sqrt{16\frac{(\delta-1)^2}{(2-\delta)^2}} \\
&= -\delta a + (2-\delta)
\end{aligned}$$

This is exactly the typical speed that we found in section 3.2 on page 39. Now filling out this typical speed in $\log(\xi(t))$, and noting that when we take $T \to \infty$, that C is equal to the equilibrium value. We see for $\log(\xi(t))$ at $t = 0$,

$$\begin{aligned}
\lim_{T\to\infty} \log(\xi(0)) &= \log\Big(\frac{-\delta a + (2-\delta) + 2E + (1+\gamma) - (1-\gamma)a}{2(1-a)}\Big) \\
&= \log\Big(\frac{-\delta a + (2-\delta) + \frac{\delta^2 C}{2-\delta} - (1-\gamma)a}{2(1-a)}\Big) \\
&= \log\Big(\frac{-\delta a + (2-\delta) + \delta - (1-\gamma)a}{2(1-a)}\Big) \\
&= \log\Big(\frac{-2a+2}{2(1-a)}\Big) = 0
\end{aligned} \tag{3.49}$$

This holds for any finite t as we can treat this point as the starting point. Say at some time $t = t^*$ we are in some point f. Now the speed in this point will be $-\delta f + (2-\delta)$ and with the same computation we see that $\log(\xi(t^*)) = 0$. This argument can also more directly be shown by looking at equation (3.21). This equation gives $x(t)$ and taking the derivative with respect to time we get $\dot{x}(t)$. Using this we see directly that for $T \to \infty$,

$$\begin{aligned}
x(t) &= (a-C)e^{-\delta t} + \frac{2-\delta}{\delta} \\
\dot{x}(t) &= -\delta(a-C)e^{-\delta t} = -\delta x + (2-\delta)
\end{aligned} \tag{3.50}$$

and thus, that $\dot{x}(t)$ has the typical speed for all finite t and $\lim_{T\to\infty}\log(\xi(t)) = 0$. Now we needed to check this by computing the integral as it may not be allowed to interchange limit and integral. In appendix B, at the end of this report, we compute the Action integral for the special case that $\gamma = 1$ (i.e. the symmetric case). In this case we are able to use equation (3.15). We show that for some values of a and n we can not just fill them out in the expression of the Action but we need a more subtle approach to attain the correct value of the Action. Furthermore we show here that for all typical paths in the symmetric case, the Action is equal to zero.

3.4 Conclusion

In this chapter we have looked at the large deviation behaviour of the continuous-time two-state Markov process. We have computed the trajectories of this process. These trajectories move, if time allows it, from there starting value a towards the equilibrium value and subsequently to there ending value b. If the ending time T is very small, the trajectories become approximately linear between begin and end-point. For large T we have seen that the relative time that the process moves towards b goes to zero, and this time halves for doubling T. After this, the Action Integral was examined. First, the Action was computed by solving the integral. After this the limiting behaviour was investigated for $T \to 0$ and $T \to \infty$. In the first case, we have seen that the limiting behaviour of $S_T(a,b)$ goes like $\frac{|a-b|}{2}\log(T^{-1})$. So, $S_T(a,b)$ gets very big for small values of T. This is intuitive, as for small values of T the paths to move from a towards b have to be very abrupt (and thus the probability of this gets very small). Only in the case that $b = a$, the path becomes typical for very small T as in this case obviously we don't have to move. In the case $T \to \infty$ we see that the Action vanishes. Also this is intuitive. We know after all that the trajectory goes to the typical path and can stay here arbitrarily long.

As a follow-up it would be interesting to look at the case where the number of states of the Markov-process get extended. In this case the dynamics can get a lot more complicated as there are more ways (via more states) to get to the equilibrium position or the ending position. However, it is suspected that also in this case, for T big enough, the system returns to equilibrium and subsequently goes to its ending value. The way it goes towards equilibrium and towards the ending value is of course the way that has the highest probability, i.e. minimizes the Action. Another interesting addition would be if the parameter γ would be dependent of the time. The difficulty that then arises however is that the Hamiltonian is not longer time independent, i.e. the energy is not a conserved quantity any more.

Appendices

Appendix A

Calculation of the Action Integral

In this first appendix we are going to derive an explicit equation for the Action integral. This can then be used for determining the asymptotic behaviour for $T \to 0$ and $T \to \infty$. To derive the explicit equation we see from (3.39) that we have to compute the integral over $\log(\xi(t))$ with respect to x. Using equation (3.21) and equation (3.30) from the previous chapter we get,

$$\begin{aligned}\log(\xi(t)) &= \log\Big(\frac{\dot{x}(t)+2E+(1+\gamma)-(1-\gamma)x(t)}{2(1-x(t))}\Big)\\ &= \log(D_1e^{\delta t}+D_2e^{-\delta t}+D_3)-\log(2(1-x(t)))\end{aligned} \tag{A.1}$$

where,

$$\begin{aligned}D_1 &= C_1\delta-(1-\gamma)C_1 = 2C_1\gamma\\ D_2 &= -C_2\delta-(1-\gamma)C_2 = -2C_2\\ D_3 &= 2E+(1+\gamma)-(1-\gamma)C\end{aligned} \tag{A.2}$$

and where the constants C_1, C_2 and C are as in the equations (3.22) p.38 in the previous section. From this it is seen that the Action is of the following form.

$$\begin{aligned}2S_T(a,b)+2ET &= \int_0^T \log(\xi(t))dx(t)\\ &= C_1\delta\int_0^T \log(D_1e^{\delta t}+D_2e^{-\delta t}+D_3)e^{\delta T}dt\\ &\quad -C_2\delta\int_0^T \log(D_1e^{\delta t}+D_2e^{-\delta t}+D_3)e^{-\delta T}dt-\int_a^b \log(2(1-x))dx\\ &= C_1\delta\int_0^T \log(D_1e^{\delta t}+D_2e^{-\delta t}+D_3)e^{\delta T}dt\\ &\quad -C_2\delta\int_0^T \log(D_1e^{\delta t}+(1-b)\log(1-b)-(1-a)\log(1-a)\\ &= C_1\delta\int_0^T \log(D_1e^{\delta t}+D_2e^{-\delta t}+D_3)e^{\delta T}dt\\ &\quad -C_2\delta\int_0^T \log(D_1e^{\delta t}+D_2e^{-\delta t}+D_3)e^{-\delta T}dt+(b-a)\log(\frac{e}{2})\\ &\quad +(1-b)\log(1-b)-(1-a)\log(1-a)\end{aligned} \tag{A.3}$$

Looking at the first integral in the bottom two lines above we find,

$$
\begin{aligned}
&C_1\delta\int_0^T \log(D_1e^{\delta t}+D_2e^{-\delta t}+D_3)e^{\delta T}\,dt\\
&\quad = C_1\int_0^T (\log(D_1e^{\delta t}+D_2e^{-\delta t}+D_3)+\delta t)\delta e^{\delta T}-\delta^2 te^{\delta T}\,dt\\
&\quad = C_1\int_0^T (\log(D_1e^{2\delta t}+D_3e^{\delta t}+D_2)\delta e^{\delta T}\,dt - C_1\int_0^T \delta^2 te^{\delta T}\,dt\\
&\quad = C_1\int_1^{e^{\delta T}} (\log(D_1u^2+D_3u+D_2)du + C_1\big(-\delta Te^{\delta T}-1+e^{\delta T}\big)\\
&\quad = C_1\Bigg[\frac{1}{D_1}\sqrt{4D_1D_2-D_3^2}\tan^{-1}\Big(\frac{2D_1u+D_3}{\sqrt{4D_1D_2-D_3^2}}\Big)\\
&\qquad -2u+\big(\frac{D_3}{2D_1}+u\big)\log(D_1u^2+D_3u+D_2)\Bigg]_1^{e^{\delta T}} + C_1\big(-\delta Te^{\delta T}-1+e^{\delta T}\big)
\end{aligned}
$$

Where in the next to last equality we substituted $u=e^{\delta T}$. Doing the same for the second integral (but now substituting $u=e^{-\delta T}$) in the last equation in (A.3) we get,

$$
\begin{aligned}
&-C_2\delta\int_0^T \log(D_1e^{\delta t}+D_2e^{-\delta t}+D_3)e^{-\delta T}\,dt\\
&\quad = C_2\int_1^{e^{-\delta T}} \log(D_2u^2+D_3u+D_1)du - C_2\big(-\delta Te^{-\delta T}+1-e^{-\delta T}\big)\\
&\quad = C_2\Bigg[\frac{1}{D_2}\sqrt{4D_1D_2-D_3^2}\tan^{-1}\Big(\frac{2D_2u+D_3}{\sqrt{4D_1D_2-D_3^2}}\Big)\\
&\qquad -2u+\big(\frac{D_3}{2D_2}+u\big)\log(D_2u^2+D_3u+D_1)\Bigg]_1^{e^{-\delta T}} - C_2\big(-\delta Te^{-\delta T}+1-e^{-\delta T}\big)
\end{aligned}
$$

Now we take the first part of the two integrals together. We will write $s:=\sqrt{4D_1D_2-D_3^2}$. We get

$$
\begin{aligned}
&C_1\left[\frac{1}{D_1}s\tan^{-1}\Big(\frac{2D_1u+D_3}{s}\Big)\right]_1^{e^{\delta T}} + C_2\left[\frac{1}{D_2}s\tan^{-1}\Big(\frac{2D_2u+D_3}{s}\Big)\right]_1^{e^{-\delta T}}\\
&= \frac{s}{2\gamma}\left(\tan^{-1}\Big(\frac{1}{s}\Big[\frac{2\gamma(b-C)e^{\delta T}-2\gamma(a-C)}{\sinh\delta T}+\frac{4\gamma C}{1-\gamma}\Big]\Big)-\tan^{-1}\Big(\frac{1}{s}\Big[\frac{2\gamma(b-C)-2\gamma(a-C)e^{-\delta T}}{\sinh\delta T}+\frac{4\gamma C}{1-\gamma}\Big]\Big)\right)\\
&\quad -\frac{s}{2}\left(\tan^{-1}\Big(\frac{1}{s}\Big[\frac{2(b-C)e^{-\delta T}-2(a-C)}{\sinh\delta T}+\frac{4\gamma C}{1-\gamma}\Big]\Big)-\tan^{-1}\Big(\frac{1}{s}\Big[\frac{2(b-C)-2(a-C)e^{\delta T}}{\sinh\delta T}+\frac{4\gamma C}{1-\gamma}\Big]\Big)\right)
\end{aligned}
$$

For s the following formula holds,

$$s = \left[\frac{4\gamma\Big((b-C)^2 + (a-C)^2 - 2(b-C)(a-C)\cosh(\delta T)}{\sinh^2(\delta T)} - \frac{16\gamma^2 C^2}{(2-\delta)^2} \right]^{\frac{1}{2}} \tag{A.4}$$

Note that C in the above equations is the same as in the previous section (C_3 in (3.22), p.38) which we made fully explicit. Now we write the second parts of the two integrals together, where again after some algebra we obtain

$$\begin{aligned}
&C_1 \left[-2u + \big(\frac{D_3}{2D_1} + u\big)\log(D_1u^2 + D_3u + D_2) \right]_1^{e^{\delta T}} + C_2 \left[-2u + \big(\frac{D_3}{2D_2} + u\big)\log(D_2u^2 + D_3u + D_1) \right]_1^{e^{-\delta T}} \\
&= \Big[\frac{C}{2-\delta} + \frac{(b-C)e^{\delta T} - (a-C)}{2\sinh(\delta T)}\Big] \log\Big(\frac{(b-C)(\gamma e^{2\delta T} + 1) - (a-C)e^{\delta T}\delta}{\sinh(\delta T)} + \frac{4e^{\delta T}\gamma C}{2-\delta}\Big) \\
&+ \Big[-\frac{C\gamma}{2-\delta} + \frac{(a-C) - (b-C)e^{-\delta T}}{2\sinh(\delta T)}\Big] \log\Big(\frac{(b-C)(e^{-2\delta T} + \gamma) - (a-C)e^{-\delta T}\delta}{\sinh(\delta T)} + \frac{4e^{-\delta T}\gamma C}{2-\delta}\Big) \\
&- a\log\Big(\frac{(b-C)\delta - (a-C)(\gamma e^{-\delta T} + e^{\delta T})}{\sinh(\delta T)} + \frac{4\gamma C}{(2-\delta)}\Big) + 2(a-b)
\end{aligned}$$

Finally, looking at the third parts of the two integrals we get,

$$\begin{aligned}
&C_1\big(-\delta T e^{\delta T} - 1 + e^{\delta T}\big) - C_2\big(-\delta T e^{-\delta T} + 1 - e^{-\delta T}\big) \\
&= \frac{(b-C)(e^{\delta T} - e^{-\delta T}) - (a-C)(e^{\delta T} - e^{-\delta T}) - (b-C)\delta T(e^{\delta T} + e^{-\delta T}) + 2\delta T(a-C)}{2\sinh(\delta T)} \\
&= (b-a) - \delta T(b-C)\tanh^{-1}(\delta T) + \delta T(a-C)\sinh^{-1}(\delta T)
\end{aligned}$$

Taking everything together that we found in (A.3) we obtain the following formula for $S_T(a,b) + ET$ only dependent on a, b, γ, T, s and the explicit constant C.

$$\begin{aligned}
&2S_T(a,b) + 2ET = \\
&\frac{s}{2\gamma}\left(\tan^{-1}\Big(\frac{1}{s}\Big[\frac{2\gamma(b-C)e^{\delta T} - 2\gamma(a-C)}{\sinh\delta T} + \frac{4\gamma C}{1-\gamma}\Big]\Big) - \tan^{-1}\Big(\frac{1}{s}\Big[\frac{2\gamma(b-C) - 2\gamma(a-C)e^{-\delta T}}{\sinh\delta T} + \frac{4\gamma C}{1-\gamma}\Big]\Big)\right) \\
&- \frac{s}{2}\left(\tan^{-1}\Big(\frac{1}{s}\Big[\frac{2(b-C)e^{-\delta T} - 2(a-C)}{\sinh\delta T} + \frac{4\gamma C}{1-\gamma}\Big]\Big) - \tan^{-1}\Big(\frac{1}{s}\Big[\frac{2(b-C) - 2(a-C)e^{\delta T}}{\sinh\delta T} + \frac{4\gamma C}{1-\gamma}\Big]\Big)\right) \\
&+ \Big[\frac{C}{2-\delta} + \frac{(b-C)e^{\delta T} - (a-C)}{2\sinh(\delta T)}\Big] \log\Big(\frac{(b-C)(\gamma e^{2\delta T} + 1) - (a-C)e^{\delta T}\delta}{\sinh(\delta T)} + \frac{4e^{\delta T}\gamma C}{2-\delta}\Big) \\
&+ \Big[-\frac{C\gamma}{2-\delta} + \frac{(a-C) - (b-C)e^{-\delta T}}{2\sinh(\delta T)}\Big] \log\Big(\frac{(b-C)(e^{-2\delta T} + \gamma) - (a-C)e^{-\delta T}\delta}{\sinh(\delta T)} + \frac{4e^{-\delta T}\gamma C}{2-\delta}\Big) \\
&- a\log\Big(\frac{(b-C)\delta - (a-C)(\gamma e^{-\delta T} + e^{\delta T})}{\sinh(\delta T)} + \frac{4\gamma C}{(2-\delta)}\Big) + (a-b)\log(2) - \delta T(b-C)\tanh^{-1}(\delta T) \\
&+ \delta T(a-C)\sinh^{-1}(\delta T) + (1-b)\log(1-b) - (1-a)\log(1-a)
\end{aligned} \tag{A.5}$$

Appendix B

The Action for the Symmetric case

Now we will compute the Action in the special case that $\gamma = 1$, i.e. the symmetric case. In order to be able to use equation (3.15) we compute the action for the symmetric case separately. Now we can combine (3.14) and (3.15) to obtain $\xi(t)$ and recalculate the Action. We have,

$$\begin{aligned} \log(\xi(t)) &= \log\Big(\frac{\dot{x}(t)+\sqrt{\dot{x}^2+4(1-x^2)}}{2(1-x)}\Big) \\ &= \log\Big(\frac{2C_1e^{2t}-2C_2e^{-2t}+\sqrt{(4-16C_1C_2)}}{2(1-C_1e^{2t}-C_2e^{-2t})} \end{aligned} \tag{B.1}$$

So, to obtain the action we compute the following integral,

$$\begin{aligned} &\int_0^T \log\Big(\frac{\dot{x}(t)+\sqrt{\dot{x}^2+4(1-x^2)}}{2(1-x)}\Big)dx = \int_0^T \log\Big(\frac{2C_1e^{2t}-2C_2e^{-2t}+\sqrt{(4-16C_1C_2)}}{2(1-C_1e^{2t}-C_2e^{-2t})}\Big)\Big(2C_1e^{2t}-2C_2e^{-2t}\Big)dt \\ &= C_1\int_1^{e^{\delta T}} \log\Big(2C_1u^2+\sqrt{(4-16C_1C_2)}u-2C_2\Big)-\log\Big(2u-2C_1u^2-2C_2\Big)du \\ &\quad + C_2\int_1^{e^{-\delta T}} \log\Big(-2C_2u^2+\sqrt{(4-16C_1C_2)}u+2C_1\Big)-\log\Big(2u-2C_2u^2-2C_1\Big)du \end{aligned} \tag{B.2}$$

Where for the last equality we used the same integration by parts strategy as before. Note that when a and b are zero also C_1 and C_2 are zero and the above integrals vanish. Not only because the constants before the integrals are zero but also because the terms inside the integrals exactly cancel each other ($\sqrt{4-16C_1C_2} = 2$ for a and b equal to zero). Computing the above integral gives,

$$\begin{aligned} 2S_T(a,b)+2ET = &\Bigg[\frac{\sqrt{-4}}{2}\tan^{-1}\Big(\frac{4C_1u+k}{\sqrt{-4}}\Big)+\frac{k}{4}\log(2C_2-u(2C_1u+k))+C_1u\log(2C_1u^2+ku-2C_2) \\ &-C_1u\log(-2C_1u^2+2u-2C_2)+\frac{1}{2}\log(C_1u^2-u+C_2)-\sqrt{4C_1C_2-1}\tan^{-1}\Big(\frac{2C_1u-1}{\sqrt{4C_1C_2-1}}\Big)\Bigg]_1^{e^{\delta T}} \\ &+\Bigg[\frac{\sqrt{-4}}{2}\tan^{-1}\Big(\frac{4C_2u-k}{\sqrt{-4}}\Big)-(\frac{k}{4}-C_2u)\log(2C_1+uk-2C_2u^2)-C_2u\log(-2C_1-2C_2u^2+2u) \\ &+\frac{1}{2}\log(C_1+C_2u^2-u)-\sqrt{4C_1C_2-1}\tan^{-1}\Big(\frac{2C_2u-1}{\sqrt{4C_1C_2-1}}\Big)\Bigg]_1^{e^{-\delta T}} \end{aligned} \tag{B.3}$$

$$= \left[\frac{1}{2}\log\left(\frac{1+2C_1u+\frac{1}{2}k}{1-2C_1u-\frac{1}{2}k}\right)+\frac{k}{4}\log(2C_2-u(2C_1u+k))+C_1u\log(2C_1u^2+ku-2C_2)\right.$$
$$\left.-C_1u\log(-2C_1u^2+2u-2C_2)+\frac{1}{2}\log(C_1u^2-u+C_2)+\frac{k}{4}\log\left(\frac{\frac{k}{2}+2C_1u-1}{\frac{k}{2}-2C_1u+1}\right)\right]_1^{e^{\delta T}}$$
$$+\left[\frac{1}{2}\log\left(\frac{1+2C_2u-\frac{1}{2}k}{1-2C_2u+\frac{1}{2}k}\right)-(\frac{k}{4}-C_2u)\log(2C_1+uk-2C_2u^2)-C_2u\log(-2C_1-2C_2u^2+2u)\right.$$
$$\left.+\frac{1}{2}\log(C_1+C_2u^2-u)+\frac{k}{4}\log\left(\frac{\frac{k}{2}+2C_2u-1}{\frac{k}{2}-2C_2u+1}\right)\right]_1^{e^{-\delta T}}$$

Where we use $k := \sqrt{4-16C_1C_2}$. Furthermore we used again the fact that the inverse tangens of a complex number can be written as a complex natural logarithm as follows,

$$\tan^{-1}(z) = \frac{1}{2\cdot i}\log\left(\frac{1+i\cdot z}{1-i\cdot z}\right).$$

Finally we used that $\sqrt{4C_1C_2-1} = -i\cdot\sqrt{1-4C_1C_2}$.

Now we take the special situation with $a=b=0$ and consequently $C_1=C_2=0$ and $k=2$. Filling this out directly in the above solution of the integral, a lot of terms cancel and the following remains,

$$\begin{aligned}2S_T(0,0) &= \frac{1}{2}\log(-2e^{\delta T})-\frac{1}{2}\log(-2)+\frac{1}{2}\log(-e^{\delta T})-\frac{1}{2}\log(-1)-\frac{1}{2}\log(2e^{-\delta T})+\frac{1}{2}\log(2)\\ &\quad+\frac{1}{2}\log(-e^{-\delta T})-\frac{1}{2}\log(-1)\\ &= \frac{1}{2}\log(e^{\delta T})+\frac{1}{2}\log(e^{\delta T})+\frac{1}{2}\log(e^{\delta T})+\frac{1}{2}\log(e^{-\delta T})\\ &= \delta T\end{aligned} \tag{B.4}$$

Now we see, we get two contradicting results. On the one hand, before calculating the integral, we see that filling out $a=b=0$ in (B.2) we get two definite integrals over the zero function which clearly should be zero. On the other hand, after calculating the integral and filling out these values for a and b directly not all terms cancel and a δT term remains. The reason for this is that we can not fill out a and b directly. This can be seen by the fact that although the logarithms that replaced the inverse tangens' do not depend on u when $a=b=0$ and thus drop out by taking the bounds of integration, the logarithms are taken over 0 or $\frac{1}{0}$ respectively. To properly calculate the action we need to use a more subtle approach. Note that in general, the value of b is the typical value to be attained after time T if $b = ae^{-2T}$ (Note that $a=b=0$ is a special case of this typical behaviour). Now We can fill out this value for b, and look what happens with the expression for $S_T(a,b)$. Again we suspect the action has to vanish as the Lagrangian of this system is the zero function. Filling out this value for C_1 and C_2 we get $C_1=0$ and $C_2=a$. So in this case still $k=2$. We fill this out in equation (B.3) and the following remains,

$$\begin{aligned}2S_T(a,ae^{-2T})+2ET &= \left[\frac{1}{2}\log(2a-2u)+\frac{1}{2}\log(a-u)\right]_1^{e^{\delta T}}+\left[\frac{1}{2}\log\left(\frac{2au}{2-2au}\right)-(\frac{1}{2}-au)\log(2u-2au^2)\right.\\ &\quad\left.-au\log(-2au^2+2u)+\frac{1}{2}\log(au^2-u)+\frac{1}{2}\log\left(\frac{2au}{2-2au}\right)\right]_1^{e^{-\delta T}}\end{aligned} \tag{B.5}$$

$$= \left[\frac{1}{2}\log(2(a-u)^2)\right]_1^{e^{\delta T}} + \left[\log\left(\frac{2au}{2-2au}\right) - \frac{1}{2}\log(2u-2au^2) + \frac{1}{2}\log(au^2-u)\right]_1^{e^{-\delta T}}$$

$$= \frac{1}{2}\log\left(\frac{2(a-e^{\delta T})^2}{2(a-1)^2}\right) + \log\left(\frac{2ae^{-\delta T}}{2-2ae^{-\delta T}}\frac{2-2a}{2a}\right)$$

This holds for all values of a and b such that $b = ae^{-2T}$, so specifically for $a = b = 0$. Now taking the limit of $a \to 0$ (and thus also $b \to 0$) we see that we get,

$$\lim_{a\to 0} 2S_T(a, ae^{-2T}) + 2ET = \lim_{a\to 0}\frac{1}{2}\log\left(\frac{(a-e^{\delta T})^2}{(a-1)^2}\right) + \log\left(\frac{e^{-\delta T} - ae^{-\delta T}}{1-ae^{-\delta T}}\right) = \delta T - \delta T = 0 \qquad \text{(B.6)}$$

But of course we should have that the action is always zero for $b = ae^{-2T}$, not only when taking the limit for $a \to 0$. To see that this is the case note that,

$$\left(\frac{(a-e^{\delta T})^2}{(a-1)^2}\right)^{-\frac{1}{2}} = \frac{a-1}{a-e^{\delta T}} = \frac{ae^{-\delta T} - e^{-\delta T}}{ae^{-\delta T} - 1} = \frac{e^{-\delta T} - ae^{-\delta T}}{1-ae^{-\delta T}}.$$

From this it immediately follows that the action is indeed zero for all a and b such that $b = ae^{-2T}$ (in the symmetric case).

Bibliography

[1] Amir Dembo and Ofer Zeitouni, *Large Deviations Techniques and Applications*, 1998, Springer-Verlag New York

[2] Frank den Hollander, *Large Deviations*, 2000, The American Mathematical Society, Providence Rhode Island

[3] David Tong, *Classical Dynamics, University of Cambridge part II Mathematical Tripos*, 2004, University of Cambridge

[4] Frank Redig, *Hamiltonian Dynamics and optimal transport for probability measures on two-point sets*, 2015, Delft Institute of Applied Mathematics

[5] Frank Redig and Feijia Wang, *Gibbs-Non-Gibbs Transitions via Large Deviations: Computable Examples*, 2012, Springer

[6] Frank Redig, *Basic Techniques in interacting particle systems*, 2014, Delft Institute of Applied Mathematics

[7] Jean-Dominique Deuschel and Daniel W. Stroock, *Large Deviations*, 1989, Academic Press INC. London LTD.

[8] Jin Feng and Thomas G. Kurtz, *Large Deviations for Stochastic Processes*, 2005, American Mathematical Society, Providence RI