Some Large Deviation Results for Diffusion Processes

Jochen Voß
Universität Kaiserslautern

May 7, 2004

Contents

Introduction **3**

1 Diffusion Processes **7**

2 Large Deviations **11**

2.1 Introduction 11
2.2 General Principles 12
2.3 The LDP for Stationary Distributions 16
2.4 The LDP for Empirical Distributions 19
2.5 Sample Path LDP 20

3 The Ornstein-Uhlenbeck Process **23**

3.1 Introduction 23
3.2 Strong Drift 24

4 Tauberian Theorems **27**

4.1 De Bruijn's Theorem 27
4.2 An LDP for Brownian Paths with small L^2-Norm 28
4.3 Upper and Lower Limits 32

5 Diffusions with Strong Drift **37**

5.1 Reaching the Final Point 38
5.2 Staying Near the Equilibrium 52
5.3 The LDP for the Endpoint 62

6 Asymptotic Separation of Processes **75**

6.1 The Bayes Risk 75
6.2 Asymptotic Separation of OU Processes 77
6.3 Asymptotic Separation of Continuous Time Markov Chains 79

7 Computational Experiments **85**

7.1 The Euler-Maruyama Method 85
7.2 Importance Sampling 86
7.3 The Rejection Method 88
7.4 Sampling Bridges 91

List of Symbols **95**

Index **97**

Bibliography **99**

Introduction

The theory of large deviations is concerned with the study of the probabilities of very rare events. In this text we present some results which can be obtained by applying large deviation techniques to the study of diffusion processes.

The basic result in this area is Schilder's theorem about large deviations of scaled down Brownian motion. With this theorem one can calculate the exponential decay rates for probabilities of the form $P(\sqrt{\varepsilon}B \in A)$ for $\varepsilon \downarrow 0$ where B is a Brownian motion and A is a set of paths. This result is generalised by the Freidlin-Wentzell theory to the case of stochastic differential equations with small noise. The theory describes how solutions of a stochastic differential equation like

$$dX_t = b(X) \bullet dt + \sqrt{\varepsilon}\, dB$$

on a fixed time interval $[0;t]$ behave for small ε.

In the present text we place our main focus on the case of strong drift instead of small noise, i.e. on solutions of the stochastic differential equation

$$dX_t^\vartheta = \vartheta b(X^\vartheta) \bullet dt + dB$$

for large ϑ. Using a time change one can transform the case of strong drift into the case of small noise, but unfortunately the resulting equation is defined on a time interval whose length depends on the parameter ε, so the Freidlin-Wentzell theory cannot easily be applied to the time-changed process.

We will derive a large deviation result for the behaviour of X_t^ϑ for fixed t when ϑ becomes large by using a different technique. The proof uses the fact that we know the density of the distribution of X^ϑ with respect to the Wiener measure on the path space from the Girsanov formula. Assuming $b = \operatorname{grad} \Phi$ this density is

$$\varphi = \exp\bigl(\vartheta F - \vartheta^2 G\bigr)$$

with

$$F = \Phi(0) - \Phi(B_t) + \frac{1}{2}\int_0^t \Delta\Phi(B_s)\, ds$$

and

$$G = \frac{1}{2}\int_0^t b^2(B_s)\, ds.$$

For large ϑ the term ϑF can be neglected and we can use the approximation $P(X_t^\vartheta \in A) \approx E\bigl(\exp(-\vartheta^2 G) 1_A(B_t)\bigr)$. The right hand side of this relation can be considered as a Laplace transform of G, the large deviation behaviour of $P(X_t^\vartheta \in A)$ for $\vartheta \to \infty$ can be expressed in terms of the tail-behaviour of this Laplace transform. Using a Tauberian theorem we can translate questions about this tail behaviour into questions about the behaviour of the distribution of G near the origin.

Following this programme we have to estimate probabilities of the form $P(G < \varepsilon)$ for small ε. The random variable G is small when the Brownian motion spends most of the time near an equilibrium point of the drift b. To estimate the probabilities for this event we have to study two different aspects: Firstly, the process has to reach the equilibrium point very quickly. This can be treated with the help of Schilder's theorem. And secondly, once near the equilibrium point the process has to stay close to this point most of the time. Here we can use Taylor approximation of the drift to replace the process with an Ornstein-Uhlenbeck process, which is much easier to work with.

The building blocks for this proof are developed during the first chapters. The final result about the large deviation behaviour of X_t^ϑ for strong drift is presented as theorem 5.19 together with two corollaries.

The text is structured as follows. The first chapter gives a characterisation of diffusion processes as solutions of stochastic differential equations. We will only consider equations of the form

$$dX_t = b(X_t) \bullet dt + \sigma(X_t) \bullet dB_t,$$

which lead to Markovian solutions. The chapter summarises some results about these processes.

The second chapter gives a very short introduction into the theory of large deviations. We present some tools from the literature, with special emphasis on techniques which will be useful when applied to families of diffusion processes. Because diffusion processes are complex objects, large deviation results can be applied on different levels. We give results about the behaviour of stationary distributions when the drift becomes strong, about empirical distributions when the process is observed over long time intervals, and about the paths of the process when the noise is small.

Chapter 3 is devoted to an example: the Ornstein-Uhlenbeck process is the solution of the stochastic differential equation

$$dX_t = -\alpha X_t \bullet dt + dB_t$$

for some positive parameter α. Because of the simple structure of the process it is possible to calculate may things explicitly here. We derive an large deviation result for the behaviour of X_t for fixed $t \in \mathbb{R}$ when the parameter α becomes large. This is a simplified version of our main result.

In chapter 4 we present a Tauberian theorem of exponential type as another tool to obtain large deviation results. The theorem provides a connection between the behaviour of a probability distribution near the origin and the Laplace transform near infinity. Using the theorem allows us to deduce results like

$$\lim_{\varepsilon \downarrow 0} \varepsilon \cdot \log P\left(\int_0^t B_s^2 \, ds \leq \varepsilon\right) = -\frac{t^2}{8}$$

where B is a Brownian motion. As an application we derive a large deviation theorem for Brownian paths with small L^2-norm. The last part of chapter 4 derives a result about upper and lower limits in the Tauberian theorem.

The central part of this text is chapter 5. Here we combine many results from the previous chapters to derive the large deviation result for the behaviour of the endpoint of a diffusion under strong drift. It transpires that a typical path under strong drift and with given endpoint runs towards an equilibrium point of the drift quickly, stays there until near the end of the time interval, and only then moves quickly to the given endpoint. The initial and final pieces of the path can be treated with Schilder's theorem about pathwise large deviations for scaled down Brownian motion. The middle piece of each path can be treated with the Tauberian theorem from chapter 4. We give separate results for the case of attracting drift and for the case of repelling drift.

Chapter 6 gives another application of large deviation results, namely to determine the exponential decay rate for the Bayes risk when separating two different processes. Since the Bayes risk is a measure of how close the probability distributions of the two processes are, this rate describes how fast we can gain information about the processes by looking at the paths.

Finally chapter 7 describes some techniques which help to experiment with rare events for diffusion processes by means of computer simulations. We describe the Euler-Maruyama method to simulate solutions of stochastic differential equations. In subsequent sections of chapter 7 we describe how the importance sampling method can be used to estimate small probabilities and how the rejection method can be used to sample from conditional distributions where the condition has very low probability. Finally we describe how the Langevin method can be used to sample paths of a diffusion with given end point.

The table at page 95 explains some symbols and some notation used throughout the text. There is also an index which might help to access the text.

I wish to thank my supervisor Professor H. v. Weizsäcker for all his help and support in writing this text and Martin Hairer for his advice.

Chapter 1

Diffusion Processes

This introductory chapter will summarise some results about diffusion processes, which we will use later. My main reference here is the book of v. Weizsäcker and Winkler [WW90]. Other useful references are the books of Karatzas and Shreve [KS91], of Ikeda and Watanabe [IW89], and of Stroock and Varadhan [SV79].

There are many ways to characterise diffusion processes. In this thesis we will use the characterisation of a diffusion␣process as the solution of the stochastic differential equation

$$dX_t = b(X_t) \bullet dt + \sigma(X_t) \bullet dB_t \tag{1.1}$$

for some initial value $X_0 \in L^1$, where B is an n-dimensional Brownian motion, $b\colon \mathbb{R}^d \to \mathbb{R}^d$ is some drift␣function, and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times n}$ is the diffusion␣coefficient.

The basic results about existence and uniqueness of solutions for this equation (theorems IV.2.4 and IV.3.1 from [IW89]) are as follows:

Theorem 1.1. *Let $b\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times n}$ be continuous, satisfy the growth condition*

$$\|\sigma(x)\|^2 + \|b(x)\|^2 \leq K\big(1 + |x|^2\big) \quad \text{for all } x \in \mathbb{R}^d \tag{1.2}$$

for some $K > 0$, and let $E|X_0|^2 < \infty$. Then the corresponding SDE has a solution with $E|X_t|^2 < \infty$ for all $t \geq 0$.

Theorem 1.2. *Let $b\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times n}$ satisfy the following local Lipschitz condition: for every $N \in \mathbb{N}$ there is a $K_N > 0$ with*

$$\|\sigma(x) - \sigma(y)\|^2 + \|b(x) - b(y)\|^2 \leq K_N|x - y|^2 \quad \text{for every } x, y \in K_N \tag{1.3}$$

where K_N is the closed ball with radius N.Then the corresponding SDE has a unique strong solution.

We will mostly consider the case where B is a d-dimensional Brownian motion and $\sigma(x) = I_d$ for all $x \in \mathbb{R}^d$. If then for example b is globally Lipschitz, i.e. there is a $c > 0$ with $|b(x) - b(y)| < c|x - y|$ for all $x, y \in \mathbb{R}^d$, then we have

$$\begin{aligned}
\|\sigma(x)\|^2 + \|b(x)\|^2 &\leq 1^2 + \big(\|b(x) - b(0)\| + \|b(0)\|\big)^2 \\
&\leq 1 + 2c^2|x - 0|^2 + 2b^2(0) \\
&\leq \max(1 + 2b^2(0), 2c^2)\big(1 + |x|^2\big)
\end{aligned}$$

and

$$\|\sigma(x) - \sigma(y)\|^2 + \|b(x) - b(y)\|^2 \leq 0 + |x - y|^2$$

for all $x, y \in \mathbb{R}^d$, i.e. equations 1.2 and 1.3 hold and the theorems guarantee the existence of a unique solution in this case.

Sometimes it is useful to rescale diffusion processes. This helps us for example, to transport diffusions on time intervals $[0;t]$ for different values of t to a common sample space. This can be done with the following lemma.

Lemma 1.3. *Let $c > 0$ and X be a solution of the SDE*

$$dX = b(X) \bullet dt + \sigma\, dB.$$

Define the rescaled process Y by $Y_t = X_{ct}/\sqrt{c}$ for every $t \geq 0$, a new Brownian motion $\tilde{B}$ by $\tilde{B}_t = B_{ct}/\sqrt{c}$ for all $t \geq 0$ and a new drift field by $\tilde{b}(x) = \sqrt{c} \cdot b(\sqrt{c} \cdot x)$ for all $x \in \mathbb{R}^d$. Then the process Y solves the SDE

$$dY = \tilde{b}(Y_t) \bullet dt + \sigma\, d\tilde{B}.$$

Proof. By the basic scaling property of Brownian motion the process $\tilde{B}$ as defined above is a Brownian motion. For any pair of stopping times S and T with $S < T$ the following holds.

$$\begin{aligned}
Y_T - Y_S &= \frac{1}{\sqrt{c}}\big(X_{cT} - X_{cS}\big) \\
&= \frac{1}{\sqrt{c}}\Big(\int_{cS}^{cT} b(X_t)\, dt + \sigma B_{cT} - \sigma B_{cS}\Big) \\
&\overset{c \cdot s = t}{=} \frac{1}{\sqrt{c}} \int_S^T b(X_{cs})\, c\, ds + \sigma \frac{1}{\sqrt{c}} B_{cT} - \sigma \frac{1}{\sqrt{c}} B_{cS} \\
&= \int_S^T \sqrt{c} b(\sqrt{c} Y_s)\, ds + \sigma(\tilde{B}_T - \tilde{B}_S).
\end{aligned}$$

This proves the claim. (qed)

The basic characterisation of reversible diffusions is the following theorem, which goes back to a result of Kolmogorov. A detailed proof is given for example in [Voß97].

Theorem 1.4. *Let $b\colon \mathbb{R}^d \to \mathbb{R}^d$ be Lipschitz continuous, B a Brownian motion with values in $\mathbb{R}^d$, and X a solution of the SDE*

$$dX = b(X) \bullet dt + dB$$

with $X_0 \in L^2$. Then the following conditions are equivalent:
(j) The process X is reversible with stationary distribution μ.
(ij) There is a function $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ with $b = -\nabla\Phi$ and $d\mu = \exp(-2\Phi(x))\, dx$.

$$\int_{\mathbb{R}^d} \exp\big(-2\Phi(x)\big)\, dx = 1.$$

Of course the condition $b = -\operatorname{grad}\Phi$ determines the potential Φ only up to a constant. So whenever $\exp(-2\Phi)$ is integrable, one can add a normalising constant to Φ, to change $\exp(-2\Phi)$ into a probability density.

The case of diffusion processes where the drift is a gradient is especially easy, because we can explicitly calculate the density of the distribution of the processes with respect to the Wiener measure. The big advantage of formula (1.4) below is, that it does not contain the stochastic integral from the Girsanov formula any more.

Lemma 1.5. *Let B be a $(\mathcal{F}_t)$-Brownian motion and let X be a solution of the stochastic differential equation*

$$\begin{aligned}
dX_t &= b(X_t) \bullet dt + dB_t \\
X_0 &= 0.
\end{aligned}$$

Furthermore let $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ *be two times continuously differentiable with* $b = -\operatorname{grad}\Phi$ *and let* b *be Lipschitz continuous. Then the distribution of* X *on* $\mathcal{F}_t$ *has a density* φ_t *with respect to the Wiener measure, defined by*

$$\varphi_t(\omega) = \exp\bigl(-\Phi(\omega_t) + \Phi(\omega_0) - \int_0^t v(\omega_s)\,ds\bigr) \qquad \text{for all } \omega \in C\bigl([0;\infty), \mathbb{R}^d\bigr) \tag{1.4}$$

where $v = \frac{1}{2}\bigl((\nabla\Phi)^2 - \Delta\Phi\bigr)$, *i. e. for every* $A \in \mathcal{F}_t$ *we have*

$$P(X \in A) = \int_A \varphi(\omega)\,d\mathbb{W}(\omega).$$

Proof. The density of $\mathcal{L}(X)$ on $\mathcal{F}_t$ is characterised by the Girsanov formula (see e.g. section 10.2 of [WW90]). The Lipschitz-continuity of b gives the necessary integrability conditions (see, e.g. corollaries 1.1 and 1.2 of [Voß97]). So we only have to check that φ_t as defined in (1.4) satisfies

$$\varphi_t(X) = \exp\bigl(\int_0^t b(X_s)\,dX_s - \frac{1}{2}\int_0^t b^2(X_s)\,ds\bigr).$$

Itô's formula gives

$$\begin{aligned} d\Phi(X) &= \operatorname{grad}\Phi(X) \bullet dX + \frac{1}{2}\Delta\Phi(X) \bullet dt \\ &= -b(X) \bullet dX + \frac{1}{2}\Delta\Phi(X) \bullet dt \end{aligned}$$

and because of

$$\begin{aligned} b(X) \bullet dX - \frac{1}{2}b^2(X) \bullet dt &= -d\Phi(X) - \frac{1}{2}\bigl((\nabla\Phi(X))^2 - \Delta\Phi(X)\bigr) \bullet dt \\ &= -d\Phi(X) - v(X) \bullet dt \end{aligned}$$

we get

$$\begin{aligned} \varphi_t(X) &= \exp\bigl(-\Phi(X_t) + \Phi(X_0) - \int_0^t v(X_s)\,ds\bigr) \\ &= \exp\bigl(\int_0^t b(X_s)\,dX_s - \frac{1}{2}\int_0^t b^2(X_s)\,ds\bigr). \end{aligned}$$

So everything is proved. (qed)

Example 1.1. Constant drift. Here we have some vector $b \in \mathbb{R}^d$ with $b(x) = b$ for all $x \in \mathbb{R}^d$. With the notation from lemma 1.5 we get

$$\begin{aligned} \Phi(x) &= -b \cdot x \\ b(x) &= -\operatorname{grad}\Phi(x) \;=\; b \\ v(x) &= \frac{1}{2}\bigl((\nabla\Phi)^2 - \Delta\Phi\bigr)(x) = \frac{b^2}{2} \end{aligned}$$

and conclude

$$\begin{aligned} \varphi_t(\omega) &= \exp\bigl(-\Phi(\omega_t) + \Phi(\omega_0) - \int_0^t v(\omega_s)\,ds\bigr) \\ &= \exp\bigl(b \cdot (\omega_t - \omega_0) - \frac{b^2}{2}t\bigr). \end{aligned}$$

This case is especially easy, because v is constant and thus the density is a function of only the endpoint of the path.

Example 1.2. The Ornstein-Uhlenbeck process. Here we have $b(x) = -\alpha x$ for some $\alpha > 0$. Using the notation of lemma 1.5 again, we get

$$\begin{aligned}
\Phi(x) &= \frac{\alpha}{2}x^2 - \frac{d}{4}\log\frac{\alpha}{\pi} \\
b(x) &= -\operatorname{grad}\Phi(x) \;=\; -\alpha x \\
v(x) &= \frac{1}{2}\big((\nabla\Phi)^2 - \Delta\Phi\big)(x) \;=\; \frac{1}{2}(\alpha^2x^2 - \alpha d).
\end{aligned}$$

The strange constant in the definition of Φ makes $\exp(-2\Phi)$ a probability density, namely the density of a normal distribution with covariance matrix $1/2\alpha \cdot I$. The process corresponding to this drift is reversible and has a stationary distribution μ. The lemma gives the density

$$\varphi_t(\omega) = \exp\Big(-\frac{\alpha}{2}(\omega_t^2 - \omega_0^2 - t\cdot d) - \frac{\alpha^2}{2}\int_0^t \omega_s^2\,ds\Big).$$

Chapter 2

Large Deviations

In this chapter I want to review some tools, which are available to study large deviations of diffusion processes. My main source here is the book of Dembo and Zeitouni [DZ98]. Other useful references include the books of Deuschel and Stroock [DS89] and den Hollander [Hol00].

2.1 Introduction

This section formulates the basic large deviation principle and gives the basic definition. For details see the references given above.

Definition 2.1. A **rate function** is a function $I\colon \mathcal{X} \to [0;\infty]$ on a Hausdorff topological space $\mathcal{X}$, which is lower semi-continuous, i.e. where all the level sets $\{\, x \in \mathcal{X} \mid I(x) \leq c \,\}$ for $c \geq 0$ are closed in $\mathcal{X}$. A rate function $I\colon \mathcal{X} \to [0;\infty]$ is called a **good rate function**, if all the level sets $\{\, x \in \mathcal{X} \mid I(x) \leq c \,\}$ for $c \geq 0$ are compact in $\mathcal{X}$.

In this text the space $\mathcal{X}$ will typically be either the Euclidean space $\mathbb{R}^n$ or a path space like $C\big([0;\infty), \mathbb{R}^d\big)$.

Definition 2.2. A family $(\mu_\varepsilon)_{\varepsilon>0}$ of probability measures on a Hausdorff topological space $\mathcal{X}$ is satisfies the **large deviation principle** (or shorter, the LDP) with rate␣function $I\colon \mathcal{X} \to [0;\infty]$, if the following two estimates hold:

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log \mu_\varepsilon(A) \leq -\inf_{x\in A} I(x) \tag{2.1}$$

for every closed set $A \subseteq \mathcal{X}$ and

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log \mu_\varepsilon(O) \geq -\inf_{x\in O} I(x) \tag{2.2}$$

for every open set $O \subseteq \mathcal{X}$.

Sometimes it is difficult to obtain a full LDP as described in the definition, but is is possible to get a weak LDP.

Definition 2.3. If the upper bound in definition 2.2 only holds for all compact (instead of all closed) sets, then the family $(\mu_\varepsilon)_{\varepsilon>0}$ satisfies the **weak large deviation principle** (or short: the weak LDP).

A weak LDP can be strengthened to a full LDP with the help of the following Lemma, which is a direct consequence of lemma 1.2.18 in [DZ98].

Definition 2.4. A family $(\mu_\varepsilon)_{\varepsilon>0}$ of probability measures on a Hausdorff topological space $\mathcal{X}$ is **exponentially tight** if for every $\alpha > 0$ there exists a compact set $K \subseteq \mathcal{X}$ with

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log \mu_\varepsilon(\mathcal{X} \setminus K) < -\alpha.$$

Lemma 2.1. *Let $(\mu_\varepsilon)_{\varepsilon>0}$ be an exponentially tight family of probability measures. If the upper bound (2.1) holds for all compact sets, then it also holds for all closed sets.*

2.2 General Principles

There are many tools, which are useful in the area of large deviations, and which can be formulated without considering a specific problem. Most of the results here help to use a exponential rate which is already known, by carrying it over to a different situation.

The following, frequently used Lemma shows that the exponential rate of a sum is just the maximum of the individual rates.

Lemma 2.2. *For any family of finitely many functions $f_1, \ldots, f_n \colon \mathbb{R}_+ \to \mathbb{R}_+$ we have*

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log\Bigl(\sum_{k=1}^{n} f_k(\varepsilon)\Bigr) \geq \max_{k=1,\ldots,n} \bigl(\liminf_{\varepsilon\downarrow 0} \varepsilon \log f_k(\varepsilon)\bigr)$$

and

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log\Bigl(\sum_{k=1}^{n} f_k(\varepsilon)\Bigr) = \max_{k=1,\ldots,n} \bigl(\limsup_{\varepsilon\downarrow 0} \varepsilon \log f_k(\varepsilon)\bigr).$$

Note the asymmetry between the $\geq$ and the $=$ sign. It is easy to construct examples, where a strict inequality holds for the first case. The next lemma illustrates two special case, where we have equality for both bounds.

Lemma 2.3. *Let $f, g \colon \mathbb{R}_+ \to \mathbb{R}_+$ be two functions and assume that either one of the two conditions $\limsup_{\varepsilon\downarrow 0} \varepsilon \log g(\varepsilon) \leq \liminf_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon)$ or $\limsup_{\varepsilon\downarrow 0} \varepsilon \log g(\varepsilon) < \liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr)$ holds. Then we have*

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr) = \liminf_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon)$$

and

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr) = \limsup_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon).$$

Proof (a). First assume $\limsup_{\varepsilon\downarrow 0} \varepsilon \log g(\varepsilon) \leq \liminf_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon)$. From lemma 2.2 we know $\liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr) \geq \liminf_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon)$. On the other hand for every $c > \liminf_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon)$ and every $E > 0$ we find an $\varepsilon < E$ with $f(\varepsilon) < \exp(c/\varepsilon)$ and, by choosing ε sufficiently small, also with $g(\varepsilon) < \exp(c/\varepsilon)$. So we can conclude

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr) \leq \liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(2\exp(c/\varepsilon)\bigr) = c$$

for all $c > \liminf_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon)$. This proves the first claim. The second claim is a direct consequence of lemma 2.2.

(b) Now assume $\limsup_{\varepsilon\downarrow 0} \varepsilon \log g(\varepsilon) < \liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr)$ and choose a $c \in \mathbb{R}$ with

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log g(\varepsilon) < c < \liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr).$$

Then there exists an $E > 0$ such that for every $\varepsilon < E$ we have both $f(\varepsilon) + g(\varepsilon) > 2\exp(c/\varepsilon)$ and $g(\varepsilon) < \exp(c/\varepsilon)$. Consequently we find $f(\varepsilon) > \exp(c/\varepsilon)$ for all $\varepsilon < E$ and thus the lower limit satisfies $\liminf_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon) \geq c$ for all $c < \liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr)$. This proves $\liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr) \leq \liminf_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon)$ and again with the first estimate from lemma 2.2 we get the equality for the lower bounds.

From the second part of lemma 2.2 we know that the rate for the sum is the maximum of the individual rate. Because of

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log g(\varepsilon) < \liminf_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr) \leq \limsup_{\varepsilon\downarrow 0} \varepsilon \log\bigl(f(\varepsilon) + g(\varepsilon)\bigr)$$

the maximum must then be attained for $\limsup_{\varepsilon\downarrow 0} \varepsilon \log f(\varepsilon)$. This completes the proof.

(qed)

For future reference we state the following basic estimate.

Lemma 2.4. *Let $c_1, \dots, c_n \in \mathbb{R}$. Then*

$$\frac{c_1^2}{\alpha_1} + \dots + \frac{c_n^2}{\alpha_n} \geq \frac{(c_1 + \dots + c_n)^2}{\sum_{k=1}^n \alpha_k}$$

for all $\alpha_1, \dots, \alpha_n > 0$ and equality holds if and only if there is a $\lambda \in \mathbb{R}$ with $\lambda\alpha_k = c_k$ for $k = 1, \dots, n$.

Proof. Let $a = \sum_{k=1}^n \alpha_k$, $p_k = \alpha_k/a$, and $d_k = c_k/p_k$ for $k = 1, \dots, n$. Then $\sum_{k=1}^n p_k = 1$ and Jensen's inequality gives

$$\sum_{k=1}^n \frac{c_k}{p_k} = \sum_{k=1}^n p_k d_k^2 \geq \Bigl(\sum_{k=1}^n p_k d_k\Bigr)^2 = \Bigl(\sum_{k=1}^n c_k\Bigr)^2 \tag{2.3}$$

where equality holds only for $d_1 = \dots = d_n$. Dividing (2.3) by a proves the claim. (qed)

A result somewhat similar to lemma 2.2 is the following. It gives an upper bound for the exponential rate of a product.

Lemma 2.5. *Let $f_1, \dots, f_n \colon \mathbb{R}_+ \to \mathbb{R}_+$ and $\limsup_{\varepsilon\downarrow 0} \varepsilon \log f_k(\varepsilon) = -c_k^2$ for $k = 1, \dots, n$. Let $\alpha_1, \dots, \alpha_k > 0$. Then the following estimate holds:*

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log \prod_{k=1}^n f_k(\alpha_k \varepsilon) \leq -\frac{\bigl(\sum_{k=1}^n c_k\bigr)^2}{\sum_{k=1}^n \alpha_k}.$$

Proof. Taking the product out of the logarithm we find

$$\begin{aligned}
\limsup_{\varepsilon\downarrow 0} \varepsilon \log \prod_{k=1}^n f_k(\alpha_k \varepsilon) &\leq \sum_{k=1}^n \limsup_{\varepsilon\downarrow 0} \varepsilon \log f_k(\alpha_k \varepsilon) \\
&= \sum_{k=1}^n \frac{1}{\alpha_k} \limsup_{\varepsilon\downarrow 0} \varepsilon \log f_k(\varepsilon) \\
&= -\sum_{k=1}^n \frac{1}{\alpha_k} c_k^2.
\end{aligned}$$

Applying the estimate from lemma 2.4 proves the claim. (qed)

The following lemma will turn out to be very useful for proving upper large deviation bounds. It helps to split an upper bound into a finite number of cases, which in turn helps to apply lemma 2.2. This trick is illustrated by the proof of proposition 2.7 below.

Lemma 2.6. *For each $\delta > 0$ there is a finite set $D_n^\delta \subseteq \{\alpha \in \mathbb{R}^n_{\geq 0} \mid \alpha_1 + \cdots + \alpha_n = 1 + \delta\}$ with*

$$\begin{aligned}&\{(\varepsilon_1, \ldots, \varepsilon_n) \in \mathbb{R}^n_{\geq 0} \mid \varepsilon_1 + \cdots + \varepsilon_n \leq \varepsilon\} \\ &\qquad \subseteq \bigcup_{\alpha \in D_n^\delta} \{(\varepsilon_1, \ldots, \varepsilon_n) \in \mathbb{R}^n_{\geq 0} \mid \varepsilon_j \leq \alpha_j \varepsilon \text{ for } j = 1, \ldots, n\}\end{aligned}$$

for all $\varepsilon > 0$.

Proof. Let $\delta > 0$. Because the simplex is compact, the covering

$$\begin{aligned}&\{(x_1, \ldots, x_n) \in \mathbb{R}^n_{\geq 0} \mid x_1 + \cdots + x_n \leq 1\} \\ &\qquad \subseteq \bigcup_{\substack{\alpha \in \mathbb{R}^n_{\geq 0} \\ \|\alpha\|_1 = 1 + \delta}} \{(x_1, \ldots, x_n) \in \mathbb{R}^n_{\geq 0} \mid x_j < \alpha_j \text{ for } j = 1, \ldots, n\}\end{aligned}$$

can be reduced to a finite one. So we can choose a finite set D_n^δ such that the inclusion still holds when the union is only taken over $\alpha \in D_n^\delta$. Multiplying both sides with $\varepsilon > 0$ gives the claim. (qed)

Lemma 2.5 is useful, if one can exploit some kind of independence structure. The following proposition features a basic application which shows the machinery at work. In chapter 5 we will use this idea in a more complicated situation.

Proposition 2.7. *Let $X_1, \ldots, X_n$ be independent, positive random variables with*

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log P(X_k \leq \varepsilon) = -c_k^2$$

and

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P(X_k \leq \varepsilon) = -b_k^2$$

where $b_k, c_k \geq 0$ for $k = 1, \ldots, n$. Then we have

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log P(X_1 + \cdots + X_n \leq \varepsilon) \leq -(c_1 + \cdots + c_n)^2.$$

and

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P(X_1 + \cdots + X_n \leq \varepsilon) \geq -(b_1 + \cdots + b_n)^2.$$

Proof. Let $\delta > 0$ and D_n^δ as in lemma 2.6. This gives

$$P(X_1 + \cdots + X_n \leq \varepsilon) \leq \sum_{\alpha \in D_n^\delta} P(X_1 \leq \alpha_1 \varepsilon, \ldots, X_k \leq \alpha_k \varepsilon).$$

For the individual terms we can use lemma 2.5. Let $\limsup_{\varepsilon \downarrow 0} \varepsilon \log P(X_k \leq \varepsilon) = -c_k^2$ with $c_k \geq 0$. Then we get

$$\begin{aligned}&\limsup_{\varepsilon \downarrow 0} \varepsilon \log P(X_1 \leq \alpha_1 \varepsilon, \ldots, X_k \leq \alpha_k \varepsilon) \\ &\qquad = \limsup_{\varepsilon \downarrow 0} \varepsilon \log \prod_{k=1}^n P(X_k \leq \alpha_k \varepsilon) \\ &\qquad \leq -\frac{(c_1 + \cdots + c_n)^2}{1 + \delta}\end{aligned}$$

for every $\alpha \in D_n^\delta$. Using lemma 2.2 we conclude

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \varepsilon \log P\bigl(X_1 + \cdots + X_n \leq \varepsilon\bigr) \\
&\qquad \leq \max_{\alpha\in D_n^\delta} \limsup_{\varepsilon\downarrow 0} \varepsilon \log P\bigl(X_1 \leq \alpha_1\varepsilon, \ldots, X_k \leq \alpha_k\varepsilon\bigr) \\
&\qquad \leq -\frac{(c_1 + \cdots + c_n)^2}{1+\delta}
\end{aligned}$$

for every $\delta > 0$ and thus

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P\bigl(X_1 + \cdots + X_n \leq \varepsilon\bigr) \leq -(c_1 + \cdots + c_n)^2.$$

For the lower bound let $\liminf_{\varepsilon\downarrow 0} \varepsilon \log P\bigl(X_k \leq \varepsilon\bigr) = -b_k^2$ with $b_k \geq 0$. From lemma 2.4 we know that we should choose α_k proportional to b_k in order to get best possible bound:

$$\begin{aligned}
&\liminf_{\varepsilon\downarrow 0} \varepsilon \log P\bigl(X_1 + \cdots + X_n \leq \varepsilon\bigr) \\
&\qquad \geq \liminf_{\varepsilon\downarrow 0} \varepsilon \log P\Bigl(X_k \leq \frac{b_k}{b_1 + \cdots + b_n}\varepsilon, k = 1, \ldots, n\Bigr) \\
&\qquad = \liminf_{\varepsilon\downarrow 0} \varepsilon \log \prod_{k=1}^n P\Bigl(X_k \leq \frac{b_k}{b_1 + \cdots + b_n}\varepsilon\Bigr) \\
&\qquad \geq \sum_{k=1}^n \frac{b_1 + \cdots + b_n}{b_k} \liminf_{\varepsilon\downarrow 0} \varepsilon \log P\bigl(X_k \leq \varepsilon\bigr) \\
&\qquad = -\sum_{k=1}^n \frac{b_1 + \cdots + b_n}{b_k} b_k^2 \\
&\qquad = -(b_1 + \cdots + b_n)^2.
\end{aligned}$$

This completes the proof. (qed)

The remaining part of this section summarises some important results from the literature. The contraction principle allows, to transform an LDP on one space to another space by means of a continuous mapping. The following theorem (theorem 4.2.1 in [DZ98]) states the result.

Theorem 2.8 *(contraction principle). Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces and $f\colon \mathcal{X} \to \mathcal{Y}$ be a continuous function. Consider a good rate function $I\colon \mathcal{X} \to [0;\infty]$.*

(a) For each $y \in \mathcal{Y}$ define

$$I'(y) = \inf\bigl\{I(x) \bigm| x \in \mathcal{X}, f(x) = y\bigr\}.$$

Then I' is a good rate function on $\mathcal{Y}$.

(b) If I controls the LDP associated with a family of probability measures (μ_ε) on $\mathcal{X}$, then I' controls the LDP associated with the family of probability measures $(\mu_\varepsilon \circ f^{-1})$ on $\mathcal{Y}$.

If f is not injective then the mapping “looses information”. Thus the contraction principle can transport an LDP from a larger space to a smaller one. The Dawson-Gärtner theorem does the opposite. It helps to transport an LDP from smaller spaces to a large space.

In order to formulate the theorem we need the concept of an projective limit. Let $(J, \leq)$ be a partially ordered set, such that whenever $i, j \in J$ there exists a $k \in J$ with $i \leq k$ and $j \leq k$.

A **projective system** is a family $(\mathcal{Y}_j)_{j\in J}$ of Hausdorff topological spaces, together with a family $(p_{ij})_{i,j\in J}$ of continuous maps $p_{ij}\colon \mathcal{Y}_j \to \mathcal{Y}_i$, such that $p_{ik} = p_{ij} \circ p_{jk}$ whenever $i, j, k \in \mathcal{Y}$ with $i \leq j \leq k$.

The **projective limit** of the system $(\mathcal{Y}_j, p_{ij})_{i,j\in J}$ is the subset $\mathcal{X}$ of all elements from the product space $\mathcal{Y} = \prod_{j\in J} \mathcal{Y}_j$, which are consistent with the maps $p_{i,j}$, i.e. a $y = (y_j)_{j\in J} \in \mathcal{Y}$ is in $\mathcal{X}$, if and only if $y_i = p_{ij}(y_j)$ whenever $i < j$. The canonical projections $p_j\colon \mathcal{X} \to \mathcal{Y}_j$ are defined by $p_j(y) = y_j$. The space $\mathcal{X}$ is equipped with the topology induced by $\mathcal{Y}$. The canonical projections are continuous because they are just the restriction of the continuous coordinate maps $\mathcal{Y} \to \mathcal{Y}_j$ to the space $\mathcal{X}$.

The typical example here is J being the set of all finite subsets of an interval $[0;t] \subseteq \mathbb{R}$, with the set inclusion as the partial ordering. The projective system is then the family of all finite dimensional spaces $\mathbb{R}^j$ for finite sets $j \subseteq [0;t]$, i.e. for $j \in J$. In this case the projective limit $\mathcal{X}$ is the space of all functions $[0;t] \to \mathbb{R}$, equipped with the topology of pointwise convergence.

The following theorem (theorem 4.6.1 in [DZ98]) explains how to transport an LDP from the spaces $\mathcal{Y}_j$ to the projective limit $\mathcal{X}$.

Theorem 2.9. *(Dawson-Gärtner) Let $(\mu_\varepsilon)_{\varepsilon>0}$ be a family of probability measures on $\mathcal{X}$. Assume that for each $j \in J$ the probability measures $(\mu_\varepsilon \circ p_j^{-1})_{\varepsilon>0}$ on $\mathcal{Y}_j$ satisfy the LDP with good rate function I_j. Then the family (μ_ε) satisfies the LDP on $\mathcal{X}$ with a good rate function I, defined by*

$$I(x) = \sup\bigl\{I_j(p_j(x)) \bigm| j \in J\bigr\} \quad \text{for all } x \in \mathcal{X}.$$

Another important tool is the Varadhan Lemma (Theorem 4.3.1 in [DZ98]). Let Z_ε be random variables, taking values in the regular topological space $\mathcal{X}$, and let μ_ε be the law of Z_ε.

Theorem 2.10. *(Varadhan␣Lemma) Suppose that (μ_ε) satisfies the LDP with a good rate function $I\colon \mathcal{X} \to [0;\infty]$, and let $\varphi\colon \mathcal{X} \to \mathbb{R}$ be any continuous function. Assume further either the tail condition*

$$\lim_{M\to\infty} \limsup_{\varepsilon\downarrow 0} \varepsilon \log E\Bigl(\exp(\varphi(Z_\varepsilon)/\varepsilon)(\varphi(Z_\varepsilon) \geq M)\Bigr) = -\infty,$$

or the following moment condition for some $\gamma > 1$,

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log E\Bigl(\exp(\gamma\varphi(Z_\varepsilon)/\varepsilon)\Bigr) < \infty.$$

Then

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log E\Bigl(\exp(\varphi(Z_\varepsilon)/\varepsilon)\Bigr) = \sup_{x\in\mathcal{X}} \bigl(\varphi(x) - I(x)\bigr).$$

2.3 The LDP for Stationary Distributions

In this section we will prove a simple large deviation principle which describes the asymptotic behaviour of the stationary distribution of a diffusion process when the drift becomes stronger.

The proof will be a simple application of the Laplace-Principle, which is formulated in the following lemma. This is similar to the Varadhan lemma (theorem 2.10). We give an independent proof here, because the proof is simple and nevertheless gives some insight.

Lemma 2.11. *(Laplace␣principle) Let $A \subseteq \mathbb{R}^d$ be measurable and $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be a measurable function with $\int_A \mathrm{e}^{-\varphi(x)}\, dx < \infty$. Then we have*

$$\lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int_A \mathrm{e}^{-\vartheta\varphi(x)}\, dx = -\operatorname*{ess\,inf}_{x\in A} \varphi(x).$$

Proof. Denote the Lebesgue measure by λ^d. First choose $c > \operatorname{ess\,inf}_{x\in A}\varphi(x)$. Then we have $\lambda^d\{\, x \in A \mid \varphi(x) < c \,\} > 0$ and thus

$$\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int_A e^{-\vartheta\varphi(x)}\, dx \;\geq\; \liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log\bigl(e^{-\vartheta c}\lambda^d\{\, x \in A \mid \varphi(x) < c \,\}\bigr) \;=\; -c.$$

This holds for all $c > \operatorname{ess\,inf}_{x\in A}\varphi(x)$, which gives

$$\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int_A e^{-\vartheta\varphi(x)}\, dx \;\geq\; -\operatorname*{ess\,inf}_{x\in A}\varphi(x).$$

For $\operatorname{ess\,inf}_{x\in A}\varphi(x) = -\infty$ everything is clear. Assume $\operatorname{ess\,inf}_{x\in A}\varphi(x) > -\infty$. Then by adding a constant to φ we can assume $\operatorname{ess\,inf}_{x\in A}\varphi(x) = 0$ without loss of generality. Because $e^{-\vartheta\varphi(x)}$ for $\vartheta \to \infty$ converges a.s. monotonically from above to the indicator function of the set $\{\, x \in A \mid \varphi(x) = 0 \,\}$ and is bounded by the integrable function $e^{-\varphi(x)}$ we have

$$\lim_{\vartheta\to\infty} \int_A e^{-\vartheta\varphi(x)}\, dx \;=\; \lambda^d\{\, x \in A \mid \varphi(x) = 0 \,\} \;\leq\; \lambda^d\{\, x \in A \mid \varphi(x) \leq 0 \,\} \;<\; \infty$$

and thus we get the upper bound

$$\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int_A e^{-\vartheta\varphi(x)}\, dx \leq 0 = -\operatorname*{ess\,inf}_{x\in A}\varphi(x).$$

(qed)

As an almost trivial conclusion of the Laplace principle we can derive a large deviation principle for standard normal distributed random variables in $\mathbb{R}$. This is illustrated in the following corollary.

Corollary 2.12. *If X is a standard normal random variable, then*

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log P(\sqrt{\varepsilon}X \in A) = -\operatorname*{ess\,inf}_{x\in A} \frac{x^2}{2}$$

for every measurable set $A \subseteq \mathbb{R}$.

Proof. The distribution of $\sqrt{\varepsilon}X$ has density

$$\psi_{0,\varepsilon}(x) = \frac{1}{\sqrt{2\pi\varepsilon}} \exp\bigl(-\frac{x^2}{2\varepsilon}\bigr).$$

Thus with $\vartheta = 1/\varepsilon$ the Laplace principle gives

$$\begin{aligned}\lim_{\varepsilon\downarrow 0} \varepsilon \log P(\sqrt{\varepsilon}X \in A) &= \lim_{\varepsilon\downarrow 0} \varepsilon \log \int_A \exp\bigl(-\frac{x^2}{2\varepsilon}\bigr)\, dx - \lim_{\varepsilon\downarrow 0} \varepsilon \log \sqrt{2\pi\varepsilon} \\ &= -\operatorname*{ess\,inf}_{x\in A} \frac{x^2}{2} + 0.\end{aligned}$$

(qed)

Example 2.1. Let B be a one-dimensional Brownian motion. Then B_t is Gaussian distributed with expectation 0 and variance t. Using the corollary we can examine the large deviation behaviour of the event $\{|B_t| > c\}$ for $c \to \infty$. With $\varepsilon = 1/c^2$ we have

$$|B_t| > c \iff \sqrt{\varepsilon}\cdot B_t/\sqrt{t} \in \bigl\{x \in \mathbb{R} \bigm| |x| > 1/\sqrt{t}\bigr\}$$

and thus

$$\lim_{c\to\infty} \frac{1}{c^2} \log P\Bigl(|B_1| > c\Bigr) = -\frac{1}{2t}.$$

For comparison with example 2.2 below we also note that the probability of $|B_t| < \varepsilon$ for $\varepsilon \downarrow 0$ decays slower the exponentially. We have

$$\begin{aligned}
&\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log P\Big(|B_1| < \varepsilon\Big) \\
&\qquad = \lim_{\varepsilon\downarrow 0} \varepsilon^2 \log \int_{-\varepsilon}^{\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\
&\qquad = \lim_{\varepsilon\downarrow 0} \varepsilon^2 \log\big(2\varepsilon \frac{1}{\sqrt{2\pi}}\big) = 0.
\end{aligned}$$

The main result of this section is as follows.

Theorem 2.13. *Let $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ be differentiable and such that $\exp(-2\Phi(x))$ is a probability density on $\mathbb{R}^d$. Let Φ be bounded from below with $\Phi_* = \inf\{\, \Phi(x) \mid x \in \mathbb{R}^d \,\} > -\infty$. Finally let $b = -\operatorname{grad}\Phi$ be Lipschitz continuous.*

Then for every $\vartheta \geq 1$ the stochastic differential equation

$$dX^\vartheta \;=\; \vartheta\, b(X^\vartheta) \bullet dt + dW$$

has a stationary distribution μ_ϑ and for every measurable set $A \subseteq \mathbb{R}^d$ we have

$$\lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log \mu_\vartheta(A) \;=\; -\operatorname*{ess\,inf}_{x\in A} 2\big(\Phi(x) - \Phi_*\big).$$

Proof. Again, let λ^d denote the Lebesgue measure on $\mathbb{R}^d$ and define

$$Z_\vartheta = \int_{\mathbb{R}^d} \exp\big(-2\vartheta\Phi(x)\big)\, dx.$$

Then we have

$$\begin{aligned}
Z_\vartheta &= \int_{\{\Phi>0\}} \exp\big(-2\vartheta\Phi(x)\big)\, dx + \int_{\{\Phi\leq 0\}} \exp\big(-2\vartheta\Phi(x)\big)\, dx \\
&\leq \int_{\{\Phi>0\}} \exp\big(-2\Phi(x)\big)\, dx + \lambda^d\{\Phi \leq 0\} \cdot \exp(-2\vartheta\Phi_*) \\
&< \infty.
\end{aligned}$$

Using Z_ϑ we can define $\Phi_\vartheta = \vartheta \cdot \Phi + \ln\sqrt{Z_\vartheta}$. This new potential has

$$\int \exp(-2\Phi_\vartheta)\, dx = \int \exp(-2\vartheta\Phi) \exp(-\ln Z_\vartheta)\, dx = 1$$

and

$$-\operatorname{grad}\Phi_\vartheta = -\vartheta \operatorname{grad}\Phi = \vartheta \cdot b.$$

By Kolmogorov's theorem (theorem 1.4) the process X^ϑ is reversible and has a stationary distribution μ_ϑ with density $\exp(-2\vartheta\Phi)/Z_\vartheta$.

Using this density we get

$$\begin{aligned}
\lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log \mu_\vartheta(A) &= \lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int_A \exp(-2\vartheta\Phi(x)) \frac{1}{Z_\vartheta}\, dx \\
&= \lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int_A \exp\big(-2\vartheta\Phi(x)\big)\, dx - \lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log Z_\vartheta \\
&= -\operatorname*{ess\,inf}_{x\in A} 2\Phi(x) + \operatorname*{ess\,inf}_{x\in\mathbb{R}^d} 2\Phi(x),
\end{aligned}$$

where the last line is a consequence of lemma 2.11. Filling in the definition of Φ_* finishes the proof. (qed)

Both, the lemma and the theorem have results of the form

$$\lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log \mu_\vartheta(A) = -\operatorname*{ess\,inf}_{x\in A} I(x) \tag{2.4}$$

for all measurable sets $A \subseteq \mathbb{R}^d$ and for some function I. Assume that a family of measures has the property (2.4). Then we can use $\inf_{x\in A} I(x) \leq \operatorname{ess\,inf}_{x\in A} I(x)$ to get the usual upper bound from the large deviation principle: for closed sets $A \subseteq \mathbb{R}^d$ we have

$$\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \mu_\vartheta(A) = -\operatorname*{ess\,inf}_{x\in A} I(x) \leq -\inf_{x\in A} I(x).$$

On the other hand we don't get the lower bound in the general case. Only in case that we have $\inf_{x\in O} I(x) = \operatorname{ess\,inf}_{x\in O} I(x)$ for every open set $O \subseteq \mathbb{R}^d$, e.g. if I is continuous, we can conclude the lower bound

$$\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log \mu_\vartheta(O) \geq -\inf_{x\in O} I(x)$$

for every open set O (actually in this case we have even equality here). So at least for continuous rate functions I the condition (2.4) implies the LDP with rate function I.

The is not true in general. If we have an LDP in the form of definition 2.2, then the limit

$$\lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log \mu_\vartheta(A)$$

does not necessarily exist. A set A where we have

$$\inf_{x\in A^\circ} I(x) = \inf_{x\in \bar{A}} I(x),$$

i.e. where the limit does exist, is called a **continuity set** of the rate function.

2.4 The LDP for Empirical Distributions

The empirical distribution of a reversible diffusion converges to the stationary distribution when $t \to \infty$. In this section we will give a large deviation result for this case.

The **empirical distribution** $L_t^\omega \in \operatorname{Prob}(\mathbb{R}^d)$ of a process X with values in $\mathbb{R}^d$ is defined by

$$L_t^\omega(A) = \frac{1}{t}\lambda^d\{s \in [0,t] \mid \omega_s \in A\} \quad \text{for all } \omega \in C\big([0;\infty),\mathbb{R}^d\big),\ A \in \mathcal{B}(\mathbb{R}^d),$$

where λ^d denotes the d-dimensional Lebesgue measure. Since L_t is a mapping from the path space $C\big([0;\infty),\mathbb{R}^d\big)$ into $\operatorname{Prob}(\mathbb{R}^d)$ we can understand L_t as a random probability measure with corresponding probability space $C\big([0;\infty),\mathbb{R}^d\big)$.

Let X be a reversible diffusion process with stationary distribution μ. In this situation theorem 6.2.21 of Deuschel and Stroock [DS89] applies. The notation there is a little bit different from ours: their V is our $2v$ and U there corresponds to 2Φ here. The theorem expresses the rate with the help of the Dirichlet form $\mathcal{E}$ (cf. below), which is associated with the process X: define $J_\mathcal{E}\colon \operatorname{Prob}(\mathbb{R}^d) \to [0;\infty]$ by

$$J_\mathcal{E}(\nu) := \begin{cases} \mathcal{E}(f,f), & \text{if } \nu \ll \mu \text{ with } d\nu = f^2 d\mu \text{ for a } f \in \mathcal{D}(\mathcal{E}), \text{ and} \\ +\infty & \text{else.} \end{cases}$$

Theorem 2.14. *Let P_ν be the distribution of a reversible solution of the stochastic differential equation*

$$dX_t = b(X_t) \bullet dt + dB_t \tag{2.5}$$

with initial distribution $\mathcal{L}(X_0) = \nu$ and stationary distribution μ. Define $v = (b^2 + \operatorname{div} b)/2$ and assume that $\{ x \in \mathbb{R}^d \mid v(x) \leq c \}$ is compact for each $c \geq 0$. Consider $\operatorname{Prob}(\mathbb{R}^d)$ *equipped* *with the weak topology and the Borel-σ-algebra. Then $J_{\mathcal{E}}$ is a good rate function and the following LDP holds:*

$$
\begin{aligned}
-\inf_{\Gamma^\circ} J_{\mathcal{E}} &\leq \inf_{\nu\in\mathrm{Prob}(\mathbb{R}^d)} \liminf_{t\to\infty} \frac{1}{t}\log P_\nu\bigl(\{\,\omega \mid L_t^\omega \in \Gamma\,\}\bigr) \\
&\leq \sup_{\nu\in\mathrm{Prob}(\mathbb{R}^d)} \limsup_{t\to\infty} \frac{1}{t}\log P_\nu\bigl(\{\,\omega \mid L_t^\omega \in \Gamma\,\}\bigr) \leq -\inf_{\overline{\Gamma}} J_{\mathcal{E}}
\end{aligned}
$$

for all measurable sets $\Gamma \subseteq \mathrm{Prob}(\mathbb{R}^d)$.

The Dirichlet-form $\mathcal{E}$ for the process X can be constructed as follows. Via

$$P_t f(x) = E_x\bigl(f(X_t)\bigr)$$

for all $x \in \mathbb{R}^d$ and all $t \geq 0$ we get a strongly continuous operator semigroup $(P_t)_{t\geq 0}$ on the space $\mathcal{L}^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$. The generator L of this semigroup satisfies

$$Lf = b\cdot\nabla f + \frac{1}{2}\Delta f$$

for all $f \in C_c^2(\mathbb{R}^d)$. Because X is reversible, the generator L is self-adjoint. Now define a quadratic form $\mathcal{E}_0$ by

$$\mathcal{E}_0(f,g) := \frac{1}{2}\int_{\mathbb{R}^d} \nabla f\cdot\nabla g\,d\mu$$

for all $f, g \in C_c^\infty(\mathbb{R}^d, \mathbb{R})$.

Lemma 2.15. *Let X be a reversible solution of the SDE* (2.5) *with Lipschitz-continuous drift $b\colon \mathbb{R}^d \to \mathbb{R}^d$ and stationary distribution μ. Then the quadratic form $\mathcal{E}_0$ satisfies*

$$\mathcal{E}_0(f,g) = (-Lf, g)_\mu$$

for all $f, g \in C_c^\infty(\mathbb{R}^d)$, where $(\,\cdot\,,\,\cdot\,)_\mu$ is the scalar product on $\mathcal{L}^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$.

According to theorem X.23 of Reed and Simon [RS72] resp. theorem 6.2.9 of Deuschel and Stroock, the quadratic form $\mathcal{E}_0$ has a closure $\bigl(\mathcal{E}, \mathcal{D}(\mathcal{E})\bigr)$. We have $f \in \mathcal{D}(\mathcal{E})$ if and only if there are $f_n \in C_c^\infty(\mathbb{R}^d)$ and $g_1, \ldots, g_d \in \mathcal{L}^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ with $f_n \to f$ and $\partial_j f_n \to g_j$ in $\mathcal{L}^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ for $n \to \infty$ and $j = 1, \ldots, d$. In this case we get

$$\mathcal{E}(f,f) = \lim_{n\to\infty} \mathcal{E}_0(f_n, f_n) = \frac{1}{2}\int_{\mathbb{R}^d} \sum_{j=1}^d g_j^2\,d\mu.$$

2.5 Sample Path LDP

While the previous sections treated the large deviation behaviour of some derived properties of a process, it is also possible to consider the large deviation behaviour of the process itself, i.e. for the distribution of the processes paths on the space of all sample paths.

The basic large deviation result here is Schilder's theorem about large deviations for scaled down Brownian motion (see theorem 5.2.1 in [DZ98]). By $C_0([0;t], \mathbb{R}^d)$ we denote the space of all continuous functions $\omega\colon [0;t] \to \mathbb{R}^d$ starting in 0, equipped with the supremum norm.

Theorem 2.16 *(Schilder). Let B be a standard Brownian motion. For $\varepsilon > 0$ let $\mathbb{W}_\varepsilon$ be the law of the scaled down process $\sqrt{\varepsilon}B$. Then the measures $\mathbb{W}_\varepsilon$ satisfy on $\bigl(C_0([0;t], \mathbb{R}^d), \|\cdot\|_\infty\bigr)$ an LDP with good rate function*

$$I(\omega) = \begin{cases} \frac{1}{2}\int_0^t |\dot{\omega}_s|^2\,ds, & \textit{if } \omega \textit{ is absolutely continuous, and} \\ +\infty & \textit{else.} \end{cases}$$

Example 2.2. Consider a one-dimensional Brownian motion on the time interval $[0;t]$. We can use Schilder's theorem to calculate the exponential decay rates of the probability $P\big(\|B\|_\infty > c\big)$ for $c \to \infty$.

With $\varepsilon = 1/c^2$ we have

$$\sup_{0\leq s\leq t} |B_s| > c \iff \sqrt{\varepsilon} B \in \big\{\omega \bigm| |\omega_s| > 1 \text{ for some } s \in [0;t]\big\} =: A$$

and because A is a continuity set of the rate function I we find

$$\begin{aligned}
&\lim_{c\to\infty} \frac{1}{c^2} \log P\big(\|B\|_\infty > c\big) \\
&\qquad = \lim_{\varepsilon\downarrow 0} \varepsilon \log P\big(\sqrt{\varepsilon} B \in A\big) \\
&\qquad = -\inf\Big\{\frac{1}{2}\int_0^t \dot{\omega}_s^2\, ds \Bigm| \omega \in A\Big\} = -\frac{1}{2}\int_0^t 1/t^2\, dt = -\frac{1}{2t},
\end{aligned}$$

where we used the fact that the infimum is attained for $\omega_s = s/t$. Comparing this with example 2.1 we notice that the exponential rate for $\sup_{0\leq s\leq t} |B_s| > c$ is the same as the exponential rate for the event $|B_t| > c$.

On the other hand it is not possible to treat $P\big(\|B\|_\infty < \varepsilon\big)$ for $\varepsilon \downarrow 0$ the same way. Since we have

$$\sup_{0\leq s\leq t} |B_s| < \varepsilon \iff \frac{1}{\varepsilon} B \in \big\{\omega \bigm| \sup_{0\leq s\leq t} |\omega_s| < 1\big\}$$

here, we would need the large deviation behaviour for the blown-up Brownian motion instead of for the the scaled-down Brownian motion. We can derive the large deviation behaviour of this special event nevertheless, because an explicit formula for the probability is known. In section X.5 (p. 342) of [Fel71] it is shown that

$$\begin{aligned}
&P\big(|B_s| \leq \varepsilon \text{ for all } s \in [0;t]\big) \\
&\qquad = \frac{4}{\pi}\sum_{n=0}^\infty \frac{1}{2n+1} \exp\Big(-\frac{(2n+1)^2\pi^2}{8\varepsilon^2} t\Big) \sin\Big(\frac{(2n+1)\pi}{2}\Big).
\end{aligned}$$

The dominating term in this sum is

$$\frac{4}{\pi}\frac{1}{2\cdot 0+1} \exp\Big(-\frac{(2\cdot 0+1)^2\pi^2}{8\varepsilon^2} t\Big) \sin\Big(\frac{(2\cdot 0+1)\pi}{2}\Big) = \frac{4}{\pi}\exp\Big(-\frac{\pi^2}{8\varepsilon^2} t\Big)$$

which corresponds to $n = 0$. For the tail of the sum we find the estimate

$$\begin{aligned}
&\sum_{n=1}^\infty \frac{1}{2n+1} \exp\Big(-\frac{(2n+1)^2\pi^2}{8\varepsilon^2} t\Big) \sin\Big(\frac{(2n+1)\pi}{2}\Big) \\
&\qquad \leq \sum_{n=1}^\infty \exp\Big(-\frac{(2n+1)^2\pi^2}{8\varepsilon^2} t\Big) \\
&\qquad \leq \sum_{n=1}^\infty \exp\Big(-\frac{\pi^2 t}{2\varepsilon^2} n\Big) = \frac{\exp\big(-\frac{\pi^2 t}{2\varepsilon^2}\big)}{1-\exp\big(-\frac{\pi^2 t}{2\varepsilon^2}\big)}
\end{aligned}$$

and thus

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \varepsilon^2 \log \frac{4}{\pi}\sum_{n=1}^\infty \frac{1}{2n+1} \exp\Big(-\frac{(2n+1)^2\pi^2}{8\varepsilon^2} t\Big) \sin\Big(\frac{(2n+1)\pi}{2}\Big) \\
&\qquad \leq -\frac{\pi^2 t}{2} < -\frac{\pi^2 t}{8}.
\end{aligned}$$

Using lemma 2.3 we get

$$\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log P\big(\|B\|_\infty < \varepsilon\big) = \lim_{\varepsilon\downarrow 0} \varepsilon^2 \log \frac{4}{\pi}\exp\Big(-\frac{\pi^2}{8\varepsilon^2} t\Big) = -\frac{\pi^2 t}{8}.$$

The result is different than in example 2.1. The probability of $\sup_{0\leq s\leq t} |B_s| < \varepsilon$ decays exponentially for $\varepsilon \downarrow 0$ while the probability for $|B_t| < \varepsilon$ does not.

A generalisation of theorem 2.16 is the so-called Freidlin-Wentzell␣Theory (see chapter 5.6 of [DZ98] for a detailed explanation). There, one considers a stochastic differential equation of the form

$$\begin{aligned} dX_t^\varepsilon &= b(X_t^\varepsilon) \bullet dt + \sqrt{\varepsilon}\, dW_t \quad \text{for } t \in [0;1], \text{ and} \\ X_0^\varepsilon &= 0, \end{aligned}$$

where $b\colon \mathbb{R} \to \mathbb{R}$ is uniformly Lipschitz. Via an application of the contraction principle one can conclude from theorem 2.16 the following result.

Theorem 2.17 *(Freidlin-Wentzell). The family (X_t^ε) satisfies the LDP in $C_0[0;t]$ with the good rate function*

$$I(\omega) = \begin{cases} \frac{1}{2}\int_0^t |\dot{\omega}_s - b(\omega_s)|^2\, ds, & \textit{if } \omega \in H_1, \textit{ and} \\ +\infty & \textit{else.} \end{cases}$$

For the proof, the contraction principle is applied to the map $F\colon C_0[0;t] \to C_0[0;t]$ where $f = F(g)$ is defined to be the unique solution of the ordinary differential equation

$$f(t) = \int_0^t b(f(s))\, ds + g(t) \quad \text{for all } t \in [0;1].$$

Further generalisations are possible, e.g. to the case of the SDE

$$\begin{aligned} dX_t^\varepsilon &= b(X_t^\varepsilon) \bullet dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon) \bullet dW_t \quad \text{for } t \in [0;1], \text{ and} \\ X_0^\varepsilon &= x, \end{aligned}$$

where $x \in \mathbb{R}^d$, $b\colon \mathbb{R}^d \to \mathbb{R}^d$ is uniformly Lipschitz, and where all components of the matrix σ are bounded, uniformly Lipschitz continuous functions.

Chapter 3

The Ornstein-Uhlenbeck Process

In this chapter we will use the the Ornstein-Uhlenbeck process to explain some of the questions which we will answer more generally in later chapters. It will turn out that the simple structure of the Ornstein-Uhlenbeck process will ease many calculations, but we still can see what kind of results we can expect for the general case.

3.1 Introduction

Let B be a d-dimensional Brownian motion and $\alpha > 0$ be a real valued parameter. Then the solution of the stochastic differential equation

$$\begin{aligned} dX_t &= -\alpha X_t \bullet dt + dB_t, \\ X_0 &= x_0 \in \mathbb{R}^d \end{aligned} \tag{3.1}$$

is called the **Ornstein-Uhlenbeck process** with parameter α and start in x_0 (see, for example, v. Weizsäcker and Winkler [WW90], p. 10). Mostly we will consider the case $x_0 = 0$.

Using the results from chapter 1 it is clear that this equation has a unique solution. But using the variation of constants method we can even explicitly solve this equation. Let

$$X_t = e^{-\alpha t} X_0 + B_t - \alpha \int_0^t e^{-\alpha(t-s)} B_s \, ds$$

or equivalently

$$X_t = e^{-\alpha t} X_0 + \int_0^t e^{-\alpha(t-s)} \, dB_s \tag{3.2}$$

for all $t \geq 0$. Then it is easy to check that the process X defined by this, is a solution of the SDE (3.1). Figure 3.1 shows five paths of an Ornstein-Uhlenbeck process with parameter $\alpha = 1$ and start in 0.

If we define the potential $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ as

$$\Phi(x) = \frac{\alpha}{2}|x|^2 - \frac{d}{4} \log \frac{\alpha}{\pi} \quad \text{for all } x \in \mathbb{R}^d,$$

then the drift $b(x) = -\alpha x$ can be expressed as $b = -\operatorname{grad} \Phi$ and we find

$$\exp\bigl(-2\Phi(x)\bigr) = \Bigl(\frac{\alpha}{\pi}\Bigr)^{d/2} \exp\Bigl(-\alpha|x|^2\Bigr) = \frac{1}{\bigl(2\pi\frac{1}{2\alpha}\bigr)^{d/2}} \exp\Bigl(-\frac{|x|^2}{2\frac{1}{2\alpha}}\Bigr).$$

With theorem 1.4 we can conclude that the process X is reversible and has a d-dimensional normal distribution with covariance matrix $\frac{1}{2\alpha} I_d$ as its stationary distribution. From chapter 1 we already know the density of the distribution on the path space:

$$\varphi_t(\omega) = \exp\Bigl(-\frac{\alpha}{2}(\omega_t^2 - \omega_0^2 - t \cdot d) - \frac{\alpha^2}{2} \int_0^t \omega_s^2 \, ds\Bigr). \tag{3.3}$$

Figure 3.1: *This figure shows five paths of an Ornstein-Uhlenbeck process on the interval* $[0; 10]$ *with parameter* $\alpha = 1$.

3.2 Strong Drift

In chapter 5 we will answer the following question: what is the large deviations behaviour of the endpoint X_t of a diffusion process, if we increase the drift? For an Ornstein-Uhlenbeck process this means that we replace the constant α with $\vartheta\alpha$ for $\vartheta > 0$ and take $\vartheta \to \infty$ then.

Let X^ϑ be the solution of (3.1) with parameter $\vartheta\alpha$ and start in 0. The process has the density

$$\begin{aligned}\varphi_t^\vartheta(\omega) &= \exp\Big(-\frac{\vartheta\alpha}{2}(\omega_t^2 - \omega_0^2 - t\cdot d) - \frac{\vartheta^2\alpha^2}{2}\int_0^t \omega_s^2\, ds\Big) \\ &= \exp\big(\vartheta F(\omega) - \vartheta^2 G(\omega)\big)\end{aligned}$$

with

$$\begin{aligned}F(\omega) &= \frac{\alpha}{2}(\omega_0^2 - \omega_t^2 + d) \quad \text{and} \\ G(\omega) &= \frac{\alpha^2}{2}\int_0^t \omega_s^2\, ds.\end{aligned}$$

The case of an Ornstein-Uhlenbeck process is especially simple here, because we know the explicit distribution of X_t^ϑ. It is a d-dimensional Gaussian distribution with expectation 0 and covariance matrix

$$\Sigma = \frac{1}{2\alpha\vartheta}\big(1 - \exp(-2\alpha\vartheta\, t)\big)\cdot I_d$$

(see for example section 8.3 of [Arn74]). Using the Laplace Principle (lemma 2.11) one can easily see that

$$\lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log P(X_t^\vartheta \in A) = -\operatorname*{ess\,inf}_{x\in A}\alpha x^2 \tag{3.4}$$

for all $t > 0$ holds. Note that this rate does not depend on t.

Here we want to give another proof for the one-dimensional case, which does not use the explicit distribution of X_t^ϑ but can be generalised to more general drift fields b. To ease the notation we only consider the case $t = 1$. Formula (1–1.9.7) from [BS96] states

$$\begin{aligned}&E_x\Big(\exp\big(-\frac{\gamma^2}{2}\int_0^t B_s^2\, ds\big); B_t \in dz\Big) \\ &\quad= \frac{\sqrt{\gamma}}{\sqrt{2\pi\sinh(t\gamma)}}\exp\Big(-\frac{(x^2+z^2)\gamma\cosh(t\gamma) - 2xz\gamma}{2\sinh(t\gamma)}\Big),\end{aligned}$$

which gives

$$\begin{aligned}\int 1_A(\omega_1)\exp\bigl(-\vartheta^2\frac{\alpha^2}{2}\int_0^1 W_s^2\,ds\bigr)\,d\mathbb{W}(\omega)\\ = \int_A \frac{\sqrt{\vartheta\alpha}}{\sqrt{2\pi\sinh(\vartheta\alpha)}}\exp\Bigl(-\frac{z^2\vartheta\alpha\cosh(\vartheta\alpha)}{2\sinh(\vartheta\alpha)}\Bigr)\,dz\end{aligned}$$

in our case. This is a generalisation of the Cameron-Martin-Formula

$$E\bigl(e^{-\lambda\int_0^1 B_t^2\,dt}\bigr) = \bigl(\cosh\sqrt{2\lambda}\bigr)^{-1/2}$$

(see [RY99], chapter XI). We are interested in the exponential tails of this expression for $\vartheta\to\infty$.

Recalling the definitions

$$\sinh(x) = \frac{e^x - e^{-x}}{2} \quad \text{and} \quad \cosh(x) = \frac{e^x + e^{-x}}{2}$$

we observe that there are constants $0 < c_1 < c_2$ with

$$c_1 e^{-\alpha\vartheta/2} \le \frac{\sqrt{\alpha}}{\sqrt{2\pi\sinh(\vartheta\alpha)}} \le c_2 e^{-\alpha\vartheta/2} \quad \text{for all } \vartheta > 1. \tag{3.5}$$

(The value 1 is arbitrary, any positive number would do.) Also we find

$$\frac{\cosh(\vartheta\alpha)}{\sinh(\vartheta\alpha)} = \frac{e^{\vartheta\alpha}+e^{-\vartheta\alpha}}{e^{\vartheta\alpha}-e^{-\vartheta\alpha}} \longrightarrow 1 \quad \text{for } \vartheta\to\infty. \tag{3.6}$$

Because (3.5) and (3.6) are independent of and thus uniform in z we can conclude

$$\begin{aligned}\lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log\int 1_A(\omega_1)\exp\bigl(-\vartheta^2\frac{\alpha^2}{2}\int_0^1 W_s^2\,ds\bigr)\,d\mathbb{W}(\omega)\\ = \lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log\sqrt{\vartheta}\int_A\frac{\sqrt{\alpha}}{\sqrt{2\pi\sinh(\vartheta\alpha)}}\exp\Bigl(-\frac{z^2\alpha}{2}\cdot\frac{\cosh(\vartheta\alpha)}{\sinh(\vartheta\alpha)}\cdot\vartheta\Bigr)\,dz\\ = \lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log\int_A\exp\Bigl(-\frac{\alpha}{2}\vartheta - \frac{z^2\alpha}{2}\vartheta\Bigr)\,dz\\ = -\operatorname*{ess\,inf}_{z\in A}\bigl(\frac{\alpha}{2}+\frac{z^2\alpha}{2}\bigr)\\ = -\frac{\alpha}{2}(1+\operatorname*{ess\,inf}_{z\in A} z^2).\end{aligned} \tag{3.7}$$

This is the asymptotic behaviour for the $\exp(\vartheta^2 G)$ term.

Now, we consider the full density by fitting in the factor $\exp(\vartheta F)$. This is easy because the function F only depends on the endpoint z of the path. We get

$$\begin{aligned}\lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log P(X_1^\vartheta\in A)\\ = \lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log\int 1_A(\omega_1)\exp\bigl(\vartheta F(\omega)-\vartheta^2 G(\omega)\bigr)\,d\mathbb{W}(\omega)\\ = \lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log\sqrt{\vartheta}\int_A\frac{\sqrt{\alpha}}{\sqrt{2\pi\sinh(\vartheta\alpha)}}\\ \cdot\exp\Bigl(\frac{\alpha}{2}(1-z^2)\cdot\vartheta - \frac{z^2\alpha}{2}\cdot\frac{\cosh(\vartheta\alpha)}{\sinh(\vartheta\alpha)}\cdot\vartheta\Bigr)\,dz\\ = \lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log\int_A\exp\Bigl(-\frac{\alpha}{2}\vartheta + (\frac{\alpha}{2}-\frac{z^2\alpha}{2})\cdot\vartheta - \frac{z^2\alpha}{2}\vartheta\Bigr)\,dz\\ = -\operatorname*{ess\,inf}_{z\in A}\alpha z^2.\end{aligned}$$

This is the expected result.

As mentioned above the stationary distribution μ_ϑ of the Ornstein-Uhlenbeck process with parameter $\vartheta\alpha$ is a d-dimensional Gaussian distribution with mean 0 and covariance matrix $\frac{1}{2\vartheta\alpha} I_d$. Thus the result from theorem 2.13 coincides with the large deviation result about Gaussian distributions from corollary 2.12 in this case. Using either result we find

$$\lim_{\vartheta\to\infty} \mu_\vartheta(A) = -\inf_{z\in A} \alpha z^2$$

for every measurable set $A \subseteq \mathbb{R}$. Since the right hand side coincides with the rate from (3.4), for any $t > 0$ the large deviation behaviour of the point X_t^ϑ for $\vartheta \to \infty$ is the same as the large deviation behaviour of the stationary distribution.

Chapter 4

Tauberian Theorems of Exponential Type

In this chapter we study for positive random variables the relation between the behaviour of the Laplace transform near infinity and the distribution near zero. Theorems of this kind are called Tauberian theorems.

A result of de Bruijn shows that

$$E(e^{-\lambda X}) \sim e^{r\sqrt{\lambda}} \quad \text{for } \lambda \to \infty \qquad \text{and} \qquad P(X \le \varepsilon) \sim e^{s/\varepsilon} \quad \text{for } \varepsilon \downarrow 0$$

are in some sense equivalent and gives a relation between the constants r and s. We give sharp bounds for the upper and lower limits for this relation. This result will turn out to be a powerful tool to determine the large deviation behaviour of random variables where the Laplace transform is known.

4.1 De Bruijn's Theorem

From de Bruijn's Tauberian theorem we can easily conclude the following result.

Theorem 4.1. *Let $X \ge 0$ be a random variable and A an event with $P(A) > 0$. Then the limit*

$$r = \lim_{\lambda\to\infty} \frac{1}{\sqrt{\lambda}} \log E(e^{-\lambda X} \cdot 1_A)$$

exists if and only if

$$s = \lim_{\varepsilon\to 0} \varepsilon \log P(X \le \varepsilon, A)$$

exists and in this case we have $s = -r^2/4$.

Proof. In Theorem 4.12.9 from [BGT87] let $\alpha = -1$, $\phi(x) = 1/x$, $\psi(x) = 1/x^2$, and $B = |s|$. This gives the case of $A = \Omega$. For general sets A we switch to the measure $Q(\,\cdot\,) = P(\,\cdot\, \cap A)/P(A)$ and the corresponding expectation. This reduces the case of general A to the first case. (qed)

While dealing with expressions like the $P(X \le \varepsilon, A)$ above we will frequently use the following trivial scaling property.

Lemma 4.2. *(a) Assume that $s = \lim_{\varepsilon\downarrow 0} \varepsilon \log P(X \le \varepsilon)$ exists. Then for every $c > 0$, we have*

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log P(cX \le \varepsilon) = sc$$

and

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log P(X \le c\varepsilon) = s/c.$$

(b) The same relations hold for the lim sup *and the* lim inf.

4.2 An LDP for Brownian Paths with small L^2-Norm

In this section we give a first application of the Tauber theorem from section 4.1. By applying the theorem to the random variable

$$X = \int_0^t B_s^2 \, ds$$

where B is an one-dimensional Brownian motion we can derive an LDP for Brownian paths conditioned to have small L^2-norm.

The first step of this programme is to calculate the tails of the Laplace transform of X. Formula (1.9.7) from [BS96] states

$$E_x\Big(\exp\big(-\frac{\gamma^2}{2}\int_0^t B_s^2\, ds\big); B_t \in dz\Big) = \varphi(x;t,z)$$

where

$$\varphi(x;t,z) = \frac{\sqrt{\gamma}}{\sqrt{2\pi \sinh(t\gamma)}} \exp\Big(-\frac{(x^2+z^2)\gamma\cosh(t\gamma) - 2xz\gamma}{2\sinh(t\gamma)}\Big).$$

For a starting point x, measurable sets $A_1, \ldots, A_n \subseteq \mathbb{R}$, and fixed times $0 < t_1 < \cdots < t_n = t$ the Markov property of Brownian motion gives then

$$\begin{aligned} &E_x\Big(\exp\big(-\frac{\gamma^2}{2}\int_0^t B_s^2\, ds\big) 1_{A_1}(B_{t_1}) \cdots 1_{A_n}(B_{t_n})\Big) \\ &= \int_{A_1} \cdots \int_{A_n} \varphi(x; t_1, z_1)\varphi(z1; t_2 - t_1, z_2) \\ &\qquad\qquad \cdots \varphi(z_{n-1}; t_n - t_{n-1}, z_n)\, dz_n \cdots dz_1. \end{aligned}$$

We are only interested in the exponential tails of this expression for $\gamma \to \infty$. First observe that there are constants $0 < c_1 < c_2$ and $G > 0$ with

$$c_1 e^{-\gamma t/2} \le \frac{1}{\sqrt{2\pi\sinh(\gamma t)}} \le c_2 e^{-\gamma t/2} \quad \text{for all } \gamma > G.$$

Then we can use the relation $|2xy| \le x^2 + y^2$ to get

$$\frac{(x^2+z^2)}{2} \cdot \frac{\cosh(\gamma t) - 1}{\sinh(\gamma t)} \le \frac{(x^2+z^2)\cosh(\gamma t) - 2xz}{2\sinh(\gamma t)} \le \frac{(x^2+z^2)}{2} \cdot \frac{\cosh(\gamma t) + 1}{\sinh(\gamma t)}$$

for all $x, z \in \mathbb{R}$.

Let $\varepsilon > 0$. Because of

$$\frac{\cosh(\gamma t) \pm 1}{\sinh(\gamma t)} = \frac{e^{\gamma t} + e^{-\gamma t} \pm 1}{e^{\gamma t} - e^{-\gamma t}} \longrightarrow 1 \quad \text{for } \gamma \to \infty.$$

we can then find a $\gamma_0 > 0$, such that whenever $\gamma > \gamma_0$ the estimate

$$\frac{(x^2+z^2)}{2} \cdot (1-\varepsilon) \le \frac{(x^2+z^2)\cosh(\gamma t) - 2xz}{2\sinh(\gamma t)} \le \frac{(x^2+z^2)}{2} \cdot (1+\varepsilon)$$

holds for all $x, z \in \mathbb{R}$. We conclude

$$\begin{aligned}
&\limsup_{\gamma\to\infty} \frac{1}{\gamma} \log E_x\Big(\exp\big(-\frac{\gamma^2}{2}\int_0^t B_s^2\,ds\big)1_{A_1}(B_{t_1})\cdots 1_{A_n}(B_{t_n})\Big) \\
&\quad\le \lim_{\gamma\to\infty} \frac{1}{\gamma} \log \gamma^{n/2} c_2^n \int_{A_1}\cdots\int_{A_n} e^{-\gamma t_1/2} \exp\Big(-\gamma\frac{x^2+z_1^2}{2}(1-\varepsilon)\Big) \\
&\qquad\qquad \cdot e^{-\gamma(t_2-t_1)/2} \exp\Big(-\gamma\frac{z_1^2+z_2^2}{2}(1-\varepsilon)\Big)\cdot\cdots \\
&\qquad\qquad \cdot e^{-\gamma(t_n-t_{n-1})/2} \exp\Big(-\gamma\frac{z_{n-1}^2+z_n^2}{2}(1-\varepsilon)\Big)\,dz_n\cdots dz_1 \\
&\quad= \lim_{\gamma\to\infty} \frac{1}{\gamma} \log \int_{A_1}\cdots\int_{A_n} \exp\big(-\gamma t_n/2 - \gamma(x^2/2 + z_1^2 + \cdots \\
&\qquad\qquad \cdots + z_{n-1}^2 + z_n^2/2)(1-\varepsilon)\big)\,dz_n\cdots dz_1.
\end{aligned}$$

Note the special rôle of the endpoint $t_n = t$. With the help of the Laplace principle (see lemma 2.11) we can calculate the limit on the right hand side to get

$$\begin{aligned}
&\limsup_{\gamma\to\infty} \frac{1}{\gamma} \log E_x\Big(\exp\big(-\frac{\gamma^2}{2}\int_0^t B_s^2\,ds\big)1_{A_1}(B_{t_1})\cdots 1_{A_n}(B_{t_n})\Big) \\
&\quad\le -\operatorname*{ess\,inf}_{z_1\in A_1,\ldots,z_n\in A_n} \Big(t/2 + (x^2/2 + z_1^2 + \cdots + z_{n-1}^2 + z_n^2/2)(1-\varepsilon)\Big).
\end{aligned}$$

for all $\varepsilon > 0$ and thus

$$\begin{aligned}
&\limsup_{\gamma\to\infty} \frac{1}{\gamma} \log E_x\Big(\exp\big(-\frac{\gamma^2}{2}\int_0^t B_s^2\,ds\big)1_{A_1}(B_{t_1})\cdots 1_{A_n}(B_{t_n})\Big) \\
&\quad\le -\operatorname*{ess\,inf}_{z_1\in A_1,\ldots,z_n\in A_n} (t/2 + x^2/2 + z_1^2 + \cdots + z_{n-1}^2 + z_n^2/2).
\end{aligned}$$

A completely analogous calculation (using the lower bounds from above) gives

$$\begin{aligned}
&\liminf_{\gamma\to\infty} \frac{1}{\gamma} \log E_x\Big(\exp\big(-\frac{\gamma^2}{2}\int_0^t B_s^2\,ds\big)1_{A_1}(B_{t_1})\cdots 1_{A_n}(B_{t_n})\Big) \\
&\quad\ge -\operatorname*{ess\,inf}_{z_1\in A_1,\ldots,z_n\in A_n} (t/2 + x^2/2 + z_1^2 + \cdots + z_{n-1}^2 + z_n^2/2).
\end{aligned}$$

and together this shows

$$\begin{aligned}
&\lim_{\gamma\to\infty} \frac{1}{\gamma} \log E_x\Big(\exp\big(-\frac{\gamma^2}{2}\int_0^t B_s^2\,ds\big)1_{A_1}(B_{t_1})\cdots 1_{A_n}(B_{t_n})\Big) \\
&\quad= -\operatorname*{ess\,inf}_{z_1\in A_1,\ldots,z_n\in A_n} (t/2 + x^2/2 + z_1^2 + \cdots + z_{n-1}^2 + z_n^2/2).
\end{aligned} \tag{4.1}$$

Corollary 4.3. *Let B be a one-dimensional Brownian Motion. Then*

$$\lim_{\varepsilon\downarrow 0} \varepsilon\cdot\log P_x\left(\int_0^t B_s^2\,ds \le \varepsilon, B_t \in A\right) = -\frac{(t + x^2 + \operatorname{ess\,inf}_{z\in A} z^2)^2}{8}$$

for every $x \in \mathbb{R}$ and every set A with $P(B_t \in A) > 0$ and in particular

$$\lim_{\varepsilon\downarrow 0} \varepsilon\cdot\log P\left(\int_0^t B_s^2\,ds \le \varepsilon\right) = -\frac{t^2}{8}.$$

Proof. Setting $\lambda = \gamma^2/2$ in equation (4.1) gives

$$r = \lim_{\lambda\to\infty} \frac{1}{\sqrt{\lambda}} \log E_x(e^{-\lambda\int_0^t B_s^2\,ds}\cdot 1_A(B_t)) = -\frac{1}{\sqrt{2}}(t + x^2 + \operatorname*{ess\,inf}_{z\in A} z^2).$$

Now we can use the Tauber theorem to get the first equality. The second claim follows by taking $x = 0$ and $A = \mathbb{R}$. (qed)

At a first glance it may seem strange that the rate is quadratic in the interval length t. But the following heuristic reveals that this actually makes sense: write the time interval $[0;t]$ as the disjoint union of intervals $I_1, \dots, I_n$ and assume for the moment that the events $\int_{I_k} B_s^2\,ds < \varepsilon$ are asymptotically independent. Further it makes sense to assume that the contribution of an interval I_k to the integral $\int_0^t B_s^2\,ds$ is proportional to its length. So we consider

$$\begin{aligned}\lim_{\varepsilon\downarrow 0} \varepsilon \cdot \log P\Bigl(\bigcap_{k=1}^{n} \Bigl\{\int_{I_k} B_s^2\,ds \leq \frac{|I_k|}{t}\cdot\varepsilon\Bigr\}\Bigr) \\ = -\frac{1}{8}\sum_{k=1}^{n} |I_k|^2\, t/|I_k| = -\frac{1}{8}\,t^2,\end{aligned}$$

where the rates were calculated using the scaling property from lemma 4.2. The result is the same as the rate which we got for the full interval. So the quadratic dependency on t is compatible with the assumption that the contributions of the intervals $I_1, \dots, I_n$ are asymptotically independent and proportional to the interval length.

Example 4.1. With the help of corollary 4.3 we can reproduce the results of example 2.2 for the L^2-norm instead of the supremum norm. The exponential rate of $P\bigl(\|B\|_2 > c\bigr)$ for $c \to \infty$ is again calculated with Schilder's theorem. With $\varepsilon = 1/c^2$ we have

$$\int_0^t B_s^2\,ds > c^2 \iff \sqrt{\varepsilon}B \in \bigl\{\omega \bigm| \int_0^t \omega_s^2\,ds > 1\bigr\} =: A.$$

The rate function $I(\omega)$ under the constraint $\int_0^t \omega_t^2\,dt = \beta$ is minimal for the function $\tilde\omega$ with

$$\tilde\omega_s = \sqrt{2\beta/t}\,\sin(s\pi/2t)$$

for all $s \in [0;t]$ and we get the minimal value

$$I(\tilde\omega) = \frac{1}{2}\int_0^t \frac{2\beta}{t}\cdot\frac{\pi^2}{4t^2}\cos^2\bigl(\frac{s\pi}{2t}\bigr)\,ds = \frac{\beta\pi^2}{8t^2}.$$

Thus the set A is a continuity set of the rate function and we find

$$\begin{aligned}&\lim_{c\to\infty}\frac{1}{c^2}\log P\bigl(\|B\|_2 > c\bigr) \\ &= \lim_{\varepsilon\downarrow 0}\varepsilon\log P\bigl(\sqrt{\varepsilon}B \in A\bigr) \\ &= -\inf\Bigl\{\frac{1}{2}\int_0^t \dot\omega_s^2\,ds \Bigm| \omega \in A\Bigr\} \\ &= -\inf_{\beta>1}\frac{\beta\pi^2}{8t^2} = -\frac{\pi^2}{8t^2}.\end{aligned}$$

This is consistent with the results of example 2.2. Since the event $\|B\|_2 > c$ implies $\|B\|_\infty > c/\sqrt{t}$ we expect

$$\begin{aligned}&\lim_{c\to\infty}\frac{1}{c^2}\log P\bigl(\|B\|_2 > c\bigr) \\ &\leq \lim_{c\to\infty}\frac{1}{c^2}\log P\bigl(\|B\|_\infty > c/\sqrt{t}\bigr) = \frac{1}{t}\lim_{c\to\infty}\frac{1}{c^2}\log P\bigl(\|B\|_\infty > c\bigr)\end{aligned}$$

and indeed the correspoding rates are $-\pi^2/8t^2$ for the left hand side and $-4/8t^2$ for the right-hand side.

The large deviation behaviour of $\|B\|_2 < \varepsilon$ for $\varepsilon \downarrow 0$ is described by corollary 4.3. We get

$$\lim_{\varepsilon\downarrow 0}\varepsilon^2\log P\bigl(\|B\|_2 < \varepsilon\bigr) = \lim_{\varepsilon\downarrow 0}\varepsilon^2\log P\bigl(\int_0^t B_s^2\,ds < \varepsilon^2\bigr) = -\frac{t^2}{8}.$$

Again, we can compare this with the results from example 2.2. Whenever we have $\|B\|_\infty < \varepsilon/\sqrt{t}$ we also have $\|B\|_2 < \varepsilon$ and thus we should have

$$\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log P\big(\|B\|_2 < \varepsilon\big) \\ \geq \lim_{\varepsilon\downarrow 0} \varepsilon^2 \log P\big(\|B\|_\infty < \varepsilon/\sqrt{t}\big) = t \lim_{\varepsilon\downarrow 0} \varepsilon^2 \log P\big(\|B\|_\infty < \varepsilon\big).$$

The results from above and from example 2.2 are $-t^2/8$ for the left hand side and $-\pi^2 t^2/8$ for the right hand side, so everything fits together well.

Later we will need the results of Corollary 4.3 uniformly in the initial condition x. To achive this uniformity we use a version of Anderson's inequality (this is corollary 5 in Anderson's original paper [And55]):

Lemma 4.4. *Let $(X_s)_{0\leq s\leq t}$ and $(Y_s)_{0\leq s\leq t}$ be two separable Gaussian processes and $k \in [0;1]$ with $E(X_s) = kE(Y_s)$ and $\mathrm{Cov}(X_r, X_s) = \mathrm{Cov}(Y_r, Y_s) = C(r,s)$ for all $0 \leq r, s \leq t$. Assume that C is continuous. Then*

$$P\Big(\int_0^t X_s^2\,ds \leq \varepsilon\Big) \geq P\Big(\int_0^t Y_s^2\,ds \leq \varepsilon\Big)$$

and

$$P\Big(\sup_{0\leq s\leq t} |X_s| \leq \varepsilon\Big) \geq P\Big(\sup_{0\leq s\leq t} |Y_s| \leq \varepsilon\Big)$$

for all $\varepsilon > 0$.

Lemma 4.5. *Let B be a one-dimensional Brownian Motion and $A \subseteq \mathbb{R}$ closed. Then*

$$\lim_{\varepsilon\downarrow 0} \varepsilon \cdot \log \sup_{x\in A} P_x\Big(\int_0^t B_s^2\,ds \leq \varepsilon\Big) = -\inf_{x\in A} \frac{(t+x^2)^2}{8}.$$

Proof. Let $x, y \in A$ with $0 < |x| < |y|$. Then lemma 4.4 applied to $X = B + |x|$, $Y = B + |y|$ and $k = |x/y|$ and the symmetry of Brownian motion gives

$$P_x\Big(\int_0^t B_s^2\,ds \leq \varepsilon\Big) \geq P_y\Big(\int_0^t B_s^2\,ds \leq \varepsilon\Big). \tag{4.2}$$

Now choose $x \in A$ with $|x| = \inf\{\,|y| \mid y \in A\,\}$. Then the estimate (4.2) gives

$$P_x\Big(\int_0^t B_s^2\,ds \leq \varepsilon\Big) = \sup_{y\in A} P_y\Big(\int_0^t B_s^2\,ds \leq \varepsilon\Big)$$

and the claim follows with corollary 4.3. (qed)

Let $\mathcal{X}$ be the space of all maps $\omega\colon [0;t] \to \mathbb{R}$ such that $\omega_0 = 0$ equipped with the topology of pointwise convergence. On $\mathcal{X}$ define the family $(P_\varepsilon)_{\varepsilon>0}$ of measures by

$$P_\varepsilon(A) = \mathbb{W}\Big(A \;\Big|\; \int_0^t W_s^2\,ds \leq \varepsilon\Big)$$

for all measurable $A \subseteq \mathcal{X}$.

Theorem 4.6. *On the space $\mathcal{X}$ the family $(P_\varepsilon)_{\varepsilon>0}$ satisfies the LDP with the good rate function*

$$I(\omega) = \sup\big\{(t + 2\omega_{t_1}^2 + \cdots + 2\omega_{t_n}^2 + \omega_t^2)^2/8 - t^2 \;\big|\; n \in \mathbb{N}, 0 < t_1 < \cdots < t_n < t\big\}$$

for all $\omega \in \mathcal{X}$.

Proof. For measurable sets $A_1, \dots, A_n \subseteq \mathbb{R}$ and fixed times $0 < t_1 < \cdots < t_n = t$ the Tauber theorem 4.1 applied to equation (4.1) gives

$$\begin{aligned}
&\lim_{\varepsilon \downarrow 0} \varepsilon \cdot \log P\Big((B_{t_1}, B_{t_2}, \dots, B_{t_n}) \in A_1 \times A_2 \times \cdots \times A_n \;\Big|\; \int_0^t B_s^2 \, ds \leq \varepsilon \Big) \\
&\quad = \lim_{\varepsilon \downarrow 0} \varepsilon \cdot \log P\Big(B_{t_1} \in A_1, B_{t_2} \in A_2, \dots, B_{t_n} \in A_n, \int_0^t B_s^2 \, ds \leq \varepsilon \Big) \\
&\qquad\qquad - \lim_{\varepsilon \downarrow 0} \varepsilon \cdot \log P\Big(\int_0^t B_s^2 \, ds \leq \varepsilon \Big) \\
&\quad = -\big(t + \operatorname*{ess\,inf}_{z \in A_1 \times A_2 \times \cdots \times A_n} (2z_1^2 + \cdots + 2z_{n-1}^2 + z_n^2) \big)^2 / 8 + t^2/8.
\end{aligned} \tag{4.3}$$

Using $A_n = \mathbb{R}$ we can drop the assumption $t_n = t$ and arrive at the following result. For all measurable sets $A_1, \dots, A_n \subseteq \mathbb{R}$ and fixed times $0 < t_1 < \cdots < t_n \leq t$ we have

$$\begin{aligned}
&\lim_{\varepsilon \downarrow 0} \varepsilon \cdot \log P\Big((B_{t_1}, B_{t_2}, \dots, B_{t_n}) \in A_1 \times A_2 \times \cdots \times A_n \;\Big|\; \int_0^t B_s^2 \, ds \leq \varepsilon \Big) \\
&\quad = - \operatorname*{ess\,inf}_{z \in A_1 \times A_2 \times \cdots \times A_n} I_{t_1, \dots, t_n}(z)
\end{aligned}$$

where $I_{t_1, \dots, t_n} : \mathbb{R}^n \to \mathbb{R}_+$ is defined by

$$I_{t_1, \dots, t_n}(z) = \frac{1}{8} \begin{cases} \big(t + 2z_1^2 + \cdots + 2z_n^2 \big)^2 - t^2, & \text{if } t_n < t, \text{ and} \\ \big(t + 2z_1^2 + \cdots + 2z_{n-1}^2 + z_n^2 \big)^2 - t^2 & \text{for } t_n = t. \end{cases}$$

With the exception of the endpoint t the actual positions of the t_i have no influence on the rate. The endpoint t is special, because the process does not need to return to the origin quickly after a visit in A_n at time t, so at the end of the interval it is "cheaper" to be far away from the origin.

Because the rate function $I_{t_1, \dots, t_n}$ is continuous we get an LDP on $\mathbb{R}^n$ as in the remark on page 19. From this we can get the LDP on the path space with rate function

$$I(\omega) = \sup \big\{ I_{t_1, \dots, t_n}(\omega_{t_1}, \dots, \omega_{t_n}) \;\big|\; n \in \mathbb{N}, 0 < t_1 < \cdots < t_n \leq t \big\}$$

by applying the Dawson-Gärtner theorem about large deviations for projective limits (see theorem 2.9). (qed)

Note that the rate function I in the theorem will typically take its infimum for a non-continuous path ω. Assume ω is continuous and non-zero. Let $\varepsilon = \|\omega\|_\infty / 2$. Then we find infinitely many distinct times t with $\omega_t^2 > \varepsilon^2$ and thus $I(\omega) = +\infty$.

4.3 Upper and Lower Limits

The remaining part of this chapter contains the proof of a theorem about upper and lower limits in the Tauberian theorem. In contrast to theorem 4.1 the result of this section applies without any assumption on the distribution of X.

Theorem 4.7. *Let $X \geq 0$ be a random variable and A an event with $P(A) > 0$. Define the upper and lower limits*

$$\bar{r} = \limsup_{\lambda \to \infty} \frac{1}{\sqrt{\lambda}} \log E(e^{-\lambda X} \cdot 1_A) \quad \text{and} \quad \underline{r} = \liminf_{\lambda \to \infty} \frac{1}{\sqrt{\lambda}} \log E(e^{-\lambda X} \cdot 1_A)$$

as well as

$$\bar{s} = \limsup_{\varepsilon \to 0} \varepsilon \log P(X \leq \varepsilon, A) \quad \text{and} \quad \underline{s} = \liminf_{\varepsilon \to 0} \varepsilon \log P(X \leq \varepsilon, A).$$

Then $-\bar{r}^2/4 = \bar{s}$ and for the lower limits we have the sharp estimates $-\underline{r}^2 \leq \underline{s} \leq -\underline{r}^2/4$.

Proof. As in the proof of theorem 4.1 it is enough to consider the case $A = \mathbb{R}$. First note that, because X is positive, the expectation $E(e^{-\lambda X})$ exists for all $\lambda \geq 0$ and is a number between 0 and 1. So all the values $\bar{r}$, $\underline{r}$, $\bar{s}$, and $\underline{s}$ will be negative.

The estimate $\bar{s} \leq -\bar{r}^2/4$ follows from the exponential Markov inequality: Let $\varepsilon > 0$. From

$$E(e^{-\lambda X}) \geq e^{-\lambda\varepsilon} P\big(e^{-\lambda X} \geq e^{-\lambda\varepsilon}\big) = e^{-\lambda\varepsilon} P\big(X \leq \varepsilon\big)$$

we get $P(X \leq \varepsilon) \leq e^{\lambda\varepsilon} E(e^{-\lambda X})$ and thus

$$\varepsilon \log P(X \leq \varepsilon) \leq \varepsilon\big(\lambda\varepsilon + \log E(e^{-\lambda X})\big) \quad \text{for all } \lambda \geq 0.$$

For $\lambda = \bar{r}^2/4\varepsilon^2$ the bound becomes

$$\varepsilon \log P(X \leq \varepsilon) \leq \bar{r}^2/4 + \varepsilon \log E(e^{-X\bar{r}^2/4\varepsilon^2}).$$

Taking upper limits we get

$$\begin{aligned}
\bar{s} = \limsup_{\varepsilon\downarrow 0} \varepsilon \cdot \log P(X \leq \varepsilon) &\leq \frac{\bar{r}^2}{4} + \limsup_{\varepsilon\downarrow 0} \varepsilon \cdot \log E(e^{-X\bar{r}^2/4\varepsilon^2}) \\
&= \frac{\bar{r}^2}{4} - \frac{\bar{r}}{2} \limsup_{\varepsilon\downarrow 0} \frac{2\varepsilon}{|\bar{r}|} \log E(e^{-X(\bar{r}/2\varepsilon)^2}) \\
&= \frac{\bar{r}^2}{4} - \frac{\bar{r}}{2} \cdot \bar{r} = -\frac{\bar{r}^2}{4}.
\end{aligned}$$

Replacing all upper limits in the previous argument with lower limits gives $\underline{s} \leq -\underline{r}^2/4$.

A more careful analysis is necessary to prove $\bar{s} \geq -\bar{r}^2/4$. We can express $\bar{r}$ via the lower tails of X:

$$\begin{aligned}
\bar{r} &= \limsup_{\lambda\to\infty} \frac{1}{\sqrt{\lambda}} \log E(e^{-\lambda X}) \\
&= \limsup_{\lambda\to\infty} \frac{1}{\sqrt{\lambda}} \log \int_0^1 P(e^{-\lambda X} \geq t)\, dt \\
&\overset{t=e^{-u}}{=} \limsup_{\lambda\to\infty} \frac{1}{\sqrt{\lambda}} \log \int_0^\infty P(X \leq u/\lambda) e^{-u}\, du \\
&= \limsup_{\varepsilon\downarrow 0} \varepsilon \log \int_0^\infty P(X \leq u\varepsilon^2) e^{-u}\, du.
\end{aligned}$$

The definition of $\bar{s}$ gives that for every δ with $0 < \delta < |\bar{s}|$ there exists an $E > 0$, such that for every $\eta < E$ we have

$$P(X \leq \eta) \leq \eta^{-3/2} e^{(\bar{s}+\delta)/\eta}.$$

I want to use the relation

$$\int_0^\infty z u^{-3/2} \exp\big(-\frac{z^2}{u} - u\big)\, du = \sqrt{\pi} e^{-2z}.$$

In the context of the above estimate this gives

$$\begin{aligned}
\int_0^\infty P(X \leq u\varepsilon^2) e^{-u}\, du &\leq \int_0^{E/\varepsilon^2} (\varepsilon^2 u)^{-3/2} \exp\Big(-\big(\frac{\sqrt{|\bar{s}+\delta|}}{\varepsilon}\big)^2 \cdot \frac{1}{u} - u\Big)\, du \\
&\quad + \int_{E/\varepsilon^2}^\infty 1 \cdot e^{-u}\, du \\
&\leq \varepsilon^{-3} \frac{\varepsilon}{\sqrt{|\bar{s}+\delta|}} \sqrt{\pi} e^{-2\sqrt{|\bar{s}+\delta|}/\varepsilon} + e^{-E/\varepsilon^2}
\end{aligned}$$

The sum is dominated by the first term, so we get

$$\bar{r} \leq -2\sqrt{|\bar{s}+\delta|} \quad \text{whenever } 0 < \delta < |\bar{s}|$$

and thus $\bar{r} \leq -2\sqrt{|\bar{s}|}$. Because both, $\bar{r}$ and $\bar{s}$, are negative this shows $\bar{s} \geq -\bar{r}^2/4$.

Finally, we can prove $-\underline{r}^2 \leq \underline{s}$. Using the estimate $e^{-\lambda x} \leq 1_{[0;\varepsilon]}(x) + e^{-\lambda\varepsilon} 1_{(\varepsilon;\infty)}(x)$ for all $x \geq 0$ with $\lambda = |\underline{s}|/\varepsilon^2$ gives

$$E\big(e^{-|s|X/\varepsilon^2}\big) \leq P(X \leq \varepsilon) + e^{-|\underline{s}|\varepsilon/\varepsilon^2} P(X > \varepsilon) \leq P(X \leq \varepsilon) + e^{-|\underline{s}|/\varepsilon}.$$

Because for the second term in the sum the limit $\lim_{\varepsilon\downarrow 0} \varepsilon \log e^{-|\underline{s}|/\varepsilon} = -|\underline{s}|$ exists, we can conclude

$$\begin{aligned} -|\underline{r}|\sqrt{|\underline{s}|} &= \liminf_{\varepsilon\downarrow 0} \varepsilon \log E\big(e^{-|s|X/\varepsilon^2}\big) \\ &\leq \max\Big(\liminf_{\varepsilon\downarrow 0} \varepsilon \log P(X \leq \varepsilon)\,,\ \lim_{\varepsilon\downarrow 0} \varepsilon \log e^{-|\underline{s}|/\varepsilon}\Big) \\ &= \max\big(-|\underline{s}|\,,\ -|\underline{s}|\big) = -|\underline{s}|. \end{aligned}$$

Taking squares the estimate becomes $\underline{r}^2 \geq |\underline{s}|$ and multiplication with -1 gives the result.

The upper bound on $\underline{s}$ is sharp, because in the case of theorem 4.1 we have equality there. The fact that the lower bound for the lower limit $\underline{s}$ is sharp is shown by the example at the end of this section. (qed)

Corollary 4.8. *Under the assumptions of theorem 4.7 we have $\bar{r} = -2\sqrt{|\bar{s}|}$ and for the lower limits we have the sharp estimates $-2\sqrt{|\underline{s}|} \leq \underline{r} \leq -\sqrt{|\underline{s}|}$.*

Proof. On $(-\infty; 0]$ the map $x \mapsto -\sqrt{|x|}$ is strictly monotonically increasing. Thus for $r, s \leq 0$ we have $s \leq -r^2$ if and only if $-\sqrt{|s|} \leq r$. Applying this to the results of theorem 4.7 proves the corollary. (qed)

Note that theorem 4.7 does not directly imply theorem 4.1. If the limit s from theorem 4.1 exists, then we get

$$s \leq -\underline{r}^2/4 \leq -\bar{r}^2/4 = s,$$

i.e. the limit r also exists and satisfies $s = -r^2/4$. But if we assume that r exists, then theorem 4.7 only gives

$$-r^2 \leq \underline{s} \leq \bar{s} = -r^2/4$$

and we cannot directly conclude that the limit s from theorem 4.1 exists.

The following example shows that for general random variables X the lower bound $-\underline{r}^2 \leq \underline{s}$ on the lower limit $\underline{s}$ is sharp.

Example 4.2. Let $s < 0$ and $(\varepsilon_n)_{n\in\mathbb{N}_0}$ be a strictly decreasing sequence with $\varepsilon_0 = \infty$ and $\lim_{n\to\infty} \varepsilon_n = 0$. Then we have

$$\sum_{n\in\mathbb{N}} \Big(e^{-|s|/\varepsilon_{n-1}} - e^{-|s|/\varepsilon_n}\Big) = e^{-|s|/\varepsilon_0} - \lim_{n\to\infty} e^{-|s|/\varepsilon_n} = 1 - 0 = 1$$

and we can define a random variable X with values in the set $\{\, \varepsilon_n \mid n \in \mathbb{N} \,\}$ by

$$P(X = \varepsilon_n) = e^{-|s|/\varepsilon_{n-1}} - e^{-|s|/\varepsilon_n}$$

for all $n \in \mathbb{N}$. This random variable has

$$P(X \leq \varepsilon) = \sum_{n=n(\varepsilon)}^{\infty} \Big(e^{-|s|/\varepsilon_{n-1}} - e^{-|s|/\varepsilon_n}\Big) = e^{-|s|/\varepsilon_{n(\varepsilon)-1}}$$

with $n(\varepsilon) = \min\{\, n \in \mathbb{N} \mid \varepsilon_n \leq \varepsilon \,\}$ and consequently

$$\varepsilon \log P(X \leq \varepsilon) = -|s| \frac{\varepsilon}{\varepsilon_{n(\varepsilon)-1}}.$$

By definition of $n(\varepsilon)$ we have $\varepsilon_{n(\varepsilon)} \leq \varepsilon < \varepsilon_{n(\varepsilon)-1}$. This allows us to calculate the exponential tail rates $\underline{s} = s$ and, because s is negative, $\bar{s} = s \cdot \liminf_{n\to\infty} \varepsilon_n/\varepsilon_{n-1}$.

Choosing different sequences (ε_n) leads to different values for $\bar{s}$, $\bar{r}$, and $\underline{r}$. For our example let $q < 1$ and define $\varepsilon_n = q^n$ for all $n \in \mathbb{N}$. The above calculation shows $\underline{s} = s$ and $\bar{s} = qs$. The theorem gives $\bar{r} = -2\sqrt{q|s|}$ and $\underline{r} \in [-2\sqrt{|s|}; -\sqrt{|s|}]$. We want to further examine $\underline{r}$. The Laplace transform of X calculates as

$$\begin{aligned} E(e^{-\lambda X}) &= \sum_{n\in\mathbb{N}} e^{-\lambda q^n}\big(e^{-|s|/q^{n-1}} - e^{-|s|/q^n}\big) \\ &= \sum_{n\in\mathbb{N}} e^{-\lambda q^n - |s|/q^{n-1}}\big(1 - e^{-|s|(1-q)/q^n}\big). \end{aligned}$$

Because of $\exp(-|s|(1-q)/q^n) \to 0$ for $n \to \infty$ we have $1/2 < 1 - \exp(-|s|(1-q)/q^n) < 1$ for sufficiently large n. Define $n(\lambda)$ by $q^{n(\lambda)} \in [q\sqrt{|s|/\lambda}; \sqrt{|s|/\lambda})$. With $f(x) = \exp(-\lambda x - q|s|/x)$ we have

$$E(e^{-\lambda X}) > \exp\big(-\lambda q^{n(\lambda)} - |s|/q^{n(\lambda)-1}\big)\frac{1}{2} = \frac{1}{2} f(q^{n(\lambda)})$$

for sufficiently large λ. Because f is increasing on the interval $(0; \sqrt{q|s|/\lambda}]$ and decreasing on $[\sqrt{q|s|/\lambda}; \infty)$ we can estimate f on $[q\sqrt{|s|/\lambda}; \sqrt{|s|/\lambda})$ by its values on the boundaries. This leads to

$$E(e^{-\lambda X}) > \frac{1}{2}\min\big(f(q\sqrt{|s|/\lambda}), f(\sqrt{|s|/\lambda})\big) = \frac{1}{2}\exp\big(-(1+q)\sqrt{\lambda|s|}\big)$$

for sufficiently large λ. Taking lower limits we get

$$-\sqrt{|s|} \geq \underline{r} \geq -(1+q)\sqrt{|s|}$$

where the first inequality comes from the theorem or equivalently

$$-\underline{r}^2 \leq \underline{s} \leq -\underline{r}^2/(1+q)^2.$$

This shows that by choosing small values of q we can force $\underline{s}$ to be arbitrarily close to $-\underline{r}^2$ without $-\underline{r}^2$ being close to 0. So the lower bound on $\underline{s}$ from the theorem is sharp.

Chapter 5

Diffusions with Strong Drift

In this chapter we derive an LDP for the behaviour of the endpoint X_t of a diffusion when the drift is strong. This is a generalisation of the result for the Ornstein-Uhlenbeck process in chapter 3.2.

We want to determine the large deviations behaviour for the endpoint X_t of solutions of the $\mathbb{R}$-valued stochastic differential equation

$$\begin{aligned} dX^\vartheta_s &= \vartheta b(X_s) \bullet ds + dB_s \quad \text{on } [0;t] \\ X^\vartheta_0 &= z \in \mathbb{R} \end{aligned} \tag{5.1}$$

for large values ϑ.

The situation here is different from the situation in the Freidlin-Wentzell theorem. In our case the length t of the interval is fixed, and instead ϑ goes to infinity. One can rescale equation (5.1) as follows. Define $Y^\vartheta_s = X^\vartheta_{t/\vartheta}$ for all $s \in [0;\vartheta t]$. Then by lemma 1.3 the process Y^ϑ is a solution of the SDE

$$\begin{aligned} dY^\vartheta_s &= b(Y_s) \bullet ds + \frac{1}{\sqrt{\vartheta}} dB_s \quad \text{on } [0;\vartheta t] \\ Y^\vartheta_0 &= 0 \end{aligned}$$

and we have

$$P(X^\vartheta_t \in A) = P(Y^\vartheta_{\vartheta t} \in A).$$

The rescaled problem looks more similar to the situation from the Freidlin-Wentzell theory, because now the noise decreases. But the length of the transformed time interval depends on ϑ, so the Freidlin-Wentzell theorem still cannot be applied easily.

From Lemma 1.5 we know the density of the distribution of X^ϑ_t: assuming $X^\vartheta_0 = 0$ and $b = -\nabla\Phi$ we get

$$P(X^\vartheta_t \in A) = \int 1_A(\omega_t) \exp\bigl(\vartheta F(\omega) - \vartheta^2 G(\omega)\bigr)\, d\mathbb{W}(\omega)$$

where

$$\begin{aligned} F(\omega) &= \Phi(\omega_0) - \Phi(\omega_t) + \frac{1}{2}\int_0^t \Delta\Phi(\omega_s)\, ds \quad \text{and} \\ G(\omega) &= \frac{1}{2}\int_0^t b^2(\omega_s)\, ds. \end{aligned}$$

It will turn out, that the only paths which contribute for the large deviations behaviour of X^ϑ_t are those, which correspond to very small values of G. This chapter consists of three parts. First we examine the situation that during a short time interval the process runs over a long distance while still keeping $\int b^2(\omega_s)\,ds$ small. This will be used for the initial and the final piece of the path. In the second section we examine the situation that $\int b^2(\omega_s)\,ds$ is small over a long interval of time. This will be used to treat the main piece of the path. In the third section we fit these two results together in order to find the exponential rate for the LDP.

5.1 Reaching the Final Point

The following part helps to estimate the probability that the path travels quickly between an equilibrium point of the drift and the final resp. initial point. Here Schilder's theorem can be applied and we will reduce the evaluation of the rate function to a variational problem.

The key for evaluating the rate function in proposition 5.3 below is the following lemma. We defer the proof of the lemma until the end of the section.

Lemma 5.1. *Let $v\colon \mathbb{R} \to [0;\infty)$ be a positive, two times continuously differentiable function with $\liminf_{|x|\to\infty} v(x) > 0$ and $m \in \mathbb{R}$ with $v(x) = 0$ if and only if $x = m$ and $v''(m) > 0$. For $a, z \in \mathbb{R}$ and $\beta \geq 0$ define*

$$M_t^{a,z,\beta} = \Big\{\omega \in C[0;t] \;\Big|\; \omega_0 = 0, \omega_t = a - z, \frac{1}{2}\int_0^t v(\omega_s + z)\,ds = \beta\Big\}$$

and

$$J(a,z) = \frac{1}{4}\Big(\big|\int_z^m \sqrt{v(x)}\,dx\big| + \big|\int_m^a \sqrt{v(x)}\,dx\big|\Big)^2.$$

Consider the rate function

$$I_t(\omega) = \begin{cases} \frac{1}{2}\int_0^t |\dot\omega|^2\,ds, & \text{if } \omega \text{ is absolutely continuous, and} \\ +\infty & \text{else.} \end{cases}$$

Let $K_1, K_2 \subseteq \mathbb{R}$ be compact sets with $0 \notin K_1 \cap K_2$ and $B \subseteq \mathbb{R}_+$ be bounded with $0 \in B$. Then we have

$$\inf\Big\{I_t(\omega) \;\Big|\; \omega \in \bigcup_{\beta\in B} M_t^{a,z,\beta}\Big\} \longrightarrow \frac{1}{\sup B} J(a,z) \quad \text{for } t \to \infty,$$

uniformly over $a \in K_2$ and $z \in K_1$.

Lemma 5.2. *Let $M_t^{a,z,\beta}$ be as in lemma 5.1. Then for every pair $K_1, K_2 \subseteq \mathbb{R}$ of compact sets the set*

$$M = \bigcup_{z\in K_1}\bigcup_{a\in K_2}\bigcup_{0\leq\beta\leq 1} M_t^{a,z,\beta}$$

is closed in $C_0([0;t],\mathbb{R})$.

Proof. By definition of the sets $M_t^{a,z,\beta}$ we have

$$M = \bigcup_{z\in K_1}\Big\{\omega \in C[0;t] \;\Big|\; \omega_0 = 0, \omega_t + z \in K_2, \frac{1}{2}\int_0^t v(\omega_r + z)\,dr \leq 1\Big\}.$$

Assume that $\omega \in C_0([0;t],\mathbb{R}) \setminus M$. The either $\omega_t + z \notin K_2$ for all $z \in K_1$, i.e. ω_t lies outside the compact set $K_2 - K_1$, or

$$\frac{1}{2}\int_0^t v(\omega_r + z)\,dr > 1$$

for every $z \in K_2$, i.e.

$$\inf_{z\in K_2}\frac{1}{2}\int_0^t v(\omega_r + z)\,dr > 1$$

because K_2 is compact and v and the integral are continuous. In both cases we can find an $\varepsilon > 0$, such that the ball $B(\omega,\varepsilon)$ also lies in $C_0([0;t],\mathbb{R}) \setminus M$. Thus M is the complement of an open set. (qed)

The main result of this section is the following proposition.

Proposition 5.3. *Let B be a one-dimensional Brownian motion with start in $z \in \mathbb{R}$ and $b\colon \mathbb{R} \to \mathbb{R}$ be a two times continuously differentiable function with $\liminf_{|x|\to\infty} |b(x)| > 0$ and $m \in \mathbb{R}$ with $b(x) = 0$ if and only if $x = m$ and $b'(m) \neq 0$. Then for every pair of compact sets $K_1, K_2 \subseteq \mathbb{R}$ we have*

$$\begin{aligned}\limsup_{t\to\infty} \limsup_{\varepsilon\downarrow 0} \sup_{z\in K_1} \varepsilon \log P_z\Big(\frac{1}{2}\int_0^{t\varepsilon} b^2(B_s)\, ds \le \varepsilon, B_{t\varepsilon} \in K_2\Big) \\ \le -\frac{1}{4} \inf_{z\in K_1} \inf_{a\in K_2} \Big(\Big|\int_z^m |b(x)|\, dx\Big| + \Big|\int_m^a |b(x)|\, dx\Big|\Big)^2\end{aligned}$$

and for every $z \in \mathbb{R}$ and every open set $O \subseteq \mathbb{R}$ we have

$$\begin{aligned}\liminf_{t\to\infty} \liminf_{\varepsilon\downarrow 0} \varepsilon \log P_z\Big(\frac{1}{2}\int_0^{t\varepsilon} b^2(B_s)\, ds \le \varepsilon, B_{t\varepsilon} \in O\Big) \\ \ge -\frac{1}{4} \inf_{a\in O} \Big(\Big|\int_z^m |b(x)|\, dx\Big| + \Big|\int_m^a |b(x)|\, dx\Big|\Big)^2.\end{aligned}$$

The modulus of the integrals is taken to properly handle the cases $m < z$ and $a < m$.

Proof. We want to apply Schilder's theorem and to evaluate the rate function using lemma 5.1. Let $K_1, K_2 \subseteq \mathbb{R}$ be compact. Define the process $\tilde{B}$ by setting $\tilde{B}_r = (B_{r\varepsilon} - z)/\sqrt{\varepsilon}$ for every $r > 0$. Then $\tilde{B}$ is a Brownian motion with start in 0 and we get

$$\begin{aligned}&P_z\Big(B_{t\varepsilon} \in K_2, \frac{1}{2}\int_0^{t\varepsilon} b^2(B_s)\, ds \le \varepsilon\Big) \\ &\overset{s=r\varepsilon}{=} P_z\Big(B_{t\varepsilon} \in K_2, \frac{1}{2}\int_0^{t} b^2(B_{r\varepsilon})\, dr \le 1\Big) \\ &= P\Big(\sqrt{\varepsilon}\tilde{B}_t + z \in K, \frac{1}{2}\int_0^{t} b^2(\sqrt{\varepsilon}\tilde{B}_r + z)\, dr \le 1\Big) \\ &= P\Big(\sqrt{\varepsilon}\tilde{B} \in \bigcup_{a\in K_2}\bigcup_{\beta\le 1} M_t^{a,z,\beta}\Big)\end{aligned}$$

and thus

$$\begin{aligned}\sup_{z\in K_1} P_z\Big(B_{t\varepsilon} \in K_2, \frac{1}{2}\int_0^{t\varepsilon} b^2(B_s)\, ds \le \varepsilon\Big) \\ \le P\Big(\sqrt{\varepsilon}\tilde{B} \in \bigcup_{z\in K_1}\bigcup_{a\in K_2}\bigcup_{\beta\le 1} M_t^{a,z,\beta}\Big).\end{aligned} \tag{5.2}$$

From lemma 5.2 we know that the set $\bigcup_{z\in K_1}\bigcup_{a\in K_2}\bigcup_{\beta\le 1} M_t^{a,z,\beta}$ is closed in the path space $\big(C_0[0;t], \|\cdot\|_\infty\big)$, so we can apply Schilder's theorem (theorem 2.16) to get

$$\begin{aligned}&\limsup_{\varepsilon\downarrow 0} \varepsilon \log \sup_{z\in K_1} P\Big(\sqrt{\varepsilon}\tilde{B} \in \bigcup_{z\in K_1}\bigcup_{a\in K_2}\bigcup_{\beta\le 1} M_t^{a,z,\beta}\Big) \\ &\le -\inf\Big\{I_t(\omega) \,\Big|\, \omega \in \bigcup_{z\in K_1}\bigcup_{a\in K_2}\bigcup_{\beta\le 1} M_t^{a,z,\beta}\Big\} \\ &= -\inf_{z\in K_1}\inf_{a\in K_2}\inf_{\beta\le 1}\inf\big\{I_t(\omega) \,\big|\, \omega \in M_t^{a,z,\beta}\big\}.\end{aligned}$$

First assume $m \in K_1 \cap K_2$. Define the path ω by $\omega_s = 0$ for all $s \in [0;t]$. Then clearly we have $\omega \in M_t^{m,m,0}$ for every t and since we find $I(\omega) = 0$ we have

$$\inf\Big\{I_t(\omega) \,\Big|\, \omega \in \bigcup_{\beta\in B} M_t^{a,z,\beta}\Big\} = 0$$

for all $t \geq 0$. On the other hand we have $J(m,m) = 0$.

Otherwise the evaluation of the infimum is done in lemma 5.1. Using $v(x) = b^2(x)$ we can for every $\eta > 0$ find a $t_0 > 0$, such that

$$\inf_{\beta \leq 1} \inf\bigl\{ I_t(\omega) \bigm| \omega \in M_t^{a,z,\beta} \bigr\} \geq J(a,z) - \eta$$

for all $z \in K_1$, $a \in K_2$ and $t \geq t_0$. This gives

$$\begin{aligned}
\limsup_{\varepsilon \downarrow 0} \varepsilon \log \sup_{z \in K_1} P\Bigl(\sqrt{\varepsilon}\tilde{B} \in \bigcup_{z \in K_1} \bigcup_{a \in K_2} \bigcup_{\beta \leq 1} M_t^{a,z,\beta}\Bigr) \\
&\leq - \inf_{z \in K_1} \inf_{a \in K_2} \inf_{m \in N} J(a,z) + \eta \\
&= -\frac{1}{4} \inf_{z \in K_1} \inf_{a \in K_2} \Bigl(\bigl| \int_z^m |b(x)|\, dx \bigr| + \bigl| \int_m^a |b(x)|\, dx \bigr| \Bigr)^2 + \eta
\end{aligned}$$

for every $\eta > 0$. Together with the relation (5.2) this proves the upper bound.

For the lower bound we follow the same procedure. Without loss of generality we can assume that O is bounded. Here we get

$$\begin{aligned}
P_z\Bigl(B_{t\varepsilon} \in O, \frac{1}{2} \int_0^{t\varepsilon} b^2(B_s)\, ds \leq \varepsilon\Bigr) \\
&\geq P_z\Bigl(B_{t\varepsilon} \in O, \frac{1}{2} \int_0^{t\varepsilon} b^2(B_s)\, ds < \varepsilon\Bigr) \\
&= P\Bigl(\sqrt{\varepsilon}\tilde{B} \in \bigcup_{a \in O} \bigcup_{\beta < 1} M_t^{a,z,\beta}\Bigr)
\end{aligned}$$

where the set

$$\bigcup_{a \in O} \bigcup_{\beta < 1} M_t^{a,z,\beta} = \Bigl\{ \omega \in C[0;t] \Bigm| \omega_0 = 0, \omega_t \in O - z, \frac{1}{2} \int_0^t b^2(\omega_r + z)\, dr < 1 \Bigr\}$$

is open in $\bigl(C_0[0;t], \|\cdot\|_\infty\bigr)$. So we can use the lower bound from Schilder's theorem and lemma 5.1 to finish the proof. (qed)

Corollary 5.4. *Under the assumptions of proposition 5.3 we have*

$$\begin{aligned}
\lim_{\eta \downarrow 0} \liminf_{t \to \infty} \liminf_{\varepsilon \downarrow 0} \varepsilon \log \inf_{m - \eta \leq z \leq m + \eta} P_z\Bigl(\frac{1}{2} \int_0^{t\varepsilon} b^2(B_s)\, ds \leq \varepsilon, B_{t\varepsilon} \in O\Bigr) \\
&\geq -\frac{1}{4} \inf_{a \in O} \Bigl(\int_m^a |b(x)|\, dx \Bigr)^2
\end{aligned}$$

for every open set $O \subseteq \mathbb{R}$.

Proof. For $z \in \mathbb{R}$ define

$$M_t^z = \Bigl\{ \omega \in C[0;t] \Bigm| \omega_0 = 0, \omega_t + z \in O, \frac{1}{2} \int_0^t b^2(\omega_s + z)\, ds < 1 \Bigr\}.$$

Let $\delta > 0$. Choose an $\tilde{\omega} \in M_t^m$ with $I_t(\tilde{\omega}) < \inf\{ I_t(\omega) \mid \omega \in M_t^m \} + \delta$. Because O is open and b and the integral are continuous we can find an $E > 0$, such that for every $\eta < E$ the ball $B_\eta(\tilde{\omega}) \subseteq C_0([0;t], \mathbb{R})$ is contained in the set M_t^z. This gives

$$\begin{aligned}
\liminf_{\varepsilon \downarrow 0} \varepsilon \log \inf_{m - \eta \leq z \leq m + \eta} P_z\Bigl(\frac{1}{2} \int_0^{t\varepsilon} b^2(B_s)\, ds \leq \varepsilon, B_{t\varepsilon} \in O\Bigr) \\
&= \liminf_{\varepsilon \downarrow 0} \varepsilon \log \inf_{m - \eta \leq z \leq m + \eta} P_z\Bigl(\sqrt{\varepsilon} B \in M_t^z\Bigr) \\
&\geq \liminf_{\varepsilon \downarrow 0} \varepsilon \log \inf_{m - \eta \leq z \leq m + \eta} P_z\Bigl(\sqrt{\varepsilon} B \in B_\eta(\tilde{\omega})\Bigr).
\end{aligned}$$

and using Schilder's theorem and the relation

$$-\inf\bigl\{I_t(\omega)\bigm|\omega\in B_\eta(\tilde\omega)\bigr\}\geq -I_t(\tilde\omega) > -\inf\bigl\{I_t(\omega)\bigm|\omega\in M_t^m\bigr\}-\delta$$

we find

$$\begin{aligned}\liminf_{\varepsilon\downarrow 0}\varepsilon\log\inf_{m-\eta\leq z\leq m+\eta}&P_z\Bigl(\frac12\int_0^{t\varepsilon}b^2(B_s)\,ds\leq\varepsilon, B_{t\varepsilon}\in O\Bigr)\\ &\geq -\inf\bigl\{I_t(\omega)\bigm|\omega\in B_\eta(\tilde\omega)\bigr\}\\ &> -\inf\bigl\{I_t(\omega)\bigm|\omega\in M_t^m\bigr\}-\delta.\end{aligned}$$

Now we can evaluate the infimum on the right hand side as we did in proposition 5.3. We get

$$\begin{aligned}\liminf_{t\to\infty}\liminf_{\varepsilon\downarrow 0}\varepsilon\log\inf_{m-\eta\leq z\leq m+\eta}&P_z\Bigl(\frac12\int_0^{t\varepsilon}b^2(B_s)\,ds\leq\varepsilon, B_{t\varepsilon}\in O\Bigr)\\ &\geq -\frac14\inf_{a\in O}\Bigl(\int_m^a|b(x)|\,dx\Bigr)^2-\delta\end{aligned}$$

for every $\eta < E$. Taking the limit $\delta\downarrow 0$ finishes the proof. (qed)

Before we can prove lemma 5.1, we need some preparations. For the remaining part of this section we assume throughout that v is non-negative and two times continuously differentiable and that $a, z\in\mathbb{R}$ are fixed.

Notation: For $x, y\in\mathbb{R}$ we will write $[x;y]$ for the closed interval between x and y; in the case $x<y$ this is to be read as $[y;x]$ instead.

As a first step towards the proof of lemma 5.1 we get rid of the parameter β.

Lemma 5.5. *Let $\{0\}\subset B\subseteq\mathbb{R}_+$ be bounded. Assume that*

$$\lim_{t\to\infty}\inf\bigl\{I_t(\omega)\bigm|\omega\in M_t^{a,z,1}\bigr\}=J(a,z)$$

locally uniform in $a, z\in\mathbb{R}$. Then lemma 5.1 holds.

Proof. Let $\beta>0$. For $\omega\in C_0[0;t]$ define $\tilde\omega\in C_0[0;t/\beta]$ by

$$\tilde\omega_r=\omega_{r\beta}\quad\text{for all } r\in[0;t/\beta].$$

Then we have $\tilde\omega_0=0$, $\tilde\omega_{t/\beta}=\omega_t$, and

$$\frac12\int_0^{t/\beta}v(\tilde\omega_r+z)\,dr\overset{s=r\beta}{=}\frac1\beta\cdot\frac12\int_0^t v(\omega_s+z)\,ds.$$

Thus $\omega\mapsto\tilde\omega$ is a one-to-one mapping from $M_t^{a,z,\beta}$ onto $M_{t/\beta}^{a,z,1}$.

Because of

$$I_{t/\beta}(\tilde\omega)=\frac12\int_0^{t/\beta}\dot{\tilde\omega}_r^2\,dr=\frac{\beta^2}{2}\int_0^{t/\beta}\dot\omega_{r\beta}^2\,dr\overset{s=r\beta}{=}\frac\beta2\int_0^t\dot\omega_s^2\,ds=\beta I_t(\omega)$$

we find

$$\inf\bigl\{I_t(\omega)\bigm|\omega\in M_t^{a,z,\beta}\bigr\}=\frac1\beta\inf\bigl\{I_t(\omega)\bigm|\omega\in M_{t/\beta}^{a,z,1}\bigr\}.$$

Since $m \notin K_1 \cap K_2$ every continuous path ω with $\omega_0 = 0$ and $\omega_t = a - z$ has

$$\frac{1}{2}\int_0^t v(\omega_s + z)\,ds > 0,$$

the set $M_t^{a,z,0}$ is empty in this case and we find

$$\begin{aligned}\inf\Big\{I_t(\omega)\;\Big|\;\omega \in \bigcup_{\beta\in B} M_t^{a,z,\beta}\Big\} &= \inf_{\beta\in B}\inf\big\{I_t(\omega)\;\big|\;\omega\in M_t^{a,z,\beta}\big\}\\ &= \inf_{\beta\in B}\frac{1}{\beta}\inf\big\{I_t(\omega)\;\big|\;\omega\in M_{t/\beta}^{a,z,1}\big\}.\end{aligned}$$

Now let $K_1, K_2 \subseteq \mathbb{R}$ be compact. Let $\eta > 0$ and choose a $t_0 > 0$ with

$$\Big|\inf\big\{I_t(\omega)\;\big|\;\omega\in M_t^{a,z,1}\big\} - J(a,z)\Big| < \eta \sup B$$

for all $t > t_0$, $z \in K_1$, and $a \in K_2$. Then for every $t > t_0 \sup B$ we have

$$\Big|\frac{1}{\beta}\inf\big\{I_t(\omega)\;\big|\;\omega\in M_{t/\beta}^{a,z,1}\big\} - \frac{1}{\beta}J(a,z)\Big| < \frac{\eta\cdot\sup B}{\beta}$$

and taking the infimum over all $\beta \in B$ on both sides gives

$$\Big|\inf\Big\{I_t(\omega)\;\Big|\;\omega\in\bigcup_{\beta\in B} M_t^{a,z,\beta}\Big\} - \frac{1}{\sup B}J(a,z)\Big| \le \eta$$

for all $t > t_0 \sup B$, $z \in K_1$, and $a \in K_2$. Because η was arbitrary this finishes the proof. (qed)

Because $I_t(\omega + z) = I_t(\omega)$ we can shift every path from $M_t^{a,z,1}$ by z and get

$$\inf\big\{I_t(\omega)\;\big|\;\omega\in M_t^{a,z,1}\big\} = \inf\Big\{I_t(\omega)\;\Big|\;\omega_0 = z, \omega_t = a, \frac{1}{2}\int_0^t v(\omega_s)\,ds = \beta\Big\}.$$

For the moment assume that there is a path $\tilde\omega$ with $I_t(\tilde\omega) = \inf\big\{I_t(\omega)\;\big|\;\omega\in M_t^{a,z,1}\big\}$. Later we will show, that such an $\tilde\omega$ in fact does exist. In order to evaluate the rate function I_t for this path $\tilde\omega$, we solve the Euler-Lagrange␣equations (see section 12 of [GF63]) for extremal values of I_t under the constraint

$$K(\omega) = \frac{1}{2}\int_0^t v(\omega_s)\,ds \overset{!}{=} 1$$

and with the boundary conditions

$$\omega_0 = z \quad\text{and}\quad \omega_t = a.$$

Because $v \in C^2(\mathbb{R})$ we can use theorem 1 from section 12.1 of [GF63] to find, that for every extremal point ω of I under the given constraints there is a constant λ, such that ω solves the equations

$$\ddot\omega_s = \lambda v'(\omega_s + z) \quad \text{for all } s \in (0;t], \text{ and} \quad \omega_0 = z \tag{5.3a}$$

$$\frac{1}{2}\int_0^t v(\omega_s)\,ds = 1 \tag{5.3b}$$

$$\omega_t = a. \tag{5.3c}$$

Existence of solutions: the autonomous second order equation (5.3a) describes the motion of a classical particle on the real line in the potential $-\lambda v$. The differential equation can be reduced to an autonomous first order equation in the plane with the usual trick: defining $x(s) = (\omega_s, \dot\omega_s)$ and $F(x_1, x_2) = \bigl(x_2, \lambda v'(x_1)\bigr)$ the equation becomes

$$\dot x(s) = F(x(s)) \quad \text{for all } s \in [0;t].$$

See e.g. section 5.3 of [BR89] for details. Because v' and thus F is locally Lipschitz continuous, for every pair $\omega_0 = z$, $\dot\omega_0 = v_0$ of initial conditions and every bounded region we find a unique solution of the ODE at least up to the boundary of that region (see theorem 8 in section 6.9 of [BR89]).

There are two degrees of freedom in (5.3a) because we can choose $\dot\omega_0$ and λ. In the following we will show, that the two additional conditions (5.3b) and (5.3c) guarantee the existence of a unique solution to the system (5.3).

The interpretation as the motion of a classical particle helps us to determine the behaviour of the solutions. We can use conservation of energy: Because of

$$\partial_s\bigl(\frac{1}{2}\dot\omega_s^2 - \lambda v(\omega_s)\bigr) = \dot\omega_s\ddot\omega_s - \lambda v'(\omega_s)\dot\omega_s = \dot\omega_s\bigl(\ddot\omega_s - \lambda v'(\omega_s)\bigr) \overset{(5.3a)}{=} 0$$

we have

$$\frac{1}{2}\dot\omega_s^2 - \lambda v(\omega_s) \;=\; \frac{1}{2}\dot\omega_0^2 - \lambda v(\omega_0) \;=:\; E \quad \text{for all } s \in [0;t]. \tag{5.4}$$

This conservation law describes the speed for any point of the path: the speed of the path at point ω_s is

$$|\dot\omega_s| = \sqrt{2(E + \lambda v(\omega_s))}. \tag{5.5}$$

Thus the rate function I_t can be expressed as a function of E and λ as follows.

$$\begin{aligned} I_t(\omega) &= \frac{1}{2}\int_0^t \dot\omega_s^2\,ds = \int_0^t E + \lambda v(\omega_s)\,ds \\ &= tE + 2\lambda, \end{aligned} \tag{5.6}$$

where λ and E are determined by equations (5.3b) and (5.3c).

Because of relation (5.4) we find, that whenever ω is a solution of (5.3a) we have $E \geq -\lambda v(\omega_s)$ for all $s \in [0;t]$ and the path can only stop and turn at points x with $-\lambda v(x) = E$. Let $x \in \mathbb{R}$ be such a point and assume $v'(x) = 0$. Then η with $\eta_s = x$ for all $s \geq 0$ is the unique solution of (5.3a) with $\eta_0 = x$ and $\dot\eta_0 = 0$. Now assume that $\omega_s = x$ for some $s > 0$. Then $(\omega_{s-r})_{r\in[0;s]}$ is also a solution of (5.3a) with start in x and initial speed 0, so we have $\omega_{s-r} = \eta_r = x$ for all $r \in [0;s]$. This shows that a point $x \neq z$ with $E = -\lambda v(x)$ and $v'(x) = 0$ cannot be reached by a solution ω of (5.3a). Thus whenever a non-constant path reaches an $x \in \mathbb{R}$ with $E = -\lambda v(x)$ then we have $\ddot\omega_s = \lambda v'(\omega_s) \neq 0$ and the path always changes direction there. Figure 5.1 illustrates two different kinds of solution, one where ω_s moves monotonically and one where the path reaches a point b with $-\lambda v(b) = E$ and turns there.

Since the differential equation (5.3a) is autonomous and since a solution ω changes direction every time is reaches a point x with $-\lambda v(x) = E$, the path can reach at most two distinct points of these nature. In this case the solution oscillates between these points periodically. Thus every solution of (5.3a) changes direction only a finite number of times before time t.

In order to find the path which minimises the rate function I_t we need to keep track of the different possible traces of the path. For the remaining part of this section we use the following notation. The path $(\omega_s)_{0\leq s\leq t}$ is said to have **trace** $T = (x_0, x_1, \ldots, x_n)$ when $\omega_0 = x_0$, $\omega_t = x_n$, and the path ω moves monotonically in either direction from x_{i-1} to x_i for $i = 1, \ldots, n$ in order and changes direction only at the points $x_1, \ldots, x_{n-1}$. We use the abbreviation

$$|T| = \sum_{i=1}^n |x_i - x_{i-1}|$$

Figure 5.1: *This figure illustrates two types of solution for equation* (5.3a)*. Here we only consider the case* $\lambda > 0$*. The curved line is the graph of the function* $x \mapsto -\lambda v(x)$*. The bold part of the lines corresponds to the points visited by the path* ω*. The thick dots are* $\big(\omega_0, -\lambda v(\omega_0)\big)$ *and* $\big(\omega_t, -\lambda v(\omega_t)\big)$*. Both solutions start at* $z \in K_1$*, head towards a neighbourhood of the zero* m*, and finally reach a point* $a \in K_2$*. The left hand image shows a free solution, i.e. one with* $E > 0$*, the right hand image shows a bound solution, i.e. one with* $E \leq 0$ *where the path* ω *turns at the point* b *with* $-\lambda v(b) = E$*.*

for the length of the trace and sometimes identify T with the set $\bigcup_{i=1}^n [x_{i-1}, x_i]$ of covered points to write $\min T$, $\max T$, $v|_T$, or $\inf_{x \in T} v(x)$. For positive functions $f\colon \mathbb{R} \to \mathbb{R}$ we use the notation

$$\int_T f(x)\,dx := \sum_{i=1}^n \Big|\int_{x_{i-1}}^{x_i} f(x)\,dx\Big|.$$

The absolute values are taken to make the integral positive even when $x_i < x_{i-1}$. If a solution ω of (5.3a) has trace $T = (x_0, x_1, \dots, x_n)$, this then implies that $v(x_1) = \cdots = v(x_{n-1}) = -E/\lambda$ and each of the $x_1, \dots, x_{n-1}$ is either $\min T$ or $\max T$. Between the points x_i the path is strictly monotonic, i.e. after the start in z it oscillates zero or more times between $\min T$ and $\max T$ before it reaches a at time t. Using this notation we can formulate the following Lemma.

Lemma 5.6. *Let* $\lambda, E \in \mathbb{R}$ *and a trace* $T = (x_0, \dots, x_n)$ *be given. Then the following two conditions are equivalent.*

(j) The unique solution $\omega\colon [0;t] \to \mathbb{R}$ *of*

$$\ddot\omega_s = \lambda v'(\omega_s) \quad \text{for all } s \in [0;t]$$

with initial conditions $\omega_0 = z$ *and* $\dot\omega_0 = \operatorname{sgn}(x_1 - x_0)\sqrt{2(E + \lambda v(0))}$ *has trace* T *and solves* (5.3b) *and* (5.3c)*.*

(ij) We have $x_0 = z$*,* $x_n = a$*,* $E = -\lambda v(x_i)$ *for* $i = 1, \dots, n-1$*, as well as* $E > -\lambda v(x)$ *for all* $\min T < x < \max T$*, and the pair* (λ, E) *solves*

$$\int_T \frac{v(x)}{\sqrt{E + \lambda v(x)}}\,dx = \sqrt{8} \tag{5.7a}$$

and

$$\int_T \frac{1}{\sqrt{E + \lambda v(x)}}\,dx = \sqrt{2}t. \tag{5.7b}$$

Proof. Assume the conditions from (*j*). Then ω is a solution of (5.3a), there are times $t_0, t_1, \dots, t_n$ with $\omega_{t_i} = x_i$ for $i = 0, \dots, n$ and between the times t_i the process moves monotonically. For any integrable, positive function $g\colon \mathbb{R} \to \mathbb{R}$ substitution using (5.5) yields

$$\begin{aligned}\int_0^t g(\omega_s)\,ds &= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} g(\omega_s)\,ds \\ &= \sum_{i=1}^n \int_{x_{i-1}}^{x_i} g(x)\,\frac{dx}{\operatorname{sgn}(x_i - x_{i-1})\sqrt{2(E+\lambda v(x))}} \\ &= \int_T \frac{g(x)}{\sqrt{2(E+\lambda v(x))}}\,dx. \end{aligned} \tag{5.8}$$

Applying (5.8) to the function $g = v$ gives

$$1 \overset{(5.3b)}{=} \frac{1}{2}\int_0^t v(\omega_s)\,ds \overset{(5.8)}{=} \frac{1}{\sqrt{8}}\int_T \frac{v(x)}{\sqrt{E+\lambda v(x)}}\,dx.$$

This is equation (5.7a). Applying (5.8) to the constant function $g = 1$ gives

$$t = \int_0^t 1\,ds \overset{(5.8)}{=} \frac{1}{\sqrt{2}}\int_0^a \frac{1}{\sqrt{E+\lambda v(x)}}\,dx,$$

which is equation (5.7b).

Now assume condition (ij). For $i = 1, \ldots, n$ define the function F_i by

$$F_i(x) = \frac{1}{\sqrt{2}}\Big|\int_{x_{i-1}}^x \frac{1}{\sqrt{E+\lambda v(x)}}\,dx\Big|$$

for all x between x_{i-1} and x_i. Then F_i is finite because of (5.7b), strictly monotonic (increasing if $x_i > x_{i-1}$ and decreasing else), and has $F_i(x_{i-1}) = 0$. Further define

$$t_k = \sum_{i=1}^k F_i(x_i).$$

Equation (5.7b) gives $t_n = t$. Because the functions F_i are monotonic they have inverse functions F_i^{-1} and we can define $\omega\colon [0;t] \to \mathbb{R}$ by

$$\omega(s) = F_i^{-1}(s - t_{i-1}) \quad \text{for all } s \in [t_{i-1}, t_i].$$

We will prove, that ω satisfies all the conditions from (j).

Because we have $t_i - t_{i-1} = F_i(x_i)$ and thus $F_i^{-1}(t_i - t_{i-1}) = x_i = F_{i+1}^{-1}(t_i - t_i)$ the function ω is well-defined on the connection points at times t_i and is continuous. This also shows $\omega_{t_i} = x_i$ for $i = 0, 1, \ldots, n$ and especially $\omega_0 = x_0 = z$ and $\omega_t = x_n = a$.

Because the F_i are differentiable at all points x strictly between x_{i-1} and x_i, the function ω is differentiable on the intervals $(t_{i-1}; t_i)$ with derivative

$$\dot\omega_s = \frac{1}{F_i'(\omega_s)} = \operatorname{sgn}(x_i - x_{i-1})\sqrt{2(E+\lambda v(\omega_s))}.$$

Because ω is continuous and the limits $\lim_{s\to t_i}\dot\omega_s$ exist, we see that ω is even differentiable on $[0;t]$ with $\dot\omega_0 = \operatorname{sgn}(x_1 - x_0)\sqrt{2(E+\lambda v(0))}$ and $\dot\omega_{t_i} = 0$ for $i = 1, \ldots, n-1$.

Using the same kind of argument again, we find

$$\ddot\omega_s = \frac{\operatorname{sgn}(x_i - x_{i-1})}{2\sqrt{2(E-\lambda v(\omega_s))}} \cdot 2\lambda v'(\omega_s) \cdot \operatorname{sgn}(x_i - x_{i-1})\sqrt{2(E-\lambda v(\omega_s))} = \lambda v'(\omega_s),$$

first between the t_i and then on the whole interval $[0;t]$. Thus ω really solves the differential equation from (j).

Using the substitution

$$\frac{1}{2}\int_0^t v(\omega_s)\,ds = \frac{1}{\sqrt{8}}\int_T \frac{v(x)}{\sqrt{E+\lambda v(x)}}\,dx$$

as in the first part, we also get back (5.3b) from (5.7a). (qed)

Figure 5.2: *This figure illustrates the domain H_T of the functions f and g. The domain is unbounded in directions $\lambda \to \infty$ and $E \to \infty$. It is bounded from below by $\lambda \mapsto -\inf_{x\in T} \lambda v(x)$, which is equal to $-\lambda \sup_{x\in T} v(x)$ for $\lambda \leq 0$ and to $-\lambda \inf_{x\in T} v(x)$ for $\lambda \geq 0$.*

Now we have reduced the problem of minimising $I_t(\omega)$ for solutions ω of the system (5.3) to the problem of minimising

$$I_t(E, \lambda) = tE + 2\lambda$$

for solutions (E, λ) of the system (5.7).

For a trace T define

$$H_T = \big\{(E, \lambda) \bigm| E \geq -\inf_{x\in T} \lambda v(x)\big\} \subseteq \mathbb{R}^2$$

and furthermore define the functions $f, g\colon H_t \to [0; \infty]$ by

$$f(E, \lambda) = \int_T \frac{1}{\sqrt{E + \lambda v(x)}}\, dx$$

and

$$g(E, \lambda) = \int_T \frac{v(x)}{\sqrt{E + \lambda v(x)}}\, dx.$$

Figure 5.2 illustrates the domain H_T. Both functions are finite in the interior of the domain, but can be infinite at the boundary. The equations (5.7) are equivalent to $f(E_\lambda, \lambda) = \sqrt{2}t$ and $g(E, \lambda) = \sqrt{8}$. For paths which change direction at some point we will find solutions (E, λ) of (5.7), which lay on the boundary of H_T. For paths which go straight from z to a we will find solutions (E, λ) in the interior of H_T.

Lemma 5.7. *Let $t > 0$ and T be a trace from $z \in \mathbb{R}$ to $a \in \mathbb{R}$ such that $v|_T$ is not constant. Then there is at most one solution (E, λ) of (5.7).*

Proof. For $E > -\inf_{x\in T} \lambda v(x)$ we can choose an E_* between $-\inf_{x\in T} \lambda v(x)$ and E. Then $v(x)/(E_* + \lambda v(x))^{3/2}$ is an integrable upper bound of $v(x)/(e + \lambda v(x))^{3/2}$ for all e in a $(E - E_*)$-Neighbourhood of E. So we can use the theorem about interchanging the Lebesgue-integral with derivatives to get

$$\frac{\partial}{\partial E} g(E, \lambda) = -\frac{1}{2} \int_T \frac{v(x)}{\big(E + \lambda v(x)\big)^{3/2}}\, dx < 0.$$

So for every λ the map $E \mapsto g(E, \lambda)$ is strictly decreasing and there can be at most one E_λ with $g\big(E_\lambda, \lambda\big) = \sqrt{8}$.

With the help of the implicit function theorem we can calculate the derivative of E_λ. Interchanging the integral with the derivative as above we get

$$\begin{aligned}\frac{\partial}{\partial\lambda}E_\lambda &= -\frac{\frac{\partial}{\partial\lambda}g(E_\lambda,\lambda)}{\frac{\partial}{\partial E}g(E_\lambda,\lambda)} \\ &= -\frac{(-\frac{1}{2})\int_T v^2(x)\big(E_\lambda+\lambda v(x)\big)^{-3/2}\,dx}{(-\frac{1}{2})\int_T v(x)\big(E_\lambda+\lambda v(x)\big)^{-3/2}\,dx} \\ &= -\frac{\int_T v^2(x)\,d\mu(x)}{\int_T v(x)\,d\mu(x)}\end{aligned}$$

where μ is the probability measure, with density

$$\frac{d\mu}{dx} = \frac{1}{Z}\big(E_\lambda+\lambda v(x)\big)^{-3/2}$$

and the normalisation constant is

$$Z = \int_T \big(E_\lambda+\lambda v(y)\big)^{-3/2}\,dy.$$

Furthermore for $(E,\lambda)\in(H_T)^\circ$ we have

$$\frac{\partial}{\partial E}f(E,\lambda) = -\frac{1}{2}\int_T\big(E+\lambda v(x)\big)^{-3/2}\,dx = -\frac{Z}{2}$$

and thus

$$\begin{aligned}\frac{\partial}{\partial\lambda}\big(f(E_\lambda,\lambda)\big) &= \frac{\partial f}{\partial E}(E_\lambda,\lambda)\cdot\frac{\partial}{\partial\lambda}E_\lambda + \frac{\partial f}{\partial\lambda}(E_\lambda,\lambda) \\ &= \frac{Z}{2}\cdot\frac{\int_T v^2(x)\,d\mu(x)}{\int_T v(x)\,d\mu(x)} - \frac{Z}{2}\int_T v(x)\,d\mu(x) \\ &= \frac{Z}{2}\cdot\frac{\int_T v^2(x)\,d\mu(x) - \big(\int_T v(x)\,d\mu(x)\big)^2}{\int_T v(x)\,d\mu(x)} \\ &\geq 0.\end{aligned}$$

Equality would only hold for the case of constant $v|_T$. So the map $\lambda\mapsto f(E_\lambda,\lambda)$ is strictly increasing and there can be at most one λ with $f\big(E_\lambda,\lambda\big)=\sqrt{2}t$. This finishes the proof.

(qed)

Lemma 5.8. *Let T a trace with $m\in T$ and $t\geq 2|T|/\int_T v(x)\,dx$. Then equation (5.7) has a solution (E,λ) with with $E,\lambda>0$.*

Proof. Define $\lambda^* = (\int_T\sqrt{v(x)}\,dx)^2/8$ and assume $0<\lambda\leq\lambda^*$. Then we have

$$g(0,\lambda) = \int_T\frac{v(x)}{\sqrt{\lambda v(x)}}\,dx = \frac{1}{\sqrt{\lambda}}\int_T\sqrt{v(x)}\,dx \geq \sqrt{8}.$$

and the dominated convergence theorem gives

$$\lim_{E\to\infty}g(E,\lambda)=0.$$

Thus for all $0<\lambda\leq\lambda^*$ there exists an $E_\lambda\geq 0$ with $g(E_\lambda,\lambda)=\sqrt{8}$.

Because of $g(0,\lambda^*)=\sqrt{8}$ we have $E_{\lambda^*}=0$. Fatou's lemma then gives

$$\liminf_{\lambda\uparrow\lambda^*}f(E_\lambda,\lambda)\geq\int_T\frac{1}{\sqrt{\lambda^* v(x)}}\,dx.$$

Because v is positive and $v(m) = 0$ we have $v'(m) = 0$ and $v''(m) \geq 0$. Then by Taylor's theorem there exists a $c > 0$ and a closed interval $I \subseteq \mathbb{R}$ with $m \in I \subseteq T$, such that $v(x) \leq c^2(x-m)^2$ for all $x \in I$. Therefore we find

$$\int_T \frac{1}{\sqrt{v(x)}}\,dx \geq \int_I \frac{1}{\sqrt{c^2(x-m)^2}}\,dx = \int_I \frac{1}{c|x-m|}\,dx = +\infty$$

and thus $\lambda \mapsto f(E_\lambda, \lambda)$ is a continuous function with

$$\lim_{\lambda \uparrow \lambda^*} f(E_\lambda, \lambda) = +\infty.$$

On the other hand because of $g(E_0, 0) = \sqrt{8}$ we have $E_0 = (\int_T v(x)\,dx)^2/8$. So for $\lambda = 0$ we get

$$f(E_0, 0) = \int_T \frac{1}{\sqrt{E_0}}\,dx = \frac{\sqrt{8}}{\int_T v(x)\,dx}|T|.$$

Together this shows, that for all

$$t \geq \frac{2|T|}{\int_T v(x)\,dx}$$

there exists a solution (E_λ, λ) with $f(E_\lambda, \lambda) = \sqrt{2}t$. (qed)

Lemma 5.9. *There are numbers $\varepsilon, c_1, c_2 > 0$, such that the following holds: For every trace T starting in K_1, ending in K_2, and visiting the ball $B_\varepsilon(m)$ there is a non-empty, closed interval $A \subseteq \mathbb{R}$, such that $A \subseteq T$, $|A| = \varepsilon$ and we have $c_1 \leq v(x) \leq c_2$ for every $x \in A$.*

Proof. Because $m \notin K_1 \cap K_2$ either K_1 or K_2 has a positive distance from m. Let ε be one third of this distance. Define $A' = \{ x \in \mathbb{R} \mid \varepsilon \leq |x-m| \leq 2\varepsilon \}$ and let $c_1 = \inf\{ v(x) \mid x \in A' \}$ and $c_2 = \sup\{ v(x) \mid x \in A' \}$.

Each trace starting in K_1, ending in K_2, and visiting the ball $B_\varepsilon(m)$ either crosses $[m - 2\varepsilon; m-\varepsilon]$ or $[m+\varepsilon; m+2\varepsilon]$. Let A be the crossed interval. Then clearly $|A| = \varepsilon$ and and because of $A \subseteq A'$ the estimates for v on A hold. (qed)

Lemma 5.10. *For every $\eta > 0$ there is a $t_1 > 0$, such that whenever $t \geq t_1$, T is a trace from $z \in K_1$ to $a \in K_2$ with $m \in [z; a]$, and (E, λ) solves (5.7), then we have*

$$\Big| I_t(E, \lambda) - \frac{1}{4}\Big(\int_T \sqrt{v(x)}\,dx\Big)^2 \Big| \leq \eta.$$

Proof. This case is illustrated in the left hand image of figure 5.1. Because $m \in [z; a]$ any path from z to a must visit m and thus we find $E > -\lambda v(m) = 0$. Thus the only possible trace in this case is $T = (z, a)$, because the process could only turn at points x where $-\lambda v(x) = E$.

Now let $\eta > 0$. Define $L = \sup\big\{ |a - z| \bigm| z \in K_1, a \in K_2 \big\}$. Then we get

$$\sqrt{2}t = \int_T \frac{1}{\sqrt{E + \lambda v(x)}}\,dx \leq \int_z^a \frac{1}{\sqrt{E}}\,dx \leq \frac{L}{\sqrt{E}}$$

and thus

$$E \leq \frac{L^2}{2t^2}.$$

So we can find a $t_1 > 0$ with

$$E \cdot t < \eta \tag{5.9}$$

whenever $t \geq t_1$.

Choosing A, c_1, and c_2 as in lemma 5.9 we get

$$\sqrt{8} = \int_T \frac{v(x)}{\sqrt{E + \lambda v(x)}}\,dx \geq \int_A \frac{c_1}{\sqrt{E + \lambda c_2}}\,dx = \frac{c_1 |A|}{\sqrt{E + \lambda c_2}}$$

and thus

$$\lambda \geq \frac{c_1^2|A|^2 - E}{8c_2} \geq \frac{c_1^2|A|^2 - L^2/2t^2}{8c_2}.$$

So we can choose a small $c_3 > 0$ and increase t_1 to achieve $\lambda > c_3$ whenever $t \geq t_1$.

Because of

$$\lim_{E \downarrow 0} \int_0^a \frac{v(x)}{\sqrt{E + v(x)}}\, dx = \int_0^a \sqrt{v(x)}\, dx$$

we can find a $c_4 > 0$ with

$$\int_0^a \frac{v(x)}{\sqrt{E + v(x)}}\, dx \geq \sqrt{1 - \eta/J(z,a)} \int_0^a \sqrt{v(x)}\, dx$$

for all $E < c_4$. Increase t_1 until

$$\frac{L^2}{2t^2 c_3} < c_4$$

and thus

$$\begin{aligned}
\sqrt{8} &= \int_0^a \frac{v(x)}{\sqrt{E + \lambda v(x)}}\, dx \\
&\geq \frac{1}{\sqrt{\lambda}} \int_0^a \frac{v(x)}{\sqrt{L^2/2t^2\lambda + v(x)}}\, dx \\
&\geq \frac{1}{\sqrt{\lambda}} \int_0^a \frac{v(x)}{\sqrt{c_4 + v(x)}}\, dx \\
&\geq \frac{1}{\sqrt{\lambda}} \sqrt{1 - \eta/J(z,a)} \int_0^a \sqrt{v(x)}\, dx
\end{aligned}$$

for all $t \geq t_1$. Solving this for λ we get

$$2\lambda \geq (1 - \eta/J(z,a))J(z,a) = J(z,a) - \eta. \tag{5.10}$$

Because E is positive we also find

$$\sqrt{8} = \int_0^a \frac{v(x)}{\sqrt{E + \lambda v(x)}}\, dx \leq \frac{1}{\sqrt{\lambda}} \int_0^a \sqrt{v(x)}\, dx$$

and thus

$$2\lambda \leq J(z,a). \tag{5.11}$$

For the rate function I_t equation (5.10) gives

$$I_t(E,\lambda) = E \cdot t + 2\lambda \geq J(z,a) - \eta$$

and equations (5.9) and (5.11) give

$$I_t(E,\lambda) = E \cdot t + 2\lambda \leq J(z,a) + \eta$$

for all $t > t_1$. (qed)

Lemma 5.11. *For every $\eta > 0$ there is a $t_2 > 0$, such that whenever $t \geq t_2$, T is a trace from $z \in K_1$ to $a \in K_2$ with $m \notin [z;a]$, and (E,λ) solves (5.7), then we have*

$$\left| I_t(E,\lambda) - \frac{1}{4}\Big(\int_T \sqrt{v(x)}\, dx\Big)^2 \right| \leq \eta.$$

Proof. This case is illustrated in the right hand image of figure 5.1. Because the path has to change direction we will have $E < 0$ in this case. Without loss of generality we can assume that $m < a, z$. We call a value $b \in \mathbb{R}$ admissible if it lies in the interval $(m; \min(a, z))$ and if additionally $v(x) > v(b)$ for all $x > b$ holds. For admissible values b consider the trace $T = (z, b, a)$ and define

$$h_{z,a}(b) = 2\frac{\int_{(z,b,a)} \frac{1}{\sqrt{v(x)-v(b)}}\,dx}{\int_{(z,b,a)} \frac{v(x)}{\sqrt{v(x)-v(b)}}\,dx}.$$

Using Taylor approximation as in lemma 5.8, one sees that for $b \to m$ the numerator converges to $+\infty$ and by dominated convergence the denominator converges to $\int_{(0,m,a)} \sqrt{v(x)}\,dx$. So h is a continuous function with $h_{z,a}(b) \to \infty$ for $b \to m$.

Let ε, c_1, and c_2 and A be as in lemma 5.9. We would like to find a $b \in B_\varepsilon(m)$ with $h_{z,a}(b) = t$, so we need an upper bound on

$$\inf_{b\in(m;m+\varepsilon)} h_{a,z}(b) \tag{5.12}$$

which is uniform in a and z. We find

$$h_{z,a}(b) \leq 2\frac{\sup_{z\in K_1, a\in K_2} \int_{(z,b,a)} \frac{1}{\sqrt{v(x)-v(b)}}\,dx}{\int_A \frac{c_1}{\sqrt{c_2}}\,dx}. \tag{5.13}$$

Because $v''(m) > 0$ and $\liminf_{|x|\to\infty} v(x) > 0$ we can decrease ε to ensure that $v'(x) \geq v''(m)(x-m)/2$ for all $x \in [m; m+\varepsilon]$ and $v(x) \geq v(m+\varepsilon)$ for all $x \geq m+\varepsilon$. Using Taylor's theorem again we get

$$v(x) - v(b) = v'(\xi)(x-b) \geq \frac{v''(m)(b-m)}{2}(x-b)$$

for some $\xi \in [b; x]$ for all $x \in [m; m+\varepsilon]$. Thus we can conclude

$$\begin{aligned}
&\int_{(z,b,a)} \frac{1}{\sqrt{v(x)-v(b)}}\,dx \\
&\quad\leq 2\int_b^{m+\varepsilon} \frac{1}{\sqrt{\frac{v''(m)(b-m)}{2}(x-b)}}\,dx \\
&\qquad + \int_{m+\varepsilon}^{z} \frac{1}{\sqrt{v(m+\varepsilon)-v(b)}}\,dx + \int_{m+\varepsilon}^{a} \frac{1}{\sqrt{v(m+\varepsilon)-v(b)}}\,dx \\
&\quad\leq 2\sqrt{\frac{2}{v''(m)(b-m)}} \cdot \sqrt{m+\varepsilon-b} \\
&\qquad + 2\frac{1}{\sqrt{v(m+\varepsilon)-v(b)}} \sup\bigl\{|x-m| \bigm| x \in K_1 \cup K_2\bigr\}.
\end{aligned} \tag{5.14}$$

The right hand side of (5.14) is independent of a and z. So we can take the infimum over all $b \in (m; m+\varepsilon)$ and use (5.13) to get the uniform upper bound on (5.12). Call this bound t_2.

Now let $t > t_2$. Then for every $z \in K_1$ and $a \in K_2$ we can find a $b \in (m; m+\varepsilon)$ with $h_{z,a}(b) = t$. Further define $\lambda > 0$ by

$$\sqrt{\lambda} = \frac{1}{\sqrt{8}} \int_{(z,b,a)} \frac{v(x)}{\sqrt{v(x)-v(b)}}\,dx$$

and E by

$$E = -\lambda v(b).$$

Then for the trace $T = (z, b, a)$ these values E and λ solve

$$E + \lambda v(b) = 0,$$

$$\int_{(z,b,a)} \frac{v(x)}{\sqrt{E + \lambda v(x)}}\, dx = \frac{1}{\sqrt{\lambda}} \int_{(z,b,a)} \frac{v(x)}{\sqrt{v(x) - v(b)}}\, dx = \sqrt{8}$$

and

$$\int_{(z,b,a)} \frac{1}{\sqrt{E + \lambda v(x)}}\, dx = \frac{1}{\sqrt{\lambda}} \int_{(z,b,a)} \frac{1}{\sqrt{v(x) - v(b)}}\, dx = \sqrt{2}t.$$

For $t \to \infty$ we have $b \to m$ uniformly in a and z,

$$\lambda \to \frac{1}{8}\Big(\int_{(z,m,a)} \frac{v(x)}{\sqrt{v(x) - v(m)}}\, dx\Big)^2 = \frac{1}{8}\Big(\int_{(z,m,a)} \sqrt{v(x)}\, dx\Big)^2,$$

and again $E \to 0$ (this time from below). This gives

$$I_t(E, \lambda) = \frac{1}{2} \int_T \sqrt{2(E + \lambda v(x))}\, dx \to \frac{1}{4}\Big(\int_{(z,m,a)} \sqrt{v(x)}\, dx\Big)^2$$

which proves the lemma. (qed)

With all these preparations in place we are now ready to calculate the asymptotic lower bound from lemma 5.1.

Proof (of lemma 5.1). Because of lemma 5.5 we can restrict ourselves to the case $\beta = 1$, i.e. we have to prove

$$\lim_{t\to\infty} \inf\big\{I_t(\omega) \bigm| \omega \in M_t^{a,z,1}\big\} = J(a, z)$$

locally uniformly in $a, z \in \mathbb{R}$.

Let $K_1, K_2 \subseteq \mathbb{R}$ be compact with $0 \notin K_1 \cap K_2$ and $\eta > 0$. Furthermore let $z \in K_1$ and $a \in K_2$.

Assume first the case $m \in [z; a]$. From lemma 5.10 we get a $t_0 > 0$, such that for every $t > t_0$ there exists a solution (E, λ) of (5.7) for the trace $T = (z, a)$ with $\big|I_t(E, \lambda) - J(a, z)\big| \leq \eta$. This t_0 only depends on K_1 and K_2, but not on z and a.

Now assume the case $m \notin [z; a]$. From lemma 5.11 we again get a $t_0 > 0$, such that for every $t > t_0$ there exists a solution (E, λ) of (5.7) for a trace $T = (z, x_1, a)$ with $\big|I_t(E, \lambda) - J(a, z)\big| \leq \eta$ and t_0 only depends on K_1 and K_2, but not on z and a.

In either case we can use lemma 5.6 to conclude, that there exists an ω, which solves (5.3a), (5.3b), and (5.3c). Because of (5.6) this path has

$$\big|I_t(\omega) - J(a, z)\big| \leq \eta.$$

Let $c = \inf\big\{I_t(\omega) \bigm| \omega \in M_t^{a,z,1}\big\}$. Because the path ω constructed just now is both, in $M_t^{a,z,1}$ and absolutely continuous, we have $c < \infty$. Let $M_n = M_t^{a,z,1} \cap \{\,\omega \mid I_t(\omega) < c + 1/n\,\}$. Because $M_t^{a,z,1}$ is closed and I_t is a good rate function, the sets M_n are compact, non-empty, and satisfy $M_n \supseteq M_{n+1}$ for every $n \in \mathbb{N}$. So the intersection $M = \bigcap_{n\in\mathbb{N}} M_n$ is again non-empty. Because every $\tilde\omega \in M$ has $I_t(\tilde\omega) = c$, we see, that there in fact exists a path $\tilde\omega$ for which the infimum is attained. From the Euler-Lagrange method we know, that $\tilde\omega$ also solves equations (5.3a), (5.3b), and (5.3c). From lemmas 5.6 and 5.7 we know, that the solution is unique, so $\tilde\omega$ must coincide with our path ω constructed above and we get

$$\Big|\inf\big\{I_t(\omega) \bigm| \omega \in M_t^{a,z,1}\big\} - J(a, z)\Big| \leq \eta$$

for all $z \in K_1$, $a \in K_2$ and $t \geq t_0$. Since $\eta > 0$ was arbitrary this completes the proof. (qed)

5.2 Staying Near the Equilibrium

In this section we want to study the event that for some drift function $b\colon \mathbb{R} \to \mathbb{R}$ the integral $\frac{1}{2}\int_0^t b^2(B_s)\,ds$ is small. In contrast to the previous section here we are considering long time intervals, but have no conditions on the final point.

The proof uses Taylor approximation around the zeros of b to reduce the problem to the case of linear b which was already studied in Corollary 4.3. In order to make the Taylor approximation work we have need upper bounds on the probability that the process leaves a neighbourhood of the zero of b. This is given by the following lemma.

Lemma 5.12. *Let B be a Brownian motion, $a, t > 0$, and $v\colon \mathbb{R} \to \mathbb{R}$ be a function with $v(x) \geq x^2 \wedge a^2$ for every $x \in \mathbb{R}$. Then we have*

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log \sup_{x\in\mathbb{R}} P_x\Big(\int_0^t v(B_s)\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| > a\Big) \leq -\frac{1}{8}\Big(t + \frac{1}{2}a^2\Big)^2.$$

In the following lemmas we will only need the fact that the lim sup from the lemma is strictly smaller than $-t^2/8$. At a first glance it seems clear that this is true: because we are considering small values of ε, the event $\{\int_0^t v(B_s)\,ds \leq \varepsilon\}$ forces the process to spend most of the time near 0 and so the event should typically occur together with $\{\sup_{0\leq s\leq t} |B_s| \leq a\}$ but not with $\{\sup_{0\leq s\leq t} |B_s| > a\}$. So recalling corollary 4.3 one would guess that that

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \varepsilon \log P\Big(\int_0^t v(B_s)\,ds \leq \varepsilon\Big) \\
&\qquad = \limsup_{\varepsilon\downarrow 0} \varepsilon \log P\Big(\int_0^t v(B_s)\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| \leq a\Big) \\
&\qquad \leq \limsup_{\varepsilon\downarrow 0} \varepsilon \log P\Big(\int_0^t B_s^2\,ds \leq \varepsilon\Big) \\
&\qquad = -\frac{1}{8}t^2
\end{aligned}$$

and that the condition $\sup_{0\leq s\leq t} |B_s| > a$ will cause an additional cost which makes this rate smaller than $-t^2/8$. Converting this idea into a formal proof turns out to be cumbersome and we defer the proof to the end of this section.

A first consequence of lemma 5.12 is the following statement.

Lemma 5.13. *For every $a > 0$ and every $x \in (-a/\sqrt{2}; +a/\sqrt{2})$ we have*

$$\begin{aligned}
&\lim_{\varepsilon\downarrow 0} \varepsilon \log P_x\Big(\int_0^t B_s^2\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| \leq a\Big) \\
&\qquad = \lim_{\varepsilon\downarrow 0} \varepsilon \log P_x\Big(\int_0^t B_s^2\,ds \leq \varepsilon\Big) = -\frac{\big(t + x^2\big)^2}{8}.
\end{aligned}$$

Proof. The second equality is proved in corollary 4.3. Applying lemma 5.12 to the function $v(x) = x^2$ we see that

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \varepsilon \log P_x\Big(\int_0^t B_s^2\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| > a\Big) \\
&\qquad \leq -\frac{1}{8}\Big(t + \frac{1}{2}a^2\Big)^2 \\
&\qquad < \liminf_{\varepsilon\downarrow 0} \varepsilon \log P_x\Big(\int_0^t B_s^2\,ds \leq \varepsilon\Big).
\end{aligned}$$

Thus we can use lemma 2.3 to prove the first equality. (qed)

When proving estimates about probabilities like the ones in lemma 5.13 coupling arguments are a useful tool. The technique is based on the following lemma.

Lemma 5.14. *Given $x, y \in \mathbb{R}$ with $|x| \geq |y|$ we can choose two Brownian motions B^x and B^y with $B_0^x = x$, $B_0^y = y$, and $|B_t^x| \geq |B_t^y|$ for all $t \geq 0$.*

Proof. Let B^x be any Brownian motion with start in x and B be another one with start in y. Define the stopping time T by

$$T = \inf\bigl\{t \geq 0 \bigm| |B_t^x| = |B_t|\bigr\}$$

and the random variable σ by $\sigma = 1$ if $B_T^x = B_T$ and $\sigma = -1$ else. Then the process B^y defined by

$$B_t^y = \begin{cases} B_t & \text{if } t \leq T, \text{ and} \\ B_T + \sigma(B_t^x - B_T^x) & \text{if } t > T \end{cases}$$

is a Brownian motion with $|B_t^y| < |B_t^x|$ for $t < T$ and either $B_t^y = B_t^x$ or $B_t^y = -B_t^x$ for $t \geq T$. This proves the claim. (qed)

The main result of this section is the following generalisation of lemma 5.13.

Proposition 5.15. *Let $b\colon \mathbb{R} \to \mathbb{R}$ be a differentiable function with $b(0) = 0$, $b'(0) \neq 0$ and $\liminf_{|x|\to\infty} |b(x)| > 0$. Then for every $\eta > 0$ we have*

$$\begin{aligned} &\lim_{\varepsilon\downarrow 0} \varepsilon \cdot \log P\Bigl(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| \leq \eta\Bigr) \\ &\qquad = \lim_{\varepsilon\downarrow 0} \varepsilon \cdot \log P\Bigl(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon\Bigr) = -\frac{|b'(0)|^2 t^2}{16}. \end{aligned}$$

Proof. Choose some $0 < \delta < |b'(0)|$. Using the Taylor formula $b(x) = b'(0)\cdot x + o(x)$ we find an $a > 0$ with

$$\bigl(|b'(0)| + \delta\bigr)^2 x^2 \geq b^2(x) \geq \bigl(|b'(0)| - \delta\bigr)^2 x^2 \quad \text{for all } x \in [-a; a]. \tag{5.15}$$

Without loss of generality we may assume that a is smaller than η and also small enough to permit $|b(x)| \geq a\bigl(|b'(0)| - \delta\bigr)$ for all $x \in \mathbb{R}$ with $|x| > a$.

We have to calculate the exponential rates of

$$\begin{aligned} P\Bigl(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon\Bigr) &= P\Bigl(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| \leq a\Bigr) \\ &\qquad + P\Bigl(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| > a\Bigr). \end{aligned} \tag{5.16}$$

Whenever $\sup_{0\leq s\leq t} |B_s| \leq a$ we can approximate $b(x)$ by $b'(0)x$ as in (5.15). This gives

$$\begin{aligned} &P\Bigl(\frac{1}{2}\int_0^t \bigl(|b'(0)| + \delta\bigr)^2 B_s^2\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| \leq a\Bigr) \\ &\qquad \leq P\Bigl(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| \leq a\Bigr) \\ &\qquad \leq P\Bigl(\frac{1}{2}\int_0^t \bigl(|b'(0)| - \delta\bigr)^2 B_s^2\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| \leq a\Bigr). \end{aligned}$$

Both bounds of this estimate can be handled using

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log P\Bigl(\int_0^t c B_s^2\,ds \leq \varepsilon,\ \sup_{0\leq s\leq t} |B_s| \leq a\Bigr) = -c\cdot\frac{t^2}{8},$$

which is a consequence of lemma 5.13 and the scaling property 4.2.

For the lower bound this gives

$$\begin{aligned}
&\liminf_{\varepsilon\downarrow 0} \varepsilon \log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon\Big) \\
&\quad \ge \liminf_{\varepsilon\downarrow 0} \varepsilon \log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| \le a\Big) \\
&\quad \ge -\frac{\big(|b'(0)|+\delta\big)^2}{16} t^2
\end{aligned}$$

whenever $\delta > 0$. For the upper bound we find

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| \le a\Big) \le -\frac{\big(|b'(0)|-\delta\big)^2}{16} t^2. \tag{5.17}$$

Define $v(x) = b^2(x)/\big(|b'(0)|-\delta\big)^2$. Then by our choice of a we have $v(x) \ge x^2 \wedge a^2$ and lemma 5.12 together with 4.2 gives

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \varepsilon \log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| > \eta\Big) \\
&\quad \le \limsup_{\varepsilon\downarrow 0} \varepsilon \log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| > a\Big) \\
&\quad \le -\frac{1}{8}\big(t+\frac{1}{2}a^2\big)^2 \frac{\big(|b'(0)|-\delta\big)^2}{2} \\
&\quad < -\frac{\big(|b'(0)|-\delta\big)^2}{16} t^2.
\end{aligned} \tag{5.18}$$

Using only the last three lines of equation (5.18) we see that the upper bound for (5.16) is dominated by (5.17) and using lemma 2.3 we get

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon\Big) \le -\frac{\big(|b'(0)|-\delta\big)^2}{16} t^2$$

for all $\delta > 0$. Letting $\delta \downarrow 0$ finishes the proof of

$$\lim_{\varepsilon\downarrow 0} \varepsilon \cdot \log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon\Big) = -\frac{|b'(0)|^2 t^2}{16}.$$

Using lemma 2.3 again, but this time with the full equation (5.18) also proves the first equality of the proposition's claim. (qed)

Using a coupling argument we can get the following refinement of proposition 5.15.

Lemma 5.16. *Let $b\colon \mathbb{R} \to \mathbb{R}$ be a differentiable function with $b(0) = 0$, $b'(0) \ne 0$ and $\liminf_{|x|\to\infty} |b(x)| > 0$. Then for every $\eta > 0$ we have*

$$\begin{aligned}
&\lim_{\zeta\downarrow 0} \liminf_{\varepsilon\downarrow 0} \varepsilon \cdot \log \inf_{-\zeta<z<\zeta} P_z\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| \le \eta\Big) \\
&\quad = -\frac{|b'(0)|^2 t^2}{16}
\end{aligned}$$

and

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \varepsilon \cdot \log \sup_{y\in\mathbb{R}} P_y\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| \le \eta\Big) \\
&\quad = -\frac{|b'(0)|^2 t^2}{16}.
\end{aligned}$$

Proof. We start by proving the claim about the lim inf. Using proposition 5.15 we find

$$\begin{aligned}
&\liminf_{\varepsilon\downarrow 0} \varepsilon \cdot \log \inf_{-\zeta<z<\zeta} P_z\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B_s| \leq \eta\Big) \\
&\quad\leq \lim_{\varepsilon\downarrow 0} \varepsilon \cdot \log P_0\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B_s| \leq \eta\Big) \\
&\quad= -\frac{|b'(0)|^2 t^2}{16}
\end{aligned}$$

for every $\zeta > 0$.

Now let $\kappa > 0$ and choose a $\delta > 0$ with

$$-\big(|b'(0)|+\delta\big)^2 \frac{(t+\delta^2)^2}{16} > -\frac{|b'(0)|^2 t^2}{16} - \kappa.$$

As in the proof of proposition 5.15 we can use Taylor approximation to find an $a > 0$ with

$$b^2(x) \leq \big(|b'(0)|+\delta\big)^2 x^2$$

for all $x \in [-a; a]$. Without loss of generality we may assume $a \leq \min(2\delta, \eta)$.

Let $\zeta < a/2$ and $z \in [-\zeta; +\zeta]$. Then we can use lemma 5.14 to choose two Brownian motions B^ζ and B^z with $B^\zeta_0 = \zeta$, $B^z_0 = z$, and $|B^\zeta_t| \geq |B^z_t|$ for all $t \geq 0$. We find

$$\begin{aligned}
&P\Big(\frac{1}{2}\int_0^t b^2(B^z_s)\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B^z_s| \leq \eta\Big) \\
&\quad\geq P\Big(\frac{1}{2}\int_0^t b^2(B^z_s)\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B^z_s| \leq a\Big) \\
&\quad\geq P\Big(\frac{1}{2}\int_0^t \big(|b'(0)|+\delta\big)^2 (B^z_s)^2\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B^z_s| \leq a\Big) \\
&\quad\geq P\Big(\frac{1}{2}\int_0^t \big(|b'(0)|+\delta\big)^2 (B^\zeta_s)^2\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B^\zeta_s| \leq a\Big)
\end{aligned}$$

for every $z \in [-\zeta; +\zeta]$, and thus

$$\begin{aligned}
&\liminf_{\varepsilon\downarrow 0} \varepsilon \cdot \log \inf_{-\zeta<z<\zeta} P_z\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B_s| \leq \eta\Big) \\
&\quad\geq \liminf_{\varepsilon\downarrow 0} \varepsilon \cdot \log P_\zeta\Big(\frac{1}{2}\int_0^t \big(|b'(0)|+\delta\big)^2 B_s^2\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B_s| \leq a\Big) \\
&\quad= \frac{1}{2}\big(|b'(0)|+\delta\big)^2 \liminf_{\varepsilon\downarrow 0} \varepsilon \cdot \log P_\zeta\Big(\int_0^t B_s^2\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B_s| \leq a\Big).
\end{aligned}$$

Because $\zeta < a/2 < \delta$ we can use lemma 5.13 to get

$$\begin{aligned}
&\liminf_{\varepsilon\downarrow 0} \varepsilon \cdot \log \inf_{-\zeta<z<\zeta} P_z\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \leq \varepsilon, \sup_{0\leq s\leq t} |B_s| \leq \eta\Big) \\
&\quad\geq -\frac{1}{2}\big(|b'(0)|+\delta\big)^2 \frac{(t+\zeta^2)^2}{8} \\
&\quad\geq -\frac{1}{2}\big(|b'(0)|+\delta\big)^2 \frac{(t+\delta^2)^2}{8} \\
&\quad> -\frac{|b'(0)|^2 t^2}{16} - \kappa
\end{aligned}$$

for all sufficiently small $\kappa > 0$. Letting $\zeta \downarrow 0$ completes the proof of the first claim.

For the second claim first note that

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \varepsilon \cdot \log \sup_{y\in\mathbb{R}} P_y\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| \le \eta\Big) \\
&\qquad \ge \limsup_{\varepsilon\downarrow 0} \varepsilon \cdot \log P_0\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| \le \eta\Big) \\
&\qquad = -\frac{|b'(0)|^2 t^2}{16},
\end{aligned}$$

again by proposition 5.15.

Let $\kappa > 0$ and choose $\delta > 0$ with

$$-\big(|b'(0)| - \delta\big)^2 \frac{t^2}{16} < -\frac{|b'(0)|^2 t^2}{16} + \kappa.$$

Using Taylor approximation we can find an $a > 0$ with

$$b^2(x) \ge \big(|b'(0)| - \delta\big)^2 x^2$$

for all $x \in [-a; a]$ and by choosing a small enough we can find a smooth, monotone function $\varphi\colon \mathbb{R} \to \mathbb{R}$ with $|b(x)| > |\varphi(x)|$ for all $x \in \mathbb{R}$ and $\varphi'(0) = |b'(0)| - \delta$.

Using the coupling argument and proposition 5.15 again, we get

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \varepsilon \cdot \log \sup_{y\in\mathbb{R}} P_y\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| \le \eta\Big) \\
&\qquad \le \limsup_{\varepsilon\downarrow 0} \varepsilon \cdot \log \sup_{y\in\mathbb{R}} P_y\Big(\frac{1}{2}\int_0^t \varphi^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| \le \eta\Big) \\
&\qquad \le \limsup_{\varepsilon\downarrow 0} \varepsilon \cdot \log P_0\Big(\frac{1}{2}\int_0^t \varphi^2(B_s)\,ds \le \varepsilon,\ \sup_{0\le s\le t} |B_s| \le \eta\Big) \\
&\qquad = -\frac{\Big(|b'(0)| - \delta\Big)^2 t^2}{16} \\
&\qquad < -\frac{|b'(0)|^2 t^2}{16} + \kappa
\end{aligned}$$

for all $\kappa > 0$. Taking the limit $\kappa \downarrow 0$ completes the proof. (qed)

Everything left to do now, is to add the proof of lemma 5.12.

Proof (of lemma 5.12). We need to find an upper bound on the exponential rate for the probability of the event

$$A^\varepsilon = \Big\{\int_0^t v(B_s)\,ds \le \varepsilon,\ \sup_{0\le s<t} |B_s| > a\Big\},$$

which is uniform in the initial point $B_0 = x$. First define two interlaced sequences of stopping times $(S_j)_{j\in\mathbb{N}}$ and $(T_j)_{j\in\mathbb{N}_0}$ by letting $T_0 = 0$ and

$$\begin{aligned}
S_j &= \inf\big\{s > T_{j-1} \bigm| |B_s| \ge a\big\} \\
T_j &= \inf\big\{s > S_j \bigm| |B_s| = a/2\big\}
\end{aligned}$$

for all $j \in \mathbb{N}$. If the initial point $B_0 = x$ has $|x| > a$ we have $S_0 = 0$ and $|B_{S_0}| > a$. Except for this we have $|B_{S_j}| = a$. For $s \in [S_j; T_j]$ we have $|B_s| \ge a/2$ and thus $v(B_s) \ge a^2/4$. Outside these intervals we have $|B_s| < a$ and thus $v(B_s) \ge B_s^2$. Therefore we can conclude

$$\Big\{\int_{S_j}^{T_j} v(B_s)\,ds \le \varepsilon\Big\} \subseteq \Big\{\int_{S_j}^{T_j} a^2/4\,ds \le \varepsilon\Big\} = \Big\{T_j - S_j \le 4\varepsilon/a^2\Big\}$$

Figure 5.3: *Five paths of a Brownian motion on the interval* $[0;1]$*, conditioned on the event that* $\int_0^1 B_s^2\,ds \leq 0.05$ *and that we have* $\sup_{0\leq s\leq 1} B_s > 1$*. One can see that the typical path under these conditions reaches its maximum near the end of the interval. This behaviour fits well with the special rôle of the endpoint in formula* (4.3)*.*

and for $d > 0$ also

$$\Bigl\{\int_{T_{j-1}}^{S_j} v(B_s)\,ds \leq \varepsilon, S_j - T_{j-1} \geq d\Bigr\} \subseteq \Bigl\{\int_{T_{j-1}}^{S_j} B_s^2\,ds \leq \varepsilon, S_j - T_{j-1} \geq d\Bigr\}$$
$$\subseteq \Bigl\{\int_{T_{j-1}}^{T_{j-1}+d} B_s^2\,ds \leq \varepsilon\Bigr\}.$$

As an abbreviation define $J = \lceil 2t/a^2 \rceil + 1$. We want to split the set A^ε into the two parts

$$A^\varepsilon = \bigl(A^\varepsilon \cap \{T_J \leq t\}\bigr) \cup \bigl(A^\varepsilon \cap \{T_J > t\}\bigr).$$

The first part corresponds to the case that there are at least J excursions up to the level $|B_s| = a$ and then back to $|B_s| = a/2$ before time t. For this case we will get an upper bound on the probability from the fact that the process has to move very fast during the intervals $[S_j; T_j]$. The second part corresponds to the case that there is at least one but that there are at most $J-1$ such excursions. This case is more difficult, because we have to take the intervals between the excursions into account.

First consider the case $T_J \leq t$. Here we have

$$A^\varepsilon \cap \{T_J \leq t\} \subseteq \Bigl\{\sum_{j=1}^{J} \int_{S_j}^{T_j} v(B_s)\,ds \leq \varepsilon\Bigr\} \subseteq \Bigl\{\sum_{j=1}^{J} (T_j - S_j) \leq 4\varepsilon/a^2\Bigr\}.$$

Using the strong Markov property for Brownian motion and the reflection principle we find

$$\begin{aligned} P_x\bigl(T_j - S_j \leq \varepsilon\bigr) &\leq P\bigl(\sup_{0\leq s\leq \varepsilon} B_s > a/2\bigr) \\ &= 2P\bigl(B_\varepsilon > a/2\bigr) \\ &= 2P\bigl(\sqrt{\varepsilon} B_1 > a/2\bigr) \end{aligned}$$

for all $x \in \mathbb{R}$. The basic large deviation result for the standard normal distribution on $\mathbb{R}$ (corollary 2.12) now gives

$$\lim_{\varepsilon\downarrow 0} \varepsilon \cdot \log \sup_{x\in\mathbb{R}} P_x\bigl(T_j - S_j \leq \varepsilon\bigr) \leq -\frac{1}{2}\bigl(a/2\bigr)^2 = -\frac{a^2}{8}.$$

In this situation we can apply proposition 2.7 to get

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \varepsilon \log \sup_{x\in\mathbb{R}} P_x\big(A^\varepsilon \cap \{T_J \leq t\}\big) \\
&\qquad\leq \limsup_{\varepsilon\downarrow 0} \varepsilon \log P_x\Big(\sum_{j=1}^J (T_j - S_j) \leq 4\varepsilon/a^2\Big) \\
&\qquad= \frac{a^2}{4} \limsup_{\varepsilon\downarrow 0} \varepsilon \log P_x\Big(\sum_{j=1}^J (T_j - S_j) \leq \varepsilon\Big) \\
&\qquad\leq -\frac{a^2}{4}\Big(\sum_{j=1}^J \frac{a}{\sqrt{8}}\Big)^2 \leq -\frac{1}{8}\big(t + \frac{1}{2}a^2\big)^2.
\end{aligned} \tag{5.19}$$

Now consider the case $T_J > t$. Choose $n \in \mathbb{N}$ with $n > 2J$ and $\varepsilon > 0$ with $4\varepsilon/a^2 < t/n$. Define $\Delta t = t/n$, the intervals $I_1 = [0; \Delta t]$ and $I_k = \big((k-1)\Delta t; k\Delta t\big]$ for $k = 2, \ldots, n$, the index set

$$Q = \Big\{(k_1, \ldots, k_\ell) \in \mathbb{N}^\ell \;\Big|\; \ell \in \{1, \ldots, J\}, 1 \leq k_1 \leq \cdots \leq k_\ell \leq n\Big\},$$

and the event

$$A^\varepsilon_{(k_1,\ldots,k_\ell)} = A^\varepsilon \cap \big\{S_j \in I_{k_j} \text{ for } j = 1, \ldots, \ell \text{ and } S_{\ell+1} > t\big\}.$$

Then we have

$$A^\varepsilon \cap \{T_J > t\} = \bigcup_{q\in Q} A^\varepsilon_q.$$

Choose $(k_1, \ldots, k_\ell) \in Q$. As we have seen above the condition $\int_{S_j}^{T_j} v(B_s)\,ds \leq \varepsilon$ implies $T_j - S_j \leq 4\varepsilon/a^2 \leq \Delta t$. Thus on A^ε_q we have

$$S_j - T_{j-1} \geq \max\big((k_j - k_{j-1} - 2)\Delta t, 0\big) =: d_{j-1} \tag{5.20}$$

for $j = 1, \ldots, \ell - 1$, where we use the convention $k_0 = 0$. If $k_\ell < n$ then we use 5.20 also for $j = \ell$ and we have

$$t - T_\ell \geq \max\big((n - k_\ell - 2)\Delta t, 0\big) =: d_\ell.$$

For $k_\ell = n$ it will turn out that we need to treat the right endpoint of the interval specially, here we define $d_{\ell-1} = \max\big((n - k_{\ell-1} - 3)\Delta t, 0\big)$.

Let $\delta > 0$ and define $D^\delta_{2\ell+1}$ as in lemma 2.6. For $\alpha \in D^\delta_{2\ell+1}$ further define

$$\begin{aligned}
A^{\alpha\varepsilon}_{(k_1,\ldots,k_\ell)} = \Big\{ &\int_{T_0}^{S_1} v(B_s)\,ds \leq \alpha_1\varepsilon, \int_{S_1}^{T_1} v(B_s)\,ds \leq \alpha_2\varepsilon, S_1 \in I_{k_1}, \\
&\vdots \\
&\int_{T_{\ell-1}}^{S_\ell} v(B_s)\,ds \leq \alpha_{2\ell-1}\varepsilon, \int_{S_\ell}^{T_\ell} v(B_s)\,ds \leq \alpha_{2\ell}\varepsilon, S_\ell \in I_{k_\ell}, \\
&\int_{T_\ell}^{t} v(B_s)\,ds \leq \alpha_{2\ell+1}\varepsilon, S_{\ell+1} > t\Big\}
\end{aligned}$$

if $k_\ell < n$ and

$$\begin{aligned}
A^{\alpha\varepsilon}_{(k_1,\ldots,k_\ell)} = \Big\{ &\int_{T_0}^{S_1} v(B_s)\,ds \leq \alpha_1\varepsilon, \int_{S_1}^{T_1} v(B_s)\,ds \leq \alpha_2\varepsilon, S_1 \in I_{k_1}, \\
&\vdots \\
&\int_{T_{\ell-1}}^{S_\ell} v(B_s)\,ds \leq \alpha_{2\ell-1}\varepsilon, S_\ell \in I_n, S_{\ell+1} > t\Big\}
\end{aligned}$$

else. Then we have

$$A^\varepsilon \cap \{T_J > t\} = \bigcup_{q\in Q} A_q^\varepsilon \subseteq \bigcup_{q\in Q} \bigcup_{\alpha\in D^\delta_{2\ell+1}} A_q^{\alpha\varepsilon}.$$

Assume first the case $k_\ell < n$. Then we get

$$\begin{aligned}
P_x\big(A^{\alpha\varepsilon}_{(k_1,\dots,k_\ell)}\big) \le P_x\Big(&\int_{T_0}^{T_0+d_0} B_s^2\,ds \le \alpha_1\varepsilon, T_1 - S_1 \le 4\alpha_2\varepsilon/a^2, S_1\in I_{k_1},\\
&\vdots\\
&\int_{T_{\ell-1}}^{T_{\ell-1}+d_{\ell-1}} B_s^2\,ds \le \alpha_{2\ell-1}\varepsilon, T_\ell - S_\ell \le 4\alpha_{2\ell}\varepsilon/a^2, S_\ell\in I_{k_\ell},\\
&\int_{T_\ell}^{T_\ell+d_\ell} B_s^2\,ds \le \alpha_{2\ell+1}\varepsilon, S_{\ell+1} > t\Big).
\end{aligned}$$

Now we use the strong Markov property of Brownian motion for the stopping times S_j and T_j. Because $|B_{T_j}| = a/2$ and $|B_{S_j}| = a$ are deterministic and the Brownian motion is symmetric we get

$$\begin{aligned}
P_x\big(A^{\alpha\varepsilon}_{(k_1,\dots,k_\ell)}\big) \le P_x\Big(&\int_{T_0}^{T_0+d_0} B_s^2\,ds \le \alpha_1\varepsilon, T_1 - S_1 \le 4\alpha_2\varepsilon/a^2, S_1\in I_{k_1},\\
&\vdots\\
&\int_{T_{\ell-1}}^{T_{\ell-1}+d_{\ell-1}} B_s^2\,ds \le \alpha_{2\ell-1}\varepsilon, T_\ell - S_\ell \le 4\alpha_{2\ell}\varepsilon/a^2, S_\ell\in I_{k_\ell}\Big)\\
\cdot P_{\frac{a}{2}}\Big(&\int_0^{d_\ell} B_s^2\,ds \le \alpha_{2\ell+1}\varepsilon\Big)\\
\le\ P_x\Big(&\int_{T_0}^{T_0+d_0} B_s^2\,ds \le \alpha_1\varepsilon, T_1 - S_1 \le 4\alpha_2\varepsilon/a^2, S_1\in I_{k_1},\\
&\vdots\\
&\int_{T_{\ell-1}}^{T_{\ell-1}+d_{\ell-1}} B_s^2\,ds \le \alpha_{2\ell-1}\varepsilon, S_\ell\in I_{k_\ell}\Big)\\
\cdot P_0\Big(&\sup_{0\le s\le 4\alpha_{2\ell}\varepsilon/a^2} B_s > a/2\Big)\\
\cdot P_{\frac{a}{2}}\Big(&\int_0^{d_\ell} B_s^2\,ds \le \alpha_{2\ell+1}\varepsilon\Big).
\end{aligned}$$

Repeating these two steps for $j = \ell-1,\dots,0$ finally gives

$$\begin{aligned}
P_x\big(A^{\alpha\varepsilon}_{(k_1,\dots,k_\ell)}\big) \le P_x\Big(&\int_0^{d_0} B_s^2\,ds \le \alpha_1\varepsilon\Big)\\
&\cdot \prod_{j=1}^{\ell} P_{\frac{a}{2}}\Big(\int_0^{d_j} B_s^2\,ds \le \alpha_{2j+1}\varepsilon\Big)\\
&\cdot \prod_{j=1}^{\ell} P_0\Big(\sup_{0\le s\le 4\alpha_{2j}\varepsilon/a^2} B_s > a/2\Big).
\end{aligned}$$

In order to use lemma 2.5 we have to calculate the individual rates for the factors on the right-hand side. Using corollary 4.5 we get

$$\lim_{\varepsilon\downarrow 0} \varepsilon\cdot\log\sup_{x\in\mathbb{R}} P_x\Big(\int_0^d B_s^2\,ds \le \varepsilon\Big) = -\frac{1}{8}d^2. \tag{5.21}$$

Using the reflection principle and the basic scaling property of Brownian motion we find

$$\begin{aligned}P_0\Big(\sup_{0\le s\le 4\varepsilon/a^2} B_s > a/2\Big) &= 2P\big(B_{4\varepsilon/a^2} > a/2\big)\\ &= 2P\big(\sqrt{4\varepsilon/a^2}B_1 > a/2\big) = 2P\big(\sqrt{\varepsilon}B_1 > a^2/4\big).\end{aligned}$$

The large deviation principle for the standard normal distribution on $\mathbb{R}$ (corollary 2.12) now gives

$$\lim_{\varepsilon\downarrow 0}\varepsilon\cdot\log P_0\Big(\sup_{0\le s\le 4\varepsilon/a^2} B_s > a/2\Big) = -\frac{1}{2}\big(a^2/4\big)^2 = -\frac{1}{8}\Big(\frac{a^2}{2}\Big)^2. \tag{5.22}$$

Now we can apply lemma 2.5 to get the combined rate. The result is

$$\lim_{\varepsilon\downarrow 0}\varepsilon\log\sup_{x\in\mathbb{R}} P_x\big(A^{\alpha\varepsilon}_{(k_1,\ldots,k_\ell)}\big) \le -\frac{1}{1+\delta}\frac{1}{8}\Big(\sum_{j=0}^{\ell} d_j + n_1\frac{a^2}{4} + \ell\frac{a^2}{2}\Big)^2,$$

where $n_1 = \big|\{j = 1,\ldots,\ell \mid d_j > 0\}\big|$. Because each of the intervals $[S_j;T_j]$ can have a non-empty intersection with at most two of the n intervals I_k we have $\sum_{j=0}^{\ell} d_j \ge n - 2J$ and thus $n_1 \ge 1$. So we find

$$\lim_{\varepsilon\downarrow 0}\varepsilon\log\sup_{x\in\mathbb{R}} P_x\big(A^{\alpha\varepsilon}_{(k_1,\ldots,k_\ell)}\big) \le -\frac{1}{1+\delta}\frac{1}{8}\Big(\frac{n-2J}{n}t + \frac{a^2}{4} + \ell\frac{a^2}{2}\Big)^2 \tag{5.23}$$

for all $\alpha \in D^{\delta}_{2\ell+1}$ and all $\delta > 0$.

Now assume $k_\ell = n$. This case is similar, but needs an additional argument to take care of the case $t \in [S_\ell;T_\ell)$. Here we can no longer use (5.22) for the interval $[S_\ell;T_\ell)$. To work around this we define a stopping time R by

$$R = \inf\big\{s \ge \max(T_{\ell-1},(n-2)\Delta t) \;\big|\; |B_s| = a/2\big\}.$$

Given the event $A^{\alpha\varepsilon}_{(k_1,\ldots,k_\ell)}$ the process cannot have $|B_s| > a/2$ for a period of time of length Δt and using the special definition of $d_{\ell-1}$ for this case we get $T_\ell - 1 + d_{\ell-1} \le R \le S_\ell$.

Similar to the other case we get then

$$\begin{aligned}P_x\big(A^{\alpha\varepsilon}_{(k_1,\ldots,k_\ell)}\big) \le P_x\Big(&\int_{T_0}^{T_0+d_0} B_s^2\,ds \le \alpha_1\varepsilon, T_1 - S_1 \le 4\alpha_2\varepsilon/a^2, S_1 \in I_{k_1},\\ &\vdots\\ &\int_{T_{\ell-2}}^{T_{\ell-2}+d_{\ell-2}} B_s^2\,ds \le \alpha_{2\ell-3}\varepsilon,\\ &\qquad T_{\ell-1} - S_{\ell-1} \le 4\alpha_{2\ell-2}\varepsilon/a^2, S_{\ell-1} \in I_{k_{\ell-1}},\\ &\int_{T_{\ell-1}}^{T_{\ell-1}+d_{\ell-1}} B_s^2\,ds \le \alpha_{2\ell-1}\varepsilon, S_\ell - R \le 4\alpha_{2\ell}\varepsilon/a^2, S_\ell \in I_n\Big).\end{aligned}$$

Using the strong Markov property for the stopping time R first gives

$$\begin{aligned}P_x\big(A^{\alpha\varepsilon}_{(k_1,\ldots,k_\ell)}\big) \le P_x\Big(&\int_{T_0}^{T_0+d_0} B_s^2\,ds \le \alpha_1\varepsilon, T_1 - S_1 \le 4\alpha_2\varepsilon/a^2, S_1 \in I_{k_1},\\ &\vdots\\ &\int_{T_{\ell-2}}^{T_{\ell-2}+d_{\ell-2}} B_s^2\,ds \le \alpha_{2\ell-3}\varepsilon,\\ &\qquad T_{\ell-1} - S_{\ell-1} \le 4\alpha_{2\ell-2}\varepsilon/a^2, S_{\ell-1} \in I_{k_{\ell-1}},\\ &\int_{T_{\ell-1}}^{T_{\ell-1}+d_{\ell-1}} B_s^2\,ds \le \alpha_{2\ell-1}\varepsilon\Big)\\ &\cdot P_0\Big(\sup_{0\le s\le 4\alpha_{2\ell}\varepsilon/a^2} B_s > a/2\Big).\end{aligned}$$

Now we can continue splitting of terms as in the first case to get

$$\begin{aligned} P_x\big(A^{\alpha\varepsilon}_{(k_1,\dots,k_\ell)}\big) \leq P_x\Big(\int_0^{d_0} B_s^2\,ds \leq \alpha_1\varepsilon\Big) \\ \cdot \prod_{j=1}^{\ell-1} P_{\frac{a}{2}}\Big(\int_0^{d_j} B_s^2\,ds \leq \alpha_{2j+1}\varepsilon\Big) \\ \cdot \prod_{j=1}^{\ell} P_0\Big(\sup_{0\leq s\leq 4\alpha_{2j}\varepsilon/a^2} B_s > a/2\Big). \end{aligned}$$

Using equations (5.21), (5.22) and lemma 2.5 as in the first case we get

$$\begin{aligned} \lim_{\varepsilon\downarrow 0} \varepsilon \log \sup_{x\in\mathbb{R}} P_x\big(A^{\alpha\varepsilon}_{(k_1,\dots,k_\ell)}\big) &\leq -\frac{1}{1+\delta}\Big(\sum_{j=0}^{\ell-1} d_j + n_1\frac{a^2}{4} + \ell\frac{a^2}{2}\Big)^2 \\ &\leq -\frac{1}{1+\delta}\frac{1}{8}\Big(\frac{n-2J-1}{n}t + \ell\frac{a^2}{2}\Big)^2 \end{aligned} \tag{5.24}$$

for all $\alpha \in D^\delta_{2\ell+1}$ and all $\delta > 0$. Note that in this case $n_1 = 0$ is possible; this occurs in the case $\ell = 1$ and $S_1 \in I_n$, because I_n was the interval we treated specially.

To estimate the upper exponential rate of $A^\varepsilon \cap \{T_J > t\}$ with lemma 2.2 we need to compare all the rates from 5.23 and 5.24. We get

$$\begin{aligned} &\limsup_{\varepsilon\downarrow 0} \varepsilon \log \sup_{x\in\mathbb{R}} P_x\big(A^\varepsilon \cap \{T_J > t\}\big) \\ &\quad= \max_{q\in Q} \max_{\alpha\in D^\delta_{2\ell+1}} \limsup_{\varepsilon\downarrow 0} \varepsilon \log P_x\big(A^{\alpha\varepsilon}_q\big) \\ &\quad\leq -\frac{1}{1+\delta}\frac{1}{8}\Big(\frac{n-2J-1}{n}t + \frac{a^2}{2}\Big)^2 \end{aligned}$$

for all $\delta > 0$ and large enough n, where the largest bound came from the case $\ell = 1$, $k_1 = n$. Letting first $\delta \downarrow 0$ and then $n \to \infty$ shows

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P_x\big(A^\varepsilon \cap \{T_J > t\}\big) \leq \frac{1}{8}\big(t + \frac{a^2}{2}\big)^2. \tag{5.25}$$

Another application of lemma 2.2 gives the upper bound for $P(A^\varepsilon)$. Using the estimates (5.19) and (5.25) we find

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log \sup_{x\in\mathbb{R}} P_x(A^\varepsilon) \leq \frac{1}{8}\big(t + \frac{a^2}{2}\big)^2.$$

This completes the proof of the lemma 5.12. (qed)

5.3 The LDP for the Endpoint

In this section we use the results of the previous section to finally derive the LDP for the endpoint X_t^ϑ of the solution of

$$\begin{aligned} dX_s^\vartheta &= \vartheta b(X_s) \bullet ds + dB_s \quad \text{on } [0;t] \\ X_0^\vartheta &= z \in \mathbb{R} \end{aligned} \tag{5.26}$$

for $\vartheta \to \infty$. The main result is theorem 5.19 at the end of this section.

We assume $b = -\Phi'$ for a C^2-function $\Phi\colon \mathbb{R} \to \mathbb{R}$ with bounded second derivative Φ''. Then the drift b is Lipschitz continuous and the SDE (5.26) has a unique solution X^ϑ.

Notation: To avoid complicated and hard to read expressions in small print we sometimes write (A) for the indicator function of the event A during this section.

Lemma 5.17. *Let $\Phi\colon \mathbb{R} \to \mathbb{R}$ be a C^2 function with bounded Φ'' and let $b = -\Phi'$. Assume there is an $m \in \mathbb{R}$ with $b(x) = 0$ if and only if $x = m$ and $\liminf_{|x|\to\infty} |b(x)| > 0$. Further assume that there is a rate function $I\colon \mathbb{R} \to [0;\infty]$ with*

$$\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log E\bigl(\exp(-\frac{\vartheta^2}{2}\int_0^t b^2(\omega_s)\,ds) 1_O(B_t)\bigr) \geq -\inf_{x\in O} I(x)$$

for every open set $O \subseteq \mathbb{R}$ and

$$\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log E\bigl(\exp(-\frac{\vartheta^2}{2}\int_0^t b^2(\omega_s)\,ds) 1_K(B_t)\bigr) \leq -\inf_{x\in K} I(x)$$

for every compact set $K \subseteq \mathbb{R}$. For $\vartheta > 0$ let X^ϑ be a solution of the SDE (5.26) with start in $X_0^\vartheta = 0$. Then for $\vartheta \to \infty$ the family $(X_t^\vartheta)_\vartheta$ satisfies the weak LDP with rate function J, where J is defined by

$$J(x) = \Phi(x) - \Phi(0) - \frac{1}{2} \cdot t \cdot \Phi''(m) + I(x).$$

Proof. From Lemma 1.5 we know the density of the distribution of this solution X_t^ϑ:

$$P(X_t^\vartheta \in A) = \int 1_A(\omega_t) \exp\bigl(\vartheta F(\omega) - \vartheta^2 G(\omega)\bigr)\, d\mathbb{W}(\omega) \tag{5.27}$$

where

$$\begin{aligned} F(\omega) &= \Phi(\omega_0) - \Phi(\omega_t) + \frac{1}{2}\int_0^t \Phi''(\omega_s)\,ds \quad \text{and} \\ G(\omega) &= \frac{1}{2}\int_0^t b^2(\omega_s)\,ds. \end{aligned}$$

First let O be open, $x \in O$ and $\delta > 0$. Then we can find an η with $0 < \eta < \delta$, $B_\eta(x) \subseteq O$, and $|\Phi(y) - \Phi(x)| \leq \delta$ for all $y \in B_\eta(x)$. Define

$$F^*(x) = \Phi(0) - \Phi(x) + \frac{1}{2} t\Phi''(m).$$

Then we find

$$\begin{aligned}
&\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log P(X_t^\vartheta \in O) \\
&\quad\geq \liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log P(X_t^\vartheta \in B_\eta(x)) \\
&\quad= \liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_\eta(x)}(\omega_t) \exp\bigl(\vartheta F(\omega) - \vartheta^2 G(\omega)\bigr)\, d\mathbb{W}(\omega) \\
&\quad\geq \liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_\eta(x)}(\omega_t) \exp\bigl(\vartheta (F^*(x) - 2\delta) - \vartheta^2 G(\omega)\bigr) \\
&\qquad\qquad\qquad \cdot \bigl(|F(\omega) - F^*(x)| \leq 2\delta\bigr)\, d\mathbb{W}(\omega) \\
&\quad= F^*(x) - 2\delta + \liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_\eta(x)}(\omega_t) \exp\bigl(-\vartheta^2 G(\omega)\bigr) \\
&\qquad\qquad\qquad \cdot \bigl(|F(\omega) - F^*(x)| \leq 2\delta\bigr)\, d\mathbb{W}(\omega).
\end{aligned}$$

By definition of $F^*(x)$ we have

$$\begin{aligned}
\bigl|F(\omega) - F^*(x)\bigr| &= \Bigl|\Phi(0) - \Phi(\omega_t) + \frac{1}{2}\int_0^t \Phi''(\omega_s)\, ds \\
&\qquad - \Phi(0) + \Phi(x) - \frac{1}{2} t \Phi''(m)\Bigr| \\
&\leq \bigl|\Phi(x) - \Phi(\omega_t)\bigr| + \frac{1}{2}\int_0^t \bigl|\Phi''(\omega_s) - \Phi''(m)\bigr|\, ds.
\end{aligned}$$

Thus whenever $\omega_t \in B_\eta(x)$ and $\bigl|F(\omega) - F^*(x)\bigr| \geq 2\delta$ we find

$$\frac{1}{2}\int_0^t \bigl|\Phi''(\omega_s) - \Phi''(m)\bigr|\, ds \geq 2\delta - \delta = \delta.$$

Because Φ'' is bounded the above estimate implies that we can find an $\varepsilon > 0$ with

$$\Bigl|\bigl\{ s \in [0;t] \bigm| |\omega_s - m| \geq \delta/t \bigr\}\Bigr| > \varepsilon$$

for all paths ω with $\omega_t \in B_\eta(x)$ and $\bigl|F(\omega) - F^*(x)\bigr| \geq 2\delta$. Because m is the only zero of b and because $\liminf_{|x|\to\infty} |b(x)| > 0$ we have

$$\inf\bigl\{ b^2(x) \bigm| |x - m| \geq \delta/t \bigr\} > 0,$$

i.e. we can find a $g > 0$ with $G(\omega) > g$ for all paths ω with $\omega_t \in B_\eta(x)$ and $\bigl|F(\omega) - F^*(x)\bigr| \geq 2\delta$. Together this gives

$$\begin{aligned}
&\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_\eta(x)}(\omega_t) \exp\bigl(-\vartheta^2 G(\omega)\bigr) (|F(\omega) - F^*(x)| > 2\delta)\, d\mathbb{W}(\omega) \\
&\quad\leq \limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int \exp\bigl(-\vartheta^2 g\bigr)\, d\mathbb{W}(\omega) \\
&\quad= -\infty.
\end{aligned}$$

So we can use lemma 2.3 to conclude

$$\begin{aligned}
&\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_\eta(x)}(\omega_t) \exp\bigl(-\vartheta^2 G(\omega)\bigr)\, d\mathbb{W}(\omega) \\
&\quad= \liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_\eta(x)}(\omega_t) \exp\bigl(-\vartheta^2 G(\omega)\bigr) (|F(\omega) - F^*(x)| \leq 2\delta)\, d\mathbb{W}(\omega)
\end{aligned}$$

and get

$$\begin{aligned}
&\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log P(X_t^\vartheta \in O) \\
&\quad \geq F^*(x) - 2\delta + \liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_\eta(x)}(\omega_t) \exp\bigl(-\vartheta^2 G(\omega)\bigr)\, d\mathbb{W}(\omega) \\
&\quad \geq F^*(x) - 2\delta - \inf_{y\in B_\eta(x)} I(y) \\
&\quad \geq F^*(x) - 2\delta - I(x)
\end{aligned}$$

for all $\delta > 0$. Letting $\delta \downarrow 0$ gives

$$\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log P(X_t^\vartheta \in O) \geq F^*(x) - I(x)$$

and taking the supremum over all $x \in O$ on the right hand side proves the lower bound.

Now let $K \subseteq \mathbb{R}$ be compact and $\delta > 0$. For each $x \in K$ we can find an $\eta > 0$ with $|\Phi(y) - \Phi(x)| \leq \delta$ whenever $y \in B_\eta(x)$. Because I is lower semi-continuous we can assume $I(y) \geq I(x) - \delta$ for every $y \in \overline{B}_\eta(x)$ by choosing η small enough. Using the compactness of K we can cover K with a finite number of such balls: there are $x_1, \ldots, x_n \in K$ and $0 < \eta_1, \ldots, \eta_n < \delta$ with

$$K \subseteq \bigcup_{k=1}^n B_{\eta_k}(x_k)$$

and the above assumption on Φ and I hold for each k. For $k = 1, \ldots, n$ consider $F^*(x_k)$ as defined above. This time we find

$$\begin{aligned}
&\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log P(X_t^\vartheta \in K) \\
&\quad \leq \limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \sum_{k=1}^n P(X_t^\vartheta \in B_{\eta_k}(x_k)) \\
&\quad = \max_{k=1,\ldots,n} \limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_{\eta_k}(x_k)}(\omega_t) \exp\bigl(\vartheta F(\omega) - \vartheta^2 G(\omega)\bigr)\, d\mathbb{W}(\omega).
\end{aligned}$$

Because F is bounded on $\{\omega_t \in B_{\eta_k}(x_k)\}$ we can use lemma 2.3 as above to conclude

$$\begin{aligned}
&\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_{\eta_k}(x_k)}(\omega_t) \exp\bigl(\vartheta F(\omega) - \vartheta^2 G(\omega)\bigr)\, d\mathbb{W}(\omega) \\
&\qquad = \limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_{\eta_k}(x_k)}(\omega_t) \exp\bigl(\vartheta F(\omega) - \vartheta^2 G(\omega)\bigr) \\
&\qquad\qquad\qquad \cdot (|F(\omega) - F^*(x_k)| \leq 2\delta)\, d\mathbb{W}(\omega)
\end{aligned}$$

for $k = 1, \ldots, n$. This gives

$$\begin{aligned}
&\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log P(X_t^\vartheta \in K) \\
&\quad \leq \max_{k=1,\ldots,n} \limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_{\eta_k}(x_k)}(\omega_t) \exp\bigl(\vartheta F(\omega) - \vartheta^2 G(\omega)\bigr) \\
&\qquad\qquad \cdot (|F(\omega) - F^*(x_k)| \leq 2\delta)\, d\mathbb{W}(\omega) \\
&\quad \leq \max_{k=1,\ldots,n} \limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{B_{\eta_k}(x_k)}(\omega_t) \exp\bigl(\vartheta (F^*(x_k) + 2\delta) - \vartheta^2 G(\omega)\bigr) \\
&\qquad\qquad \cdot \bigl(|F(\omega) - F^*(x_k)| \leq 2\delta\bigr)\, d\mathbb{W}(\omega) \\
&\quad \leq \max_{k=1,\ldots,n} F^*(x_k) + 2\delta \\
&\qquad\qquad + \limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log \int 1_{\overline{B}_{\eta_k}(x_k)}(\omega_t) \exp\bigl(-\vartheta^2 G(\omega)\bigr)\, d\mathbb{W}(\omega).
\end{aligned}$$

Now we can use the upper bound on the rate of the integral and our choice of η_k to get

$$\begin{aligned}\limsup_{\vartheta\to\infty}\frac{1}{\vartheta}\log P(X_t^\vartheta\in K)\\ &\leq \max_{k=1,\dots,n} F^*(x_k)+2\delta-\inf_{y\in\overline{B}_\delta(x_k)} I(y)\\ &\leq \max_{k=1,\dots,n} F^*(x_k)+2\delta-I(x_k)+\delta.\end{aligned}$$

and letting $\delta\downarrow 0$ finishes the proof for compact sets. (qed)

To get the upper bound for general closed sets we have to show exponential tightness of the $\mathcal{L}(X_t^\vartheta)$, i.e. we have to show that for every $\alpha\in\mathbb{R}$ there exists a compact set K_α, such that

$$\limsup_{\vartheta\to\infty}\frac{1}{\vartheta}\log P(X_t^\vartheta\notin K_\alpha)<-\alpha.$$

The upper bound for arbitrary closed sets would then follow from lemma 2.1.

Note that we can learn some properties of I from the fact that J is a rate function. As a rate function J is positive. So we can conclude that the function I from the lemma must satisfy

$$I(x)\geq-\Phi(x)+\Phi(0)+\frac{1}{2}\cdot t\cdot\Phi''(m)$$

for all $x\in\mathbb{R}$.

The following lemma is a generalisation of corollary 4.3. It helps to determine the rate function I which is needed to apply lemma 5.17.

Lemma 5.18. *Let $m\in\mathbb{R}$ and $b\colon\mathbb{R}\to\mathbb{R}$ be a C^2-function with $b(x)=0$ if and only if $x=m$, $b'(m)\neq 0$, and $\liminf_{|x|\to\infty}|b(x)|>0$. Then for any compact set $K\subseteq\mathbb{R}$ we have*

$$\begin{aligned}\limsup_{\varepsilon\to 0}\varepsilon\log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds\leq\varepsilon, B_t\in K\Big)\\ &\leq-\frac{1}{4}\inf_{a\in K}\Big(\Big|\int_0^m|b(x)|\,dx\Big|+\frac{1}{2}|b'(m)|t+\Big|\int_m^a|b(x)|\,dx\Big|\Big)^2\end{aligned}$$

and for any open set $O\subseteq\mathbb{R}$ we have

$$\begin{aligned}\liminf_{\varepsilon\to 0}\varepsilon\log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds\leq\varepsilon, B_t\in O\Big)\\ &\geq-\frac{1}{4}\inf_{a\in O}\Big(\Big|\int_0^m|b(x)|\,dx\Big|+\frac{1}{2}|b'(m)|t+\Big|\int_m^a|b(x)|\,dx\Big|\Big)^2.\end{aligned}$$

Proof. As an abbreviation define $v(x)=b^2(x)/2$ for all $x\in\mathbb{R}$. For the proof of the upper bound choose a compact set K, let $\delta,\eta,\tau>0$ and choose D_3^δ as in lemma 2.6. Then for $\varepsilon<t/2\tau$ we have

$$\begin{aligned}\Big\{\int_0^t v(B_s)\,ds\leq\varepsilon, B_t\in K\Big\}\\ &\subseteq\bigcup_{\alpha\in D_3^\delta}\Big\{\int_0^{\varepsilon\tau}v(B_s)\,ds\leq\alpha_1\varepsilon,\int_{\varepsilon\tau}^{t-\varepsilon\tau}v(B_s)\,ds\leq\alpha_2\varepsilon,\\ &\qquad\int_{t-\varepsilon\tau}^t v(B_s)\,ds\leq\alpha_3\varepsilon, B_t\in K\Big\}.\end{aligned}$$

Writing (A) for the indicator function of A and using the strong Markov property of Brownian motion this gives

$$
\begin{aligned}
&P\Big(\int_0^t v(B_s)\,ds \le \varepsilon, B_t \in K\Big) \\
&\quad \le \sum_{\alpha\in D_3^\delta} E\Big((\textstyle\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_1\varepsilon)(\int_{\varepsilon\tau}^{t-\varepsilon\tau} v(B_s)\,ds \le \alpha_2\varepsilon) \\
&\qquad\qquad E\big((\textstyle\int_{t-\varepsilon\tau}^{t} v(B_s)\,ds \le \alpha_3\varepsilon, B_t \in K) \mid \mathcal{F}_{t-\varepsilon\tau}\big)\Big) \\
&\quad = \sum_{\alpha\in D_3^\delta} E\Big((\textstyle\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_1\varepsilon)(\int_{\varepsilon\tau}^{t-\varepsilon\tau} v(B_s)\,ds \le \alpha_2\varepsilon) \\
&\qquad\qquad E_{B_{t-\varepsilon\tau}}\big((\textstyle\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_3\varepsilon, B_{\varepsilon\tau} \in K)\big)\Big) \\
&\quad =: \sum_{\alpha\in D_3^\delta} p(\alpha,\varepsilon)
\end{aligned}
$$

Now let $\alpha \in D_3^\delta$ be fixed and $a > 0$. We split the corresponding event further by distinguishing the two cases $\big\{\sup_{\varepsilon\tau\le s\le t-\varepsilon\tau} |B_s - m| > a\big\}$ and $\big\{\sup_{\varepsilon\tau\le s\le t-\varepsilon\tau} |B_s - m| \le a\big\}$. Since omitting some conditions makes the probability only larger we get

$$p(\alpha,\varepsilon) \le p_1(\alpha,\varepsilon) + p_2(\alpha,\varepsilon)$$

with

$$p_1(\alpha,\varepsilon) = \sup_{y\in\mathbb{R}} P_y\Big(\int_0^{t-2\varepsilon\tau} v(B_s)\,ds \le \alpha_2\varepsilon, \sup_{0\le s\le t-2\varepsilon\tau} |B_s - m| > a\Big)$$

and

$$
\begin{aligned}
p_2(\alpha,\varepsilon) = P\Big(&\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_1\varepsilon, |B_{\varepsilon\tau} - m| \le a\Big) \\
&\cdot \sup_{y\in\mathbb{R}} P_y\Big(\int_0^{t-2\varepsilon\tau} v(B_s)\,ds \le \alpha_2\varepsilon, \sup_{0\le s\le t-2\varepsilon\tau} |B_s - m| \le a\Big) \\
&\cdot \sup_{|z-m|\le a} P_z\Big(\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_3\varepsilon, B_{\varepsilon\tau} \in K\Big).
\end{aligned}
$$

To calculate the rate for the sum $p_1(\alpha,\varepsilon) + p_2(\alpha,\varepsilon)$ we have to calculate the rates of the individual terms. For p_1 we can use lemma 5.12 to get

$$
\begin{aligned}
&\limsup_{\varepsilon\to 0} \varepsilon \log p_1(\alpha,\varepsilon) \\
&\quad \le \limsup_{\varepsilon\to 0} \varepsilon \log \sup_{|y-m|<a/2} P_y\Big(\int_0^{t-\eta} v(B_s)\,ds \le \alpha_2\varepsilon, \sup_{0\le s\le t-\eta} |B_s - m| > a\Big), \\
&\quad \le -\frac{1}{8}\Big(t - \eta + \frac{1}{2}a^2\Big)^2.
\end{aligned}
$$

Since for fixed η this rate become arbitrary small when a becomes large, we can choose a large enough that the rate of $p_1(\alpha,\varepsilon) + p_2(\alpha,\varepsilon)$ is dominated by p_2.

To treat the p_2-term we apply lemma 2.5 for the rate of a product. From proposition 5.3 we know the individual rates

$$
\begin{aligned}
\limsup_{\varepsilon\to 0} \varepsilon \log P\Big(&\int_0^{\varepsilon\tau} v(B_s)\,ds \le \varepsilon, |B_{\varepsilon\tau} - m| \le a\Big) \\
&\le -\frac{1}{4}\Big(\int_0^m |b(x)|\,dx\Big)^2 \cdot r_1^2(\tau)
\end{aligned}
$$

and

$$\begin{aligned}\limsup_{\varepsilon\to 0}\varepsilon\log\sup_{|z-m|\le a}P_z\Big(\int_0^{\varepsilon\tau}v(B_s)\,ds\le\varepsilon,B_{\varepsilon\tau}\in K\Big)\\ \le-\frac{1}{4}\inf_{a\in K}\Big(\int_m^a|b(x)|\,dx\Big)^2\cdot r_2^2(\tau)\end{aligned}$$

where $\lim_{\tau\to\infty}r_1(\tau)=\lim_{\tau\to\infty}r_2(\tau)=1$, and lemma 5.16 gives

$$\begin{aligned}&\limsup_{\varepsilon\to 0}\varepsilon\log\sup_{y\in\mathbb{R}}P_y\Big(\int_0^{t-2\varepsilon\tau}v(B_s)\,ds\le\varepsilon,\sup_{0\le s\le t-2\varepsilon\tau}|B_s-m|\le a\Big)\\ &\quad\le\limsup_{\varepsilon\to 0}\varepsilon\log\sup_{|y-m|<a/2}P_y\Big(\int_0^{t-\eta}v(B_s)\,ds\le\varepsilon,\sup_{0\le s\le t-\eta}|B_s-m|\le a\Big)\\ &\quad\le-\frac{|b'(m)|^2(t-\eta)^2}{16}.\end{aligned}$$

Using lemma 2.5 we get the combined rate

$$\begin{aligned}&\limsup_{\varepsilon\to 0}\varepsilon\log p(\alpha,\varepsilon)\\ &\quad\le-\frac{1}{1+\delta}\Big(\frac{1}{2}\Big|\int_0^m|b(x)|\,dx\Big|\cdot r_1(\tau)\\ &\qquad\qquad+\frac{1}{4}|b'(m)|\cdot(t-\eta)+\frac{1}{2}\inf_{a\in K}\Big|\int_m^a|b(x)|\,dx\Big|\cdot r_2(\tau)\Big)^2\end{aligned}$$

for all $\alpha\in D_3^\delta$.

The rate for the sum over all $\alpha\in D_3^\delta$ can be estimated with lemma 2.2. The result is

$$\begin{aligned}&\limsup_{\varepsilon\to 0}\varepsilon\log P\Big(\int_0^t v(B_s)\,ds\le\varepsilon,B_t\in K\Big)\\ &\quad\le-\frac{1}{1+\delta}\Big(\frac{1}{2}\Big|\int_0^m|b(x)|\,dx\Big|\cdot r_1(\tau)\\ &\qquad\qquad+\frac{1}{4}|b'(m)|\cdot(t-\eta)+\frac{1}{2}\inf_{a\in K}\Big|\int_m^a|b(x)|\,dx\Big|\cdot r_2(\tau)\Big)^2\end{aligned}$$

for all $\eta>0$, $\delta>0$, and $\tau>0$. Letting finally $\tau\to\infty$, $\delta\downarrow 0$, and $\eta\downarrow 0$ gives

$$\begin{aligned}&\limsup_{\varepsilon\to 0}\varepsilon\log P\Big(\frac{1}{2}\int_0^t b^2(B_s)\,ds\le\varepsilon,B_t\in K\Big)\\ &\quad\le-\frac{1}{4}\Big(\frac{1}{2}\Big|\int_0^m|b(x)|\,dx\Big|+\frac{1}{2}|b'(m)|t+\inf_{a\in K}\Big|\int_m^a|b(x)|\,dx\Big|\Big)^2.\end{aligned}$$

This proves the upper bound.

For the lower bound: Let $\zeta,\eta,\tau>0$ and $\alpha_1,\alpha_2,\alpha_3\in\mathbb{R}$ with $\alpha_1+\alpha_2+\alpha_3=1$. Then for $\varepsilon<t/2\tau$ we have

$$\begin{aligned}&\Big\{\int_0^t v(B_s)\,ds\le\varepsilon,B_t\in O\Big\}\\ &\quad\supseteq\Big\{\int_0^{\varepsilon\tau}v(B_s)\,ds\le\alpha_1\varepsilon,|B_{\varepsilon\tau}-m|<\zeta\Big\}\\ &\qquad\cap\Big\{\int_{\varepsilon\tau}^{t-\varepsilon\tau}v(B_s)\,ds\le\alpha_2\varepsilon,|B_{t-\varepsilon\tau}-m|<\eta\Big\}\\ &\qquad\cap\Big\{\int_{t-\varepsilon\tau}^{t}v(B_s)\,ds\le\alpha_3\varepsilon,B_t\in O\Big\}\end{aligned}$$

and thus we get

$$
\begin{aligned}
P\Big(\int_0^t & v(B_s)\,ds \le \varepsilon, B_t \in O\Big) \\
&\ge E\Big(\big(\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_1\varepsilon, |B_{\varepsilon\tau} - m| < \zeta\Big\}\big) \\
&\qquad \cdot \big(\int_{\varepsilon\tau}^{t-\varepsilon\tau} v(B_s)\,ds \le \alpha_2\varepsilon, |B_{t-\varepsilon\tau} - m| < \eta\big) \\
&\qquad \cdot E\Big(\big(\int_{t-\varepsilon\tau}^{t} v(B_s)\,ds \le \alpha_3\varepsilon, B_t \in O\Big) \,\Big|\, \mathcal{F}_{t-\varepsilon\tau}\Big)\Big) \\
&\ge E\Big(\big(\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_1\varepsilon, |B_{\varepsilon\tau} - m| < \zeta\big) \\
&\qquad \cdot E\Big(\big(\int_{\varepsilon\tau}^{t-\varepsilon\tau} v(B_s)\,ds \le \alpha_2\varepsilon, |B_{t-\varepsilon\tau} - m| < \eta\big) \,\Big|\, \mathcal{F}_{\varepsilon\tau}\Big)\Big) \\
&\qquad \cdot \inf_{m-\eta<y<m+\eta} P_y\Big(\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_3\varepsilon, B_{\varepsilon\tau} \in O\Big) \\
&\ge P_0\Big(\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_1\varepsilon, B_{\varepsilon\tau} \in (m-\zeta; m+\zeta)\Big) \\
&\qquad \cdot \inf_{m-\zeta<z<m+\zeta} P_z\Big(\int_0^{t-2\varepsilon\tau} v(B_s)\,ds \le \alpha_2\varepsilon, |B_{t-2\varepsilon\tau} - m| < \eta\Big) \\
&\qquad \cdot \inf_{m-\eta<y<m+\eta} P_y\Big(\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_3\varepsilon, B_{\varepsilon\tau} \in O\Big).
\end{aligned}
$$

First take lower exponential rates for $\varepsilon \downarrow 0$. The lower exponential rate of the left-hand side is greater or equal to the sum of the lower rates of the right-hand side. This inequality holds for all $\eta, \tau > 0$ and $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ with $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Then let $\tau \to \infty$. We treat the three terms on the right hand side individually. First term: from Lemma 5.1 we know

$$
\begin{aligned}
\lim_{\tau\to\infty} \liminf_{\varepsilon\downarrow 0} \varepsilon \log P_0 & \Big(\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_1\varepsilon, B_{\varepsilon\tau} \in (m-\zeta; m+\zeta)\Big) \\
&\ge -\frac{1}{\alpha_1}\frac{1}{4} \inf_{m-\zeta<a<m+\zeta} \Big(\big|\int_0^m |b(x)|\,dx\big| + \big|\int_m^a |b(x)|\,dx\big|\Big)^2 \\
&= -\frac{1}{\alpha_1}\frac{1}{4}\Big(\big|\int_0^m |b(x)|\,dx\big|\Big)^2 \cdot r_1(\zeta)
\end{aligned}
$$

where $\lim_{\zeta\downarrow 0} r_1(\zeta) = 1$.

Second term: we can make the probability smaller by replacing $t - 2\varepsilon\tau$ with t. Then the term is no longer τ-dependent and using lemma 5.16 we get

$$
\begin{aligned}
\liminf_{\varepsilon\downarrow 0} \varepsilon \log \inf_{m-\zeta<z<m+\zeta} & P_z\Big(\int_0^{t-2\varepsilon\tau} v(B_s)\,ds \le \alpha_2\varepsilon, |B_{t-2\varepsilon\tau} - m| < \eta\big)\Big) \\
&\ge -\frac{1}{\alpha_2}\frac{|b'(m)|^2}{16} t^2 \cdot r_2(\zeta)
\end{aligned}
$$

where $\lim_{\zeta\downarrow 0} r_2(\zeta) = 1$.

Third term: using corollary 5.4 we get

$$
\begin{aligned}
\liminf_{\varepsilon\downarrow 0} \varepsilon \log \inf_{m-\eta<y<m+\eta} & P_y\Big(\int_0^{\varepsilon\tau} v(B_s)\,ds \le \alpha_3\varepsilon, B_{\varepsilon\tau} \in O\Big) \\
&\ge -\frac{1}{\alpha_3}\frac{1}{4} \inf_{a\in O}\Big(\int_m^a |b(x)|\,dx\Big)^2 \cdot r_3(\eta)
\end{aligned}
$$

where $\lim_{\eta\downarrow 0} r_3(\eta) = 1$.

Combining the three rates we get

$$\begin{aligned}\liminf_{\varepsilon\downarrow 0} \varepsilon \log P\Big(B_t \in O, \int_0^t v(B_s)\,ds \le \varepsilon\Big) \\ \ge -\frac{1}{\alpha_1}\frac{1}{4}\Big(\Big|\int_0^m |b(x)|\,dx\Big|\Big)^2 \cdot r_1(\zeta) \\ -\frac{1}{\alpha_2}\frac{|b'(m)|^2}{16}t^2 \cdot r_2(\zeta) \\ -\frac{1}{\alpha_3}\frac{1}{4}\inf_{a\in O}\Big(\Big|\int_m^a |b(x)|\,dx\Big|\Big)^2 \cdot r_3(\eta).\end{aligned}$$

and letting first $\zeta \downarrow 0$ and then $\eta \downarrow 0$ yields

$$\begin{aligned}\liminf_{\varepsilon\downarrow 0} \varepsilon \log P\Big(B_t \in O, \int_0^t v(B_s)\,ds \le \varepsilon\Big) \\ \ge -\frac{1}{\alpha_1}\Big(\frac{1}{2}\int_0^m |b(x)|\,dx\Big)^2 \\ -\frac{1}{\alpha_2}\Big(\frac{|b'(m)|}{4}t\Big)^2 \\ -\frac{1}{\alpha_3}\Big(\frac{1}{2}\inf_{a\in O}\int_m^a |b(x)|\,dx\Big)^2\end{aligned}$$

for all $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ with $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Choosing optimal α_1, α_2, and α_3 as described in lemma 2.4 we get

$$\begin{aligned}\liminf_{\varepsilon\downarrow 0} \varepsilon \log P\Big(B_t \in O, \frac{1}{2}\int_0^t b^2(B_s)\,ds \le \varepsilon\Big) \\ \ge -\Big(\frac{1}{2}\Big|\int_0^m |b(x)|\,dx\Big| + \frac{|b'(m)|}{4}t + \frac{1}{2}\inf_{a\in O}\Big|\int_m^a |b(x)|\,dx\Big|\Big)^2 \\ = -\frac{1}{4}\Big(\Big|\int_0^m |b(x)|\,dx\Big| + \frac{|b'(m)|}{2}t + \inf_{a\in O}\Big|\int_m^a |b(x)|\,dx\Big|\Big)^2.\end{aligned}$$

This completes the proof. (qed)

The main result of this chapter is the following theorem together with the corollaries 5.20 and 5.21.

Theorem 5.19. *Let $\Phi\colon \mathbb{R} \to \mathbb{R}$ be a C^3-function with bounded Φ'' and $b = -\Phi'$. Assume there is an $m \in \mathbb{R}$ with $b(x) = 0$ if and only if $x = m$, $b'(m) \neq 0$, and $\liminf_{|x|\to\infty} |b(x)| > 0$. Then for every $t > 0$ the solution X^ϑ of*

$$\begin{aligned}dX_s^\vartheta &= \vartheta b(X_s) \bullet ds + dB_s \quad \text{for } s \in [0;t], \text{ and} \\ X_0^\vartheta &= z \in \mathbb{R}\end{aligned}$$

satisfies the following weak LDP: for every compact set $K \subseteq \mathbb{R}$ we have

$$\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log P(X_t^\vartheta \in K) \le -\inf_{x\in K} J_t(x)$$

and for every open set $O \subseteq \mathbb{R}$ we have

$$\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log P(X_t^\vartheta \in O) \ge -\inf_{x\in O} J_t(x),$$

where the rate function is

$$J_t(x) = V_z^m(\Phi) - \Phi(z) + t(\Phi''(m))^- + V_m^x(\Phi) + \Phi(x). \tag{5.28}$$

In the theorem $V_a^b(\Phi)$ denotes the total variation of Φ between a and b. It can be interpreted as the "cost" of the process going from a to b. Because $b = -\Phi'$ we have

$$V_a^b(\Phi) = \Big|\int_a^b |b(x)|\,dx\Big|$$

for any $a, b \in \mathbb{R}$. The notation $(\Phi''(m))^-$ denotes the negative part of $\Phi''(m)$, i.e. $(\Phi''(m))^- = 0$ if $\Phi''(m) \geq 0$ and $(\Phi''(m))^- = |\Phi''(m)|$ if $\Phi''(m) < 0$. This can be interpreted as the "cost" of staying near m for a unit of time. This term only occurs, if the equilibrium point m is unstable.

Proof. Since the rate function J_t is invariant under space shifts we can without loss of generality assume $z = 0$ by replacing Φ with the shifted function $\Phi(\,\cdot\, + z)$ and starting the SDE in 0. Since most of the work was already done in the previous section, the proof consists only of three steps.

First define

$$\begin{aligned} H(x) &= \frac{1}{4}\Big(\Big|\int_0^m |b(y)|\,dy\Big| + \frac{1}{2}|b'(m)|t + \Big|\int_{[m;x]} |b(y)|\,dy\Big|\Big)^2 \\ &= \frac{1}{4}\Big(V_0^m(\Phi) + \frac{1}{2}|b'(m)|t + V_m^x(\Phi)\Big)^2 \end{aligned}$$

and $v(x) = b^2(x)/2$ for all $y \in \mathbb{R}$. From lemma 5.18 we know that for every compact set $K \subseteq \mathbb{R}$ we have

$$\limsup_{\varepsilon\to 0} \varepsilon \log P\Big(\int_0^t v(B_s)\,ds \leq \varepsilon, B_t \in K\Big) \leq -\inf_{a\in K} H(a)$$

and for every open set $O \subseteq \mathbb{R}$ we have

$$\liminf_{\varepsilon\to 0} \varepsilon \log P\Big(\int_0^t v(B_s)\,ds \leq \varepsilon, B_t \in O\Big) \geq -\inf_{a\in O} H(a).$$

Second, let

$$I(x) = 2\sqrt{H(x)} = V_0^m(\Phi) + \frac{1}{2}|b'(m)|t + V_m^x(\Phi)$$

for all $x \in \mathbb{R}$. Then for every set $A \subseteq \mathbb{R}$ we find

$$-2\sqrt{\big|-\inf_{x\in A} H(x)\big|} = -2\sqrt{\inf_{x\in A} H(x)} = -\inf_{x\in A} I(x)$$

and corollary 4.8 allows us to conclude

$$\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log E\big(\exp(-\vartheta^2 \int_0^t v(\omega_s)\,ds) 1_K(B_t)\big) \leq -\inf_{x\in K} I(x)$$

for every compact set $K \subseteq \mathbb{R}$ and

$$\liminf_{\vartheta\to\infty} \frac{1}{\vartheta} \log E\big(\exp(-\vartheta^2 \int_0^t v(\omega_s)\,ds) 1_O(B_t)\big) \geq -\inf_{x\in O} I(x)$$

for every open set $O \subseteq \mathbb{R}$.

Finally we can use lemma 5.17 to conclude that the family $(X_t^\vartheta)_{\vartheta>0}$ satisfies the weak LDP with rate function

$$\begin{aligned} J_t(x) &= \Phi(x) - \Phi(0) - \frac{1}{2}\cdot t\cdot \Phi''(m) + I(x) \\ &= \Phi(x) - \Phi(0) + V_0^m(\Phi) + t(\Phi''(m))^- + V_m^x(\Phi). \end{aligned}$$

This completes the proof. (qed)

Given the sign of $b'(m)$ the rate function from the theorem can be simplified because the drift b has only one zero. The following corollary describes the case of $b'(m) < 0$, which corresponds to attracting drift. In this case the weak LDP from the theorem can be strengthend to the full LDP.

Corollary 5.20. *Let $\Phi\colon \mathbb{R} \to \mathbb{R}$ be a C^3-function with bounded Φ'' and $b = -\Phi'$. Assume there is an $m \in \mathbb{R}$ with $b(x) = 0$ if and only if $x = m$, $b'(m) < 0$, and $\liminf_{|x|\to\infty} |b(x)| > 0$. Furthermore let X^ϑ be the solution of*

$$\begin{aligned} dX_t^\vartheta &= \vartheta b(X_t) \bullet dt + dB_t, \\ X_0^\vartheta &= z \in \mathbb{R}. \end{aligned} \tag{5.29}$$

Then the following claims hold:

a) For every $t > 0$ the family $(X_t^\vartheta)_{\vartheta>0}$ for $\vartheta \to \infty$ satisfies the weak LDP on $\mathbb{R}$ with rate function

$$J_t(x) = 2\big(\Phi(x) - \Phi(m)\big) \quad \text{for all } x \in \mathbb{R}. \tag{5.30}$$

b) If b is monotone, then the family $(X_t^\vartheta)_{\vartheta>0}$ satisfies the full LDP with rate function J_t.

Proof. a) Since we assume that m is the only zero of the drift b, for $b'(m) < 0$ the point m is the minimum of Φ. In this case we have $V_z^m(\Phi) = \Phi(z) - \Phi(m)$, $V_m^x(\Phi) = \Phi(x) - \Phi(m)$ and $\Phi''(m) > 0$, so the rate function simplifies to the expression given in formula (5.30).

b) To strengthen the weak LDP to the full LDP we have to check the exponential tightness condition from lemma 2.1, i.e. we have to show that for every $c > 0$ there is an $a > 0$ with

$$\limsup_{\vartheta\to\infty} \frac{1}{\vartheta} \log P\big(|X_t^\vartheta - m| > a\big) < -c. \tag{5.31}$$

We use a comparison argument to obtain this estimate.

Using the assumption $\liminf_{|x|\to\infty} |b(x)| > 0$ and theorem 1.4 we find that the SDE (5.29) has a stationary distribution with density $\exp\big(-2\vartheta\Phi(x)\big)$. Let X^ϑ be a solution of (5.29) with start in z and Y^ϑ be a stationary solution, both with respect to the same Brownian motion. Then we get the deterministic differential equation

$$\frac{d}{dt}(X_t^\vartheta - Y_t^\vartheta) = \vartheta\big(b(X_t^\vartheta) - b(Y_t^\vartheta)\big)$$

for the difference between the processes. First assume $X_0^\vartheta - Y_0^\vartheta \geq 0$. Because for $X_t^\vartheta - Y_t^\vartheta = 0$ the right hand side vanishes, the process $X_t^\vartheta - Y_t^\vartheta$ can never change its sign and stays positive. Since b is decreasing we have $b(X_t^\vartheta) - b(Y_t^\vartheta) \leq 0$ and we can conclude

$$0 \leq X_t^\vartheta - Y_t^\vartheta \leq X_0^\vartheta - Y_0^\vartheta.$$

For the case $X_0^\vartheta - Y_0^\vartheta \leq 0$ we can interchange the roles of X and Y to obtain the estimate

$$0 \leq Y_t^\vartheta - X_t^\vartheta \leq Y_0^\vartheta - X_0^\vartheta.$$

Combining these two cases gives

$$|Y_t^\vartheta - X_t^\vartheta| \leq |Y_0^\vartheta - X_0^\vartheta| = |Y_0^\vartheta - z|.$$

Using

$$\begin{aligned} |X_t^\vartheta - m| &\leq |X_t^\vartheta - Y_t^\vartheta| + |Y_t^\vartheta - m| \\ &\leq |z - Y_0^\vartheta| + |Y_t^\vartheta - m| \\ &\leq |z - m| + |Y_0^\vartheta - m| + |Y_t^\vartheta - m| \end{aligned}$$

Figure 5.4: *This figure illustrates the potential use of a comparison theorem for solutions of the SDE* (5.29)*. The thick line is the original drift b. The thin line is the new drift φ. The solution for drift b should be closer to m than the solution for drift φ.*

we can conclude

$$\begin{aligned} P\big(|X_t^\vartheta - m| > a\big) &\leq P\big(|Y_0^\vartheta - m| + |Y_t^\vartheta - m| > a - |z-m|\big) \\ &\leq P\Big(|Y_0^\vartheta - m| > \frac{a-|z-m|}{2}\Big) \\ &\qquad\qquad + P\Big(|Y_t^\vartheta - m| > \frac{a-|z-m|}{2}\Big) \\ &= 2P\Big(|Y_0^\vartheta - m| > \frac{a-|z-m|}{2}\Big). \end{aligned}$$

Now let $c > 0$. Then using theorem 2.13 we can find an $a > 0$ with

$$\lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log P\Big(|Y_0^\vartheta - m| > \frac{a-|z-m|}{2}\Big) \leq -c$$

and using the above estimate we get

$$\lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log P\big(|X_t^\vartheta - m| > a\big) \leq -c.$$

Since this is the exponential tightness condition (5.31) we can use lemma 2.1 to derive the full LDP and to complete the proof. (qed)

Remarks. 1) Note that in this case the rate function is independent of the interval length t. Because we have $\liminf_{|x|\to\infty} |b(x)| > 0$ the potential Φ converges to $+\infty$ for $|x| \to \infty$ and J_t is a good rate function. In fact the rate function coincides with the rate function of the LDP for the stationary distribution from theorem 2.13. This makes sense, because for strong drift we would expect the process to reach the equilibrium very quickly.

2) Using the assumptions on b we can find a monotonically decreasing, differentiable function $\varphi\colon \mathbb{R} \to \mathbb{R}$ which satisfies

$$|b(x)| \geq |\varphi(x)| \quad \text{for all } x \in \mathbb{R}$$

and has $\varphi'(m) < 0$. This is illustrated in figure 5.4.

Because the drift b pushes the process stronger towards m than the drift φ does, one could guess that when Y^ϑ is a solution of the SDE with drift $\vartheta\varphi$ instead of ϑb we would have $P\big(|X_t^\vartheta - m| > a\big) \leq P\big(|Y_t^\vartheta - m| > a\big)$. This would show that for finding an upper bound on $P\big(|X_t^\vartheta - m| > a\big)$ we could without loss of generality assume b to be monotonically decreasing.

Example 5.1. For the Ornstein-Uhlenbeck process we have $\Phi(x) = \alpha x^2/2$. Thus, using corollary (5.20) we get the rate function

$$J_t(x) = \alpha x^2$$

which coincides with the previous result from formula (3.4).

The case of repelling drift, i.e. of $b'(m) > 0$ is described in the following corollary.

Corollary 5.21. *Let $\Phi\colon \mathbb{R} \to \mathbb{R}$ be a C^3-function with bounded Φ'' and $b = -\Phi'$. Assume there is an $m \in \mathbb{R}$ with $b(x) = 0$ if and only if $x = m$, $b'(m) > 0$, and $\liminf_{|x|\to\infty} |b(x)| > 0$. Then for every $t > 0$ the solution X^ϑ of*

$$\begin{aligned} dX_s^\vartheta &= \vartheta b(X_s) \bullet ds + dB_s \quad \text{for } s \in [0;t], \text{ and} \\ X_0^\vartheta &= z \in \mathbb{R} \end{aligned}$$

satisfies the weak LDP on $\mathbb{R}$ with constant rate function

$$J_t(x) = 2\big(\Phi(m) - \Phi(z)\big) - t\Phi''(m). \tag{5.32}$$

Proof. In the case $b'(m) > 0$ the point m is the maximum of Φ and because of $V_z^m(\Phi) = \Phi(m) - \Phi(z)$, $V_m^x(\Phi) = \Phi(m) - \Phi(x)$ and $\Phi''(m) < 0$ we get

$$\begin{aligned} J_t(x) &= \big(\Phi(m) - \Phi(z)\big) - \Phi(z) - t\Phi''(m) + \big(\Phi(m) - \Phi(x)\big) + \Phi(x) \\ &= 2\big(\Phi(m) - \Phi(z)\big) - t\Phi''(m) \end{aligned}$$

for all $x \in \mathbb{R}$. (qed)

Remarks. The corollary shows that in the case of repelling drift the rate function does not depend on x. In particular it is not a good rate function. Also in this case it is impossible to strengthen the weak LDP to the full LDP because we have

$$\lim_{\vartheta\to\infty} \frac{1}{\vartheta} \log P(X_t^\vartheta \in \mathbb{R}) = 0 \quad \neq \quad 2\big(\Phi(m) - \Phi(z)\big) - t\Phi''(m).$$

Example 5.2. For $\vartheta > 0$ consider the solution of the SDE

$$\begin{aligned} dX_t^\vartheta &= \vartheta X_t^\vartheta \bullet dt + dB_t, \\ X_0^\vartheta &= z \in \mathbb{R}. \end{aligned} \tag{5.33}$$

This time the equilibrium point 0 is unstable, as soon as the process leaves 0 the drift will drive it further and further away. Here we have $\Phi(x) = -x^2/2$ and using corollary 5.21 we can determine the rate function for the large deviation behaviour of X_t^ϑ as

$$J_t(x) = t + z^2. \tag{5.34}$$

As in the case of the Ornstein-Uhlenbeck process (see page 24) we can explicitly determine the distribution of X_t^ϑ and verify the somewhat surprising result 5.34 manually. Since the derivation of formula (3.2) did not depend on the sign of the drift we get

$$X_t = e^{\vartheta t} z + \int_0^t e^{\vartheta(t-s)}\, dB_s. \tag{5.35}$$

Because the SDE (5.33) is linear, we know that X_t^ϑ has a Gaussian distribution and from (5.35) we find the expectation

$$\mu = E(X_t^\vartheta) = e^{\vartheta t} z$$

and the variance

$$\begin{aligned}\sigma^2 = E\big((X_t^\vartheta - \mu)^2\big) &= E\Big(\big(\int_0^t e^{\vartheta(t-s)}\,dB_s\big)^2\Big)\\ &= \int_0^t e^{2\vartheta(t-s)}\,ds = \frac{1}{2\vartheta}(e^{2\vartheta t}-1).\end{aligned}$$

Thus for every measurable set $A \subseteq \mathbb{R}$ we have

$$\begin{aligned}P(X_t^\vartheta \in A) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_A \exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)\,dx\\ &= \sqrt{\frac{\vartheta}{\pi}}\big(e^{2\vartheta t}-1\big)^{-1/2}\int_A \exp\Big(-\vartheta\frac{(e^{-\vartheta t}x - z)^2}{1-e^{-2\vartheta t}}\Big)\,dx.\end{aligned}$$

Now we can calculate the exponential rates of this expression for $\vartheta \to \infty$. For the normalising constant we find

$$\lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log\sqrt{\frac{\vartheta}{\pi}}\big(e^{2\vartheta t}-1\big)^{-1/2} = -\frac{1}{2}2t = -t.$$

If $A \subseteq \mathbb{R}$ is bounded, then

$$\sup_{x\in A}\Big|\frac{(e^{-\vartheta t}x - z)^2}{1-e^{-2\vartheta t}} - z^2\Big| \longrightarrow 0$$

for $\vartheta \to \infty$ and thus for compact sets $K \subseteq \mathbb{R}$ we find

$$\begin{aligned}&\lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log P(X_t^\vartheta \in K)\\ &\qquad = -t + \lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log\int_K \exp(-z^2)\,dx = -(t+z^2).\end{aligned}$$

For open sets $O \subseteq \mathbb{R}$ we can choose any bounded subset $A \subseteq O$ with non-zero Lebesgue measure, to get

$$\begin{aligned}&\liminf_{\vartheta\to\infty}\frac{1}{\vartheta}\log P(X_t^\vartheta \in O)\\ &\qquad \geq \liminf_{\vartheta\to\infty}\frac{1}{\vartheta}\log P(X_t^\vartheta \in A)\\ &\qquad = -t + \lim_{\vartheta\to\infty}\frac{1}{\vartheta}\log\int_A \exp(-z^2)\,dx = -(t+z^2).\end{aligned}$$

This reproduces the result from the corollary.

Chapter 6

Asymptotic Separation of Processes

One application of large deviation results is to determine the exponential decay rate of the Bayes risk for the separation of two processes which are observed over long intervals of time. This rate is a measure of how easy it is to distinguish between two processes while only looking at the paths.

6.1 The Bayes Risk

We consider two mechanisms to generate a stochastic process, e.g. two different drift fields, which are indexed by a parameter $\vartheta \in \Theta = \{0, 1\}$. For $\vartheta \in \Theta$ let $P_t^\vartheta = \mathcal{L}(X)\big|_{\mathcal{F}_t}$ be the law of the corresponding process X^ϑ observed up to time t.

The **Bayes risk** $B(\lambda, t)$ with a priory distribution $\lambda \in \mathrm{Prob}(\Theta)$ is defined by

$$B(\lambda, t) = \int \min_{\vartheta \in \Theta} \frac{d\lambda_\vartheta P_t^\vartheta}{dP_t} \, dP_t$$

where $P_t = \lambda_1 P_t^1 + \lambda_0 P_t^0$. Assuming that the distribution $\mathcal{L}(X_t)\big|_{\mathcal{F}_t}$ on the path space has a density φ_t^ϑ with respect to Wiener measure $\mathbb{W}$ we find

$$B(\lambda, t) = \lambda_1 P_t^1(\lambda_1 \varphi_t^1 < \lambda_0 \varphi_t^0) + \lambda_0 P_t^0(\lambda_1 \varphi_t^1 \geq \lambda_0 \varphi_t^0). \tag{6.1}$$

Thus $B(\lambda, t)$ can be seen as the total probability of error for a likelihood ratio test for the parameter ϑ, where ϑ is chosen randomly according to the distribution λ. Figure 6.1 gives a geometric interpretation of the Bayes risk in this situation.

We want to calculate the exponential decay rates

$$\lim_{t \to \infty} \frac{1}{t} \log B(\lambda, t) \tag{6.2}$$

of the Bayes risk. This rate is measure of how fast the Bayes risk decays when the observation time t tends to infinity. Thus it is a measure of how easy the two processes can be separated by looking only at the paths. A very negative Bayes risk indicates, that the processes are very different and thus easy to distinguish, a Bayes risk that is close to zero indicates processes which are similar.

In corollary 5.6 of [KW97] is is shown that if the limit in (6.2) exists, then it does not depend on λ (except for the pathologic cases $\lambda_0 = 0$ or $\lambda_1 = 0$). So we can choose for example $\lambda = (1/2, 1/2)$ to calculate the rate, which gives

$$B(t) := B\big((1/2, 1/2), t\big) = \frac{1}{2} P_t^0(\varphi_t^0 \leq \varphi_t^1) + \frac{1}{2} P_t^1(\varphi_t^0 > \varphi_t^1)$$

Figure 6.1: *This figure illustrates the geometric interpretation of the Bayes risk. The two curves are the weighted densities of the distributions on the probability space Ω. The Bayes risk is the size of the hatched area. It is large if the distributions are similar.*

and using lemma 2.2 we get

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{t} \log B(\lambda, t) = \max\Bigl(&\lim_{t\to\infty} \frac{1}{t} \log P_t^0(\varphi_t^0 \leq \varphi_t^1), \\ &\lim_{t\to\infty} \frac{1}{t} \log P_t^1(\varphi_t^0 > \varphi_t^1)\Bigr)\end{aligned} \tag{6.3}$$

for all $\lambda \in \mathrm{Prob}(\vartheta)$.

Here we consider two reversible diffusions with different drift fields. For $\vartheta \in \Theta = \{0, 1\}$ let $\Phi^\vartheta \colon \mathbb{R}^d \to \mathbb{R}$ be two time continuously differentiable, $b^\vartheta = -\operatorname{grad} \Phi^\vartheta$, and X^ϑ be a solution of the SDE

$$\begin{aligned} dX^\vartheta &= b^\vartheta(X^\vartheta) \bullet dt + dB \\ X_0^\vartheta &= 0. \end{aligned} \tag{6.4}$$

From lemma 1.5 we know the explicit form of the densities φ_t^ϑ. They can be defined by

$$\varphi_t^\vartheta(\omega) = \exp\Bigl(-\Phi^\vartheta(\omega_t) + \Phi^\vartheta(0) - \int_0^t v^\vartheta(\omega_s)\, ds\Bigr)$$

for all $\omega \in C\bigl([0;\infty), \mathbb{R}^d\bigr)$ where $v^\vartheta = \bigl((\nabla\Phi^\vartheta)^2 - \Delta\Phi^\vartheta\bigr)/2$. Because the exponential function is monotonically increasing, the event $\varphi_t^0(X^0) \leq \varphi_t^1(X^0)$ in equation (6.3) can be expressed as

$$-\Phi^0(X_t^0) + \Phi^0(X_0^0) - \int_0^t v^0(X_s^0)\, ds \leq -\Phi^1(X_t^0) + \Phi^1(X_0^0) - \int_0^t v^1(X_s^0)\, ds$$

or equivalently as

$$\frac{1}{t}\int_0^t \bigl(v^1 - v^0\bigr)(X_s^0)\, ds + \frac{1}{t}\bigl(\Phi^1 - \Phi^0\bigr)(X_t^0) - \frac{1}{t}\bigl(\Phi^1 - \Phi^0\bigr)(X_0^0) \leq 0. \tag{6.5a}$$

The opposite event $\varphi_t^0(X^1) > \varphi_t^1(X^1)$ becomes

$$\frac{1}{t}\int_0^t \bigl(v^1 - v^0\bigr)(X_s^1)\, ds + \frac{1}{t}\bigl(\Phi^1 - \Phi^0\bigr)(X_t^1) - \frac{1}{t}\bigl(\Phi^1 - \Phi^0\bigr)(X_0^1) > 0. \tag{6.5b}$$

In order to calculate the rate for the Bayes risk we have to consider large deviations for the events (6.5a) and (6.5b). In general this is a difficult problem. The following example illustrates the procedure for a very simple case.

Example 6.1 (Constant drift). Assume that we have fixed vectors $b^0, b^1 \in \mathbb{R}^d$ with $b^\vartheta(x) = b^\vartheta$ for all $x \in \mathbb{R}^d$. From example 1.1 we know that we have $\Phi^\vartheta(x) = -b^\vartheta \cdot x$ and $v^\vartheta(x) = |b^\vartheta|^2/2$ here. Thus the densities φ^ϑ only depend on the endpoint X_t^ϑ of the path and we get

$$\varphi_t^0(X^0) \leq \varphi_t^1(X^0) \quad \Longleftrightarrow \quad \frac{b^1 + b^0}{2} \cdot (b^1 - b^0) - \frac{1}{t} X_t^0 \cdot (b^1 - b^0) \leq 0.$$

Because the drift is constant b^ϑ the value X_t^0 is $\mathcal{N}(t\cdot b^0, t)$-distributed, thus X_t^0/t is $\mathcal{N}(b^0, 1/t)$-distributed and the above condition means that X_t^0/t is contained in the half-plane with normal vector $b^1 - b^0$ which does not contain the vector b^0. Corollary 2.12 gives

$$\lim_{t\to\infty} \frac{1}{t} \log P\big(\varphi_t^0(X^0) \leq \varphi_t^1(X^0)\big) = -\frac{1}{2}\Big|\frac{b^1 + b^0}{2}\Big|^2 = -\frac{|b^1 - b^0|^2}{8}$$

and a very similar calculation also shows

$$\lim_{t\to\infty} \frac{1}{t} \log P\big(\varphi_t^0(X^1) > \varphi_t^1(X^1)\big) = -\frac{|b^1 - b^0|^2}{8}.$$

Thus both rates from the right hand side of (6.3) coincide and we get the result

$$\lim_{t\to\infty} \frac{1}{t} \log B(\lambda, t) = -\frac{|b^1 - b^0|^2}{8}.$$

Details about this can be found in [Voß97].

6.2 Asymptotic Separation of OU Processes

In this section we determine the exponential rate for the decay of the Bayes risk when distinguishing two Ornstein-Uhlenbeck processes with different parameters α_0 and α_1.

As we have seen in chapter 3 the density of a d-dimensional Ornstein-Uhlenbeck process with parameter α_ϑ on the path space is

$$\varphi_t^\vartheta(\omega) = \exp\Big(-\alpha_\vartheta \frac{\omega_t^2}{2} - \frac{1}{2}\int_0^t \alpha_\vartheta^2 \omega_s^2 - \alpha_\vartheta \, ds\Big).$$

The representation of the event $\varphi_t^0(X^0) \leq \varphi_t^1(X^0)$ from (6.5a) becomes

$$\frac{\alpha_1^2 - \alpha_0^2}{2} \frac{1}{t} \int_0^t (X_s^0)^2 \, ds - \frac{\alpha_1 - \alpha_0}{2} d + (\alpha_1 - \alpha_0) \frac{(X_t^0)^2}{2t} \leq 0$$

and assuming $\alpha_0 > \alpha_1 > 0$ we can divide by $\alpha_1 - \alpha_0 < 0$ to get the condition

$$\frac{\alpha_1 + \alpha_0}{2} \frac{1}{t} \int_0^t (X_s^0)^2 \, ds - \frac{d}{2} + \frac{(X_t^0)^2}{2t} \geq 0. \tag{6.6a}$$

The distribution of the process X converges to a d-dimensional Gaussian distribution with expectation 0 and covariance matrix $\frac{1}{2\alpha} I_d$. Almost surely we have

$$\frac{1}{t}\int_0^t (X_s)^2 \, ds \to \frac{d}{2\alpha} \quad \text{and} \quad \frac{(X_t)^2}{t} \to 0$$

for $t \to \infty$, so the left hand side of (6.6a) a.s. converges to

$$\frac{\alpha_1 + \alpha_0}{2} \frac{d}{2\alpha_0} - \frac{d}{2} + 0 < \frac{2\alpha_0}{2} \frac{d}{2\alpha_0} - \frac{d}{2} = 0$$

and we can see that the probability of the event (6.6a) at least converges to 0.

Similarly we find that for $\alpha_0 > \alpha_1 > 0$ the event $\varphi_t^0(X^1) > \varphi_t^1(X^1)$ is equivalent to

$$\frac{\alpha_1 + \alpha_0}{2} \frac{1}{t} \int_0^t (X_s^1)^2 \, ds - \frac{d}{2} + \frac{(X_t^1)^2}{2t} < 0. \tag{6.6b}$$

The following theorem states the main result of the section.

Figure 6.2: *This figure sketches the rate function I_c for Y_t/t from formula (6.7). The process converges to $-(c \cdot 1/2\alpha + 1/2) = (\alpha_0 - \alpha_1)/4\alpha_1$. We will consider the event $Y_t/t < 0$.*

Theorem 6.1. *The exponential decay rate of the Bayes risk $B(\lambda, t)$ for the distinction between two one-dimensional Ornstein-Uhlenbeck processes with parameters $\alpha_0, \alpha_1 > 0$ is*

$$\lim_{t\to\infty} \frac{1}{t} \log B(\lambda, t) = -\frac{(\alpha_1 - \alpha_0)^2}{8(\alpha_1 + \alpha_0)}.$$

for every a priory distribution $\lambda \in \mathrm{Prob}\big(\{0,1\}\big)$ with $\lambda_0 \neq 0$ and $\lambda_1 \neq 0$.

Proof. We calculate the two rates R_1 and R_0 in the maximum from formula (6.3) separately. Without loss of generality we may assume $\alpha_0 > \alpha_1 > 0$ and using equation (6.6b) we get

$$\begin{aligned} R_1 &:= \lim_{t\to\infty} \frac{1}{t} \log P_t^1\left(\varphi_t^0 > \varphi_t^1\right) \\ &= \lim_{t\to\infty} \frac{1}{t} \log P_t^1\Big(\frac{\alpha_1 + \alpha_0}{2}\frac{1}{t}\int_0^t X_s^2\, ds - \frac{1}{2} + \frac{X_t^2}{2t} < 0\Big) \\ &= \lim_{t\to\infty} \frac{1}{t} \log P_t^1\Big(\frac{X_t^2}{2t} - \frac{1}{t}\int_0^t -\frac{\alpha_1 + \alpha_0}{2} X_s^2 + \frac{1}{2}\, ds < 0\Big) \end{aligned}$$

The large deviation behaviour of the random variables $Y_t(c)/t$ with

$$Y_t(c) = \frac{1}{2}X_t^2 - \int_0^t cX_s^2 + \frac{1}{2}\, ds$$

where X is an Ornstein Uhlenbeck process is examined by Florens-Landais and Pham in [FLP99]. We use the first case of theorem 2.2 from their article: Let X be an Ornstein-Uhlenbeck process with parameter α. Then for every $c \leq -\alpha/2$ the family $Y_t(c)/t$ satisfies the large deviation principle on $\mathbb{R}$ with the good rate function I_c defined by

$$I_c(y) = \begin{cases} -\dfrac{\alpha^2}{c}\, \dfrac{(y + \frac{c+\alpha}{2\alpha})^2}{2y+1} & \text{if } y > -\frac{1}{2}, \text{ and} \\ +\infty & \text{else.} \end{cases} \tag{6.7}$$

The rate function for this case is sketched in figure 6.2.

Since in our situation we have $\alpha = \alpha_1$ and $c = -(\alpha_1 + \alpha_0)/2$ the rate function is decreasing to the left of $-(c+\alpha)/2\alpha = (\alpha_0 - \alpha_1)/4\alpha_1 > 0$ and is increasing to the right of this point. Because $(-\infty; 0)$ is a continuity set of I_c we get

$$\begin{aligned} R_1 &= \lim_{t\to\infty} \frac{1}{t} \log P_t^1\Big(\frac{Y_t(c)}{t} < 0\Big) \\ &= -I_c(0) = -\frac{(c+\alpha)^2}{4c} = -\frac{(\alpha_0 - \alpha_1)^2}{8(\alpha_0 + \alpha_1)}. \end{aligned}$$

Now we have to calculate the other rate from formula (6.3). Assuming $\alpha_0 > \alpha_1 > 0$ again, we get

$$\begin{aligned} R_0 &= \lim_{t\to\infty} \frac{1}{t} \log P_t^0\big(\varphi_t^0 \le \varphi_t^1\big) \\ &= \lim_{t\to\infty} \frac{1}{t} \log P_t^0\Big(\frac{Y_t(c)}{t} \ge 0\Big), \end{aligned}$$

but this time with $\alpha = \alpha_0$ and $c = -(\alpha_1 + \alpha_0)/2$. The rate function is decreasing to the left of $-(c+\alpha)/2\alpha = (\alpha_1 - \alpha_0)/4\alpha_0 < 0$ and is increasing to the right of this point. So we get again the result

$$R_0 = -I_c(0) = -\frac{(\alpha_0 - \alpha_1)^2}{8(\alpha_0 + \alpha_1)}.$$

This proves that both terms in the maximum on the right hand side of formula (6.3), and thus the maximum itself, are equal to the rate from our claim. (qed)

6.3 Asymptotic Separation of Continuous Time Markov Chains

In this section we determine the exponential decay rate for the Bayes risk when separating two continuous time Markov chains. These processes are no diffusion processes in the sense of chapter 1, but the concept of the Bayes risk of course also makes sense here. We will see, that the exponential rate of the Bayes risk is $e^r - 1$, where r is the rate for separation of the embedded Markov chains.

Let X be a continuous time Markov chain with finite state space S and generator $q \in \mathbb{R}^{S\times S}$. For the technical details about continuous time Markov chains and their generators we refer to [Law95]. Because q is a generator we have $q_{ij} \ge 0$ for $i \ne j$ and $q_{ii} = -\sum_{j\ne i} q_{ij} < 0$ for all $i \in S$. The process X can be described with the help of an embedded Markov chain Y: whenever X reaches a state $i \in S$, the process stays there for an $\mathrm{Exp}(-q_{ii})$-distributed time and then jumps into a randomly chosen new state. The new state is $j \ne i$ with probability $q_{ij}/\sum_{k\ne i} q_{ik}$.

Here we restrict ourselves to the case $q_{ii} = -1$ for all $i \in S$, i.e. to the case of equal and homogeneous jump rates. Let $T(t)$ be the number of jumps up to time t and Y_n for $n \in \mathbb{N}_0$ be the state of X after the nth jump. Then $T(t)$ is Poisson distributed with parameter t and Y is a Markov chain with transition matrix

$$\pi_{ij} = \begin{cases} q_{ij}, & \text{if } i \ne j \text{, and} \\ 0 & \text{else.} \end{cases}$$

Now consider two irreducible Markov chains X^0 and X^1 with different transition rates q^0 and q^1. Our task is to observe one path and to determine which transition mechanism generated this path.

Distinguishing two non-equivalent Markov chains is easy: as soon as a transition occurs, which is only possible for one chain but not for the other, we have identified the transition mechanism with probability one. So here we assume that the processes are equivalent, i.e. we consider the case $q_{ij}^0 > 0 \Leftrightarrow q_{ij}^1 > 0$.

Again we consider the Bayes risk

$$B(t) = \frac{1}{2} P_t^0\Big(\frac{dP_t^1}{dP_t^0} \ge 1\Big) + \frac{1}{2} P_t^1\Big(\frac{dP_t^1}{dP_t^0} \le 1\Big), \tag{6.8}$$

where P_t^ϑ is the distribution of the path $X_s^\vartheta\big|_{0\le s\le t}$. We want to calculate the exponential decay rate $\lim_{t\to\infty} \frac{1}{t} \log B(t)$. Using lemma 2.2 we can reduce this problem to calculation the rates for the individual terms in the sum (6.8), as we did in (6.3).

Because the processes have coinciding jump rates, the whole information about ϑ is contained in the transition frequencies between the different states. Thus define the empirical pair measure $\hat{m}^n$ by $\hat{m}^n_{ij} = \frac{1}{n}\sum_{k=1}^n 1_{\{(i,j)\}}(Y_{k-1}, Y_k)$ for all $i, j \in S$. Then we have

$$\frac{dP^1_t}{dP^0_t} = \frac{\pi^1_{Y_0Y_1}\pi^1_{Y_1Y_2}\cdots\pi^1_{Y_{N(t)-1}Y_{N(t)}}}{\pi^0_{Y_0Y_1}\pi^0_{Y_1Y_2}\cdots\pi^0_{Y_{N(t)-1}Y_{N(t)}}} = \prod_{i,j\in S}\left(\frac{\pi^1_{ij}}{\pi^0_{ij}}\right)^{N(t)\cdot\hat{m}^{N(t)}_{ij}},$$

where we use the convention $(0/0)^0 = 1$. Defining

$$A^1 = \big\{a \in \mathbb{R}^{S\times S} \mid \sum_{i,j} a_{ij}\log(\pi^1_{ij}/\pi^0_{ij}) \le 0\big\}$$

we can express the probabilities from (6.8) as

$$P^0_t\Big(\frac{dP^1_t}{dP^0_t} \ge 1\Big) = P^0_t\Big(\sum_{i,j\in S}\hat{m}^{N(t)}_{ij}\log\frac{\pi^1_{ij}}{\pi^0_{ij}} \ge 0\Big) = P^0_t\big(\hat{m}^{N(t)} \in A^1\big).$$

Because the jumps of the embedded Markov chain Y are independent of the jumping times, and because $N(t)$ is Poisson distributed, we find

$$P^0_t\Big(\frac{dP^1_t}{dP^0_t} \ge 1\Big) = \sum_{n=0}^{\infty} P\big(N(t) = n\big)P^0\big(\hat{m}^n \in A^1\big) = \sum_{n=0}^{\infty} \mathrm{e}^{-t}\frac{t^n}{n!}P^0\big(\hat{m}^n \in A^1\big).$$

From Peter Scheffel's thesis [Sch97] we know the exponential rate for the separation of two Markov chains. It can be expressed by the spectral radius ρ (i.e. by the maximum of the absolute values of the eigenvalues) of the matrices $\pi^{(\lambda)}$ with $\pi^{(\lambda)}_{i,j} = (\pi^1_{ij})^{\lambda}(\pi^0_{ij})^{(1-\lambda)}$ for all $\lambda \in [0; 1]$. For irreducible, equivalent Markov chains the following result holds true:

$$\lim_{n\to\infty}\frac{1}{n}\log P^0\big(\hat{m}^n \in A^1\big) = \log\inf_{0<\lambda<1}\rho(\pi^{(\lambda)}).$$

With the help of the following elementary lemma we can transfer this result to our situation.

Lemma 6.2. *Let (a_n) be a sequence of positive real numbers. Then*

$$\liminf_{t\to\infty}\frac{1}{t}\log\sum_{n=1}^{\infty}\mathrm{e}^{-t}\frac{t^n}{n!}a_n \ge \exp\Big(\liminf_{n\to\infty}\frac{1}{n}\log a_n\Big) - 1$$

and

$$\limsup_{t\to\infty}\frac{1}{t}\log\sum_{n=1}^{\infty}\mathrm{e}^{-t}\frac{t^n}{n!}a_n \le \exp\Big(\limsup_{n\to\infty}\frac{1}{n}\log a_n\Big) - 1.$$

Proof. According to Stirling's formula we have

$$n! \sim \sqrt{2\pi}\,\frac{n^{n+1/2}}{\mathrm{e}^n},$$

where $\sim$ indicates, that the quotient of both sides converges to 1 as $n \to \infty$. Thus we get

$$\begin{aligned}\mathrm{e}^{-t}\frac{t^n}{n!}a_n &\sim \frac{\mathrm{e}^{-t}t^n\mathrm{e}^n}{n^n}\,\frac{a_n}{\sqrt{2\pi n}}\\ &= \exp\big(-t - n\log n + n\log t + n + \log\frac{a_n}{\sqrt{2\pi n}}\big)\\ &= \exp\Big(t\big(-1 - \frac{n}{t}\log\frac{n}{t} + \frac{n}{t} + \frac{n}{t}\,c_n\big)\Big)\end{aligned} \tag{6.9}$$

where
$$c_n = \frac{1}{n} \log \frac{a_n}{\sqrt{2\pi n}}.$$
Here the quotient of both sides does not depend on t, i.e. the convergence for $n \to \infty$ is uniform in t.

Now I want to express the right hand side as a function of n/t. In order to do so define the function g by
$$g_c(x) = -1 - x \log x + x + x \cdot c$$
for all $x > 0$. This function is monotonically increasing in c. We will use it to get bounds on (6.9) in the situation when the sequence $(c_n)_{n \in \mathbb{N}}$ is bounded. Since
$$g_c'(x) = -\log x - \frac{x}{x} + 1 + c = c - \log x,$$
the derivative g_c' is strictly decreasing with a zero at $x = e^c$. Thus g_c attains its global maximum at the point e^c. The value of the maximum is
$$g_c(e^c) = -1 - e^c \log e^c + e^c + e^c \cdot c = e^c - 1.$$

Now let $a = \liminf_{n\to\infty} \frac{1}{n} \log a_n$ and $\varepsilon > 0$. Then there is an $N \in \mathbb{N}$ with
$$\frac{1}{n} \log \frac{a_n}{\sqrt{2\pi n}} > a - \varepsilon \quad \text{for all } n \geq N.$$
Because lemma 2.2 only applies to finite sums, we will split the infinite sum from the claim into the summands for $n = 1, 2, \ldots, N-1$ and the remaining tail. For $n < N$ we find
$$\lim_{t\to\infty} \frac{1}{t} \log e^{-t} \frac{t^n}{n!} a_n = -1 + \lim_{t\to\infty} \frac{n}{t} \log t + \lim_{t\to\infty} \frac{1}{t} \log \frac{a_n}{n!} = -1 + 0 + 0 = -1$$
and with $n_t = \lfloor e^{a-\varepsilon} \cdot t \rfloor$ we conclude for the tail
$$\begin{aligned}
&\liminf_{t\to\infty} \frac{1}{t} \log \sum_{n=N}^{\infty} e^{-t} \frac{t^n}{n!} a_n \\
&\geq \liminf_{t\to\infty} \frac{1}{t} \log\Big(e^{-t} \frac{t^{n_t}}{n_t!} a_{n_t}\Big) \\
&= \liminf_{t\to\infty} \frac{1}{t} \Big(t\big(-1 - \frac{n_t}{t} \log \frac{n_t}{t} + \frac{n_t}{t} + \frac{n_t}{t} \frac{1}{n_t} \log \frac{a_{n_t}}{\sqrt{2\pi n_t}}\big)\Big) \\
&\geq \liminf_{t\to\infty} g_{a-\varepsilon}\big(n_t/t\big) \\
&= g_{a-\varepsilon}\big(e^{a-\varepsilon}\big) \\
&= e^{a-\varepsilon} - 1 \quad \text{for all } \varepsilon > 0
\end{aligned}$$
i.e.
$$\liminf_{t\to\infty} \frac{1}{t} \log \sum_{n=N}^{\infty} e^{-t} \frac{t^n}{n!} a_n \geq e^a - 1.$$
Lemma 2.2 gives now the result. Because of $e^a - 1 > -1$ the first $N-1$ terms are not important and the rate is determined by the tail.

Now let $b = \limsup_{n\to\infty} \frac{1}{n} \log a_n$ and $\varepsilon > 0$. Then there is an $N \in \mathbb{N}$ with
$$\frac{1}{n} \log \frac{a_n n^2}{\sqrt{2\pi n}} < b + \varepsilon \quad \text{for all } n \geq N.$$
For the upper bound we have to consider all terms in the tail: multiplying the numerator and denominator of equation (6.9) with n^2 shows, that one can choose N large to obtain
$$e^{-t} \frac{t^n}{n!} a_n \leq \frac{1}{n^2} \exp\Big(t\big(-1 - \frac{n}{t} \log \frac{n}{t} + \frac{n}{t} + \frac{n}{t} \frac{1}{n} \log \frac{a_n n^2}{\sqrt{2\pi n}}\big)\Big) \cdot (1 + \varepsilon)$$

for all $n \geq N$. (Instead of $1/n^2$ we could have chosen any other summable sequence.) Then we have

$$\begin{aligned}
&\limsup_{t\to\infty} \frac{1}{t} \log \sum_{n=N}^{\infty} e^{-t} \frac{t^n}{n!} a_n \\
&\quad \leq \limsup_{t\to\infty} \frac{1}{t} \log \sum_{n=N}^{\infty} \frac{1}{n^2} \exp\bigl(t g_{b+\varepsilon}(n/t)\bigr)(1+\varepsilon) \\
&\quad \leq \limsup_{t\to\infty} \frac{1}{t} \log\Bigl(\bigl(\sum_{n=N}^{\infty} \frac{1}{n^2}\bigr) \exp\bigl(t g_{b+\varepsilon}(e^{b+\varepsilon})\bigr)(1+\varepsilon)\Bigr) \\
&\quad = e^{b+\varepsilon} - 1 \quad \text{for all } \varepsilon > 0,
\end{aligned}$$

i.e.

$$\limsup_{t\to\infty} \frac{1}{t} \log \sum_{n=N}^{\infty} e^{-t} \frac{t^n}{n!} a_n \leq e^b - 1.$$

Using lemma 2.2 again, we also get the second part of the claim. (qed)

With the help of the lemma we get the main result of this section.

Theorem 6.3. *Let X^0, X^1 be two irreducible, equivalent continuous time Markov chains with finite state space S and generators $q^0, q^1 \in \mathbb{R}^{S\times S}$. Further assume $q^{\vartheta}_{ii} = -1$ for all $i \in S$. For $\lambda \in [0;1]$ let the matrix $\pi^{(\lambda)}$ be defined as above and let ρ denote the spectral radius. Then the Bayes risk $B(t)$ for the separation of X^0 and X^1 has exponential decay rate*

$$\lim_{t\to\infty} \frac{1}{t} \log B(t) = \inf_{0<\lambda<1} \rho(\pi^{(\lambda)}) - 1.$$

Proof. Substituting Peter Scheffel's result for Markov chains into lemma 6.2 gives

$$\begin{aligned}
\lim_{t\to\infty} \frac{1}{t} \log P^0_t\Bigl(\frac{dP^1_t}{dP^0_t} \geq 1\Bigr) &= \exp\Bigl(\log \inf_{0<\lambda<1} \rho(\pi^{(\lambda)})\Bigr) - 1 \\
&= \inf_{0<\lambda<1} \rho(\pi^{(\lambda)}) - 1.
\end{aligned}$$

Analogous one also gets

$$\begin{aligned}
\lim_{t\to\infty} \frac{1}{t} \log P^1_t\Bigl(\frac{dP^1_t}{dP^0_t} \leq 1\Bigr) &= \exp\Bigl(\log \inf_{0<\lambda<1} \rho(\pi^{(\lambda)})\Bigr) - 1 \\
&= \inf_{0<\lambda<1} \rho(\pi^{(\lambda)}) - 1.
\end{aligned}$$

Using lemma 2.2 finished the proof. (qed)

First notice that, in contrast to the Markov chain case, the rate cannot drop below -1. On a closer look this is not surprising. As we saw in the proof of lemma 6.2, the value -1 is just the exponential rate for the event, that the process has no jump (or at most N jumps) up to time t. In this case of course we cannot gather any information to distinguish between the two processes.

Thus the situation is as follows: when the processes are very similar, then the rate ρ for the separation of the embedded Markov chains is close to 0 and the rate for the continuous time case is $e^{\rho} - 1 \approx \rho$, i.e. it is mostly determined by the transition mechanism. In the case of very different Markov chains, on the other hand, ρ is significantly smaller then 0 and for continuous time we get the rate $e^{\rho} - 1 \approx -1$. Here the rate is mainly determined by the jump mechanism.

Figure 6.3: *The diagram shows one path of a continuous time Markov chain, generated by the transition mechanism from example 6.2. The question is, whether the process has generator q^0 or q^1.*

Example 6.2. This example demonstrates, that given the transition matrices it is easy to explicitly calculate the exponential decay rates for the Bayes risk. Consider the transition matrices

$$\pi^1 = \begin{pmatrix} 0 & 1/3 & 2/3 \\ 2/3 & 0 & 1/3 \\ 1/3 & 2/3 & 0 \end{pmatrix}$$

and

$$\pi^0 = \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 0 & 2/3 \\ 2/3 & 1/3 & 0 \end{pmatrix}.$$

Defining $\pi^{(\lambda)}$ as above we get the spectral radius $\rho(\pi^{(\lambda)}) = (2^{1-\lambda} + 2^{\lambda})/3$ and consequently $\inf_{0<\lambda<1} \rho(\pi^{(\lambda)}) = \rho(\pi^{1/2}) = \sqrt{8}/3$. From Peter Scheffel's result we get the decay rate of the Bayes risk for the separation of the two corresponding discrete time Markov chains:

$$\lim_{n\to\infty} \frac{1}{n} \log B(\lambda, n) = \log(\sqrt{8}/3) \approx -0.059.$$

The generators for the corresponding continuous time Markov chains are $q^\vartheta = \pi^\vartheta - I$ for $\vartheta \in \{0, 1\}$. Figure 6.3 illustrates one instantiation of this process. From theorem 6.3 we know the rate for the separation of the continuous time Markov chains. It is

$$\lim_{t\to\infty} \frac{1}{t} \log B(t) = \inf_{0<\lambda<1} \rho(\pi^{(\lambda)}) - 1 = \sqrt{8}/3 - 1 \approx -0.057.$$

Chapter 7

Computational Experiments

In the process of understanding complicated stochastic mechanisms computer simulations can be a useful tool. The area of large deviation problems places special challenges here. Because the whole point of large deviation problems is to handle extremely small probabilities, naive approaches tend to fail here. Some solutions to the resulting problems are illustrated in the following examples.

The easy way to estimate the probability of an event is to generate many random samples, to count the number of occurrences of the event in question, and finally to use the law of large numbers to estimate the probability with the relative frequency. This works well if the probability is reasonably large. But if n is the maximum number of samples the computer will generate in the time we are willing to wait, this method won't work for probabilities smaller than $1/n$, because typically we would observe no occurrences of the event.

Another problem occurs when one tries to sample according to a conditional distribution. The easy way to do this is to just sample from the full distribution, and to reject every value which does not meet the condition. But then, again, this will only work if the probability is not too small, because otherwise we will just have to reject every sample and get no values which meet the condition.

This chapter presents some methods which are useful to overcome these problems.

7.1 The Euler-Maruyama Method

The basic method to numerically solve stochastic differential equations is the Euler-Maruyama␣method or stochastic␣Euler ␣␣method. The method is, for example, described in [KP99]. We can use this method to generate random paths from the solution of a SDE. The results of the subsequent sections can then be used to estimate probabilities or to sample from conditional distributions.

Consider the stochastic Differential equation

$$\begin{aligned} dX_t &= b(X_t, t) \bullet dt + \sigma(X_t, t) \bullet dB_t \quad \text{for } 0 < t \le T \\ X_0 &= z \in \mathbb{R}^d, \end{aligned} \tag{7.1}$$

where B is a d-dimensional Brownian motion, $b\colon \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ is some drift function, and $\sigma\colon \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^{d\times n}$ is the diffusion coefficient.

Solutions to this method can be approximated as follows: let $N \in \mathbb{N}$ and $\Delta t = T/N$. Then define

$$\tilde{X}_0 = z$$

and iteratively

$$\tilde{X}_n = \tilde{X}_{n-1} + b\big(\tilde{X}_{n-1}, (n-1)\Delta t\big) \cdot \Delta t + \sigma\big(\tilde{X}_{n-1}, (n-1)\Delta t\big)\xi_n \cdot \sqrt{\Delta t}$$

for $n = 1, \ldots, N$, where $\xi_1, \ldots, \xi_N$ are d-dimensional, i.i.d. standard normal random variables. Then the distribution of $\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_N$ is an approximation for the distribution of the values $X_0, X_{1\cdot\Delta t}, \ldots, X_{N\cdot\Delta t}$.

One of the basic results about this method is the following theorem, which is a direct consequence of theorem 10.2.2 from [KP99].

Theorem 7.1. *Let X be a solution of (7.1) for a Brownian motion B. Define $\xi_n = (B_{n\Delta t} - B_{(n-1)\Delta t})/\sqrt{\Delta t}$ and let $\tilde{X}_n$, $n = 0, \ldots, N$ be defined by the Euler-Maruyama method with step size $\Delta t = T/N$ as above. Furthermore assume that*

$$\begin{gathered} \bigl|b(x,t) - b(y,t)\bigr| + \bigl|\sigma(x,t) - \sigma(y,t)\bigr| \leq K_1|x-y| \\ \bigl|b(x,t)\bigr| + \bigl|\sigma(x,t)\bigr| \leq K_2\bigl(1+|x|\bigr) \\ \bigl|b(x,s) - b(x,t)\bigr| + \bigl|\sigma(x,s) - \sigma(x,t)\bigr| \leq K_3\bigl(1+|x|\bigr)|s-t|^{1/2} \end{gathered}$$

for all $x, y \in \mathbb{R}^d$ and all $s, t \in [0;T]$, where the constants K_1, K_2, and K_3 do not depend on N. Then there is a constant K_4, which is also independent of N, such that the Euler approximation $\tilde{X}$ satisfies

$$E\bigl(|X_T - \tilde{X}_N|\bigr) \leq K_4 \Delta t^{1/2}.$$

The theorem shows, that the Euler-Maruyama method gives a pathwise approximation to the solution. The expected error goes to zero with order 0.5.

Example 7.1. For the Ornstein-Uhlenbeck process with parameter α we have $b(x,t) = -\alpha x$ and $\sigma(x,t) = 1$ for all $x \in \mathbb{R}$, $t \geq 0$. The conditions of the theorem are satisfied for $K_1 = \alpha$, $K_2 = \max(1, \alpha)$ and $K_3 = 0$, so the Euler-Maruyama method will converge pathwise.

The result of a numerical simulation with $\Delta t = 0.005$ is shown in figure 3.1 (page 24).

7.2 Importance Sampling

Importance Sampling is a variation of Monte-Carlo sampling, where we use some knowledge about the integrated function to reduce the variance of the estimate. This is useful, because small variance means small errors in the estimate. The basics of this method are explained in [KW86].

Assume that $X, X_1, X_2, \ldots$ is an i.i.d. sequence of random variables and f is a measurable function. Basic Monte-Carlo integration uses the law of large numbers in the form of

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \longrightarrow E\bigl(f(X)\bigr). \tag{7.2}$$

The sum of the left hand side is used as an approximation for the expectation on the right hand side. The speed of convergence is determined by the variance of the left-hand side:

$$\operatorname{Var}\Bigl(\frac{1}{n}\sum_{k=1}^{n} f(X_k)\Bigr) = \frac{1}{n}\operatorname{Var}\bigl(f(X)\bigr). \tag{7.3}$$

Now let $Y_1, Y_2, \ldots$ be another sequence of i.i.d. random variables, such that the distribution of Y_k has a density g with respect to the distribution of X:

$$g = \frac{d\mathcal{L}(Y_k)}{d\mathcal{L}(X)} \quad \text{for all } k \in \mathbb{N}.$$

Then the law of large numbers gives

$$\frac{1}{n}\sum_{k=1}^{n}\frac{f(Y_k)}{g(Y_k)} \longrightarrow E\Big(\frac{f(Y_1)}{g(Y_1)}\Big) \tag{7.4}$$

$$= \int \frac{f(y)}{g(y)}\, g(y) d\mathcal{L}(X)(y)$$

$$= E\big(f(X)\big) \tag{7.5}$$

for $n \to \infty$. Again, the sum of the left hand side can be used as an approximation for the expectation on the right hand side. The variance of $f(Y_k)/g(Y_k)$ is small, if f and g are approximately proportional to each other. The boundary case is

$$g(y) = \frac{f(y)}{E\big(f(X)\big)}.$$

Then all the information about $E\big(f(X)\big)$ is already contained in g and

$$\frac{f(Y_k)}{g(Y_k)} = E\big(f(X)\big)$$

is constant for all $k \in \mathbb{N}$.

Of course this trick does not change the order of the method. The variance of the estimate is

$$\operatorname{Var}\Big(\frac{1}{n}\sum_{k=1}^{n}\frac{f(Y_k)}{g(Y_k)}\Big) = \frac{1}{n}\operatorname{Var}\Big(\frac{f(Y_1)}{g(Y_1)}\Big), \tag{7.6}$$

i.e. the method is still of order $1/\sqrt{n}$, but sometimes one can choose a function g to obtain a much better constant in the variance.

We are interested in estimating the probability of the event $\{X \in A\}$ for a measurable set A, i.e. in the case $f = 1_A$. Here the Monte-Carlo method (7.2) becomes

$$\frac{1}{n}\sum_{k=1}^{n} 1_A(X_k) \longrightarrow P(X \in A)$$

and the variance (7.3) for this estimate is

$$\sigma^2_{\mathrm{MC}} = \operatorname{Var}\Big(\frac{1}{n}\sum_{k=1}^{n} 1_A(X_k)\Big) = \frac{1}{n}\big(P(A) - P(A)^2\big).$$

To make importance sampling useful for $P(X \in A) \approx 0$, we choose random variables Y_k with $P(Y_k \in A) \gg P(X \in A)$. The importance sampling method (7.4) is

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1_A(Y_k)}{g(Y_k)} \longrightarrow P(X \in A) \quad \text{for } n \to \infty$$

and the corresponding variance from (7.6) becomes

$$\sigma^2_{\mathrm{Imp}} = \operatorname{Var}\Big(\frac{1}{n}\sum_{k=1}^{n}\frac{1_A(Y_k)}{g(Y_k)}\Big) = \frac{1}{n}\operatorname{Var}\Big(\frac{1_A(Y_k)}{g(Y_k)}\Big)$$

$$= \frac{1}{n}\Big(E\big(\frac{1_A(X)}{g(X)}\big) - P(A)^2\Big).$$

The ratio of the variances for the importance sampling method and the Monte-Carlo method is

$$\frac{\sigma^2_{\mathrm{Imp}}}{\sigma^2_{\mathrm{MC}}} = \frac{E\big(\frac{1_A(X)}{g(X)}\big) - P(A)^2}{P(A) - P(A)^2}.$$

If g is large on A this ratio becomes small, i.e. in these cases the estimate from the importance sampling method has a better variance.

Figure 7.1: *This figure illustrates the variance reduction, which can be achieved by deploying importance sampling. In the upper pictures we estimate the probability that a Brownian motion exceeds the level* 3 *before time 1 by generating a sample of 10000 Brownian paths, sampled with a step size of* $DT = 0.001$, *and counting how many of these reach a value greater than 3. The upper picture gives the histogram for the distribution of 2500 estimates generated in this way. The lower picture gives the histogram for 2500 estimates obtained by the importance sampling method from example 7.2. Again each estimate is calculated from a sample of 10000 paths, but one can see that the estimates obtained by importance sampling are much better concentrated around the theoretical value* $2.69 \cdot 10^{-3}$.

Example 7.2. To test the importance sampling method we try to use it to estimate the probability, that a Brownian motion exceeds the level 3 before time 1. We define

$$A = \bigl\{\omega\colon [0;1] \to \mathbb{R} \bigm| \sup_{0 \le t \le 1} \omega_t > 3\bigr\}.$$

From the reflection principle for a Brownian Motion X we know the exact value of this probability:

$$P(X \in A) = P\bigl(\sup_{0 \le t \le 1} X_t > 3\bigr) = 2P(X_1 > 3) = 0.00269\ldots.$$

For the process Y we can use Brownian Motion with a constant drift, i.e. $Y_t = X_t + bt$. From chapter 1 we know the density

$$g(\omega) = \frac{d\mathcal{L}(Y)}{d\mathcal{L}(X)}(\omega) = \exp(b \cdot \omega_t - b^2/2).$$

The result of a numerical simulation is displayed in figure 7.1.

7.3 The Rejection Method

The rejection method is a technique, which can be used to generate samples according to a distribution where the density with respect to some original measure is given. Further details can be found in [Knu81]. The book [PTV92] contains a sample implementation.

Theorem 7.2 *(rejection method). Let f, g be probability densities on some measurable space $(\mathcal{X}, \mathcal{F}, \mu)$ and $\lambda \geq 1$ be a number with $\lambda f \geq g$. Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of random variables with values in $\mathcal{X}$ and density f, and let $(U_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of random variables, uniformly distributed on the interval $[0;1]$, independently of the (X_n). Define $N = \min\{\, n \in \mathbb{N} \mid \lambda U_n f(X_n) \leq g(X_n)\,\}$. Then the distribution of X_N has density g on $(\mathcal{X}, \mathcal{F}, \mu)$.*

Thus the algorithm works as follows. Assume that we want to generate random values according to a probability density g. We first have to find another density f and a number $\lambda \geq 1$ with $\lambda f \geq g$, where we can already generate random values according to the probability density f. To sample for the density g one has to perform the following steps.

step 1: generate a random value X according to f

step 2: generate a random value U, uniformly distributed on $[0;1]$

step 3: if $U \cdot \lambda f(X) > g(X)$ go back to step 1

step 4: emit X

Because it gives some insights I reproduce the proof of the theorem here.

Proof. Given the value of X_n, it is accepted with probability $g(X_n)/\lambda f(X_n)$. So the total probability that X_n is accepted, is

$$P\big(\lambda U_n f(X_n) \leq g(X_n)\big) = \int_{\mathcal{X}} \frac{g(x)}{\lambda f(x)} f(x)\, d\mu(x) = 1/\lambda.$$

The value N is geometrically distributed with parameter $1/\lambda$ and for every set $A \in \mathcal{F}$ and $n \in \mathbb{N}$ we get

$$\begin{aligned} P(X_n \in A, N = n) &= \int_A \Big(1 - \frac{1}{\lambda}\Big)^{k-1} \frac{g(x)}{\lambda f(x)} f(x)\, d\mu(x) \\ &= \Big(1 - \frac{1}{\lambda}\Big)^{k-1} \frac{1}{\lambda} \cdot \int_A g(x)\, d\mu(x). \end{aligned}$$

Summation over n gives

$$P(X_N \in A) = \sum_{n\in\mathbb{N}} P(X_n \in A, N = n) = \int_A g(x)\, d\mu(x).$$

This proves the claim. (qed)

If we want to use this method to simulate a conditional distribution $P(\,\cdot \mid A)$ we proceed as follows. We choose some density φ where A has a high enough probability wrt. φ and where we can generate random values which are distributed according to φ. For the algorithm we choose the densities f and g with

$$f(x) = \frac{\varphi(x) 1_A(x)}{\int_A \varphi\, dP} \quad \text{and} \quad g(x) = \frac{1_A(x)}{P(A)}.$$

Because A has a high enough probability wrt. φ we can get samples according to f from the naive algorithm. The density g is the density of the conditional distribution which we are interested in. For λ we can choose

$$\lambda = \operatorname*{ess\,sup}_{x\in A} g(x)/f(x) = \frac{\int_A \varphi\, dP}{P(A)\, \operatorname{ess\,inf}_{x\in A} \varphi(x)}.$$

One good thing about the algorithm is, that we do not need to know the probabilities $P(A)$ and $\int_A \varphi\, dP$ in order to apply it: the condition

$$\lambda U f(X) \leq g(X)$$

for accepting a value becomes

$$U\varphi(X) \leq \operatorname*{ess\,inf}_{x\in A} \varphi(x) \tag{7.7}$$

here.

To generate one random value which is distributed according to f, we need in the mean $1/\int_A \varphi\,dP$ values which are distributed according to φ. From the proof above we know that in the theorem the number N of necessary input sample values is geometrically distributed with parameter $1/\lambda$. The mean value is $E(N) = \lambda$. So in the mean we have to generate

$$m = \frac{1}{\int_A \varphi\,dP} \cdot \lambda = \frac{1}{P(A)\operatorname{ess\,inf}_{x\in A}\varphi(x)} \tag{7.8}$$

values distributed according to φ in order to get one value distributed according to g. If φ is concentrated near A this can be much better than the value $1/P(A)$ from the naive algorithm.

Example 7.3. We can use the rejection method to simulate a Brownian Motion on the time interval $[0;t]$ conditioned on the event that $\int_0^t B_s^2\,ds < \varepsilon$ for a small value of ε and $|B_t| \leq c$ for some $c > 0$.

Because the integral condition is only satisfied for paths which stay most of the time near the origin, we use an Ornstein-Uhlenbeck process to sample the original random paths. From formula (3.3) we know that the density of an one-dimensional Ornstein-Uhlenbeck process with parameter $\alpha > 0$ is

$$\varphi_t(\omega) = \exp\Big(\frac{\alpha}{2}(t-\omega_t^2) - \frac{\alpha^2}{2}\int_0^t \omega_s^2\,ds\Big)$$

for all $\omega \in C([0;t],\mathbb{R})$. On the set

$$A_\varepsilon = \Big\{\omega \in C([0;t],\mathbb{R}) \;\Big|\; \int_0^t \omega_s^2\,ds < \varepsilon, |\omega_t| \leq c\Big\}$$

we find

$$\operatorname*{ess\,inf}_{x\in A_\varepsilon} \varphi_t(x) = \exp\Big(\frac{\alpha}{2}(t-c^2) - \frac{\alpha^2}{2}\varepsilon\Big).$$

Thus from condition (7.7) we conclude, that we should accept a path of the Ornstein-Uhlenbeck process, if it is in A_ε and additionally satisfies

$$U\exp\Big(\frac{\alpha}{2}(t-X_t^2) - \frac{\alpha^2}{2}\int_0^t X_s^2\,ds\Big) \leq \exp\Big(\frac{\alpha}{2}(t-c^2) - \frac{\alpha^2}{2}\varepsilon\Big)$$

or equivalently

$$U \leq \exp\Big(-\frac{\alpha}{2}(c^2-X_t^2) - \frac{\alpha^2}{2}\big(\varepsilon - \int_0^t X_s^2\,ds\big)\Big).$$

We want to keep the mean number of samples used from equation 7.8 small. Assume $c^2 < t$, now. Because of

$$\begin{aligned} m &= \frac{1}{P(A_\varepsilon)\operatorname{ess\,inf}_{x\in A_\varepsilon}\varphi(x)} \\ &= \frac{1}{P(A_\varepsilon)}\exp\Big(\frac{\alpha^2}{2}\varepsilon - \frac{\alpha}{2}(t-c^2)\Big) \\ &= \frac{1}{P(A_\varepsilon)}\exp\Big(\big(\alpha\sqrt{\varepsilon/2} - (t-c^2)/\sqrt{8\varepsilon}\big)^2 - \big(t-c^2\big)^2/8\varepsilon\Big) \end{aligned}$$

we will then choose

$$\alpha = \frac{t-c^2}{\sqrt{8\varepsilon}} \cdot \sqrt{\frac{2}{\varepsilon}} = \frac{t-c^2}{2\varepsilon}$$

Figure 7.2: *Sample path of a Brownian motion on the interval* $[0;1]$, *conditioned on the event that* $\int_0^1 B_s^2\,ds \leq 0.01$ *and* $|B_1| \leq 0.5$. *This figure was created with the rejection method described in example 7.3.*

to get the optimal mean number of samples used to produce one path from the conditioned Brownian motion, which is

$$m^*(\varepsilon) = \frac{\exp\bigl(-(t-c^2)^2/8\varepsilon\bigr)}{P(A_\varepsilon)}.$$

Using lemma 4.3 we can conclude

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log m^*(\varepsilon) = -\bigl(t-c^2\bigr)^2/8 - \lim_{\varepsilon\downarrow 0} \varepsilon \log P(A_\varepsilon) = \frac{t^2-(t-c^2)^2}{8}.$$

So the number of necessary samples still grows exponentially for $\varepsilon \downarrow 0$, but with a better exponential rate than the original $t^2/8$.

To illustrate the effect we can try this with $t = 1$, $c = 0.5$, and $\varepsilon = 0.01$. I simulated 1000000 path with step size $\Delta t = 10^{-5}$ each. The results are summarised in the following table. One of the resulting paths is shown in figure 7.2.

	naive method	rejection method
input samples	1000000	1000000
samples according to f	—	130813
accepted samples	0	914

So this is one of the cases where the rejection method works quite well, but the naive approach fails.

7.4 Sampling Bridges

Using the Euler-Maruyama from section 7.1 works well for ordinary stochastic differential equations, but it does not allow to sample a process conditioned on a given value for the end point. In order to simulate the process conditioned on the end point (a bridge), we need more sophisticated methods. This section describes such a method.

The basic principle used here is the Langevin_method. Given a probability distribution μ with density φ on $\mathbb{R}^d$, we consider the stochastic differential equation

$$dZ_t = \operatorname{grad} \log \varphi(Z_t) \bullet dt + \sqrt{2} \bullet dB_t. \tag{7.9}$$

From theorem 1.4 we know, that this process has a stationary distribution μ. Assuming ergodicity for Z we can approximate the distribution μ by simulating a solution of (7.9) and taking Z_t for large t, and we can approximate expectations $\int f\,d\mu$ by numerically evaluating $\int_0^t f(Z_s)\,ds$ for large t.

For constants $c \in \mathbb{R}$ we have $\operatorname{grad}\log(c\varphi) = \operatorname{grad}(\log c + \log\varphi) = \operatorname{grad}\log\varphi$, i.e. we can determine the drift for the SDE (7.9) even if we only know the density φ up to a constant.

Now assume we want to simulate solutions of

$$\begin{aligned} dX_t &= f(X_t)\bullet dt + dB_t \quad \text{for } 0 < t \leq T \\ X_0 &= a \in \mathbb{R}, \end{aligned} \tag{7.10}$$

where B is a 1-dimensional Brownian motion and $f\colon \mathbb{R} \to \mathbb{R}$ is some drift function, but conditioned on $X_T = b \in \mathbb{R}$. We can get an approximation $(\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_N) \in \mathbb{R}^{N+1}$ of the unconditioned solution via Euler method. $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_N)$ is a random vector in $\mathbb{R}^N$ with density

$$\varphi(x_1, \ldots, x_N) = \frac{1}{(2\pi\Delta t)^{N/2}} \exp\Big(-\sum_{n=1}^{N} \frac{\big(x_n - x_{n-1} - f(x_{n-1})\Delta t\big)^2}{2\Delta t}\Big)$$

where we use $x_0 = a$ as an abbreviation. The conditional density of the vector $(\tilde{X}_1, \ldots, \tilde{X}_{N-1})$, conditioned on $\tilde{X}_N = b$ is then

$$\varphi(x_1, \ldots, x_{N-1} | x_N = b) = c\exp\Big(-\sum_{n=1}^{N} \frac{\big(x_n - x_{n-1} - f(x_{n-1})\Delta t\big)^2}{2\Delta t}\Big)$$

where on the right hand side $x_0 = a$, $x_N = b$, and c is the normalising constant, which makes the function a probability density again. We want to apply the Langevin method to this probability density.

Define

$$I(x_1, \ldots, x_{N-1}) = \sum_{n=1}^{N} \frac{\big(x_n - x_{n-1} - f(x_{n-1})\Delta t\big)^2}{2\Delta t},$$

then the drift for the Langevin equation is $\operatorname{grad}\log\varphi(\,\cdot\,|x_N = b) = -\operatorname{grad} I$ and we get the Langevin equation

$$d\tilde{Z}_s = -\nabla I(\tilde{Z}_s)\bullet ds + \sqrt{2}d\tilde{B}_s.$$

For a $(N-1)$-dimensional Brownian motion $\tilde{B}$. I use s for the time in the Langevin equation here to distinguish it from the time t in the SDE (7.10).

Because here we are only interested in the stationary distribution of this SDE, we can freely rescale time in the Langevin equation. After the s-time transformation $Z_s = \tilde{Z}_{s/\Delta t}$ and $B_s = \tilde{B}_{s/\Delta t}$ we get the SDE

$$dZ_s = -\frac{1}{\Delta t}\nabla I(Z_s)\bullet ds + \sqrt{\frac{2}{\Delta t}}dB_s. \tag{7.11}$$

where B is another $(N-1)$-dimensional Brownian motion. A direct calculation gives

$$\begin{aligned} -\frac{1}{\Delta t}\partial_n I(x) = {} & \frac{x_{n+1} - 2x_n + x_{n-1}}{\Delta t^2} \\ & - f(x_n)f'(x_n) \\ & - \frac{f(x_n) - f(x_{n-1})}{\Delta t} + f'(x_n)\frac{x_{n+1} - x_n}{\Delta t} \end{aligned} \tag{7.12}$$

where we again use the abbreviations $x_0 = a$ and $x_N = b$. The reason for rescaling the s-time is, that now the first term on the right hand side is a discretized version of the Laplace operator and one could hope that for $N \to \infty$ the $(N-1)$-dimensional SDE (7.11) converges to a stochastic partial differential equation on $\mathbb{R}_+ \times [0;T]$.

The method to simulate a solution X of the SDE (7.9) conditioned on the end point X_T works now as follows. First calculate the corresponding drift for the Langevin equation

Figure 7.3: *This figure shows a path of an Ornstein-Uhlenbeck process with parameter* $\alpha = 5$ *and start in* 2*, conditioned on* $X_1 = 5$*. The simulation was done using the Langevin method from example 7.4. The simulation parameters are* $N = 1000$*,* $\Delta t = 0.001$*, and* $\Delta s = 5 \cdot 10^{-7}$*.*

using formula (7.12). Then use this drift to simulate a solution (Z_s) of (7.11) using the Euler-Maruyama method. The initial value Z_0 is arbitrary. One could, for example, use linear interpolation between the a and b. Now get Z_s for a large time s. Then the components $Z_{s,1}, \dots, Z_{s,N-1}$ are an approximation for $X_{1 \cdot \Delta t}, \dots, X_{(N-1) \cdot DT}$. Of course the problem with this method is to find good values of s.

Note that quite a small step size in s-direction is necessary, in order to keep the method stable. If the running time of the program is an important factor, then methods like the implicit Euler-Maruyama method or the Crank-Nicholson scheme, which allow greater step sizes in s-direction might be advantageous.

Example 7.4 (Ornstein-Uhlenbeck bridges). We can use the method described in this section to simulate paths from the Ornstein-Uhlenbeck process with given initial and final values.

Consider the SDE

$$dX_t = -\alpha X_t \bullet dt + dB_t \quad \text{for } 0 < t < 1$$

with $\alpha > 0$ and boundary conditions $X_0 = a$ and $X_1 = b$. Then we have $f(x) = -\alpha x$ and equation (7.12) becomes

$$\begin{aligned} -\frac{1}{\Delta t}\partial_n I(x) = {} & \frac{x_{n+1} - 2x_n + x_{n-1}}{\Delta t^2} \\ & - \alpha^2 x_n \\ & + \alpha \frac{-x_{n+1} + 2x_n - x_{n-1}}{\Delta t} \\ = {} & (1 - \alpha\Delta t)\frac{x_{n+1} - 2x_n + x_{n-1}}{\Delta t^2} - \alpha^2 x_n. \end{aligned}$$

Figure 7.3 shows the result of a numerical simulation, using this method.

List of Symbols

1_A the indicator function of the set A, i.e. $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ else
(A) the indicator function of the set A, used if A is an complicated expression (page 62)
$B_\delta(x)$ the open ball with radius δ around x
$B(\lambda, t)$ the Bayes risk for the separation of two processes, observed until time t, for a priory distribution λ (p. 75)
$\mathcal{B}(\mathbb{R}^d)$ the Borel-σ-algebra on $\mathbb{R}^d$
$C_0([0;t], \mathbb{R}^d)$ the space of all continuous functions from the interval $[0;t]$ into $\mathbb{R}^d$ (page 20)
$\mathrm{Exp}(\lambda)$ the exponential distribution with parameter λ
I_d the $d \times d$ identity matrix
K_r the closed ball with radius r around the origin: $K_r = \{ x \mid |x| \leq r \}$.
λ^d the d-dimensional Lebesgue measure (p. 16)
$\mathbb{N}$ the natural numbers $1, 2, 3, \ldots$
$\mathbb{N}_0$ the natural numbers including the zero: $0, 1, 2, \ldots$
$\mathcal{N}(\mu, \sigma^2)$ the Gaussian distribution with expectation μ and variance σ^2
$\mathrm{Prob}(\mathbb{R}^d)$ the space of all probability measures on $\mathbb{R}^d$
$\mathbb{R}_+$ the positive real numbers, i.e. $\mathbb{R}_+ = [0; \infty)$
$\rho(A)$ the spectral radius of the matrix A, i.e. the maximum of the absolute values of the eigenvalues of A
$s \wedge t$ the minimum of the numbers s and t
$s \vee t$ the maximum of the numbers s and t
$V_a^b(f)$ the total variation of the function f between a and b (p. 69)
$\mathbb{W}$ the Wiener measure on the path space
$\mathbb{W}_\varepsilon$ the law of scaled down Brownian motion (p. 20)
$\lfloor x \rfloor$ the largest integer smaller or equal to the real number x
$\lceil x \rceil$ the smallest integer greater or equal to the real number x
x^+, x^- the positive part resp. negative part of $x \in \mathbb{R}$. This is defined as $x^+ = \max(0, x)$ and $x^- = \max(0, -x)$.

Index

Anderson's inequality 31

Bayes risk 75
— for constant drift 76
— for Markov chains 82
— for OU processes 78
beautiful proof 14
bridges 91

Cameron-Martin-Formula 25
continuity set 19
contraction principle 15

diffusion coefficient 7
diffusion process 7
—, density of 9
drift function 7
—, constant 9, 76

empirical distribution 19
Euler-Lagrange equations 42
Euler-Maruyama method 85

Freidlin-Wentzell Theory 22

Girsanov-Formula 8

importance sampling 86

Langevin method 91
Laplace principle 16
Laplace transform 27
LDP 11
—, weak 11
— for empirical distributions 19
— for paths with small L^2-norm 31
— for sample paths 20, 22
— for stationary distributions 18
— for strong drift 25, 69, 71, 73
Lipschitz condition 7

occupation measure, empirical 19
Ornstein-Uhlenbeck process 10, 23
—, separation of 77
— conditioned on the endpoint 93
— with strong drift 24

projective limit 16

rate function 11
—, good 11
rejection method 89
reversible diffusion 8

small noise 20, 22
stationary distribution 8, 16
— of the OU process 23
stochastic Euler method 85
strong drift 16, 24, 37

theorem, Tauberian 27
— of Dawson-Gärtner 16
— — de Bruijn 27
— — Kolmogorov 8
— — Schilder 20
tightness, exponential 12
trace of a path 43

ugly calculation 50

Varadhan Lemma 16

Bibliography

[And55] T. W. Anderson. The integral of symmetric unimodular functions over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6:170–176, 1955.

[Arn74] Ludwig Arnold. *Stochastic Differential Equations: Theory and Applications.* John Wiley & Sons, 1974.

[BGT87] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation.* Cambridge University Press, 1987.

[BR89] Garrett Birkhoff and Gian-Carlo Rota. *Ordinary Differential Equations.* John Wiley & Sons, fourth edition, 1989.

[BS96] Andrei N. Borodin and Paavo Salminen. *Handbook of Brownian Motion — Facts and Formulae.* Probability and Its Applications. Birkhäuser, Basel, 1996.

[DS89] J.-D. Deuschel and D.W. Stroock. *Large Deviations.* Academic Press, 1989.

[DZ98] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Applications of Mathematics.* Springer, second edition, 1998.

[Fel71] William Feller. *An Introduction to Probability Theory and its Applications*, volume II. John Wiley & Sons, second edition, 1971.

[FLP99] Danielle Florens-Landais and Huyên Pham. Large deviations in estimation of an Ornstein-Uhlenbeck model. *J. Appl. Prob.*, 36(1):60–77, 1999.

[GF63] I. M. Gelfand and S. V. Fomin. *Calculus of Variations.* Prentice-Hall, 1963.

[Hol00] Frank den Hollander. *Large Deviations*, volume 14 of *Fields Institute Monographs.* American Mathematical Society, 2000.

[IW89] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library.* North-Holland, 1989.

[Knu81] Donald E. Knuth. *Seminumerical Algorithms*, volume 2 of *The art of computer programming.* Addison-Wesley, second edition, 1981.

[KP99] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations.* Number 23 in Applications of Mathematics. Springer, 1999. Corrected Third Printing.

[KS91] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer, second edition, 1991.

[KW86] Malvin H. Kalos and Paula A. Whitlock. *Monte Carlo Methods, Volume I: Basics.* John Wiley & Sons, 1986.

[KW97] J. Krob and H.v. Weizsäcker. On the rate of information gain in experiments with a finite parameter set. *Statistics and Decisions*, 15:281–294, 1997.

[Law95] Gregory F. Lawler. *Introduction to Stochastic Processes.* Chapman & Hall, 1995.

[PTV92] William H. Press, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C.* Cambridge University Press, second edition, 1992.

[RS72] Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics*, volume I. Academic Press, New York, 1972.

[RY99] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion.* Number 293 in Grundlehren der Mathematischen Wissenschaften. Springer, Berlin, third edition, 1999.

[Sch97] Peter Scheffel. *Exponential Risk Rates in Discrete Markov Models.* PhD thesis, Universität Kaiserslautern, 1997.

[SV79] D.W. Stroock and S.R.S. Varadhan. *Multidimensional diffusion processes.* Number 233 in Grundlehren der Mathematischen Wissenschaften. Springer, 1979.

[Voß97] Jochen Voß. Über die Asymptotik des Bayesrisikos bei Diffusionsprozessen. Diplomarbeit, Universität Kaiserslautern, 1997.

[WW90] H.v. Weizsäcker and G. Winkler. *Stochastic Integrals, An Introduction.* Advanced Lectures in Mathematics. Vieweg, Braunschweig-Wiesbaden, 1990.

About the Author

name: Jochen Voß
born: 29. September 1970
email: voss@seehuhn.de

Education and Work Experience

Oct. 2003 - June 2004 University of Warwick
Marie-Curie scholarship

Sept. 1998 - Aug. 2003 University of Kaiserslautern
Employed by the Department of Mathematics, University of Kaiserslautern as a research assistant ("wissenschaftlicher Mitarbeiter").

Feb. 1998 - Mar. 1998
Employed by the Department of Mathematics, University of Kaiserslautern as a teaching assistant ("wissenschaftliche Hilfskraft").

1991 - 1997 University of Kaiserslautern
"Diplom" in Mathematics with secondary topic Physics
Thesis "Über die Asymptotik des Bayesrisikos bei Diffusionsprozessen" (On the Asymptotical Behaviour of the Bayes Risk for Diffusion Models)

1981 - 1990 Albert Schweitzer Gymnasium, Kaiserslautern

Publications

Jochen Voß. Über die Asymptotik des Bayesrisikos bei Diffusionsprozessen. Diplomarbeit, Universität Kaiserslautern, 1997.

Andreas Voß, Klaus Rothermund, Jochen Voß. Interpreting the parameters of the diffusion model: An empirical validation. *Memory&Cognition*, to appear.