Large deviations for stochastic processes on Riemannian manifolds

Versendaal, R.

DOI
10.4233/uuid:1489bacb-4e60-47fa-9409-3866a164efcd

Publication date
2020

Document Version
Final published version

Citation (APA)
Versendaal, R. (2020). *Large deviations for stochastic processes on Riemannian manifolds*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:1489bacb-4e60-47fa-9409-3866a164efcd

Large deviations for stochastic processes on Riemannian manifolds

COLOPHON

Rik Versendaal
Large deviations for stochastic processes on Riemannian manifolds, Delft, 2020
ISBN: 000-00-0000-000-0

Cover art:
Cover design by
Printed by

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

Large deviations for stochastic processes on Riemannian manifolds

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 25 september 2020 om 10:00 uur

door

Rik VERSENDAAL

Master of Science in Mathematics
Technische Universiteit Delft, Nederland,
geboren te Dordrecht, Nederland.

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. F.H.J. Redig en Prof. dr. J.M.A.M. van Neerven

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. F.H.J. Redig,	Technische Universiteit Delft
Prof. dr. J.M.A.M. van Neerven,	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. A. Thalmaier,	Université du Luxembourg
Prof. dr. W.Th.F. den Hollander,	Universteit Leiden
Dr. B. Janssens,	Technische Universiteit Delft
Prof. dr. ir. G. Jongbloed,	Technische Universiteit Delft
Prof. dr. ir. M.C. Veraar,	Technische Universiteit Delft, reservelid
Overige leden:	
Dr. R.C. Kraaij,	Technische Universiteit Delft

Dit onderzoek werd gefinancierd door het Peter Paul Peterich fonds via het TU Delft Universiteitsfonds.

Contents

I

Introduction

1

Introduction

The main topic of this thesis is large deviations for stochastic processes in a geometric setting, such as a sphere. Large deviations is a mathematical theory that is concerned with quantifying the exponentially small probabilities of rare events, in particular deviations from the typical behaviour.

This chapter serves the purpose of providing a panoramic overview of the subjects treated in this thesis. Before we give an outline of the thesis, we embark on a journey to get an understanding of what large deviations are. We start with some fundamental examples and results. Based on these, we explain how to extend the problems to a geometric setting, which are studied in this thesis. Besides that, we also refer to other related directions which have been investigated in this area.

1.1. Large deviations for random walks

Arguably the most well known example of a probabilistic experiment is the tossing of a coin. Suppose we play a game in which we win 1 euro if the coin lands on heads, while we lose 1 euro if it lands on tails. If we keep on playing this game, we can win quite some money, but we can also lose it. Hence, we are interested in the behaviour of our profit after a (large) number of games. We explain how to study this in a variety of ways.

First, let us state the problem mathematically. We denote by X_n our winnings for the n-th toss. Since the coin is fair, we have

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}.$$

Our profit after n tosses is then given by the sum $S_n = \sum_{i=1}^n X_i$. Such a sum is often referred to as a *random walk*. To understand why, we can plot the value of S_n against the time n. At each time step, the value of S_n either moves up 1 or moves down 1. After n steps, we then have a trajectory moving up and down. The randomness comes from the fact that we toss a coin to decide if we move upwards or downwards.

The goal is to study the behaviour of S_n. A first way to do this is to consider the average profit $\frac{1}{n}\sum_{i=1}^n X_i$. Intuitively, if we perform a large number of tosses, we expect approximately an equal amount of heads and tails. This translates to an equal amount of times gaining or losing 1 euro, so that the profit will be close to 0. This result is known as the *law of large numbers.* More precisely, it states that

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to 0.$$

The law of large numbers shows that the probability for S_n to deviate order n from the expected behaviour goes to 0 when n becomes large.

We can of course also study fluctuations of different sizes around the expected behaviour. In the law of large numbers, the limit is deterministic, which shows that the variance has vanished. Since S_n consists of n independent tosses, its variance is precisely n times the variance of a single toss. Since the variance measures the expected squared deviation from the mean, it follows that $\frac{1}{\sqrt{n}}S_n$ has constant variance. The *central limit theorem* states that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \to Z,$$

where Z has a normal distribution with mean 0 and variance equal to the variance of a single toss. The central limit theorem allows us to study the probability of deviations of S_n from its expectation of order $\sqrt{n}$.

In contrast to the law of large numbers, the central limit theorem provides us with more specific information on the probabilities of deviations of order $\sqrt{n}$. One can wonder whether this is also possible on the scale of the law of large numbers, i.e., for deviations of order n. Such deviations are referred to as *large deviations*, since a sum of n terms typically has a size of at most order n. Where the law of large numbers only tells that the probability of large deviations goes to 0, the theory of large deviations is concerned with how fast this convergence is. More precisely, it quantifies the limiting behaviour of the exponentially small probabilities.

For the coin flipping example with which we started, one can show that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \approx x\right) \approx e^{-nI(x)} \tag{1.1.1}$$

where $I(x) = \frac{1}{2}(1+x)\log(1+x) + \frac{1}{2}(1-x)\log(1-x)$. This result should be interpreted as follows: the probability that the average profit $\frac{1}{n}\sum_{i=1}^{n} X_i$ is close to x decays exponentially in n with rate $I(x)$. In particular, we have that $I(0) = 0$, meaning that the probability that $\frac{1}{n}\sum_{i=1}^{n} X_i$ is close to 0 converges to 1. This is exactly what the law of large numbers tells us. Furthermore, the farther we go from $x = 0$, the larger I gets. This confirms our common sense that the larger the deviation from 0, the less likely it is to occur.

The result in (1.1.1) is a special case of a more general result known as Cramér's theorem, see Theorem 2.1.10. Moreover, Cramér's theorem shows how to compute the function I from the distribution of the random variables $\{X_n\}_{n\geqslant 1}$. Furthermore, the result also holds when the X_n are d-dimensional vectors. The large deviations for empirical averages were first proven in [25] and improved upon in [20] to hold for more general distributions of the random vectors. In all these cases, the random variables $\{X_n\}_{n\geqslant 1}$ have to be independent and have the same distribution. This was relaxed in [50] and further generalized in [34]. This result is known as the Gärtner-Ellis theorem, see Theorem 2.1.12 for the version from [50].

1.1.1. Areas of application

To explain the relevance of large deviations, we continue with the example we introduced in the previous section, i.e., we have random variables $\{X_n\}_{n\geqslant 1}$ with

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}.$$

Now consider the random variable

$$Z_n = \prod_{i=1}^{n} 2^{X_i}.$$

One can think of Z_n as follows: If the coin flip lands on heads we multiply by 2, and if it comes up tails, we multiply by $\frac{1}{2}$. Since we expect approximately an equal amount of heads and tails, we expects Z_n to be approximately 1. However, the expectation of 2^{X_i} is $\frac{5}{4}$, so that the expected value of Z_n is $(\frac{5}{4})^n$. We thus see that Z_n is expected to be exponentially large, and certainly not close to 1. To understand what happens, note that we can write $Z_n = 2^{S_n}$, where $S_n = \sum_{i=1}^{n} X_i$. This shows that the large deviation events of the sum S_n control the behaviour of the expectation of Z_n. The reason for this is that, although the large deviations for S_n have an exponentially small probability to occur, they have an exponentially large contribution to the expected value of Z_n. Even though this is a toy-example, the observations we make are certainly relevant. For example, they play a role in the entropy-energy balance in statistical mechanics, a field of research which provides a wealth of applications of large deviation theory, see e.g. [90].

Another, fairly early area of application in which one is interested in large deviations is *information theory*, which was introduced by Shannon in [84]. The idea is that we want to transmit information over a noisy channel. We can think of this information as a string of zeros and ones, and for each bit there is a certain probability that we make an error in its transmission. Too many errors in the transmission can result in a wrong transmission of the message, and the risk of this happening is relevant to know. Early results in this direction can be found in [43].

One can also use large deviations in risk assessment, for example in the context of insurance claims, see [78] among others. Let us sketch a simplified version of this application. For this, let X_n be the amount of the n-th insurance claim done by any of the customers. Assume that all insurance claims have the same distribution, and are independent of each other. The sum $S_n = \sum_{i=1}^{n} X_i$ now represents the total amount of the first n insurance claims. If this value is excessively large, it is impossible for the insurance company to pay out all claims, with bankruptcy following. It is therefore worthwhile to know the risk of an excessive number of high claims, which can be estimated with the large deviations for $\frac{1}{n}\sum_{i=1}^{n} X_i$.

Finally, we also want to mention the area of statistics, in which we try to estimate for example parameters based on a certain amount of data. These estimators converge to the true value if we let the amount of data grow. However, the estimator is still random, and we would like to understand the probability that our estimate is far off. In a certain way, this quantifies the risk of a 'wrong' estimate.

1.2. Large deviations for trajectories

Large deviations can also be studied for objects other than empirical averages. In general, as long as some version of a law of large numbers is satisfied, one can ask the question if there is also some form of large deviations. We will explain large deviations for trajectories of processes. More specifically, we consider trajectories of random walks and of Brownian motion.

1.2.1. Large deviations for trajectories of random walks

Sometimes we are not only interested in the end point of a random walk, but also want to understand how we got there. In order to study the behaviour of the trajectory of a random walk, we define for every $t \in [0, 1]$ the random variable

$$Z_n(t) = \frac{1}{n}\sum_{i=1}^{\lfloor nt \rfloor} X_i.$$

Note that the path $Z_n(t)$ is constant on time-intervals $[\frac{i-1}{n}, \frac{i}{n})$, and steps occur at times $t_{i,n} = \frac{i}{n}$.

If we assume that the random variables $\{X_n\}_{n \geqslant 1}$ are independent, identically distributed with mean 0, then the law of large numbers gives us that

$$Z_n(\cdot) \to 0.$$

More precisely, the trajectories $Z_n(\cdot)$ converge to the trajectory which is constant 0.

Likewise, we also have an analogue of the central limit theorem. This is called the *invariance principle*, which states that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor n\cdot \rfloor} X_i \to W(\cdot)$$

in distribution, where W is a Brownian motion, see e.g. [12]. This is one of the reasons why Brownian motion is sometimes viewed as the path-space analogue of the normal distribution.

Since the trajectories $Z_n(\cdot)$ satisfy the law of large numbers, we can also study their large deviation behaviour. We explain heuristically how these can be obtained from the large deviations for random walks.

For every individual time $t \in [0,1]$, we obtain from the large deviations for random walks that

$$\mathbb{P}(Z_n(t) \approx x) \approx e^{-nI_t(x)}.$$

Since the increments of the random walk are independent and identically distributed, one can prove that $I_t(x) = tI(t^{-1}x)$, where $I = I_1$.

For two times $t_1, t_2 \in [0,1]$ with $t_1 < t_2$, the tuple $(Z_n(t_1), Z_n(t_2))$ also satisfies the large deviation principle:

$$\mathbb{P}(Z_n(t_1) \approx x_1, Z_n(t_2) \approx x_2) \approx e^{-nI_{t_1,t_2}(x)}.$$

Since $Z_n(t_2)$ depends on $Z_n(t_1)$, the rate function I_{t_1,t_2} is not the sum of I_{t_1} and I_{t_2}. However, since the increments of the random walk are independent, the increment $Z_n(t_2) - Z_n(t_1)$ is independent of $Z_n(t_1)$. Furthermore, the increments all have the same distribution, so that $Z_n(t_2) - Z_n(t_1)$ has the same distribution as $Z_n(t_2 - t_1)$. Therefore, heuristically we have

$$\begin{aligned}
\mathbb{P}(Z_n(t_1) \approx x_1, Z_n(t_2) \approx x_2) &= \mathbb{P}(Z_n(t_1) \approx x_1, Z_n(t_2) - Z_n(t_1) \approx x_2 - x_1) \\
&= \mathbb{P}(Z_n(t_1) \approx x_1)\mathbb{P}(Z_n(t_2) - Z_n(t_1) \approx x_2 - x_1) \\
&= \mathbb{P}(Z_n(t_1) \approx x_1)\mathbb{P}(Z_n(t_2 - t_1) \approx x_2 - x_1) \\
&\approx e^{-nI_{t_1}(x_1)} e^{-nI_{t_2-t_1}(x_2-x_1)}.
\end{aligned}$$

Remembering that $I_t(x) = tI(t^{-1}x)$, we thus find that

$$I_{t_1,t_2}(x_1,x_2) = I_{t_1}(x_1) + I_{t_2-t_1}(x_2 - x_1) = t_1 I\left(\frac{x_1}{t_1}\right) + (t_2 - t_1) I\left(\frac{x_2 - x_1}{t_2 - t_1}\right).$$

Continuing this idea, for a curve γ and partition $0 = t_0 < t_1 < \cdots < t_k \leqslant 1$ we find that

$$\mathbb{P}(Z_n(t_1) \approx \gamma(t_1), \ldots, Z_n(t_k) \approx \gamma(t_k)) \approx e^{-nI_{t_1,\ldots,t_k}(\gamma(t_1),\ldots,\gamma(t_k))},$$

where

$$I_{t_1,\ldots,t_k}(\gamma(t_1), \ldots, \gamma(t_k)) = \sum_{i=1}^{k} (t_i - t_{i-1}) I\left(\frac{\gamma(t_i) - \gamma(t_{i-1})}{t_i - t_{i-1}}\right).$$

Under some conditions on γ, if we let the mesh-size of the partition tend to 0, we have

$$\sum_{i=1}^{k} (t_i - t_{i-1}) I\left(\frac{\gamma(t_i) - \gamma(t_{i-1})}{t_i - t_{i-1}}\right) \to \int_0^1 I(\dot{\gamma}(t))\,\mathrm{d}t.$$

This suggest that

$$\mathbb{P}(Z_n(\cdot) \approx \gamma) \approx e^{-n\int_0^1 I(\dot{\gamma}(t))\,\mathrm{d}t}. \tag{1.2.1}$$

This can be made precise, and was proven in [74]. The result is known as Mogulskii's theorem, see Theorem 2.1.13.

The form of the rate function in Mogulskii's theorem is a special case of a more general form given by

$$\mathcal{I}(\gamma) = \int_0^1 \mathcal{L}(\gamma(t), \dot{\gamma}(t))\,\mathrm{d}t.$$

The function $\mathcal{L}$ is called the Lagrangian, and the function $\mathcal{I}$ is then interpreted as an 'action'. When considering Brownian motion in the next section, we will see this form again.

1.2.2. Large deviations for Brownian motion with small variance

On the level of processes, arguably the most important stochastic process is Brownian motion. As mentioned earlier, it acts as the analogue of the normal distribution on process level, as is for example justified by the invariance principle.

The increments of Brownian motion are independent, stationary and have a normal distribution. Therefore, if we take an appropriate scaling $W_n(t) = a(n)W(t)$ of a Brownian motion $W(t)$, it should be possible to approximate $W_n(t)$ by

$$Z_n(t) = \frac{1}{n}\sum_{i=1}^{\lfloor nt \rfloor} X_i,$$

where $\{X_n\}_{n\geqslant 1}$ is a sequence of independent, standard normal random variables. To find the correct scaling $a(n)$, observe that by the invariance principle we have

$$\sqrt{n}Z_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt \rfloor} X_i \to W(t)$$

in distribution. This implies that for n large, $Z_n(t)$ is approximately equal to $\frac{1}{\sqrt{n}}W(t)$ in distribution. We thus should take $a(n) = \frac{1}{\sqrt{n}}$. This is also supported when we study the increments of $Z_n(t)$ and $W_n(t)$. Indeed, with this specific choice of $a(n)$, $W_n(\frac{i}{n}) - W_n(\frac{i-1}{n})$ has a normal distribution with mean 0 and variance $a(n)^2\frac{1}{n} = \frac{1}{n^2}$. The increments of $Z_n(t)$ also follow this distribution, so that $Z_n(t)$ is a piecewise constant approximation of $W_n(t)$.

It is possible to prove that $Z_n(t)$ approximates $W_n(t)$ well enough, such that their limiting behaviour on an exponential scale is the same, i.e., they follow the same large deviation principle. It is therefore enough to understand the large deviations of $Z_n(\cdot)$, which follow from Mogulskii's theorem. One can compute that the function I in Mogulskii's theorem is given by $I(x) = \frac{1}{2}|x|^2$ in the case of standard normal random variables. As a consequence, we have

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}W(\cdot) \approx \gamma\right) \approx \mathbb{P}(Z_n(\cdot) \approx \gamma) \approx e^{-n\frac{1}{2}\int_0^1 |\dot{\gamma}(t)|^2\,\mathrm{d}t}.$$

This result is due to Schilder, see [83]. We also give the precise statement in Theorem 2.1.14. Observe that the rate of a trajectory γ is given by the action $\frac{1}{2}\int_0^1 |\dot{\gamma}(t)|^2\,\mathrm{d}t$ obtained from the kinetic energy. More precisely, the higher the action, the less likely the trajectory.

1.3. Random walks in curved spaces

Large deviations for random walks have been studied in a variety of settings. In theoretical context, the result of Cramér's theorem also holds in Banach spaces, see [31]. Furthermore, one can also consider large deviations for empirical measures, the result on which is known as Sanov's theorem, see e.g. [56, 29]. Recently, a lot of attention has also gone to large deviations for random walks in random environments, see e.g. [94, 100] for some initial results. Finally, we mention the recent development of obtaining large deviations for Markov processes based on convergence of non-linear semigroups and viscosity solutions of Hamilton-Jacobi equations. This method was introduced in [39]. Among a wealth of applications, it is for example applied to obtain results concerning empirical measures of Markov jump processes, see [62]. We will also make use of this method to obtain some of our results.

In this thesis, we consider random walks in curved spaces, i.e., manifolds. Such random walks are mainly studied in the context of approximating diffusions on manifolds. The origin of this can be found in [58], where the central limit problem is considered. This result has been extended to a time-inhomogeneous setting in [64]. Other recent results concerning approximating solutions to stochastic differential equations on manifolds can be found in [1].

However, it seems that the large deviations of such random walks have not been considered. Therefore, our aim is to obtain results similar to those in (1.1.1) and (1.2.1) for random walks in manifolds. In order to do this, we first have to define random walks in manifolds. Indeed, if we would simply copy the approach from the Euclidean case, a problem we immediately run into is that we cannot add points in a manifold together and rescale by a factor. This problem already occurs when one considers the sphere, which is the prototypical example of a manifold. We thus need to find a suitable generalization of $\frac{1}{n}\sum_{i=1}^{n} X_i$ in such spaces. For this, we will use the viewpoint of random walks.

1.3.1. Geodesic random walks

The increments $\{X_n\}_{n\geqslant 1}$ of the random walk $\sum_{i=1}^{n} X_i$ may be thought of as vectors. The addition of such a vector then amounts to following the straight line in the direction of the vector for time 1 to assure that we add the entire vector. See the left picture in Figure 1.1 for a visualization of this interpretation.

On a manifold, vectors providing directions are precisely the tangent vectors. Therefore, to make a 'step' of the random walk, we take a random tangent vector. We then have to follow the 'straight line' in that direction. In Euclidean space, straight lines are lines of shortest distance between points, i.e., they are geodesics. This explains that in the manifold, following the 'straight line' means that we have to follow the geodesic in that direction. We again do this for time 1, to 'add' the entire vector. We now construct a random walk by concatenating a number of random steps. Since each time we are at a different point, we need for every point on the manifold a distribution on the tangent space to tell us how to sample the next direction. In

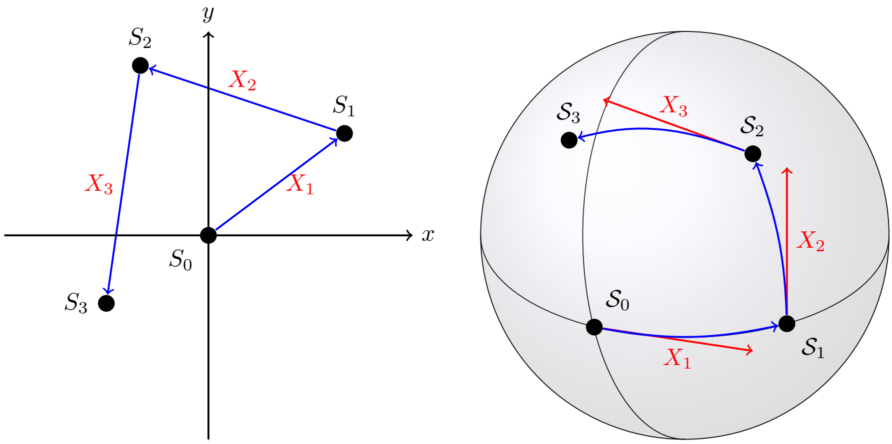

Figure 1.1: Visualization of the construction of geodesic random walks. On the left, we see the interpretation of a random walk in Euclidean space as repeatedly following straight lines in the direction of vectors. On the right, this idea is extended to the sphere, where we follow geodesics in the direction of tangent vectors.

Figure 1.1 this construction is shown for the sphere.

To summarize, to construct a random walk on a manifold, we first take on every tangent space a probability distribution. Then, to take a step, we sample a tangent vector at the point where the random walk is, and then follow the geodesic in that direction for time 1. We will denote the random walk after n steps by $\mathcal{S}_n$. Since we 'walk' along geodesics, we will refer to $\mathcal{S}_n$ as a *geodesic random walk*.

What remains is to define how we can rescale the random walk by a factor $\frac{1}{n}$. Since we cannot rescale $\mathcal{S}_n$, what we do, is we rescale the tangent vectors we sample instead. Equivalently, we can also follow the geodesics for time $\frac{1}{n}$ instead of time 1. We denote the rescaled random walk by $(\frac{1}{n} * \mathcal{S})_n$.

Example: the sphere

As an example, we can consider the sphere as a 2-dimensional manifold, see Figure 1.1. To start the random walk, we need to select a point $\mathcal{S}_0 = x_0$ on the sphere. Furthermore, we have to define a probability distribution on every tangent space. For this, we can for example say we always take a tangent vector with a uniformly random direction and a fixed length. Since geodesics on the sphere are the great circles, the geodesic random walk then consists of following pieces of great circles of equal length, in random directions. This approach of defining random walks on a sphere agrees with early definitions made specifically in this case, see e.g. [81].

Random walks using grids

For completeness, we also shortly discuss another approach to defining random walks in curved spaces. For this, one takes a collection of points $\{p_i\}_i$ in the manifold, which together form a grid. To define a random walk on the manifold now reduces to defining a random walk on the grid. One can do this by assigning to each pair of points (p_i, p_j) a probability to jump from point p_i to point p_j.

In Euclidean space, we usually take an equidistant grid, for example consisting of all points with integer coordinates. If we then want to consider random walks with small stepsize, we can take the grid containing points with coordinates which are multiples of $\frac{1}{n}$. For general curved spaces, such regular grids do not necessarily exist, and choosing appropriate grids is more involved. Also, in order to obtain a grid with small stepsizes, we cannot simply rescale, and must for example add points to the grid to make sure that grid points tend to be closer to each other.

Grids for manifolds naturally arise when we have a collection of data points from a certain manifold. Laplacian based machine learning algorithms rely on the convergence of the discrete Laplacian on the approximating grid to the Laplace-Beltrami operator on the manifold. We refer to [85] and references therein. Furthermore, grids may be used to study interacting particle systems on manifolds, see e.g. [45].

1.3.2. Applications of probability theory in manifolds

A question we have to ask ourselves is whether it is necessary to complicate matters and take curvature into account. Indeed, while the Earth is spherical, if we look around us, we perceive it as flat. If we would then zoom in enough, we may just as well locally approximate our curved space with a flat one.

However, our perception of the Earth as flat is a matter of scale. If we for example would like to predict the trajectory of a hurricane or of streams in the ocean, the curvature does become relevant. Scale is also important if we study the behaviour of systems in nano-biology. Limit results then help us to understand macroscopic behaviour from the (stochastic) microscopic behaviour of the system.

Furthermore, as already mentioned above, manifolds occur naturally when considering data. The problem of manifold learning or visualization is concerned with retrieving the manifold structure of the data, which is usually of a much lower dimension than the data itself. This is for example treated in [89, 101] among others.

Related to this problem is the problem of sampling from a distribution on a manifold. In Euclidean space, this can often be done effectively using Markov Chain Monte Carlo. The idea is essentially to construct a Markov chain which has the target distribution as its invariant distribution. This approach can also be taken in the manifold setting, see e.g. [16].

Finally, we also mention the role probability theory and geometry play in shape analysis. One can for example consider an object, such as a human organ, that deforms over time. This deformation may be modelled as a stochastic process. However, since we cannot measure continuously, the problem is now that given observations at different time points, we would like to reconstruct the underlying process of deformation. This can for example be done by constructing diffusion bridges. We refer to [4] among others.

1.4. Brownian motion in Riemannian manifolds

In addition to Cramér's and Mogulskii's theorem, we also wish to extend Schilder's theorem to a geometric setting. For this, it is necessary to have a notion of Brownian motion in curved spaces.

In the Euclidean setting, Brownian motion $W(t)$ is usually defined as the unique continuous process with independent, stationary increments such that $W(t)-W(s)$ has a normal distribution with mean 0 and variance $t-s$. Since there is no clear way to define increments of a manifold-valued process, this approach is not suitable to define a Brownian motion in a manifold.

It is thus necessary to consider other characterizations of Brownian motion. As we have seen before, it follows from the invariance principle that Brownian motion is the limit of random walks of the form

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i.$$

Furthermore, since Brownian motion is a Markov process, we can also consider its generator. The generator of a Markov process describes in a certain way the infinitesimal evolution of the process. For Brownian motion, the generator is given by $\frac{1}{2}\Delta$, where Δ is the Laplacian, i.e.,

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}.$$

Finally, Brownian motion is also uniquely characterized as the martingale $W(t)$ having quadratic variation $[W]_t = t$.

We will explain how each of these characterizations can be used to define Brownian motion in manifolds. Furthermore, we also introduce a geometric construction.

It turns out that in order to define Brownian motion in a manifold, we need some additional structure. A Riemannian metric on a manifold is a smooth selection of inner-products on the tangent spaces. A *Riemannian manifold* is a manifold, together with a Riemannian metric. It is possible to define a notion of Brownian motion in a Riemannian manifold. However, in contrast to the Euclidean case, this process can blow up in finite time. One can define Riemannian Brownian motion either geometrically or in a probabilistic way. We discuss both approaches.

Geometric construction of Brownian motion

Firstly, we discuss a purely geometric way of defining Brownian motion in Riemannian manifolds. This method is due to Eells-Elworthy-Malliavin, see [35, 71, 57]. For simplicity, we again consider the sphere, which is a two-dimensional manifold. The entire procedure explained works equally well in general Riemannian manifolds.

The idea is that we transfer a curve γ from the plane $\mathbb{R}^2$ to the sphere by rolling the sphere along the curve, without slipping. The contact point between the plane and the sphere then traces a curve along the sphere, which we call the *development* of γ onto the sphere. In this rolling procedure, 'without slipping' intuitively means that the motion of the contact point between the plane and sphere is only influenced by the velocity of the curve γ and the curvature of the sphere. This procedure can be made mathematically precise, the details of which can be found in Section 2.3.

The idea is now to start with a Brownian motion $B(t)$ in the plane, and develop this onto the sphere. Unfortunately, as described above, we need to know the velocity of a curve if we want to develop it onto the sphere. An insight by Malliavin, called *Malliavin's transfer principle* shows that in a suitable way, the same procedure may also be carried out for stochastic processes. The 'velocity' of the stochastic process is then replaced by the Stratonovich differential. This is extensively explained in Section 2.4. A Brownian motion on the sphere is now obtained by considering the development of a Brownian motion in the plane.

A probabilistic approach to Riemannian Brownian motion

It is also possible to define Brownian motion in a Riemannian manifold in a probabilistic way. The different probabilistic approaches are based on the characterizations of Brownian motion in the Euclidean case.

In the Euclidean setting, Brownian motion is a Markov process generated by $\frac{1}{2}\Delta$. A Riemannian manifold possesses a natural analogue of the Laplacian, namely the Laplace-Beltrami operator, which we denote by Δ_M. Since the notion of a generator can be extended to manifold-valued Markov processes, we can define Riemannian Brownian motion as the continuous process generated by $\frac{1}{2}\Delta_M$.

Furthermore, it was shown in [58] that the invariance principle also holds in Riemannian manifolds. Therefore, Brownian motion can be obtained as the limit of geodesic random walks which are scaled by $\frac{1}{\sqrt{n}}$.

Finally, we also mention the extension of the idea that Brownian motion is a martingale $W(t)$ with quadratic variation $[W]_t = t$. For this, one first defines a notion of manifold-valued semimartingales and a notion of quadratic variation. One then uses the Levi-Civita connection of the Riemannian manifold to define manifold-valued martingales. Finally, Brownian motion is then characterized as a martingale with a specific quadratic variation in terms of the Riemannian metric. For details on this approach, see [36, 57].

1.4.1. Schilder's theorem for Riemannian manifolds

With a Riemannian Brownian motion at hand, we can pose the question if an analogue of Schilder's theorem also holds in Riemannian manifolds. For this, we first of all should notice that if $W(t)$ is Riemannian Brownian motion, then $\frac{1}{\sqrt{n}}W(t)$ is not defined. Instead, observe that in the Euclidean case, $\frac{1}{\sqrt{n}}W(t) = W(\frac{1}{n}t)$ in distribution. This motivates that in order to study large deviations, we should consider the processes $W_n(t) = W(\frac{1}{n}t)$.

Let us motivate the generalization of Schilder's theorem. As a Markov process, Brownian motion in Euclidean space $\mathbb{R}^d$ possesses a transition density $p(t,x,y)$ given by

$$p(t,x,y) = \frac{1}{(2\pi t)^{\frac{d}{2}}} e^{-\frac{1}{2t}|x-y|^2}.$$

From this it follows that

$$\lim_{t\to 0} t \log p(t,x,y) = -\frac{1}{2}|x-y|^2,$$

which describes the short time behaviour of the transition density of Brownian motion.

A Riemannian Brownian motion also possesses a transition density $p_M(t,x,y)$. However, contrary to the Euclidean case, we cannot give an explicit expression. Nonetheless, Varadhan studied the short time behaviour of the transition density in [93], obtaining that

$$\lim_{t\to 0} p_M(t,x,y) = -\frac{1}{2} d(x,y)^2.$$

Here, $d(x,y)$ is the so called Riemannian distance between points x and y, which in the Euclidean case is precisely $|x-y|$.

Recall that in Euclidean space we can prove the large deviation principle for Brownian motion by approximating with polygonal paths over meshes with size tending to zero. The similarity in the short time behaviour of the transition densities then suggests that in the Riemannian setting, we should be able to obtain a similar large deviation result for Riemannian Brownian motion. More precisely, if $W(t)$ is a Riemannian Brownian motion, then for $W_n(t) = W(\frac{1}{n}t)$ we have

$$\mathbb{P}(W_n(\cdot) \approx \gamma) \approx e^{-n\frac{1}{2}\int_0^1 |\dot\gamma(t)|^2_{g(\gamma(t))}\,\mathrm{d}t}.$$

Here, g denotes the Riemannian metric, and $|\dot\gamma(t)|_{g(\gamma(t))}$ is the norm of $\dot\gamma(t)$ with respect to the inner product $g(\gamma(t))$.

The rate function $\mathcal{I}(\gamma) = \frac{1}{2}\int_0^1 |\dot\gamma(t)|^2_{g(\gamma(t))}\,\mathrm{d}t$ is again given by the action of the path γ. Different from the Euclidean case is that we evaluate the norm of $\dot\gamma(t)$ with respect to the inner product $g(\gamma(t))$. The precise statement is given in Theorem 5.1.3 and can also already be found in [9, 41]. We also refer to [13] for related results. Although the result is already known, we provide several novel approaches for proving it, see Chapter 5. While developed to prove Schilder's theorem for Riemannian Brownian motion, the approaches are interesting in itself and can be applied to a wider variety of problems.

1.4.2. Brownian motion in evolving manifolds

The final generalization we consider in this thesis is Riemannian Brownian motion in a time-evolving manifold. More precisely, this means that we study manifolds with a Riemannian metric which changes over time. One can for example think of a sphere, whose radius varies in time. Furthermore, one could think of studying the random movements of proteins in cell membranes. Cells usually deform over time, and this influences the stochastic process that describes the movement of the proteins. Additionally, it is also possible that the parameter space of some model forms a manifold, and the relation between different parameters changes over time.

We describe the time-evolution of the Riemannian manifold by letting the Riemannian metric $g(t)$ depend on time. The geometric and probabilistic approaches to define Riemannian Brownian motion in the time-homogeneous setting may be adapted to the time-inhomogeneous case, see Chapter 7. In this way we can define Riemannian Brownian motion in an evolving manifold.

It was recently shown in [64] that the invariance principle for geodesic random walks also holds in this time-inhomogeneous setting. Other work in this direction mainly focusses on functional estimates, such as gradient estimates for the heat semigroup, to characterize curvature and solutions to the Ricci flow. A selection of references includes [18, 19, 54]. A result in the direction of large deviations can be found in [24], where the probability for Brownian motion to be in a small band around some given curve is studied.

To find the analogue of Schilder's theorem in this time-inhomogeneous setting, we cannot simply consider the process $W_n(t) = W(\frac{1}{n}t)$ for $W(t)$ a Riemannian Brownian motion with respect to the evolving metric $g(t)$. Indeed, in the limit of n to infinity, we will only notice the contribution of the metric $g(0)$. To solve this, we also have to scale the time-dependence of the metric. More precisely, we first define $\tilde{W}_n(t)$ as a Riemannian Brownian motion with respect to the evolving metric $g_n(t) = g(nt)$. We can then study the large deviations for the processes $W_n(t) = \tilde{W}_n(\frac{1}{n}t)$. This is done in Chapter 7. It turns out that the idea of the rate function being the action of the path carries over, i.e.,

$$\mathcal{I}(\gamma) = \frac{1}{2}\int_0^1 |\dot{\gamma}(t)|^2_{g(t)(\gamma(t))}\,\mathrm{d}t.$$

The difference with the time-homogeneous case is that we evaluate the norm of $\dot{\gamma}(t)$ with respect to the metric $g(t)$.

1.5. Outline of the thesis

We conclude the introduction by providing an outline of the thesis, and shortly summarizing the content of each chapter. The thesis consists of three main parts:

I **Introduction**: The current chapter with a general introduction to the topics studied, and a chapter providing some necessary mathematical background. (Chapters 1 & 2)

II **Large deviations for processes in Riemannian manifolds**: Extensions of classical large deviation result to a geometric setting. (Chapters 3-5)

III **Large deviations in a time-inhomogeneous setting**: Large deviations for random walks with time-inhomogeneous increments in the Euclidean setting, and Brownian motion with respect to a time-evolving metric in the geometric setting. (Chapters 6 & 7)

In addition to the general introduction to the topic we have given here, Chapter 2 gives a more rigorous introduction to the mathematical concepts used in this thesis. We introduce the concept of large deviations, and provide some basic results from the literature. Furthermore, we discuss the relevant notions from (Riemannian) geometry. Additionally, we explain the notion of horizontal lift and (anti-)development of curves via the frame bundle over a manifold. We conclude by extending these notions to hold also for stochastic processes.

In part II we treat extensions of classical large deviation results to the geometric setting. We start in Chapter 3 by extending Cramér's theorem to random walks in Riemannian manifolds. For this, we first introduce geodesic random walks in Riemannian manifolds. To prove the analogue of Cramér's theorem for geodesic random walks, we show how to identify the random walk in the manifold with a process in some tangent space. This way, we can use Cramér's theorem in vector spaces. To get this identification, we perform a careful geometric analysis of geodesic random walks.

Chapter 4 is also concerned with random walks in manifolds, but now specifically in Lie groups. The additional group structure allows for a slightly different and simpler definition of a random walk. In some cases, this coincides with the notion of a geodesic random walk. We discuss when exactly this is the case. With or without this identification, we show that a roughly similar approach as taken for geodesic random walks also results in the large deviations for random walks in Lie groups. However, the estimates we have to make for this are different from the ones for geodesic random walks.

In Chapter 5 we focus on path-space large deviations for processes in Riemannian manifolds. More precisely, we study the analogues of Mogulskii's and Schilder's theorem. We take two approaches of studying such large deviations. The first approach is based on the convergence of non-linear semigroups and viscosity solutions for Hamilton-Jacobi equations as introduced in [39]. Without going into details, we only state the results we need for our purposes. The second approach relies on lifting the process in the manifold to the frame bundle, and is only used for the analogue of Schilder's theorem. For Riemannian Brownian motion, the lifted process satisfies a globally defined stochastic differential equation. We prove the large deviations for this by embedding the frame bundle in Euclidean space and using Freidlin-Wentzell theory.

In part III, we generalize classical large deviation results to a time-inhomogeneous setting. In Chapter 6 we start by studying random walks in Euclidean space with time-inhomogeneous increments. Under suitable condition on the time-dependence, we prove the analogues of Cramér's and Mogulskii's theorem. As a step up towards the next chapter, we also prove the large deviation principle for processes generated by weighted Laplacians, where the weight depends (only) on time.

The latter is a special case of a Riemannian Brownian motion with respect to a time-evolving metric. We study the large deviations for such processes in Chapter 7. In order to do this, we extend the notions of horizontal lift and (anti)-development to time-dependent connections. We then show that the embedding approach used in the time-homogeneous case can also be used in the time-inhomogeneous setting.

2

Mathematical background

This chapter serves the purpose of introducing the various mathematical topics that are necessary in the main part of this work. Furthermore, it allows us to fix the notation. Before we get to the individual topics, we first discuss some generalities that do not belong to any of the treated subjects in particular.

First of all, we use Einstein's summation convention whenever there is no confusion. This means that if an index occurs twice in an expression, once as subscript and once as superscript, this index is summed over. For example, if $\{e_1, \ldots, e_d\}$ denotes the standard basis of $\mathbb{R}^d$, then for $v \in \mathbb{R}^d$ we write

$$v = v^i e_i.$$

Furthermore, we define the function spaces that we will encounter. For the set of bounded, measurable function on $\mathbb{R}^d$ we write $L^\infty(\mathbb{R}^d)$. We denote by $C(\mathbb{R}^d)$ the set of continuous functions and we write $C_b(\mathbb{R}^d)$ for the set of bounded, continuous functions. Furthermore, we denote by $C^p(\mathbb{R}^d)$ the set of p-times continuously differentiable functions, and by $C^\infty(\mathbb{R}^d)$ the set of smooth functions, i.e., infinitely differentiable functions. A subscript c denotes that we only consider functions with compact support, i.e., we write $C_c(\mathbb{R}^d)$, $C_c^p(\mathbb{R}^d)$ and $C_c^\infty(\mathbb{R}^d)$. If we work in a space different from $\mathbb{R}^d$, but in which any of the notions make sense (think of a manifold M, see Section 2.2), we use the same notation, with $\mathbb{R}^d$ replaced by the given space.

Additionally, we also need to define spaces of curves. For an interval $[a, b] \subset \mathbb{R}$ we write $L^\infty([a, b]; \mathbb{R}^d)$ for the set of bounded, measurable curves $\gamma : [a, b] \to \mathbb{R}^d$. We denote the continuous curves by $C([a, b]; \mathbb{R}^d)$. Furthermore, we write $L^1([a, b]; \mathbb{R}^d)$ for the set of integrable curves.

We say a curve $\gamma : [a, b] \to \mathbb{R}^d$ is *absolutely continuous* if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for any partition $a \leqslant s_1 < t_1 \leqslant \cdots \leqslant s_n < t_n \leqslant b$ satisfying $\sum_{i=1}^n (t_i - s_i) < \delta$ it holds that

$$\sum_{i=1}^n |\gamma(t_i) - \gamma(s_i)| < \varepsilon.$$

Equivalently, a curve $\gamma : [a, b] \to \mathbb{R}^d$ is absolutely continuous if γ is almost everywhere differentiable with $\dot{\gamma} \in L^1([a,b];\mathbb{R}^d)$ and such that

$$\gamma(t) = \gamma(a) + \int_a^t \dot{\gamma}(s)\,\mathrm{d}s.$$

We write $AC([a,b];\mathbb{R}^d)$ for the set of absolutely continuous curves $\gamma : [a,b] \to \mathbb{R}^d$. Finally, we define the space $H^1([a,b];\mathbb{R}^d)$ by

$$H^1([a,b];\mathbb{R}^d) := \left\{ \gamma : [a,b] \to \mathbb{R}^d \,\middle|\, \gamma \text{ differentiable a.e.}, \int_a^b |\dot{\gamma}(t)|^2\,\mathrm{d}t < \infty \right\}.$$

In the case of curves, if we only consider curves γ with a given initial point $\gamma(a) = x$, we write $C_x([a,b];\mathbb{R}^d)$, $AC_x([a,b];\mathbb{R}^d)$, $H^1_x([a,b];\mathbb{R}^d)$. Again, whenever each notion makes sense for spaces other than $\mathbb{R}^d$, we replace $\mathbb{R}^d$ in the notation accordingly.

In the remainder of this chapter we provide a mathematical introduction to the topics we are studying. In Section 2.1 we discuss the large deviation principle, together with some useful and noteworthy results. Section 2.2 is devoted to introducing the necessary basics from (Riemannian) geometry, and most importantly, fixing the notation we will use. In Section 2.3 we study the frame bundle over a manifold, and define the notions of horizontal lift, development and anti-development of curves. Finally, in Section 2.4 we discuss some stochastic calculus in manifolds.

2.1. Large deviations

The theory of large deviations is concerned with the limiting behaviour on an exponential scale of a sequence of random variables $\{X_n\}_{n\geqslant 1}$ in some state space $\mathcal{X}$. Examples of sequences for which this problem can be studied include empirical averages and diffusions with decreasing variance.

In this chapter, we define the notion of a large deviation principle in general. We also collect some useful results from the theory that will be of later use. Finally, we state the classical results concerning large deviations for empirical averages and diffusions with decreasing variance, the extensions of which to geometric and time-inhomogeneous settings are the main topic of this thesis.

2.1.1. Large deviation principle

We begin with the basic definition of a large deviation principle. For our purposes, we will restrict ourselves to processes taking values in a metric space $\mathcal{X}$.

Definition 2.1.1 (Rate function)**.** *A* rate function *is a lower-semicontinuous function $I : \mathcal{X} \to [0,\infty]$. A rate function is* good *if its level sets $\{x \in \mathcal{X} | I(x) \leqslant \alpha\}$ are compact. The domain $\mathcal{D}_I$ of a rate function I is the subset of $\mathcal{X}$ where I is finite, i.e., $\mathcal{D}_I = \{x \in \mathcal{X} | I(x) < \infty\}$.*

The rate function governs the exponential rate of decay in the large deviation principle, which we define next.

Definition 2.1.2 (Large deviation principle). *Let $\{X_n\}_{n\geqslant 1}$ be a sequence of random variables with values in $\mathcal{X}$. The sequence $\{X_n\}_{n\geqslant 1}$ satisfies the* large deviation principle (LDP) *in $\mathcal{X}$ with rate function I if the following are satisfied:*

1. *(Upper bound) For any $F \subset \mathcal{X}$ closed we have*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X_n \in F) \leqslant - \inf_{x\in F} I(x).$$

2. *(Lower bound) For any $G \subset \mathcal{X}$ open we have*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X_n \in G) \geqslant - \inf_{x\in G} I(x).$$

Remark 2.1.3. The large deviation principle is in a way the exponential version of the notion of weak convergence. Indeed, by Portmanteau's theorem (see e.g [12]), X_n converges weakly to X if and only if for all closed sets F we have

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leqslant \mathbb{P}(X \in F),$$

or equivalently,

$$\liminf_{n\to\infty} \mathbb{P}(X_n \in G) \geqslant \mathbb{P}(X \in G)$$

for all open sets G.

Furthermore, the infimum occurs in the upper and lower bound of the large deviation principle, since only the largest exponential contribution will determine the rate. This follows from the *Laplace principle*, which states that

$$\lim_{n\to\infty} \frac{1}{n} \log(e^{na} + e^{nb}) = \max\{a, b\}.$$

Theoretical results in large deviation theory

We now discuss some theoretical results that will help us in proving large deviation principles. Furthermore, we discuss how to obtain new large deviation principles from old ones.

In many cases, it is easier to prove the upper bound for compact sets, rather than general closed sets. If the lower bound of the large deviation principle holds, and the upper bound holds only for compact sets, we say the sequence $\{Z_n\}_{n\geqslant 1}$ satisfies the *weak large deviation principle*. If the mass of the random variables is then concentrated enough on compact sets, then the upper bound may actually be extended to all closed sets. We have the following definition.

Definition 2.1.4 (Exponential tightness). *A sequence $\{X_n\}_{n\geqslant 1}$ is* exponentially tight *if for every $\alpha > 0$ there exists a compact set $K_\alpha \subset \mathcal{X}$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Z_n \in K_\alpha^c) < -\alpha.$$

We have the following proposition, which can for example be found in Section 1.2 in [29].

Proposition 2.1.5. *Let $\{X_n\}_{n\geqslant 1}$ be a sequence of random variables satisfying the weak large deviation principle in $\mathcal{X}$ with rate function I. Assume furthermore that the sequence is exponentially tight. Then $\{X_n\}_{n\geqslant 1}$ satisfies the (full) large deviation principle in $\mathcal{X}$ with the same rate function I.*

One can obtain new large deviation principles from given ones by applying continuous functions to them. The following is Theorem 4.2.1 in [29].

Theorem 2.1.6 (Contraction principle)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces and let $f : \mathcal{X} \to \mathcal{Y}$ be continuous. Suppose $I : \mathcal{X} \to [0, \infty]$ is a good rate function.*

1. *Define $I' : \mathcal{Y} \to [0, \infty]$ by*
$$I'(y) = \inf\{I(x)|x \in \mathcal{X}, f(x) = y\}.$$
Then I' is a good rate function on $\mathcal{Y}$. Here, the infimum of the empty set is taken to be infinite, as usual.

2. *Suppose $\{X_n\}_{n\geqslant 1}$ satisfies the large deviation principle in $\mathcal{X}$ with rate function I. Then $\{f(X_n)\}_{n\geqslant 1}$ satisfies the large deviation principle in $\mathcal{Y}$ with rate function I'.*

Finally, there are also conditions under which two different sequences of random variables satisfy the same large deviation principle.

Definition 2.1.7 (Exponential equivalence)**.** *Let $(\mathcal{X}, d)$ be a metric space, and let $\{X_n\}_{n\geqslant 1}$ and $\{Y_n\}_{n\geqslant 1}$ be two sequences of random variables with values in $\mathcal{X}$. The sequences $\{X_n\}_{n\geqslant 1}$ and $\{Y_n\}_{n\geqslant 1}$ are* exponentially equivalent *if there exists a sequence $\{\mathbb{P}_n\}_{n\geqslant 1}$ of joint distributions of $\{X_n\}_{n\geqslant 1}$ and $\{Y_n\}_{n\geqslant 1}$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n \left(d(X_n, Y_n) > \delta\right) = -\infty$$

for every $\delta > 0$.

If two sequences of random variables are exponentially equivalent, then in the limit they are indistinguishable on an exponential scale. The following is Theorem 4.2.13 in [29].

Theorem 2.1.8. *Suppose $\{X_n\}_{n\geqslant 1}$ satisfies the large deviation principle with good rate function I and let $\{Y_n\}_{n\geqslant 1}$ be exponentially equivalent to $\{X_n\}_{n\geqslant 1}$. Then $\{Y_n\}_{n\geqslant 1}$ also satisfies the large deviation principle with rate function I.*

2.1.2. Large deviations for empirical averages

Let $\{X_n\}_{n\geqslant 1}$ be a sequence of independent, identically distributed random variables taking values in $\mathbb{R}^d$. Define $S_n = \sum_{i=1}^n X_i$ and consider the sequence $\left\{\frac{1}{n} S_n\right\}_{n\geqslant 1}$ of empirical averages. If $\mathbb{E}(X_1) < \infty$, then by the law of large numbers we have

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_1)$$

in probability.

Cramér's Theorem is concerned with the large deviations for the sequence $\left\{\frac{1}{n}S_n\right\}_{n\geqslant 1}$. Define $M(\lambda) = \mathbb{E}\left(e^{\langle\lambda, X_1\rangle}\right)$, the moment generating function of X_1 and set $\Lambda(\lambda) = \log M(\lambda)$. Λ is called the *log-moment generating function*, and is also known as the cumulant generating function.

The rate of the large deviation principle for $\{\frac{1}{n}S_n\}_{n\geqslant 1}$ is governed by the Legendre transform of the log-moment generating function, which we define next.

Definition 2.1.9 (Legendre transform)**.** *The Legendre transform $\Lambda^* : \mathbb{R}^d \to [0, \infty]$ of a function $\Lambda : \mathbb{R}^d \to \mathbb{R}$ is defined by*

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}^d} \left\{\langle\lambda, x\rangle - \Lambda(\lambda)\right\}.$$

The following is Cramér's theorem, see e.g. Theorem 2.2.3 in [29] or Theorem 1.4 in [56].

Theorem 2.1.10 (Cramér)**.** *Let $\{X_n\}_{n\geqslant 1}$ be a sequence of independent, identically distributed random variables taking values in $\mathbb{R}^d$. Denote by Λ the log-moment generating function of X_1 and assume that Λ is everywhere finite. Then $\left\{\frac{1}{n}S_n\right\}_{n\geqslant 1}$ satisfies the large deviation principle in $\mathbb{R}^d$ with good rate function I given by*

$$I(x) = \Lambda^*(x) = \sup_{\lambda\in\mathbb{R}^d} \left\{\langle\lambda, x\rangle - \Lambda(\lambda)\right\}.$$

The conditions on Λ in Theorem 2.1.10 may be weakened significantly. It can be shown that it suffices to assume that 0 is in the interior of the domain of Λ.

Beyond independent, identically distributed increments

Apart from weakening the condition on Λ in Theorem 2.1.10, it is also possible to weaken the conditions on the sequence $\{X_n\}_{n\geqslant 1}$. To this end, we present a more general result, which includes the case of empirical averages of a sequence of increments which are not necessarily independent and identically distributed.

Let $\{Z_n\}_{n\geqslant 1}$ be a sequence of random variables in $\mathbb{R}^d$. For every $n \geqslant 1$, define

$$\Lambda_n(\lambda) = \log \mathbb{E}\left(e^{\langle\lambda, Z_n\rangle}\right),$$

the log-moment generating function of Z_n.

Assumption 2.1.11. *For every $\lambda \in \mathbb{R}^d$,*

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

exists and Λ is differentiable.

Theorem 2.1.12 (Gärtner-Ellis). *Let $\{Z_n\}_{n\geqslant 1}$ be a sequence of $\mathbb{R}^d$-valued random variables. Suppose that Assumption 2.1.11 is satisfied. Then $\{Z_n\}_{n\geqslant 1}$ satisfies the large deviation principle in $\mathbb{R}^d$ with rate function I given by*

$$I(x) = \Lambda^*(x) = \sup_{\lambda\in\mathbb{R}^d} \{\langle \lambda, x\rangle - \Lambda(\lambda)\}$$

As for Cramér's theorem, the conditions on Λ in Assumption 2.1.11 can be weakened in order for Theorem 2.1.12 to still hold. This relies on some intricate convex analysis, which is beyond the scope of this exposition. We refer to Section 2.3 in [29].

2.1.3. Path-space large deviations

The study of large deviations is not restricted to empirical averages of sequences of random variables. We will also study large deviations on the level of trajectories. We do this for trajectories of random walks, as well as trajectories of diffusions with small variance.

Path-space large deviations for empirical averages

For a sequence $\{X_n\}_{n\geqslant 1}$ of $\mathbb{R}^d$-valued random variables, the sum $S_n = \sum_{i=1}^n X_i$ may be considered as a random walk in $\mathbb{R}^d$. Therefore, the sequence of empirical averages $\left\{\frac{1}{n}S_n\right\}$ may also be considered as a random walk of n steps with size of order $\frac{1}{n}$. Cramér's theorem can now also be used to obtain the large deviations for other points of this random walk, not simply the endpoint. More generally, for every $t \in [0,1]$, Cramér's theorem gives the large deviations for

$$Z_n(t) = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} X_i, \tag{2.1.1}$$

where $\lfloor x\rfloor$ denotes the largest integer below x. Given times $0 \leqslant t_1 < t_2 < \cdots t_l \leqslant 1$, we can use the Gärtner-Ellis theorem to obtain the large deviations for the sequence $\{(Z_n(t_1), \cdots, Z_n(t_l))\}_{n\geqslant 1}$. By making the partition ever finer, we finally obtain the large deviations for $\{Z_n(\cdot)\}_{n\geqslant 1}$ as random variables in $L^\infty([0,1];\mathbb{R}^d)$. This is known as Mogulskii's theorem, see e.g. Theorem 5.1.2 in [29].

Theorem 2.1.13 (Mogulskii). *Let $\{X_n\}_{n\geqslant 1}$ be a sequence of independent, identically distributed random variables in $\mathbb{R}^d$. Assume that the log-moment generating function Λ of X_1 is everywhere finite. Define $Z_n(t)$ for $t \in [0,1]$ as in (2.1.1). Then $\{Z_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $L^\infty([0,1];\mathbb{R}^d)$ with good rate function I given by*

$$I(\gamma) = \begin{cases} \int_0^1 \Lambda^*(\dot\gamma(t))\,\mathrm{d}t, & \gamma \in AC_0([0,1];\mathbb{R}^d) \\ \infty & \textit{otherwise.} \end{cases}$$

Large deviations for Brownian motion with small variance

Let $\{W(t)\}_{t\geqslant 0}$ be a standard Brownian motion in $\mathbb{R}^d$. Note that for every t we can write

$$W(t) = W(t) - W\left(\frac{\lfloor nt \rfloor}{n}\right) + \sum_{i=1}^{\lfloor nt \rfloor} \left\{ W\left(\frac{i}{n}\right) - W\left(\frac{i-1}{n}\right) \right\}.$$

Since the increments $W\left(\frac{i}{n}\right) - W\left(\frac{i-1}{n}\right)$ are independent with a normal distribution with mean 0 and variance $\frac{1}{n}$, we find that $\sqrt{n}\left(W\left(\frac{i}{n}\right) - W\left(\frac{i-1}{n}\right)\right)$ follows a standard normal distribution. Hence, Mogulskii's theorem (Theorem 2.1.13) implies that $\{Z_n(\cdot)\}_{n\geqslant 1}$, where

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \left\{ W\left(\frac{i}{n}\right) - W\left(\frac{i-1}{n}\right) \right\}, \tag{2.1.2}$$

satisfies the large deviation principle in $L^\infty([0,1];\mathbb{R}^d)$ with rate function

$$I(\gamma) = \begin{cases} \frac{1}{2}\int_0^1 |\dot{\gamma}(t)|^2 \,\mathrm{d}t, & \gamma \in AC_0([0,1];\mathbb{R}^d) \\ \infty & \text{otherwise.} \end{cases}$$

Here, the form of the rate function follows from the fact that for a standard normal distribution we have $\Lambda(\lambda) = \frac{1}{2}|\lambda|^2$, so that $\Lambda^*(x) = \frac{1}{2}|x|^2$.
Now define $W_n(t) = \frac{1}{\sqrt{n}} W(t)$. Then $Z_n(t)$ in (2.1.2) can be written as

$$Z_n(t) = \sum_{i=1}^{\lfloor nt \rfloor} \left\{ W_n\left(\frac{i}{n}\right) - W_n\left(\frac{i-1}{n}\right) \right\}.$$

It can be shown that $\{W_n(\cdot)\}_{n\geqslant 1}$ and $\{Z_n(\cdot)\}_{n\geqslant 1}$ are exponentially equivalent in $L^\infty([0,1];\mathbb{R}^d)$. As a consequence, we obtain the large deviations for $\{W_n(\cdot)\}_{n\geqslant 1}$ in $L^\infty([0,1];\mathbb{R}^d)$. Since the paths of Brownian motion are almost surely continuous, the large deviation principle actually holds in $C([0,1];\mathbb{R}^d)$, see Lemma 4.1.5 in [29]. This result was proved in [83] and is known as Schilder's theorem.

Theorem 2.1.14 (Schilder)**.** *Let $\{W(t)\}_{t\geqslant 0}$ be a standard Brownian motion in $\mathbb{R}^d$. For every $n \geqslant 1$ and $t \in [0,1]$, define $W_n(t) = \frac{1}{\sqrt{n}} W(t)$. Then $\{W_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];\mathbb{R}^d)$ with good rate function I given by*

$$I(\gamma) = \begin{cases} \frac{1}{2}\int_0^1 |\dot{\gamma}(t)|^2 \,\mathrm{d}t, & \gamma \in H_0^1([0,1];\mathbb{R}^d) \\ \infty & \textit{otherwise.} \end{cases}$$

Freidlin-Wentzell theory

Brownian motion with small variance is an example of a diffusion process with a small diffusion constant. The study of the large deviations for diffusions with small variance is known as Freidlin-Wentzell theory, see [41].

Let $\{W(t)\}_{t\geqslant 0}$ be a standard Brownian motion in $\mathbb{R}^d$, and let $b : \mathbb{R}^d \to \mathbb{R}^d$ be Lipschitz continuous. Let $X_n(t)$ be the solution of the stochastic differential equation

$$\mathrm{d}X_n(t) = b(X_n(t))\,\mathrm{d}t + \frac{1}{\sqrt{n}}\mathrm{d}W(t), \qquad X_n(0) = 0.$$

Define the map $F : C_0([0,1];\mathbb{R}^d) \to C([0,1];\mathbb{R}^d)$ given by $F(g) = f$, where f is the solution of the integral equation

$$f(t) = \int_0^t b(f(s))\,\mathrm{d}s + g(t),$$

for all $t \in [0,1]$. Then $X_n = F(W_n)$ and it can be shown (see Theorem 5.6.3 in [29]) that F is continuous. It now follows from the contraction principle (Theorem 2.1.6) together with Schilder's theorem that $\{X_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];\mathbb{R}^d)$ with good rate function I given by

$$I(\gamma) = \begin{cases} \frac{1}{2}\int_0^1 |\dot\gamma(t) - b(\gamma(t))|^2\,\mathrm{d}t, & \gamma \in H_0^1([0,1];\mathbb{R}^d) \\ \infty, & \text{otherwise.} \end{cases}$$

We conclude the discussion by also considering the case where the diffusion constant depends on space, i.e., it is a map $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$. Assume furthermore that σ is Lipschitz continuous. Let $Y_n(t)$ be the solution of

$$\mathrm{d}Y_n(t) = b(Y_n(t))\mathrm{d}t + \frac{1}{\sqrt{n}}\sigma(Y_n(t))\mathrm{d}W(t), \qquad Y_n(0) = y \in \mathbb{R}^d. \tag{2.1.3}$$

In this case, Y_n is no longer a continuous function of a rescaled Brownian motion. However, it can be approximated well enough by processes which are a continuous function of rescaled Brownian motion. This is shown in the proof of Theorem 5.6.7 in [29], which states the following.

Theorem 2.1.15 (Freidlin-Wentzell)**.** *Let $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be Lipschitz continuous and bounded. Fix $y \in \mathbb{R}^d$ and for every $n \geqslant 1$, let $Y_n(t)$ be the process defined by* (2.1.3)*. Then $\{Y_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];\mathbb{R}^d)$ with good rate function I given by*

$$I(\gamma) = \inf\left\{ \frac{1}{2}\int_0^1 |\dot\varphi(t)|^2\,\mathrm{d}t \,\middle|\, \varphi \in H^1([0,1];\mathbb{R}^d), \right.$$
$$\left. \gamma(t) = y + \int_0^t b(\gamma(s))\,\mathrm{d}s + \int_0^t \sigma(\gamma(s))\dot\varphi(s))\,\mathrm{d}s \right\}.$$

2.2. Some differential geometry

In this section we introduce the required notions from differential geometry, see for example [86] for a general introduction. Our main focus is towards Riemannian geometry, for which we refer to [69] among others.

2.2.1. Generalities

A topological space M is a *manifold* if for every point $x \in M$ there is a neighbourhood U which is homeomorphic to some Euclidean space. Such a neighbourhood, together with the homeomorphism is called a *chart*, which provides coordinates for the points in U. A collection of charts covering M is called an *atlas*. We call a manifold *second countable* if there exists an atlas of countably many charts. The *dimension* of the manifold at a point $x \in M$ is given by the dimension of the Euclidean space to which it is locally homeomorphic. We say the manifold M has dimension d if it has dimension d at every point. Finally, a manifold is called *smooth* if the transition maps between different charts are all smooth. In what follows, whenever we consider a manifold, we always consider it to be smooth, second countable and of finite dimension, unless otherwise stated.

For $x \in M$, the *tangent space* T_xM consists of all possible derivatives of curves through x. Elements of T_xM are called *tangent vectors*. In coordinates, if we write $\{e_1, \dots, e_d\}$ for the standard basis of $\mathbb{R}^d$, then we define $\frac{\partial}{\partial x^i}$ to be the tangent vector of a curve whose coordinates only move in the direction of e_i. For notational purposes, we often write ∂_i for $\frac{\partial}{\partial x^i}$. The tangent vectors $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^d}$ form a basis for T_xM. This allows us to write $v \in T_xM$ as

$$v = v^i \frac{\partial}{\partial x^i}.$$

This shows that equivalently, we can define tangent vectors in T_xM as derivations at x.

We define the *tangent bundle* TM to be the vector bundle with fibres T_xM, i.e.

$$TM = \bigsqcup_{x \in M} T_xM.$$

Here, the $\bigsqcup$ denotes the disjoint union. To avoid cumbersome notation, we will consider an element $v \in T_xM$ also as element of TM, where the base point $x \in M$ is implicit in the notation when considering $v \in TM$. A *section* of TM is a map $v : M \to TM$ such that $v(x) \in T_xM$ for every $x \in M$. A smooth section of TM is called a *vector field*. The set of all vector fields on M is denoted by $\Gamma(TM)$.

The dual of T_xM, i.e., the set of linear functions on T_xM, is denoted by T_x^*M. We refer to T_x^*M as the *cotangent space*, and to the elements as *cotangent vectors*. The vector bundle

$$T^*M = \bigsqcup_{x \in M} T_x^*M$$

is called the *cotangent bundle*. Smooth section of the cotangent bundle are called *1-forms*.

Finally, consider a smooth function $f : M \to N$ between two manifolds. The derivative of f, also called the *differential*, is a map $\mathrm{d}f : TM \to TN$ defined as

$$\mathrm{d}f(x)(v) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} f(\gamma(t)),$$

where $\gamma : (-\varepsilon, \varepsilon) \to M$ is such that $\gamma(0) = x$ and $\dot{\gamma}(0) = v$. In particular, for every $x \in M$ we have that $\mathrm{d}f(x) : T_xM \to T_{f(x)}N$. Furthermore, for every vector field v on M, $w(x) := \mathrm{d}f(x)(v(x))$ defines a vector field on N. We denote this vector field by $\mathrm{d}f(v)$ (and sometimes also by $f_*(v)$) and it is called the *push-forward* of v along f.

2.2.2. Connections, geodesics and parallel transport

Let $\pi : E \to M$ be a vector bundle over M. A connection on E is a way to differentiate smooth sections of E, which we denote by $\Gamma(E)$. We have the following definition.

Definition 2.2.1 (Connection)**.** *Let $\pi : E \to M$ be a vector bundle over M. A* connection *on E is a map $\nabla : \Gamma(TM) \times \Gamma(E) \to \Gamma(E)$, denoted by $(X, Y) \mapsto \nabla_X Y$ satisfying the following:*

1. *$\nabla_X Y$ is C^∞-linear in X, i.e., for all $X \in \Gamma(TM)$, $Y \in \Gamma(E)$ and $f \in C^\infty(M)$ we have*
$$\nabla_{fX} Y = f \nabla_X Y.$$
2. *$\nabla_X Y$ is linear in Y.*
3. *∇ satisfies the Leibniz rule:*
$$\nabla_X(fY) = (Xf)Y + f\nabla_X Y$$
for all $f \in C^\infty(M)$.

We call $\nabla_X Y$ the covariant derivative *of Y in the direction of X.*

If we take $E = TM$ in Definition 2.2.1, we obtain a connection ∇ on TM, which is sometimes also referred to as a *linear connection*. It provides a way to differentiate vector fields on M. When there is no confusion, we say that ∇ is a connection on M. In coordinates around $x \in M$, writing $\partial_i = \frac{\partial}{\partial x^i} \in T_xM$, we have that $\nabla_{\partial_i}\partial_j \in T_xM$. Since $\{\partial_1, \ldots, \partial_d\}$ is a basis for T_xM, there exist coefficients $\Gamma^k_{ij}(x)$ such that
$$\nabla_{\partial_i}\partial_j = \Gamma^k_{ij}(x)\partial_k.$$

We refer to the coefficients $\Gamma^k_{ij}(x)$ as the *connection coefficients*. It follows from the Leibniz-rule and linearity of the connection that we can use the connection coefficients to express $\nabla_v w$ in coordinates for general vector fields $v, w \in \Gamma(TM)$. Indeed, if we write $v = v^i\partial_i$ and $w = w^j\partial_j$, then
$$\nabla_v w = (v(w^k) + v^i w^j \Gamma^k_{ij})\partial_k.$$

Example 2.2.2. As example, let us consider the canonical connection on $M = \mathbb{R}^d$. For vector fields v, w on $\mathbb{R}^d$ we can write $v(x) = v_i(x)\frac{\partial}{\partial x_i}$ and likewise $w(x) = w_i(x)\frac{\partial}{\partial x_i}$. We then define the connection $\overline{\nabla}$ on $\mathbb{R}^d$ by
$$\overline{\nabla}_v w(x) := v_i(x)\frac{\partial w}{\partial x_i}(x) = v_i(x)\frac{\partial w_j}{\partial x_i}(x)\frac{\partial}{\partial x_j}.$$

In particular, writing $\partial_i = \frac{\partial}{\partial x^i}$, we have

$$\overline{\nabla}_{\partial_i}\partial_j = 0.$$

This shows that the connection coefficients of $\overline{\nabla}$ are 0.

A curve in M is a map $\gamma : I \to M$, where I is some real interval. Curves are always assumed to be smooth, unless otherwise stated. A *vector field along* γ is a smooth map $v : I \to TM$ with $v(t) \in T_{\gamma(t)}M$ for all $t \in I$. We denote the space of vector fields along γ by $\Gamma(T\gamma)$. A connection ∇ on TM allows us to differentiate vector fields v along γ. The following is Lemma 4.9 in [69].

Proposition 2.2.3. *Let ∇ be a connection on M, and let $\gamma : I \to M$ be a curve. There exists a unique linear map $D_t : \Gamma(T\gamma) \to \Gamma(T\gamma)$ satisfying the following:*

1. *D_t satisfies the Leibniz rule*

$$D_t(fv) = f'v + fD_tv$$

   for all $f \in C^\infty(I)$.

2. *If $v \in \Gamma(T\gamma)$ extends to a vector field $\tilde{v} \in \Gamma(TM)$ on M, then*

$$D_tv(t) = \nabla_{\dot{\gamma}(t)}\tilde{v}$$

Abusing notation, we sometimes write $\nabla_{\dot{\gamma}(t)}v(t)$ even if the vector field v along γ does not extend to a vector field on M. Furthermore, we sometimes write $\dot{v}(t)$ instead of $D_tv(t)$.

Using the derivative of vector fields along a curve, we define parallel vector fields and geodesics.

Definition 2.2.4. *A vector field v along a curve $\gamma : I \to M$ is called* parallel *if $D_tv(t) = 0$ for all $t \in I$. A curve γ is called a* geodesic *if the vector field $\dot{\gamma}$ is parallel along γ.*

Equivalent to having a connection is having a notion of parallel transport. Given a curve $\gamma : [a, b] \to M$ and $v \in T_{\gamma(a)}M$, we can consider the solution $v(t)$ of the differential equation

$$\nabla_{\dot{\gamma}(t)}v(t) = 0, \qquad v(0) = v. \tag{2.2.1}$$

In coordinates, equation (2.2.1) is a system of linear differential equations, so that the solution is unique, and exists for all time. This allows us to define a linear map

$$\tau_{\gamma(a)\gamma(t);\gamma} : T_{\gamma(a)}M \to T_{\gamma(t)}M$$

by setting $\tau_{\gamma(a)\gamma(t);\gamma}v = v(t)$. The map $\tau_{\gamma(a)\gamma(t);\gamma}$ is called *parallel transport* along γ. We omit the reference to the curve γ when it is understood.

We can use parallel transport to compute covariant derivatives. To this end, let $v, w \in \Gamma(TM)$ be vector fields and $x \in M$. Let γ be a curve with $\gamma(0) = x$ and $\dot{\gamma}(0) = v(x)$. Then

$$\nabla_v w(x) = \lim_{h \to 0} \frac{\tau^{-1}_{x\gamma(h)}w(\gamma(h)) - w(x)}{h}. \tag{2.2.2}$$

2.2.3. Riemannian geometry

For every $x \in M$, we can equip T_xM with an inner product $g(x)$, which is a positive definite, symmetric bilinear form. For $v, w \in T_xM$ we write $g(x)(v,w) = \langle v, w \rangle_{g(x)}$. The reference to the point x is omitted whenever the tangent space is understood. In coordinates, the inner product $g(x)$ is given by a matrix $G(x) = (g_{ij}(x))$ such that

$$g(x)(v,w) = g_{ij}(x)v^i w^j,$$

where $v = v^i \partial_i$ and $w = w^j \partial_j$. A *Riemannian metric* on M is a smooth selection $g = \{g(x)\}_{x \in M}$ of inner products on the tangent spaces. More precisely, for every $v, w \in \Gamma(TM)$, the map $x \mapsto \langle v, w \rangle_{g(x)}$ is smooth. A manifold M with Riemannian metric g is called a *Riemannian manifold* and is denoted by (M, g).

For general vector spaces, an inner product can be used to identify the dual of the vector space with the vector space itself. Using the Riemannian metric, this allows us to identify T_x^*M with T_xM for every $x \in M$. For an element $\omega \in T_x^*M$, we define $\omega^\# \in T_xM$ as the unique element satisfying

$$\omega(v) = \langle \omega^\#, v \rangle_{g(x)}$$

for all $v \in T_xM$. This procedure is sometimes referred to as 'raising an index', which has to do with the fact that coefficients of cotangent vectors are written with lower indices, while for tangent vectors the coordinates are written with upper indices. Likewise, for $w \in T_xM$ we can define $w^\flat \in T_x^*M$ by setting

$$w^\flat(v) = \langle w, v \rangle_{g(x)}$$

for all $v \in T_xM$. This procedure is known as 'lowering an index' for the same reasons as explained above.

By identifying T_x^*M with T_xM via the inner product $g(x)$, we can define an inner product on T_x^*M. Indeed, for $\omega, \eta \in T_x^*M$ we define

$$\langle \omega, \eta \rangle_{g(x)} := \langle \omega^\#, \eta^\# \rangle_{g(x)}.$$

If $G(x) = (g_{ij}(x))$ are the coordinates of the inner product $g(x)$ on T_xM, one can show that the coordinates of the inner product on T_x^*M are given by $G^{-1}(x) = (g^{ij}(x))$, i.e., we have

$$\langle \omega, \eta \rangle_{g(x)} = g^{ij}(x)\omega_i \eta_j,$$

where $\omega = \omega_i \mathrm{d}x^i$ and $\eta = \eta_j \mathrm{d}x^j$.

Finally, the identification of cotangent vectors with tangent vectors may also be used to define parallel transport of cotangent vectors. For $\omega \in T_x^*M$ and a curve γ with $\gamma(0) = x$ and $\gamma(1) = y$ we define $\tau_{xy;\gamma}\omega \in T_y^*M$ by the relation

$$(\tau_{xy;\gamma}\omega)^\# = \tau_{xy;\gamma}(\omega^\#).$$

In particular, this implies that

$$(\tau_{xy;\gamma}\omega)(v) = \omega(\tau_{xy;\gamma}^{-1} v)$$

for all $v \in T_yM$.

The Levi-Civita connection

Associated to a Riemannian metric is a unique connection on M which behaves well with respect to the metric. This is stated in the following theorem, see e.g. Theorem 5.4 in [69].

Theorem 2.2.5 (Fundamental lemma of Riemannian geometry)**.** *Let (M, g) be a Riemannian manifold. There exists a unique connection ∇ on M satisfying the following:*

1. *Compatibility: for all $X, Y, Z \in \Gamma(TM)$,*
$$X\langle Y, Z\rangle_g = \langle \nabla_X Y, Z\rangle_g + \langle Y, \nabla_X Z\rangle_g.$$
2. *Symmetric: for all $X, Y \in \Gamma(TM)$,*
$$\nabla_X Y - \nabla_Y X = [X, Y].$$
Here, $[X, Y] = XY - YX$, the commutator of X and Y.

This connection is called the Levi-Civita connection *of g.*

The symmetry property of the Levi-Civita connection is sometimes also referred to as the Levi-Civita connection being *torsion-free*. Since the Levi-Civita connection is compatible with the metric g, the inner product of parallel vector fields along a curve is constant. In particular, this implies that the associated parallel transport is an isometry, see e.g. Lemma 5.2 in [69]. Finally, we call the connection coefficients of the Levi-Civita connection *Christoffel symbols*.

Given a Riemannian metric g on M, we define the *length* of $v \in T_x M$ by its usual formula
$$|v|_{g(x)} = \sqrt{\langle v, v\rangle_{g(x)}}.$$
We omit the reference to the point $x \in M$ whenever the tangent space is understood. Given a curve $\gamma : [a, b] \to M$, we define its *length* by
$$L(\gamma; [a, b]) = \int_a^b |\dot{\gamma}(t)|_{g(\gamma(t))}\, \mathrm{d}t.$$

Using this length function, we define the *Riemannian distance* d on M as
$$d(x, y) := \inf\{L(\gamma) | \gamma : [a, b] \to M, \gamma(a) = x, \gamma(b) = y, \gamma \text{ piecewise smooth}\}. \tag{2.2.3}$$

One can prove that the Riemannian distance d is a metric on M, which generates a topology which coincides with the topology of M as manifold, see Lemma 6.2 in [69]. In particular, this shows that all Riemannian distances on M generate the same topology. Furthermore, it can be shown (Theorem 6.6 in [69]) that optimal paths for the distance between points in M are geodesics with respect to the Levi-Civita connection.

Riemannian exponential map

Given $x \in M$, define for every $v \in T_xM$ the geodesic γ_v satisfying $\gamma_v(0) = x$ and $\dot{\gamma}_v(0) = v$. By Theorem 4.10 in [69] geodesics are unique, but generally only exist on a small time interval. We say that the manifold M is *geodesically complete* if every such geodesic can be extended indefinitely. By the Hopf-Rinow theorem (see e.g. Theorem 6.13 in [69]), this is equivalent to the completeness of M as a metric space with the Riemannian distance d defined in (2.2.3).

We define the *Riemannian exponential map* $\mathrm{Exp}_x : \mathcal{E}(x) \to M$ by setting $\mathrm{Exp}_x v = \gamma_v(1)$, where $\mathcal{E}(x) \subset T_xM$ contains all $v \in T_xM$ for which γ_v as above exists at least on $[0,1]$. If M is complete, we have $\mathcal{E}(x) = T_xM$. If additionally M is simply connected, it holds that Exp_x is surjective.

However, due to curvature, the exponential map is not necessarily injective. For $x \in M$ we define the *injectivity radius* $\iota(x) \in (0,\infty]$ as

$$\iota(x) = \sup \{\delta > 0 | \mathrm{Exp}_x \text{ is injective on } B(0,\delta)\}.$$

That $\iota(x) > 0$ for all $x \in M$ follows from the fact that the Riemannian exponential map is a local diffeomorphism.

Given a set $A \subset M$, the *injectivity radius* of A is defined by

$$\iota(A) = \inf \{\iota(x) | x \in A\}. \tag{2.2.4}$$

We have the following result, which can be found in e.g. [59].

Proposition 2.2.6. *The injectivity radius $\iota(x)$ depends continuously on x. In particular, this implies that if $K \subset M$ is compact, then $\iota(K) > 0$.*

Example 2.2.7. As an example, we derive the injectivity radius for points on a sphere of radius R. For any point x on the sphere, the antipodal point is responsible for the Riemannian exponential map failing to be injective. This implies that Exp_x is injective on $B(0,\pi R)$, but not on any larger set. We conclude that $\iota(x) = \pi R$. Since the injectivity radius is independent of x, we find that $\iota(A) = \pi R$ for every subset A of the sphere.

2.2.4. Curvature

The idea of differential geometry is that it allows us to study non-flat spaces. Curvature is introduced as a measure of how non-flat a space is. This can be quantified in different ways, which we discuss here. However, first we should say what we consider to be a flat space. A Riemannian manifold is called *flat* if it locally isometric to Euclidean space with the usual Euclidean inner product. Not only Euclidean space is flat, but for example a cylinder is as well.

It turns out that a space is flat if and only if we have that (see e.g. [69, Theorem 7.3])

$$\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z = 0, \tag{2.2.5}$$

where $[X,Y] = XY - YX$ is the commutator of X and Y. This leads us to define the *Riemann curvature endomorphism* $R : \Gamma(TM) \times \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$

given by

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

The Riemann curvature endomorphism can be visualized as the way a tangent vector changes when it is parallel transported along an infinitesimal parallelogram. Using the result in (2.2.5) we find that a manifold is flat if and only if the Riemann curvature endomorphism is 0.

Associated to the Riemann curvature endomorphism is the *Riemann curvature tensor*, obtained by lowering an index of the Riemann curvature endomorphism. More precisely, it is the map $\mathrm{Rm} : \Gamma(TM) \times \Gamma(TM) \times \Gamma(TM) \times \Gamma(TM) \to C^\infty(M)$ given by

$$\mathrm{Rm}(X,Y,Z,W) = \langle R(X,Y)Z, W\rangle_g.$$

By taking the trace of the Riemann curvature tensor on its first and last index, we obtain the Ricci curvature. More precisely, the *Ricci curvature* is defined as the map $\mathrm{Ric} : \Gamma(TM) \times \Gamma(TM) \to C^\infty(M)$, where $\mathrm{Ric}(Y,Z)$ is given as the trace of the linear map $X \mapsto R(X,Y)Z$. The Ricci curvature measures how the volume of a small piece of a geodesic ball differs from its Euclidean counterpart. Here, a *geodesic ball* around $x \in M$ of radius $\varepsilon > 0$ is defined as the set $\mathrm{Exp}_x(B(0,\varepsilon))$.

Finally, by taking the trace of the Ricci-curvature, we obtain a function on M which is called the *scalar curvature* which is denoted by S. More precisely, for every $X \in \Gamma(TM)$ we have that $\mathrm{Ric}(X,\cdot) \in \Gamma(T^*M)$ and hence $\mathrm{Ric}(X,\cdot)^\# \in \Gamma(TM)$. The scalar curvature S is then given as the trace of the linear map $X \mapsto \mathrm{Ric}(X,\cdot)^\#$. In particular, this shows that the scalar curvature depends on the Riemannian metric, since we raise an index. In two dimensions, positive scalar curvature means that the surface bends away from the outward facing normal, while negative curvature means exactly the opposite. The sphere is a prototypical example of a manifold with constant positive curvature.

2.3. Horizontal lift to the frame bundle

In order to study trajectories in a d-dimensional manifold M, it can be worthwhile to identify curves in M with curves in $\mathbb{R}^d$. However, not any such identification will be useful, since we want to preserve certain properties of the curves.

A natural way of transferring a curve from a manifold to $\mathbb{R}^d$ is by 'rolling' the manifold over $\mathbb{R}^d$ along the curve. The curve in the manifold serves as the contact points on the manifold. The curve of contact points in $\mathbb{R}^d$ is then the resulting curve with which we identify our original curve in M. This procedure should only be influenced by the velocity of the curve and the curvature properties of the manifold. Therefore, the displacement of the contact point should only be caused by these properties, and should not be influenced by any external forces. Hence, we have to roll the manifold along the curve without 'slipping'.

To state this mathematically, the aim is to identify a curve γ in a d-dimensional manifold M with a curve w in $\mathbb{R}^d$ and vice versa. As mentioned above, the velocity of γ should be one of the determining factors for the velocity of w. Note that

$\dot{\gamma}(t) \in T_{\gamma(t)}M$. Hence, if we are given a basis $u(t) = \{u_1(t), \dots, u_d(t)\}$ of $T_{\gamma(t)}M$, we can identify $\dot{\gamma}(t)$ with its vector of coordinates in the basis $u(t)$. We then define $\dot{w}(t)$ to be exactly this vector of coordinates. The curve w is then obtained by integration (which makes sense in $\mathbb{R}^d$). For this procedure, we need to choose bases $u(t)$ for every tangent space $T_{\gamma(t)}M$. Since we require the procedure to furthermore only depend on the curvature of M, $u(t)$ cannot be chosen freely. Instead, when moving along the curve, we should not 'twist' the coordinate system. Therefore, the bases $u(t)$ should be parallel along $\gamma(t)$.

A coordinate system $u(t)$ for $T_{\gamma(t)}M$ will be called a frame. The curve $\gamma(t)$ with a parallel collection of frames $u(t)$ attached to it will be called the horizontal lift of γ. The curve w in $\mathbb{R}^d$ is called the anti-development. It turns out that this procedure is invertible. More precisely, if we start with a curve w in $\mathbb{R}^d$, we can find a curve $\gamma(t)$ in M and a parallel collection of frames $u(t)$ attached to $\gamma(t)$ such that $\dot{w}(t)$ are the coordinates of $\dot{\gamma}(t)$ in the frame $u(t)$. In this section we define these notions and the sketched procedure rigorously. References include [61, 87, 57].

2.3.1. Frame bundle

For $x \in M$, a *frame* for T_xM is a linear isomorphism $u : \mathbb{R}^d \to T_xM$. This can be thought of as providing a basis for the tangent space T_xM. Indeed, if we denote by $e_1, \dots, e_d$ the standard basis of $\mathbb{R}^d$, then $ue_1, \dots, ue_d$ is a basis for T_xM. The collection of frames for T_xM is denoted by F_xM, i.e.,

$$F_xM = \{u : \mathbb{R}^d \to T_xM | u \text{ linear isomorphism}\}.$$

Denote by $GL(d, \mathbb{R})$ the general linear group over $\mathbb{R}$, i.e., the group of invertible $d \times d$ matrices with real entries. If $g \in GL(d, \mathbb{R})$ and $u \in F_xM$, then the composition ug is again a frame for T_xM. Therefore, $GL(d, \mathbb{R})$ acts on F_xM by right multiplication.

The *frame bundle* FM over M is the bundle with fibres F_xM, sometimes denoted as

$$FM = \bigsqcup_{x \in M} F_xM.$$

Here $\bigsqcup$ denotes the disjoint union. The frame bundle can be made into a manifold of dimension $d + d^2$, with the projection $\pi : FM \to M$ being a smooth map. Furthermore, its tangent bundle can be split in two parts, namely in directions in M and in the direction of the frames, i.e., vectors tangent to the fibres of FM. If $V \in T_uFM$ is tangent to the fibre $F_{\pi u}M$, then V is said to be *vertical*. More precisely, $V \in T_uFM$ is vertical if and only if it is the tangent vector of a curve that remains inside $F_{\pi u}M$. We denote the vertical subspace of T_uFM by V_uFM. Since $F_{\pi u}M$ has dimension d^2 (as manifold), we find that V_uFM is a subspace of dimension d^2. Now consider left multiplication $L_u : GL(d, \mathbb{R}) \to F_{\pi u}M$ defined as $L_u g = ug$. Then $\mathrm{d}L_u(I) : T_I GL(d, \mathbb{R}) \to T_u F_{\pi u}M$, where we note that $T_I GL(d, \mathbb{R}) = M(d, \mathbb{R})$, the set of all $d \times d$-matrices (which is the Lie algebra of $GL(d, \mathbb{R})$). Using this, a basis of V_uFM is given by

$$V_{ij}(u) = [\mathrm{d}L_u(I)](E_{ij}), \tag{2.3.1}$$

where E_{ij} is the matrix of all zeros, except for a 1 in position (i, j).

Horizontal lift

Since the subspace V_uFM of vertical vectors has dimension d^2, there are still d independent directions left in T_uFM. These will represent the directions along M in the frame bundle FM, and we want to call these directions horizontal. However, different choices of d independent vectors in T_uFM, independent of V_uFM, span different subspaces of T_uFM. We will explain how to make an appropriate choice if we are given a connection ∇ on M. In Section 2.3.2 we show that conversely, every choice of horizontal subspaces satisfying certain consistency assumptions gives rise to a connection on M.

We now define the notion of horizontal curves, which we need to define horizontal vectors.

Definition 2.3.1. *Let ∇ be a connection on M and $\gamma : [0,1] \to M$ a curve in M. A* horizontal lift *of γ (with respect to ∇) is a curve $u : [0,1] \to FM$ with $\pi u(t) = \gamma(t)$ and*

$$\nabla_{\dot{\gamma}(t)} u(t)a = 0 \tag{2.3.2}$$

for all $t \in [0,1]$ and all $a \in \mathbb{R}^d$. A curve $u : [0,1] \to FM$ is said to be horizontal *if it is a horizontal lift of the curve $\pi u(t)$ in M.*

Since locally, $u(t)$ satisfies a system of ordinary differential equations, we have local existence, and uniqueness once an initial frame $u_0 \in F_{\gamma(0)}M$ is given. As we will see in Section 2.3.3, the horizontal lift actually exists for all time. In what follows now, the local existence is sufficient.

If $u(t)$ is a horizontal lift of γ, then for every $a \in \mathbb{R}^d$ the condition in (2.3.2) implies that $u(t)a$ is parallel along γ. As a consequence, $u(t)a \in T_{\gamma(t)}M$ is the parallel transport of $u(0)a \in T_{\gamma(0)}M$. We thus see that the horizontal lift encodes parallel vector fields along γ by a single vector in $\mathbb{R}^d$. Moreover, since equation (2.2.1) has a unique solution, we find that parallel transport along γ is given by

$$\tau_{\gamma(0)\gamma(t);\gamma} = u(t)u(0)^{-1}. \tag{2.3.3}$$

Using horizontal lifts of curves, we can define a notion of horizontal lifts of tangent vectors.

Definition 2.3.2. *For $p \in M$, let $X \in T_pM$ and $u \in F_pM$. Let γ be a curve in M with $\gamma(0) = p$ and $\dot{\gamma}(0) = X$ and denote by $u(t)$ its horizontal lift satisfying $u(0) = u$. Then the* horizontal lift *of X via the frame u, denoted by $X^*(u)$, is defined as $X^*(u) = \dot{u}(0)$.*

A vector $W \in T_uFM$ is said to be *horizontal* if $W = X^*(u)$ for some $X \in T_{\pi u}M$. Equivalently, W is horizontal if it is the tangent vector of a horizontal curve through u. We write H_uFM for the set of horizontal vectors in T_uFM and refer to this as the *horizontal subspace* of T_uFM. Note that this definition depends on the connection ∇, and that in general different connections lead to different horizontal subspaces. The following lemma justifies this definition.

Lemma 2.3.3. *For every $u \in FM$ we have that $T_uFM = H_uFM \oplus V_uFM$.*

Proof. Since V_uFM is d^2-dimensional, and T_uFM has dimension $d+d^2$, it suffices to prove that H_uFM is a subspace of at least dimension d, which is linearly independent from V_uFM.
First of all, note that for every $X \in T_{\pi u}M$ we have that $\mathrm{d}\pi(X^*(u)) = X$. Indeed, let $\gamma(t)$ be a curve with $\gamma(0) = \pi u$ and $\dot\gamma(0) = X$ and denote by $u(t)$ its horizontal lift with $u(0) = u$. By definition of $X^*(u)$, $\dot u(0) = X^*(u)$. From this it follows that

$$\mathrm{d}\pi(X^*(u)) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \pi(u(t)) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \gamma(t) = X.$$

This proves that H_uFM is at least d-dimensional. Furthermore, it shows that if $W \in H_uFM$ is such that $\mathrm{d}\pi(W) = 0$, then $W = 0$. A similar computation shows that for all $V \in V_uFM$ we have that $\mathrm{d}\pi(V) = 0$. Combining everything, we find that $H_uFM \cap V_uFM = \{0\}$. □

Finally, given $a \in \mathbb{R}^d$ and $u \in F_pM$, we have that $ua \in T_pM$, so that we can define its horizontal lift. We denote this by $H(u)a$, which is thus given by

$$H(u)a := (ua)^*(u). \tag{2.3.4}$$

If we again denote by $e_1, \ldots, e_d$ the standard basis of $\mathbb{R}^d$, then the horizontal vectors $H_1(u), \ldots, H_d(u)$ given by

$$H_i(u) := H(u)e_i \tag{2.3.5}$$

form a basis for H_uFM. The vector fields $H_1, \ldots, H_d$ so defined are referred to as the *canonical horizontal vector fields.*

Development and anti-development

A horizontal lift of a curve assigns to a curve γ in M a horizontal curve u in FM. For every t, $u(t)$ is a frame for the tangent space $T_{\gamma(t)}M$. This allows us to convert the velocity of γ to a velocity in $\mathbb{R}^d$. This observation can be used to associate to a curve in M a curve in $\mathbb{R}^d$ and vice versa. We have the following definition.

Definition 2.3.4. *Let $\gamma : [0,1] \to M$ be a curve in M and let $u(t)$ be a horizontal lift of γ. The* anti-development of γ *is defined as the curve $w : [0,1] \to \mathbb{R}^d$ given by*

$$w(t) = \int_0^t u(s)^{-1}\dot\gamma(s)\,\mathrm{d}s. \tag{2.3.6}$$

If we fix a frame $u \in F_{\gamma(0)}M$, we can speak about the anti-development of γ via u since in that case the horizontal lift of γ satisfying $u(0) = u$ is unique.

If $w(t)$ is the anti-development of $\gamma(t)$ via the horizontal lift $u(t)$, then (2.3.6) implies that

$$\dot w(t) = u(t)^{-1}\dot\gamma(t),$$

which rewrites to

$$\dot{\gamma}(t) = u(t)\dot{w}(t).$$

Since both sides are elements of $T_{\gamma(t)}M$, we can consider their horizontal lifts, which must be equal (see (2.3.4) for a definition of $H(u(t))$):

$$H(u(t))\dot{w}(t) = (u(t)\dot{w}(t))^* = (\dot{\gamma}(t))^* = \dot{u}(t). \tag{2.3.7}$$

Here, the last equality holds because $u(t)$ is the horizontal lift of γ. We thus obtained a differential equation for the horizontal lift $u(t)$ in terms of the anti-development w. This shows how to invert the operation of taking the anti-development of a curve. We make the following definition.

Definition 2.3.5. *Let $w : [0,1] \to \mathbb{R}^d$ be a curve in $\mathbb{R}^d$ and fix $u_0 \in F_pM$. Let $u : [0,1] \to FM$ be the solution of*

$$\dot{u}(t) = H(u(t))\dot{w}(t)$$

with $u(0) = u_0$, where $H(u(t))$ is as defined in (2.3.4). *Then the curve $\gamma(t) = \pi u(t)$ is called the* development *of w onto M.*

Sometimes, the curve u is referred to as the development of w, rather than the projection of u onto M.

To summarize, given a curve γ in M and frame $u_0 \in F_{\gamma(0)}M$, there exists a unique horizontal curve u in FM with with $u(0) = u_0$ and $\pi u(t) = \gamma(t)$ for all t. The curve u is the horizontal lift of γ via the frame u_0. Using the horizontal lift, we can define the curve w in $\mathbb{R}^d$ by

$$w(t) = \int_0^t u^{-1}(s)\dot{\gamma}(s)\,\mathrm{d}s.$$

The curve w is the anti-development of γ.

Conversely, given a curve w in $\mathbb{R}^d$ and $u_0 \in F_xM$, there exists a unique horizontal curve u in FM with $u(0) = u_0$ and satisfying

$$\dot{u}(t) = H(u(t))\dot{w}(t)$$

for all t. Here H is as defined in (2.3.4). The curve γ given by $\gamma(t) = \pi u(t)$ for all t is the development of w onto the manifold M.

2.3.2. Connection on the frame bundle

In the previous section we have seen how to define horizontal tangent vectors in FM when we are given a connection on M. In this section we show that these two approaches are equivalent, in the sense that we can also first define a suitable notion of horizontal tangent vectors on FM, and use these to define a connection on M.

Using a connection on M we defined the collection $\{H_uFM\}_{u\in FM}$ of horizontal subspaces by

$$H_uFM = \{X^*(u) | X \in T_{\pi u}M\}. \tag{2.3.8}$$

One can show that these subspaces depend smoothly on u.
For every $g \in GL(d,\mathbb{R})$, let $R_g : FM \to FM$ denote right-multiplication by g, i.e. $R_g u = ug$ for all $u \in FM$. The collection $\{H_uFM\}_{u\in FM}$ satisfies the following consistency property.

Proposition 2.3.6. *Let $g \in GL(d,\mathbb{R})$ and $u \in FM$. Then $\mathrm{d}R_g(H_uFM) = H_{ug}FM$.*

Proof. Let γ be a curve in M with $\gamma(0) = \pi u$ and $\dot\gamma(0) = X$ and let $u(t)$ be its horizontal lift with $u(0) = u$. For $a \in \mathbb{R}^d$ we have $ga \in \mathbb{R}^d$, and hence

$$\nabla_{\dot\gamma(t)} u(t) ga = 0,$$

because $u(t)$ is horizontal. It follows that $u(t)g$ is again a horizontal lift of γ. From this, it follows that

$$\mathrm{d}R_g(X^*(u)) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} R_g u(t) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} u(t)g \in H_{ug}FM,$$

where the latter holds since $u(0)g = ug$. We conclude that $\mathrm{d}R_g(H_uFM) \subset H_{ug}FM$. In the same way we find that $\mathrm{d}R_{g^{-1}}(H_{ug}FM) \subset H_uFM$. Since

$$\mathrm{d}R_g(\mathrm{d}R_{g^{-1}}(H_{ug}FM)) = \mathrm{d}(R_g \circ R_{g^{-1}})(H_{ug}FM) = H_{ug}FM,$$

we find that $H_{ug}FM \subset \mathrm{d}R_g(H_uFM)$, concluding the proof. □

Collecting everything, it turns out the collection $\{H_uFM\}_{u\in FM}$ of horizontal subspaces defines a so-called principal connection which we define next (see e.g. [61, Chapter 2] or [87, Chapter 8]).

Definition 2.3.7 (Principal connection)**.** *For every $u \in FM$, let $H_u \subset T_uFM$ be a subspace. The collection $\{H_u\}_{u\in FM}$ is called a* (principal) connection *on FM if the following hold:*

1. *$T_uFM = H_u \oplus V_uFM$ for all $u \in FM$,*
2. *$\mathrm{d}R_g(H_u) = H_{ug}$ for all $u \in FM, g \in GL(d,\mathbb{R})$,*
3. *H_u depends smoothly on u.*

The subspaces H_u are called horizontal. *For $X \in T_uFM$ we write $X = h(X) + v(X)$ for the unique decomposition in a horizontal and vertical vector.*

Now the notion of horizontal lift of curves and tangent vectors can also be defined in terms of a principal connection on FM. A curve $u : [0,1] \to FM$ is called *horizontal* if $\dot u(t) \in H_{u(t)}$ for all $t \in [0,1]$. A *horizontal lift* of a curve $\gamma : [0,1] \to M$ is a horizontal curve $u : [0,1] \to M$ such that $\pi u(t) = \gamma(t)$ for all $t \in [0,1]$. This notion of horizontal lift coincides with the notion of a horizontal lift in Definition 2.3.1 if we take the collection $\{H_uFM\}_{u\in FM}$ of horizontal subspaces defined in (2.3.8). Therefore, existence of horizontal lifts for all time with respect to a connection ∇ on

M follows from the existence for all time of horizontal lifts defined via a principal connection on FM. We sketch how to prove the latter. We refer to [61, Chapter 2] or [87, Chapter 8] for the details.

Given a curve γ in M, by working locally and patching pieces smoothly together, we can find a curve $v : [0,1] \to FM$ with $v(0) = u_0$ and $\pi v(t) = \gamma(t)$ for all $t \in [0,1]$. However, $v(t)$ need not be horizontal and hence, we need to adapt the curve. More precisely, we look for a curve $a : [0,1] \to GL(d,\mathbb{R})$ such that $u(t) = v(t)a(t)$ is horizontal. By the Leibniz rule we have

$$\dot{u}(t) = \mathrm{d}L_{v(t)}(a(t))(\dot{a}(t)) + \mathrm{d}R_{a(t)}(v(t))(\dot{v}(t)). \tag{2.3.9}$$

For $\dot{u}(t)$ to be horizontal, the vertical part of the right-hand side in (2.3.9) has to vanish. Since $\mathrm{d}L_{v(t)}(a(t))(\dot{a}(t))$ is vertical, we must have

$$\mathrm{d}L_{v(t)}(a(t))(\dot{a}(t)) = -v(\mathrm{d}R_{a(t)}(v(t))(\dot{v}(t))) \in V_{v(t)a(t)}FM.$$

Using that $V_{v(t)a(t)}FM$ can be identified with $M(d,\mathbb{R})$ using the map $\mathrm{d}L^{-1}_{v(t)a(t)}(I)$, we obtain a differential equation for $a(t)$ in $M(d,\mathbb{R})$. It can be shown (see e.g. [61, Lemma on p.69]) that this differential equation has a unique solution which exists for all $t \in [0,1]$ and takes values in $GL(d,\mathbb{R})$. But then $u(t) = v(t)a(t)$ also exists for all time, and by construction, this is a horizontal lift of γ, which concludes the sketch of the proof.

Finally, we can also define a connection on M when starting with a principal connection on FM. By (2.2.2) it is sufficient to define parallel transport along curves. For this, let u be the horizontal lift of γ. Inspired by (2.3.3), we define the map $\tau_{\gamma(0)\gamma(t);\gamma} : T_{\gamma(0)}M \to T_{\gamma(t)}M$ by

$$\tau_{\gamma(0)\gamma(t);\gamma} = u(t)u(0)^{-1}.$$

It can be shown (although this is not trivial, see again [61, 87]) that this defines a genuine parallel transport, meaning that it defines a connection on M (in the sense of Definition 2.2.1) via formula (2.2.2).

2.3.3. Principal bundles

The notion of a horizontal lift can also be defined in a more general setting than the frame bundle. It suffices to have a bundle over M and a Lie group (see Chapter 4) that acts on the fibres of the bundle by right multiplication. More precisely, we have the following definition, see for example [61].

Definition 2.3.8 (Principal bundle)**.** *Let M and P be manifolds. Furthermore, let G be a Lie group (see Chapter 4). The manifold P is called a* principal bundle over M with structure group G *if the following are satisfied:*

1. *G acts freely on P by right multiplication, i.e. $(u,g) \in P \times G$ implies $ug \in P$ and if $ug = u$ for all $u \in P$ then $g = e$, the identity element of G.*

2. M is diffeomorphic to the quotient P/G under the equivalence relation $\sim$ given by $u \sim v$ if and only if there exists a $g \in G$ such that $u = vg$, and the induced projection $\pi : P \to M$ is smooth.

3. P is locally trivial: for all $x \in M$ there exists a $U \subset M$ open and a diffeomorphism $\psi : \pi^{-1}(U) \to U \times G$ of the form $\psi(u) = (\pi(u), \varphi(u))$ with φ a map $\varphi : \pi^{-1}(U) \to G$ satisfying $\varphi(ua) = \varphi(u)a$ for all $u \in \pi^{-1}(U)$ and $a \in G$.

Adapting Definition 2.3.7 by replacing FM with P and $GL(d, \mathbb{R})$ with the Lie group G gives us the definition of a principal connection on P. Following the same procedure as in Section 2.3.2, a principal connection gives rise to a notion of horizontal lift of a curve in M to a curve in P.

We conclude this section with some examples of principal bundles that we will encounter in future chapters. The frame bundle FM is a prototypical example of a principal bundle over M, its structure group being $GL(d, \mathbb{R})$.
If M is equipped with a Riemannian metric, for every $p \in M$ we can consider the collection of orthonormal frames given by

$$O_pM := \{u \in F_pM | u : \mathbb{R}^d \to (T_pM, g) \text{ isometry}\}.$$

Here, we consider $\mathbb{R}^d$ to carry the standard Euclidean inner product. The bundle OM with fibres O_pM is called the *orthonormal frame bundle* and is also denoted as

$$OM = \bigsqcup_{p \in M} O_pM.$$

Here, $\bigsqcup$ denotes the disjoint union. The orthonormal frame bundle is a principal bundle with structure group $O(d)$, the orthogonal group.
Finally, the frame bundle can also be considered as bundle over $\mathbb{M} := \mathbb{R} \times M$. More precisely, we consider the bundle $\mathcal{F}$ with fibres given by

$$\mathcal{F}_{(t,p)} = F_pM$$

for all $p \in M$ and $t \in \mathbb{R}$. This is a principal bundle with structure group $GL(d, \mathbb{R})$. In this case, the time-coordinate does not really add anything new. However, suppose we are given a collection of Riemannian metrics $\{g(t)\}_{t \in \mathbb{R}}$. Then for every $t \in \mathbb{R}$, the orthonormal frame bundle is different. Let us set

$$\mathcal{O}_{(t,p)} = \{u \in F_pM | u : \mathbb{R}^d \to (T_pM, g(t)) \text{ isometry}\}$$

for all $p \in M$ and $t \in \mathbb{R}$. The bundle $\mathcal{O}$ with fibres $\mathcal{O}_{(t,p)}$, also denoted by

$$\mathcal{O} = \bigsqcup_{(t,p) \in \mathbb{M}} \mathcal{O}_{(t,p)},$$

is a principal bundle over $\mathbb{M}$ with structure group $O(d)$. This bundle will be important when we study large deviations for diffusions in evolving Riemannian manifolds, see Chapter 7.

2.4. Stochastic analysis in manifolds

One way of studying stochastic processes in manifolds, is to generalize the notion of horizontal lift and anti-development of curves to processes. The idea is that we can then derive properties of the process from its anti-development. Since the anti-development is a process in some Euclidean space, this is either easier to study, or has already been studied extensively.

The Malliavin transfer principle (see e.g. [72]) states that any construction for smooth curves in a manifold can be extended to processes by replacing the ordinary differential equations by Stratonovich stochastic differential equations. This is a heuristic principle which is motivated by the fact that Stratonovich stochastic differential equations satisfy the usual fundamental theorem of calculus.

First, we discuss how to define Stratonovich stochastic integrals in Euclidean space in terms of Itô integrals. In a similar way, we also define Stratonovich stochastic differential equations, and explain some of its properties. These properties inspire the definition of stochastic differential equations on manifolds, which we give next. We then use these to generalize the notion of horizontal lift and (anti-)development to stochastic processes. We conclude by defining Riemannian Brownian motion. For further reading, we refer to [57, 36]

2.4.1. Stochastic differential equations on $\mathbb{R}^d$

In this section we consider stochastic calculus in Euclidean space. For a general introduction, we refer to [38].

Let X_t be an $\mathbb{R}^d$-valued semimartingale, i.e., the sum of a martingale and a process of bounded variation. By Itô's formula, we have for $f \in C_c^\infty(\mathbb{R}^d)$ that

$$f(X_t) = f(X_0) + \sum_{i=1}^{d} \int_0^t \frac{\partial f}{\partial x_i}(X_s)\mathrm{d}Z_s^i + \frac{1}{2}\sum_{i,j=1}^{d} \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\, \mathrm{d}[X^i, X^j]_s. \quad (2.4.1)$$

Here, $[X^i, X^j]_s$ is the *quadratic variation* of X^i and X^j. It is defined as the limit in probability of

$$\sum_{l=1}^{k} (X^i_{t_l} - X^i_{t_{l-1}})(X^j_{t_l} - X^j_{t_{l-1}})$$

when the mesh-size of $0 = t_0 = 0 < t_1 < \cdots < t_k = s$ tends to zero.

The stochastic integrals in (2.4.1) are Itô-integrals. We see that such integrals do not follow the ordinary rules of calculus, since we have to correct with a quadratic variation term. The idea is to define an alternative stochastic integral in terms of the Itô integral which does satisfy the ordinary fundamental theorem of calculus. We refer to Chapter 6 in [38].

Definition 2.4.1. *Let X_t be an $\mathbb{R}^d$-valued semimartingale and let $V : \mathbb{R}^d \to \mathbb{R}^d$ be* *smooth. We define*

$$\int_0^t V(X_s) \circ \mathrm{d}X_s := \int_0^t V(X_s)\mathrm{d}X_s + \frac{1}{2}\sum_{i,j=1}^{d} \int_0^t \frac{\partial V_i}{\partial x_j}(X_s)\,\mathrm{d}[X^i, X^j]_s.$$

The stochastic integral so defined is called the Stratonovich stochastic integral.

In Definition 2.4.1 we interpret $V(x)$ as a row vector. This is in order to match the dimensions of the integrand and the integrator in the stochastic integrals.
For $f \in C_c^\infty(\mathbb{R}^d)$, we can consider the vector field $V = \nabla f$, where we view the gradient as row vector. The following is an immediate consequence of the definition of the Stratonovich integral for this particular vector field and Itô's formula in (2.4.1).

Proposition 2.4.2. *Let X_t be an $\mathbb{R}^d$-valued semimartingale, and let $f \in C_c^\infty(\mathbb{R}^d)$. Then*

$$f(X_t) - f(X_0) = \int_0^t \nabla f(X_s) \circ \mathrm{d}X_s.$$

Next, we want to consider a Stratonovich stochastic differential equation of the form

$$\mathrm{d}X_t = V(X_t) \circ \mathrm{d}W_t, \tag{2.4.2}$$

where W_t is a standard, $\mathbb{R}^l$-valued Brownian motion and $V : \mathbb{R}^d \mapsto \mathbb{R}^{d\times l}$. We can think of the columns of V as a collection of l vector fields $V_1, \dots, V_l$. A precise definition of such an equation is given in [38]. For our purposes, it suffices to know to which Itô stochastic differential equation the equation in (2.4.2) is equivalent. The following theorem can be found in [38].

Theorem 2.4.3. *Let W_t be a standard Brownian motion with values in $\mathbb{R}^l$ and let $V : \mathbb{R}^d \to \mathbb{R}^{d\times l}$ be smooth. Denote by V_j the j-th column of V. A process X_t is a* solution of the Stratonovich stochastic differential equation

$$\mathrm{d}X_t = V(X_t) \circ \mathrm{d}W_t$$

if it is a solution of the Itô stochastic differential equation

$$\mathrm{d}X_t = V(X_t)\,\mathrm{d}W_t + \frac{1}{2}\sum_{j=1}^{l} DV_j(X_t)V_j(X_t)\,\mathrm{d}t.$$

The following result is an extension of Proposition 2.4.2.

Proposition 2.4.4. *Let W_t be a standard, $\mathbb{R}^l$-valued Brownian motion and let $V : \mathbb{R}^d \to \mathbb{R}^{d\times l}$ be smooth. Suppose X_t satisfies*

$$\mathrm{d}X_t = V(X_t) \circ \mathrm{d}W_t.$$

Then for every $f \in C_c^\infty(\mathbb{R}^d)$ we have

$$\mathrm{d}f(X_t) = Vf(X_t) \circ \mathrm{d}W_t.$$

Here Vf should be interpreted as $(V_1 f, \dots, V_l f) \in \mathbb{R}^l$, where $V_1, \dots, V_l$ are the columns of V.

Proof. By Proposition 2.4.2 we have that

$$\mathrm{d}f(X_t) = \nabla f(X_t) \circ \mathrm{d}X_t.$$

Since $\mathrm{d}X_t = V(X_t) \circ \mathrm{d}W_t$, we thus find

$$\mathrm{d}f(X_t) = \nabla f(X_t)V(X_t) \circ \mathrm{d}W_t = Vf(X_t) \circ \mathrm{d}W_t,$$

concluding the proof. □

The generator of a solution

Let A be an operator on $C_b(\mathbb{R}^d)$ with domain $\mathcal{D}(A)$ containing $C_c^\infty(\mathbb{R}^d)$. A Markov process X_t solves the *martingale problem* for A if for all initial distributions of X_0 and all $f \in C_c^\infty(\mathbb{R}^d)$ the process

$$M_t^f := f(X_t) - f(X_0) - \int_0^t Af(X_s)\,\mathrm{d}s$$

is a martingale. In this case, we say that X_t is *generated* by A, or that A is the *generator* of X_t. For more details, we refer to [37].

For solutions of Stratonovich stochastic differential equations, the generator has a particularly simple form.

Proposition 2.4.5. *Let W_t be a standard, $\mathbb{R}^l$-valued Brownian motion and let $V : \mathbb{R}^d \to \mathbb{R}^{d\times l}$ be smooth. Let X_t be the solution of*

$$\mathrm{d}X_t = V(X_t) \circ \mathrm{d}W_t.$$

Then X_t has generator A given by

$$A = \sum_{i=1}^{l} V_i^2,$$

where V_i is the i-th column of V. Here $V_i^2 f$ is given by

$$V_i^2 f(x) = \langle V_i(x), \nabla\langle V_i(x), \nabla f(x)\rangle\rangle.$$

Proof. By Theorem 2.4.3, X_t solves the Itô stochastic differential equation

$$\mathrm{d}X_t = V(X_t)\,\mathrm{d}W_t + \frac{1}{2}\sum_{i=1}^{l} DV_i(X_t)V_i(X_t)\,\mathrm{d}t.$$

From this it follows that the generator $\mathcal{A}$ of X_t is given by

$$\mathcal{A}f(x) = \sum_{j=1}^{d}\left(\frac{1}{2}\sum_{i=1}^{l} DV_i(x)V_i(x)\right)_j \frac{\partial f}{\partial x_j}(x) + \sum_{k,m=1}^{d} (V(x)V(x)^T)_{km}\frac{\partial^2 f}{\partial x_k \partial x_m}(x)$$

$$= \frac{1}{2}\sum_{i=1}^{l} \nabla f(x) DV_i(x) V_i(x) + \sum_{i=1}^{l} V_i(x)^T Hf(x) V_i(x).$$

Here, Hf denotes the Hessian of f.
We now compute $V_i^2 f$. From the product rule it follows that

$$\nabla\langle V_i(x), \nabla f(x)\rangle = \nabla f(x) DV_i(x) + Hf(x) V_i(x).$$

It follows that

$$V_i^2 f(x) = \langle V_i(x), \nabla\langle V_i(x), \nabla f(x)\rangle\rangle = \nabla f(x) DV_i(x) V_i(x) + V_i(x)^T Hf(x) V_i(x).$$

Comparing to the expression we found for $\mathcal{A}f$, we conclude that

$$\mathcal{A}f = \sum_{i=1}^{l} V_i^2 f.$$

□

Finally, like for Itô stochastic differential equations, one can also consider Stratonovich stochastic differential equations with a drift. Since the integral occurring in the drift has nothing to do with being a Stratonovich or Itô integral, this is defined in exactly the same way as usual.

2.4.2. Stochastic differential equations on manifolds

Before we can define stochastic differential equations on a manifold, we first have to define M-valued semimartingales. Since the expectation of a manifold-valued random variable is not well-defined, we cannot use the usual definition. However, a real-valued process X_t is a semimartingale if and only if $f(X_t)$ is a semimartingale for all smooth functions f. Following [57, 36], we make the following definition.

Definition 2.4.6. *Let M be a manifold, $(\Omega, \mathcal{F}, \mathbb{P})$ a filtered probability space and τ a stopping time with respect to the filtration $\mathcal{F}$. An M-*valued semimartingale *is a continuous M-valued process X_t on $[0, \tau)$ such that $f(X_t)$ is a real-valued semimartingale on $[0, \tau)$ for all $f \in C^\infty(M)$.*

Remark 2.4.7. The approach for defining manifold-valued semimartingales is not suitable for defining martingales with values in M. Indeed, in general $f(X_t)$ is not a martingale when X_t is a (real-valued) martingale. However, it is possible to define martingales on a manifold once the manifold is equipped with a connection. In that case, different connections generally give different martingales. We refer to [36] and [57, Section 2.5].

Semimartingales with values in M will serve as solutions of stochastic differential equations on M, which we define next. Let $V_1, \ldots, V_l$ be l vector fields on M and let Z be an $\mathbb{R}^l$-valued semimartingale. We consider the stochastic differential equation

$$\mathrm{d}X_t = V_i(X_t) \circ \mathrm{d}Z_t^i, \tag{2.4.3}$$

where X_0 is given, and may be random. The notion of a solution to (2.4.3) is inspired by Proposition 2.4.4.

Definition 2.4.8. *Let Z_t be an $\mathbb{R}^l$-valued semimartingale and let $V_1, \ldots, V_l$ be d vector fields on M. An M-valued semimartingale X_t defined up to a stopping time τ is a* solution *of*

$$\mathrm{d}X_t = V_i(X_t) \circ \mathrm{d}Z_t^i \tag{2.4.4}$$

up to time τ if

$$\mathrm{d}f(X_t) = \int_0^t V_i f(X_s) \circ \mathrm{d}Z_s^i$$

for all $f \in C^\infty(M)$.

In general, a solution to (2.4.4) only exists up to some stopping time $e(V, Z)$, called its *explosion time*. However, since $V_1, \ldots, V_l$ are smooth, we can show that the solution is unique. For this we use the fact Stratonovich stochastic differential equations as in (2.4.4) behave well under push-forward via a diffeomorphism. This illustrates well why we need to consider Stratonovich stochastic differential equations rather than the Itô variant. The following is Proposition 1.2.4 in [57].

Proposition 2.4.9. *Let $\phi : M \to N$ be a diffeomorphism and suppose that X is a solution of*

$$\mathrm{d}X_t = V_i \circ \mathrm{d}Z_t^i$$

on M with given initial value X_0. Then $Y = \phi(X)$ is a solution of

$$\mathrm{d}Y_t = \phi_* V_i \circ \mathrm{d}Z_t^i$$

on N with given initial value $Y_0 = \phi(X_0)$. Here, $\phi_* V_i$ denotes the push-forward of V_i via ϕ, which is sometimes also written as $\mathrm{d}\phi(V_i)$.

To show that equation (2.4.4) has a unique solution, we take the following approach. First observe that by Whitney's embedding theorem, we can embed M into some Euclidean space $\mathbb{R}^N$. We denote this embedding by ι. Since ι is a diffeomorphism, using Proposition 2.4.9, we have that X_t is a solution of (2.4.4) if and only if $\iota(X_t)$ is a solution of

$$\mathrm{d}Y_t = \iota_* V_i \circ \mathrm{d}Z_t^i.$$

Since $\iota(M)$ is closed in $\mathbb{R}^N$, we can extend the vector fields $\iota_* V_i$ to vector fields $\tilde{V}_i$ on $\mathbb{R}^N$. Since these vector fields are locally Lipschitz, the solution of the so obtained stochastic differential equation on $\mathbb{R}^N$ is unique. Since the vector fields $\tilde{V}_i$ are tangent to $\iota(M)$, every solution started in $\iota(M)$ will remain inside $\iota(M)$. Because the solutions of the equation in $\mathbb{R}^N$ are in one to one correspondence with solutions of (2.4.4), the uniqueness carries over.

2.4.3. Stochastic horizontal lift and development

Given a smooth curve $w : [0,1] \to \mathbb{R}^d$, its development onto the frame bundle of M is the solution of the differential equation

$$\dot{u}(t) = H(u(t))\dot{w}(t),$$

see Definition 2.3.5. Using the canonical horizontal vector fields defined in (2.3.5), we can also write this equation as

$$\dot{u}(t) = H_i(u(t))\dot{w}^i(t).$$

Here, $\dot{w}^i$ denotes the i-th coordinate of $\dot{w}$.

Now, let Z_t be a semimartingale with values in $\mathbb{R}^d$. With the transfer principle of Malliavin in mind (see e.g. [72]), we consider the stochastic differential equation on the frame bundle FM given by

$$\mathrm{d}U_t = \sum_{i=1}^{d} H_i(U_t) \circ \mathrm{d}Z_t^i. \tag{2.4.5}$$

In analogy to the case of smooth curves, the solution U_t to (2.4.5) should be a horizontal process. We use this as our definition, see also [57].

Definition 2.4.10. *An FM-valued semimartingale U_t is called* horizontal *if there exists an $\mathbb{R}^d$-valued semimartingale Z_t such that* (2.4.5) *is satisfied. The process Z_t is called an* anti-development *of U_t. Furthermore, we call U_t the* development *of Z_t onto FM and $X_t := \pi U_t$ the* development *of Z_t onto M.*

It follows from Theorem 2.3.4 in [57] that the anti-development Z_t of a horizontal semimartingale U_t is unique once Z_0 is fixed.

With a notion of a horizontal process at hand, we can define what we mean by a horizontal lift of a process in M.

Definition 2.4.11. *Let X_t be a semimartingale with values in M. A semimartingale U_t in FM is a* horizontal lift *of X_t if U_t is horizontal and $\pi U_t = X_t$.*

Following the relation between parallel transport and a horizontal lift in (2.3.3), we can use the stochastic horizontal lift to define parallel transport along a semimartingale.

Definition 2.4.12. *Let X_t be a semimartingale with values in M and let U_t be a horizontal lift. We define* parallel transport *along X_t as the map $\tau_{X_0 X_t} : T_{X_0}M \to T_{X_t}M$ by*

$$\tau_{X_0 X_t} = U_t U_0^{-1}.$$

2.4.4. Riemannian Brownian motion

In Euclidean space, a standard Brownian motion is generated by $\frac{1}{2}\Delta$, where Δ denotes the Laplacian. In order to have a notion of a Laplacian on a manifold, we need a Riemannian metric. A Riemannian manifold possesses a natural analogue of the Laplacian, namely the Laplace-Beltrami operator. It is denoted by Δ_M and is defined in coordinates by

$$\Delta_M = \frac{1}{\sqrt{\det G}} \frac{\partial}{\partial x^i} \left(\sqrt{\det G} g^{ij} \frac{\partial}{\partial x^j} \right).$$

Here $G = (g_{ij})$ is the matrix of coefficients of the Riemannian metric with inverse denoted by $G^{-1} = (g^{ij})$. One can show that the definition of Δ_M is independent of the coordinates used.

With the Laplace-Beltrami operator defined, we make the following definition, inspired by the Euclidean setting.

Definition 2.4.13 (Riemannian Brownian motion). *A continuous M-valued process W_t is a* Riemannian Brownian motion *if for all $f \in C_c^\infty(M)$,*

$$f(W_t) - f(W_0) - \frac{1}{2} \int_0^t \Delta_M f(W_s) \, \mathrm{d}s$$

is a local martingale up to the explosion time of W_t.

Note that a priori there is no guarantee that Riemannian Brownian motion is defined for all times $t > 0$. It turns out that this depends on the geometry of M. We make the following definition.

Definition 2.4.14 (Stochastic completeness). *We say that a Riemannian manifold is* stochastically complete *if the explosion time of its Riemannian Brownian motion is almost surely infinite.*

The following proposition gives an important sufficient geometric condition for stochastic completeness (see e.g. [57, Section 4.2]).

Proposition 2.4.15. *Let (M, g) be a Riemannian manifold. Assume there exists a finite constant $L \in \mathbb{R}$ such that* Ric $\geqslant L$*. Then M is stochastically complete.*

Using the stochastic development, it is possible to develop a standard Euclidean Brownian motion onto M. In order for the resulting process in M to be a Riemannian Brownian motion, we need to restrict the development procedure to only use orthonormal frames. The reason for this is that the equality in (2.4.6) only holds for orthonormal frames.

For this, we define the set of orthonormal frames for $(T_xM, g(x))$ by

$$O_xM = \{u \in F_xM | u : \mathbb{R}^d \to (T_xM, g(x)) \text{ isometry}\}.$$

The *orthonormal frame bundle* is the bundle OM with fibres O_xM. It is a principal bundle with structure group $O(d)$, the orthogonal group. As explained in Section 2.3.3, all constructions of horizontal lift, development and anti-development are also valid in OM.

If W_t is a standard Euclidean Brownian motion, then its stochastic development onto OM satisfies the equation

$$\mathrm{d}U_t = H_i(U_t) \circ \mathrm{d}W_t^i$$

with $U_0 \in OM$ almost surely. Recalling Proposition 2.4.5, we see that the process U_t has a generator given by

$$\Delta_{OM} := \frac{1}{2}\sum_{i=1}^{d} H_i^2,$$

which is known as Bochner's horizontal Laplacian. It is the horizontal lift of the Laplace Beltrami operator Δ_M in the following sense: for all $f \in C^\infty(M)$ we have (see [57, Proposition 3.1.2])

$$\Delta_M f \circ \pi = \Delta_{OM}(f \circ \pi) \tag{2.4.6}$$

on OM. This motivates the following result, which is Proposition 3.2.1 in [57].

Proposition 2.4.16. *Let W_t be a standard Euclidean Brownian motion, and let U_t be a stochastic development of W_t onto OM. Then the process $X_t := \pi U_t$ is a Riemannian Brownian motion in (M, g).*

If X_t is the development of W_t onto M, then W_t is the anti-development of X_t. This observation gives the following result.

Proposition 2.4.17. *A continuous semimartingale X_t with values in M is a Riemannian Brownian motion if and only if its anti-development via OM is a standard Brownian motion in $\mathbb{R}^d$.*

II

Large deviations for processes in Riemannian manifolds

3

Large deviations for geodesic random walks

In this chapter we prove a generalization of Cramér's theorem (Theorem 2.1.10) to the setting of a Riemannian manifold (M, g). For this, we define the appropriate analogue of a random walk in $\mathbb{R}^N$, namely a geodesic random walk. The result is proven by performing a careful analysis of the geometry behind geodesic random walks, in conjuction with the Gärtner-Ellis theorem (Theorem 2.1.12) in Euclidean space. The results presented here are based on:

Rik Versendaal. "Large deviations for geodesic random walks". In: *Electron. J. Probab.* 24 (2019), Paper No. 93, 39 pp.

Random walks are among the most extensively studied discrete stochastic processes. Given a sequence of random variables $\{X_n\}_{n\geqslant 1}$ in some vector space V, one defines the random walk with increments $\{X_n\}_{n\geqslant 1}$ as the random variable

$$S_n = \sum_{i=1}^n X_i.$$

When rescaled by a factor $\frac{1}{n}$, one can study large deviations for the so obtained sequence $\{\frac{1}{n}S_n\}_{n\geqslant 1}$. Recall from Section 2.1.2 that when the increments are independent and identically distributed, Cramér's theorem (Theorem 2.1.10) states that the sequence $\{\frac{1}{n}S_n\}_{n\geqslant 1}$ satisfies the large deviation principle. Intuitively, this means that there is some rate function $I : V \to [0, \infty]$ such that

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

More specifically, the rate function is given as the Legendre transform of the log moment generating function of the increments, i.e.,

$$I(x) = \sup_{\lambda \in V} \left\{\langle \lambda, x\rangle - \Lambda(\lambda)\right\},$$

where $\Lambda(\lambda) = \log \mathbb{E}(e^{\langle \lambda, X_1 \rangle})$. One may weaken the independence assumption to obtain for example the Gärtner-Ellis theorem, see Theorem 2.1.12 in Section 2.1.2. Also, Cramér's theorem can be generalized to the setting of topological vector spaces or Banach spaces. Furthermore, Cramér's theorem provides a basis for path space large deviations, such as Mogulskii's theorem (random walks) and Schilder's theorem (Brownian motion), see e.g. [29, 88, 30].

However, the analogue of Cramér's theorem in the Riemannian setting was originally obtained as a consequence of the generalization of Mogulskii's theorem, as explained in Section 5.1. Indeed, since evaluation in the end point of trajectories is a continuous map, Cramér's theorem then follows from Mogulskii's theorem by an application of the contraction principle (Theorem 2.1.6). The results in Chapter 5 are obtained via a general approach using convergence of non-linear semigroups and viscosity solutions to Hamilton Jacobi equations as initiated in [39]. A drawback of this approach is that it is only suitable for Markov processes. Therefore, it does not allow for extensions to a setting where the increments of the random walk may be dependent. This causes an obstruction in finding a Riemannian analogue of the Gärtner-Ellis theorem for example. Additionally, the order of first proving Mogulskii's theorem and then deducing Cramér's theorem is historically unnatural. It is thus a fair question to ask whether there is a more direct approach to proving Cramér's theorem to fix this discrepancy.

It turns out that it is possible to only study the underlying geometry of a geodesic random walk in order to prove Cramér's theorem. This gives us new insight in what geometrical aspects allow us to still obtain the large deviation principle for rescaled geodesic random walks, even though the geodesic random walk is in general no longer a simple function of its increments. Apart from large deviations, the geometric results also allow us to obtain Gaussian concentration inequalities for geodesic random walks. Furthermore, this geometric approach does not rely on the fact that the random walk is a Markov process, and thus seems suitable to be extended to random walks with dependent increments for example.

The main difficulty in the Riemannian setting, is that we lack a vector space structure to define a random walk as sum of increments. The appropriate analogue is a geodesic random walk as introduced by Jørgensen in [58]. To define a geodesic random walk, we need to find a replacement for the additive structure, as well as a generalization of the increments. It turns out that as increments one uses tangent vectors, while the additive structure is replaced by an application of the Riemannian exponential map.

More precisely, we introduce a family of probability measures $\{\mu_x\}_{x \in M}$ such that for each $x \in M$, μ_x is a measure on $T_x M$, the tangent space at x. These measures $\{\mu_x\}_{x \in M}$ provide the space-dependent distribution of the increments. Now we start a random walk at some initial point $Z_0 = x_0 \in M$. Then recursively, we define for

$k = 0, \dots, n-1$ the random variable

$$Z_{k+1} = \mathrm{Exp}_{Z_k}\left(\frac{1}{n} X_{k+1}\right),$$

where X_{k+1} is distributed according to μ_{Z_k}. Hence, the random variable Z_n takes values in M and is the natural analogue of the empirical average of the increments $X_1, \dots, X_n$. In Euclidean space, this definition reduces to the usual one, because the Riemannian exponential map is simply vector addition, i.e.,

$$\mathrm{Exp}_x v = x + v.$$

To obtain an analogue of Cramér's theorem, we also need to generalize the notion of the increments of the random walk being identically distributed, since the increments are no longer in the same space. To compare two distributions μ_x and μ_y, we need to identify the tangent spaces T_xM and T_yM. We do this by taking a curve γ connecting x and y and using parallel transport along γ. Because different curves lead to different identifications, we say that the distributions μ_x and μ_y are identical if for all curves γ connecting x and y we have

$$\mu_x = \mu_y \circ \tau_{yx;\gamma}^{-1}.$$

Here, τ denotes parallel transport. Equivalently, one can characterize this property by assuming that the log moment generating functions are invariant under parallel transport, i.e.,

$$\Lambda_x(\lambda) = \Lambda_y(\tau_{xy;\gamma}\lambda),$$

where $\Lambda_x(\lambda) = \log \int_{T_xM} e^{\langle \lambda, v\rangle} \mu_x(\mathrm{d}v)$.

In Euclidean space, the end point of the random walk is a simple function of the increments. In the Riemannian setting, curvature ensures that this is in general no longer the case. For example, the endpoint in general depends on the order of the increments. Nonetheless, it is possible to utilize the vector space structure of the tangent spaces. Denote by $\frac{1}{n}\sum_{i=1}^n \tilde{X}_i$ the empirical average of the appropriately transported increments in $T_{x_0}M$, were x_0 is the starting point of the random walk. By controlling the error induced by the curvature, the large deviations for the geodesic random walk Z_n can be obtained from the large deviations for $\frac{1}{n}\sum_{i=1}^n \tilde{X}_i$. To support this claim, we can also define an alternative random walk in M. For this, we take a sequence of independent, identically distributed random variables $\{Y_n\}_{n\geqslant 1}$ in $T_{x_0}M$ with distribution μ_{x_0} and consider the process

$$\tilde{Z}_n = \mathrm{Exp}_{x_0}\left(\frac{1}{n}\sum_{i=1}^n Y_i\right).$$

In general, $\tilde{Z}_n$ is different from Z_n, even in distribution. Although our method of proving the large deviation principle for Z_n does not immediately allow us to conclude that Z_n and $\tilde{Z}_n$ are exponentially equivalent, the main idea of our proof does rely on the fact that we can (in some sense) relate and compare the geodesic random walk to a sum of independent, identically distributed random variables in the tangent space at x_0, following the distribution μ_{x_0}.

This chapter is organised as follows. In Section 3.1 we introduce Jacobi fields, which are essential for the geometric approach we take in this chapter. Section 3.2 introduces the geodesic random walks, which are the main objects of interest. In Section 3.3 we give the precise statement of Cramér's theorem for geodesic random walks. Additionally, we provide an overview of the various steps that are needed for the proof. In Section 3.4 we obtain a Taylor expansions of the Riemannian exponential map with appropriate error bound. Furthermore, we compare the differential of the exponential map to parallel transport. We also provide bounds for how far geodesics, possibly starting at different points, can spread in a given amount of time. Finally, we show that convex functionals which are invariant under parallel transport are minimized by geodesics. These geometric results are key ingredients in the proof of Cramér's theorem, which is given in Section 3.5. We conclude this chapter with some Gaussian concentration inequalities for geodesic random walks in Section 3.6.

3.1. Some additional Riemannian geometry

In this chapter, we work in a complete Riemannian manifold (M, g) of dimension N. Let d be the associated Riemannian distance, and ∇ the Levi-Civita connection. For a curve $\gamma : [a, b] \to M$, we write $\tau_{\gamma(a)\gamma(b);\gamma}$ for parallel transport along γ. For the sake of readability, we omit the reference to γ when the curve is understood.

Since M is complete, we have that for every $x \in M$ the Riemannian exponential map Exp_x is defined on all of T_xM. Recall that for $x \in M$, the injecitivity radius $\iota(x) \in (0, \infty]$ is defines as

$$\iota(x) = \sup\{\delta > 0 | \mathrm{Exp}_x \text{ is injective on } B(0, \delta)\}.$$

For a set $A \subset M$, the injecitivity radius of A is defined by

$$\iota(A) = \inf\{\iota(x) | x \in A\}. \tag{3.1.1}$$

It follows from Proposition 2.2.6 in Section 2.2 that if $K \subset M$ is compact, then $\iota(K) > 0$.

3.1.1. Calculus of variations

Calculus of variations is often used in optimizing functionals over trajectories, such as finding trajectories of minimum length or minimum energy. For our exposition, we follow roughly the approach taken in [69]. Other references include [86, 40].

We start out by defining what we mean by a variation of a curve.

Definition 3.1.1. *Let $\gamma : [0, 1] \to M$ be a piecewise smooth curve, i.e., γ is continuous and there exist $0 = a_0 < a_1 < \cdots < a_k = 1$ such that γ is smooth on $[a_{i-1}, a_i]$*

for all $i = 1, \ldots, k$*. A* variation *of* γ *is a continuous map* $\Gamma : (-\varepsilon, \varepsilon) \times [0,1] \to M$ *which is smooth on every rectangle* $(-\varepsilon, \varepsilon) \times [a_{i-1}, a_i]$ *and such that* $\Gamma(0,t) = \gamma(t)$ *for all* $t \in [0,1]$*.*

A variation Γ of a curve γ induces two types of curves. We refer to the curves $\Gamma_s(t) := \Gamma(s,t)$ as *main curves* and to $\Gamma_t(s) := \Gamma(s,t)$ as *transverse curves.*
We write

$$\partial_s \Gamma(s,t) := \frac{\mathrm{d}}{\mathrm{d}s} \Gamma_t(s) \qquad \text{and} \qquad \partial_t \Gamma(s,t) := \frac{\mathrm{d}}{\mathrm{d}t} \Gamma_s(t).$$

We use variations of curves to find optima of functionals of curves. For this, we would like to differentiate the functional. We therefore need the rate of change of the variation of curves in the transverse direction. We make the following definition.

Definition 3.1.2. *Let* Γ *be a variation of a curve* γ*. The vector field* V *along* γ *defined by*

$$V(t) := \partial_s \Gamma(0,t)$$

is called the variational vector field *of* Γ*.*

Furthermore, given a continuous map $\Gamma : (-\varepsilon, \varepsilon) \times [0,1] \to M$, we define a *vector field along* Γ as a continuous map $V : (-\varepsilon, \varepsilon) \times [0,1] \to TM$ with $V(s,t) \in T_{\Gamma(s,t)}M$ for every $(s,t) \in (-\varepsilon, \varepsilon) \times [0,1]$ and such that V is smooth wherever Γ is. Whenever well-defined, we write $D_s V$, respectively $D_t V$ for the covariant derivative of V in the direction of the main curves and the transverse curves of Γ. More precisely, we define

$$D_s V(s,t) := \nabla_{\partial_s \Gamma(s,t)} V(s,t) \qquad \text{and} \qquad D_t V(s,t) := \nabla_{\partial_s \Gamma(s,t)} V(s,t).$$

Because the Levi-Civita connection is symmetric, we obtain the following symmetry lemma, see e.g. [69, Lemma 6.3] or [40, Theorem 10.1].

Lemma 3.1.3 (Symmetry lemma)**.** *Let* $\gamma : [0,1] \to M$ *be a smooth curve and* $\Gamma : (-\varepsilon, \varepsilon) \times [0,1] \to M$ *a variation of* γ*. If* M *is equipped with the Levi-Civita connection, then*

$$D_s \partial_t \Gamma(s,t) = D_t \partial_s \Gamma(s,t).$$

3.1.2. Jacobi fields

Suppose $\gamma : [0,1] \to M$ is a geodesic. Let $\Gamma : (-\varepsilon, \varepsilon) \times [0,1] \to M$ be a variation of γ such that for every $s \in (-\varepsilon, \varepsilon)$, the curve $\Gamma_s(t) = \Gamma(s,t)$ is a geodesic. We call Γ a *variation of geodesics*, and the corresponding variational vector field is called a *Jacobi field* along γ.

It is possible to derive a second order differential equation satisfied by Jacobi fields. One can show (see e.g. [69, Theorem 10.2] or [40, Section 10.1]) that a Jacobi field $J(t)$ along a geodesic γ satisfies

$$D_t^2 J(t) + R(J(t), \dot{\gamma}(t))\dot{\gamma}(t) = 0. \tag{3.1.2}$$

Here, R denotes the Riemann curvature endomorphism, see Section 2.2.4. Equation (3.1.2) is called the Jacobi equation.

If $J(0) = 0$ and $\dot{J}(0)$ is given, we can give an explicit formula for the Jacobi field along γ satisfying these conditions. Note that for every $v \in T_xM$, the curve $\gamma_v(t) = \mathrm{Exp}_x(tv)$ is a geodesic. Since a Jacobi field is intuitively the derivative of a variation of geodesics, it is not surprising that the differential of the Riemannian exponential map plays a role in the theory of Jacobi fields.

The differential $\mathrm{d}(\mathrm{Exp}_x)$ of the Riemannian exponential map at x is a linear map from $T(T_xM)$ into TM. Upon identifying $T_v(T_xM)$ with T_xM, we find that for any $v \in T_xM$ we have

$$\mathrm{d}(\mathrm{Exp}_x)_v : T_xM \to T_{\mathrm{Exp}_x v}M.$$

This map is sometimes also written as $\mathrm{d}(\mathrm{Exp}_x)(v)$. We can use this map to write down Jacobi fields with $J(0) = 0$.

Proposition 3.1.4. *Let γ be a geodesic. Then*

$$J(t) = \mathrm{d}(\mathrm{Exp}_{\gamma(0)})_{t\dot{\gamma}(0)}(tv) \tag{3.1.3}$$

defines a Jacobi field along γ with $J(0) = 0$ and $\dot{J}(0) = v$.

Proof. Consider the variation of geodesic of γ given by

$$\Gamma(t, s) = \mathrm{Exp}_{\gamma(0)}(t(\dot{\gamma}(0) + sv)).$$

Using the chain rule, we find that

$$J(t) = \left.\frac{\mathrm{d}}{\mathrm{d}_s}\right|_{s=0} \mathrm{Exp}_{\gamma(0)}(t(\dot{\gamma}(0) + sv)) = \mathrm{d}(\mathrm{Exp}_{\gamma(0)})(t\dot{\gamma}(0))(tv).$$

Furthermore, we have that

$$\begin{aligned}
\dot{J}(0) &= \lim_{h\to 0} \frac{\tau_{\gamma(h)\gamma(0)}\mathrm{d}(\mathrm{Exp}_{\gamma(0)})_{h\dot{\gamma}(0)}(hv) - J(0)}{h} \\
&= \lim_{h\to 0} \tau_{\gamma(h)\gamma(0)}\mathrm{d}(\mathrm{Exp}_{\gamma(0)})_{h\dot{\gamma}(0)}(v) \\
&= \mathrm{d}(\mathrm{Exp}_{\gamma(0)})_0(v) \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \mathrm{Exp}_{\gamma(0)}(sv) \\
&= v.
\end{aligned}$$

Here we used in the second line that $J(0) = 0$ and that $\tau_{\gamma(h)\gamma(0)}\mathrm{d}(\mathrm{Exp}_{\gamma(0)})_{h\dot{\gamma}(0)}$ is a linear map. In the third line we used continuity, together with the fact that $\tau_{\gamma(0)\gamma(0)}$ is the identity. □

In Euclidean space, the Jacobi field in (3.1.3) reduces to $J(t) = t\dot{J}(0)$, which is indeed the variation field of the variation $\Gamma(t, s) = \gamma(0) + t(\dot{\gamma}(0) + s\dot{J}(0))$.

Properties of Jacobi fields

We conclude this section by collecting some properties of Jacobi fields that we need later on.

Proposition 3.1.5. *Let $\gamma : [0,1] \to M$ be a geodesic and $J(t)$ a Jacobi field along γ. Then*

$$\langle J(t), \dot\gamma(t)\rangle = t\langle \dot J(0), \dot\gamma(0)\rangle + \langle J(0), \dot\gamma(0)\rangle$$

for all $t \in [0,1]$.

Proof. Define $f(t) = \langle J(t), \dot\gamma(t)\rangle$. Then

$$f'(t) = \langle D_t J(t), \dot\gamma(t)\rangle + \langle J(t), D_t\dot\gamma(t)\rangle = \langle D_t J(t), \dot\gamma(t)\rangle,$$

because γ is a geodesic. We are done once we show that $f''(t) = 0$. For this, notice that, using (3.1.2)

$$f''(t) = \langle D_t^2 J(t), \dot\gamma(t)\rangle = -\langle R(J(t), \dot\gamma(t))\dot\gamma(t), \dot\gamma(t)\rangle = 0.$$

Here, the last step follows from the symmetry properties of the Riemann curvature tensor. ☐

Proposition 3.1.6. *Let $\gamma : [0,1] \to M$ be a geodesic and $J(t)$ a Jacobi field along γ. For every $t \in [0,1]$ there exists $\xi_t \in (0,t)$ such that*

$$|\dot J(t)| = |\dot J(0)| - t\frac{1}{|\dot J(\xi_t)|}\langle R(J(\xi_t), \dot\gamma(\xi_t))\dot\gamma(\xi_t), \dot J(\xi_t)\rangle.$$

Proof. Define $f(t) = |\dot J(t)|$. We have

$$\begin{aligned} f'(t) &= \frac{1}{|\dot J(t)|}\langle \ddot J(t), \dot J(t)\rangle \\ &= -\frac{1}{|\dot J(t)|}\langle R(J(t), \dot\gamma(t))\dot\gamma(t), \dot J(t)\rangle. \end{aligned}$$

The statement now follows from the mean-value theorem. ☐

3.2. Geodesic random walks

In order to generalize Cramér's theorem to the setting of Riemannian manifolds, we first need to introduce the appropriate analogue of the sequence $\{\frac{1}{n}\sum_{i=1}^n X_i\}_{n\geqslant 0}$ for a sequence of increments $\{X_n\}_{n\geqslant 1}$. In order to do this, we introduce geodesic random walks, following the construction in [58]. Furthermore, we generalize the notion of identically distributed increments to geodesic random walks and characterize it using log moment generating functions. We conclude by providing some examples of geodesic random walks with identically distributed increments.

3.2.1. Definition of geodesic random walks

We start by defining a *geodesic random walk* $\{\mathcal{S}_n\}_{n\geqslant 0}$ on M with increments $\{X_n\}_{n\geqslant 1}$. For this we need to generalize how to add increments together. This is achieved by using the Riemannian exponential map. Because the space variable determines in which tangent space the increment should be, we have to define the random walk recursively, which is the main difficulty in the definition below.

Definition 3.2.1. *Fix x_0 in M. A pair $(\{\mathcal{S}_n\}_{n\geqslant 0}, \{X_n\}_{n\geqslant 1})$ is called a* geodesic random walk *with increments $\{X_n\}_{n\geqslant 1}$ and started at x_0 if the following hold:*

1. $\mathcal{S}_0 = x_0$,
2. $X_{n+1} \in T_{\mathcal{S}_n}M$ *for all* $n \geqslant 0$,
3. $\mathcal{S}_{n+1} = \mathrm{Exp}_{\mathcal{S}_n}(X_{n+1})$ *for all* $n \geqslant 0$.

In what follows, the sequence $\{X_n\}_{n\geqslant 1}$ of increments will usually be omitted and we simply write that $\{\mathcal{S}_n\}_{n\geqslant 0}$ is a geodesic random walk with increments $\{X_n\}_{n\geqslant 1}$.
Note that in the above definition, we fix nothing about the distribution of the increments $\{X_n\}_{n\geqslant 1}$. The distribution is allowed to depend both on the space variable, as well as on time.
For $M = \mathbb{R}^N$, the Riemannian exponential map can be identified with addition, i.e., $\mathrm{Exp}_x(v) = x + v$. Hence, a geodesic random walk in $\mathbb{R}^N$ reduces to the usual random walk, i.e. $\mathcal{S}_n = \sum_{i=1}^n X_i$.

Next, we introduce the concept of time-homogeneous increments for geodesic random walks. For this, we need to fix the distribution of the increments independent of the time variable. Because the increments can take values in different tangent spaces, we need a collection of measures $\{\mu_x\}_{x\in M}$ such that μ_x is a probability measure on T_xM for every $x \in M$. We denote the set of probability measures on T_xM by $\mathcal{P}(T_xM)$. We have the following definition.

Definition 3.2.2. *Let $\{\mathcal{S}_n\}_{n\geqslant 0}$ be a geodesic random walk with increments $\{X_n\}_{n\geqslant 1}$ and started at x_0. Let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_xM)$ for every $x \in M$. We say the random walk $(\{\mathcal{S}_n\}_{n\geqslant 0}, \{X_n\}_{n\geqslant 1})$ is* compatible *with the collection $\{\mu_x\}_{x\in M}$ if $X_{n+1} \sim \mu_{\mathcal{S}_n}$ for every $n \geqslant 0$.*

Essentially, the collection of measures provides the distributions for the increments of the geodesic random walk. Because the collection of measures is independent of n, the increments are time-homogeneous.

Next, we want to define what it means for the increments of a geodesic random walk to be independent. Because the distribution of increment X_{n+1} depends on $\mathcal{S}_n$, we have that X_{n+1} is in general not independent of $\mathcal{A}_n = \sigma(\{X_1, \ldots, X_n\})$ in the usual sense. However, this dependence is purely geometric, as $\mathcal{S}_n$ simply determines in which tangent space we have to choose X_{n+1}. If this is the only dependence of X_{n+1} on $\mathcal{A}_n$, we say the increments of $\{\mathcal{S}_n\}_{n\geqslant 0}$ are independently distributed. We make this precise in the following definition.

Definition 3.2.3. *Let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_xM)$ for every $x \in M$. Let $\{\mathcal{S}_n\}_{n\geqslant 0}$ be a geodesic random walk with increments $\{X_n\}_{n\geqslant 1}$, compatible with $\{\mu_x\}_{x\in M}$. For every $n \geqslant 1$, define the σ-algebra $\mathcal{F}_n$ by*

$$\mathcal{F}_n = \sigma\left(\{(\mathcal{S}_0, X_1), \dots, (\mathcal{S}_{n-1}, X_n)\}\right).$$

We say the increments of $\{\mathcal{S}_n\}_{n\geqslant 0}$ are independent *if for every $n \geqslant 1$ and all bounded, continuous functions $f : M^n \to \mathbb{R}$ we have*

$$\mathbb{E}\left(f(X_1, \dots, X_n)|\mathcal{F}_{n-1}\right) = \int_{T_{\mathcal{S}_{n-1}}M} f(X_1, \dots, X_{n-1}, v)\mu_{\mathcal{S}_{n-1}}(\mathrm{d}v).$$

Remark 3.2.4. Because $\mathcal{S}_n = \mathrm{Exp}_{\mathcal{S}_{n-1}} X_n$, we have that $\mathcal{S}_n$ is $\mathcal{F}_n$-measurable. From this it follows that $\sigma(\{\mathcal{S}_0, \dots, \mathcal{S}_n\}) \subset \mathcal{F}_n$. However, equality need not hold. Indeed, if the Riemannian exponential map Exp_x is not injective, one cannot retrieve the increments $X_1, \dots, X_n$ from $\mathcal{S}_0, \dots, \mathcal{S}_n$.

Remark 3.2.5. Let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_xM)$ for all $x \in M$. Let $\{\mathcal{S}_n\}_{n\geqslant 0}$ be a geodesic random walk with increments $\{X_n\}_{n\geqslant 1}$ compatible with $\{\mu_x\}_{x\in M}$. Suppose furthermore that the increments are independent. Then $\{\mathcal{S}_n\}_{n\geqslant 0}$ is a time-homogeneous, discrete time Markov process on M with transition operator

$$Pf(x) = \mathbb{E}\left(f(\mathcal{S}_1)|\mathcal{S}_0 = x\right) = \int_{T_xM} f(\mathrm{Exp}_x(v))\mu_x(\mathrm{d}v).$$

This is the point of view taken in Chapter 5, in particular in Section 5.3.

Rescaled geodesic random walks

In Euclidean space, one commonly encounters rescaled versions of a random walk, for example for laws of large numbers and central limit theorems. On a general manifold, this rescaling cannot be achieved by multiplication.

Before we define the appropriate analogue of $\left\{\frac{1}{n}\sum_{i=1}^n X_i\right\}_{n\geqslant 0}$, we first need to define how to rescale a geodesic random walk by a factor $\alpha > 0$ independent of n. Note that in Euclidean space we can write $\alpha\sum_{i=1}^n X_i = \sum_{i=1}^n(\alpha X_i)$. This shows that we should rescale the increments of the random walk, which is possible in a manifold because the increments are tangent vectors.

Definition 3.2.6. *Fix x_0 in M and $\alpha > 0$. A pair $(\{(\alpha * \mathcal{S})_n\}_{n\geqslant 0}, \{X_n\}_{n\geqslant 1})$ is called an α-*rescaled geodesic random walk *with increments $\{X_n\}_{n\geqslant 1}$ and started at x_0 if the following hold:*

1. $(\alpha * \mathcal{S})_0 = x_0$,
2. $X_{n+1} \in T_{(\alpha*\mathcal{S})_n}M$ *for all* $n \geqslant 0$,
3. $(\alpha * \mathcal{S})_{n+1} = \mathrm{Exp}_{(\alpha*\mathcal{S})_n}(\alpha X_{n+1})$ *for all* $n \geqslant 0$.

As with geodesic random walks, we will often omit the sequence of increments and simply write that $\{(\alpha * \mathcal{S})_n\}_{n\geqslant 0}$ is an α-rescaled geodesic random walk with increments $\{X_n\}_{n\geqslant 1}$.

Note that an α-rescaled geodesic random walk can itself be considered as a geodesic random walk. Indeed, if $(\alpha * \mathcal{S})_n$ is an α-rescaled geodesic random walk with increments $\{X_n\}_{n\geqslant 1}$, then it is a geodesic random walk with increments $\{\alpha X_n\}_{n\geqslant 1}$.

As for geodesic random walks, we say that an α-rescaled geodesic random walk $\{(\alpha * \mathcal{S})_n\}_{n\geqslant 0}$ with increments $\{X_n\}_{n\geqslant 1}$ is compatible with a collection of probability measures $\{\mu_x\}_{x\in M}$ if $X_{n+1} \sim \mu_{(\alpha*\mathcal{S})_n}$ for every $n \geqslant 0$. It follows that when considered as geodesic random walk, $\{(\alpha * \mathcal{S}_n)\}_{n\geqslant 0}$ is compatible with the collection of measures $\{\mu_x^\alpha\}_{x\in M}$ given by

$$\mu_x^\alpha = \mu_x \circ m_\alpha^{-1}$$

where $m_\alpha : T_xM \to T_xM$ denotes multiplication by α, i.e., $m_\alpha(v) = \alpha v$.

Empirical average process

We conclude this section by introducing the analogue of the sequence of empirical averages $\{\frac{1}{n}\sum_{i=1}^n X_i\}_{n\geqslant 0}$ for a sequence $\{X_n\}_{n\geqslant 1}$ of random variables.

Fix $x_0 \in M$ and let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_xM)$ for all $x \in M$. For every $n \geqslant 1$, let $\{\left(\frac{1}{n} * \mathcal{S}\right)_j\}_{j\geqslant 0}$ be a $\frac{1}{n}$-rescaled geodesic random walk started at x_0 with increments $\{X_j^n\}_{j\geqslant 1}$, compatible with the measures $\{\mu_x\}_{x\in M}$. By considering the diagonal elements of $\{\left(\frac{1}{n} * \mathcal{S}\right)_j\}_{n\geqslant 1, j\geqslant 0}$, we obtain for every $n \geqslant 1$ a random variable $\left(\frac{1}{n} * \mathcal{S}\right)_n$ in M. If we now set the initial value of the sequence $\{\left(\frac{1}{n} * \mathcal{S}\right)_n\}_{n\geqslant 0}$ to be x_0, we obtain the Riemannian analogue of the sequence $\{\frac{1}{n}\sum_{i=1}^n X_i\}_{n\geqslant 0}$. We refer to this process as the *empirical average process* started at x_0, compatible with the collection of measures $\{\mu_x\}_{x\in M}$.

3.2.2. Identically distributed increments

For our purposes, we also need a notion of identically distributed increments. In general, the increments of a geodesic random walk do not live in the same tangent space. In order to overcome this problem, we use parallel transport to identify tangent spaces. Because the identification via parallel transport depends on the curve along which the vectors are transported, we need to make the following definition.

Definition 3.2.7. *Let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_xM)$ for all $x \in M$. Let $\{\mathcal{S}_n\}_{n\geqslant 0}$ be a geodesic random walk with increments $\{X_n\}_{n\geqslant 1}$, compatible with $\{\mu_x\}_{x\in M}$. We say the increments $\{X_n\}_{n\geqslant 1}$ are* identically distributed *if the measures satisfy the following* consistency property*: for any $y, z \in M$ and any smooth curve $\gamma : [a, b] \to M$ with $\gamma(a) = y$ and $\gamma(b) = z$ we have*

$$\mu_z = \mu_y \circ \tau_{yz;\gamma}^{-1}.$$

By the transitivity property of parallel transport, one can equivalently define the consistency property to hold for all piecewise smooth curves.

Note that in Euclidean space, our definition of independent increments implies that the measures are independent of the space variable, because parallel transport is the identity map. Hence, our definition reduces to the usual one, as we obtain that every increment has some fixed distribution μ.

In Section 3.2.3 we provide some examples of families of measures $\{\mu_x\}_{x\in M}$ satisfying the consistency property in Definition 3.2.7. Here, we state a noteworthy property of the expectations of such a collection of measures.

Proposition 3.2.8. *Let $\{\mu_x\}_{x\in M}$ be a collection of measures satisfying the consistency property in Definition 3.2.7. Then for every $x, y \in M$ and every curve $\gamma : [0,1] \to M$ with $\gamma(0) = x$ and $\gamma(1) = y$ we have*

$$\tau_{xy;\gamma}\left(\int_{T_xM} v\ \mu_x(\mathrm{d}v)\right) = \int_{T_yM} w\ \mu_y(\mathrm{d}w).$$

Proof. Since parallel transport is linear, we have

$$\tau_{xy;\gamma}\left(\int_{T_xM} v\ \mu_x(\mathrm{d}v)\right) = \int_{T_xM} \tau_{xy;\gamma}v\ \mu_x(\mathrm{d}v) = \int_{T_yM} w\ \mu_x \circ \tau_{xy;\gamma}^{-1}(\mathrm{d}w).$$

Since the collection of measures satisfies the consistency property in Definition 3.2.7, we have

$$\int_{T_yM} w\ \mu_x \circ \tau_{xy;\gamma}^{-1}(\mathrm{d}w) = \int_{T_yM} w,\ \mu_y(\mathrm{d}w)$$

which concludes the proof. □

The consistency property in Definition 3.2.7 may also be characterised by a consistency assumption for the corresponding log-moment generating functions $\Lambda_x : T_xM \to \mathbb{R}$ of μ_x given by

$$\Lambda_x(\lambda) = \log \int_{T_xM} e^{\langle \lambda, v\rangle}\mu_x(\mathrm{d}v).$$

This is recorded in the following proposition.

Proposition 3.2.9. *Let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_xM)$ for every $x \in M$. Assume that $\Lambda_x(\lambda) < \infty$ for all $x \in M$ and all $\lambda \in T_xM$. The following are equivalent:*

(a) The collection $\{\mu_x\}_{x\in M}$ satisfies the consistency property in Definition 3.2.7.

(b) For all $x, y \in M$ and all smooth curves $\gamma : [a,b] \to M$ with $\gamma(a) = x$ and $\gamma(b) = y$ and for all $\lambda \in T_xM$ we have

$$\Lambda_x(\lambda) = \Lambda_y(\tau_{xy;\gamma}\lambda).$$

Proof. We first prove that (a) implies (b). Fix $x, y \in M$ and $\gamma : [a,b] \to M$ a smooth curve with $\gamma(a) = x$ and $\gamma(b) = y$. Let $\lambda \in T_xM$. Writing $\tau_{xy} = \tau_{xy;\gamma}$ we find

$$\Lambda_x(\lambda) = \log \int_{T_xM} e^{\langle \lambda, v\rangle}\mu_x(\mathrm{d}v)$$

$$
\begin{aligned}
&= \log \int_{T_xM} e^{\langle \tau_{xy}\lambda, \tau_{xy}v\rangle} \mu_x(\mathrm{d}v) \\
&= \log \int_{T_yM} e^{\langle \tau_{xy}\lambda, w\rangle} \mu_y(\mathrm{d}w) \\
&= \Lambda_y(\tau_{xy}\lambda).
\end{aligned}
$$

Here, the second line follows from the fact that the inner product is invariant under parallel transport and the third line follows from the consistency assumption of the collection of measures.

For the reverse implication, fix $x, y \in M$ and let $\gamma : [a, b] \to M$ be a smooth curve with $\gamma(a) = x$ and $\gamma(b) = y$. A similar argument as above shows that the log moment generating function of $\mu_x \circ \tau_{xy}^{-1}$ coincides with the log moment generating function of μ_y. Because the moment generating function determines the distribution, we conclude that $\mu_x \circ \tau_{xy}^{-1} = \mu_y$ as desired. □

The *Legendre transform* $\Lambda_x^* : T_xM \to \mathbb{R}$ of Λ_x is defined by

$$
\Lambda_x^*(v) := \sup_{\lambda \in T_xM} \langle \lambda, v\rangle - \Lambda_x(\lambda).
$$

If the collection of log-moment generating functions $\{\Lambda_x\}_{x\in M}$ satisfies the consistency property in (b) of Proposition 3.2.9, then so does the collection $\{\Lambda_x^*\}_{x\in M}$ of their Legendre transforms.

3.2.3. Examples

We give some examples of collections of measures $\{\mu_x\}_{x\in M}$ satisfying Definition 3.2.7

Example 3.2.10 (Uniform distribution on a ball)**.** Fix $r > 0$. For any $x \in M$, let μ_x be the uniform distribution on $\{v \in T_xM \mid |v|_g \leqslant r\} \subseteq T_xM$. To see that this collection of measures satisfies the consistency property, observe that parallel transport is an isometry between tangent spaces. From this it follows that parallel transport maps balls of same radii in different tangent spaces bijectively onto each other.

The next example will be used in a later chapter to indicate the connection between Mogulskii's theorem and Schilder's theorem.

Example 3.2.11 (Normal distribution)**.** We now want to consider geodesic random walks with normally distributed increments. For this, we define what we consider to be a standard normal distribution on T_xM and show that it satisfies the consistency property. We say that V has a standard normal distribution if for some basis (equivalently, all bases) $e_1, \ldots, e_N$ of T_xM it holds that

$$
(V^1, V^2, \ldots, V^N) \sim \mathcal{N}(0, G^{-1}(x))
$$

where $V = V^i e_i$ and $G(x)$ is the matrix of the metric tensor at x with respect to the basis $e_1, \ldots, e_N$. This is well-defined, because $G^{-1}(x)$ transforms tensorially under coordinate transformations.

To show that this collection of measures satisfies the consistency property in Definition 3.2.7, we make use of Proposition 3.2.9. We compute the log moment generating function Λ_x of μ_x. For this, we will show that

$$\langle \lambda, V \rangle_{g(x)} \sim N(0, |\lambda|^2_{g(x)}).$$

for any $\lambda \in T_x M$. To this end, write $v = V^i e_i$ and $\lambda = \lambda^j e_j$. Then

$$\langle \lambda, V \rangle_{g(x)} = \lambda^j V^i g_{ij}(x)$$

Note that this has a normal distribution with mean 0 and variance

$$\lambda^T G(x) G^{-1}(x) G(x) \lambda = |\lambda|^2_{g(x)}.$$

Using this, the log moment generating function becomes

$$\Lambda_x(\lambda) = \log \int_{T_x M} e^{\langle \lambda, v \rangle} \mu_x(\mathrm{d}v) = \frac{1}{2} |\lambda|^2_{g(x)}.$$

Because parallel transport along any smooth curve is an isometry, we find that (b) of Proposition 3.2.9 is satisfied and as a consequence, the collection $\{\mu_x\}_{x \in M}$ satisfies the consistency property in Definition 3.2.7.

Remark 3.2.12. The previous example shows that if we have for all $x \in M$ that $\Lambda_x(\lambda) = F(|\lambda|_{g(x)})$ for some function F, independent of x, then the measures $\{\mu_x\}_{x \in M}$ satisfy the consistency property in Definition 3.2.7. This is for example the case if μ_x conditioned on the norm is uniformly distributed, and the norm is distributed according to a distribution ν independent of x.

Finally, we will show that if a geodesic random walk has identically distributed increments, it is sufficient to know the probability distribution in a given tangent space. This leads to an equivalent characterization of a geodesic random walk.

Example 3.2.13 (Equivalent characterization of a geodesic random walk)**.** Suppose we have fixed an initial point $x_0 \in M$ and a measure μ on $T_{x_0} M$ with the following property: *For every smooth loop $\gamma : [a, b] \to M$ with $\gamma(a) = \gamma(b) = x_0$ it holds that $\mu = \mu \circ \tau_{\gamma(a)\gamma(b);\gamma}$, i.e., μ is invariant under parallel transport along any loop.*

Given such a measure μ, we can construct a family of measures $\{\mu_x\}_{x \in M}$ which satisfies Definition 3.2.7. Indeed, given $x \in M$, we take a smooth curve $\gamma : [a, b] \to M$ with $\gamma(a) = x_0$ and $\gamma(b) = x$ and define $\mu_x = \mu \circ \tau_{\gamma(a)\gamma(b);\gamma}$. The assumption on μ implies that this is well-defined, i.e. independent of the curve γ, and that the given collection of measures satisfies the consistency property. More precisely, by arguing in a chart around x and x_0 respectively, one can make sure to concatenate a smooth curve from x to x_0 to the one from x to x_0 in a smooth way to create a smooth loop.

Now if $\{\tilde{X}_n\}_{n \geqslant 1}$ is a sequence of $T_{x_0} M$-valued random variables with distribution μ, one can parallelly transport these along the path of the geodesic random walk to obtain the sequence $\{X_n\}_{n \geqslant 1}$ of the geodesic random walk.

3.3. Sketch of the proof of Cramér's theorem for Riemannian manifolds

In this section we provide a sketch of the proof of Cramér's theorem for geodesic random walks and stress what observations and properties are important to make the proof work. Before we get to this, let us first state the exact theorem we wish to prove.

3.3.1. Statement of Cramér's theorem

Cramér's theorem is concerned with the large deviations for the empirical average process $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 1}$ with independent, identically distributed increments.
Along with the large deviation principle, we need to identify the rate function. In Euclidean space, the rate function is given by

$$I(x) = \Lambda^*(x),$$

the Legendre transform of the log moment generating function of an increment. Note here that one can consider the vector x as the tangent vector of the straight line from the origin to the point x. Using this viewpoint, the analogue of the rate function in the Riemannian setting should be

$$I(x) = \inf\left\{\Lambda^*_{x_0}(v) | \mathrm{Exp}_{x_0} v = x\right\}.$$

Here, we have to take the infimum, because the Riemannian exponential map is not necessarily injective, i.e., there may be more than one geodesic connecting x_0 and x. We will show that this is indeed the correct rate function, as collected in the following theorem.

Theorem 3.3.1 (Cramér's theorem for Riemannian manifolds)**.** *Let (M, g) be a complete Riemannian manifold. Fix $x_0 \in M$ and let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_x M)$ for all $x \in M$. For every $n \geqslant 1$, let $\{(\frac{1}{n} * \mathcal{S})_j\}_{j\geqslant 0}$ be a $\frac{1}{n}$-rescaled geodesic random walk started at x_0 with independent increments $\{X^n_j\}_{j\geqslant 1}$, compatible with $\{\mu_x\}_{x\in M}$. Let $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$ be the associated empirical average process started at x_0. Assume the increments are bounded and have expectation 0. Assume furthermore that the collection $\{\mu_x\}_{x\in M}$ satisfies the consistency property in Definition 3.2.7. Then $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$ satisfies the large deviation principle in M with good rate function*

$$I_M(x) = \inf\left\{\Lambda^*_{x_0}(v) | \mathrm{Exp}_{x_0} v = x\right\}.$$

Due to geometrical influences, which become apparent when sketching the proof, we prove Cramér's theorem only in the case when the increments are bounded. This allows for a less technical proof of the theorem, but nevertheless introduces all geometrical obstructions that have to be dealt with. The details of the proof can be found in Section 3.5.

Like in the Euclidean setting, we prove Cramér's theorem for geodesic random walks by separately proving the upper and lower bound for the large deviation principle of $\{(\frac{1}{n} * \mathcal{S})_n\}_{n \geqslant 0}$. In Section 3.3.2 we give an overview of the steps one needs to take to prove the upper bound, while in Section 3.3.3 we sketch how to prove the lower bound.

3.3.2. Sketch of the proof of the upper bound

In the Euclidean case, one proves the upper bound in Cramér's theorem by using Chebyshev's inequality. More precisely, the key step is to show that for $\Gamma \subset \mathbb{R}^d$ compact one has (see e.g. [56, 29])

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} S_n \in \Gamma\right) \leqslant -\inf_{x\in\Gamma} \sup_{\lambda \in \mathbb{R}^d} \left\{ \langle \lambda, x\rangle - \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(e^{n\langle \lambda, \frac{1}{n} S_n\rangle}\right)\right\}.$$

The upper bound is then extended to all closed sets by proving exponential tightness. The idea is to follow a similar procedure in the Riemannian case. However, because $(\frac{1}{n} * \mathcal{S})_n$ is M-valued, its moment generating function is not defined.

Step 1: Analogue of the moment generating function $\mathbb{E}(e^{n\langle \lambda, \frac{1}{n} S_n\rangle})$

To overcome the problem of not having a moment generating function of $(\frac{1}{n} * \mathcal{S})_n$, we want to identify points in M with tangent vectors in $T_{x_0}M$. For this we use the Riemannian exponential map. However, this map is not necessarily injective. Hence, we first assume that for each $n \geqslant 1$, the $\frac{1}{n}$-rescaled geodesic random walk stays within the injectivity radius $\iota(x_0)$ of its initial point x_0 up to time n. Because Exp_{x_0} is injective on $B(0, \iota(x_0)) \subset T_{x_0}M$, we can uniquely define $v_k^n \in T_{x_0}M$ satisfying $|v_k^n| < \iota(x_0)$ and

$$\mathrm{Exp}_{x_0}^{-1}(v_k^n) = \left(\frac{1}{n} * \mathcal{S}\right)_k.$$

Ideally, we would like to prove the large deviation principle for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$ by proving the large deviation principle for $\{v_n^n\}_{n\geqslant 0}$ in $T_{x_0}M$ and then apply the contraction principle (see e.g. [29, Chapter 4]) with the continuous function Exp_{x_0}. For this to work, we would need to show that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(e^{n\langle \lambda, v_n^n\rangle}\right) = \Lambda_{x_0}(\lambda).$$

Unfortunately, using the estimate for $\mathbb{E}(e^{n\langle \lambda, v_n^n\rangle})$ found in Step 2 as explained below, we are only able to show that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(e^{n\langle \lambda, v_n^n\rangle}\right) \leqslant \Lambda_{x_0}(\lambda) + C|\lambda| \tag{3.3.1}$$

and likewise

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(e^{n\langle \lambda, v_n^n\rangle}\right) \geqslant \Lambda_{x_0}(\lambda) - C|\lambda|, \tag{3.3.2}$$

where the constant only depends on the curvature and the uniform bound of the increments.

Step 2: Upper bound for the moment generating function of v_n^n

In $\mathbb{R}^d$ we simply have $v_n^n = \frac{1}{n}\sum_{i=1}^n X_i$ and hence its moment generating function is given by

$$\mathbb{E}\left(e^{n\langle\lambda,v_n^n\rangle}\right) = \prod_{i=1}^n \mathbb{E}\left(e^{\langle\lambda,X_i\rangle}\right) = \mathbb{E}\left(e^{\langle\lambda,X_1\rangle}\right)^n.$$

Here we use the fact that we can write $v_k^n = v_{k-1}^n + \frac{1}{n}X_k$. This fails in the Riemannian setting, which results in the fact that we can only estimate $\mathbb{E}(e^{n\langle\lambda,v_n^n\rangle})$ as mentioned above in (3.3.1) and (3.3.2).

In a Riemannian manifold we replace the identity $v_k^n = v_{k-1}^n + \frac{1}{n}X_k$ by the Taylor expansion of $\mathrm{Exp}_{x_0}^{-1}$ (see Section 3.4.1, Proposition 3.4.4). This results in

$$v_k^n = v_{k-1}^n + \frac{1}{n}\mathrm{d}(\mathrm{Exp}_{x_0})_{v_{k-1}^n}^{-1} X_k^n + \mathcal{O}\left(\frac{1}{n^2}\right). \tag{3.3.3}$$

Here one needs to be careful that the constant in the error term may depend on curvature properties of the manifold around $(\frac{1}{n} * \mathcal{S})_{k-1}$. Because we assume the increments are uniformly bounded, there exists a compact set $K \subset M$ such that for all $n \geqslant 1$ and all $0 \leqslant j \leqslant n$ we have $(\frac{1}{n} * \mathcal{S})_j \in K$. This allows us to control the constant in the error term.

However, the problem arises that this expression does not yet allow us to use the assumption that the increments of the geodesic random walk are identically distributed, which essentially means that the distribution of the increments is invariant under parallel transport.

Therefore, we need to argue that $\mathrm{d}(\mathrm{Exp}_{x_0})_{v_{k-1}^n}^{-1}$ can be approximated well enough by parallel transport. It turns out there exists a constant $C > 0$ such that

$$|\mathrm{d}(\mathrm{Exp}_{x_0})_{v_{k-1}^n}^{-1} X_k^n - \tau_{x_0 \frac{1}{n}\mathcal{S}_{k-1}}^{-1} X_k^n| \leqslant C|v_{k-1}^n|^2|X_k^n|, \tag{3.3.4}$$

see Section 3.4.2 for details, in particular Corollary 3.4.8. By the same reasoning as before, the constant C may be controlled independent of k.

Combining (3.3.3) and (3.3.4) and using that $v_n^n = \sum_{k=1}^n v_k^n - v_{k-1}^n$, we have

$$\left| v_n^n - \frac{1}{n}\sum_{k=1}^n \tau_{x_0 \frac{1}{n}\mathcal{S}_{k-1}}^{-1} X_k^n \right| \lesssim \frac{1}{n} + 1. \tag{3.3.5}$$

Using the Cauchy-Schwarz inequality, we now find

$$\begin{aligned}\mathbb{E}(e^{n\langle\lambda,v_n^n\rangle}) &\leqslant e^{C|\lambda|}e^{nC|\lambda|}\mathbb{E}\left(e^{\sum_{i=1}^n \langle\lambda,\tau_{x_0 \frac{1}{n}\mathcal{S}_{n-1}}^{-1} X_k^n\rangle}\right) \\ &= e^{C|\lambda|}e^{nC|\lambda|}\mathbb{E}\left(e^{\langle\lambda,X_1\rangle}\right)^n.\end{aligned} \tag{3.3.6}$$

Here, the last line uses that the increments are independent and identically distributed. From this it follows that

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}(e^{n\langle\lambda,v_n^n\rangle}) \leqslant C|\lambda| + \Lambda_{x_0}(\lambda),$$

so that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(v_n^n \in F) \leqslant -\inf_{v\in F} \sup_{\lambda\in T_{x_0}M} \{\langle \lambda, v\rangle - \Lambda_{x_0}(\lambda) - C|\lambda|\}.$$

It remains to get rid of the $C|\lambda|$ term. In the next step we show how to reduce the order n term in the upper bound in (3.3.6), so that we can still use the above estimating procedure to obtain the upper bound of the large deviation principle for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$.

Step 3: Reducing the upper bound in Step 2 by splitting the random walk in pieces

The problematic factor in estimate (3.3.6) arises from the replacement of the differential of the exponential map with parallel transport as done in Step 2. This error depends on $|v_k^n|$, i.e., the distance from x_0 to $(\frac{1}{n} * \mathcal{S})_k$. Note that in Step 2, we simply estimated $|v_k^n|$ uniformly in k. However, if we write r for the uniform bound on the increments, we actually have $|v_k^n| \leqslant \frac{k}{n} r$. Consquently, we can reduce the upper bound if the amount of steps for which we need to compare parallel transport and the differential of the exponential map becomes smaller.

To do this, the idea is to cut the random walk in finitely many pieces, say m, each consisting of (roughly) $m^{-1}n$ steps. We can then consider each of these pieces as separate random walks which we need to identify with a vector in some tangent space. In the end, we can then let the amount of pieces tend to infinity by considering the limit $m \to \infty$, so that the part of the upper bound which we want to reduce vanishes entirely.

More precisely, fix $m \in \mathbb{N}$, and define for $l = 0, \ldots, m-1$ the indices $n_l = l\lfloor m^{-1}n\rfloor$ and set $n_m = n$. This divides the random walk in m pieces, where a piece starts in $(\frac{1}{n} * \mathcal{S})_{n_l}$ and consists of $\lfloor m^{-1}n\rfloor$ increments. Now recall there is a compact set $K \subset M$ such that for all n and all $0 \leqslant j \leqslant n$ we have $(\frac{1}{n} * \mathcal{S})_j \in K$. Because $\iota(K) > 0$, we can choose m sufficiently large, such that for all n, all $l = 1, \ldots, m$ and all $k = 1, \ldots, \lfloor m^{-1}n\rfloor$ we have

$$\left(\frac{1}{n} * \mathcal{S}\right)_{n_{l-1}+k} \in B\left(\left(\frac{1}{n} * \mathcal{S}\right)_{n_{l-1}}, \iota(K)\right).$$

We may thus follow the same procedure as in Step 1, so that for every $l = 1, \ldots, m$ and every $k = 1, \ldots, \lfloor m^{-1}n\rfloor$ we can uniquely define $\tilde{v}_k^{n,m,l} \in T_{(\frac{1}{n}*\mathcal{S})_{n_{l-1}}}M$ such that

$$\tilde{v}_k^{n,m,l} \in \mathrm{Exp}^{-1}_{(\frac{1}{n}*\mathcal{S})_{n_{l-1}}}\left(\left(\frac{1}{n} * \mathcal{S}\right)_{n_{l-1}+k}\right)$$

and $|\tilde{v}_k^{n,m,l}| < \iota((\frac{1}{n} * \mathcal{S})_{n_{l-1}})$. Finally, we define $v_k^{n,m,l} \in T_{x_0}M$ by

$$v_k^{n,m,l} = \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{n_{l-1}}} \tilde{v}_k^{n,m,l},$$

where the parallel transport can be taken along any path connecting x_0 and $\left(\frac{1}{n} * \mathcal{S}\right)_{n_{l-1}}$, as long as it is measurable with respect to $\mathcal{F}_{n_{l-1}} = \sigma(X_1, \dots, X_{n_{l-1}})$.
This associates to $(\frac{1}{n} * \mathcal{S})_n \in M$ a tuple

$$\left(v^{n,m,1}_{\lfloor m^{-1} n \rfloor}, \dots, v^{n,m,m}_{\lfloor m^{-1} n \rfloor}\right) \in (T_{x_0} M)^m.$$

Following the procedure in Step 2, apart from some technical details, we find

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}\left(e^{n\langle \lambda, v^{n,m,l}_{\lfloor m^{-1} n \rfloor}\rangle}\right) \leqslant C|\lambda| \frac{1}{m^3} + \frac{1}{m} \Lambda_{x_0}(\lambda),$$

for all $\lambda \in T_{x_0} M$. From here it is possible to show that

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}\left(e^{n \sum_{l=1}^m \langle \lambda_l, v^{n,m,l}_{\lfloor m^{-1} n \rfloor}\rangle}\right) \leqslant C \frac{1}{m^3} \sum_{l=1}^m |\lambda_l| + \frac{1}{m} \sum_{l=1}^m \Lambda_{x_0}(\lambda_l)$$

for all $(\lambda_1, \dots, \lambda_m) \in (T_{x_0} M)^m$. We conclude that

$$\begin{aligned}
&\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1} n \rfloor}, \dots, v^{n,m,m}_{\lfloor m^{-1} n \rfloor}\right) \in F\right) \\
&\qquad \leqslant - \inf_{(v_1, \dots, v_m) \in F} \frac{1}{m} \sum_{l=1}^m \sup_{\lambda \in T_{x_0} M} \{\langle \lambda, m v_l \rangle - \Lambda_{x_0}(\lambda) - \frac{1}{m^2} C|\lambda|\}.
\end{aligned}$$

Step 4: Upper bound for the large deviation principle of $\{(\frac{1}{n} * \mathcal{S})_n\}_{n \geqslant 0}$
To prove the large deviation upper bound for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n \geqslant 0}$, we notice that the map sending $(v^{n,m,1}_{\lfloor m^{-1} n \rfloor}, \dots, v^{n,m,m}_{\lfloor m^{-1} n \rfloor})$ to $(\frac{1}{n} * \mathcal{S})_n$ is continuous. Hence, if $F \subset M$ is closed, there exists a closed set $\tilde{F} \subset (T_{x_0} M)^m$ such that

$$\mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in F\right) = \mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1} n \rfloor}, \dots, v^{n,m,m}_{\lfloor m^{-1} n \rfloor}\right) \in \tilde{F}\right).$$

From this it follows that

$$\begin{aligned}
&\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in F\right) \\
&\qquad \leqslant - \inf_{(v_1, \dots, v_m) \in \tilde{F}} \frac{1}{m} \sum_{l=1}^m \sup_{\lambda \in T_{x_0} M} \left\{\langle \lambda, v_l \rangle - \Lambda_{x_0}(\lambda) - \frac{1}{m^2} C|\lambda|\right\}.
\end{aligned}$$

Now note that for every $v \in \mathrm{Exp}_{x_0}^{-1} F$ we have that $(\frac{1}{m} v, \dots, \frac{1}{m} v) \in \tilde{F}$. Furthermore, by convexity, the infimum in the upper bound is attained when all v_i are equal. Therefore, the upper bound reduces to

$$\begin{aligned}
&\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in F\right) \\
&\qquad \leqslant - \inf_{v \in \mathrm{Exp}_{x_0}^{-1} F} \sup_{\lambda \in T_{x_0} M} \left\{\langle \lambda, v \rangle - \Lambda_{x_0}(\lambda) - \frac{1}{m^2} C|\lambda|\right\}.
\end{aligned}$$

The desired upper bound now follows by considering the limit $m \to \infty$.

3.3.3. Sketch of the proof of the lower bound

To prove the lower bound of the large deviation principle for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$, it suffices to show that if $G \subset M$ is open, then

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in G\right) \geqslant -I_M(x),$$

for all $x \in G$. Because $I_M(x) = \inf_{v\in \operatorname{Exp}_{x_0}^{-1} x} \Lambda^*_{x_0}(v)$, it is in fact sufficient to show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in G\right) \geqslant -\Lambda^*_{x_0}(v)$$

for any $v \in \operatorname{Exp}_{x_0}^{-1} G$. Therefore, we again need to transfer the problem to the tangent space $T_{x_0}M$.

Transfer to the tangent space $T_{x_0}M$

Similar to how estimate (3.5.2) is derived, we find that

$$\left| v^n_{\lfloor m^{-1}n\rfloor} - \frac{1}{n} \sum_{k=1}^{\lfloor m^{-1}n\rfloor} \tau^{-1}_{x_0 \frac{1}{n}\mathcal{S}_{k-1}} X^n_k \right| \lesssim \frac{1}{nm} + \frac{1}{m^3}.$$

As a consequence, by choosing m sufficiently large, we can get $v^n_{\lfloor m^{-1}n\rfloor}$ arbitrarily close to $\frac{1}{n}\sum_{k=1}^{\lfloor m^{-1}n\rfloor} \tau^{-1}_{x_0 \frac{1}{n}\mathcal{S}_{k-1}} X^n_k$. The latter is a sum of independent random variables with distribution μ_{x_0}, which is a consequence of the fact that the increments of the geodesic random walk are independent and identically distributed. Hence, by Cramér's theorem for vector spaces we obtain that for every $m \in \mathbb{N}$ the sequence $\{\frac{1}{n}\sum_{k=1}^{\lfloor m^{-1}n\rfloor} \tau^{-1}_{x_0 \frac{1}{n}\mathcal{S}_{k-1}} X^n_k\}_{n\geqslant 0}$ satisfies the large deviation principle in $T_{x_0}M$ with good rate function $I(v) = \frac{1}{m}\Lambda^*_{x_0}(mv)$.

Putting everything together, after some technicalities, we find that if $\varepsilon > 0$ is small enough, there exists a constant $c \in (0,1)$ such that for m large enough

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(v^n_{\lfloor m^{-1}n\rfloor} \in B(v,\varepsilon)) \\
&\quad\geqslant \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{k=1}^{\lfloor m^{-1}n\rfloor} \tau^{-1}_{x_0 \frac{1}{n}\mathcal{S}_{k-1}} X^n_k \in B(v, c\varepsilon^2)\right) \\
&\quad\geqslant \frac{1}{m}\Lambda^*_{x_0}(mv).
\end{aligned} \tag{3.3.7}$$

In order to make use of this fact, we again need to divide the random walk in pieces, like in Step 3 in Section 3.3.2. To this end, we again first identify $(\frac{1}{n} * \mathcal{S})_n \in M$ with a tuple

$$\left(\tilde{v}^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, \tilde{v}^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in T_{(\frac{1}{n}*\mathcal{S})_{n_0}} M \times \cdots \times T_{(\frac{1}{n}*\mathcal{S})_{n_m}} M.$$

However, this time we need to be careful how we transport these vectors to $T_{x_0}M$. Indeed, we wish to do this in such a way that

$$\left(v^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, v^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in B(v, c\varepsilon^2)^m \Rightarrow \left(\frac{1}{n} * \mathcal{S}\right)_n \in B(\mathrm{Exp}_{x_0} v, \varepsilon). \tag{3.3.8}$$

The key to making the correct choice is given by Proposition 3.4.10, which gives us control over how far geodesics can spread in a short time when starting in different points of the manifold. This result shows us how to choose the parallel transport based on the vector v, so that the curvature has only little effect. Essentially, one first transports a vector to an associated point on the geodesic with speed v which connects x_0 and x. After that, one transports the vector along this geodesic to x_0. More precisely, we do the following:

1. Consider the geodesic $\gamma(t) = \mathrm{Exp}_{x_0}(tv)$ and for $i = 0, \ldots, m$ define the points $y_i = \gamma(\frac{i}{m})$. Note that $y_0 = x_0$.
2. For every $i = 0, \ldots, m$ and every $x \in M$, choose a geodesic of minimal length connecting y_i and x and define $\tau_{y_i x}$ to be parallel transport along this geodesic.
3. Now define for $i = 1, \ldots, m$ the vector $v^{n,m,1}_{\lfloor m^{-1}n\rfloor} \in T_{x_0}M$ by

$$v^{n,m,i}_{\lfloor m^{-1}n\rfloor} = \tau^{-1}_{y_0 y_i} \tau^{-1}_{y_i (\frac{1}{n} * \mathcal{S})_{n_{i-1}}} \tilde{v}^{n,m,i}_{\lfloor m^{-1}n\rfloor}$$

Now, given $G \subset M$ open, $x \in G$ and $v \in \mathrm{Exp}^{-1}_{x_0} x$, by (3.3.8) we have

$$\mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in G\right) \geqslant \mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, v^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in B(v, c\varepsilon^2)^m\right).$$

Using this, an approach similar to the one used to obtain (3.3.7), also using that the increments are independent and identically distributed, gives us that

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in G\right) \\
&\quad \geqslant \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, v^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in B(v, c\varepsilon^2)^m\right) \\
&\quad \geqslant -\Lambda^*_{x_0}(v)
\end{aligned}$$

which is as desired.

3.4. Some geometric results

This section focuses on geometric results needed for the proof of Cramér's theorem for geodesic random walks as sketched in Section 3.3. We obtain a Taylor expansion for the inverse Riemannian exponential map and estimate the residual term.

Furthermore, we bound the difference between the differential of the Riemannian exponential map and parallel transport. This heavily relies on the theory of Jacobi fields, which have been introduced in Section 3.1.2. We also show how far geodesics can spread in a short time interval when starting in different points on the manifold. We conclude this section by proving that convex functionals are minimized by geodesics.

3.4.1. Taylor expansion of the inverse Riemannian exponential map

The Riemannian exponential map $\mathrm{Exp}_x : T_xM \to M$ is a local diffeomorphism around 0. More precisely, it is a diffeomorphism between $B(0,\iota(x)) \subset T_xM$ and $\mathrm{Exp}_x(B(0,\iota(x)))$. Now suppose $\gamma(t)$ is a curve in $\mathrm{Exp}_x(B(0,\iota(x)))$. There exists a unique curve $w(t)$ in $B(0,\iota(x)) \subset T_xM$ such that $\mathrm{Exp}_x w(t) = \gamma(t)$. Our aim is to find a Taylor expansion for $w(t)$ around $t = 0$. Although this seems to be folklore, we also find a precise estimate of the residual term of the Taylor approximation.

Before we can do this, we first need two lemmas that will help us control the error term in the first order Taylor polynomial for the inverse of the Riemannian exponential map.

Lemma 3.4.1. *Let $K \subset M$ be compact and for any $x \in K$, let $K_x \subset T_xM$ be compact. Assume the Riemannian exponential map Exp_x is defined on K_x for all $x \in K$. Assume furthermore there exists a $C > 0$ such that $K_x \subset \overline{B(0,C)}$ for any $x \in K$. Then*

$$\sup_{x\in K}\sup_{v\in K_x} |\mathrm{d}(\mathrm{Exp}_x)_v| < \infty$$

Proof. Because the sets K_x are uniformly bounded and K is compact, it follows that

$$\tilde{K} := \{(x,v) \in TM |\ x \in K, v \in K_x\}$$

is compact. By assumption, $\tilde{K}$ is contained in the domain $U \subset TM$ of the Riemannian exponential map.

Let $\pi : TM \to M$ be the canonical projection and define the vertical tangent bundle $T^VTM \subset TTM$ as the kernel of $\mathrm{d}\pi : TTM \to TM$. Furthermore, define

$$D := \{(v,w) \in TM \times TM |\ \pi(v) = \pi(w)\} \subset TM \times TM.$$

Then T^VTM is isomorphic to D via the isomorphism $\iota : D \to T^VTM$ given by

$$\iota(v,w) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} v + tw.$$

Now consider the set

$$B^V\tilde{K} := \{(v,w) \in D |\ v \in \tilde{K}, |w|_g \leqslant 1\} \subset D \simeq T^VTM.$$

Since $\tilde{K}$ is compact, so is $B^V\tilde{K}$.

Now, for $\mathrm{Exp} : U \to M \times M$ we have $\mathrm{dExp} : TU \to TM \times TM$ given by

$$\mathrm{dExp}((x,v))(\tilde{w}) = (0, \mathrm{d}(\mathrm{Exp}_x)_v w),$$

where $\iota(v,w) = (v,\tilde{w})$. From this it follows that the map $((x,v),w) \mapsto |\mathrm{d}(\mathrm{Exp}_x)_v w|_g$ is continuous on D, and as a consequence it is bounded on $B^V\tilde{K}$. This implies that

$$\sup_{x\in K}\sup_{v\in K_x}\sup_{w\in T_xM, |w|_g\leqslant 1} |\mathrm{d}(\mathrm{Exp}_x)_v w|_g < \infty,$$

from which the claim follows, since

$$\sup_{w\in T_xM, |w|_g\leqslant 1} |\mathrm{d}(\mathrm{Exp}_x)_v w|_g = |\mathrm{d}(\mathrm{Exp}_x)_v|_g.$$

□

As long as one restricts to a set where the inverse of the Riemannian exponential map is well-defined, one obtains in a similar way a bound for the differential of the inverse Riemannian exponential map.

Lemma 3.4.2. *Let $K \subset M$ be compact and for any $x \in K$, let $K_x \subset B(0, \underline{\iota(x)) \subset T_xM}$ be compact. Assume that there exists a constant $C > 0$ such that $K_x \subset \overline{B(0,C)}$ for any $x \in K$. Then*

$$\sup_{x\in K}\sup_{v\in K_x} |\mathrm{d}(\mathrm{Exp}_x)_v^{-1}| < \infty.$$

Remark 3.4.3. When we take $K = \{x_0\}$ in Lemma 3.4.2, the statement simplifies as follows: If $\tilde{K} \subset B(0,\iota(x_0))$ is compact, then

$$\sup_{v\in\tilde{K}} |\mathrm{d}(\mathrm{Exp}_{x_0})_v^{-1}| < \infty.$$

We are now in a position to find a first order Taylor expansion of the inverse Riemannian exponential map and control the error term appropriately.

Proposition 3.4.4. *Fix $x_0 \in M$ and let $K \subset B(0,\iota(x_0))$ be compact. Define $\tilde{K} = \mathrm{Exp}_{x_0} K$ and let $x \in \tilde{K}$ and $v \in T_xM$. Consider the geodesic $\gamma_v : [0,T] \to M$ defined by $\gamma_v(t) = \mathrm{Exp}_x(tv)$, where T is such that the image of γ_v is contained in $\tilde{K}$. Restrict Exp_{x_0} to K and set $w(t) = \mathrm{Exp}_{x_0}^{-1}(\gamma_v(t)) \in K \subset T_{x_0}M$. Then there exists a constant $C > 0$ such that*

$$|w(t) - w(0) - t\mathrm{d}(\mathrm{Exp}_{x_0})_{w(0)}^{-1}(v)|_{g(x_0)} \leqslant Ct^2|v|^2_{g(x_0)}$$

for all $t \in [0,T]$. Here, the constant C only depends on the compact set $\tilde{K}$ (and the dimension of M).

Proof. Let $\{e_1,\ldots,e_d\}$ be an orthonormal basis of $T_{x_0}M$ and consider the associated normal coordinates around x_0. Since $K \subset B(0,\iota(x_0))$, these normal coordinates are defined in a neighbourhood of $\tilde{K}$.

Writing $\gamma_v^k(t)$ for the coordinates of $\gamma_v(t)$, we have that $w(t) = \gamma_v^k(t)e_k$. Furthermore, by the chain rule we have

$$\dot{w}(0) = \mathrm{d}(\mathrm{Exp}_{x_0})^{-1}_{w(0)}(\dot{\gamma}_v(0)) = \mathrm{d}(\mathrm{Exp}_{x_0})^{-1}_{w(0)}(v).$$

As a consequence, we find that

$$\begin{aligned}|w(t) - w(0) - t\mathrm{d}(\mathrm{Exp}_{x_0})^{-1}_{w(0)}(v)|_{g(x_0)} &= |w(t) - w(0) - t\dot{w}(0)|_{g(x_0)} \\ &\leqslant \sum_{k=1}^{d} |\gamma_v^k(t) - \gamma_v^k(0) - t\dot{\gamma}_v^k(0)|.\end{aligned}$$

Now, by Taylor's theorem we have for all $k = 1, \ldots, d$ that

$$|\gamma_v^k(t) - \gamma_v^k(0) - t\dot{\gamma}_v^k(0)| \leqslant \frac{1}{2}t^2|\ddot{\gamma}^k(\xi_{t,k})|$$

for some $\xi_{t,k} \in (0, t)$.
Because γ_v is a geodesic, its coordinates satisfy the geodesic equations

$$\ddot{\gamma}_v^k(t) + \dot{\gamma}_v^i(t)\gamma^j(t)\Gamma_{ij}^k(\gamma_v(t)) = 0.$$

In particular, we find that

$$|\ddot{\gamma}_v^k(t)| \leqslant |\dot{\gamma}_v^i(t)\gamma^j(t)\Gamma_{ij}^k(\gamma_v(t))|.$$

Now observe that $(x, v) \mapsto v^i v^j \Gamma_{ij}^k(\gamma_v(t))$ is continuous, and therefore bounded on the compact set

$$\{(x, v) \in TM|\ x \in \tilde{K}, |v|_{g(x)} = 1\}.$$

In particular, this implies that there exists a constant $C > 0$ such that

$$|v^i v^j \Gamma_{ij}^k(\gamma_v(t))| \leqslant C|v|_g^2$$

for every $k = 1, \ldots, d$. Using this, we obtain that

$$|\ddot{\gamma}_v^k(t)| \leqslant C|\dot{\gamma}_v(t)|_g^2 = C|v|_g^2,$$

where we used that γ_v is a geodesic with $\dot{\gamma}_v(0) = v$.
Combining everything, we find that

$$|w(t) - w(0) - t\mathrm{d}(\mathrm{Exp}_{x_0})^{-1}_{w(0)}(v)|_{g(x_0)} \leqslant \sum_{k=1}^{d} |\gamma_v^k(t) - \gamma_v^k(0) - t\dot{\gamma}_v^k(0)| \leqslant dCt^2|v|_{g(x_0)}$$

as desired. □

3.4.2. Differential of the Riemannian exponential map and parallel transport

Next, we wish to understand the relation between the differential of the Riemannian exponential map and parallel transport. Before we can make the appropriate comparison, we first need a version of Taylor's theorem suitable for vector fields along a curve on a manifold.

Proposition 3.4.5 (Taylor's theorem)**.** *Let γ be a curve in M and v a vector field along γ. Define $D_t v(t) := \nabla_{\dot\gamma(t)} v(t)$ and D_t^k as the k-th covariant derivative in this way. Fix $n \in \mathbb{N}$. For every $t > 0$ there exists $\xi_t \in (0,t)$ such that*

$$v(t) = \sum_{k=0}^{n} \frac{t^k}{k!} \tau_{\gamma(0)\gamma(t)} D_t^k v(0) + \frac{t^{k+1}}{(k+1)!} \tau_{\gamma(\xi_t)\gamma(t)} D_t^{k+1} v(\xi_t).$$

Proof. Consider the map $f(t) = \tau^{-1}_{\gamma(0)\gamma(t)} v(t)$, mapping into $T_{\gamma(0)}M$. Because f is smooth, by Taylor's theorem, given $n \in \mathbb{N}$ and $t > 0$, there exists $\xi_t \in (0,t)$ such that

$$f(t) = \sum_{k=0}^{n} \frac{t^k}{k!} f^{(k)}(0) + \frac{t^{k+1}}{(k+1)!} f^{(k+1)}(\xi_t).$$

Let us compute the derivatives of f. Note that

$$\begin{aligned} f'(t) &= \lim_{h \to 0} \frac{f(t+h) - f(t)}{h} \\ &= \lim_{h \to 0} \frac{\tau^{-1}_{\gamma(0)\gamma(t+h)} v(t+h) - \tau^{-1}_{\gamma(0)\gamma(t)} v(t)}{h} \\ &= \tau^{-1}_{\gamma(0)\gamma(t)} \lim_{h \to 0} \frac{\tau^{-1}_{\gamma(t)\gamma(t+h)} v(t+h) - v(t)}{h} \\ &= \tau^{-1}_{\gamma(0)\gamma(t)} D_t v(t). \end{aligned}$$

Using induction, one can show that

$$f^{(k)}(t) = \tau^{-1}_{\gamma(0)\gamma(t)} D_t^k v(t)$$

for all $k \in \mathbb{N}$. But then we find that

$$\tau^{-1}_{\gamma(0)\gamma(t)} v(t) = \sum_{k=0}^{n} \frac{t^k}{k!} D_t^k v(0) + \frac{t^{k+1}}{(k+1)!} \tau^{-1}_{\gamma(0)\gamma(\xi_t)} D_t^{k+1} v(\xi_t).$$

Applying $\tau_{\gamma(0)\gamma(t)}$ to both sides and observing that $t > \xi_t$ gives the desired result. ☐

We are now able to compare the differential of the Riemannian exponential map and parallel transport. The Taylor series of the differential of the exponential map may be found in e.g [98, Appendix A]. The error term for finite Taylor polynomials seems to belong to folklore, but we insert a proof here for the reader's convenience.

Proposition 3.4.6. *Let* $x_0 \in M$ *and take* $w, u \in T_{x_0}M$. *Consider the geodesic* $\gamma_w : [0,1] \to M$ *given by* $\gamma_w(t) = \operatorname{Exp}_{x_0}(tw)$. *For every* $t \in [0,1]$ *there exists* $\xi_t \in (0,t)$ *such that*

$$\begin{aligned}&\mathrm{d}(\operatorname{Exp}_{x_0})_{tw}(u)\\&\qquad = \tau_{\gamma_w(0)\gamma_w(t)}u + \frac{1}{2}t\tau_{\gamma_w(\xi_t)\gamma_w(t)}R_{\gamma_w(\xi_t)}(\mathrm{d}(\operatorname{Exp}_{x_0})_{\xi_t w}(\xi_t u), \dot\gamma_w(\xi_t))\dot\gamma_w(\xi_t).\end{aligned}$$

Proof. Consider the vector field $J(t) = \mathrm{d}(\operatorname{Exp}_{x_0})_{tw}(tu)$ along $\gamma_w(t)$. It follows from Proposition 3.1.4 that $J(t)$ is a Jacobi field along $\gamma(t)$ with $J(0) = 0$ and $\dot J(0) = u$. By the Jacobi equation (3.1.2), the second derivative is given by

$$D_t^2 J(t) = -R_{\gamma_w(t)}(J(t), \dot\gamma_w(t))\dot\gamma_w(t).$$

Therefore, by Proposition 3.4.5 we find there exists some $\xi_t \in (0,t)$ such that

$$J(t) = t\tau_{\gamma_w(0)\gamma_w(t)}u - \frac{1}{2}t^2\tau_{\gamma_w(\xi_t)\gamma_w(t)}R_{\gamma_w(\xi_t)}(\mathrm{d}(\operatorname{Exp}_{x_0})_{\xi_t w}(\xi_t u), \dot\gamma_w(\xi_t))\dot\gamma_w(\xi_t).$$

The result now follows after dividing by t. □

This proposition allows us to obtain the following estimate.

Corollary 3.4.7. *Fix* $x_0 \in M$ *and let* $w \in B(0, \iota(x_0)) \subset T_{x_0}M$. *Define the geodesic* $\gamma_w : [0,1] \to M$ *by* $\gamma_w(t) = \operatorname{Exp}_{x_0}(tw)$. *There exists a constant* $C > 0$ *only depending on some compact set containing* γ_w *such that*

$$|\mathrm{d}(\operatorname{Exp}_{x_0})_w(u) - \tau_{\gamma_w(0)\gamma_w(1)}u|_{g(\gamma_w(1))} \leqslant C|u|_{g(x_0)}|w|^2_{g(x_0)}$$

for all $u \in T_{x_0}M$.

Proof. By Proposition 3.4.6 there exists $\xi \in (0,1)$ such that

$$\begin{aligned}&\mathrm{d}(\operatorname{Exp}_{x_0})_w(u) - \tau_{\gamma_w(0)\gamma_w(1)}u\\&\qquad = -\frac{1}{2}\tau_{\gamma_w(\xi),\gamma_w(1)}R_{\gamma_w(\xi)}(\mathrm{d}(\operatorname{Exp}_{x_0})_{\xi w}(\xi u), \dot\gamma_w(\xi))\dot\gamma_w(\xi).\end{aligned}$$

Now taking norms on both sides, we first observe that the norm of the Riemann curvature endomorphism is bounded on compact sets, because it is continuous (in coordinates the norm can be expressed as a continuous functions of the coefficients). Furthermore, by Lemma 3.4.1 we have that $w \mapsto |\mathrm{d}(\operatorname{Exp}_{x_0})_w|$ is bounded on compact sets. We thus obtain constants $C_1, C_2 > 0$, only depending on some compact set containing the curve γ_w such that

$$\begin{aligned}&|\mathrm{d}(\operatorname{Exp}_{x_0})_w(u) - \tau_{\gamma_w(0)\gamma_w(1)}u|_{g(\gamma_w(1))}\\&\qquad \leqslant \frac{1}{2}|R_{\gamma_w(\xi)}(\mathrm{d}(\operatorname{Exp}_{x_0})_{\xi w}(\xi u), \dot\gamma_w(\xi))\dot\gamma_w(\xi)|_{g(\gamma_w(\xi))}\\&\qquad \leqslant C_1|\mathrm{d}(\operatorname{Exp}_{x_0})_{\xi w}(\xi u)|_{g(\gamma_w(\xi))}|\dot\gamma_w(\xi)|^2_{g(\gamma_w(\xi))}\\&\qquad \leqslant C_1C_2|u|_{g(x_0)}|w|^2_{g(x_0)}.\end{aligned}$$

Here, in the last line we used that $\xi < 1$ and the fact that γ_w is a geodesic. □

The result in the latter corollary can also be used to compare the inverse of the differential of the exponential map to the inverse of parallel transport, which itself is parallel transport, but in the opposite direction.

Corollary 3.4.8. *Let $x_0 \in M$ and fix $w \in B(0, \iota(x_0)) \subset T_{x_0}M$. Define the geodesic $\gamma_w : [0,1] \to M$ by $\gamma_w(t) = \mathrm{Exp}_{x_0}(tw)$. Then there exists a constant $C > 0$ only depending on some compact set containing γ_w, such that*

$$|\mathrm{d}(\mathrm{Exp}_{x_0})_w^{-1}(u) - \tau_{\gamma_w(0)\gamma_w(1)}^{-1} u|_{g(\gamma_w(1))} \leqslant C|u|_{g(\gamma_w(1))}|w|^2_{g(x_0)}$$

for all $u \in T_{\gamma_w(1)}M$.

Proof. Fix $u \in T_{\gamma_w(1)}M$ and consider $\mathrm{d}(\mathrm{Exp}_{x_0})_w^{-1}u \in T_{x_0}M$. By Corollary 3.4.7, there exists a constant $C > 0$ only depending on a compact set containing γ_w such that

$$|u - \tau_{\gamma_w(0)\gamma_w(1)}\mathrm{d}(\mathrm{Exp}_{x_0})_w^{-1}u|_{g(\gamma_w(1))} \leqslant C|\mathrm{d}(\mathrm{Exp}_{x_0})_w^{-1}u|_{g(x_0)}|w|^2_{g(x_0)}.$$

Because parallel transport is an isometry, the left hand side is equal to

$$|\tau_{\gamma_w(1)\gamma_w(0)}u - \mathrm{d}(\mathrm{Exp}_{x_0})_w^{-1}u|_{g(\gamma_w(1))}.$$

For the right hand side, we observe that by Lemma 3.4.2 there exists a constant $\tilde{C} > 0$, only depending on some compact set containing γ_w such that

$$|\mathrm{d}(\mathrm{Exp}_{x_0})_w^{-1}u|_{g(x_0)} \leqslant \tilde{C}|u|_{g(\gamma_w(1))}.$$

Putting everything together, we find

$$|\tau_{\gamma_w(1)\gamma_w(0)}u - \mathrm{d}(\mathrm{Exp}_{x_0})_w^{-1}u|_{g(\gamma_w(1))} \leqslant C\tilde{C}|u|_{g(\gamma_w(1))}|w|^2_{g(x_0)}$$

as desired. □

3.4.3. Spreading of geodesics

We conclude this section with a result on how far geodesics, possibly starting in different points, can spread in a given amount of time. To shed some light on the upcoming result, we first consider the Euclidean case. For this, let $\gamma(t) = \gamma(0)+t\dot{\gamma}(0)$ and $\phi(t) = \phi(0) + t\dot{\phi}(t)$ be two straight lines. Then

$$|\gamma(t) - \phi(t)|^2 = |\gamma(0) - \phi(0)|^2 + 2t\langle\dot{\gamma}(0) - \dot{\phi}(0), \gamma(0) - \phi(0)\rangle + t^2|\dot{\gamma}(t) - \dot{\phi}(t)|^2.$$

It turns out that in a Riemannian manifold, this formula is analogous up to first order. The curvature terms show up in the second order term. Before we prove this, we first need a lemma.

Lemma 3.4.9. *Let $K \subset M$ be compact and fix $L > 0$. Let $0 < r < \iota(K)$. Let $\phi : [0,T] \to M$ and $\gamma : [0,T] \to M$ be two geodesics contained in K. Assume that $d(\phi(0), \gamma(0)) \leqslant \frac{r}{2}$ and $|\dot{\phi}(0)|, |\dot{\gamma}(0)| \leqslant L$. Then there exists a $t_0 > 0$, only depending on K, L and r, such that for all $0 \leqslant t \leqslant t_0$ we have*

$$d(\phi(t), \gamma(t)) < r.$$

Proof. Because $d : M \times M \to \mathbb{R}$ is continuous, and $K \times K$ is compact, $d(\cdot,\cdot)$ is uniformly continuous on $K \times K$. We can thus pick $\varepsilon > 0$ such that $|d(x,y) - d(x',y')| < \frac{r}{2}$, whenever $d(x,x') < \varepsilon$ and $d(y,y') < \varepsilon$.

Now observe that $d(\phi(t),\phi(0)) \leqslant t|\dot{\phi}(0)| \leqslant tL$ and likewise $d(\gamma(t),\gamma(0)) \leqslant tL$. Hence, if we take $t_0 < \varepsilon L^{-1}$, then for all $0 \leqslant t \leqslant t_0$ we have $d(\phi(t),\phi(0)) < \varepsilon$ and $d(\gamma(t),\gamma(0)) < \varepsilon$. By the choice of ε, it follows that

$$|d(\phi(0),\gamma(0)) - d(\phi(t),\gamma(t))| < \frac{r}{2}.$$

Since $d(\phi(0),\gamma(0)) \leqslant \frac{1}{2}r$, the above then implies that $d(\phi(t),\gamma(t)) < r$ as desired. □

Proposition 3.4.10. *Let $K \subset M$ be compact and fix $L > 0$. Let $0 < r < \iota(K)$ and fix $t_0 > 0$ as in Lemma 3.4.9. Let $\phi : [0,t_0] \to M$ and $\gamma : [0,t_0] \to M$ be two geodesics in K such that $d(\gamma(0),\phi(0)) \leqslant \frac{r}{2}$ and $|\dot{\phi}(0)|, |\dot{\gamma}(0)| \leqslant L$. Finally, let $\tilde{K}$ be a compact set containing all geodesics of minimal length between points in K. Then for all $0 \leqslant t \leqslant t_0$ we have*

$$\begin{aligned} &d(\gamma(t),\phi(t))^2 \\ &\quad \leqslant d(\gamma(0),\phi(0))^2 + 2t\langle \tau^{-1}_{\phi(0)\gamma(0)}\dot{\gamma}(0) - \dot{\phi}(0), \mathrm{Exp}^{-1}_{\phi(0)}\gamma(0)\rangle + t^2 C(|\dot{\gamma}(0)| + |\dot{\phi}(0)|), \end{aligned}$$

where the constant $C > 0$ only depends on $\tilde{K}, L$ and r.

Proof. Define $f(t) = d(\gamma(t),\phi(t))^2$. By the choice of t_0, Lemma 3.4.9 gives us that

$$d(\phi(t),\gamma(t)) < r < \iota(K)$$

for every $0 \leqslant t \leqslant t_0$. This implies that $\phi(t)$ and $\gamma(t)$ may be joined by a unique geodesic of minimal length. Moreover, by restricting Exp, we have $f(t) = |\mathrm{Exp}^{-1}_{\phi(t)}\gamma(t)|^2$. Using this, we can compute

$$\begin{aligned} f'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}|\mathrm{Exp}^{-1}_{\phi(t)}\gamma(t)|^2 \\ &= 2\langle \nabla_{\dot{\phi}(t)}\mathrm{Exp}^{-1}_{\phi(t)}\gamma(t), \mathrm{Exp}^{-1}_{x_0}\gamma(t)\rangle. \end{aligned}$$

Now define the variation of curves $\Gamma : [0,t_0] \times [0,1] \to M$ by

$$\Gamma(t,s) = \mathrm{Exp}_{\phi(t)}(s\mathrm{Exp}^{-1}_{\phi(t)}\gamma(t)).$$

Then for each t, the curve $s \mapsto \Gamma(t,s)$ is the geodesic of minimal length between $\phi(t)$ and $\gamma(t)$. Hence, $\Gamma([0,t_0] \times [0,1]) \subset \tilde{K}$. Furthermore, because Γ is a variation of geodesics, the vector field

$$J_t(s) = \partial_t \Gamma(t,s)$$

is a Jacobi field along the curve $\Gamma_t(s) := \Gamma(t,s)$ for all $0 \leqslant t \leqslant t_0$.

Now note that by the Symmetry Lemma (Lemma 3.1.3), we have

$$\nabla_{\dot{\phi}(t)}\mathrm{Exp}^{-1}_{\phi(t)}\gamma(t) = D_t\partial_s\Gamma(t,0) = D_s\partial_t\Gamma(t,0) = \dot{J}_t(0).$$

From this, we obtain

$$f'(t) = 2\langle \dot{J}_t(0), \mathrm{Exp}_{x_0}^{-1}\gamma(t)\rangle = 2\langle \dot{J}_t(0), \partial_s\Gamma(t,0)\rangle.$$

By Proposition 3.1.5 we find

$$\begin{aligned} f'(t) &= 2\langle \dot{J}_t(0), \partial_s\Gamma(t,0)\rangle \\ &= 2\langle J_t(1), \partial_s\Gamma(t,1)\rangle - 2\langle J_t(0), \partial_s\Gamma(t,0)\rangle \\ &= 2\langle \dot{\gamma}(t), -\mathrm{Exp}_{\gamma(t)}\phi(t)\rangle - 2\langle \dot{\phi}(t), \mathrm{Exp}_{\phi(t)}^{-1}\gamma(t)\rangle \\ &= 2\langle \tau_{\phi(t)\gamma(t)}^{-1}\dot{\gamma}(t) - \dot{\phi}(t), \mathrm{Exp}_{\phi(t)}^{-1}\gamma(t)\rangle. \end{aligned}$$

In particular, we have

$$f'(0) = 2\langle \tau_{\phi(0)\gamma(0)}^{-1}\dot{\gamma}(0) - \dot{\phi}(0), \mathrm{Exp}_{\phi(0)}^{-1}\gamma(0)\rangle.$$

By Taylor's theorem, we find that

$$\begin{aligned} &d(\gamma(t),\phi(t))^2 \\ &\quad \leqslant d(\gamma(0),\phi(0))^2 + 2t\langle \tau_{\phi(0)\gamma(0)}^{-1}\dot{\gamma}(0) - \dot{\phi}(0), \mathrm{Exp}_{\phi(0)}^{-1}\gamma(0)\rangle + \frac{1}{2}t^2 \sup_{\xi\in[0,t]} |f''(\xi)|. \end{aligned}$$

We now turn to estimating the residual term. For this, we compute $f''(t)$ as follows:

$$\begin{aligned} \frac{1}{2}f''(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\langle \dot{\gamma}(t), -\mathrm{Exp}_{\gamma(t)}\phi(t)\rangle - \frac{\mathrm{d}}{\mathrm{d}t}\langle \dot{\phi}(t), \mathrm{Exp}_{\phi(t)}^{-1}\gamma(t)\rangle \\ &= -\langle \dot{\gamma}(t), \nabla_{\dot{\gamma}(t)}\mathrm{Exp}_{\gamma(t)}^{-1}\phi(t)\rangle - \langle \dot{\phi}(t), \nabla_{\dot{\phi}(t)}\mathrm{Exp}_{\phi(t)}^{-1}\gamma(t)\rangle \\ &= \langle \dot{\gamma}(t), \partial_t\Gamma(t,1)\rangle - \langle \dot{\phi}(t), \partial_t\Gamma(t,0)\rangle \\ &= \langle \dot{\gamma}(t), \dot{J}_t(1)\rangle - \langle \dot{\phi}(t), \dot{J}_t(0)\rangle. \end{aligned}$$

Here we used that $\nabla_{\dot{\phi}(t)}\dot{\phi}(t) = \nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = 0$, since ϕ and γ are geodesics. It follows that

$$\frac{1}{2}|f''(t)| \leqslant |\dot{\gamma}(t)||\dot{J}_t(1)| + |\dot{\phi}(t)||\dot{J}_t(0)| = |\dot{\gamma}(0)||\dot{J}_t(1)| + |\dot{\phi}(0)||\dot{J}_t(0)|,$$

where we again used that γ and ϕ are geodesics. It follows that we are done once we can bound $|\dot{J}_t(0)|$ and $|\dot{J}_t(1)|$. For this, we first obtain a more specific expression for the Jacobi field J_t. To this end, we define for every $0 \leqslant t \leqslant t_0$ the vector fields

$$J_t^1(s) = \mathrm{d}(\mathrm{Exp}_{\phi(t)})_{s\partial_s\Gamma(t,0)}(s\dot{J}_t^1(0))$$

and

$$J_t^2(s) = \mathrm{d}(\mathrm{Exp}_{\gamma(t)})_{-s\partial_s\Gamma(t,1)}(s\dot{J}_t^2(0)),$$

where

$$\dot{J}_t^1(0) = \mathrm{d}(\mathrm{Exp}_{\phi(t)})_{\mathrm{Exp}_{\phi(t)}^{-1}\gamma(t)}^{-1}\dot{\gamma}(t) \in T_{\phi(t)}M$$

and likewise

$$\dot{J}_t^2(0) = \mathrm{d}(\mathrm{Exp}_{\gamma(t)})^{-1}_{\mathrm{Exp}^{-1}_{\gamma(t)}\phi(t)}\dot{\phi}(t) \in T_{\gamma(t)}M.$$

It follows from Proposition 3.1.4 that J_t^1 and J_t^2 are Jacobi fields along Γ_t. Note that $J_t^1(0) = J_t^2(0) = 0$ and $J_t^1(1) = \dot{\gamma}(t)$ and $J_t^2(1) = \dot{\phi}(t)$. Because J_t is the unique Jacobi field along Γ_t with $J_t(0) = \dot{\phi}(t)$ and $J_t(1) = \dot{\gamma}(t)$, it follows that

$$J_t(s) = J_t^1(s) + J_t^2(1-s).$$

Using the above decomposition, we show how to bound $|\dot{J}_t(0)|$. The bound for $|\dot{J}_t(1)|$ may be obtained similarly. By the triangle inequality, we have

$$|\dot{J}_t(0)| \leqslant |\dot{J}_t^1(0)| + |\dot{J}_t^2(1)|.$$

Note that

$$|\dot{J}_t^1(0)| = |\mathrm{d}(\mathrm{Exp}_{\phi(t)})^{-1}_{\mathrm{Exp}^{-1}_{\phi(t)}\gamma(t)}\dot{\gamma}(t)| \leqslant |\mathrm{d}(\mathrm{Exp}_{\phi(t)})^{-1}_{\mathrm{Exp}^{-1}_{\phi(t)}\gamma(t)}||\dot{\gamma}(t)|.$$

Therefore, by Lemma 3.4.2 there exists a constant $C > 0$ only depending on K and r (since $|\mathrm{Exp}^{-1}_{\phi(t)}\gamma(t)| = d(\phi(t), \gamma(t)) \leqslant r$) such that

$$|\dot{J}_t^1(0)| \leqslant C|\dot{\gamma}(t)| = C|\dot{\gamma}(0)|.$$

For the other term, it follows from Proposition 3.1.6 that

$$\begin{aligned}
|\dot{J}_t^2(1)| &\leqslant |\dot{J}_t^2(0)| + \sup_{s\in[0,1]} |R_{\Gamma(t,s)}(J_t^2(s), \partial_s\Gamma(t,s))\partial_s\Gamma(t,s)| \\
&\leqslant C|\dot{\phi}(0)| + |\partial_s\Gamma(t,0)|^2 \sup_{s\in[0,1]} |R_{\psi_t(s)}||J_t^2(s)| \\
&\leqslant C|\dot{\phi}(0)| + \tilde{C}d(\gamma(t), \phi(t))^2 \sup_{s\in[0,1]} |J_t^2(s)| \\
&\leqslant C|\dot{\phi}(0)| + \tilde{C}r^2 \sup_{s\in[0,1]} |J_t^2(s)|.
\end{aligned}$$

Here we used in the second line again Lemma 3.4.2 as above, together with the fact that the curves $\Gamma_t(s)$ are geodesics. Furthemore, we used that the curvature is continuous, and hence bounded on compact sets, so that $\tilde{C}$ only depends on $\tilde{K}$, since the variation Γ is contained in $\tilde{K}$. In the last line, we used that $d(\gamma(t), \phi(t)) \leqslant r$ for all $0 \leqslant t \leqslant t_0$ by choice of t_0.
Finally, we have for any $s \in [0,1]$

$$\begin{aligned}
|J_t^2(s)| &= |\mathrm{d}(\mathrm{Exp}_{\gamma(t)})_{-s\partial_s\Gamma(t,1)}(s\dot{J}_t^2(0))| \\
&\leqslant s|\mathrm{d}(\mathrm{Exp}_{\gamma(t)})_{-s\partial_s\Gamma(t,1)}||\dot{J}_t^2(0))| \\
&\leqslant C'|\dot{\phi}(0)|,
\end{aligned}$$

where in the last line we used Lemma 3.4.1. Collecting everything, there exists a constant $C > 0$, only depending on $\tilde{K}$ and r, such that

$$|\dot{J}_t^2(1)| \leqslant C|\dot{\phi}(0)|.$$

Putting everything together, we find that

$$|\dot{J}_t(0)| \leqslant |\dot{J}_t^1(0)| + |\dot{J}_t^2(1)| \leqslant C(|\dot{\gamma}(0)| + |\dot{\phi}(0)|)$$

for some $C > 0$ only depending on $\tilde{K}$ and r. Obtaining a similar bound for $|\dot{J}_t(1)|$ now proves the claim. □

3.4.4. Minimizing trajectories for convex functionals

In this section we generalize the result that if $F : \mathbb{R}^N \to \mathbb{R}$ is convex, the integral

$$I(\gamma) = \int_0^1 F(\dot{\gamma}(t))\,\mathrm{d}t$$

considered for curves with $\gamma(0) = x$ and $\gamma(1) = y$ is minimized by the straight line connecting x and y.

In the context of Riemannian geometry, this concept arises naturally when minimizing the Riemannian distance between points $x, y \in M$. Indeed, the distance is given by (see Section 2.2)

$$d(x,y)^2 = \inf\left\{ \int_0^1 |\dot{\gamma}(t)|^2 \middle| \gamma : [0,1] \to M, \gamma(0) = x, \gamma(1) = y, \gamma \text{ piecewise smooth} \right\}.$$

It can be shown that the optimal trajectories for the distance between x and y are geodesics, see e.g. [69, Chapter 6]. Note that in this case, we consider the integral of the function $F : TM \to \mathbb{R}$ given by $F(x,v) = |v|^2_{g(x)}$. For every $x \in M$, the map $v \mapsto F(x,v)$ is strictly convex. Furthermore, F is invariant under parallel transport in the sense that for all $x, y \in M$ and $v \in T_xM$ we have $F(x,v) = F(y, \tau_{xy}v)$. The next result states that these two conditions on F are in general sufficient to conclude that geodesics are minimizing trajectories for the integral of F.

Proposition 3.4.11. *Let $F : TM \to \mathbb{R}$ be a smooth function satisfying the following properties:*

1. *For every $x \in M$, the map $F(x,\cdot) : T_xM \to \mathbb{R}$ is strictly convex.*
2. *For every $x, y \in M$ and smooth curve γ connecting x and y, we have*

   $$F(x,v) = F(y, \tau_{xy;\gamma}v)$$

   for all $v \in T_xM$.

Define the functional I by

$$I(\gamma) := \int_0^1 F(\gamma(t), \dot{\gamma}(t))\,\mathrm{d}t.$$

Then for any $x, y \in M$ we have

$$\begin{aligned}
&\inf\{I(\gamma) | \gamma : [0,1] \to M, \gamma(0) = x, \gamma(1) = y, \gamma \text{ piecewise smooth}\} \\
&\quad = \inf\{I(\gamma) | \gamma : [0,1] \to M, \gamma(0) = x, \gamma(1) = y, \gamma \text{ geodesic}\} \\
&\quad = \inf\{F(v) | v \in \operatorname{Exp}_x^{-1} y\}.
\end{aligned}$$

Before we get to the proof, we first need a lemma.

Lemma 3.4.12. *Let F be as in Proposition 3.4.11. Fix $x, y \in M$ and let $\gamma : [0,1] \to M$ be any curve with $\gamma(0) = x$ and $\gamma(1) = y$. Denote by τ_{xy} parallel transport from $T_x M$ to $T_y M$ along γ. For every $x \in M$, define $F_x : T_x M \to \mathbb{R}$ by $F_x(v) := F(x, v)$. Then for any $v \in T_x M$ we have*

$$\tau_{xy} \nabla F_x(v) = \nabla F_y(\tau_{xy} v).$$

Proof. By the consistency property 2 of F we have $F_x = F_y \circ \tau_{xy}$. Applying the chain rule, we find

$$\mathrm{d}F_x(v) = \mathrm{d}(F_y \circ \tau_{xy})(v) = \mathrm{d}F_y(\tau_{xy} v) \circ \mathrm{d}\tau_{xy}(v) = \mathrm{d}F_y(\tau_{xy} v) \circ \tau_{xy}.$$

Here we used in the final step that τ_{xy} is linear. But then we find for every $w \in T_x M$ that

$$\begin{aligned}
\langle \nabla F_x(v), w \rangle = \mathrm{d}F_x(v)(w) &= \mathrm{d}F_y(\tau_{xy} v)(\tau_{xy} w) \\
&= \langle \nabla F_y(\tau_{xy} v), \tau_{xy} w \rangle = \langle \tau_{yx} \nabla F_y(\tau_{xy} v), w \rangle.
\end{aligned}$$

This implies that $\nabla F_x(v) = \tau_{yx} \nabla F_y(\tau_{xy} v)$, from which the desired equality follows by applying τ_{xy}. □

We now turn to the proof of Proposition 3.4.11. The proof is similar to the variational approach in proving that geodesics are optimal trajectories for the length functional, see e.g. [69, Chapter 6]. For the notation from the calculus of variations, we refer to Section 3.1.1.

Proof of Proposition 3.4.11. Fix $x, y \in M$ and let $\gamma : [0,1] \to M$ be a curve with $\gamma(0) = x$ and $\gamma(1) = y$. Let $\Gamma : (-\varepsilon, \varepsilon) \to M$ be a variation of γ with $\Gamma(s, 0) = x$ and $\Gamma(s, 1) = y$ for all $s \in (-\varepsilon, \varepsilon)$. If γ minimizes I, then

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} I(\Gamma(s, \cdot)) = 0.$$

On the other hand, we have

$$\frac{\mathrm{d}}{\mathrm{d}s} I(\Gamma(s, \cdot)) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}s} F(\Gamma(s,t), \partial_t \Gamma(s,t)) \, \mathrm{d}t.$$

Hence, we need to compute

$$\frac{\mathrm{d}}{\mathrm{d}s} F(\Gamma(s,t), \partial_t \Gamma(s,t)).$$

Note that for every $s \in (-\varepsilon, \varepsilon)$ we have

$$F(\Gamma(s,t), \partial_t \Gamma(s,t)) = F(\gamma(t), \tau_{\Gamma(s,t)\Gamma(0,t)} \partial_t \Gamma(s,t)).$$

Using this, we compute

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s} F(\Gamma(s,t), \partial_t \Gamma(s,t)) &= \frac{\mathrm{d}}{\mathrm{d}s} F(\gamma(t), \tau_{\Gamma(s,t)\Gamma(0,t)} \partial_t \Gamma(s,t)) \\
&= \left\langle \nabla F_{\gamma(t)} \left(\tau_{\Gamma(s,t)\Gamma(0,t)} \partial_t \Gamma(s,t) \right), \frac{\mathrm{d}}{\mathrm{d}s} \tau_{\Gamma(s,t)\Gamma(0,t)} \partial_t \Gamma(s,t) \right\rangle.
\end{aligned}$$

Now

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s} \tau_{\Gamma(s,t)\Gamma(0,t)} \partial_t \Gamma(s,t) &= \lim_{h \to 0} \frac{\tau_{\Gamma(s+h,t)\Gamma(0,t)} \partial_t \Gamma(s+h,t) - \tau_{\Gamma(s,t)\Gamma(0,t)} \partial_t \Gamma(s,t)}{h} \\
&= \tau_{\Gamma(s,t)\Gamma(0,t)} \lim_{h \to 0} \frac{\tau_{\Gamma(s+h,t)\Gamma(s,t)} \partial_t \Gamma(s+h,t) - \partial_t \Gamma(s,t)}{h} \\
&= \tau_{\Gamma(s,t)\Gamma(0,t)} D_s \partial_t \Gamma(s,t).
\end{aligned}$$

Combining the above equations, we find

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s} F(\Gamma(s,t), \partial_t \Gamma(s,t)) &= \left\langle \nabla F_{\gamma(t)} \left(\tau_{\Gamma(s,t)\Gamma(0,t)} \partial_t \Gamma(s,t) \right), \tau_{\Gamma(s,t)\Gamma(0,t)} D_s \partial_t \Gamma(s,t) \right\rangle \\
&= \left\langle \nabla F_{\Gamma(s,t)} \left(\partial_t \Gamma(s,t) \right), D_s \partial_t \Gamma(s,t) \right\rangle \\
&= \left\langle \nabla F_{\Gamma(s,t)} \left(\partial_t \Gamma(s,t) \right), D_t \partial_s \Gamma(s,t) \right\rangle.
\end{aligned}$$

Here, the one but last line follows from Lemma 3.4.12. The last line follows from the Symmetry lemma (Lemma 3.1.3).
Now define the variational vector field $V(t)$ of Γ by $V(t) = \partial_s \Gamma(0,t)$. Then for $s = 0$ we obtain

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} F(\Gamma(s,t), \partial_t \Gamma(s,t)) = \left\langle \nabla F_{\gamma(t)}(\dot{\gamma}(t)), \nabla_{\dot{\gamma}(t)} V(t) \right\rangle.$$

Collecting everything, we obtain

$$\begin{aligned}
0 &= \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} I(\Gamma(s,\cdot)) \\
&= \int_0^1 \left\langle \nabla F_{\gamma(t)}(\dot{\gamma}(t)), \nabla_{\dot{\gamma}(t)} V(t) \right\rangle \mathrm{d}t \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \langle \nabla F_{\gamma(t)}(\dot{\gamma}(t)), V(t) \rangle - \left\langle \nabla_{\dot{\gamma}(t)} \nabla F_{\gamma(t)}(\dot{\gamma}(t)), V(t) \right\rangle \mathrm{d}t \\
&= \left\langle \nabla F_{\gamma(1)}(\dot{\gamma}(1)), V(1) \right\rangle - \left\langle \nabla F_{\gamma(0)}(\dot{\gamma}(0)), V(0) \right\rangle - \int_0^1 \left\langle \nabla_{\dot{\gamma}(t)} \nabla F_{\gamma(t)}(\dot{\gamma}(t)), V(t) \right\rangle \mathrm{d}t \\
&= -\int_0^1 \left\langle \nabla_{\dot{\gamma}(t)} \nabla F_{\gamma(t)}(\dot{\gamma}(t)), V(t) \right\rangle \mathrm{d}t,
\end{aligned}$$

because $V(0) = V(1) = 0$, since $\Gamma(s,0) = x$ and $\Gamma(s,1) = y$ for all $s \in (-\varepsilon, \varepsilon)$.

Now pick $\varphi : [0,1] \to \mathbb{R}$ smooth with $\varphi(0) = \varphi(1) = 0$ and $\varphi(t) > 0, t \in (0,1)$. Consider the vector field $V(t) = \varphi(t)\nabla_{\dot\gamma(t)}\nabla F_{\gamma(t)}(\dot\gamma(t))$. Constructing Γ with this variational vector field, we obtain

$$0 = -\int_0^1 \varphi(t)|\nabla_{\dot\gamma(t)}\nabla F_{\gamma(t)}(\dot\gamma(t))|^2 \,\mathrm{d}t$$

from which it follows that

$$\nabla_{\dot\gamma(t)}\nabla F_{\gamma(t)}(\dot\gamma(t)) = 0$$

for all $t \in [0,1]$. We thus have that $\nabla F_{\gamma(t)}(\dot\gamma(t))$ is parallel along $\gamma(t)$, i.e.

$$\tau_{\gamma(t)\gamma(0)}\nabla F_{\gamma(t)}(\dot\gamma(t)) = \nabla F_{\gamma(0)}(\dot\gamma(0)).$$

On the other hand, by Lemma 3.4.12 we have

$$\tau_{\gamma(t)\gamma(0)}\nabla F_{\gamma(t)}(\dot\gamma(t)) = \nabla F_{\gamma(0)}(\tau_{\gamma(t)\gamma(0)}\dot\gamma(t)).$$

Because $F_{\gamma(0)}$ is strictly convex, its derivative is injective, so that

$$\nabla F_{\gamma(0)}(\dot\gamma(0)) = \nabla F_{\gamma(0)}(\tau_{\gamma(t)\gamma(0)}\dot\gamma(t))$$

implies that

$$\dot\gamma(0) = \tau_{\gamma(t)\gamma(0)}\dot\gamma(t).$$

As this holds for all $t \in [0,1]$, we conclude that γ is a geodesic. □

3.5. Proof of Cramér's theorem for geodesic random walks

In this section we provide a proof of Cramér's theorem for geodesic random walks with independent and identically distributed increments, which are bounded and have expectation 0. The proof relies on an analysis of the geometric properties of a geodesic random walk. To prove the theorem, we follow the steps as discussed in Section 3.3. We provide the details and show how we use the geometric results from Section 3.4. For completeness, let us recall the statement of the theorem.

Theorem 3.5.1 (Cramér's theorem for Riemannian manifolds). *Let (M,g) be a complete Riemannian manifold. Fix $x_0 \in M$ and let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_xM)$ for all $x \in M$. For every $n \geqslant 1$, let $\{(\frac{1}{n} * \mathcal{S})_j\}_{j\geqslant 0}$ be a $\frac{1}{n}$-rescaled geodesic random walk started at x_0 with independent increments $\{X_j^n\}_{j\geqslant 1}$, compatible with $\{\mu_x\}_{x\in M}$. Let $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$ be the associated empirical average process started at x_0. Assume the increments are bounded and have expectation 0. Assume furthermore that the collection $\{\mu_x\}_{x\in M}$ satisfies the consistency property in Definition 3.2.7. Then $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$ satisfies the large deviation principle in M with good rate function*

$$I_M(x) = \inf\{\Lambda^*_{x_0}(v) | v \in \mathrm{Exp}^{-1}_{x_0} x\}. \tag{3.5.1}$$

In Section 3.5.1 we prove the upper bound of the large deviation principle for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 1}$ in M, while in Section 3.5.2 we prove the lower bound. More specifically, Theorem 3.5.1 follows immediately from Proposition 3.5.8 together with Proposition 3.5.10.

However, before we can prove the upper and lower bound of the large deviation principle for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 1}$, we first need some general results and estimates. From here on, we fix $r > 0$ to be the uniform bound on the increments of the random walk. By the triangle inequality, we find

$$d\left(\left(\frac{1}{n} * \mathcal{S}\right)_k, x_0\right) \leqslant \frac{1}{n}\sum_{l=1}^{k} |X_k^n| \leqslant \frac{k}{n} r \leqslant r$$

for all $0 \leqslant k \leqslant n$. Therefore, for every $n \geqslant 0$ and $1 \leqslant k \leqslant n$ we have

$$\left(\frac{1}{n} * \mathcal{S}\right)_k \in \overline{B(x_0, r)} =: K.$$

By completeness of M, K is compact since it is closed and bounded.
Now consider the process Z_n in $T_{x_0}M$ given by

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X_k^n.$$

Here, the parallel transport $\tau_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}$ is considered along the piecewise geodesic path traced out by the geodesic random walk. From Cramér's theorem for vector spaces it follows that $\{Z_n\}_{n\geqslant 0}$ satisfies the large deviation principle in $T_{x_0}M$, which we will show in the following proposition.

Proposition 3.5.2. *Let the assumptions of Theorem 3.5.1 be satisfied. For every $n \geqslant 0$, define $Z_n = \frac{1}{n}\sum_{k=1}^{n} \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X_k^n \in T_{x_0}M$. Let $\Lambda_{x_0}(\lambda) = \log \mathbb{E}(e^{\langle \lambda, X_1\rangle})$ be the log moment generating function of the increments. Then $\{Z_n\}_{n\geqslant 0}$ satisfies the large deviation principle in $T_{x_0}M$ with good rate function*

$$I(v) = \Lambda^*_{x_0}(v) := \sup_{\lambda \in T_{x_0}M} \{\langle \lambda, v\rangle - \Lambda_{x_0}(\lambda)\}.$$

Proof. Define $Y_k^n = \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X_k^n \in T_{x_0}M$. We compute for any $\lambda \in T_{x_0}M$

$$\begin{aligned}
\mathbb{E}(e^{\langle \lambda, Y_k^n\rangle}) &= \mathbb{E}\left(\mathbb{E}\left(e^{\langle \lambda, \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X_k^n\rangle} \middle| \mathcal{F}_{k-1}\right)\right) \\
&= \mathbb{E}\left(\int_{T_{(\frac{1}{n}*\mathcal{S})_{k-1}}M} e^{\langle \lambda, \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} v\rangle} \mu_{(\frac{1}{n}*\mathcal{S})_{k-1}}(\mathrm{d}v)\right) \\
&= \mathbb{E}\left(\int_{T_{x_0}M} e^{\langle \lambda, v\rangle} \mu_{x_0}(\mathrm{d}v)\right)
\end{aligned}$$

$$= \int_{T_{x_0}M} e^{\langle \lambda, v\rangle} \mu_{x_0}(\mathrm{d}v).$$

Here we used in the second line that $\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}$ is measurable with respect to $\mathcal{F}_{k-1}$, together with the fact that the increments are independent (see Definition 3.2.3). In the third line we applied Proposition 3.2.9, using that the increments are identically distributed. It follows that Y_k^n is distributed according to μ_{x_0}.
As a consequence, the result follows from Cramér's theorem (Theorem 2.1.10) once we show that Y_k^n and Y_l^n are independent whenever $k \neq l$. To this end, assume without loss of generality that $l < k$. Then for measurable sets $A, B \subset T_{x_0}M$ we find in a similar way as above that

$$\begin{aligned}
&\mathbb{P}(Y_l^n \in A, Y_k^n \in B)\\
&\quad = \mathbb{E}(I(Y_l^n \in A)\mathbb{E}(I(Y_k^n \in B)|\mathcal{F}_{k-1}))\\
&\quad = \mathbb{E}\left(I(Y_l^n \in A)\int_{T_{(\frac{1}{n}*\mathcal{S})_{k-1}}M} I\left(\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} v \in B\right)\mu_{(\frac{1}{n}*\mathcal{S})_{k-1}}(\mathrm{d}v)\right)\\
&\quad = \mathbb{E}\left(I(Y_l^n \in A)\int_{T_{x_0}M} I\left(v \in B\right)\mu_{x_0}(\mathrm{d}v)\right)\\
&\quad = \mathbb{E}(I(Y_l^n \in A))\mathbb{E}(I(Y_k^n \in B))\\
&\quad = \mathbb{P}(Y_l^n \in A)\mathbb{P}(Y_k^n \in B),
\end{aligned}$$

where I denotes the indicator function. Above, we used in the one but last line that Y_k^n is distributed according to μ_{x_0}. We conclude that the Y_l^n and Y_k^n are independent. □

Remark 3.5.3. Note that in the proof of Proposition 3.5.2 we did not use along which path we performed the parallel transport $\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}$, only that it was measurable with respect to $\mathcal{F}_{k-1}$. Therefore, the result holds for any choice of parallel transport, as long as it is measurable with respect to $\mathcal{F}_{k-1}$.

Proposition 3.5.2 suggests we should try to map the sequence $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$ from M to $T_{x_0}M$ in such a way that it will be close to the sequence $\{Z_n\}_{n\geqslant 0}$.
To this end, recall that if we assume that $r < \iota(x_0)$, then for all n and all $0 \leqslant k \leqslant n$ we can uniquely define

$$v_k^n \in \mathrm{Exp}_{x_0}^{-1}\left(\left(\frac{1}{n} * \mathcal{S}\right)_k\right) \subset T_{x_0}M$$

with $|v_k^n| < \iota(x_0)$, because $d((\frac{1}{n} * \mathcal{S})_k, x_0) \leqslant r < \iota(x_0)$.
As explained in Step 2 of Section 3.3.2, we have the following estimate. The first term of the upper bound in (3.5.2) follows from replacing v_n^l with a sum of differentials of the Riemannian exponential map, while the second term follows from replacing these differentials with parallel transport.

Proposition 3.5.4. *Let the assumptions of Theorem 3.5.1 be satisfied. Additionally, let r be the uniform bound of the increments and assume that $r < \iota(x_0)$. Then there exists a constant $C > 0$ such that*

$$\left| v_l^n - \frac{1}{n} \sum_{k=1}^{l} \tau_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}^{-1} X_k^n \right| \leqslant C\frac{l}{n^2} + Cr^2\frac{l^3}{n^3} \tag{3.5.2}$$

for all n and all $1 \leqslant l \leqslant n$.

Proof. Recall that for all n and all $0 \leqslant k \leqslant n$ we have that $(\frac{1}{n}*\mathcal{S})_k$ is in the compact set $K = \overline{B(x_0, r)}$. This implies that

$$v_k^n \in \overline{B(0,r)} \subset T_{x_0}M$$

for all n and all $0 \leqslant k \leqslant n$. But then it follows from Proposition 3.4.4 that for every $0 \leqslant k \leqslant n$ there exists a constant $C_k > 0$ only depending on the norms of v_k^n, v_{k+1}^n and X_k^n such that

$$\left| v_{k+1}^n - \left(v_k^n + \frac{1}{n} \mathrm{d}(\mathrm{Exp}_{x_0})_{v_k^n}^{-1} X_{k+1}^n \right) \right| \leqslant C_k \frac{1}{n^2}. \tag{3.5.3}$$

Because each of the norms $|v_k^n|, |v_{k+1}^n|$ and $|X_k^n|$ are bounded by r, we conclude that we can take $C_k = C$ independent of k.

Turning to the proof of the statement, by the triangle inequality we have

$$\begin{aligned} &\left| v_l^n - \frac{1}{n} \sum_{k=1}^{l} \tau_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}^{-1} X_k^n \right| \\ &\leqslant \left| v_l^n - \frac{1}{n} \sum_{k=1}^{l} \mathrm{d}(\mathrm{Exp}_{x_0})_{v_{k-1}^n}^{-1} X_k^n \right| + \frac{1}{n} \sum_{k=1}^{l} \left| \mathrm{d}(\mathrm{Exp}_{x_0})_{v_{k-1}^n}^{-1} X_k^n - \tau_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}^{-1} X_k^n \right|. \end{aligned}$$

We estimate both terms separately.
For the first term, we write v_l^n as the telescoping sum

$$v_l^n = \sum_{k=1}^{l} (v_k^n - v_{k-1}^n).$$

Using this, we obtain

$$\begin{aligned} \left| v_l^n - \frac{1}{n} \sum_{k=1}^{l} \mathrm{d}(\mathrm{Exp}_{x_0})_{v_{k-1}^n}^{-1} X_k^n \right| &\leqslant \sum_{k=1}^{l} |v_k^n - v_{k-1}^n - \mathrm{d}(\mathrm{Exp}_{x_0})_{v_{k-1}^n}^{-1} X_k^n| \\ &\leqslant C\frac{l}{n^2}, \end{aligned}$$

where the last line follows from the estimate in (3.5.3).

For the other term, observe that by Corollary 3.4.8, there exists a constant $C > 0$ only depending on the compact set $\overline{B(0,r)}$ and r, such that

$$|\mathrm{d}(\mathrm{Exp}_{x_0})^{-1}_{v^n_{k-1}} X^n_k - \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X^n_k| \leqslant C|v^n_{k-1}|^2$$

But then we find

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{l} |\mathrm{d}(\mathrm{Exp}_{x_0})^{-1}_{v^n_{k-1}} X^n_k - \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X^n_k| &\leqslant C\frac{1}{n}\sum_{k=1}^{l} |v^n_{k-1}|^2 \\ &\leqslant Cr^2\frac{l^3}{n^3},\end{aligned}$$

where in the last line we used that $|v^n_{k-1}| \leqslant r\frac{k-1}{n} \leqslant r\frac{l}{n}$ for any $1 \leqslant k \leqslant l$. □

One might hope to combine Propositions 3.5.2 and 3.5.4 to prove that $\{v^n_n\}_{n\geqslant 0}$ satisfies in $T_{x_0}M$ the large deviation principle. Unfortunately, the upper bound found in Proposition 3.5.4 gives an unwanted contribution on the exponential scale. Indeed, taking $l = n$, we find that the upper bound in (3.5.2) is $\mathcal{O}(1)$, which results in the fact that we get stuck with a constant as explained in Step 1 of Section 3.3.2. In an attempt to reduce this term in the upper bound, we cut up the random walk in finitely many pieces and analyse the pieces separately.
To this end, recall that

$$d\left(\left(\frac{1}{n} * \mathcal{S}\right)_k, x_0\right) \leqslant \frac{1}{n}\sum_{l=1}^{k} |X^n_k| \leqslant \frac{k}{n}r.$$

Now observe that $\iota(\overline{B(x_0,r)}) > 0$, because $\overline{B(x_0,r)}$ is compact (see (3.1.1) for the definition of the injectivity radius of a set). Therefore, if $k \leqslant \frac{n\iota(\overline{B(x_0,r)})}{2r}$, then

$$d\left(\left(\frac{1}{n} * \mathcal{S}\right)_k, x_0\right) \leqslant \frac{\iota(\overline{B(x_0,r)})}{2} < \iota(\overline{B(x_0,r)}). \tag{3.5.4}$$

Now let $m \in \mathbb{N}$ such that $m \geqslant \frac{2r}{\iota(\overline{B(x_0,r)})}$. For $0 \leqslant l \leqslant m-1$ we define $n_l = l\lfloor m^{-1}n\rfloor$ and $n_m = n$. By (3.5.4), for every $0 \leqslant l \leqslant m-1$ and $1 \leqslant k \leqslant n_{l+1} - n_l$ we can uniquely define

$$\tilde{v}^{n,m,l}_k \in \mathrm{Exp}^{-1}_{(\frac{1}{n}*\mathcal{S})_{n_l}}\left(\left(\frac{1}{n} * \mathcal{S}\right)_{n_l+k}\right) \subset T_{(\frac{1}{n}*\mathcal{S})_{n_l}}M \tag{3.5.5}$$

with $|\tilde{v}^{n,m,l}_k| < \iota((\frac{1}{n} * \mathcal{S})_{n_l})$, because $n_{l+1} - n_l \leqslant nm^{-1} \leqslant \frac{n\iota(\overline{B(x_0,r)})}{2r}$. Finally, we set

$$v^{n,m,l}_k = \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{n_l}} \tilde{v}^{n,m,l}_k \in T_{x_0}M,$$

where parallel transport $\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{n_l}}$ is taken along the piecewise geodesic path through the points $(\frac{1}{n} * \mathcal{S})_{n_1}, \ldots, (\frac{1}{n} * \mathcal{S})_{n_{l-1}}$.

Alongside this division of the random walk into pieces, we define a map $\Psi_m : (T_{x_0}M)^m \to M$ to identify the tuple $(v^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, v^{n,m,m}_{\lfloor m^{-1}n\rfloor})$ with $(\frac{1}{n} * \mathcal{S})_n$, just like we used the Riemannian exponential map to identify v^n_n and $(\frac{1}{n} * \mathcal{S})_n$ before. Essentially, Ψ_m is an m time recursive application of the Riemannian exponential map.

More precisely, let $(v_1, \dots, v_m) \in (T_{x_0}M)^m$ be given and define $x_1 = \mathrm{Exp}_{x_0}(v_1)$. Now, suppose $x_1, \dots, x_i$ are given. Denote by $\tau_{x_0 x_i}$ parallel transport along the constructed piecewise geodesic path via $x_1, \dots, x_{i-1}$. Then we define $\tilde{v}_{i+1} = \tau_{x_0 x_i} v_{i+1}$ and set $x_{i+1} = \mathrm{Exp}_{x_i}(\tilde{v}_{i+1})$. Finally, we define $\Psi_m(v_1, \dots, v_m) = x_m$. In particular, we have for every $x \in M$ and $v \in \mathrm{Exp}^{-1}_{x_0} x$ that $(\frac{1}{m} v, \dots, \frac{1}{m} v) \in \Psi^{-1}_m x$. To see this, observe that the path that Ψ_m constructs is precisely the geodesic $\gamma_v(t) = \mathrm{Exp}_{x_0}(tv)$, because the speed of a geodesic is parallel along the geodesic. Furthermore, the map Ψ_m is continuous as a composition of continuous maps.

Remark 3.5.5. Strictly speaking, if we divide the random walk into m pieces as above, for the last piece we can only guarantee that it has at most $\lfloor m^{-1}n\rfloor + m$ increments, since n need not be divisible by m. Additionally, this implies that $\Psi_m(v^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, v^{n,m,m}_{\lfloor m^{-1}n\rfloor})$ is only equal to $\left(\frac{1}{n} * \mathcal{S}\right)_n$ when n is divisible by m. However, for every $m \in \mathbb{N}$ it holds that

$$d\left(\Psi_m(v^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, v^{n,m,m}_{\lfloor m^{-1}n\rfloor}), \left(\frac{1}{n} * \mathcal{S}\right)_n\right) = \mathcal{O}\left(\frac{1}{n}\right).$$

Since in the proofs to follow we always first let n tend to infinity before m, this has no influence on the results and arguments. Therefore, to avoid unnecessarily complicated notation and reasoning, we proceed with the above.

3.5.1. Upper bound of the large deviation principle for the sequence $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$

In this section we prove the large deviation upper bound for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$. Before we can do this, we first need some preliminary results.

Proposition 3.5.6 (Upper bound for $\mathbb{E}(e^{n\langle \lambda, v^n_n\rangle})$)**.** *Let the assumptions of Theorem 3.5.1 be satisfied. Additionally, let r be the uniform bound of the increments and assume that $r < \iota(x_0)$. Then there exists a constanct $C > 0$ such that for all n and all $1 \leqslant l \leqslant n$*

$$\mathbb{E}(e^{n\langle \lambda, v^n_l\rangle}) \leqslant e^{ln^{-1}|\lambda|C} e^{|\lambda|Cr^2 l^3 n^{-2}} M_{x_0}(\lambda)^l$$

for all $\lambda \in T_{x_0}M$. Here, $M_{x_0}(\lambda) = \int_{T_{x_0}M} e^{\langle \lambda, v\rangle} \mu_{x_0}(\mathrm{d}v)$.

Proof. By Proposition 3.5.4 and the Cauchy-Schwarz inequality, there exists a constant $C > 0$ such that

$$\begin{aligned}\langle \lambda, v^n_l\rangle - \frac{1}{n}\sum_{k=1}^{l} \langle \lambda, \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X^n_k\rangle &\leqslant |\lambda| \left| v^n_l - \frac{1}{n}\sum_{k=1}^{l} \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X^n_k \right| \\ &\leqslant C|\lambda|\frac{l}{n^2} + C|\lambda| r^2 \frac{l^3}{n^3}.\end{aligned}$$

But then we can estimate

$$\mathbb{E}\left(e^{n\langle\lambda,v_l^n\rangle}\right) = \mathbb{E}\left(e^{\sum_{k=1}^l \langle\lambda,\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}X_k^n\rangle} e^{n\langle\lambda,v_l^n\rangle - \sum_{k=1}^l \langle\lambda,\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}X_k^n\rangle}\right)$$
$$\leqslant e^{C|\lambda|ln^{-1}} e^{C|\lambda|r^2l^3n^{-2}}\mathbb{E}\left(e^{\sum_{k=1}^l \langle\lambda,\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}X_k^n\rangle}\right).$$

As shown in the proof of Proposition 3.5.2, for every $1 \leqslant k \leqslant n$ we have that $\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}X_k^n$ is distributed accodring to μ_{x_0} and is independent of $\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{l-1}}X_l^n$ for any $l \neq k$. Consequentely, we find that

$$\mathbb{E}\left(e^{\sum_{k=1}^l \langle\lambda,\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}X_k^n\rangle}\right) = \prod_{k=1}^l \mathbb{E}\left(e^{\langle\lambda,\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}}X_k^n\rangle}\right)$$
$$= M_{x_0}(\lambda)^l,$$

where the last step follows from Proposition 3.2.9. □

Using Proposition 3.5.6, we obtain the following inequality, which is key in deriving the large deviation upper bound for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$.

Proposition 3.5.7. *Let the assumptions of Theorem 3.5.1 be satisfied. Denote by r the uniform bound on the increments of the geodesic random walk. Then for any $m \in \mathbb{N}$ such that $m \geqslant \frac{2r}{\iota(\overline{B(x_0,r)})}$ and any closed $F \subset (T_{x_0}M)^m$ we have*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, v^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in F\right)$$
$$\leqslant -\inf_{(v_1,\ldots,v_m)\in F} \sup_{(\lambda_1,\ldots,\lambda_m)\in(T_{x_0}M)^m} \frac{1}{m}\sum_{i=1}^m \left\{\langle\lambda_i, mv_i\rangle - \Lambda_{x_0}(\lambda_i) - m^{-2}C|\lambda_i|r^2\right\}.$$

Here, C is a constant depending on the curvature of the compact set $\overline{B(0,r)}$ and the bound r.

Proof. We first prove the upper bound for compact sets, so let $\Gamma \subset (T_{x_0}M)^m$ be compact. Following the proof of Cramér's theorem (see e.g. [29, 56]) for the vector space $(T_{x_0}M)^m$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, v^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in \Gamma\right)$$
$$\leqslant -\inf_{(v_1,\ldots,v_m)\in\Gamma} \sup_{(\lambda_1,\ldots,\lambda_m)\in(T_{x_0}M)^m} \left\{\sum_{i=1}^m\langle\lambda_i, v_i\rangle - \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left(e^{n\sum_{i=1}^m\langle\lambda_i, v^{n,m,i}_{\lfloor m^{-1}n\rfloor}\rangle}\right)\right\}.$$

Recall that for $0 \leqslant i \leqslant m-1$ we write $n_i = i\lfloor m^{-1}n\rfloor$ and $n_m = n$. By Proposition 3.5.4 (which we may apply, because m is chosen large enough) there exists a $C > 0$

such that for any $1 \leqslant i \leqslant m$ we have

$$\begin{aligned}\left|\tilde{v}^{n,m,i}_{\lfloor m^{-1}n\rfloor} - \frac{1}{n}\sum_{k=n_{i-1}+1}^{n_i} \tau^{-1}_{(\frac{1}{n}*\mathcal{S})_{n_{i-1}}(\frac{1}{n}*\mathcal{S})_k} X^n_k\right| &\leqslant C\frac{\lfloor m^{-1}n\rfloor}{n^2} + Cr^2\frac{\lfloor m^{-1}n\rfloor^3}{n^3} \\ &\leqslant C\frac{1}{nm} + Cr^2\frac{1}{m^3}.\end{aligned}$$

But then we also have that

$$\left|v^{n,m,i}_{\lfloor m^{-1}n\rfloor} - \frac{1}{n}\tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{n_{i-1}}}\sum_{k=n_{i-1}+1}^{n_i} \tau^{-1}_{(\frac{1}{n}*\mathcal{S})_{n_{i-1}}(\frac{1}{n}*\mathcal{S})_k} X^n_k\right| \leqslant C\frac{1}{nm} + Cr^2\frac{1}{m^3}, \quad (3.5.6)$$

because parallel transport is an isometry.
Now define

$$Y^n_i = \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{n_{i-1}}}\sum_{k=n_{i-1}+1}^{n_i} \tau^{-1}_{(\frac{1}{n}*\mathcal{S})_{n_{i-1}}(\frac{1}{n}*\mathcal{S})_k} X^n_k \in T_{x_0}M.$$

Using (3.5.6), it follows from the Cauchy-Schwarz inequality and the triangle inequality that

$$\left|\sum_{i=1}^m \langle \lambda_i, v^{n,m,i}_{\lfloor m^{-1}n\rfloor}\rangle - \frac{1}{n}\sum_{i=1}^m \langle \lambda_i, Y^n_i\rangle\right| \leqslant C\left(\frac{1}{nm} + r^2\frac{1}{m^3}\right)\sum_{i=1}^m |\lambda_i|.$$

As a consequence, we find that

$$\mathbb{E}\left(e^{n\sum_{i=1}^m \langle \lambda_i, v^{n,m,i}_{\lfloor m^{-1}n\rfloor}\rangle}\right) \leqslant e^{Cm^{-1}\sum_{i=1}^m |\lambda_i|} e^{Cr^2 m^{-3} n \sum_{i=1}^m |\lambda_i|}\,\mathbb{E}\left(e^{\sum_{i=1}^m \langle \lambda_i, Y^n_i\rangle}\right).$$

Now note that, like in the proof of Proposition 3.5.2, we can show that for $i \neq j$ the random variables Y^n_i and Y^n_j are independent. Therefore, we have that

$$\mathbb{E}\left(e^{\sum_{i=1}^m \langle \lambda_i, Y^n_i\rangle}\right) = \prod_{i=1}^m \mathbb{E}\left(e^{\langle \lambda_i, Y^n_i\rangle}\right).$$

Moreover, again following the proof of Proposition 3.5.2, one can show that

$$\mathbb{E}\left(e^{\langle \lambda_i, Y^n_i\rangle}\right) = M_{x_0}(\lambda_i)^{\lfloor m^{-1}n\rfloor}.$$

Combining everything, we find that

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left(e^{n\sum_{i=1}^m \langle \lambda_i, v^{n,m,i}_{\lfloor m^{-1}n\rfloor}\rangle}\right) \\ &\qquad \leqslant \limsup_{n\to\infty}\left\{\frac{C}{mn}\sum_{i=1}^m |\lambda_i| + \frac{Cr^2}{m^3}\sum_{i=1}^m |\lambda_i| + \frac{\lfloor m^{-1}n\rfloor}{n}\sum_{i=1}^m \Lambda_{x_0}(\lambda_i)\right\}\end{aligned}$$

$$= \frac{Cr^2}{m^3}\sum_{i=1}^{m}|\lambda_i| + \frac{1}{m}\sum_{i=1}^{m}\Lambda_{x_0}(\lambda_i).$$

Putting everything together, we obtain

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1}n\rfloor},\ldots,v^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right)\in\Gamma\right)$$
$$\leqslant -\inf_{(v_1,\ldots,v_m)\in\Gamma}\sup_{(\lambda_1,\ldots,\lambda_m)\in(T_{x_0}M)^m}\sum_{i=1}^{m}\left\{\langle\lambda_i,v_i\rangle - m^{-1}\Lambda_{x_0}(\lambda_i) - m^{-3}Cr^2|\lambda_i|\right\}.$$

This concludes the proof of the upper bound for compact sets.
To extend the upper bound to all closed sets, one should simply notice that

$$\left(v^{n,m,1}_{\lfloor m^{-1}n\rfloor},\ldots,v^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right)\in\overline{B(0,r)}^m$$

almost surely, where r is the uniform bound of the increments. Since M is complete, $\overline{B(0,r)}^m$ is compact, so that the sequence is exponentially tight. □

It now remains to transfer the upper bound in Proposition 3.5.7 for the process in $(T_{x_0}M)^m$ related to $\{\left(\frac{1}{n}*\mathcal{S}\right)_n\}_{n\geqslant 0}$ to the upper bound of the large deviation principle for $\{\left(\frac{1}{n}*\mathcal{S}\right)_n\}_{n\geqslant 0}$. With all preparations done, the only thing that remains to be shown, is that the upper bound has the desired form.

Proposition 3.5.8. *Let the assumptions of Theorem 3.5.1 be satisfied. Then for any $F\subset M$ closed we have*

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(\frac{1}{n}*\mathcal{S}\right)_n\in F\right)\leqslant -\inf_{x\in F}I_M(x),$$

where

$$I_M(x) = \inf\{\Lambda^*_{x_0}(v)|v\in\operatorname{Exp}^{-1}_{x_0}x\}.$$

Proof. Let $F\subset M$ be closed and pick $m\in\mathbb{N}$ such that $m\geqslant\frac{2r}{\iota(\overline{B(x_0,r)})}$, where r denotes the uniform bound of the increments. Let $\Psi_m:(T_{x_0}M)^m\to M$ be the recursive application of the Riemannian exponential map defined just above Section 3.5.1. Because Ψ_m is continuous, we have that $\Psi_m^{-1}F\subset(T_{x_0}M)^m$ is closed. Hence, by Proposition 3.5.7 we find that

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(\frac{1}{n}*\mathcal{S}\right)_n\in F\right)$$
$$\leqslant\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1}n\rfloor},\ldots,v^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right)\in\Psi_m^{-1}F\right)$$
$$\leqslant -\inf_{(v_1,\ldots,v_m)\in\Psi_m^{-1}F}\sup_{(\lambda_1,\ldots,\lambda_m)\in(T_{x_0}M)^m}\frac{1}{m}\sum_{i=1}^{m}\left\{\langle\lambda_i,mv_i\rangle - \Lambda_{x_0}(\lambda_i) - m^{-2}Cr^2|\lambda_i|\right\}.$$

Now observe that for every $\lambda \in T_{x_0}M$ we have $|\lambda| \leqslant |\lambda|^2 + 1$. Plugging this into the above estimate, keeping in mind the minus sign in front, we find that

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in F\right) \\ &\leqslant \frac{Cr^2}{m^2} - \inf_{(v_1,\ldots,v_m)\in\Psi_m^{-1}F} \sup_{(\lambda_1,\ldots,\lambda_m)\in(T_{x_0}M)^m} \frac{1}{m}\sum_{i=1}^m \left\{\langle\lambda_i, mv_i\rangle - \Lambda_{x_0}(\lambda_i) - m^{-2}Cr^2|\lambda_i|^2\right\}. \end{aligned}$$

We now focus on the infimum in the above expression. The necessity of replacing $|\lambda|$ with $|\lambda|^2$, and making the upper bound slightly worse, will become clear when we try to calculate this infimum further.

First, consider the map $\Lambda_m : T_{x_0}M \to \mathbb{R}$ defined by

$$\Lambda_m(\lambda) = \Lambda_{x_0}(\lambda) + \frac{1}{m^2}Cr^2|\lambda|^2.$$

and denote by Λ_m^* its Legendre transform. Then

$$\sup_{(\lambda_1,\ldots,\lambda_m)\in(T_{x_0}M)^m} \frac{1}{m}\sum_{i=1}^m \left\{\langle\lambda_i, mv_i\rangle - \Lambda_{x_0}(\lambda_i) - m^{-2}Cr^2|\lambda_i|\right\} = \frac{1}{m}\sum_{i=1}^m \Lambda_m^*(mv_i).$$

The latter may be interpreted as

$$\int_0^1 \Lambda_m^*(\dot\gamma_m(t))\,\mathrm{d}t,$$

where γ_m is piecewise geodesic on intervals of the form $[\frac{(i-1)}{m}, \frac{i}{m}]$ with speed $m\tilde{v}_i$, where $\tilde{v}_i = \tau_{x_0\gamma_m(\frac{(i-1)}{m})}v_i$.

Now note that since Λ_{x_0} is differentiable and convex, we find that Λ_m is differentiable and strictly convex. Furthermore, we have for every $u \in T_{x_0}M$ that

$$\begin{aligned} \Lambda_m^*(u) &= \sup_{\lambda\in T_{x_0}M}\left\{\langle\lambda, u\rangle - \Lambda_{x_0}(\lambda) - \frac{1}{m^2}Cr^2|\lambda|^2\right\} \\ &\leqslant \sup_{\lambda\in T_{x_0}M}\left\{\langle\lambda, u\rangle - \frac{1}{m^2}Cr^2|\lambda|^2\right\} \\ &< \infty. \end{aligned}$$

Here we used that Λ_{x_0} is non-negative, because the expectation of μ_{x_0} is 0. We conclude that Λ_m^* is everywhere finite. Note that this does not contradict the fact that the rate function might be infinite, since Λ_m^* merely provides a lower bound of the rate function. Because Λ_m^* is everywhere finite, it follows from Lemma 3.8.1 that Λ_m^* is strictly convex and differentiable.

The above shows that we can apply Proposition 3.4.11, giving us that minimizing trajectories for the functional

$$\int_0^1 \Lambda_m^*(\dot\gamma(t))\,\mathrm{d}t$$

are geodesics. Because for every $x \in F$ and every $v \in \operatorname{Exp}_{x_0}^{-1} x$ we have that $(\frac{1}{m}v, \ldots, \frac{1}{m}v) \in \Psi_m^{-1} F$, we find that

$$
\begin{aligned}
&- \inf_{(v_1,\ldots,v_m)\in\Psi_m^{-1}F} \sup_{(\lambda_1,\ldots,\lambda_m)\in(T_{x_0}M)^m} \frac{1}{m}\sum_{i=1}^m \left\{\langle \lambda_i, m v_i\rangle - \Lambda_{x_0}(\lambda_i)\right\} - m^{-2}Cr^2|\lambda_i|^2\right\} \\
&\qquad = - \inf_{v\in\operatorname{Exp}_{x_0}^{-1}F} \sup_{\lambda\in T_{x_0}M} \left\{\langle \lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda|^2\right\}.
\end{aligned}
$$

Now note that

$$
\begin{aligned}
\lim_{m\to\infty} &\sup_{\lambda\in T_{x_0}M} \left\{\langle \lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda|^2\right\} \\
&= \sup_{\lambda\in T_{x_0}M} \lim_{m\to\infty} \left\{\langle \lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda|^2\right\} \qquad (3.5.7)\\
&= \sup_{\lambda\in T_{x_0}M} \left\{\langle \lambda, v\rangle - \Lambda_{x_0}(\lambda)\right\},
\end{aligned}
$$

because $\Lambda_m(\lambda) = \langle \lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda|^2$ is increasing in m for every $\lambda \in T_{x_0}M$. Furthermore, we have

$$
\langle \lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda| \geqslant \langle \lambda, v\rangle - r|\lambda| - m^{-2}Cr^2|\lambda|^2,
$$

because the support of μ_{x_0} is contained in $\overline{B(0,r)}$. Furthermore, one may compute that if $|v| > r$, then

$$
\sup_{\lambda\in T_{x_0}M} \left\{\langle \lambda, v\rangle - r|\lambda| - m^{-2}Cr^2|\lambda|^2\right\} = \frac{m^2}{4Cr^2}(|v| - r)^2. \qquad (3.5.8)
$$

Now write

$$
\operatorname{Exp}_{x_0}^{-1} F = \left(\operatorname{Exp}_{x_0}^{-1} F \cap \overline{B(0,2r)}\right) \cup \left(\operatorname{Exp}_{x_0}^{-1} F \cap \overline{B(0,2r)}^C\right).
$$

Note that by (3.5.8), we find that

$$
\begin{aligned}
\lim_{m\to\infty} &\inf_{v\in\operatorname{Exp}_{x_0}^{-1}F\cap\overline{B(0,2r)}^C} \sup_{\lambda\in T_{x_0}M} \left\{\langle \lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda|^2\right\} \\
&\geqslant \lim_{m\to\infty} \inf_{v\in\operatorname{Exp}_{x_0}^{-1}F\cap\overline{B(0,2r)}^C} \sup_{\lambda\in T_{x_0}M} \left\{\langle \lambda, v\rangle - r|\lambda| - m^{-2}Cr^2|\lambda|^2\right\} \\
&\geqslant \lim_{m\to\infty} \frac{m^2}{4Cr^2} r^2 \\
&= \infty,
\end{aligned}
$$

where we used in the one but last line that $|v| \geqslant 2r$. Also, because $|v| \geqslant 2r \geqslant r$, we have

$$
\sup_{\lambda\in T_{x_0}M} \left\{\langle \lambda, v\rangle - \Lambda_{x_0}(\lambda)\right\} = \infty,
$$

so that

$$\lim_{m\to\infty} \inf_{v\in \mathrm{Exp}_{x_0}^{-1}F\cap \overline{B(0,2r)}^C} \sup_{\lambda\in T_{x_0}M} \left\{\langle\lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda|^2\right\}$$
$$= \inf_{v\in \mathrm{Exp}_{x_0}^{-1}F\cap \overline{B(0,2r)}^C} \sup_{\lambda\in T_{x_0}M} \{\langle\lambda, v\rangle - \Lambda_{x_0}(\lambda)\}\,.$$

For the other part, because $\mathrm{Exp}_{x_0}^{-1}F \cap \overline{B(0,2r)}$ is compact, it follows from (3.5.7) that

$$\lim_{m\to\infty} \inf_{v\in \mathrm{Exp}_{x_0}^{-1}F\cap \overline{B(0,2r)}} \sup_{\lambda\in T_{x_0}M} \left\{\langle\lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda|^2\right\}$$
$$= \inf_{v\in \mathrm{Exp}_{x_0}^{-1}F\cap \overline{B(0,2r)}} \sup_{\lambda\in T_{x_0}M} \{\langle\lambda, v\rangle - \Lambda_{x_0}(\lambda)\}\,.$$

Collecting everything, we find that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in F\right) \\
&\leqslant \lim_{m\to\infty}\left(\frac{Cr^2}{m^2} - \inf_{v\in \mathrm{Exp}_{x_0}^{-1}F} \sup_{\lambda\in T_{x_0}M} \left\{\langle\lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda|^2\right\}\right) \\
&= -\lim_{m\to\infty} \inf_{v\in \mathrm{Exp}_{x_0}^{-1}F\cap \overline{B(0,2r)}} \sup_{\lambda\in T_{x_0}M} \left\{\langle\lambda, v\rangle - \Lambda_{x_0}(\lambda) - m^{-2}Cr^2|\lambda|^2\right\} \\
&= -\inf_{v\in \mathrm{Exp}_{x_0}^{-1}F} \sup_{\lambda\in T_{x_0}M} \{\langle\lambda, v\rangle - \Lambda_{x_0}(\lambda)\} \\
&= -\inf_{x\in F} I_M(x),
\end{aligned}$$

which concludes the proof. □

3.5.2. Lower bound of the large deviation principle for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$

In this section we prove the large deviation lower bound for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$. In order to do this, we need a refinement of Proposition 3.5.2, which may be proven in a similar way.

Proposition 3.5.9. *Let the assumptions of Theorem 3.5.1 be satisfied. Let $m \in \mathbb{N}$ and set $Z_n^m = \frac{1}{n}\sum_{k=1}^{\lfloor m^{-1}n\rfloor} \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X_k^n$. Finally, define $\Lambda_{x_0}(\lambda) = \log \mathbb{E}(e^{\langle\lambda, X_1\rangle})$. Then $\{Z_n^m\}_{n\geqslant 1}$ satisfies in $T_{x_0}M$ the large deviation principle with good rate function*

$$I_m(v) = \frac{1}{m}\Lambda_{x_0}^*(mv),$$

where $\Lambda_{x_0}^*(v) = \sup_{\lambda\in T_{x_0}M}\{\langle\lambda, v\rangle - \Lambda_{x_0}(\lambda)\}$.

We are now able to prove the large deviation lower bound for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$.

Proposition 3.5.10. *Let the assumptions of Theorem 3.5.1 be satisfied. Then for any $G \subset M$ open,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in G\right) \geqslant -\inf_{x\in G} I_M(x),$$

where I_M is as in (3.5.1).

Proof. It suffices to show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in G\right) \geqslant -I_M(x)$$

for every $x \in G$.

So fix $x \in G$ and pick $v \in \mathrm{Exp}_{x_0}^{-1} x$. Because G is open, there exists an $\varepsilon > 0$ such that $B(x, \varepsilon) \subset G$. Let $m \in \mathbb{N}$ such that $m \geqslant \frac{2r}{\iota(\overline{B(x_0,r)})}$, where r is the uniform bound on the increments of the geodesic random walk.

We again need to identify the geodesic random walk with a tuple in $(T_{x_0}M)^m$. However, this time the parallel transport back to $T_{x_0}M$ is carried out by first transporting to a well-chosen point on the geodesic $\gamma_v(t) = \mathrm{Exp}_{x_0}(tv)$ and then to x_0 along this geodesic.

More precisely, we define a map $\Psi_{m,x,v} : (T_{x_0}M)^m \to M$ that allows us to identify the random variable $(\frac{1}{n} * \mathcal{S})_n \in M$ with a vector of random variables in $(T_{x_0}M)^m$. To this end, define for $0 \leqslant i \leqslant m$ the points $y_i = \mathrm{Exp}_{x_0}(\frac{i}{m} v)$. For $1 \leqslant i \leqslant m$ we define $\tau_{x_0 y_i}$ as parallel transport along the geodesic $\mathrm{Exp}_{x_0}(tv)$. Furthermore, for every $z \in M$ and every $0 \leqslant i \leqslant m-1$, we choose a geodesic $\gamma_{y_i x}$ of minimum length and denote by $\tau_{y_i x}$ parallel transport along this geodesic. We now define $\Psi_{m,x,v}(v_1, \dots, v_m)$ as follows. Define $x_1 = \mathrm{Exp}_{x_0}(\frac{1}{m} v_1)$ and if x_i is defined, we set $\tilde{v}_{i+1} = \tau_{y_i x_i} \tau_{x_0 y_i} v_i$ and define $x_{i+1} = \mathrm{Exp}_{x_i}(\frac{1}{m} \tilde{v}_{i+1})$. Finally, we set $\Psi_{m,x,v}(v_1, \dots, v_m) = x_m$.

Now note that by the triangle inequality, we have

$$d(x_i, x_0) \leqslant \frac{1}{m} \sum_{j=1}^{i} |v_j| \leqslant \frac{1}{m} \sum_{j=1}^{m} |v_j|$$

for any $1 \leqslant i \leqslant m$. Therefore, if $(v_1, \dots, v_m) \in B(v, 1)^m$, then we have $|v_j| \leqslant |v| + 1$, so that

$$d(x_i, x_0) \leqslant |v| + 1$$

for any $1 \leqslant i \leqslant m$. Because also $d(x_0, y_i) \leqslant \frac{i}{m}|v| \leqslant |v|$, we find that $x_i, y_i \in \overline{B(x_0, |v| + 1)}$ for all $1 \leqslant i \leqslant m$.

Writing $\eta = |v| + 1$, we will show that there exists a constant $m_0 \in \mathbb{N}$ such that for all $m \geqslant m_0$ we have

$$(v_1, \dots, v_m) \in B(v, \varepsilon^2/(8\eta))^m \Rightarrow \Psi_{m,x,v}(v_1, \dots, v_m) \in B(x, \varepsilon), \tag{3.5.9}$$

whenever $\varepsilon > 0$ is small enough.

To this end, let $K \subset M$ be a compact set, such that all geodesics of minimal length between points $x, y \in \overline{B(x_0, \eta)}$ are contained in K. Because K is compact, its injectivity radius $\iota(K)$ is strictly positive.

Fix $0 < \delta < \iota(K)$. We first show that for ε small enough and m large enough we have

$$d(x_i, y_i)^2 \leqslant \frac{i-1}{2m}\varepsilon^2 + \frac{i}{m^2}C \tag{3.5.10}$$

for $1 \leqslant i \leqslant m$. Here, $C > 0$ is some constant only depending on K and δ. We proceed by induction.

First consider the case $i = 1$. By taking m large enough, we can apply Proposition 3.4.10 to obtain a constant $C > 0$ (depending only on K and δ) such that

$$d(x_1, y_1)^2 = d\left(\mathrm{Exp}_{x_0}\left(\frac{1}{m}v_1\right), \mathrm{Exp}_{x_0}\left(\frac{1}{m}v\right)\right) \leqslant \frac{1}{m^2}C.$$

Now suppose that $d(x_i, y_i)^2 \leqslant \frac{i-1}{2m}\varepsilon^2 + \frac{i}{m^2}C$. Then in particular we have

$$d(x_i, y_i)^2 \leqslant \frac{\varepsilon^2}{2} + \frac{1}{m}C,$$

which can be made smaller than $\frac{\delta}{2}$ by taking ε sufficiently small and m sufficiently large. In that case, we may again apply Proposition 3.4.10, so that for the same constant $C > 0$ as above, we have

$$\begin{aligned}
d(x_{i+1}, y_{i+1})^2 &= d\left(\mathrm{Exp}_{x_i}\left(\frac{1}{m}\tau_{y_i x_i}\tau_{x_0 y_i}v_{i+1}\right), \mathrm{Exp}_{y_i}\left(\frac{1}{m}\tau_{x_0 y_i}v\right)\right)\\
&\leqslant d(x_i, y_i)^2 + 2\frac{1}{m}\langle \tau_{x_0 y_i}v_{i+1} - \tau_{x_0 y_i}v, \mathrm{Exp}_{y_i}^{-1}x_i\rangle + \frac{1}{m^2}C\\
&\leqslant \frac{i-1}{2m}\varepsilon^2 + \frac{i}{m^2}C + 2\frac{1}{m}|\tau_{x_0 y_i}v_{i+1} - \tau_{x_0 y_i}v||\mathrm{Exp}_{y_i}^{-1}x_i| + \frac{1}{m^2}C\\
&= \frac{i-1}{2m}\varepsilon^2 + \frac{i+1}{m^2}C + \frac{2}{m}|v_i - v|d(x_i, y_i).
\end{aligned}$$

Now, observe that $d(x_i, y_i) \leqslant 2\eta$ since $x_i, y_i \in \overline{B(x_0, \eta)}$. Using this, together with the induction hypothesis and the fact that $|v_i - v| \leqslant \frac{\varepsilon^2}{8\eta}$, we find

$$d(x_i, y_i)^2 \leqslant \frac{i-1}{2m}\varepsilon^2 + \frac{i+1}{m^2}C + \frac{1}{2m}\varepsilon^2 = \frac{2i}{2m}\varepsilon^2 + \frac{i+1}{m^2}C,$$

as desired.

Now taking $i = m$ in (3.5.10), we obtain

$$d(x_m, y_m)^2 \leqslant \frac{\varepsilon^2}{2} + C\frac{1}{m},$$

whenever $(v_1, \ldots, v_m) \in B(v, \varepsilon^2/(8\eta))^m$. It follows that if we take $m_0 > \frac{2C}{\varepsilon^2}$, then we obtain for $m > m_0$ that

$$d\left(\Psi_{m,x,v}(v_1, \ldots, v_m), x\right)^2 = d(x_m, y_m)^2 < \frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2} = \varepsilon^2$$

as desired.

Fixing m_0 and C as above, let $m \geqslant m_0$ be large enough so that we can define

$$\tilde{v}_k^{n,m,l} \in \mathrm{Exp}^{-1}_{(\frac{1}{n}*\mathcal{S})_{n_l}} \left(\left(\frac{1}{n} * \mathcal{S} \right)_{n_l+k} \right) \subset T_{(\frac{1}{n}*\mathcal{S})_{n_l}} M$$

like in (3.5.5). Different from before, we now define the vectors

$$v_k^{n,m,l} = \tau^{-1}_{x_0 y_{n_l}} \tau^{-1}_{y_{n_l}(\frac{1}{n}*\mathcal{S})_{n_l}} \tilde{v}_k^{n,m,l} \in T_{x_0} M, \tag{3.5.11}$$

using the parallel transport procedure used in the definition of the map $\Psi_{m,x,v}$. As a consequence, by construction we obtain

$$\Psi_{m,x,v}\left(v^{n,m,1}_{\lfloor m^{-1} n \rfloor}, \ldots, v^{n,m,m}_{\lfloor m^{-1} n \rfloor} \right) = \left(\frac{1}{n} * \mathcal{S} \right)_n .$$

Using this, together with the implication in (3.5.9), we find

$$\begin{aligned} \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S} \right)_n \in G \right) &\geqslant \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S} \right)_n \in B(x, \varepsilon) \right) \\ &\geqslant \mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1} n \rfloor}, \ldots, v^{n,m,m}_{\lfloor m^{-1} n \rfloor} \right) \in B(v, \varepsilon^2/(8\eta))^m \right). \end{aligned}$$

Now define for $1 \leqslant i \leqslant m$ the random variables

$$Y_i^n = \tau^{-1}_{x_0 y_{n_{i-1}}} \tau^{-1}_{y_{n_{i-1}}(\frac{1}{n}*\mathcal{S})_{n_{i-1}}} \sum_{k=n_{i-1}+1}^{n_i} \tau^{-1}_{(\frac{1}{n}*\mathcal{S})_{n_{i-1}}(\frac{1}{n}*\mathcal{S})_{k-1}} X_k^n \in T_{x_0} M,$$

where the parallel transport $\tau^{-1}_{(\frac{1}{n}*\mathcal{S})_{n_{i-1}}(\frac{1}{n}*\mathcal{S})_{k-1}}$ is carried out along the trajectory of the geodesic random walk. The sum is then transported from $T_{(\frac{1}{n}*\mathcal{S})_{n_{i-1}}} M$ to $T_{x_0} M$ as in the definition of $v_k^{n,m,l}$ in (3.5.11).

In the same way as we obtained (3.5.6) in the proof of Proposition 3.5.7, we find that there exists a constant $\tilde{C} > 0$ such that

$$\left| v^{n,m,1}_{\lfloor m^{-1} n \rfloor} - Y_i^n \right| \leqslant \tilde{C} \frac{1}{nm} + \tilde{C} r^2 \frac{1}{m^3}.$$

Hence, we can take m large enough such that almost surely we have

$$\left| v^{n,m,1}_{\lfloor m^{-1} n \rfloor} - Y_i^n \right| < \frac{\varepsilon^2}{16\eta}.$$

But then we find that if $Y_i^n \in B(v, \varepsilon^2/(16\eta))$, then $v^{n,m,1}_{\lfloor m^{-1} n \rfloor} \in B(v, \varepsilon^2/(8\eta))$. This implies that

$$\begin{aligned}\mathbb{P}\left(\left(v^{n,m,1}_{\lfloor m^{-1}n\rfloor},\ldots,v^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right)\in B(v,\varepsilon^2/(8\eta))^m\right)&\\ \geqslant \mathbb{P}\left((Y_1^n,\ldots,Y_m^n)\in B(v,\varepsilon^2/(16\eta))^m\right).&\end{aligned}$$

Now note that, like in the proof of Proposition 3.5.2, we can show that the random variables Y_i^n and Y_j^n are independent and identically distributed for $i \neq j$, so that

$$\begin{aligned}\mathbb{P}\left((Y_1^n,\ldots,Y_m^n)\in B(v,\varepsilon^2/(16\eta))^m\right) &= \prod_{i=1}^m \mathbb{P}\left(Y_i^n \in B(v,\varepsilon^2/(16\eta))\right)\\ &= \mathbb{P}\left(Y_1^n \in B(v,\varepsilon^2/(16\eta))\right)^m.\end{aligned}$$

Furthermore, by Proposition 3.5.9 we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(Y_1^n\in B(v,\varepsilon^2/(16\eta))\right)\geqslant -\frac{1}{m}\Lambda^*_{x_0}(v).$$

Combining everything, we find that

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(\frac{1}{n}*\mathcal{S}\right)_n\in G\right)&\geqslant m\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(Y_1^n\in B(v,\varepsilon^2/(16\eta))\right)\\ &\geqslant -\Lambda^*_{x_0}(v).\end{aligned}$$

Since this holds for all $v\in \mathrm{Exp}_{x_0}^{-1}x$, we find that

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(\frac{1}{n}*\mathcal{S}\right)_n\in G\right)\geqslant -\inf_{v\in\mathrm{Exp}_{x_0}^{-1}x}\Lambda^*_{x_0}(v) = -I_M(x),$$

which concludes the proof. □

3.6. Concentration inequalities

Concentration inequalities are concerned with the problem of where the mass of a given probability measure is concentrated, see e.g. [67]. In this section we derive Gaussian concentration inequalities for geodesic random walks with bounded increments. In contrast to the large deviation principle which only holds asymptotically, Gaussian concentration inequalities provide exponential estimates for every n large enough. However, these are only estimates for probabilities of deviating from the expected value, while the large deviation principle gives us the asymptotic behaviour of the probability of any event.

We prove the Gaussian concentration inequalities for geodesic random walks by applying the Azuma-Hoeffding inequality (see e.g. [2]) to a suitable supermartingale. This supermartingale is obtained by using Proposition 3.4.10, as we will show in the following proposition.

Proposition 3.6.1. *Let (M,g) be a complete Riemannian manifold. Fix $x_0\in M$ and let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x\in\mathcal{P}(T_xM)$ for all $x\in M$.*

Fix $n \geqslant 1$ and let $\{(\frac{1}{n} * \mathcal{S})_j\}_{j\geqslant 1}$ be a $\frac{1}{n}$-rescaled geodesic random walk started at x_0 with independent increments $\{X_j^n\}_{j\geqslant 1}$, compatible with $\{\mu_x\}_{x\in M}$. Assume the increments are bounded. Assume furthermore that the collection $\{\mu_x\}_{x\in M}$ satisfies the consistency property in Definition 3.2.7. Define the curve $E_n : [0,n] \to M$ by

$$E_n(t) = \mathrm{Exp}_{x_0}\left(\frac{t}{n}\mathbb{E}(X_1^n)\right).$$

Then there exists a constant $C > 0$ only depending on the bound of the increments, such that

$$M_k = d\left(\left(\frac{1}{n} * \mathcal{S}\right)_k, E(k)\right)^2 - \frac{k}{n^2}C$$

is a super-martingale up to time n with respect to the filtration $\{\mathcal{F}_k^n\}_{0\leqslant k\leqslant n}$ given by

$$\mathcal{F}_k^n = \sigma(X_1^n, \dots, X_k^n).$$

Proof. Because the increments are bounded and are identically distributed, they are uniformly bounded. Writing $r > 0$ for this bound, we have

$$d\left(\left(\frac{1}{n} * \mathcal{S}\right)_k, x_0\right) \leqslant \frac{1}{n}\sum_{l=1}^{k}|X_l^n| \leqslant \frac{kr}{n} \leqslant r$$

for any $0 \leqslant k \leqslant n$. From this it follows that up to time n, the rescaled geodesic random walk is almost surely contained in the compact set $K = \overline{B(x_0, r)}$, which does not depend on n. Therefore, by Proposition 3.4.10 there exists a constant $C > 0$ such that

$$\begin{aligned} d&\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k+1}, E(k+1)\right)^2 - d\left(\left(\frac{1}{n} * \mathcal{S}\right)_k, E(k)\right)^2 \\ &\leqslant -2\frac{1}{n}\left\langle \tau^{-1}_{E(k)(\frac{1}{n}*\mathcal{S})_k} X_{k+1}^n - \tau_{x_0E(k)}\mathbb{E}(X_1^n), \mathrm{Exp}^{-1}_{\left(\frac{1}{n}*\mathcal{S}\right)_k} E(k)\right\rangle + \frac{1}{n^2}C. \quad (3.6.1)\end{aligned}$$

From the independence of the increments, together with the fact that $(\frac{1}{n} * \mathcal{S})_k$ is measurable with respect to $\mathcal{F}_k^n$, it follows that

$$\mathbb{E}(\tau^{-1}_{E(k)(\frac{1}{n}*\mathcal{S})_n} X_{k+1}^n | \mathcal{F}_k^n) = \tau^{-1}_{E(k)(\frac{1}{n}*\mathcal{S})_n}\mathbb{E}(X_{k+1}^n).$$

Since the increments are identically distributed, it follows from Proposition 3.2.8 that their expectations are invariant under parallel transport. This implies that

$$\tau^{-1}_{E(k)(\frac{1}{n}*\mathcal{S})_n}\mathbb{E}(X_{k+1}^n) = \tau_{x_0E(k)}\mathbb{E}(X_1^n).$$

Collecting everything, we obtain that

$$\mathbb{E}\left(\left\langle \tau^{-1}_{E(k)(\frac{1}{n}*\mathcal{S})_k} X_{k+1}^n - \tau_{x_0E(k)}\mathbb{E}(X_1^n), \mathrm{Exp}^{-1}_{\left(\frac{1}{n}*\mathcal{S}\right)_k} E(k)\right\rangle \middle| \mathcal{F}_k^n\right) = 0.$$

Combined with the estimate in (3.6.1), this implies that

$$\mathbb{E}\left(d\left(\left(\frac{1}{n}*\mathcal{S}\right)_{k+1}, E(k+1)\right)^2 \middle| \mathcal{F}_k^n\right) \leqslant d\left(\left(\frac{1}{n}*\mathcal{S}\right)_k, E(k)\right)^2 + \frac{1}{n^2}C \quad (3.6.2)$$

Now define the process

$$M_k = d\left(\left(\frac{1}{n}*\mathcal{S}\right)_k, E(k)\right)^2 - \frac{k}{n^2}C.$$

Using the estimate in (3.6.2), we find

$$\mathbb{E}(M_{k+1}|\mathcal{F}_k) \leqslant d\left(\left(\frac{1}{n}*\mathcal{S}\right)_k, x_0\right)^2 - \frac{k+1}{n^2}C + \frac{1}{n^2}C = M_k,$$

showing that M_k is a super-martingale. □

Remark 3.6.2. In the case of Euclidean space, one can actually obtain a martingale by taking $C = \mathbb{E}(|X_1|^2)$.

We are now able to derive Gaussian concentration inequalities for $(\frac{1}{n}*\mathcal{S})_n$.

Proposition 3.6.3. *Let (M, g) be a complete Riemannian manifold. Fix $x_0 \in M$ and let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_xM)$ for all $x \in M$. For every $n \geqslant 1$, let $\{(\frac{1}{n}*\mathcal{S})_j\}_{j\geqslant 1}$ be a $\frac{1}{n}$-rescaled geodesic random walk started at x_0 with independent increments $\{X_j^n\}_{j\geqslant 1}$, compatible with $\{\mu_x\}_{x\in M}$. Assume the increments are bounded. Assume furthermore that the collection $\{\mu_x\}_{x\in M}$ satisfies the consistency property in Definition 3.2.7. Define for every $n \geqslant 1$ the curve $E_n : [0, n] \to M$ by*

$$E_n(t) = \mathrm{Exp}_{x_0}\left(\frac{t}{n}\mathbb{E}(X_1^n)\right).$$

Then there exists a constant $L > 0$ such that for every $\varepsilon > 0$ there exists a $N \in \mathbb{N}$ such that for all $n \geqslant N$ and $k = 1, \ldots, n$ we have

$$\mathbb{P}\left(d\left(\left(\frac{1}{n}*\mathcal{S}\right)_k, E_n(k)\right) > \varepsilon\right) \leqslant e^{-\frac{1}{8kL^2}n^2\varepsilon^4}.$$

The constant L can be chosen to only depend on the bound of the increments.

Proof. By Proposition 3.6.1, there exists a constant $C > 0$ only depending on the bound of the increments, such that for every n

$$M_k^n = d\left(\left(\frac{1}{n}*\mathcal{S}\right)_k, E_n(k)\right)^2 - \frac{k}{n^2}C$$

is a super-martingale up to time n.
Note that $M_0^n = 0$. Furthermore,

$$|M_{k+1}^n - M_k^n|$$

$$\begin{aligned}
&\leqslant \left| d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k+1}, E_n(k+1)\right)^2 - d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k}, E_n(k)\right)^2 \right| + \frac{1}{n^2}C \\
&= \left| d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k+1}, E_n(k+1)\right) - d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k}, E_n(k)\right) \right| \times \\
&\quad \times \left| d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k+1}, E_n(k+1)\right) + d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k}, E_n(k)\right) \right| + \frac{1}{n^2}C.
\end{aligned}$$

Writing r for the bound of the increments, the triangle inequality (via x_0) gives us that

$$\left| d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k+1}, E_n(k+1)\right) + d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k}, E_n(k)\right) \right| \leqslant 4r.$$

Again applying the triangle inequality, we also obtain that

$$\begin{aligned}
&\left| d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k+1}, E_n(k+1)\right) - d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k}, E_n(k)\right) \right| \\
&\quad \leqslant \left| d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k+1}, E_n(k)\right) - d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k}, E_n(k)\right) \right| \\
&\qquad + \left| d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k+1}, E_n(k+1)\right) - d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k+1}, E_n(k)\right) \right| \\
&\quad \leqslant \frac{1}{n}|X_{k+1}^n| + \frac{1}{n}|\mathbb{E}(X_1^n)| \\
&\quad \leqslant \frac{2r}{n}.
\end{aligned}$$

Collecting everything, we find that

$$|M_{k+1}^n - M_k^n| \leqslant \frac{8r^2}{n} + \frac{1}{n^2}C \leqslant L\frac{1}{n}$$

for some $L > 0$.
By the Azuma-Hoefdding inequality (see e.g. [2]) we obtain

$$\mathbb{P}\left(d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k}, E_n(k)\right)^2 - \frac{k}{n^2}C \geqslant \rho \right) \leqslant e^{-\frac{1}{2kL^2}n^2\rho^2}$$

for every $k = 1, \ldots, n$.
Now fix $\varepsilon > 0$ and take N large enough such that $\frac{1}{n}C < \frac{\varepsilon^2}{2}$ for all $n \geqslant N$. In particular, this implies $\frac{k}{n^2}C < \frac{\varepsilon^2}{2}$ for all $k = 1, \ldots, n$.
But then we find for $n \geqslant N$ that

$$\begin{aligned}
&\mathbb{P}\left(d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k}, E_n(k)\right) > \varepsilon \right) \\
&\quad \leqslant \mathbb{P}\left(d\left(\left(\frac{1}{n} * \mathcal{S}\right)_{k}, E_n(k)\right)^2 - \frac{k}{n^2}C \geqslant \varepsilon^2/2 \right) + \mathbb{P}\left(\frac{k}{n^2}C \geqslant \varepsilon^2/2 \right)
\end{aligned}$$

$$\leqslant e^{-\frac{1}{8kL^2}n^2\varepsilon^4}$$

which is the desired estimate. Because L depends only on the constants r and C, it may be chosen as claimed. □

Remark 3.6.4. Note that the curve E_n defined in Propositions 3.6.1 and 3.6.3 only depends on n via a rescaling in time. More precisely, define the curve $E : [0,1] \to M$ by

$$E(t) = \mathrm{Exp}_{x_0}\left(\frac{t}{n}\mathbb{E}(X_1^1)\right).$$

Since the increments are identically distributed, we have $\mathbb{E}(X_1^n) = \mathbb{E}(X_1^1)$ for all n. This shows that $E_n(t) = E\left(\frac{t}{n}\right)$. In particular, this shows that the image $E_n([0,n])$ is the same curve for every $n \geqslant 1$.

Remark 3.6.5. Because we assume the increments of the random walk are bounded, Proposition 3.6.3 is only interesting for small ε. In Euclidean space we can actually improve the concentration inequality by having ε^2 in the exponential rather than ε^4. To obtain this, one utilizes the additive structure of Euclidean space, which lacks in the Riemannian setting.

3.7. Concluding remarks

We conclude this chapter by discussing possible extensions of the results obtained in this chapter.

First of all, in Cramér's theorem for geodesic random walks we assume the increments are bounded and have expectation 0. The use of the boundedness of the increments is two-fold. Firstly, it assures that the rescaled geodesic random walks remain almost surely in some compact subset of the manifold. As a consequence, the exponential tightness of the sequence $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$ is immediate. In the Euclidean setting, the exponential tightness follows from the fact that the moment generating function of the increments is everywhere finite. It should be possible to obtain exponential tightness in the Riemannian setting with a similar argument.

Secondly, we use the fact that the random walk remains in a compact set to be able to cut the random walk into pieces, each of which we can then pull-back to a tangent space and use Cramér's theorem there. However, this containment in a compact set need not be almost sure, but it should occur with very high probability. The finiteness of moment generating functions should also allow us to construct such sets, possibly growing with n in an appropriate manner. All in all, there is reason to believe that the boundedness assumption can be replaced with an assumption on the moment generating functions of the measures μ_x.

Overcoming the assumption that the increments need to be centered is part of another possible extension, namely to geodesic random walks with drift. To obtain such random walks, we take a (deterministic) vector field V. If we then sample an increment according to the distribution μ_x on T_xM, we add $V(x)$. The drift we have added to the random walk is then formed by the flow of the vector field V. If we start with a geodesic random walk with independent, identically distributed increments, and add a (deterministic) drift, it is expected that one can again prove a large deviation principle.

Finally, we remark that the estimate in Proposition 3.5.4 is one of the most important ingredients of the proof of Theorem 3.5.1. Indeed, it allows us in some sense to connect large deviations for $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 0}$ in M to large deviations for the sums $\{\frac{1}{n}\sum_{k=1}^n \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X_k^n\}_{n\geqslant 0}$ in the tangent space $T_{x_0}M$. Therefore, by making appropriate assumptions on the sequence $\{\frac{1}{n}\sum_{k=1}^n \tau^{-1}_{x_0(\frac{1}{n}*\mathcal{S})_{k-1}} X_k^n\}_{n\geqslant 0}$, for example in the spirit of the Gärtner-Ellis theorem (Theorem 2.1.12), we can obtain more general results than Cramér's theorem for geodesic random walks in a similar way.

3.8. Appendix: Some convex analysis

In this appendix we collect a result from convex analysis. Although this is probably well-known, we include it for the sake of being self-contained.

Lemma 3.8.1. *Let V be a vector space, and let $F : V \to \mathbb{R}$ be strictly convex and differentiable. Then its Legendre transform F^* is strictly convex and differentiable on the interior of its domain $\mathcal{D}^\circ_{F^*}$.*

Proof. The differentiability of F^* follows from [82][Theorem 26.3].
For the strict convexity, we first prove that for each $v \in \mathcal{D}^\circ_{F^*}$, there exists a $\lambda^*_v \in V$ such that

$$F^*(v) = \langle \lambda^*_v, v\rangle - F(\lambda^*_v).$$

Indeed, suppose this is not the case. Because $F^*(v) < \infty$, we can find a sequence λ_n such that

$$F^*(v) = \lim_{n\to\infty} \langle \lambda_n, v\rangle - F(\lambda_n).$$

Because the map $\lambda \mapsto \langle \lambda, v\rangle - F_x(\lambda)$ is continuous, the sequence λ_n cannot contain a convergent subsequence, else the limit of this subsequence would serve as λ^*_v. We conclude that $\lim_{n\to\infty} |\lambda_n| = \infty$.
But then there exists a $w \in V$ such that $\lim_{n\to\infty}\langle \lambda_n, w\rangle = \infty$. To see this, suppose such a w does not exist. Denoting by $e_1, \dots, e_d$ a basis of V, we must have that $\langle \lambda_n, e_i\rangle$ is a bounded sequence for all $i = 1, \dots, d$. But then, by taking subsequences, we find $\langle \lambda_n, e_i\rangle$ converges for all $i = 1, \dots, d$, which contradicts the fact that $\lim_{n\to\infty} |\lambda_n| = \infty$.
Now consider $v + \varepsilon w \in V$ and let λ_n be the sequence found above. We have that

$$F^*(v + \varepsilon w) \geqslant \lim_{n\to\infty} \langle \lambda_n, v + \varepsilon w\rangle - F(\lambda_n) = F^*(v) + \varepsilon \lim_{n\to\infty} \langle \lambda_n, w\rangle = \infty.$$

We conclude that $v + \varepsilon w \notin \mathcal{D}_{F*}$ for any $\varepsilon > 0$, which contradicts the assumption that $v \in \mathcal{D}^{\circ}_{F*}$.

We are now ready to prove that F^* is strictly convex on $\mathcal{D}^{\circ}_{F*}$. To this end, fix $v, w \in \mathcal{D}^{\circ}_{F*}$, $v \neq w$ and $t \in (0,1)$ and assume that

$$F^*(tv + (1-t)w) = tF^*(v) + (1-t)F^*(w).$$

Now let λ_t^* be such that

$$F^*(tv + (1-t)w) = \langle tv + (1-t)w, \lambda_t^* \rangle - F(\lambda_t^*).$$

We find that

$$tF^*(v) + (1-t)F^*(w) = t(\langle \lambda_t^*, v \rangle - F(\lambda_t^*)) + (1-t)(\langle \lambda_t^*, w \rangle - F(\lambda_t^*)).$$

But then we find that

$$F^*(v) = \langle v, \lambda_t^* \rangle - F(\lambda_t^*)$$

and

$$F^*(w) = \langle w, \lambda_t^* \rangle - F(\lambda_t^*).$$

Now, because F is everywhere differentiable, it must be that $\nabla F(\lambda_t^*) = v$ and $\nabla F(\lambda_t^*) = w$, which contradicts the assumption that $v \neq w$. We conclude that F^* is strictly convex on $\mathcal{D}^{\circ}_{F*}$. □

4

Large deviations for random walks in Lie groups

In this chapter we study random walks in Lie groups. More precisely, we are interested in the large deviations for rescaled random walks, in a similar sense as in Chapter 3. The results in this chapter seem very similar, but as we will argue in Section 4.2.1, the random walks we study in Lie groups cannot always be considered as geodesic random walks. The main reason for this is that the Lie group exponential map does not necessarily coincide with the Riemannian exponential map. The results we present in this chapter are based on:

Rik Versendaal. "Large deviations for random walks on Lie groups". In: *Preprint; ArXiv:* 1909.05065 (2019). ArXiv: `1909.05065`.

Since the middle of the previous century, the study of random matrices has received a lot of attention. Of particular interest is the limiting behaviour of products of random matrices. This finds its applications for example in the study of disordered spin systems (see e.g. [27]), where the associated partition function is a product of transfer matrices, which are random because of the disorder. Certain physical quantities of the system are obtained by a limiting procedure, which in the case of disorder, is replaced by a law of large numbers or central limit theorem. Another application can be found in studying solutions to difference equations. One can for example think about the Schrödinger equation on a one-dimensional lattice with random vector potentials, see e.g. [14].
The limiting behaviour of products of random matrices was first studied in [11] and further developed in (among others) [42]. In these works, one takes a sequence $\{M_n\}_{n\geqslant 1}$ of matrix valued random variables and studies the product

$$\mathcal{S}_n = M_1 \cdots M_n.$$

In order to say anything about the limiting behaviour of the random variable $\mathcal{S}_n$, we take a matrix norm and consider the sequence of real-valued random variables given by $\log ||\mathcal{S}_n||$. It is then shown that under mild conditions we have

$$\lim_{n\to\infty} \frac{1}{n} \log ||\mathcal{S}_n|| = \gamma$$

almost surely, which is the analogue of the law of large numbers. The constant γ is referred to as the upper Lyapunov exponent. Furthermore, in [66] (see also [14]) it is shown that under additional assumptions, $\log ||\mathcal{S}_n||$ also satisfies the central limit theorem, i.e.,

$$\frac{\log ||\mathcal{S}_n|| - n\gamma}{\sqrt{n}}$$

converges in distribution to a Gaussian random variable. The same work also verifies the large deviation properties of the sequence $\{\log ||\mathcal{S}_n x||\}_{n\geqslant 1}$ of random variables, where x is some vector.

It is possible to go beyond matrix groups, and study products of elements of a general Lie group. For a random sequence $\{g_n\}_{n\geqslant 1}$ in a Lie group G, using the group operation, we can define the product

$$\mathcal{S}_n = g_1 g_2 \cdots g_n.$$

We will refer to this as a random walk in the Lie group G.

In order to study limit theorems like the law of large numbers and central limit theorem, we can no longer use a norm, since G is not necessarily a normed space. Instead, we can equip G with a left-invariant Riemannian metric with associated Riemannian distance d and study the real-valued random variables $d(\mathcal{S}_n, e)$, where e is the identity element of the group G. It is shown in [48] that if G is locally compact, there exists a $\gamma \geqslant 0$ such that almost surely

$$\lim_{n\to\infty} \frac{1}{n} d(\mathcal{S}_n, e) = \gamma.$$

Furthermore, the central limit theorem, i.e., the convergence of

$$\frac{d(\mathcal{S}_n, e) - \gamma n}{\sqrt{n}}$$

in distribution to a normal distribution, is studied in [91].

Another approach to study limit theorems, which we will be considering here, is not to transfer the problem to a real-valued setting, but to find a suitable way of rescaling the random walk in the Lie group G itself. For this, we slightly modify the definition of a random walk. Let $\mathfrak{g}$ denote the Lie algebra of G, and let $\{X_n\}_{n\geqslant 1}$ be a sequence of random variables in $\mathfrak{g}$. We define the random walk in G as

$$\mathcal{S}_n = \exp(X_1) \cdots \exp(X_n),$$

where $\exp : \mathfrak{g} \to G$ denotes the exponential map. Because $\mathfrak{g}$ is a vector space, we can rescale the sequence $\{X_n\}_{n\geqslant 1}$, allowing us to define the rescaled random walk by

$$\sigma_n^n = \exp\left(\frac{1}{n} X_1\right) \cdots \exp\left(\frac{1}{n} X_n\right).$$

However, from the Baker-Campbell-Hausdorff formula it follows after a formal computation that

$$\sigma_n^n = \exp\left(\frac{1}{n}\sum_{i=1}^{n} X_i + \mathcal{O}(1)\right),$$

which one obtains by counting the number of commutators. Therefore, it is not obvious how to use known results regarding the limiting behaviour of $\frac{1}{n}\sum_{i=1}^{n} X_i$ in order to study the limiting behaviour of σ_n^n. To overcome this problem, instead of simply rescaling the elements $\mathfrak{g}$ by $\frac{1}{n}$, one uses so-called dilations $D_{\frac{1}{n}} : \mathfrak{g} \to \mathfrak{g}$ as done in [15, 10, 44, 76]. The idea is to decompose an element $Y \in \mathfrak{g}$ as $Y = \sum_{i\geqslant 1} Y_i$, where Y_i is an i-th order commutator, meaning it is of the form $[Y_i^1, [\cdots, [Y_i^{i-1}, Y_i^i]]]$, where none of the Y_i^j are commutators. We call a Lie algebra nilpotent if there is some $l \in \mathbb{N}$ such that all commutators of order l vanish. In that case, Y may be written as a finite sum $Y = \sum_{i=1}^{l} Y_i$ and we define the dilation $D_{\frac{1}{n}} Y$ of Y by

$$D_{\frac{1}{n}} Y = \sum_{i=1}^{l} \frac{1}{n^i} Y_i$$

So essentially, we dilate the elements of $\mathfrak{g}$ in such a way that the problematic parts, being the (higher order) commutators, are scaled away in the limit by multiplying those by higher powers of $\frac{1}{n}$. Now the Baker-Campbell-Hausdorff formula will give us after a formal computation that

$$\prod_{i=1}^{n} \exp\left(D_{\frac{1}{n}} X_i\right) = \exp\left(\frac{1}{n}\sum_{i=1}^{n} X_i + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

This makes it at least more viable that in the limit this product should indeed behave like $\exp\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$. It is shown in [44, 76] that the law of large numbers is satisfied: if $\{X_n\}_{n\geqslant 1}$ are independent, identically distributed with $\mathbb{E}(X_1) = 0$ and with finite moment generating function in a neighbourhood of the origin, then almost surely

$$\lim_{n\to\infty} \prod_{i=1}^{n} \exp\left(D_{\frac{1}{n}} X_i\right) = 0.$$

The large deviations for the sequence

$$\left\{\prod_{i=1}^{n} \exp\left(D_{\frac{1}{n}} X_i\right)\right\}_{n\geqslant 0} \tag{4.0.1}$$

are studied in [10]. The proof uses path-space large deviations, first transferring the problem to $\mathbb{R}^d$ to use Mogulskii's theorem, followed by the contraction principle to get the large deviations for the end-point of the random walk.

However, if G admits a bi-invariant metric (which implies that its Lie algebra $\mathfrak{g}$ is reductive, see e.g. [60]), the processes $\mathcal{S}_n$ and σ_n^n are special cases of geodesic random walks as defined in [58]. The large deviations for these have been studied in Chapter 3. Therefore, if G admits a bi-invariant metric, Theorem 3.3.1 applies to the sequence $\{\sigma_n^n\}_{n\geqslant 1}$, so that the sequence satisfies in G the large deviation principle. Moreover, the corresponding rate function coincides with the rate function for the sequence of random variables in (4.0.1), where the higher order commutators are scaled away.

This raises the question whether the sequence $\{\sigma_n^n\}_{n\geqslant 1}$ also satisfies a large deviation principle when G does not necessarily admit a bi-invariant metric. This would cover the result for all (connected, finite-dimensional) Lie groups. Following the approach in Chapter 3, we will show that under some assumptions on the sequence $\{X_n\}_{n\geqslant 1}$, this is indeed the case. More precisely, we will prove that if $\{X_n\}_{n\geqslant 1}$ is a sequence of bounded, independent and identically distributed $\mathfrak{g}$-valued random variables, with $\mathbb{E}(X_1) = 0$ and everywhere finite moment generating function, then the sequence $\{\sigma_n^n\}_{n\geqslant 0}$ satisfies in G the large deviation principle with rate function I given by

$$I(g) = \inf\left\{\int_0^1 \Lambda^*(\dot\gamma(t))\,\mathrm{d}t \,\middle|\, \gamma \in AC([0,1];G), \gamma(0) = e, \gamma(1) = g\right\}.$$

Here, $\Lambda(\lambda) = \log\mathbb{E}(e^{\langle\lambda, X_1\rangle})$ denotes the log moment generating function, and Λ^* its Legendre transform given by

$$\Lambda^*(X) := \sup_{\lambda\in\mathfrak{g}}\{\langle\lambda, X\rangle - \Lambda(\lambda)\}.$$

This chapter is organised as follows. First, in Section 4.1 we introduce some theory on Lie groups and Lie algebras and fix the notation we use in what follows. With the notation fixed, we define in Section 4.2 the random walks in Lie groups we will be studying, and discuss their relation to geodesic random walks. In Section 4.3 we state our main theorem and give a sketch of its proof. Additionally, we discuss an example by considering the stochastic group. Section 4.4 is devoted to important estimates following from the Baker-Campbell-Hausdorff formula. Finally, we use these estimates to prove our main theorem in Section 4.5.

4.1. Lie groups and Lie algebras

In this section we collect the necessary notation and theory on Lie groups and Lie algebras. For more details, we refer to [21, 60, 68, 99] for general Lie group theory and to [51] for a treatment of matrix Lie groups.

Let G be a finite-dimensional Lie group, i.e., a finite dimensional group with a smooth manifold structure such that the group operations of multiplication and inversion are smooth. We write e for the identity element of G. The *Lie algebra* $\mathfrak{g}$ of G is defined as the tangent space T_eG at the identity.

Next, we want to equip $\mathfrak{g}$ with a *Lie bracket* $[\cdot,\cdot]$, which is a map from $\mathfrak{g}\times\mathfrak{g}$ into $\mathfrak{g}$

which is bilinear, skew-symmetric and satisfies the Jacobi identity:

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$$

for all $X, Y, Z \in \mathfrak{g}$. In order to construct such a Lie bracket, we need a different interpretation of the Lie algebra $\mathfrak{g}$.

To this end, we denote by $L_g : G \to G$ left multiplication with g. A vector field V on G is called *left-invariant* if for all $g, h \in G$ we have $\mathrm{d}L_g(h)(V(h)) = V(gh)$.

For every $X \in \mathfrak{g}$, we can define a left-invariant vector field X^L on G by setting

$$X^L(g) = \mathrm{d}L_g(e)(X). \tag{4.1.1}$$

This is a vector space isomorphism between the Lie algebra $\mathfrak{g}$ and the set of left-invariant vector fields over G. Indeed, its inverse is given by the evaluation of the vector field at the identity e. Therefore, the Lie algebra $\mathfrak{g}$ of G may be identified with the set of left-invariant vector fields over G. This set forms a Lie algebra under the Lie bracket $[V,W] = VW - WV$. Using this, we define the Lie bracket $[X,Y]$ for $X, Y \in \mathfrak{g}$ by $[X,Y] := [X^L, Y^L](e)$.

The above procedure also shows us that for every $g \in G$ we can identify the tangent space T_gM with $\mathfrak{g}$ via the isomorphism $\mathrm{d}L_g(e) : \mathfrak{g} \to T_gM$. Whenever we consider a tangent vector $X \in T_gM$ as element of $\mathfrak{g}$ or vice versa, we have this identification in mind.

4.1.1. Exponential map

We now define an important map that allows us to map elements of the Lie algebra to the Lie group. For this, first observe that for every $X \in \mathfrak{g}$, there exists a curve $\gamma_X : \mathbb{R} \to G$ satisfying $\gamma_X(0) = e$ and

$$\dot\gamma_X(t) = X^L(\gamma_X(t)). \tag{4.1.2}$$

In particular, $\dot\gamma_X(0) = X$. The fact that the curve γ_X exists for all time can be seen as follows: Suppose γ_X exists on $[-\varepsilon, \varepsilon]$. For $t_0 \in [-\varepsilon, \varepsilon]$, define $\phi : [-\varepsilon, \varepsilon] \to G$ given by $\phi(t) = \gamma_X(t_0)\gamma_X(t)$. Since

$$\dot\phi(t) = \mathrm{d}L_{\gamma_X(t_0)}(\gamma_X(t))(\dot\gamma_X(t)),$$

it follows from the left-invariance of X^L that ϕ again satisfies (4.1.2), with $\phi(0) = \gamma(t_0)$. Repeating this procedure, we can construct a solution for all time.

Using such curves, we make the following definition.

Definition 4.1.1. *The* exponential map *is a map* $\exp : \mathfrak{g} \to G$ *given by*

$$\exp(X) = \gamma_X(1),$$

where γ_X *is the curve satisfying* (4.1.2) *with* $\gamma_X(0) = e$.

For every $X \in \mathfrak{g}$ we have

$$\mathrm{d}\exp(0)(X) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \exp(tX) = X$$

so that $\mathrm{d}\exp(0) = I$. Therefore, by the inverse function theorem, there exists an $r > 0$ such that exp is a homeomorphism from $B(0,r)$ onto its image. The inverse of the exponential map is referred to as the *logarithm map*, and is denoted by log. We have the following proposition.

Proposition 4.1.2. *For every $r > 0$ such that* exp *is a homeomorphism on $\overline{B(0,r)}$, there exists an $\varepsilon > 0$ such that* log *is well-defined on $\overline{B(e,\varepsilon)}$ and for all $g \in \overline{B(e,\varepsilon)}$ we have $|\log(g)| \leqslant r$.*

Proof. Because exp is a homeomorphism on $\overline{B(0,r)}$, it is an open map, and hence $\exp(\overline{B(0,r)})$ contains some open ball $B(e,\varepsilon)$. Because $\exp(\overline{B(0,r)})$ is closed, it must be that $\overline{B(e,\varepsilon)} \subset \exp(\overline{B(0,r)})$ so that log is well defined on $\overline{B(e,\varepsilon)}$ and $\log(\overline{B(e,\varepsilon)}) \subset \overline{B(0,r)}$ as desired. □

4.1.2. Riemannian metric

For reasons that will become clear later, we equip $\mathfrak{g}$ with an inner product $\langle\cdot,\cdot\rangle$. This induces on $\mathfrak{g}$ a norm $|\cdot|$ given by $|X| = \sqrt{\langle X, X\rangle}$. Because $\mathfrak{g}$ is finite-dimensional, all norms are equivalent, and hence, our results will not depend on the choice of inner product. For more details regarding the Riemmanian structure of Lie groups, we refer to e.g. [73].

The inner product on $\mathfrak{g}$ may be extended to a Riemannian metric on G. For this, we use the fact that T_gM may be identified with $\mathfrak{g}$ via the isomorphism $\mathrm{d}L_g(e)$. With this identification in mind, we define an inner product on T_gG by setting

$$\langle X, Y\rangle_g := \langle \mathrm{d}L_g(e)^{-1}X, \mathrm{d}L_g(e)^{-1}Y\rangle$$

for $X, Y \in T_gM$. The assumption that the group operations are smooth implies that this defines a Riemannian metric on G. By construction this Riemannian metric is left-invariant, i.e., for all $g, h \in G$ and for all $X, Y \in T_gG$ we have

$$\langle \mathrm{d}L_h(g)X, \mathrm{d}L_h(g)Y\rangle_{hg} = \langle X, Y\rangle_g.$$

This also shows that $\mathrm{d}L_h(g) : T_gG \to T_hG$ is an isometry for all $g, h \in G$. In particular, the identification $\mathrm{d}L_g(e) : \mathfrak{g} \to T_gG$ of T_gG with the Lie algebra $\mathfrak{g}$ is an isometry. Therefore, if we consider $X \in T_gG$ as element in $\mathfrak{g}$, its norm can also be taken as element of $\mathfrak{g}$.

To the Riemannian metric we can associate a Riemannian distance $d : G \times G \to \mathbb{R}$ given by the usual formula (see Section 2.2):

$$d(g,h) = \inf\left\{ \int_0^1 |\dot\gamma(t)|\,\mathrm{d}t \,\middle|\, \gamma : [0,1] \to G \text{ piecewise smooth}, \gamma(0) = g, \gamma(1) = h \right\}.$$

Because the Riemannian metric is left-invariant, it follows that for all $f, g, h \in G$ we have

$$d(g, h) = d(fg, fh).$$

This shows that the distance between elements of G is preserved under left-multiplication.

4.2. Random walks in Lie groups

In this section we introduce random walks in a general (connected) Lie group G. We will explain how these random walks relate to geodesic random walks defined in Chapter 3, and argue that the two notions only coincide when we equip the Lie group with a bi-invariant metric.

We start by defining a random walk in a Lie group G. To this end, let $\{X_n\}_{n\geqslant 1}$ be a sequence of $\mathfrak{g}$-valued random variables. We define the random walk $\mathcal{S}_n \in G$ with increments $\{X_n\}_{n\geqslant 1}$ by

$$\mathcal{S}_n = \exp(X_1)\exp(X_2)\cdots\exp(X_n). \tag{4.2.1}$$

Furthermore, we define the rescaled random walk by

$$\sigma_n^n = \exp\left(\frac{1}{n}X_1\right)\exp\left(\frac{1}{n}X_2\right)\cdots\exp\left(\frac{1}{n}X_n\right). \tag{4.2.2}$$

4.2.1. Relation to geodesic random walks

In order to relate the random walk defined in (4.2.1) to the concept of a geodesic random walk in Chapter 3, we need to argue how one-parameter subgroups of the form $\gamma(t) = g\exp(tX)$ can be interpreted as geodesics. To this end, we need some additional theory from Lie groups.

Definition 4.2.1. *Let ∇ be a connection on a Lie group G. ∇ is said to be* left-invariant *if for any two left-invariant vector fields X^L and Y^L (see* (4.1.1)*) with $X, Y \in \mathfrak{g}$ we have that $\nabla_{X^L}Y^L$ is also left-invariant.*

Among the left-invariant connections, there are special connections for which the one-parameter subgroups form geodesics.

Definition 4.2.2. *A* Cartan connection *on a Lie group G is a left-invariant connection satisfying the property that the subgroup $\gamma(t) = \exp(tX)$ is a geodesic for every $X \in \mathfrak{g}$.*

One question that arises, is whether such connections always exist. This is indeed the case, as the following result from [73] states.

Proposition 4.2.3. *For any Lie group G there exists a unique symmetric Cartan connection ∇ given by*

$$\nabla_{X^L}Y^L = \frac{1}{2}[X, Y]^L \tag{4.2.3}$$

for any $X, Y \in \mathfrak{g}$.

By definition, a random walk on G is a geodesic random walk when we equip G with a Cartan connection. Unfortunately, the Cartan connection given in Proposition 4.2.3 is in general not compatible with the Riemannian metric. It can be shown that the connection in (4.2.3) is compatible with the metric if and only if the metric is bi-invariant, see e.g. [70, 80, 73]. In this case, the exponential map $\exp : \mathfrak{g} \to G$ coincides with the Riemannian exponential map.

We will now connect the result of Theorem 4.3.1 to the result in Theorem 3.3.1. For this, let μ be a probability measure $\mathfrak{g}$. We need to find a collection of probability measures $\{\mu_g\}_{g\in G}$ with μ_g a measure on T_gM, such that if we identify an increment $X \in \mathfrak{g}$ as element of T_gM, then X has distribution μ_g. Because we identify the tangent space T_gM with the Lie algebra $\mathfrak{g}$ via the map $\mathrm{d}L_g(e)^{-1}$, the measure μ_g is given by

$$\mu_g = \mu \circ \mathrm{d}L_g(e)^{-1}. \tag{4.2.4}$$

From this definition, it immediately follows that the collection $\{\mu_g\}_{g\in G}$ is left-invariant, in the sense that

$$\mu_{gh} = \mu_h \circ \mathrm{d}L_g(h)^{-1},$$

for all $g, h \in G$.

In order for the random walk to have identically distributed increments in the sense of Definition 3.2.7, the collection of measures $\{\mu_g\}_{g\in G}$ has to be invariant under parallel transport. In the case of a bi-invariant metric, it can be shown that parallel transport along a geodesic of the form $\gamma(t) = \exp(tX)$ is given by (see e.g. [70, 55])

$$\tau_{\gamma(0)\gamma(t);\gamma}Y = \mathrm{d}L_{\exp(tX/2)}(\exp(tX/2))(\mathrm{d}R_{\exp(tX/2)}(e)(Y)).$$

Here, $R_g : G \to G$ denotes right-multiplication, i.e., $R_g h = hg$. Therefore, in order for the collection of measures $\{\mu_g\}_{g\in G}$ to be invariant under parallel transport, one also needs that the collection is right-invariant, meaning that

$$\mu_{hg} = \mu_h \circ \mathrm{d}R_g(h)^{-1},$$

for all $g, h \in G$. Since the metric is bi-invariant, a sufficient condition for this is that μ only depends on the norm of $X \in \mathfrak{g}$. Another example is when the Lie algebra is abelian, in which case left- and right-multiplication coincide. However, in that case the random walk σ_n^n reduces to

$$\sigma_n^n = \exp\left(\frac{1}{n}\sum_{i=1}^n X_i\right),$$

in which case Theorem 4.3.1 immediately follows from Cramér's theorem (Theorem 2.1.10), together with the contraction principle (Theorem 2.1.6).

Collecting everything, we see that the random walk in (4.2.1) only coincides with the notion of a geodesic random walk with independent, identically distributed increments as in Chapter 3, when the metric on G is bi-invariant, and the collection $\{\mu_g\}_{g\in G}$ defined in (4.2.4) is right-invariant. In that case, although the proof might be somewhat simpler due to the extra group structure, the results in this chapter do not add anything over the results in Chapter 3. The novelty of the work in this chapter is in the case when either of those conditions is not satisfied. Most importantly, the result is new when no bi-invariant metric exists, in which case the random walk in (4.2.1) cannot be interpreted as a geodesic random walk with respect to some Riemannian metric.

4.3. Main theorem, sketch of the proof and an example

With all the notation fixed, we are ready to state the main theorem that we are going to prove. Because the proof consists of a number of steps, we also provide a sketch of the proof so that the main steps are clear. The precise proof will be given in Section 4.5. We conclude this section by showing how the main theorem can be applied if we consider the Lie group of stochastic matrices.

4.3.1. Statement of the main theorem

Let G be a Lie group with Lie algebra $\mathfrak{g}$ equipped with an inner product $\langle\cdot,\cdot\rangle$. Let $\{X_n\}_{n\geqslant 1}$ be a sequence of independent, identically distributed random variables in the Lie algebra $\mathfrak{g}$ and denote by σ_n^n the rescaled random walk as in (4.2.2). We are going to prove that under some assumptions on the increments $\{X_n\}_{n\geqslant 1}$, the sequence $\{\sigma_n^n\}_{n\geqslant 1}$ satisfies a large deviation principle in G.

Along with the large deviation principle for $\{\sigma_n^n\}_{n\geqslant 1}$, we need to identify the corresponding rate function. If G admits a bi-invariant metric, it follows from Theorem 3.3.1 that the rate function is given by

$$I(g) = \inf\{\Lambda^*(X) \,|\, \exp(X) = g\}.$$

Here, $\Lambda(\lambda)$ is the log moment generating function of an increment, given by

$$\Lambda(\lambda) := \log \mathbb{E}\left(e^{\langle \lambda, X_1\rangle}\right),$$

while Λ^* denotes its Legendre transform, defined as

$$\Lambda^*(X) := \sup_{\lambda\in\mathfrak{g}} \{\langle \lambda, X\rangle - \Lambda(\lambda)\}.$$

Obtaining this form of the rate function relies on the fact that if we minimize $\int_0^1 \Lambda^*(\dot\gamma(t))\,\mathrm{d}t$ over curves with fixed endpoints, the minimum is attained by a geodesic. However, if G does not admit a bi-invariant metric, curves of the form $\gamma(t) = \exp(tX)$ are no longer necessarily geodesics (when taking the exponential map in the terminology of Lie groups and Lie algebra's). As a consequence, we can do no better than the expression

$$I(g) = \inf\left\{\int_0^1 \Lambda^*(\dot\gamma(t))\,\mathrm{d}t \,\middle|\, \gamma \in AC([0,1];G), \gamma(0) = e, \gamma(1) = g\right\}.$$

We now collect everything and give the statement of the theorem.

Theorem 4.3.1 (Cramér's theorem for Lie groups)**.** *Let G be a Lie group and $\mathfrak{g}$ its associated Lie algebra, equipped with an inner product $\langle\cdot,\cdot\rangle$. Let $\{X_n\}_{n\geqslant 1}$ be a sequence of random variables in $\mathfrak{g}$ and denote by σ_n^n the associated rescaled random walk as in* (4.2.2)*. Assume the increments $\{X_n\}_{n\geqslant 1}$ are independent, identically distributed and bounded. Assume furthermore that the log moment generating function $\Lambda(\lambda) = \log \mathbb{E}(e^{\langle \lambda, X_1\rangle})$ is everywhere finite. Then the sequence $\{\sigma_n^n\}_{n\geqslant 0}$ satisfies the large deviation principle in G with good rate function*

$$I_G(g) = \inf\left\{\int_0^1 \Lambda^*(\dot\gamma(t))\,\mathrm{d}t \,\middle|\, \gamma \in AC([0,1];G), \gamma(0) = e, \gamma(1) = g\right\}. \tag{4.3.1}$$

Some remarks on optimal trajectories for the rate function

In this section, we make some remarks about the optimal trajectories for the rate function in (4.3.1). Comparing to the setting of geodesic random walks in Chapter 3, one might expect that optimal trajectories are of the form $\gamma(t) = \exp(tX)$ for some $X \in \mathfrak{g}$. It turns out this is true in specific cases, but not in general.

Indeed, recall that we equipped the Lie group G with a left-invariant Riemannian metric, to which we may associate a Levi-Civita connection. Now suppose that $\Lambda^*(X)$ is given as $\Lambda^*(X) = F(|X|)$ for some convex function F. It then follows from Proposition 3.4.11 that the optimal trajectories are geodesics with respect to the Levi-Civita connection. These geodesics are of the form $\gamma(t) = \exp(tX)$, precisely when the Levi-Civita connection is a Cartan connection. As discussed in Section 4.2, the latter is true precisely when the Riemannian metric is bi-invariant. Therefore, if $\Lambda^*(X)$ only depends on the norm of X, then optimal trajectories for the rate function in (4.3.1) are of the form $\gamma(t) = \exp(tX)$ if the metric is bi-invariant. Note that in this case the random walk in (4.2.2) is a geodesic random walk as treated in Chapter 3.

However, if the Riemannian metric is only left-invariant, the Lie group structure still allows us to give a simpler expression for the rate function, given that $\Lambda^*(X)$ is a function of the norm of X. Indeed, geodesics for a left-invariant metric on a Lie group satisfy a special equation, namely the Euler-Arnold equation, see e.g. [7, 17].

4.3.2. Sketch of the proof of Theorem 4.3.1

Since the proof of Theorem 4.3.1 is rather long, we first provide a sketch. The detailed proof is given in Section 4.5. The proof is inspired by the proof of Theorem 3.3.1, and therefore, we will follow similar steps as explained in Section 3.3. Like in the proof of Cramér's theorem in Euclidean space (Theorem 2.1.10), we prove the upper and lower bound of the large deviation principle separately.

By Cramér's theorem for vector spaces, the sequence $\{\frac{1}{n}\sum_{i=1}^n X_i\}_{n\geqslant 1}$ of empirical averages satisfies the large deviation principle in $\mathfrak{g}$ with good rate function $I(X) = \Lambda^*(X)$. This, together with the contraction principle (Theorem 2.1.6), implies that the sequence $\{\Sigma_n\}_{n\geqslant 1}$ given by

$$\Sigma_n = \exp\left(\frac{1}{n}\sum_{i=1}^n X_i\right)$$

satisfies the large deviation principle in G with good rate function

$$I_G(g) = \inf\{\Lambda^*(X) |\exp(X) = g\}.$$

Unfortunately, the Baker-Campbell-Hausdorff formula shows us that in general, Σ_n and σ_n^n do not coincide. More precisely, given that the random walk stays close enough to the identity e so that logarithms are well-defined, the integral version of the Baker-Campbell-Hausdorff formula (see Theorem 4.4.1) gives us that

$$\log(\sigma_n^n) = \frac{1}{n}\sum_{i=1}^n \left(\int_0^1 \frac{\mathrm{ad}_{\log(\sigma_{i-1}^n)}}{1-e^{-\mathrm{ad}_{\log(\sigma_{i-1}^n)}}}\right) X_i \tag{4.3.2}$$

Here, the operator ad is as defined in (4.4.1) and σ_i^n is the point of the random walk after i steps, i.e.,

$$\sigma_i^n = \exp\left(\frac{1}{n}X_1\right)\cdots\exp\left(\frac{1}{n}X_i\right).$$

However, we would like to understand the difference between $\log(\sigma_n^n)$ and $\frac{1}{n}\sum_{i=1}^n X_i$. For this, we compare $\frac{1}{n}\sum_{i=1}^n X_i$ to the expression found in (4.3.2) for $\log(\sigma_n^n)$. We prove (see Proposition 4.4.2) that there exists constants $C_{|\log(\sigma_{i-1}^n)|}$ such that

$$\left|\left(\int_0^1 \frac{\mathrm{ad}_{\log(\sigma_{i-1}^n)}}{1-e^{-\mathrm{ad}_{\log(\sigma_{i-1}^n)}}}\right) X_i - X_i\right| \leqslant C_{|\log(\sigma_{i-1}^n)|}|X_i|,$$

where C_α is a constant, decreasing in α and such that $\lim_{\alpha\to 0} C_\alpha = 0$.
Using the triangle inequality and the smoothness of the logarithm, one can show that $|\log(\sigma_i^n)| \lesssim \frac{i}{n}B$, where B is the uniform bound on the increments. As a consequence, $C_{|\log(\sigma_{i-1}^n)|} \leqslant C_B$ for all $i = 1, \ldots, n$. If we now collect everything, we find

$$\left|\log(\sigma_n^n) - \frac{1}{n}\sum_{i=1}^n X_i\right| \leqslant C_B B. \tag{4.3.3}$$

Because B is fixed, this upper bound unfortunately does not show us that $\log(\sigma_n^n)$ and $\frac{1}{n}\sum_{i=1}^n X_i$ will get arbitrarily close if n tends to infinity. The key will be to decrease the constant $C_B B$ in an appropriate way.

To do this, we split the random walk into finitely many, say m, pieces, each consisting of $\lfloor m^{-1}n\rfloor$ increments. It turns out that this also takes care of the problem that the logarithms we use are not necessarily well-defined. More precisely, for $m \in \mathbb{N}$ we define the indices $n_j = j\lfloor m^{-1}n\rfloor$ for $j = 0, \dots, m-1$ and set $n_m = n$. We can prove (see (4.5.2) and (4.5.3)) that if B is the uniform bound on the increments, then for every $j = 1, \dots, m$ and $i = 1, \dots, n_j - n_{j-1}$ we have

$$d(e, (\sigma^n_{n_{j-1}})^{-1}\sigma^n_{n_{j-1}+i}) = d(\sigma^n_{n_{j-1}}, \sigma^n_{n_{j-1}+i}) \leqslant \frac{i}{n}B \leqslant \frac{1}{m}B.$$

Here, the first equality follows from the left-invariance of the metric d. This shows that if $m \in \mathbb{N}$ is large enough, then $\log((\sigma^n_{n_{j-1}})^{-1}\sigma^n_{n_{j-1}+i})$ is well-defined for every $j = 1, \dots, m$ and every $i = 1, \dots, n_j - n_{j-1}$. In particular, one may show in a similar spirit as (4.3.3), that

$$\left|\log((\sigma^n_{n_{j-1}})^{-1}\sigma^n_{n_j}) - \frac{1}{n}\sum_{i=1}^{n_j - n_{j-1}} X_{n_{j-1}+i}\right| \leqslant C_{m^{-1}B}\frac{B}{m}. \tag{4.3.4}$$

Now let us define $Y^{n,m,j}_{\lfloor m^{-1}n\rfloor} = \log((\sigma^n_{n_{j-1}})^{-1}\sigma^n_{n_j}) \in \mathfrak{g}$ for $j = 1, \dots, m$. By the above construction, we have

$$\sigma^n_n = \exp\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}\right)\cdots\exp\left(Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) =: \Psi_m\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right),$$

where $\Psi_m : \mathfrak{g}^m \to G$ is the continuous function given by

$$\Psi_m(x_1, \dots, x_m) = \exp(x_1)\cdots\exp(x_m).$$

Using this, we can prove the upper and lower bound for the large deviation principle for $\{\sigma^n_n\}_{n\geqslant 1}$, which we explain in the upcoming two sections.

Upper bound of the large deviation principle for $\{\sigma^n_n\}_{n\geqslant 1}$.
In this section we sketch the proof of the upper bound of the large deviation principle for $\{\sigma^n_n\}_{n\geqslant 1}$. For $F \subset G$ closed and every $m \in \mathbb{N}$ large enough, we have that $\Psi_m^{-1}F \subset \mathfrak{g}^m$ is closed and

$$\mathbb{P}(\sigma^n_n \in F) = \mathbb{P}\left(\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in \Psi_m^{-1}F\right). \tag{4.3.5}$$

Because $\mathfrak{g}^m$ is a vector space, we can use a similar argument as in the proof of Cramér's theorem for the Euclidean setting (see e.g. [29, 56]), to obtain for $\Gamma \subset \mathfrak{g}^m$ compact that

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in \Gamma\right)\\ \leqslant -\inf_{x\in\Gamma}\sup_{\lambda\in\mathfrak{g}^m}\left\{\langle\lambda, x\rangle - \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left(e^{n\langle\lambda,(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor},\dots,Y^{n,m,m}_{\lfloor m^{-1}n\rfloor})\rangle}\right)\right\}.\end{aligned}$$

By exponential tightness, this bound also holds for all $\tilde{F} \subset \mathfrak{g}^m$ closed.

Now one can use (4.3.4) to prove that

$$\mathbb{E}\left(e^{n\langle\lambda,(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor},\ldots,Y^{n,m,m}_{\lfloor m^{-1}n\rfloor})\rangle}\right) \leqslant e^{C_{m^{-1}B}|\lambda|Bm^{-1}}\mathbb{E}\left(e^{n\langle\lambda,(Z^{m,n}_1,\ldots,Z^{n,m}_m)\rangle}\right),$$

where

$$Z^{m,n}_j = \frac{1}{n}\sum_{i=1}^{n_j-n_{j-1}} X_{n_{j-1}+i}.$$

Because the increments $\{X_n\}_{n\geqslant 1}$ are independent, identically distributed, the random variables $Z^{n,m}_1,\ldots,Z^{n,m}_m$ are also independent and identically distributed with $\mathbb{E}(e^{n\langle\lambda,Z^{n,m}_1\rangle}) = \mathbb{E}(e^{\langle\lambda,X_1\rangle})^{\lfloor m^{-1}n\rfloor}$. Therefore, we find that

$$\mathbb{E}\left(e^{\langle\lambda,(Z^{m,n}_1,\ldots,Z^{n,m}_m)\rangle}\right) = M(\lambda_1)^{\lfloor m^{-1}n\rfloor}\cdots M(\lambda_m)^{\lfloor m^{-1}n\rfloor}.$$

Collecting everything, we find that

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\sigma^n_n\in F)\\
&\quad\leqslant -\inf_{x\in\Psi_m^{-1}F}\sup_{\lambda\in\mathfrak{g}^m}\left\{\langle\lambda,x\rangle-\frac{1}{m}\sum_{i=1}^m\Lambda(\lambda_i)-C_{m^{-1}B}|\lambda|B\frac{1}{m}\right\}\\
&\quad= -\inf_{x\in\Psi_m^{-1}F}\frac{1}{m}\sum_{j=1}^m\sup_{\lambda\in\mathfrak{g}}\left\{\langle\lambda,mx_j\rangle-\Lambda(\lambda)-C_{m^{-1}B}|\lambda|B\right\}.
\end{aligned}$$

Finally, by letting m tend to infinity, apart from some technical difficulties, one obtains

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\sigma^n_n\in F)\leqslant -\inf_{g\in G}I_G(g),$$

as desired.

Lower bound of the large deviation principle for $\{\sigma^n_n\}_{n\geqslant 1}$.

To prove the lower bound of the large deviation principle for $\{\sigma^n_n\}_{n\geqslant 1}$, we first observe that it is sufficient to show for every $U\subset G$ open and every $g\in U$ that

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\sigma^n_n\in U)\geqslant -\int_0^1\Lambda^*(\dot\gamma(t))\,\mathrm{d}t$$

for all $\gamma\in AC([0,1];G)$ with $\gamma(0)=e$ and $\gamma(1)=g$.

To do this, we fix $\gamma\in AC([0,1];G)$ with $\gamma(0)=e$ and $\gamma(1)=g$ and define for $m\in\mathbb{N}$ the vectors

$$y^m_i := \log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma\left(\frac{i}{m}\right)\right)\in\mathfrak{g}.$$

Note that $\Psi_m((y^m_1,\ldots,y^m_m)) = g$, where $\Psi_m:\mathfrak{g}^m\to G$ is as in (4.3.5). In order to continue, we need to know a bit more about the continuity properties of Ψ_m. More precisely, we will prove (see Proposition 4.5.7) that there exists a constant $C > 0$ such that for $\varepsilon > 0$ and $m \in \mathbb{N}$ large enough, we have that if

$$(x_1, \dots, x_m) \in \prod_{i=1}^{m} B(y_i^m, (Cm)^{-1}\varepsilon),$$

then

$$\Psi_m((x_1, \dots, x_m)) \in B(\Psi_m((y_1^m, \dots, y_m^m)), \varepsilon) = B(g, \varepsilon).$$

Now note that in general, contrary to the Euclidean case, we have

$$\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma\left(\frac{i}{m}\right)\right) \neq \int_{\frac{i-1}{m}}^{\frac{i}{m}} \dot{\gamma}(t)\,\mathrm{d}t.$$

We will show that under the condition that $\dot{\gamma}$ is bounded (see Proposition 4.5.8), we have

$$\left|\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma\left(\frac{i}{m}\right)\right) - \int_{\frac{i-1}{m}}^{\frac{i}{m}} \dot{\gamma}(t)\,\mathrm{d}t\right| \leqslant L_m\frac{1}{m},$$

where $\lim_{m\to\infty} L_m = 0$. In particular, if we set

$$\tilde{y}_i^m := \int_{\frac{i-1}{m}}^{\frac{i}{m}} \dot{\gamma}(t)\,\mathrm{d}t,$$

then for m large enough we have $B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon) \subset B(y_i^m, (Cm)^{-1}\varepsilon)$. We conclude that if

$$(x_1, \dots, x_m) \in \prod_{i=1}^{m} B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon),$$

then $\Psi_m((x_1, \dots, x_m)) \in B(g, \varepsilon)$.

Because U is open, there exists an $\varepsilon > 0$ such that $B(\varepsilon, g) \subset U$. Using the above continuity property, we find that

$$\begin{aligned}&\mathbb{P}(\sigma_n^n \in U)\\ &\geqslant \mathbb{P}\left(\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in B(\tilde{y}_1^m, (2Cm)^{-1}\varepsilon) \times \dots \times B(\tilde{y}_m^m, (2Cm)^{-1}\varepsilon)\right).\end{aligned}$$

Now using (4.3.4) and the fact that $\lim_{m\to\infty} C_{m^{-1}B} = 0$, we have for m large enough that

$$\left|Y^{n,m,j}_{\lfloor m^{-1}n\rfloor} - \frac{1}{n}\sum_{i=1}^{n_j - n_{j-1}} X_{n_{j-1}+i}\right| \leqslant (2Cm)^{-1}\frac{\varepsilon}{2}.$$

But then we find that

$$\mathbb{P}\left(\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in B(\tilde{y}_1^m, (2Cm)^{-1}\varepsilon) \times \dots \times B(\tilde{y}_m^m, (2Cm)^{-1}\varepsilon)\right)$$

$$\geqslant \mathbb{P}\left(\left(\frac{1}{n}\sum_{i=1}^{n_1} X_i, \ldots, \frac{1}{n}\sum_{i=1}^{n_m - n_{m-1}} X_{n_{m-1}+i}\right) \in \prod_{i=1}^{m} B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon/2)\right)$$
$$= \prod_{j=1}^{m} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n_j - n_{j-1}} X_{n_{j-1}+i} \in B(\tilde{y}_j^m, (2Cm)^{-1}\varepsilon/2)\right).$$

By Cramér's theorem for random walks in Euclidean space (Theorem 2.1.10), we find

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n_j - n_{j-1}} X_{n_{j-1}+i} \in B(y_j, (Cm)^{-1}\varepsilon)\right) \geqslant -\frac{1}{m}\Lambda^*(my_j^m).$$

Therefore, if we collect everything, we obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\sigma_n^n \in U) \geqslant -\frac{1}{m}\sum_{j=1}^{m} \Lambda^*(m\tilde{y}_j^m).$$

Finally, the convexity of Λ^* together with Jensen's inequality implies that

$$\frac{1}{m}\sum_{j=1}^{m} \Lambda^*(m\tilde{y}_j^m) \leqslant \sum_{i=1}^{m} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \Lambda^*(\dot{\gamma}(t))\,\mathrm{d}t = \int_0^1 \Lambda^*(\dot{\gamma}(t))\,\mathrm{d}t.$$

From this, we conclude

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\sigma_n^n \in U) \geqslant -\int_0^1 \Lambda^*(\dot{\gamma}(t))\,\mathrm{d}t,$$

which finishes the proof.

4.3.3. Example: Products of transition matrices

We conclude this section by discussing an example. In this example, we aim to study the limiting behaviour of products of transition matrices on a finite dimensional state space. For this, we use the stochastic group and its Lie algebra, see e.g. [47, 79]. For theory regarding matrix Lie groups, see e.g. [51].

We define the set of transition matrices $\mathcal{T}(d,\mathbb{R})$ on d states by

$$\mathcal{T}(d,\mathbb{R}) = \{P \in M(d,\mathbb{R}) | P\mathbf{1} = \mathbf{1}, P_{ij} \geqslant 0 \text{ for } 1 \leqslant i,j \leqslant d\}.$$

Here, $M(d,\mathbb{R})$ denotes the set of all $d \times d$-matrices, and $\mathbf{1}$ is the vector of all ones. Because we will be working with groups, we need inverses to be well-defined. We therefore consider the subset $\mathcal{S}_+(d,\mathbb{R})$ of invertible matrices in $\mathcal{T}(d,\mathbb{R})$, i.e.

$$\mathcal{S}_+(d,\mathbb{R}) = \{P \in \mathcal{T}(d,\mathbb{R}) | \det(P) \neq 0\}.$$

Note that $\mathcal{S}_+(d,\mathbb{R})$ is closed under matrix multiplication. Indeed, if P and Q have non-negative entries, then so does PQ. Furthermore, if $P\mathbf{1} = \mathbf{1}$ and $Q\mathbf{1} = \mathbf{1}$ then

$PQ\mathbf{1} = P\mathbf{1} = \mathbf{1}$. Finally, if P and Q are invertible, then so is PQ. However, inverses of elements in $\mathcal{S}_+(d,\mathbb{R})$ need not have only non-negative entries. It turns out that the smallest group containing $\mathcal{S}_+(d,\mathbb{R})$ is given by

$$\mathcal{S}(d,\mathbb{R}) = \{P \in M(d,\mathbb{R}) | \det(P) \neq 0, P\mathbf{1} = \mathbf{1}\}.$$

This group is called the *stochastic group*. It is in fact a Lie group. Because we are dealing with matrix Lie groups, this follows from the observation that if $P_n \to P$ element wise, and $P_n\mathbf{1} = \mathbf{1}$ for all n, then also $P\mathbf{1} = \mathbf{1}$.

The Lie algebra associated to $\mathcal{S}(d,\mathbb{R})$ is given by

$$\mathfrak{s}(d,\mathbb{R}) = \{A \in M(d,\mathbb{R}) | A\mathbf{1} = 0\}.$$

Indeed, if $A \in M(d,\mathbb{R})$ is such that $A\mathbf{1} = 0$, then

$$\exp(tA)\mathbf{1} = \mathbf{1} + \left(\sum_{n=1}^{\infty} \frac{t^n A^{n-1}}{n!}\right) A\mathbf{1} = \mathbf{1}.$$

This shows that $\exp(tA) \in \mathcal{S}(d,\mathbb{R})$ for all t, implying that

$$\{A \in M(d,\mathbb{R}) | A\mathbf{1} = 0\} \subset \mathfrak{s}(d,\mathbb{R}).$$

Conversely, if $\exp(tA)\mathbf{1} = \mathbf{1}$ for all $t \in \mathbb{R}$, then

$$A\mathbf{1} = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \exp(tA)\mathbf{1} = 0,$$

so that

$$\mathfrak{s}(d,\mathbb{R}) \subset \{A \in M(d,\mathbb{R}) | A\mathbf{1} = 0\}.$$

In order to consider random walks in the Lie group $\mathcal{S}(d,\mathbb{R})$ which only use invertible transition matrices, i.e., elements from $\mathcal{S}_+(d,\mathbb{R})$, we need to find a subset of $\mathfrak{s}(d,\mathbb{R})$ which is mapped by the exponential map into $\mathcal{S}_+(d,\mathbb{R})$. To this end, consider the set

$$\mathfrak{s}_+(d,\mathbb{R}) = \{A \in \mathfrak{s}(d,\mathbb{R}) | A_{ij} \geqslant 0 \text{ whenever } i \neq j\}.$$

We will prove that for all $A \in \mathfrak{s}_+(d,\mathbb{R})$ we have $\exp(A) \in \mathcal{S}_+(d,\mathbb{R})$. It suffices to prove that $\exp(A)$ has nonnegative entries. To show this, we fix $k = \max_{i=1}^d |A_{ii}|$. Then the matrix $B = A + kI$ has nonnegative entries, from which it follows, using the Taylor series expression, that $\exp(B)$ has nonnegative entries. Because A and I commute, we have

$$\exp(B) = \exp(A)\exp(kI) = e^k \exp(A),$$

so that $\exp(A) = e^{-k}\exp(B)$. The latter now has nonnegative entries because $e^{-k} > 0$ and $\exp(B)$ has nonnegative entries.

If we now take a measure μ on $\mathfrak{s}(d, \mathbb{R})$ supported in $\mathfrak{s}_+(d, \mathbb{R})$, then the random walk $\mathcal{S}_n$ associated to an independent, identically distributed sequence $\{X_n\}_{n\geqslant 1}$ will remain in $\mathcal{S}_+(\mathbb{R}, d)$. This random walk may be thought of as the (random) n-step transition matrix of a Markov process with state space $\Omega = \{1, \ldots, d\}$.

From an increment $A \in \mathfrak{s}_+(d, \mathbb{R})$ of such a random walk, we can deduce some qualitative behaviour of the random walk. Indeed, for a state $i \in \{1, \ldots, d\}$ we have that the larger $|A_{ii}|$, the more mass remains at site i after that iteration. The remainder of the mass at state i is then distributed over the states $j \neq i$ according to the relative size of the A_{ij}.

A concrete example

To get a better understanding of these random walks in $\mathcal{S}(d, \mathbb{R})$ and their limiting behaviour, we do the calculations for a specific example. For this, we take $d = 2$ and $\alpha, \beta > 0$. Consider the matrices

$$A = \begin{pmatrix} -\alpha & \alpha \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ \beta & -\beta \end{pmatrix}.$$

Let $\{X_n\}_{n\geqslant 1}$ be a sequence of independent, identically distributed random variables with

$$\mathbb{P}(X_1 = A) = \mathbb{P}(X_1 = B) = \frac{1}{2}.$$

One may compute the exponentials of the matrices A and B to find

$$\exp\left(\frac{1}{n}A\right) = \begin{pmatrix} e^{-\frac{1}{n}\alpha} & 1 - e^{-\frac{1}{n}\alpha} \\ 0 & 1 \end{pmatrix}, \qquad \exp\left(\frac{1}{n}B\right) = \begin{pmatrix} 1 & 0 \\ 1 - e^{-\frac{1}{n}\beta} & e^{-\frac{1}{n}\beta} \end{pmatrix}.$$

Intuitively, this process chooses one of the states uniformly at random and then distributes the mass at that state over the two states according to some parameter. Additionally, one sees that if n tends to infinity, the mass that is passed between states becomes exponentially small.

Now consider the rescaled random walk

$$\sigma_n^n = \exp\left(\frac{1}{n}X_1\right) \cdots \exp\left(\frac{1}{n}X_n\right).$$

By Theorem 4.3.1, the sequence $\{\sigma_n^n\}_{n\geqslant 1}$ satisfies in $\mathcal{S}(2, \mathbb{R})$ the large deviation principle. In order to obtain an explicit expression for the rate function, we need to equip $\mathfrak{s}(2, \mathbb{R})$ with an inner product. For this, we will use the Frobenius inner product given by

$$\langle A, B \rangle = \mathrm{Tr}(A^T B) = \sum_{i,j=1}^{2} A_{ij} B_{ij}.$$

With this inner product, the log moment generating function $\Lambda : \mathfrak{s}(2, \mathbb{R}) \to \mathbb{R}$ of X_1 is given by

$$\Lambda\left(\begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}\right) = \log\left(\frac{1}{2}e^{2\alpha\lambda_1} + \frac{1}{2}e^{2\beta\lambda_2}\right).$$

Let us compute $\Lambda^* : \mathfrak{s}(2,\mathbb{R}) \to \mathbb{R}$, i.e., we want to compute

$$\begin{aligned}\Lambda^*\left(\begin{pmatrix} -x_1 & x_1 \\ x_2 & -x_2 \end{pmatrix}\right) &= \sup_{\lambda\in\mathfrak{s}(2,\mathbb{R})} \langle \lambda, x\rangle - \Lambda(\lambda) \\ &= \sup_{\lambda_1,\lambda_2\in\mathbb{R}} 2\lambda_1 x_1 + 2\lambda_2 x_2 - \log\left(\frac{1}{2}e^{2\alpha\lambda_1} + \frac{1}{2}e^{2\beta\lambda_2}\right).\end{aligned}$$

Here we used that every $\lambda \in \mathfrak{s}(2,\mathbb{R})$ may be characterized by two elements $\lambda_1, \lambda_2 \in \mathbb{R}$. By taking $\lambda_2 = 0$ and letting $|\lambda_1|$ tend to infinity, we see that Λ^* is infinite whenever $x_1 \notin [0,\alpha]$. In a similar way one can show that Λ^* is infinite if $x_2 \notin [0,\beta]$.
Next, we show that Λ^* is also infinite if $\alpha x_2 + \beta x_1 \neq \alpha\beta$. To see this, take λ_1, λ_2 such that $\alpha\lambda_1 - \beta\lambda_2 = \alpha\beta$. Writing everything in terms of λ_2, we find that

$$\begin{aligned}\Lambda^*\left(\begin{pmatrix} -x_1 & x_1 \\ x_2 & -x_2 \end{pmatrix}\right) &\geqslant 2\lambda_1 x_1 + 2\lambda_2 x_2 - \log\left(\frac{1}{2}e^{2\alpha\lambda_1} + \frac{1}{2}e^{2\beta\lambda_2}\right) \\ &= 2x_1\left(\beta + \frac{\beta}{\alpha}\lambda_2\right) + 2\lambda_2 x_2 - \log\left(\frac{1}{2}e^{2\beta\lambda_2}\left(e^{\alpha\beta}+1\right)\right) \\ &= 2\left(\frac{\beta}{\alpha}x_1 + x_2 - \beta\right)\lambda_2 + 2x_1 - \log\left(\frac{1}{2}\left(e^{\alpha\beta}+1\right)\right).\end{aligned}$$

By letting $|\lambda_2|$ tend to infinity, we see that if we maximize the above over $\lambda_2 \in \mathbb{R}$, it will only be finite when

$$\frac{\beta}{\alpha}x_1 + x_2 - \beta = 0,$$

which is equivalent to

$$\beta x_1 + \alpha x_2 = \alpha\beta.$$

Let us now compute the finite values of Λ^*. To this end, first consider the case $x_1 \in (0,\alpha)$ and $x_2 \in (0,\beta)$ with $\beta x_1 + \alpha x_2 = \alpha\beta$. Let us define

$$F(\lambda_1,\lambda_2) = 2\lambda_1 x_1 + 2\lambda_2 x_2 - \log\left(\frac{1}{2}e^{2\alpha\lambda_1} + \frac{1}{2}e^{2\beta\lambda_2}\right).$$

Computing the gradient, and equating to 0, we find for the critical points of F that

$$x_1 = \frac{\alpha}{e^{2\alpha\lambda_1} + e^{2\beta\lambda_2}} e^{2\alpha\lambda_1}$$

and

$$x_2 = \frac{\beta}{e^{2\alpha\lambda_1} + e^{2\beta\lambda_2}} e^{2\beta\lambda_2}.$$

Using that $\beta x_1 + \alpha x_2 = \alpha\beta$, we find that the above set of equations is solved by

$$\lambda_1^* = \frac{1}{2\alpha}\log(\beta x_1), \qquad \lambda_2^* = \frac{1}{2\beta}\log(\alpha x_2).$$

It follows that

$$\begin{aligned}\Lambda^*\left(\left(\begin{array}{cc} -x_1 & x_1 \\ x_2 & -x_2 \end{array}\right)\right) &= F(\lambda_1^*, \lambda_2^*) \\ &= \frac{1}{\alpha}\log(\beta x_1)x_1 + \frac{1}{\beta}\log(\alpha x_2)x_2 - \log\left(\frac{1}{2}\beta x_1 + \frac{1}{2}\alpha x_2\right) \\ &= \frac{1}{\alpha}\log(\beta x_1)x_1 + \frac{1}{\beta}\log(\alpha x_2)x_2 - \log\left(\frac{1}{2}\alpha\beta\right),\end{aligned}$$

where in the final step we used again that $\beta x_1 + \alpha x_2 = \alpha\beta$.
Now, in the case that $x_1 = 0$ and $x_2 = \beta$ we have

$$\begin{aligned}\Lambda^*\left(\left(\begin{array}{cc} 0 & 0 \\ \beta & -\beta \end{array}\right)\right) &= \sup_{\lambda_2\in\mathbb{R}}\sup_{\lambda_1\in\mathbb{R}}\left\{2\lambda_2\beta - \log\left(\frac{1}{2}e^{2\alpha\lambda_1} + \frac{1}{2}e^{2\beta\lambda_2}\right)\right\} \\ &= \sup_{\lambda_2\in\mathbb{R}}\left\{2\lambda_2\beta - \log\left(\frac{1}{2}e^{2\beta\lambda_2}\right)\right\} \\ &= \log(2).\end{aligned}$$

Likewise, we also have

$$\Lambda^*\left(\left(\begin{array}{cc} -\alpha & \alpha \\ 0 & 0 \end{array}\right)\right) = \log(2).$$

Now, the rate function for the large deviation principle for $\{\sigma_n^n\}_{n\geqslant 1}$ is given by

$$I(M) = \inf\left\{\int_0^1 \Lambda^*(\dot{\gamma}(t))\,\mathrm{d}t \,\middle|\, \gamma \in AC([0,1];\mathcal{S}(2,\mathbb{R})), \gamma(0) = I, \gamma(1) = M\right\}.$$

To get a more specific expression, we calculate the rate function further in the case where $\alpha = \beta$. Let $\gamma \in AC([0,1];\mathcal{S}(2,\mathbb{R}))$ with $\gamma(0) = I$. Then we can write

$$\gamma(t) = \left(\begin{array}{cc} 1-\gamma_1(t) & \gamma_1(t) \\ \gamma_2(t) & 1-\gamma_2(t) \end{array}\right),$$

so that

$$\dot{\gamma}(t) = \left(\begin{array}{cc} -\dot{\gamma}_1(t) & \dot{\gamma}_1(t) \\ \dot{\gamma}_2(t) & -\dot{\gamma}_2(t) \end{array}\right) \in T_{\gamma(t)}\mathcal{S}(2,\mathbb{R}).$$

Now recall that we may identify $T_{\gamma(t)}\mathcal{S}(2,\mathbb{R})$ with $\mathfrak{s}(2,\mathbb{R})$ using the map $\mathrm{d}L_{\gamma(t)}^{-1} = \mathrm{d}L_{\gamma(t)^{-1}}$. Because $\mathcal{S}(2,\mathbb{R})$ is a matrix Lie group, we have

$$\mathrm{d}L_{\gamma(t)^{-1}}(X) = \gamma(t)^{-1}X.$$

Therefore, as element of $\mathfrak{s}(2,\mathbb{R})$, the curve tangent to γ is given by

$$\begin{aligned}&\mathrm{d}L_{\gamma(t)^{-1}}(\dot{\gamma}) = \gamma(t)^{-1}\dot{\gamma}(t) \\ &= \frac{1}{1-\gamma_1(t)-\gamma_2(t)}\left(\begin{array}{cc} -(1-\gamma_2(t))\dot{\gamma}_1(t) - \gamma_1(t)\dot{\gamma}_2(t) & (1-\gamma_2(t))\dot{\gamma}_1(t) + \gamma_1(t)\dot{\gamma}_2(t) \\ (1-\gamma_1(t))\dot{\gamma}_2(t) + \gamma_2(t)\dot{\gamma}_1(t) & -(1-\gamma_1(t))\dot{\gamma}_2(t) - \gamma_2(t)\dot{\gamma}_1(t) \end{array}\right)\end{aligned}$$

Now, in order for

$$\int_0^1 \Lambda^*(\gamma(t)^{-1}\dot{\gamma}(t))\,\mathrm{d}t$$

to be finite, we need to have

$$\beta\frac{(1-\gamma_2(t))\dot{\gamma}_1(t)+\gamma_1(t)\dot{\gamma}_2(t)}{1-\gamma_1(t)-\gamma_2(t)}+\alpha\frac{(1-\gamma_1(t))\dot{\gamma}_2(t)+\gamma_2(t)\dot{\gamma}_1(t)}{1-\gamma_1(t)-\gamma_2(t)}=\alpha\beta,$$

because, as we have seen above, only then $\Lambda^*(\gamma(t)^{-1}\dot{\gamma}(t))<\infty$. Since $\alpha=\beta$, after some calculations, the above may be rewritten as

$$\dot{\gamma}_1(t)+\dot{\gamma}_2(t)=\alpha(1-(\gamma_1(t)+\gamma_2(t))).$$

If we now write $\psi(t)=\gamma_1(t)+\gamma_2(t)$, the previous equality gives a differential equation for ψ, namely

$$\dot{\psi}(t)=\alpha(1-\psi(t)),$$

with $\psi(0)=\gamma_1(0)+\gamma_2(0)=0$. The solution to this equation is given by

$$\psi(t)=1-e^{-\alpha t}.$$

In particular, this implies that

$$\gamma_1(1)+\gamma_2(1)=\psi(1)=1-e^{-\alpha}.$$

From this, we deduce that $I(M)$ is only finite for matrices satisfying $M_{12}+M_{21}=1-e^{-\alpha}$. Now, if M is such a matrix, the convexity of Λ^* together with Jensen's inequality, implies that

$$\inf\left\{\int_0^1 \Lambda^*(\dot{\gamma}(t))\,\mathrm{d}t\,\middle|\,\gamma\in\mathcal{AC}([0,1];\mathcal{S}(2,\mathbb{R})),\gamma(0)=I,\gamma(1)=M\right\}$$

is attained when taking $\gamma_1^*(t)=c\psi(t)$ and $\gamma_2^*(t)=(1-c)\psi(t)$. Since we need that $\gamma_1^*(1)=M_{12}$, we take

$$\gamma_1^*(t)=\frac{M_{12}}{M_{12}+M_{21}}\psi(t)=\frac{M_{12}}{1-e^{-\alpha}}(1-e^{-\alpha t}),$$

in which case

$$\gamma_2^*(t)=\frac{M_{21}}{1-e^{-\alpha}}(1-e^{-\alpha t}).$$

Using the expression for Λ^* we derived above, one obtains after some computations that

$$\begin{aligned}I(M)&=\int_0^1\Lambda^*\left(\begin{pmatrix}-\dot{\gamma}_1^*(t) & \dot{\gamma}_1^*(t)\\ \dot{\gamma}_2^*(t) & -\dot{\gamma}_2^*(t)\end{pmatrix}\right)\mathrm{d}t\\&=\alpha^2 M_{12}\log\left(\frac{\alpha M_{12}}{1-e^{-\alpha}}\right)-M_{12}-\frac{\alpha e^{-\alpha}M_{12}}{1-e^{-\alpha}}\end{aligned}$$

$$+\alpha^2 M_{21}\log\left(\frac{\alpha M_{21}}{1-e^{-\alpha}}\right) - M_{21} - \frac{\alpha e^{-\alpha}M_{21}}{1-e^{-\alpha}} - \log\left(\frac{1}{2}\alpha^2\right)$$
$$= \alpha^2 M_{12}\log\left(\frac{\alpha M_{12}}{1-e^{-\alpha}}\right) + \alpha^2 M_{21}\log\left(\frac{\alpha M_{21}}{1-e^{-\alpha}}\right)$$
$$+ (1-\alpha)e^{-\alpha} - \log\left(\frac{1}{2}\alpha^2\right) - 1$$

if $M_{12}+M_{21} = 1-e^{-\alpha}$. Otherwise, we have $I(M)=\infty$.

4.4. Some estimation results from Lie group theory

In this section we use the integral version of the Baker-Campbell-Hausdorff formula to derive a key estimate we need for proving Theorem 4.3.1. Essentially, we will show that for $X, Y \in \mathfrak{g}$ small enough, we can bound the difference between $\log(\exp(X)\exp(Y))$ and $X+Y$. These results are likely to be known to experts, however, we did not find a version in which the estimates are quantified precisely enough for our purposes. Estimates which are closely related, and obtained using a similar approach may be found in e.g. [46, Section 3].

4.4.1. Baker-Campbell-Hausdorff formula

Before we can state the Baker-Campbell-Hausdorff formula, we first need to introduce some linear operators on $\mathfrak{g}$.

For every $X \in \mathfrak{g}$, we define the adjoint map $\mathrm{ad}_X : \mathfrak{g} \to \mathfrak{g}$ by

$$\mathrm{ad}_X(Y) := [X, Y]. \tag{4.4.1}$$

Because the map $(X,Y) \mapsto \mathrm{ad}_X Y$ is smooth, it follows that $||\mathrm{ad}_X||$ depends continuously on X. In particular, this implies that

$$\sup_{X\in K} ||\mathrm{ad}_X|| < \infty$$

for all $K \subset \mathfrak{g}$ compact. Additionally, it also gives us that

$$\lim_{X\to 0} ||\mathrm{ad}_X|| = ||\mathrm{ad}_0|| = 0. \tag{4.4.2}$$

Because ad_X is a bounded operator, we can define the operator $e^{t\mathrm{ad}_X}$ by

$$e^{t\mathrm{ad}_X} = \sum_{m=0}^{\infty} \frac{t^m \mathrm{ad}_X^m}{m!}.$$

Similarly, for $f(z) = \frac{1-e^{-z}}{z} = \sum_{m=0}^{\infty} \frac{(-1)^m}{(m+1)!} z^m$ we define the operator

$$\frac{I-e^{-\mathrm{ad}_X}}{\mathrm{ad}_X} = f(\mathrm{ad}_X) = \sum_{m=0}^{\infty} \frac{(-1)^m}{(m+1)!}\mathrm{ad}_X^m.$$

From this series representation, we find that

$$\left\|I - \frac{I - e^{-\mathrm{ad}_X}}{\mathrm{ad}_X}\right\| \leqslant \sum_{k=1}^{\infty} \frac{||\mathrm{ad}_X||^k}{(k+1)!} \leqslant e^{||\mathrm{ad}_X||} - 1.$$

Now, by (4.4.2) the upper bound goes to 0 if $X \to 0$. This implies that if $|X|$ is small enough, then

$$\frac{I - e^{-\mathrm{ad}_X}}{\mathrm{ad}_X}$$

is invertible, with inverse given by

$$\frac{\mathrm{ad}_X}{I - e^{-\mathrm{ad}_X}} = g(e^{\mathrm{ad}_X}) \tag{4.4.3}$$

where $g(z) = \frac{z\log(z)}{z-1} = 1 + \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m(m+1)}(z-1)^m$ for $|z-1| < 1$.
With all relevant operators defined, we can state the integral form of the Baker-Campbell-Hausdorff formula, see e.g. [51, 92].

Theorem 4.4.1 (Baker-Campbell-Hausdorff)**.** *There exists an $r > 0$ such that for all $X, Y \in \mathfrak{g}$ with $|X|, |Y| \leqslant r$ we have that $\log(\exp(X)\exp(tY))$ is well-defined for all $t \in [0,1]$ and is given by*

$$\log(\exp(X)\exp(tY)) = X + \left(\int_0^t g(e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y})\,\mathrm{d}s\right) Y,$$

where $g(z) = \frac{z\log(z)}{z-1} = 1 + \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m(m+1)}(z-1)^m$ for $|z-1| < 1$.

We will use this formula to deduce approximations for the logarithm of a product of exponentials.

4.4.2. Logarithm of a product of exponentials

In this section we aim to control the difference

$$|\log(\exp(X)\exp(Y)) - X - Y|$$

for X and Y small enough. We do this using the Baker-Campbell-Hausdorff formula. We have the following proposition.

Proposition 4.4.2. *There exists an $r > 0$ such that for every $X \in \mathfrak{g}$ with $|X| \leqslant r$ there exists a constant $C_X > 0$ such that*

$$|\log(\exp(X)\exp(Y)) - X - Y| \leqslant C_X|Y|$$

for all $|Y| \leqslant r$. Moreover, the constants C_X may be chosen to only depend on $|X|$ and such that $\lim_{X\to 0} C_X = 0$.

Proof. By Theorem 4.4.1 there exists an $r > 0$ such that for $X, Y \in \mathfrak{g}$ with $|X|, |Y| \leqslant r$ we have

$$\log(\exp(X)\exp(Y)) = X + Y + \left(\int_0^1 \sum_{m=1}^{\infty} \frac{(-1)^m}{m(m+1)} (e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} - I)^m \, \mathrm{d}s\right) Y.$$

From this, it follows that

$$\begin{aligned}
&|\log(\exp(X)\exp(Y)) - X - Y| \\
&= \left| \left(\int_0^1 \sum_{m=1}^{\infty} \frac{(-1)^m}{m(m+1)} (e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} - I)^m \, \mathrm{d}s\right) Y \right| \\
&\leqslant \int_0^1 \sum_{m=1}^{\infty} \frac{1}{m(m+1)} ||e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} - I||^{m-1} |(e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} - I)Y| \, \mathrm{d}s.
\end{aligned}$$

Because $\mathrm{ad}_Y Y = 0$, we find that

$$e^{s\mathrm{ad}_Y} Y = Y + \sum_{m=1}^{\infty} \frac{s^m \mathrm{ad}_Y^{m-1}}{m!} \mathrm{ad}_Y Y = Y,$$

so that

$$|(e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} - I)Y| = |(e^{\mathrm{ad}_X} - I)Y| \leqslant ||e^{\mathrm{ad}_X} - I|||Y| \leqslant (e^{||\mathrm{ad}_X||} - 1)|Y|.$$

Here, the latter follows from

$$||e^{\mathrm{ad}_X} - I|| \leqslant \sum_{m=1}^{\infty} \frac{||\mathrm{ad}_X||^m}{m!} = e^{||\mathrm{ad}_X||} - 1.$$

Now define $Z(t) = \log(\exp(X)\exp(tY))$. Then

$$e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} = e^{\mathrm{ad}_{Z(s)}},$$

see e.g. [51, Chapter 5] or [92, Chapter 2]. From this we deduce

$$||e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} - I|| \leqslant e^{||\mathrm{ad}_{Z(s)}||} - 1.$$

By (4.4.2), we find $r' > 0$ such that $||\mathrm{ad}_{Z(s)}|| \leqslant \frac{\log(2)}{2}$ whenever $|Z(s)| \leqslant r'$. By Proposition 4.1.2, there is $r'' > 0$ such that latter holds whenever $d(e, \exp(Z(s))) \leqslant r''$.

Now we have

$$\begin{aligned}
d(e, \exp(Z(s))) &= d(e, \exp(X)\exp(tY)) \\
&\leqslant d(e, \exp(X)) + d(\exp(X), \exp(X)\exp(tY)) \\
&= d(e, \exp(X)) + d(e, \exp(tY)) \\
&\leqslant |X| + t|Y|,
\end{aligned}$$

where we used the triangle inequality and left-invariance of the metric. The last step follows by noticing that if $\gamma(t) = \exp(tX)$, then

$$d(e, \exp(X)) \leqslant \int_0^1 |\dot{\gamma}(t)| \, \mathrm{d}t = |X|.$$

Therefore, if $|X|, |Y| \leqslant \frac{1}{2} r''$, then $d(e, \exp(Z(s))) \leqslant r''$ so that $||\mathrm{ad}_{Z(s)}|| \leqslant \frac{\log(2)}{2}$ for all $s \in (0,1)$. But then $||e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} - I|| \leqslant \sqrt{2} - 1 < 1$, and hence

$$\begin{aligned}
&\int_0^1 \sum_{m=1}^{\infty} \frac{1}{m(m+1)} ||e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} - I||^{m-1} |(e^{\mathrm{ad}_X} e^{s\mathrm{ad}_Y} - I)Y| \, \mathrm{d}s \\
&\qquad \leqslant (e^{||\mathrm{ad}_X||} - 1)|Y| \int_0^1 \sum_{m=1}^{\infty} \frac{(\sqrt{2}-1)^{m-1}}{m(m+1)} \, \mathrm{d}s \\
&\qquad \leqslant (e^{||\mathrm{ad}_X||} - 1)|Y| \sum_{m=1}^{\infty} \frac{(\sqrt{2}-1)^{m-1}}{m(m+1)}.
\end{aligned}$$

We may thus take

$$C_X = (e^{||\mathrm{ad}_X||} - 1) \sum_{m=1}^{\infty} \frac{(\sqrt{2}-1)^{m-1}}{m(m+1)} < \infty.$$

Because $\lim_{|X| \to 0} ||\mathrm{ad}_X|| = 0$, it follows that $\lim_{X \to 0} C_X = 0$, and that C_X may be chosen to depend only on $|X|$. □

We conclude this section with the following result, which shows a Lipschitz-like estimate for the logarithm of a product of two exponentials.

Proposition 4.4.3. *There exist constants $r > 0$ and $C > 0$ such that for all $X, Y \in \mathfrak{g}$ with $|X|, |Y| \leqslant r$ we have*

$$|\log(\exp(X)\exp(-Y))| \leqslant C|X - Y|$$

Proof. Following the same reasoning as in the proof of Proposition 4.4.2, there exists an $r > 0$ such that

$$\log(\exp(X)\exp(-Y)) = X - Y - \left(\int_0^1 \sum_{m=1}^{\infty} \frac{(I - e^{\mathrm{ad}_X} e^{-t\mathrm{ad}_Y})^m}{m(m+1)} \, \mathrm{d}t \right) Y$$

whenever $|X|, |Y| \leqslant r$.
As before, we have $e^{-t\mathrm{ad}_Y} Y = Y$ and similarly $e^{\mathrm{ad}_X} X = X$. As a consequence, we can write

$$(I - e^{\mathrm{ad}_X} e^{-t\mathrm{ad}_Y})Y = (I - e^{\mathrm{ad}_X})Y = (I - e^{\mathrm{ad}_X})(Y - X),$$

from which it follows that

$$|(I - e^{\mathrm{ad}_X} e^{-t\mathrm{ad}_Y})^m Y| \leqslant ||I - e^{\mathrm{ad}_X} e^{-t\mathrm{ad}_Y}||^{m-1} ||I - e^{\mathrm{ad}_X}|| |Y - X|.$$

By similar reasoning as in the proof of Proposition 4.4.2, after possibly shrinking r, there exist constants $C, \tilde{C} > 0$ such that $|X|, |Y| \leqslant r$ implies that

$$\left| \left(\int_0^1 \sum_{m=1}^{\infty} \frac{(I - e^{\mathrm{ad}_X} e^{-t\mathrm{ad}_Y})^m}{m(m+1)} \, \mathrm{d}t \right) Y \right| \leqslant \tilde{C} ||I - e^{\mathrm{ad}_X}||| Y - X| \leqslant C|X - Y|.$$

By the triangle inequality we then find that

$$|\log(\exp(X)\exp(-Y))| \leqslant (C+1)|X - Y|$$

as desired. □

4.5. Proof of Theorem 4.3.1

As explained in Section 4.3.2, we prove the upper bound and lower bound for the large deviation principle of $\{\sigma_n^n\}_{n \geqslant 0}$ separately. More precisely, Theorem 4.3.1 follows immediately from Propositions 4.5.3 and 4.5.9. Before we get to either of these, we first need two general results, which we use in both the proof of the upper and lower bound.

First of all, define for every $n \in \mathbb{N}$ and every $1 \leqslant k \leqslant n$ the random variable

$$\sigma_k^n = \exp\left(\frac{1}{n} X_1\right) \cdots \exp\left(\frac{1}{n} X_k\right) \in G,$$

i.e., the point of the rescaled random walk after k increments. Finally, we set $\sigma_0^n = e$, the identity element of G. We have the following estimate.

Proposition 4.5.1. *Let the assumptions of Theorem 4.3.1 be satisfied. Then for every m large enough, there exists a constant $C_m > 0$ such that for all $1 \leqslant k \leqslant \lfloor m^{-1} n \rfloor$, $\log(\sigma_k^n)$ is well-defined and*

$$\left| \log\left(\sigma_k^n\right) - \frac{1}{n} \sum_{i=1}^{k} X_i \right| \leqslant C_m \frac{1}{m}.$$

Moreover, the constants C_m may be chosen so that $\lim_{m \to \infty} C_m = 0$.

Proof. First note that by the triangle inequality we have for any n and $1 \leqslant k \leqslant n$ that

$$d(\sigma_k^n, e) \leqslant \sum_{i=1}^{k} d(\sigma_i^n, \sigma_{i-1}^n).$$

Considering the curve $\gamma_i(t) = \sigma_{i-1}^n \exp(tX_i)$ in G, we obtain

$$d(\sigma_i^n, \sigma_{i-1}^n) \leqslant \int_0^{\frac{1}{n}} |\dot{\gamma}_i(t)| \, \mathrm{d}t = \frac{1}{n} |X_i|.$$

Hence, if we write B for the uniform bound on the increments, we find

$$d(\sigma_k^n, e) \leqslant \frac{k}{n} B. \tag{4.5.1}$$

But then we have for $1 \leqslant k \leqslant \lfloor m^{-1} n \rfloor$ that

$$d(\sigma_k^n, e) \leqslant \frac{\lfloor m^{-1} n \rfloor}{n} B \leqslant \frac{1}{m} B. \tag{4.5.2}$$

Thus if we choose m large enough, we can assure that σ_k^n is sufficiently close to e for $k = 1, \ldots, \lfloor m^{-1} n \rfloor$, so that $\log(\sigma_k^n)$ is well-defined for $1 \leqslant k \leqslant \lfloor m^{-1} n \rfloor$.

Turning to the proof of the estimate, first note that we may write

$$\log(\sigma_k^n) = \sum_{i=1}^k \log(\sigma_i^n) - \log\left(\sigma_{i-1}^n\right)$$

so that

$$\left| \log(\sigma_k^n) - \frac{1}{n} \sum_{i=1}^k X_i \right| \leqslant \sum_{i=1}^k \left| \log(\sigma_i^n) - \log\left(\sigma_{i-1}^n\right) - \frac{1}{n} X_i \right|.$$

By Proposition 4.1.2, for every $r > 0$ there exists an $\varepsilon > 0$ such that $d(e, g) \leqslant \varepsilon$ implies that $|\log(g)| \leqslant r$. Therefore, it follows from (4.5.2) that for $1 \leqslant k \leqslant \lfloor m^{-1} n \rfloor$, $|\log(\sigma_k^n)|$ can be made arbitrarily small by taking m large enough. Furthermore, because $|X_i| \leqslant B$, we find that $\frac{1}{n} X_i$ becomes small for large n. Hence, for m and n large enough we can apply Proposition 4.4.2 to obtain constants C_m with $\lim_{m \to \infty} C_m = 0$ such that

$$\left| \log(\sigma_i^n) - \log\left(\sigma_{i-1}^n\right) - \frac{1}{n} X_i \right| \leqslant C_m \frac{1}{n} |X_i| \leqslant C_m \frac{1}{n} B.$$

Combining everything, we find that

$$\left| \log(\sigma_k^n) - \frac{1}{n} \sum_{i=1}^k X_i \right| \leqslant \sum_{i=1}^k C_m \frac{1}{n} = C_m \frac{k}{n} \leqslant C_m \frac{1}{m}.$$

Here we used that $k \leqslant \lfloor m^{-1} n \rfloor$ and absorbed the constant B into C_m. □

In general we do not have that $\log(\sigma_k^n)$ exists in $\mathfrak{g}$ for all n and all $1 \leqslant k \leqslant n$. Therefore, in order to be able to use some identification of the random walk with a process in the Lie algebra, we need to make sure we can actually use the logarithm map.

To this end, notice that in the previous proof we have seen in (4.5.2) that $d(\sigma_k^n, e) \leqslant \frac{1}{m} B$ for $1 \leqslant k \leqslant \lfloor m^{-1} n \rfloor$, where B is the uniform bound on the increments. With this estimate in mind, the idea is now to split the random walk into m pieces, each consisting of (approximately) $\lfloor m^{-1} n \rfloor$ increments. More precisely, we define the indices $n_l = l\lfloor m^{-1}n\rfloor$ for $l = 0, \dots, m-1$ and set $n_m = n$. Because the metric is left-invariant, it follows similarly to (4.5.2) that

$$d(e, (\sigma^n_{n_{l-1}})^{-1}\sigma^n_{n_{l-1}+k}) = d(\sigma^n_{n_{l-1}}, \sigma^n_{n_{l-1}+k}) \leqslant \frac{1}{m}B \tag{4.5.3}$$

for every $l = 1, \dots, m$ and every $k = 1, \dots, n_l - n_{l-1}$. This implies that for m large enough we can define

$$Y_k^{n,m,l} = \log\left((\sigma^n_{n_{l-1}})^{-1}\sigma^n_{n_{l-1}+k}\right) \in \mathfrak{g}$$

for every $l = 1, \dots, m$ and $k = 1, \dots, n_l - n_{l-1}$.
Note that

$$(\sigma^n_{n_{l-1}})^{-1}\sigma^n_{n_{l-1}+k} = \exp\left(\frac{1}{n}X_{n_{l-1}+1}\right)\cdots\exp\left(\frac{1}{n}X_{n_{l-1}+k}\right),$$

so that

$$Y_k^{n,m,l} = \log\left(\exp\left(\frac{1}{n}X_{n_{l-1}+1}\right)\cdots\exp\left(\frac{1}{n}X_{n_{l-1}+k}\right)\right). \tag{4.5.4}$$

For every m, this allows us to define a random vector

$$\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in \mathfrak{g}^m. \tag{4.5.5}$$

By (4.5.4), we have that $Y^{m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{m,m}_{\lfloor m^{-1}n\rfloor}$ are independent and identically distributed random variables in $\mathfrak{g}$, because the X_i are independent and identically distributed by assumption.

4.5.1. Proof of the upper bound for the large deviation principle of $\{\sigma^n_n\}_{n\geqslant 0}$

As explained in Section 4.3.2, we prove the upper bound for the large deviation principle of $\{\sigma^n_n\}_{n\geqslant 0}$ by transferring the problem to the Lie algebra and obtain suitable estimates there using a similar approach as in the Euclidean case. We start with the following result.

Proposition 4.5.2. *Let the assumptions of Theorem 4.3.1 be satisfied. Let $m \in \mathbb{N}$ be large enough so that the random vector*

$$\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in \mathfrak{g}^m$$

in (4.5.5) *is well-defined. Then there exists a constant $C_m > 0$ such that for every $F \subset \mathfrak{g}^m$ closed we have*

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in F\right)$$
$$\leqslant -\inf_{(x_1,\dots,x_m)\in F}\frac{1}{m}\sum_{i=1}^m \sup_{\lambda\in\mathfrak{g}}\left\{\langle\lambda, mx_i\rangle - \Lambda(\lambda) - C_m|\lambda|\right\}.$$

Here, $\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda X_1})$. *Moreover, the constants* C_m *may be chosen such that* $\lim_{m\to\infty} C_m = 0$.

Proof. Following the proof of Cramér's theorem for the vector space $\mathfrak{g}^m$ (see e.g. [29, 56]), we have for any $\Gamma \subset \mathfrak{g}^m$ compact that

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in \Gamma\right)\\ &\leqslant - \inf_{(x_1,\dots,x_m)\in\Gamma} \sup_{(\lambda_1,\dots,\lambda_m)\in\mathfrak{g}^m} \left\{\sum_{i=1}^m \langle \lambda_i, x_i\rangle - \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(e^{n\sum_{i=1}^m \langle \lambda_i, Y^{n,m,i}_{\lfloor m^{-1}n\rfloor}\rangle}\right)\right\}\end{aligned}$$

However, as mentioned above, the fact that the X_i are independent and identically distributed, together with (4.5.4), shows that $Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}$ are independent and identically distributed. Hence

$$\mathbb{E}\left(e^{n\sum_{i=1}^m \langle \lambda_i, Y^{n,m,i}_{\lfloor m^{-1}n\rfloor}\rangle}\right) = \prod_{i=1}^m \mathbb{E}\left(e^{n\langle \lambda_i, Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}\rangle}\right).$$

By Proposition 4.5.1, there exist constants $C_m > 0$ with $\lim_{m\to\infty} C_m = 0$ such that

$$\left| Y^{n,m,1}_{\lfloor m^{-1}n\rfloor} - \frac{1}{n}\sum_{i=1}^{\lfloor m^{-1}n\rfloor} X_i \right| \leqslant C_m \frac{1}{m}.$$

Using the Cauchy-Schwarz inequality, this gives us that

$$\begin{aligned}\mathbb{E}\left(e^{n\langle \lambda_i, Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}\rangle}\right) &\leqslant \mathbb{E}\left(e^{\sum_{j=1}^{\lfloor m^{-1}n\rfloor}\langle \lambda_i, X_j\rangle}\right) e^{n|\lambda_i| C_m m^{-1}}\\ &= e^{n|\lambda_i|C_m m^{-1}} \mathbb{E}\left(e^{\langle \lambda_i, X_1\rangle}\right)^{\lfloor m^{-1}n\rfloor}.\end{aligned}$$

Hence

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(e^{n\sum_{i=1}^m \langle \lambda_i, Y^{n,m,i}_{\lfloor m^{-1}n\rfloor}\rangle}\right) &= \sum_{i=1}^m \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(e^{n\langle \lambda_i, Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}\rangle}\right)\\ &\leqslant \sum_{i=1}^m \left\{|\lambda_i| C_m \frac{1}{m} + \frac{1}{m}\log \mathbb{E}\left(e^{\langle \lambda_i, X_1\rangle}\right)\right\}\\ &= \frac{1}{m}\sum_{i=1}^m \left\{C_m|\lambda_i| + \log \mathbb{E}\left(e^{\langle \lambda_i, X_1\rangle}\right)\right\}.\end{aligned}$$

Collecting everything, we find that

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \dots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in \Gamma\right)\\ &\qquad\leqslant - \inf_{(x_1,\dots,x_m)\in\Gamma} \sup_{(\lambda_1,\dots,\lambda_m)\in\mathfrak{g}^m} \frac{1}{m}\sum_{i=1}^m \left\{\langle \lambda_i, m x_i\rangle - \log \mathbb{E}\left(e^{\langle \lambda_i, X_1\rangle}\right) - C_m|\lambda_i|\right\}\end{aligned}$$

$$= -\inf_{(x_1,\ldots,x_m)\in\Gamma} \frac{1}{m}\sum_{i=1}^{m} \sup_{\lambda\in\mathfrak{g}} \{\langle\lambda, mx_i\rangle - \Lambda(\lambda) - C_m|\lambda|\}.$$

To extend this upper bound to all closed sets, note that the boundedness of the increments of the random walk implies that $Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}$ is bounded, and hence remains in a compact subset of $\mathfrak{g}$. Because $Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}$ are independent and identically distributed, we can conclude from this that $(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor})$ is exponentially tight in $\mathfrak{g}^m$. From this it follows that the upper bound actually holds for all closed sets, which completes the proof. □

With the preparations done, we can now turn to the proof of the upper bound of the large deviation principle for $\{\sigma^n_n\}_{n\geqslant 1}$. The main work goes into proving that we actually obtain the desired form of the upper bound.

Proposition 4.5.3. *Let the assumptions of Theorem 4.3.1 be satisfied. Then for any $F \subset G$ closed we have*

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\sigma^n_n \in F) \leqslant -\inf_{g\in F} I_G(g),$$

where I_G is the good rate function given by (4.3.1).

Proof. Let $F \subset G$ be closed. Choose $m \in \mathbb{N}$ large enough so that the random vector

$$\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in \mathfrak{g}^m$$

defined in (4.5.5) is well-defined. Let $\Psi_m : \mathfrak{g}^m \to G$ be the map given by

$$\Psi_m(x_1, \ldots, x_m) = \exp(x_1)\cdots\exp(x_m).$$

Because Ψ_m is a composition of continuous functions, it is itself continuous. Furthermore, observe that by construction

$$\Psi_m\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) = \sigma^n_n.$$

This implies that

$$\mathbb{P}\left(\sigma^n_n \in F\right) = \mathbb{P}\left(\left(Y^{n,m,1}_{\lfloor m^{-1}n\rfloor}, \ldots, Y^{n,m,m}_{\lfloor m^{-1}n\rfloor}\right) \in \Psi_m^{-1}F\right),$$

where $\Psi_m^{-1}F$ is closed, because F is closed and Ψ_m is continuous. By Proposition 4.5.2 we then find that

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left(\sigma^n_n \in F\right)$$
$$\leqslant -\inf_{(x_1,\ldots,x_m)\in\Psi_m^{-1}F} \frac{1}{m}\sum_{i=1}^{m} \sup_{\lambda\in\mathfrak{g}} \{\langle\lambda, mx_i\rangle - \Lambda(\lambda) - C_m|\lambda|\},$$

where $\lim_{m\to\infty} C_m = 0$.
The final step is now to let m tend to infinity, and show that we obtain the desired upper bound. For this, we need to show that

$$-\lim_{m\to\infty} \inf_{(x_1,\dots,x_m)\in\Psi_m^{-1}F} \frac{1}{m}\sum_{i=1}^m \sup_{\lambda\in\mathfrak{g}} \{\langle\lambda, mx_i\rangle - \Lambda(\lambda) - C_m|\lambda|\} \leqslant -\inf_{g\in F} I_G(g).$$

To this end, let $\varepsilon > 0$ be arbitrary. Because $\lim_{m\to\infty} C_m = 0$, we can find $m_0 \in \mathbb{N}$ such that $m \geqslant m_0$ implies that $C_m < \varepsilon$. In that case, we have

$$\begin{aligned} &-\inf_{(x_1,\dots,x_m)\in\Psi_m^{-1}F} \frac{1}{m}\sum_{i=1}^m \sup_{\lambda\in\mathfrak{g}} \{\langle\lambda, mx_i\rangle - \Lambda(\lambda) - C_m|\lambda|\} \\ &\qquad\leqslant -\inf_{(x_1,\dots,x_m)\in\Psi_m^{-1}F} \frac{1}{m}\sum_{i=1}^m \sup_{\lambda\in\mathfrak{g}} \{\langle\lambda, mx_i\rangle - \Lambda(\lambda) - \varepsilon|\lambda|\} \\ &\qquad= -\inf_{(x_1,\dots,x_m)\in\Psi_m^{-1}F} \frac{1}{m}\sum_{i=1}^m \Lambda_\varepsilon^*(mx_i), \end{aligned}$$

where $\Lambda_\varepsilon(\lambda) = \Lambda(\lambda) + \varepsilon|\lambda|$ and Λ_ε^* denotes its Legendre transform.
Now note that

$$\frac{1}{m}\sum_{i=1}^m \Lambda_\varepsilon^*(mx_i) = \int_0^1 \Lambda_\varepsilon^*(\dot\gamma(t))\,\mathrm{d}t,$$

where $\gamma : [0,1] \to G$ is given by $\gamma(0) = e$ and

$$\gamma(t) = \gamma\left(\frac{i-1}{m}\right)\exp\left(\left(t - \frac{i-1}{m}\right) mx_i\right), \qquad t \in \left[\frac{i-1}{m}, \frac{i}{m}\right],$$

for $i = 1, \dots, m$. Furthermore, note that $\gamma(1) = \Psi_m(x_1, \dots, x_m)$.
Using this, we find that

$$\begin{aligned} &-\inf_{(x_1,\dots,x_m)\in\Psi_m^{-1}F} \frac{1}{m}\sum_{i=1}^m \Lambda_\varepsilon^*(mx_i) \\ &\qquad\leqslant -\inf\left\{\int_0^1 \Lambda_\varepsilon^*(\dot\gamma(t))\,\mathrm{d}t \,\middle|\, \gamma \in AC([0,1];G), \gamma(0) = e, \gamma(1) = g\right\}. \end{aligned}$$

It remains to consider the limit $\varepsilon \to 0$. To this end, first suppose that $I_G(g) < \infty$. By the goodness of the rate function $\mathcal{I}_\varepsilon(\gamma) = \int_0^1 \Lambda_\varepsilon^*(\dot\gamma(t))\,\mathrm{d}t$, the sets

$$C_\varepsilon := \left\{\gamma \,\middle|\, \int_0^1 \Lambda_\varepsilon^*(\dot\gamma(t))\,\mathrm{d}t \leqslant 2I_G(g)\right\}$$

are compact. Furthermore, we have $C_{\varepsilon'} \subset C_\varepsilon$ whenever $\varepsilon' \leqslant \varepsilon$. Because lower-semicontinuous functions attain their minimum on compact sets, we have a sequence γ_ε such that

$$\int_0^1 \Lambda_\varepsilon^*(\dot\gamma_\varepsilon(t))\,\mathrm{d}t = \inf\left\{\int_0^1 \Lambda_\varepsilon^*(\dot\gamma(t))\,\mathrm{d}t \,\middle|\, \gamma \in AC([0,1];G), \gamma(0) = e, \gamma(1) = g, \right\}$$

$$=: I_\varepsilon.$$

Because the sequence C_ε is decreasing, for ε small enough, the sequence γ_ε is contained in a compact set, and hence, upon passing to subsequences, we may assume that γ_ε converges with limit γ. But then we find for every $\delta > 0$ that

$$\begin{aligned}\liminf_{\varepsilon\to 0} I_\varepsilon &= \liminf_{\varepsilon\to 0} \int_0^1 \Lambda_\varepsilon^*(\dot\gamma_\varepsilon(t))\,\mathrm{d}t \\ &\geqslant \liminf_{\varepsilon\to 0} \int_0^1 \Lambda_\delta^*(\dot\gamma_\varepsilon(t))\,\mathrm{d}t \\ &\geqslant \int_0^1 \Lambda_\delta^*(\dot\gamma(t))\,\mathrm{d}t.\end{aligned}$$

As this holds for all $\delta > 0$, by taking the limit $\delta \to 0$ we find that

$$\liminf_{\varepsilon\to 0} I_\varepsilon \geqslant \int_0^1 \Lambda^*(\dot\gamma(t))\,\mathrm{d}t \geqslant I_G(g).$$

Because also $I_\varepsilon \leqslant I_G(g)$ for every $\varepsilon > 0$, we find that $\lim_{\varepsilon\to 0} I_\varepsilon = I_G(g)$ as desired. Now consider the case that $I_G(g) = \infty$. Suppose that I_ε does not go to ∞. Then $\liminf_{\varepsilon\to 0} I_\varepsilon < \infty$. Upon passing to subsequences, suppose that $\lim_{\varepsilon\to 0} I_\varepsilon = I$. Following a similar reasoning as above, we find a sequence γ_ε converging to γ which we can use to show that

$$I_G(g) \leqslant \liminf_{\varepsilon\to 0} I_\varepsilon < \infty,$$

which is a contradiction. We conclude that $\lim_{\varepsilon\to 0} I_\varepsilon = \infty$.
Collecting everything, we have that

$$\lim_{\varepsilon\to 0}\left[\inf\left\{\int_0^1 \Lambda_\varepsilon^*(\dot\gamma(t))\,\mathrm{d}t | \gamma : [0,1] \to G, \gamma(0) = e, \gamma(1) = g, \gamma \in AC\right\}\right] = I_G(g).$$

so that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\sigma_n^n \in F) \leqslant - \inf_{g\in F} I_G(g)$$

as desired. □

4.5.2. Proof of the lower bound for the large deviation principle of $\{\sigma_n^n\}_{n\geqslant 0}$

Before we get to the proof of the lower bound for the large deviation principle of $\{\sigma_n^n\}_{n\geqslant 0}$, we first need to study more carefully the continuity properties of the maps $\Psi_m : \mathfrak{g}^m \to G$ given by

$$\Psi_m(x_1, \ldots, x_m) = \exp(x_1) \cdots \exp(x_m).$$

Before we can do this, we first need some technical lemmas.

Lemma 4.5.4. *Let $K \subset G$ be compact. Denote by $\Phi_g : G \to G$ conjugation by g, i.e., $\Phi_g h = ghg^{-1}$. Then*

$$\sup_{g \in K} ||\mathrm{d}\Phi_g(e)|| < \infty.$$

Proof. Consider the map $\Phi : G \times G \to G$ given by $\Phi(g,h) := \Phi_g h$. Because $\Phi(g,h) = R_{g^{-1}} L_g h$, the map Φ is smooth. Now note that

$$\mathrm{d}\Phi_g(e)(X) = \mathrm{d}\Phi(g,e)(0,X)$$

Because Φ is smooth, the latter is continuous in g and hence

$$\sup_{g \in K} |\mathrm{d}\Phi_g(e)X| < \infty$$

for all $X \in \mathfrak{g}$. But then it follows from the uniform boundedness principle that also

$$\sup_{g \in K} ||\mathrm{d}\Phi_g(e)|| < \infty$$

as desired. □

Lemma 4.5.5. *For every $X \in \mathfrak{g}$ and $g, h \in G$ we have*

$$|\mathrm{d}\Phi_g(h)(\mathrm{d}L_h(e)(X))| = |\mathrm{d}\Phi_g(e)(X)|.$$

Proof. Since $\mathrm{d}L^{-1}_{\Phi_g h}(\Phi_g h)$ is an isometry, we have

$$|\mathrm{d}\Phi_g(h)(\mathrm{d}L_h(e)(X))| = |\mathrm{d}L^{-1}_{\Phi_g h}(\Phi_g h)(\mathrm{d}\Phi_g(h)(\mathrm{d}L_h(e)(X)))|.$$

By the chain rule

$$\mathrm{d}L^{-1}_{\Phi_g h}(\Phi_g h)(\mathrm{d}\Phi_g(h)(\mathrm{d}L_h(e)(X))) = \mathrm{d}(L^{-1}_{\Phi_g h} \circ \Phi_g \circ L_h)(e)(X).$$

Now, consider $\gamma(t) = \exp(tX)$. Then $\gamma(0) = e$ and $\dot{\gamma}(0) = X$, which gives us that

$$\begin{aligned}
\mathrm{d}(L^{-1}_{\Phi_g h} \circ \Phi_g \circ L_h)(e)(X) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (L^{-1}_{\Phi_g h} \circ \Phi_g \circ L_h)(\exp(tX)) \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} gh^{-1}g^{-1}gh\exp(tX)g^{-1} \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} g\exp(tX)g^{-1} \\
&= \mathrm{d}\Phi_g(e)(X).
\end{aligned}$$

Combining all the equalities, we find that

$$|\mathrm{d}\Phi_g(h)(\mathrm{d}L_h(e)(X))| = |\mathrm{d}\Phi_g(e)(X)|$$

as desired. □

Lemma 4.5.6. *For every $X \in \mathfrak{g}$ and $g \in G$ we have*

$$d(e, \Phi_g \exp(X)) \leqslant |\mathrm{d}\Phi_g(e)X|,$$

where Φ_g denotes conjugation with g.

Proof. Consider the curve $\gamma(t) = \Phi_g \exp(tX)$. By definition of the Riemannian distance, we have

$$d(e, \Phi_g \exp(X)) \leqslant \int_0^1 |\dot{\gamma}(t)| \, \mathrm{d}t.$$

By the chain rule, we have that

$$\dot{\gamma}(t) = \mathrm{d}\Phi_g(\exp(tX))(\mathrm{d}L_{\exp(tX)}(e)(X)).$$

By Lemma 4.5.5 we have

$$|\mathrm{d}\Phi_g(\exp(tX))(\mathrm{d}L_{\exp(tX)}(e)(X))| = |\mathrm{d}\Phi_g(e)(X)|.$$

Combining everything, we find that

$$d(e, \Phi_g \exp(X)) \leqslant \int_0^1 |\mathrm{d}\Phi_g(e)(X)| \, \mathrm{d}t = |\mathrm{d}\Phi_g(e)(X)|,$$

which concludes the proof. □

We can now prove the following continuity property of the maps Ψ_m.

Proposition 4.5.7. *For every $r > 0$, there exists a constant $C > 0$ such that for all $\varepsilon > 0$ and $m \in \mathbb{N}$ large enough we have that if*

$$(x_1, \ldots, x_m) \in B(y_1, C^{-1}\varepsilon) \times \cdots B(y_m, C^{-1}\varepsilon),$$

then

$$\Psi_m(x_1, \ldots, x_m) \in B(\Psi_m(y_1, \ldots, y_m), \varepsilon)$$

whenever $|x_i|, |y_i| \leqslant \frac{r}{m}$.

Proof. By the triangle inequality and left-invariance of the metric, we have

$$\begin{aligned}
&d(\Psi_m(x_1, \ldots, x_m), \Psi_m(y_1, \ldots, y_m)) \\
&\qquad = d(\Psi_{m-1}(x_2, \ldots, x_m), \exp(-x_1)\exp(y_1)\Psi_{m-1}(y_2, \ldots, y_m)) \\
&\qquad \leqslant d(\Psi_{m-1}(x_2, \ldots, x_m), \Psi_{m-1}(y_2, \ldots, y_m)) \\
&\qquad\qquad + d(\Psi_{m-1}(y_2, \ldots, y_m), \exp(-x_1)\exp(y_1)\Psi_{m-1}(y_2, \ldots, y_m)) \\
&\qquad = d(\Psi_{m-1}(x_2, \ldots, x_m), \Psi_{m-1}(y_2, \ldots, y_m)) \\
&\qquad\qquad + d(e, \Psi_{m-1}(y_2, \ldots, y_m)^{-1}\exp(-x_1)\exp(y_1)\Psi_{m-1}(y_2, \ldots, y_m)).
\end{aligned}$$

Now if m is large enough, then for x_1, y_1 with $|x_1|, |y_1| \leqslant \frac{r}{m}$, we have that

$$\log(\exp(-x_1)\exp(y_1))$$

is well-defined. Furthermore, by Proposition 4.4.3 there exists a constant C such that

$$|\log(\exp(-x_1)\exp(y_1))| \leqslant C|x_1 - y_1|.$$

If we now write Φ_g for conjugation with g, it follows from Lemma 4.5.6 that

$$\begin{aligned} &d(e, \Psi_{m-1}(y_2, \dots, y_m)^{-1}\exp(-x_1)\exp(y_1)\Psi_{m-1}(y_2, \dots, y_m)) \\ &\quad \leqslant ||\mathrm{d}\Phi_{\Psi_{m-1}(y_2,\dots,y_m)^{-1}}(e)|||\log(\exp(-x_1)\exp(y_1))| \\ &\quad \leqslant C||\mathrm{d}\Phi_{\Psi_{m-1}(y_2,\dots,y_m)^{-1}}(e)|||x_1 - y_1| \end{aligned}$$

Because $y_2, \dots, y_m \in B(0, rm^{-1})$, in the same way as we obtained (4.5.1), we find that

$$|\Psi_{m-1}(y_2, \dots, y_m)| \leqslant B\sum_{i=2}^{m}|y_i| \leqslant Br.$$

Since Lie groups are complete as Riemannian manifold, the set $\overline{B(e, Br)} \subset G$ is compact. Combining everything and applying Lemma 4.5.4, there exists a constant $\tilde{C} > 0$ such that

$$||\mathrm{d}\Phi_{\Psi_{m-1}(y_2,\dots,y_m)^{-1}}(e)|| \leqslant \tilde{C}.$$

Collecting everything, and absorbing the constants into one, we find that

$$d(e, \Psi_{m-1}(y_2, \dots, y_m)^{-1}\exp(-x_1)\exp(y_1)\Psi_{m-1}(y_2, \dots, y_m)) \leqslant C|x_1 - y_1|$$

for some $C > 0$. We conclude that

$$\begin{aligned} &d(\Psi_m(x_1, \dots, x_m), \Psi_m(y_1, \dots, y_m)) \\ &\quad \leqslant d(\Psi_{m-1}(x_1, \dots, x_m), \Psi_{m-1}(y_1, \dots, y_m)) + C|x_1 - y_1|. \end{aligned}$$

Iterating this procedure, we find that

$$d(\Psi_m(x_1, \dots, x_m), \Psi_m(y_1, \dots, y_m)) \leqslant C\sum_{i=1}^{m}|x_i - y_i|.$$

It thus follows that if

$$(x_1, \dots, x_m) \in B(y_1, (Cm)^{-1}\varepsilon) \times \cdots \times B(y_m, (Cm)^{-1}\varepsilon),$$

then

$$d(\Psi_m(x_1, \dots, x_m), \Psi_m(y_1, \dots, y_m)) < C\sum_{i=1}^{m}(Cm)^{-1}\varepsilon = \varepsilon,$$

which proves the claim. ☐

We need one more result, which allows us to partition absolutely continuous curves in G in an appropriate way.

Proposition 4.5.8. *Let $\gamma \in AC([0,1];G)$ be arbitrary. Assume that $\dot{\gamma} \in L^\infty([0,1],\mathfrak{g})$. Then for each m large enough, the vectors*

$$\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma\left(\frac{i}{m}\right)\right) \in \mathfrak{g}$$

are well-defined for $i = 1,\dots,m$. Furthermore, there exist constants L_m with $\lim_{m\to\infty} L_m = 0$ such that

$$\left|\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma\left(\frac{i}{m}\right)\right) - \int_{\frac{i-1}{m}}^{\frac{i}{m}} \dot{\gamma}(t)\,\mathrm{d}t\right| \leqslant L_m\frac{1}{m}||\dot{\gamma}||_\infty.$$

Proof. First of all, because γ is continuous and $[0,1]$ is compact, it is actually uniformly continuous. Therefore, we can take $m \in \mathbb{N}$ large enough, so that for $i = 1,\dots,m$ the vectors

$$\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right) \in \mathfrak{g}$$

are well-defined for $r \in [\frac{i-1}{m},\frac{i}{m}]$.
Now consider the function $f_{i,m} : [\frac{i-1}{m},\frac{i}{m}] \to \mathfrak{g}$ given by

$$f_{i,m}(r) = \log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right). \tag{4.5.6}$$

Then

$$f'_{i,m}(r) = \mathrm{d}\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right)(\dot{\gamma}(r)),$$

where again we used the identification of $T_{\gamma(r)}G$ with $\mathfrak{g}$. Using this, we obtain

$$\begin{aligned}\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma\left(\frac{i}{m}\right)\right) &= f_{i,m}\left(\frac{i}{m}\right) - f_{i,m}\left(\frac{i-1}{m}\right)\\ &= \int_{\frac{i-1}{m}}^{\frac{i}{m}} f'_{i,m}(r)\,\mathrm{d}r\\ &= \int_{\frac{i-1}{m}}^{\frac{i}{m}} \mathrm{d}\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right)(\dot{\gamma}(r))\,\mathrm{d}r.\end{aligned}$$

With this expression at hand, we estimate

$$\begin{aligned}&\left|\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma\left(\frac{i}{m}\right)\right) - \int_{\frac{i-1}{m}}^{\frac{i}{m}} \dot{\gamma}(r)\,\mathrm{d}t\right|\\ &\qquad\leqslant \int_{\frac{i-1}{m}}^{\frac{i}{m}} \left|\left|\mathrm{d}\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right) - I\right|\right| |\dot{\gamma}(r)|\,\mathrm{d}r\end{aligned} \tag{4.5.7}$$

$$\leqslant ||\dot{\gamma}||_\infty \int_{\frac{i-1}{m}}^{\frac{i}{m}} \left\| \mathrm{d}\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right) - I \right\| \mathrm{d}r.$$

It follows from (4.4.3) that (see also [51, Chapter 5] or [92, Chapter 2])

$$\mathrm{d}\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right) - I = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k(k+1)} \left(e^{\mathrm{ad}_{f_{i,m}(r)}} - I\right)^k.$$

Here, $f_{i,m}(r)$ is as defined in (4.5.6). From this it follows that

$$\left\| \mathrm{d}\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right) - I \right\| \leqslant \sum_{k=1}^{\infty} \frac{1}{k(k+1)} \left(e^{||\mathrm{ad}_{f_{i,m}(r)}||} - 1\right)^k. \tag{4.5.8}$$

Now, fix $\varepsilon > 0$. By Proposition 4.1.2 there exists a $\delta > 0$ such that $d(e,g) < \delta$ implies that $|\log(g)| < \varepsilon$. Since γ is uniformly continuous on $[0,1]$, we can choose m large enough so that for every $i = 1,\dots,m$ it holds that $d(\gamma(\frac{i-1}{m}),\gamma(r)) < \delta$ for all $r \in [\frac{i-1}{m},\frac{i}{m}]$. Using left-invariance of the Riemannian distance, we find that $d(e,\gamma(\frac{i-1}{m})^{-1}\gamma(r)) < \delta$ for all $r \in [\frac{i-1}{m},\frac{i}{m}]$ and all $i = 1,\dots,m$. But then we have for all $i = 1,\dots,m$ and all $r \in [\frac{i-1}{m},\frac{i}{m}]$ that

$$\left|\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right)\right| < \varepsilon.$$

We conclude that

$$\lim_{m\to\infty} \sup_{1\leqslant i\leqslant m} \sup_{r\in[\frac{i-1}{m},\frac{i}{m}]} \left|\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right)\right| = 0.$$

With (4.4.2) in mind, it follows that

$$\lim_{m\to\infty} \sup_{1\leqslant i\leqslant m} \sup_{r\in[\frac{i-1}{m},\frac{i}{m}]} \left\| \mathrm{ad}_{\log\left(\gamma(\frac{i-1}{m})^{-1}\gamma(r)\right)} \right\| = 0.$$

Recalling the definition of $f_{i,m}(r)$ in (4.5.6), this in turn implies that the upper bound in (4.5.8) tends to 0 if m goes to infinity, independent of i. We can thus find constants L_m with $\lim_{m\to\infty} L_m = 0$ such that

$$\left\| \mathrm{d}\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma(r)\right) - I \right\| \leqslant L_m$$

for all $i = 1,\dots,m$ and all $r \in [\frac{i-1}{m},\frac{i}{m}]$. If we plug this into (4.5.7), we find

$$\left|\log\left(\gamma\left(\frac{i-1}{m}\right)^{-1}\gamma\left(\frac{i}{m}\right)\right) - \int_{\frac{i-1}{m}}^{\frac{i}{m}} \dot{\gamma}(r)\,\mathrm{d}t\right| \leqslant ||\dot{\gamma}||_\infty \int_{\frac{i-1}{m}}^{\frac{i}{m}} L_m\,\mathrm{d}r = L_m \frac{1}{m} ||\dot{\gamma}||_\infty$$

as desired. □

With the final preparations done, we can prove the lower bound of the large deviation principle for $\{\sigma_n^n\}_{n\geqslant 1}$.

Proposition 4.5.9. *Let the assumptions of Theorem 4.3.1 be satisfied. Then for every $U \subset G$ open we have*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\sigma_n^n \in U) \geqslant -\inf_{g\in U} I_G(g),$$

where I_G is the good rate function given by (4.3.1)*.*

Proof. Let $U \subset G$ be open. Fix $g \in U$ and a curve $\gamma \in AC([0,1];G)$ with $\gamma(0) = e$ and $\gamma(1) = g$. We will show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(\frac{1}{n} * \mathcal{S}\right)_n \in U\right) \geqslant -\int_0^1 \Lambda^*(\dot{\gamma}(t))\, \mathrm{d}t.$$

If $\int_0^1 \Lambda^*(\dot{\gamma}(t))\, \mathrm{d}t = \infty$, the above is certainly true. Hence, suppose $\int_0^1 \Lambda^*(\dot{\gamma}(t))\, \mathrm{d}t < \infty$. Because Λ is the log-moment generating function of a bounded random variable, it follows Λ^* is finite only on a bounded set, referred to as its domain. Therefore, because $\int_0^1 \Lambda^*(\dot{\gamma}(t))\, \mathrm{d}t < \infty$, it must be that $\dot{\gamma}(t)$ is in the domain of Λ^* for almost all t. But then we have that $||\dot{\gamma}||_\infty < \infty$.

By the same reasoning as in the proof of Proposition 4.5.8, we can take $m \in \mathbb{N}$ large enough, so that we can define for $i = 1, \ldots, m$ the vectors

$$y_i^m := \log\left(\gamma\left(\frac{i-1}{m}\right)^{-1} \gamma\left(\frac{i}{m}\right)\right) \in \mathfrak{g}.$$

Let $\Psi_m : \mathfrak{g}^m \to G$ be again the map given by

$$\Psi_m(x_1, \ldots, x_m) = \exp(x_1)\cdots\exp(x_m),$$

so that $g = \Psi_m(y_1^m, \ldots, y_m^m)$.

Because U is open, there exists an $\varepsilon > 0$ such that $B(g, \varepsilon) \subset U$. By Proposition 4.5.7, for m large enough, there exists a constant $C > 0$ independent of m, such that if

$$(x_1, \ldots, x_m) \in B(y_1^m, (Cm)^{-1}\varepsilon) \times \cdots \times B(y_m^m, (Cm)^{-1}\varepsilon),$$

then $\Psi_m(x_1, \ldots, x_m) \in B(g, \varepsilon)$.

Now define for $i = 1, \ldots, m$ the vectors

$$\tilde{y}_i^m := \int_{\frac{i-1}{m}}^{\frac{i}{m}} \dot{\gamma}(t)\, \mathrm{d}t.$$

By Proposition 4.5.8, for every m large enough there exists a constant L_m such that for $i = 1, \ldots, m$ we have

$$|y_i^m - \tilde{y}_i^m| \leqslant L_m \frac{1}{m} ||\dot{\gamma}||_\infty$$

and $\lim_{m\to\infty} L_m = 0$. It follows that $B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon) \subset B(y_i^m, (Cm)^{-1}\varepsilon)$ for m large enough. We conclude that if

$$(x_1, \ldots, x_m) \in B(\tilde{y}_1^m, (2Cm)^{-1}\varepsilon) \times \cdots \times B(\tilde{y}_m^m, (2Cm)^{-1}\varepsilon),$$

then $\Psi_m(x_1, \ldots, x_m) \in B(g, \varepsilon)$.
Now, let $\left(Y_{\lfloor m^{-1}n\rfloor}^{n,m,1}, \ldots, Y_{\lfloor m^{-1}n\rfloor}^{n,m,m}\right)$ be again as in (4.5.5), so that

$$\Psi_m\left(\left(Y_{\lfloor m^{-1}n\rfloor}^{n,m,1}, \ldots, Y_{\lfloor m^{-1}n\rfloor}^{n,m,m}\right)\right) = \sigma_n^n.$$

Using the above, we have

$$\begin{aligned}
\mathbb{P}\left(\sigma_n^n \in U\right) &\geqslant \mathbb{P}\left(\left(Y_{\lfloor m^{-1}n\rfloor}^{n,m,1}, \ldots, Y_{\lfloor m^{-1}n\rfloor}^{n,m,m}\right) \in \prod_{i=1}^m B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon)\right) \\
&= \prod_{i=1}^m \mathbb{P}\left(Y_{\lfloor m^{-1}n\rfloor}^{n,m,i} \in B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon)\right) \\
&= \prod_{i=1}^m \mathbb{P}\left(Y_{\lfloor m^{-1}n\rfloor}^{n,m,1} \in B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon)\right).
\end{aligned}$$

Here we used again the fact that $Y_{\lfloor m^{-1}n\rfloor}^{n,m,1}, \ldots, Y_{\lfloor m^{-1}n\rfloor}^{n,m,m}$ are independent and identically distributed, which follows from the fact that the sequence $\{X_n\}_{n\geqslant 1}$ is independent, identically distributed, together with expression (4.5.4).
Continuing, it follows from Proposition 4.5.1 that

$$\left|Y_{\lfloor m^{-1}n\rfloor}^{n,m,1} - \frac{1}{n}\sum_{j=1}^{\lfloor m^{-1}n\rfloor} X_j\right| \leqslant C_m \frac{1}{m},$$

where $\lim_{m\to\infty} C_m = 0$. As a consequence, for m large enough we have

$$\left|Y_{\lfloor m^{-1}n\rfloor}^{n,m,1} - \frac{1}{n}\sum_{j=1}^{\lfloor m^{-1}n\rfloor} X_j\right| \leqslant (2Cm)^{-1}\frac{\varepsilon}{2}.$$

In that case we find that

$$\mathbb{P}\left(Y_{\lfloor m^{-1}n\rfloor}^{n,m,1} \in B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon)\right) \geqslant \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{\lfloor m^{-1}n\rfloor} X_j \in B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon/2)\right).$$

By Cramér's theorem for vector spaces (Theorem 2.1.10), it follows that the sequence $\{\frac{1}{n}\sum_{j=1}^{\lfloor m^{-1}n\rfloor} X_j\}_{n\geqslant 0}$ satisfies the large deviation principle in $\mathfrak{g}$ with good rate function $I_m(x) = \frac{1}{m}\Lambda^*(mx)$. Hence, we obtain that

$$\liminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}\left(\sigma_n^n \in U\right) \geqslant \sum_{i=1}^m \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(Y_{\lfloor m^{-1}n\rfloor}^{n,m,1} \in B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon)\right)$$

$$\begin{aligned}
&\geqslant \sum_{i=1}^{m} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{\lfloor m^{-1}n\rfloor} X_j \in B(\tilde{y}_i^m, (2Cm)^{-1}\varepsilon/2)\right)\\
&\geqslant \sum_{i=1}^{m} -I_m(\tilde{y}_i^m)\\
&= -\frac{1}{m}\sum_{i=1}^{m} \Lambda^*(m\tilde{y}_i^m).
\end{aligned}$$

We are done once we show that

$$\frac{1}{m}\sum_{i=1}^{m} \Lambda^*(m\tilde{y}_i^m) \leqslant \int_0^1 \Lambda^*(\dot{\gamma}(t))\,\mathrm{d}t.$$

By the convexity of Λ^* and Jensen's inequality, we have

$$\Lambda^*(m\tilde{y}_i^m) = \Lambda^*\left(m\int_{\frac{i-1}{m}}^{\frac{i}{m}} \dot{\gamma}(t)\,\mathrm{d}t\right) \leqslant m\int_{\frac{i-1}{m}}^{\frac{i}{m}} \Lambda^*(\dot{\gamma}(t))\,\mathrm{d}t.$$

From this it follows that

$$\frac{1}{m}\sum_{i=1}^{m} \Lambda^*(m\tilde{y}_i^m) \leqslant \sum_{i=1}^{m}\int_{\frac{i-1}{m}}^{\frac{i}{m}} \Lambda^*(\dot{\gamma}(t))\,\mathrm{d}t = \int_0^1 \Lambda^*(\dot{\gamma}(t))\,\mathrm{d}t,$$

which concludes the proof. □

5

Path-space large deviations in Riemannian manifolds

This chapter focusses on path space large deviation results in Riemannian manifolds. We prove the analogue of Mogulskii's theorem (Theorem 2.1.13), i.e., path space large deviations for trajectories of geodesic random walks as defined in Chapter 3. The only difference is that in this chapter, we consider the moment generating function of an increment as function on the cotangent space, see Section 5.1. It turns out this is more natural in light of our method for proving the result. Furthermore, we provide two novel approaches to obtain the generalization of Schilder's theorem (Theorem 2.1.14) for Riemannian Brownian motion with small variance. The results presented in this chapter are based on:

Richard C. Kraaij, Frank Redig, and Rik Versendaal. "Classical large deviation theorems on complete Riemannian manifolds". In: *Stochastic Process. Appl.* 129.11 (2019), pp. 4294–4334. ISSN: 0304-4149. DOI: `10.1016/j.spa.2018.11.019`. URL: `https://doi.org/10.1016/j.spa.2018.11.019`.

As explained in Section 2.1.2, Mogulskii's theorem is the natural path space large deviations result accompanying Cramér's theorem. More precisely, given a sequence $\{X_n\}_{n\geqslant 1}$ of independent, identically distributed random variables, Moguslkii's theorem provides the large deviation principle for the trajectories $\{S_n(\cdot)\}_{n\geqslant 1}$ given by

$$S_n(t) = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} X_i.$$

To obtain an analogue of Mogulskii's theorem for Riemannian manifolds, we make use of geodesic random walks (see Chapter 3), which extend the notion of a random walk to Riemannian manifolds. For every $n \geqslant 1$, let $\{(\frac{1}{n} * \mathcal{S})_j\}_{j\geqslant 1}$ be a $\frac{1}{n}$-rescaled geodesic random walk started in x_0 with increments $\{X_j^n\}_{n\geqslant 1}$ as in Definition 3.2.6. If the increments are independent and identically distributed in the sense of Definitions 3.2.3 and 3.2.7, Theorem 3.3.1 gives the large deviation principle in M for the sequence $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 1}$. For the associated path space large deviations, we consider the random trajectories

$$Z_n(t) = \left(\frac{1}{n} * \mathcal{S}\right)_{\lfloor nt \rfloor}.$$

Under the same conditions on the increments as in Theorem 3.3.1 (apart from the increments being centered), we obtain the large deviation principle for the trajectories $\{Z_n(\cdot)\}_{n\geqslant 1}$. The precise statement is given in Theorem 5.1.1.

Contrary to the analogue of Moguslkii's theorem for Riemannian manifolds, the analogue of Schilder's theorem for Riemannian Brownian motion has been considered before, see e.g. [9, 41]. In this chapter, we provide two novel approaches in obtaining this result. These approaches are interesting in their own right, and find applications beyond Schilder's theorem.

To prove the analogues of Mogulskii's and Schilder's theorem for Riemannian manifolds, we make use of a general approach for studying large deviations for Markov processes introduced by Feng and Kurtz in [39]. This approach relies on the convergence of non-linear semigroups and viscosity solutions to Hamilton-Jacobi equations. Since the details of this approach are beyond the scope of this work, we only collect the relevant results from Section 7 in [63].

For Schilder's theorem, we also discuss a second approach using embeddings of manifolds into Euclidean space. This is particularly useful when the process being studied is the solution of a stochastic differential equation. As explained in Section 2.4.2, Stratonovich stochastic differential equations behave well under diffeomorphisms. Therefore, using the embedding allows us to transfer the problem from the manifold to the Euclidean setting, in which we can apply Freidlin-Wentzell theory (see Section 2.1.3). The importance of this approach lies in Chapter 7, where we extend the result further to time-evolving Riemannian manifolds.

This chapter is organized as follows. In Section 5.1 we give the precise statements of the analogues of Mogulskii's and Schilder's theorem for Riemannian manifolds. In Section 5.2 we collect the important results from the Feng-Kurtz approach to studying large deviations for Markov processes. Section 5.3 is devoted to showing how these results can be applied to prove Theorem 5.1.1 and 5.1.3. Finally, in Section 5.4 we show how we can use embeddings to provide a different proof of Schilder's theorem for Riemannian Brownian motion.

5.1. Main results

In this section we state the analogues of Mogulskii's and Schilder's theorem for Riemannian manifolds. Furthermore, we touch on the relations between Mogulskii's theorem to Cramér's theorem (Theorem 3.3.1). Finally, we work out an example to show the relations between all three theorems in the case when we consider geodesic random walks with normally distributed increments.

In the Euclidean case, the rate function in Mogulskii's theorem is given by

$$\mathcal{I}(\gamma) = \begin{cases} \int_0^1 \Lambda^*(\dot\gamma(t))\,\mathrm{d}t, & \gamma \in AC_0([0,1];\mathbb{R}^d) \\ \infty, & \text{otherwise.} \end{cases} \tag{5.1.1}$$

Here, $\Lambda(\lambda) = \log \mathbb{E}(e^{\langle \lambda, X_1\rangle})$ is the log-moment generating function of the increments, and

$$\Lambda^*(v) = \sup_{\lambda\in\mathbb{R}^d} \{\langle \lambda, v\rangle - \Lambda(\lambda)\}$$

is its Legendre transform.

However, in the Riemannian setting, the distribution of an increments depends on the position of the geodesic random walk in M. More precisely, we have a collection of measures $\{\mu_x\}_{x\in M}$ with $\mu_x \in \mathcal{P}(T_xM)$, where $\mathcal{P}(T_xM)$ denotes the set of probability measures on T_xM. We thus have for every $x \in M$ a log-moment generating function $\Lambda_x : T_x^*M \to \mathbb{R}$ given by

$$\Lambda_x(\lambda) = \log \int_{T_xM} e^{\langle v, \lambda\rangle} \mu_x(\mathrm{d}v)$$

with Legendre transform given by

$$\Lambda_x^*(v) = \sup_{\lambda\in T_xM} \{\langle v, \lambda\rangle - \Lambda_x(\lambda)\}\,.$$

Here, $\langle v, \lambda\rangle$ denotes the pairing of the cotangent vector λ with the tangent vector v. We sometimes also denote this as $\lambda(v)$. Observe that in Chapter 3 we defined the log-moment generating function Λ_x as function on T_xM, rather than T_x^*M. However, these functions are essentially the same if we identify T_x^*M with T_xM using the Riemannian metric. More precisely, abusing notation and writing Λ_x for both functions, we have that

$$\Lambda_x(\lambda) = \Lambda_x(\lambda^\#) \tag{5.1.2}$$

for all $\lambda \in T_x^*M$. Here $\lambda^\#$ is the unique tangent vector such that

$$\lambda(v) = \langle \lambda^\#, v\rangle$$

for all $v \in T_xM$ denotes the tangent vector associated to λ via the inner product, see Section 2.2.3.

For a curve $\gamma : [0,1] \to M$, we have that $\dot\gamma(t) \in T_{\gamma(t)}M$. Therefore, the appropriate analogue of the rate function in (5.1.1) is given by

$$\begin{cases} \int_0^1 \Lambda^*_{\gamma(t)}(\dot\gamma(t))\,\mathrm{d}t, & \gamma \in AC_{x_0}([0,1];M) \\ \infty, & \text{otherwise.} \end{cases}$$

We have the following theorem.

Theorem 5.1.1 (Mogulskii's theorem for Riemannian manifolds)**.** *Let (M,g) be a complete Riemannian manifold. Fix $x_0 \in M$ and let $\{\mu_x\}_{x\in M}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_xM)$ for all $x \in M$. For every $n \geqslant 1$, let $\{(\frac{1}{n} * \mathcal{S})_j\}_{j\geqslant 0}$ be a $\frac{1}{n}$-rescaled geodesic random walk started at x_0 with independent increments $\{X_j^n\}_{j\geqslant 1}$, compatible with $\{\mu_x\}_{x\in M}$. Define the random trajectories*

$$Z_n(t) = \left(\frac{1}{n} * \mathcal{S}\right)_{\lfloor nt \rfloor}$$

for $t \in [0,1]$. Assume the collection $\{\mu_x\}_{x\in M}$ satisfies the consistency property in Definition 3.2.7. Finally, assume the increments are bounded. Then $\{Z_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $L^\infty([0,1];M)$ with good rate function

$$\mathcal{I}_M(\gamma) = \begin{cases} \int_0^1 \Lambda^*_{\gamma(t)}(\dot\gamma(t))\,\mathrm{d}t, & \gamma \in AC_{x_0}([0,1];M) \\ \infty, & \textit{otherwise.} \end{cases} \tag{5.1.3}$$

It is interesting to observe how Theorem 5.1.1 relates to Cramér's theorem (Theorem 3.3.1). Since evaluation in the end-point of a trajectory is continuous, it follows from the contraction principle (Theorem 2.1.6) that Mogulskii's theorem implies the large deviations for the sequence $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 1}$. Furthermore, this shows that the rate function is given by

$$\tilde I_M(x) = \inf\left\{ \int_0^1 \Lambda^*_{\gamma(t)}(\dot\gamma(t))\,\mathrm{d}t \,\middle|\, \gamma \in AC_{x_0}([0,1];M),\ \gamma(1) = x \right\}.$$

For every $x \in M$, we have that Λ_x is convex, and hence, so is Λ_x^*. Furthermore, by Proposition 3.2.9 we have that the maps $\{\Lambda_x\}_{x\in M}$ are invariant under parallel transport, and hence, so are the maps $\{\Lambda_x^*\}_{x\in M}$. As a consequence, apart from some technicalities, it follows from Proposition 3.4.11 that the optimal trajectories for $\tilde I_M(x)$ are given by geodesics. This shows that we obtain the desired rate function as stated in Theorem 3.3.1. Let us illustrate this connection with an example.

Example 5.1.2. Let (M,g) be a compact Riemannian manifold. Let $\{\mu_x\}_{x\in M}$ be the collection of standard normal distributions as defined in Example 3.2.11. There it was shown that these measures satisfy the consistency property as in Definition 3.2.7. Furthermore, we have that $\Lambda_x(\lambda) = \frac{1}{2}|\lambda|^2_{g(x)}$, from which it follows that $\Lambda_x^*(v) = \frac{1}{2}|v|^2_{g(x)}$.
For every $n \geqslant 1$, let $\{(\frac{1}{n} * \mathcal{S})_j\}_{j\geqslant 1}$ be a $\frac{1}{n}$-rescaled geodesic random walk with increments compatible with the measures $\{\mu_x\}_{x\in M}$. By Cramér's theorem (Theorem 3.3.1), the sequence $\{(\frac{1}{n} * \mathcal{S})_n\}_{n\geqslant 1}$ satisfies the large deviation principle with good rate function

$$I_M(x) = \inf\left\{ \frac{1}{2}|v|^2_{x_0} \,\middle|\, \mathrm{Exp}_{x_0} v = x \right\}.$$

Note that $|v|_{g(x_0)}$ is the length of the geodesic $\gamma_v : [0,1] \to M$ given by $\gamma_v(t) = \mathrm{Exp}_{x_0}(tv)$. Since we take the infimum over all possible v, we also consider the geodesic of minimal length between x_0 and x and hence

$$I_M(x) = d(x, x_0)^2.$$

Furthermore, by Mogulskii's theorem (Theorem 5.1.1), we find that the process $Z_n(t) = \frac{1}{n} * \mathcal{S}_{\lfloor nt \rfloor}$ satisfies the large deviation principle in $L^\infty([0,1], M)$ with good rate function

$$\mathcal{I}(\gamma) = \begin{cases} \frac{1}{2}\int_0^\infty |\dot{\gamma}(t))|^2_{g(\gamma(t))}\,\mathrm{d}t, & H^1_{x_0}([0,1];M), \\ \infty & \textit{otherwise.} \end{cases}$$

As explained, by the contraction principle, we also have that

$$I_M(x) = \inf\left\{\frac{1}{2}\int_0^1 |\dot{\gamma}(t))|^2_{g(\gamma(t))}\,\mathrm{d}t \middle| \gamma \in AC_{x_0}([0,1];M),\ \gamma(1) = x\right\}.$$

By Jensen's inequality and the definition of the Riemannian distance, the right-hand side is indeed equal to $d(x, x_0)^2$.

In Section 2.1.3 we discussed that if we take the increments of a random walk to be standard normal, we can use Mogulskii's theorem to obtain the large deviations for the $\{\frac{1}{\sqrt{n}}W(\cdot)\}_{n \geqslant 1}$, where $W(t)$ is a standard Brownian motion. It is not clear if a similar approach works in the Riemannian setting. Indeed, the increments of Brownian motion are no longer normally distributed (in the sense of Example 3.2.11) with the desired parameters due to curvature.
However, Varadhan ([93]) showed that the short-time asymptotics of the heat kernel on a Riemannian manifold are given by

$$\lim_{t \to 0} t \log p_M(x, y, t) = -\frac{d^2(x,y)}{2}. \tag{5.1.4}$$

This suggests that for short times, the 'increments' of Riemannian Brownian motion are almost normally distributed. With the computations in Example 5.1.2 in mind, it turns out that this provides the correct intuition.

Theorem 5.1.3 (Schilder's theorem for Riemannian Brownian motion)**.** *Let (M, g) be a complete Riemannian manifold. Assume furthermore that (M, g) is stochastically complete. Let $x_0 \in M$ and let $X(t)$ be a Riemannian Brownian motion and with $X(0) = x_0$ almost surely. Define for every $n \geqslant 1$ the process $X_n(t) := X(\frac{t}{n})$. Then the sequence $\{X_n(\cdot)\}_{n \geqslant 1}$ satisfies the large deviation principle in $C([0,1]; M)$ with good rate function*

$$\mathcal{I}_{BM}(\gamma) = \begin{cases} \frac{1}{2}\int_0^\infty |\dot{\gamma}(t)|^2_{g(\gamma(t))}\,\mathrm{d}t, & \gamma \in H^1_{x_0}([0,1];M), \\ \infty, & \text{otherwise.} \end{cases}. \tag{5.1.5}$$

Remark 5.1.4. The short-time asymptotics of the heat kernel as in (5.1.4) can be used in the Euclidean case to prove the large deviation principle for processes generated by a weighted Laplacian. However, in general, these conditions are not satisfied by a Riemannian metric. Nonetheless, similarly as done in the proof of Lemma 3.1 in [93], one can use (5.1.4) to obtain the large deviations for the finite dimensional distributions of Brownian paths once Gaussian bounds for the heat kernel for a general (stochastically complete) Riemannian manifold are established (see e.g. [6]). Using Proposition 3.7 in [63] (which replaces Lemma 3.2 in [93]), one can follow the argument in proving Theorem 3.3 in [93] to obtain the large deviations upper bound in Schilder's theorem. For the lower bound, one can exactly mimic the proof of Lemma 3.4 in [93]. We show that all assumptions, apart from the stochastic completeness may be dropped, see Theorem 5.1.3.

5.2. Large deviations via Hamilton-Jacobi equations

In this section we explain the steps of the approach introduced in [39] to study large deviations for Markov processes. This approach is based on convergence of non-linear semigroups and solving Hamilton-Jacobi equations in viscosity sense. For general theory on viscosity solution for Hamilton-Jacobi equations we refer to [26]. For an extensive treatment of the relation to large deviations in the Euclidean case, apart from [39], we also refer to the Appendix in [22]. In [63], this approach has been adapted to the setting of Riemannian manifolds. Here, we state some of the results from Section 7 in [63] that we need in order to prove Theorems 5.1.1 and 5.1.3.

5.2.1. Comparison principle for Hamilton-Jacobi equations

Let $H : \mathcal{D}(H) \subset C_b(M) \to C_b(M)$ be an operator. For $h \in C_b(M)$ and $\lambda > 0$, consider the *Hamilton-Jacobi equation*

$$f - \lambda H f = h. \tag{5.2.1}$$

We want to solve (5.2.1) in the viscosity sense. We have the following definition.

Definition 5.2.1. *A function u is a* viscosity subsolution *of equation* (5.2.1) *if u is bounded, upper semi-continuous and if for every $f \in \mathcal{D}(H)$ there exists a sequence $x_n \in M$ such that*

$$\lim_{n\to\infty} u(x_n) - f(x_n) = \sup_x \{u(x) - f(x)\},$$

and

$$\lim_{n\to\infty} u(x_n) - \lambda H f(x_n) - h(x_n) \leqslant 0.$$

A function v is a viscosity supersolution *of equation* (5.2.1) *if v is bounded, lower semi-continuous and if for every $f \in \mathcal{D}(H)$ there exists a sequence $x_n \in M$ such that*

$$\lim_{n\to\infty} v(x_n) - f(x_n) = \inf_x \{v(x) - f(x)\},$$

and

$$\lim_{n\to\infty} v(x_n) - \lambda H f(x_n) - h(x_n) \geqslant 0.$$

A function u is a viscosity solution *of equation* (5.2.1) *if it is both a viscosity sub- and supersolution.*

Definition 5.2.2. *We say that* (5.2.1) *satisfies the* comparison principle *if for a subsolution u and supersolution v we have $u \leqslant v$.*

Note that if the comparison principle is satisfied, then a viscosity solution is unique.

From now on we assume that the Hamiltonian H has a special form. More precisely, it should be possible to represent H by a map on the cotangent bundle.

Assumption 5.2.3. *The operator $H : \mathcal{D}(H) \subset C_b(M) \to C_b(M)$ satisfies $C_c^\infty(M) \subseteq \mathcal{D}(H) \subseteq C_b(M) \cap C^1(M)$ and can be represented as*

$$Hf(x) = \mathcal{H}(x, \mathrm{d}f(x)),$$

where $\mathcal{H} : T^*M \to \mathbb{R}$ is continuous and for each $x \in M$ the map $p \mapsto \mathcal{H}(x,p)$ from T_x^*M to $\mathbb{R}$ is convex.

We now wish to state a sufficient condition for the comparison principle. The idea is that we want to use a subsolution u and supersolution v as test functions in Definition 5.2.1, and so obtain conditions on the map $\mathcal{H}$ evaluated in $\mathrm{d}u$ and $\mathrm{d}v$ in order to conclude that $u \leqslant v$. This is explained in detail in [63, Section 7.2]. However, this approach relies on the fact that u and v are test functions, and that there exist points x_0, y_0 such that $u(x_0) - v(x_0) = \sup_x\{u(x) - v(x)\}$ and $v(y_0) - u(y_0) = \inf_y\{v(y) - u(y)\}$. The first issue can be resolved by penalizing by a distance function, in this case the Riemannian distance. The second problem is taken care of by restricting to compact sets. For this we use what we call a compact containment function.

Definition 5.2.4. *A function $\Upsilon : M \to \mathbb{R}$ is a* good containment function *for H if it satisfies the following:*

1. *$\Upsilon \geqslant 0$ and there exists a point x_0 such that $\Upsilon(x_0) = 0$,*
2. *Υ is twice continuously differentiable,*
3. *for every $c \geqslant 0$, the set $\{x \in M \mid \Upsilon(x) \leqslant c\}$ is compact,*
4. *we have $\sup_z \mathcal{H}(z, \mathrm{d}\Upsilon(z)) < \infty$.*

Using the Riemannian distance to penalize and a good compact containment function, we obtain the following sufficient condition on $\mathcal{H}$ for the comparison principle.

Proposition 5.2.5. *Let H be an operator satisfying Assumption 5.2.3. Fix $\lambda > 0$, $h \in C_b(M)$ and consider u and v a sub- and super-solution to $f - \lambda Hf = h$. Let Υ be a good containment function. Moreover, for every $\alpha, \varepsilon > 0$ let $x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon} \in M$ be such that*

$$\frac{u(x_{\alpha,\varepsilon})}{1-\varepsilon} - \frac{v(y_{\alpha,\varepsilon})}{1+\varepsilon} - \frac{\alpha}{2}d^2(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}) - \frac{\varepsilon}{1-\varepsilon}\Upsilon(x_{\alpha,\varepsilon}) - \frac{\varepsilon}{1+\varepsilon}\Upsilon(y_{\alpha,\varepsilon})$$
$$= \sup_{x,y\in M}\left\{\frac{u(x)}{1-\varepsilon} - \frac{v(y)}{1+\varepsilon} - \frac{\alpha}{2}d^2(x,y) - \frac{\varepsilon}{1-\varepsilon}\Upsilon(x) - \frac{\varepsilon}{1+\varepsilon}\Upsilon(y)\right\}.$$

Suppose that

$$\liminf_{\varepsilon\to 0}\liminf_{\alpha\to\infty}\left\{\mathcal{H}\left(x_{\alpha,\varepsilon}, \frac{\alpha}{2}\mathrm{d}d^2(\cdot, y_{\alpha,\varepsilon})(x_{\alpha,\varepsilon})\right)\right.$$
$$\left. - \mathcal{H}\left(y_{\alpha,\varepsilon}, -\frac{\alpha}{2}\mathrm{d}d^2(x_{\alpha,\varepsilon}, \cdot)(y_{\alpha,\varepsilon})\right)\right\} \leqslant 0. \quad (5.2.2)$$

Then $u \leqslant v$, i.e. the Hamilton-Jacobi equation $f - \lambda H f = h$ satisfies the comparison principle.

Remark 5.2.6. Although the square of the Riemannian distance is not everywhere smooth, this is not a problem in Proposition 5.2.5. The squared Riemannian distance is smooth at points which are inside each others injectivity radius. Now, by Lemma 7.6 in [63], it follows that $\{x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon} | \alpha > 0\}$ is contained in a compact set for every $\varepsilon > 0$. Furthermore, we have that $\lim_{\alpha\to\infty} d(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}) = 0$ for all $\varepsilon > 0$. Since the injectivity radius of compact sets is strictly positive (Proposition 2.2.6), we thus find that for α large enough, the points $x_{\alpha,\varepsilon}$ and $y_{\alpha,\varepsilon}$ are within each others injectivity radius.

5.2.2. Compact containment and the large deviation principle

To connect the Hamilton-Jacobi equation to the large deviation principle, we introduce some additional concepts. First of all, we denote the Skorokhod space of càdlàg paths by $D([0,1]; M)$, see [37, Section 3.5]. Furthermore, we introduce a notion of operator convergence for which we consider bounded and uniform convergence on compact sets (buc). We define this next.

Definition 5.2.7. *Let $\{f_n\}_{n\geqslant 1}$ be a sequence in $C_b(M)$ and let $f \in C_b(M)$. We say that f_n converges to f boundedly, and uniformly on compacts, denoted by* $\mathrm{LIM}_n f_n = f$, *if the following are satisfied:*

1. $\sup_n ||f_n|| < \infty$,
2. *For all $K \subseteq M$ compact,*

$$\lim_{n\to\infty}\sup_{x\in K}|f_n(x) - f(x)| = 0.$$

We now define our notions of operator convergence.

Definition 5.2.8. *For every $n \geqslant 1$, let $B_n : \mathcal{D}(B_n) \subset C_b(M) \to C_b(M)$ be an operator. The* extended limit $ex - \lim_{n\to\infty} B_n$ *is defined by the collection $(f,g) \in C_b(M)\times C_b(M)$ such that there exist a sequence $\{f_n\}_{n\geqslant 1}$ with $f_n \in \mathcal{D}(B_n)$ and such that*

$$\mathrm{LIM}_n f_n = f, \qquad \mathrm{LIM}\, B_n f_n = g.$$

An operator $(B, \mathcal{D}(B))$ *is said to be contained in* $ex - \lim_{n\to\infty} B_n$ *if the graph* $\{(f, Bf) \mid f \in \mathcal{D}(B)\}$ *of* B *is a subset of* $ex - \mathrm{LIM}_n B_n$.

Before we get to any results relating Hamilton-Jacobi equations to large deviation principles, we first need to define the operators that we will be considering.

Assumption 5.2.9. *Depending on whether we consider Markov processes in continuous time or discrete time, we consider the following:*

Continuous time case *Assume that for each* $n \geqslant 1$, *we have a linear operator* $A_n \subseteq C_b(M) \times C_b(M)$ *and existence and uniqueness holds for the* $D([0,1], M)$ *martingale problem for* (A_n, μ) *for each initial distribution* $\mu \in \mathcal{P}(M)$. *Letting* $\mathbb{R}^n_y \in \mathcal{P}(D([0,1], M))$ *be the solution to the martingale problem for* (A_n, δ_y), *the mapping* $y \mapsto \mathbb{R}^n_y$ *is measurable for the weak topology on* $\mathcal{P}(D([0,1], M))$. *Let* X_n *be the solution to the martingale problem for* A_n *and set*

$$H_n f = \frac{1}{n} e^{-nf} A_n e^{nf} \qquad e^{nf} \in \mathcal{D}(A_n).$$

Discrete time case *Assume for each* $n \geqslant 1$ *we have a transition operator* $T_n : C_b(M) \to C_b(M)$ *for a Markov chain. For each* n, *let* X_n *be a discrete-time Markov chain with transition operator* T_n:

$$\mathbb{E}\left[f(X_n(t)) \mid X_n(0) = x\right] = T_n^{\lfloor nt \rfloor} f(x).$$

Set

$$H_n f = \log e^{-nf} T_n e^{nf}.$$

Suppose that we have an operator $H : \mathcal{D}(H) \subseteq C_b(M) \to C_b(M)$ *with* $\mathcal{D}(H) = C_c^\infty(M)$ *and* $H \subseteq ex - \mathrm{LIM}\, H_n$ *which satisfies Assumption 5.2.3. Finally, assume that the map* $\mathcal{H} : T^*M \to \mathbb{R}$ *is continuously differentiable.*

The following result is concerned with the limiting behaviour of the probability of sequence of processes to stay in compact sets. It is Proposition 7.15 in [63].

Proposition 5.2.10. *Suppose Assumption 5.2.9 is satisfied and assume that* Υ *is a good containment function for* H. *Then the sequence* $\{X_n\}_{n\geqslant 1}$ *satisfies the* exponential compact containment *condition: for every* $T > 0$ *and* $a \geqslant 0$, *there exists a compact set* $K_{a,T} \subseteq M$ *such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[X_n(t) \notin K_{a,T} \text{ for some } t \leqslant T\right] \leqslant -a.$$

We conclude with the result giving us conditions for the large deviation principle to hold for a sequence of Markov processes with generators satisfying Assumption 5.2.9.

Theorem 5.2.11. *Consider the setting of Assumption 5.2.9. We have the following:*

(a) *Suppose that* Υ *is a good containment function for* H. *Then the processes* $\{X_n\}_{n\geqslant 1}$ *are exponentially tight in* $D([0,1],M)$.

(b) *In addition to the assumption in (a), suppose that for each* $\lambda > 0$ *and* $h \in C_b(M)$ *the comparison principle is satisfied for* $f - \lambda H f = h$. *Then the sequence* $\{X_n\}_{n\geqslant 1}$ *satisfies the large deviation principle in* $D([0,1];M)$ *with good rate function* I *given by*

$$I(\gamma) = \begin{cases} \int_0^\infty \mathcal{L}(\gamma(s),\dot\gamma(s))\,\mathrm{d}s & \text{if } \gamma \in AC([0,1];M), \\ \infty & \text{otherwise.} \end{cases}$$

where $\mathcal{L} : TM \to [0,\infty]$ *is the Legendre transform of* $\mathcal{H}$ *given by*

$$\mathcal{L}(x,v) = \sup_{p\in T_x^* M} \left\{ \langle v,p\rangle - \mathcal{H}(x,p) \right\}.$$

5.3. Classical large deviations in Riemannian manifolds via the Feng-Kurtz formalism

In this section we prove Theorems 5.1.1 and 5.1.3 using the Feng-Kurtz approach discussed in Section 5.2. Before doing so we construct a good containment function that we will use for both proofs.

5.3.1. Good containment function

In order to use Theorem 5.2.11 to prove the analogues for Mogulskii's and Schilder's theorem for Riemannian manifolds, we need a good containment function. We construct one containment function that will suffice for both proofs. We use the following proposition, which essentially follows from the fact that the function $r(x) = d(x,x_0)$ is 1-Lipschitz.

Proposition 5.3.1. *Let* (M,g) *be a complete Riemannian manifold. Fix* $x_0 \in M$ *and define* $r(x) := d(x,x_0)$. *There exists a smooth function* $f \in C^\infty(M)$ *such that* $||f - r|| \leqslant 1$ *and* $|\mathrm{d}f| \leqslant 2$.

Consider the function f as in the above proposition and set

$$\Upsilon(x) = \log(1 + f^2(x)). \tag{5.3.1}$$

Lemma 5.3.2. *Let* $\{\mu_x\}_{x\in M}$ *be a collection of measures such that* $\mu_x \in \mathcal{P}(T_x M)$ *for all* $x \in M$. *Assume that* $\{\mu_x\}_{x\in M}$ *satisfies the consistency property in Definition 3.2.7. Let* $\mathcal{H}$ *be given by*

$$\mathcal{H}(x,p) = \log \int_{T_x M} e^{\langle v,p\rangle} \mu_x(\mathrm{d}v). \tag{5.3.2}$$

Then Υ *given in* (5.3.1) *is a good containment function for* $\mathcal{H}$.

Proof. Clearly $\Upsilon \geqslant 0$, $\Upsilon(x_0) = 0$ and $\Upsilon \in C^\infty(M)$.
Now fix $c \geqslant 0$. By the continuity of Υ, the set $\{x \in M \mid \Upsilon(x) \leqslant c\}$ is closed. Furthermore, the set is bounded since $\Upsilon(x) \leqslant c$ implies that $d(x, x_0) \leqslant 1+\sqrt{e^c - 1}$. Because M is a complete, finite dimensional manifold, it follows that $\{x \in M \mid \Upsilon(x) \leqslant c\}$ is compact.
Now consider the Hamiltonian $\mathcal{H}$ in (5.3.2). Note that for all $x \in M \backslash \{x_0\}$,

$$\mathrm{d}\Upsilon(x) = \frac{2f(x)}{1+f^2(x)}\mathrm{d}f(x).$$

This implies that $|\mathrm{d}\Upsilon(x)| \leqslant |\mathrm{d}f(x)| \leqslant 2$, where the latter holds by choice of f. But then

$$\mathcal{H}(x, \mathrm{d}\Upsilon(x)) = \log \int_{T_xM} e^{\langle v, \mathrm{d}\Upsilon(x)\rangle} \mu_x(\mathrm{d}v) \leqslant \log \int_{T_xM} e^{2|v|_{g(x)}} \mu_x(\mathrm{d}v) =: C_x < \infty,$$

where C_x is finite because we assume the log moment generating function of μ_x is everywhere finite. By the consistency property (as in Definition 3.2.7), C_x actually does not depend on x. We conclude that $\sup_{x\in M} \mathcal{H}(x, \mathrm{d}\Upsilon(x)) < \infty$. □

5.3.2. Proof of Mogulskii's Theorem, Theorem 5.1.1

In this section we prove the analogue of Mogulskii's theorem for rescaled geodesic random walks. Before we can get to the proof, we first need the following result.

Proposition 5.3.3. *Let $x, y \in M$ and assume that $d(x, y) < \iota(x)$. Then for all $v \in T_yM$ we have*

$$\mathrm{d}_y d^2(x, y)(v) = 2\langle \dot\gamma(1), v\rangle_{g(y)},$$

where $\gamma : [0, 1] \to M$ is the unique geodesic of minimal length connecting x and y. Moreover, we have

$$\tau_{xy}\mathrm{d}_x d^2(x, y) = -\mathrm{d}_y d^2(x, y).$$

Proof. For a path $h : [0, 1] \to M$, define the Lagrangian

$$L(h(t)) = \langle \dot h(t), \dot h(t)\rangle_{g(h(t))} = |\dot h(t)|^2_{g(h(t))}$$

and the action

$$S(h) = \int_0^1 L(h(t))\,\mathrm{d}t.$$

Observe that for $x, y \in M$ we have

$$d^2(x, y) = \inf\{S(h) | h(0) = x, h(1) = y, h \text{ piecewise smooth}\}.$$

Since $d(x, y) < \iota(x)$, there is an optimal path $\gamma : [0, 1] \to M$ for S, the geodesic of minimal length connecting x and y. Note that the differential of the action in the starting point equals the momentum of the optimal path γ in 0. For this we refer to [69, Chapter 6] or [77, Chapter 5] for an approach using variational calculus and to [8, Chapter 3] for the physical intuition. In coordinates one finds that the j-th component of this momentum equals

$$p_j = \frac{\partial L}{\partial \dot{h}^j(t)}(\gamma(t)) = 2\sum_{i=1}^{k} g_{ij}(\gamma(t))\dot{\gamma}^i(t) = 2(\dot{\gamma}(t))^\flat_j.$$

Here, $v^\flat \in T_x^*M$ denotes the unique cotangent vector defined by $v^\flat(w) = \langle v, w\rangle$ for all $w \in T_xM$, see also Section 2.2.3.

We thus find that $\mathrm{d}_x d^2(x,y) = 2(\dot{\gamma}(0))^\flat$, where γ is the geodesic of minimum length connecting x and y. In particular, this implies that for every $v \in T_xM$ we have $\mathrm{d}_x d^2(x,y)(v) = 2\langle v, \dot{\gamma}(0)\rangle_{g(x)}$. Defining $\gamma^-(t) := \gamma(1-t)$, γ^- is the geodesic of minimum length connecting y and x. We obtain $\mathrm{d}_y d^2(x,y) = 2(\dot{\gamma}^-(0))^\flat = -2(\dot{\gamma}(1))^\flat$. Noticing that $\dot{\gamma}(1)$ is the parallel transport of $\dot{\gamma}(0)$ now proves the claim. ☐

We are now set to prove Mogulskii's theorem for geodesic random walks.

Proof of Theorem 5.1.1. We verify the conditions for Theorem 5.2.11.

Step 1: We start by calculating H_n and its limit H. From Remark 3.2.5 it follows that for every $n \geqslant 1$ the sequence $\{\frac{1}{n} * \mathcal{S}_k\}_{k\geqslant 1}$ is a Markov chain with transition operator given by

$$T_n f(x) = \mathbb{E}\left(f\left(\frac{1}{n} * \mathcal{S}_{k+1}\right)\middle| \frac{1}{n} * \mathcal{S}_k = x\right) = \int_{T_xM} f(\mathrm{Exp}_x(n^{-1}v))\mu_x(\mathrm{d}v).$$

Using this, for every $n \geqslant 1$ we can compute the Hamiltonian

$$H_n f(x) = \log e^{-nf} T_n e^{nf}(x) = \log \int_{T_xM} e^{n(f(\exp_x(n^{-1}v)-f(x))}\mu_x(\mathrm{d}v).$$

We first establish that $\sup_n ||H_n f|| < \infty$. By the mean value theorem there exists a $t \in (0, n^{-1})$ such that

$$n(f(\exp_x(n^{-1}v) - f(x)) = \mathrm{d}f(\mathrm{Exp}_x(tv))(\tau_{x\mathrm{Exp}_x(tv)}v).$$

In particular, we find

$$|n(f(\exp_x(n^{-1}v) - f(x))| \leqslant ||\mathrm{d}f||_\infty |v|_{g(x)}.$$

For $f \in C_c^\infty(M)$ we have $||\mathrm{d}f||_\infty < \infty$, and by the consistency property from Definition 3.2.7 for $\{\mu_x\}_{x\in M}$ we have

$$H_n f(x) \leqslant \log \int_{T_xM} e^{||\mathrm{d}f||_\infty |v|_{g(x)}}\mu_x(\mathrm{d}v) = \log \int_{T_{x_0}M} e^{||\mathrm{d}f||_\infty |v|_{g(x_0)}}\mu_{x_0}(\mathrm{d}v) := C < \infty,$$

where the upper bound is finite, because Λ_{x_0} is everywhere finite. Similarly, we find

$$H_n f(x) \geqslant \log \int_{T_{x_0}M} e^{-||\mathrm{d}f||_\infty |v|}\mu_{x_0}(\mathrm{d}v) =: c > -\infty.$$

We conclude that $\sup_n ||H_n f|| < \infty$.
Furthermore, by a similar argument, we find for $f \in C_c^\infty(M)$ that

$$Hf(x) := \lim_{n\to\infty} H_n f(x) = \log \int_{T_x M} e^{\langle v, \mathrm{d}f(x)\rangle} \mu_x(\mathrm{d}v)$$

uniformly in x, so that we can take $\mathcal{D}(H) = C_c^\infty(M)$. Note that indeed H has the form $Hf(x) = \mathcal{H}(x, \mathrm{d}f(x))$ for a continuous map $\mathcal{H} : T^*M \to \mathbb{R}$ that is convex in the second coordinate. Here, the convexity follows from Hölder's inequality. This implies that Assumption 5.2.3 is satisfied and $\mathcal{H}$ is given by

$$\mathcal{H}(x,p) = \log \int_{T_x M} e^{\langle v, p\rangle} \mu_x(\mathrm{d}v) = \Lambda_x(p).$$

Step 2: By Lemma 5.3.2 we have a good containment function Υ.
Step 3: Fix $\lambda > 0$ and $h \in C_b(M)$. We verify the comparison principle for $f - \lambda Hf = h$ by an application of Proposition 5.2.5. Let $x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}$ be as in Proposition 5.2.5. We establish (5.2.2).

Fix $\varepsilon > 0$. By Lemma 7.6 in [63] there is a compact set $K^\varepsilon \subseteq M$ such that $\{x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon} | \alpha > 0\}$ is contained in K^ε. Since K^ε is compact, it follows from Proposition 2.2.6 that there exists a $\delta > 0$ such that $\iota(K^\varepsilon) \geqslant \delta > 0$. Now for $x, y \in K^\varepsilon$ with $d(x,y) < \delta$ we find

$$\begin{aligned}\mathcal{H}\left(x, \frac{\alpha}{2}(\mathrm{d}d^2(\cdot,y))(x)\right) &= \Lambda_x\left(\frac{\alpha}{2}(\mathrm{d}d^2(\cdot,y))(x)\right)\\ &= \Lambda_y\left(\frac{\alpha}{2}\tau_{xy}(\mathrm{d}d^2(\cdot,y))(x)\right)\\ &= \Lambda_y\left(-\frac{\alpha}{2}(\mathrm{d}d^2(x,\cdot))(y)\right)\\ &= \mathcal{H}\left(y, -\frac{\alpha}{2}(\mathrm{d}d^2(x,\cdot))(y)\right).\end{aligned}$$

Here τ_{xy} denotes parallel transport along the unique geodesic of minimal length connecting x and y. The second equality follows from proposition 3.2.9 with the identification in (5.1.2) in mind and the third from Proposition 5.3.3. We thus find for $x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}$ with $d(x_{\alpha,\varepsilon}, y_{\alpha,\varepsilon}) < \delta$ that

$$\mathcal{H}\left(x_{\alpha,\varepsilon}, \frac{\alpha}{2}(\mathrm{d}d^2(\cdot, y_{\alpha,\varepsilon}))(x_{\alpha,\varepsilon})\right) - \mathcal{H}\left(y_{\alpha,\varepsilon}, -\frac{\alpha}{2}(\mathrm{d}d^2(x_{\alpha,\varepsilon},\cdot))(y_{\alpha,\varepsilon})\right) = 0.$$

Therefore, by Proposition 5.2.5 we find that H satisfies the comparison principle.

Since the collection of measures $\{\mu_x\}_{x\in M}$ satisfies the consistency property as in Definition 3.2.7, it follows from Proposition 3.2.9 (with the identification as in (5.1.2) in mind) that $\mathcal{H}(x,p) = \Lambda_{x_0}(\tau_{xx_0} p)$ for all $x \in M$ and $p \in T_x^* M$. This shows that $\mathcal{H}$ is continuously differentiable. Hence, Theorem 5.2.11 implies that the sequence $\{Z_n\}_{n\geqslant 1}$ satisfies in $D([0,1], M)$ the large deviation principle with good rate function given by (5.1.3). Since the random variables $\{Z_n\}_{n\geqslant 1}$ are almost surely in $D([0,1]; M)$ and $D([0,1]; M)$ is closed in $L^\infty([0,1]; M)$, the large deviation principle also holds in $L^\infty([0,1]; M)$ with the same rate function. □

5.3.3. Proof of Schilder's Theorem, Theorem 5.1.3

In this section we prove Schilder's theorem for Riemannian Brownian motion based on Theorem 5.2.11. The proof is similar as the proof of Theorem 5.1.1.

Proof of Theorem 5.1.3. Let us verify the conditions for Theorem 5.2.11. We calculate H_n and limit H. The process $X(t)$ solves the martingale problem for the operator $\frac{1}{2}\Delta_M$ and therefore, $X_n(t)$ is generated by $\frac{1}{2n}\Delta_M$. For $f \in C_c^\infty(M)$, we find

$$\begin{aligned} H_n f &= \frac{1}{n} e^{-nf} \frac{1}{2n} \Delta_M e^{nf} \\ &= \frac{1}{n} e^{-nf} e^{nf} \frac{1}{2} (\Delta_M f + n \left|\mathrm{d}f\right|^2_{g(x)}) \\ &= \frac{1}{2n} \Delta_M f + \frac{1}{2} \left|\mathrm{d}f\right|^2_{g(x)} . \end{aligned}$$

Let $H \subseteq C_b(M) \times C_b(M)$ be the operator with $\mathcal{D}(H) = C_c^\infty(M)$ and given by

$$Hf = \tfrac{1}{2} \left|\mathrm{d}f\right|^2_{g(x)}$$

for $f \in C_c^\infty(M)$:
It follows that for all $f \in C_c^\infty(M)$,

$$\lim_{n\to\infty} ||H_n f - Hf|| = 0,$$

implying that $H \subseteq ex - \lim_{n\to\infty} H_n$. Note that $Hf(x) = \mathcal{H}(x, \mathrm{d}f(x))$ for $\mathcal{H} : T^*M \to \mathbb{R}$ of the form $\mathcal{H}(x,p) = \frac{1}{2}|p|^2_{g(x)}$.
Now consider the collection $\{\mu_x\}_{x\in M}$ of normal distributions as defined in Example 3.2.11. As shown in the example, this collection satisfies the consistency property in Definition 3.2.7. Furthermore, keeping the identification in (5.1.2) in mind, the example also shows that

$$\log \int_{T_x M} e^{\langle v,p\rangle} \mu_x(\mathrm{d}v) = \frac{1}{2} |p|^2_{g(x)} = \mathcal{H}(x,p).$$

As a consequence, by Lemma 5.3.2 we have a good containment function Υ. Furthermore, it follows from the proof of Theorem 5.1.1 that for $\lambda > 0$ and $h \in C_b(M)$, the Hamilton-Jacobi equation satisfies the comparison principle.
Finally, note that $\mathcal{H}$ is continuously differentiable. Therefore, by Theorem 5.2.11, the sequence $\{X_n\}_{n\geqslant 1}$ satisfies in $D([0,1], M)$ the large deviation principle with good rate function given by (5.1.5). Since $\{X_n\}_{n\geqslant 1}$ lies almost surely in $C([0,1];M)$ and the topology of $D([0,1],M)$ restricted to $C([0,1],M)$ reduces to the uniform topology, the same large deviation principle holds in $C([0,1],M)$. □

5.4. A proof of Schilder's theorem via embeddings

In this section we provide an alternative proof of Schilder's theorem for Riemannian Brownian motion on M (Theorem 5.1.3). This approach is relevant for the extension of Theorem 5.1.3 to the time-inhomogeneous case which we consider in Chapter 7.

The proof relies on lifting the Riemannian Brownian motion to the orthonormal frame bundle OM. This lift is a diffusion on OM driven by a Euclidean Brownian motion. We then embed OM into some Euclidean space and use Freidlin-Wentzell theory to obtain the large deviations for the embedded process. By the contraction principle, we then obtain the large deviations for the process on OM, from which the large deviations for the rescaled Riemannian Brownian motion follow (also by the contraction principle). We refer to Section 2.3 for the terminology of frame bundles and horizontal lifts.

Freidlin-Wentzell theory for Stratonovich diffusions

As explained in Section 2.4, in manifolds we work with Stratonovich stochastic differential equations. Therefore, if we want to carry out the above procedure, we need to adapt Freidlin-Wentzell theory to the setting of Stratonovich stochastic differential equations. We have the following result.

Theorem 5.4.1 (Freidlin-Wentzell, Stratonovich version). *Let W_t be an $\mathbb{R}^l$-valued standard Brownian motion. Let $b : \mathbb{R}^k \to \mathbb{R}^k$ and $\sigma : \mathbb{R}^k \to \mathbb{R}^{k\times l}$ be bounded, Lipschitz continuous functions with $D\sigma$ also Lipschitz continuous. Fix $y \in \mathbb{R}^k$ and assume that for any $n \geqslant 1$ the process Y^n_t satisfies the Stratonovich stochastic differential equation*

$$\mathrm{d}Y^n_t = b(Y^n_t)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\sigma(Y^n_t)\circ \mathrm{d}W_t \tag{5.4.1}$$

with $Y^n_0 = y$. Then the sequence $\{Y^n\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];\mathbb{R}^k)$ with good rate function I given by

$$I(\gamma) = \inf\left\{\frac{1}{2}\int_0^1 |\dot\varphi(t)|^2\,\mathrm{d}t \,\middle|\, \varphi \in H_1([0,1];\mathbb{R}^l),\right.$$
$$\left.\gamma(t) = y + \int_0^t b(\gamma(s))\,\mathrm{d}s + \int_0^t \sigma(\gamma(s))\dot\varphi(s))\,\mathrm{d}s\right\}. \tag{5.4.2}$$

Proof. By Theorem 2.4.3, equation (5.4.1) is equivalent to the Itô stochastic differential equation

$$\mathrm{d}Y^n_t = b(Y^n_t)\,\mathrm{d}t + \frac{1}{2n}\sum_{j=1}^{l} D\sigma_j(Y^n_t)\sigma_j(Y^n_t)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\sigma(Y^n_t)\,\mathrm{d}W_t,$$

where $\sigma_1,\dots,\sigma_l$ denote the columns of σ.

Now suppose that $\tilde Y^n_t$ satisfies the Itô stochastic differential equation

$$\mathrm{d}\tilde Y^n_t = b(\tilde Y^n_t)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\sigma(\tilde Y^n_t)\,\mathrm{d}W_t$$

with $\tilde Y^n_0 = y$. By Theorem 2.1.15, we find that $\tilde Y^n_t$ satisfies in $C([0,1];\mathbb{R}^k)$ the large deviation principle with good rate function I as in (5.4.2). To complete the proof, it suffices to show that the sequences $\{Y^n\}_{n\geqslant 1}$ and $\{\tilde Y^n\}_{n\geqslant 1}$ are exponentially equivalent in $C([0,1];\mathbb{R}^k)$.

Consider the joint law of Y_0^n and $\tilde{Y}_0^n$ and the following system of stochastic differential equations

$$\begin{cases} \mathrm{d}Y_t^n = b(Y_t^n)\,\mathrm{d}t + \frac{1}{2n}\sum_{j=1}^l D\sigma_j(Y_t^n)\sigma_j(Y_t^\varepsilon)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\sigma(Y_t^n)\,\mathrm{d}W_t \\ \mathrm{d}\tilde{Y}_t^n = b(\tilde{Y}_t^n)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\sigma(\tilde{Y}_t^n)\,\mathrm{d}W_t \\ \mathrm{d}Z_t^n = b(Y_t^n)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\sigma(Y_t^n)\,\mathrm{d}W_t, \end{cases}$$

with $Y_0^n = \tilde{Y}_0^n = Z_0^n = y$.
First note that

$$\mathrm{d}(Z_t^n - Y_t^n) = -\frac{1}{2n}\sum_{j=1}^l D\sigma_j(Y_t^n)\sigma_j(Y_t^n)\,\mathrm{d}t.$$

Because σ and $D\sigma$ are bounded, we can find a constant $C > 0$ such that

$$|Z_t^n - Y_t^n| \leqslant C\frac{1}{n} \tag{5.4.3}$$

for all $t \in [0,1]$.
Furthermore, we have

$$\mathrm{d}(\tilde{Y}_t^n - Z_t^n) = (b(\tilde{Y}_t^n) - b(Y_t^n))\mathrm{d}t + \frac{1}{\sqrt{n}}(\sigma(\tilde{Y}_t^n) - \sigma(Y_t^n))\,\mathrm{d}W_t.$$

If we write B for the Lipschitz constant of b and use the estimate in (5.4.3), we find that

$$\begin{aligned} |b(\tilde{Y}_t^n) - b(Y_t^n)| &\leqslant B|\tilde{Y}_t^n - Y_t^n| \\ &\leqslant B(|\tilde{Y}_t^n - Z_t^n| + |Z_t^n - Y_t^n|) \\ &\leqslant B\sqrt{2}(|\tilde{Y}_t^n - Z_t^n|^2 + |Z_t^n - Y_t^n|^2)^{1/2} \\ &\leqslant B\sqrt{2}\left(|\tilde{Y}_t^n - Z_t^n|^2 + C^2\frac{1}{n^2}\right)^{1/2}. \end{aligned}$$

A similar estimate holds with σ instead of b. But then it follows from Lemma 5.6.18 in [29] that for $\delta > 0$ we have

$$\limsup_{n\to\infty}\frac{1}{n}\log P(\sup_{0\leqslant t\leqslant 1}|\tilde{Y}_t^n - Z_t^n| \geqslant \delta) \leqslant \limsup_{n\to\infty} K + \log\left(\frac{C^2}{C^2 + n^2\delta^2}\right) = -\infty.$$

This shows that that the sequences $\{\tilde{Y}^n\}_{n\geqslant 1}$ and $\{Z^n\}_{n\geqslant 1}$ are exponentially equivalent in $C([0,1];\mathbb{R}^k)$. Furthermore, it follows from (5.4.3) that the sequences $\{Y^n\}_{n\geqslant 1}$ and $\{Z^n\}_{n\geqslant 1}$ are also exponentially equivalent in $C([0,1];\mathbb{R}^k)$. We conclude that the sequences $\{Y^n\}_{n\geqslant 1}$ and $\{\tilde{Y}^n\}_{n\geqslant 1}$ are exponentially equivalent in $C([0,1];\mathbb{R}^k)$ as desired. □

Proof of Theorem 5.1.3 using embeddings

The proof we present here is an adaptation of the proof given in Section 6 of [63]. Before we give the proof, let us first provide a short overview.

Define $X_t^n = X_{nt}$ with X_t a Riemannian Brownian motion. Observe that the horizontal lift U_t^n of X_t^n satisfies the stochastic differential equation

$$\mathrm{d}U_t^n = H_i(U_t^n) \circ \mathrm{d}W_t^{n,i} \tag{5.4.4}$$

with $U_0^n = u_0 \in O_{x_0}M$. Here, $W_t^n = \frac{1}{\sqrt{n}}W_t$ with W_t an $\mathbb{R}^k$-valued standard Brownian motion. Using Whitney's embedding theorem, we can embed OM smoothly into a Euclidean space $\mathbb{R}^N$ and push-forward equation (5.4.4), making use of proposition 2.4.9 to relate the solutions. This results in a stochastic differential equation on $\mathbb{R}^N$ driven by a Euclidean Brownian motion, the solution of which remains inside the embedding of the manifold.

In order to obtain the large deviation principle, we first restrict the vector fields in equation (5.4.4) to a compact set using bump functions. This assures that the diffusion matrix of the pushed-forward equation in $\mathbb{R}^N$ is smooth with compact support. This in turn allows us to apply Freidlin-Wentzell theory in Euclidean space, giving us the large deviation principle in $C([0,1];\mathbb{R}^N)$. By the contraction principle, this also gives us a large deviation principle for $\{U^n\}_{n\geqslant 1}$ in $C([0,1];OM)$ and hence also for $\{X_n\}_{n\geqslant 1}$ in $C([0,T];M)$, at least if we restrict to some compact set. We remove this restriction by letting the compact set grow and using a compact containment argument (Proposition 5.2.10). Let us provide the details.

Proof of Theorem 5.1.3 using embeddings. Fix $u_0 \in O_{x_0}M$ and for every $n \geqslant 1$, let U_t^n be the solution of

$$\mathrm{d}U_t^n = H_i(U_t^n) \circ \mathrm{d}W_t^{n,i}$$

with $U_0^n = u_0$. Here, $W_t^n = \frac{1}{\sqrt{n}}$ with W_t a standard $\mathbb{R}^k$-valued Brownian motion. It follows from the proof of this theorem given in Section 5.3.3 that we can apply Proposition 5.2.10 to obtain for every $\alpha > 0$ a compact set $K_\alpha \subset M$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X_t^n \notin K_\alpha \text{ for some } t \in [0,1]) \leqslant -\alpha.$$

The sets can be chosen to be increasing and such that $\bigcup_\alpha K_\alpha = M$.

Let $\varphi_\alpha : M \to \mathbb{R}$ be a smooth function with $\varphi_\alpha \equiv 1$ on K_α and with compact support. We extend φ_α to OM by defining it to be constant on fibers. Abusing notation, we call this extension φ_α as well. Consider the process $U_t^{n,\alpha}$ in OM satisfying

$$\mathrm{d}U_t^{n,\alpha} = \varphi_\alpha H_i(U_t^{n,\alpha}) \circ \mathrm{d}W_t^{n,i} \tag{5.4.5}$$

with $U_0^{n,\alpha} = u_0$.

By Whitney's embedding theorem there exists an $N \in \mathbb{N}$ and a smooth embedding $\iota : OM \to \mathbb{R}^N$. We push (5.4.5) forward to $\iota(OM)$ to obtain the stochastic differential equation

$$\mathrm{d}V_t^{n,\alpha} = \iota_*(\varphi_\alpha H_i)(V_t^{n,\alpha}) \circ \mathrm{d}W_t^{n,i} \tag{5.4.6}$$

with $V_0^{n,\alpha} = \iota(u_0)$.
Because φ_α has compact support in M, the continuity of ι implies that the vector fields $\iota_*(\varphi_\alpha H_i)$ have compact support. Since $\iota(OM)$ is closed in $\mathbb{R}^N$, they can be extended to smooth, compactly support vector fields on $\mathbb{R}^N$, which we denote by $\iota_*\widetilde{(\varphi_\alpha H_i)}$. This gives us the following stochastic differential equation on $\mathbb{R}^N$:

$$\mathrm{d}\widetilde{V}_t^{n,\alpha} = \iota_*\widetilde{(\varphi_\alpha H_i)}(\widetilde{V}_t^{n,\alpha}) \circ \mathrm{d}W_t^{n,i}$$

with $\widetilde{V}_0^{n,\alpha} = \iota(u_0)$. Since the diffusion matrix is smooth with compact support, it follows from Theorem 5.4.1 that $\{\widetilde{V}^{n,\alpha}\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];\mathbb{R}^N)$ with good rate function

$$I_{\mathbb{R}^N}^\alpha(f) = \inf\left\{\frac{1}{2}\int_0^1 |\dot{g}(t)|^2_{\mathbb{R}^k}\,\mathrm{d}t \;\middle|\; g \in H_0^1([0,1];\mathbb{R}^k), \right. \\ \left. f(0) = \iota(u_0),\ \dot{f}(t) = \dot{g}^i(t)\iota_*\widetilde{(\varphi_\alpha H_i(f(t))}\right\}.$$

Since $f(0) = \iota(u_0) \in \iota(OM)$, the existence of such a g as in the rate function implies that $f([0,1]) \subseteq \iota(OM)$, because the vector fields $\iota_*\widetilde{(\varphi_\alpha H_i)}$ are tangent to $\iota(OM)$ at points of $\iota(OM)$. For this, a similar proof (but adjusted to the deterministic case) as that of [57, Proposition 1.2.8] can be used. Hence, $I_{\mathbb{R}^N}^\alpha$ is infinite outside $C([0,1];\iota(OM))$. Since the latter is a closed subset of $C([0,1];\mathbb{R}^N)$ (as $\iota(OM)$ is closed in $\mathbb{R}^N$), we conclude that $\{V_{n,\alpha}\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];\iota(OM))$, where the rate function $I_{\iota(OM)}^\alpha$ is simply the restriction of $I_{\mathbb{R}^N}^\alpha$.
Now observe that as ι is diffeomorphism, by Proposition 2.4.9 we have that $\iota(U_t^{n,\alpha})$ solves (5.4.6) with initial value $\iota(u_0)$ if and only if $U_t^{n,\alpha}$ solves (5.4.5) with initial value u_0. Therefore, by the contraction principle (Theorem 2.1.6) we find that $\{U^{n,\alpha}\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];OM)$ with good rate function given by

$$\begin{aligned} I_{OM}^\alpha(h) &= I_{\iota(OM)}(\iota\circ h) \\ &= \inf\left\{\frac{1}{2}\int_0^1 |\dot{g}(t)|^2_{\mathbb{R}^k}\,\mathrm{d}t \;\middle|\; g \in H_0^1([0,1];\mathbb{R}^k),\ \iota\circ h(0) = \iota(u_0), \right. \\ &\qquad \left. \frac{\mathrm{d}}{\mathrm{d}t}(\iota\circ h)(t) = \dot{g}^i(t)\iota_*(\varphi_\alpha H_i)(\iota\circ h(t))\right\}. \end{aligned}$$

Now observe that since ι is a smooth embedding, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}(\iota\circ h)(t) = \dot{g}^i(t)\iota_*(\varphi_\alpha H_i)(\iota\circ h(t))$$

if and only if

$$\dot{h}(t) = \dot{g}^i(t)(\varphi_\alpha H_i)(h(t)).$$

But then we can rewrite the rate function I_{OM}^α as

$$I^{\alpha}_{OM}(h) = \inf \left\{ \frac{1}{2}\int_0^1 |\dot{g}(t)|^2_{\mathbb{R}^k}\,\mathrm{d}t \;\middle|\; g \in H^1_0([0,1];\mathbb{R}^k), \right.$$
$$\left. h(0) = u_0,\ \dot{h}(t) = \dot{g}^i(t)(\varphi_\alpha H_i)(h(t)) \right\}.$$

Now, if there exists a $g \in H^1_0([0,1];\mathbb{R}^k)$ such that $\dot{h}(t) = \dot{g}^i(t)(\varphi_\alpha H_i)(h(t))$, then $\dot{h}(t)$ is horizontal for every $t \in [0,1]$. This implies that h is a horizontal curve. It follows that $I^{\alpha}_{OM}(h)$ can only be finite if h is a horizontal curve in OM.

Now define $X^{n,\alpha}_t = \pi(U^{n,\alpha}_t)$. By the continuity of π, the contraction principle (Theorem 2.1.6) implies that $\{X^{n,\alpha}\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];M)$ with good rate function $I^{\alpha}_M(f)$ given by

$$I^{\alpha}_M(f) = \inf \left\{ I^{\alpha}_{OM}(\hat{f}) \mid \hat{f} \in C([0,1];OM) \text{ with } \pi(\hat{f}) = f \right\}.$$

We now show how to simplify this expression when $f([0,1]) \subseteq K_\alpha$. As discussed above, $I^{\alpha}_{OM}(\hat{f})$ can only be finite if $\hat{f}$ is horizontal with $\hat{f}(0) = u_0$. This implies that it suffices to consider the horizontal lift h_f of f. Furthermore, since $f([0,1]) \subset K_\alpha$, we have that $\varphi_\alpha(f(t)) = 1$ for all $t \in [0,1]$. Therefore, to compute $I^{\alpha}_{OM}(h_f)$, we need to consider the unique curve $g : [0,1] \to \mathbb{R}^k$ such that $\dot{h}_f(t) = \dot{g}^i(t)H_i(h(t))$. In particular, g is the anti-development of f via the frame u_0. From this it follows that

$$I^{\alpha}_M(f) = \frac{1}{2}\int_0^1 |\dot{g}(t)|^2_{\mathbb{R}^k}\,\mathrm{d}t,$$

where g is the anti-development of f. Furthermore, we have that $\dot{g}(t) = h_f^{-1}(t)\dot{f}(t)$ for every $t \in [0,1]$. Using that $h_f(t)$ is an orthonormal frame and thus an isometry, we find that

$$|\dot{g}(t)|_{\mathbb{R}^k} = |h_f^{-1}(t)\dot{f}(t)|_{\mathbb{R}^k} = |\dot{f}(t)|_M.$$

This shows that, at least for f such that $f([0,1]) \subset K_\alpha$, the rate function I^{α}_M is given by

$$I^{\alpha}_M(f) = \frac{1}{2}\int_0^1 |\dot{f}(t)|^2_M\,\mathrm{d}t.$$

Finally, we show how to remove the restriction to compact sets. For this, let $T^{n,\alpha}$ be the exit time of X^n_t from K_α. Observe that by definition of $X^{n,\alpha}_t$ we have that X^n_t and $X^{n,\alpha}_t$ agree up to time $T^{n,\alpha}$.

Let us first prove the upper bound of the large deviation principle for $\{X^n\}_{n\geqslant 1}$. For this, let $F \subseteq C([0,1];M)$ be closed. Then

$$\begin{aligned}
\mathbb{P}(X^n \in F) &= \mathbb{P}(X^n \in F | T^{n,\alpha} > 1)\mathbb{P}(T^{n,\alpha} > 1) + \mathbb{P}(X^n \in F | T^{n,\alpha} \leqslant 1)\mathbb{P}(T^{n,\alpha} \leqslant 1) \\
&\leqslant \mathbb{P}(X^n \in F, T^{n,\alpha} > 1) + \mathbb{P}(T^{n,\alpha} \leqslant 1) \\
&= \mathbb{P}(X^{n,\alpha} \in F, T^{n,\alpha} > 1) + \mathbb{P}(T^{n,\alpha} \leqslant 1) \\
&\leqslant \mathbb{P}(X^{n,\alpha} \in F \cap C([0,1];K_\alpha)) + \mathbb{P}(T^{n,\alpha} \leqslant 1).
\end{aligned}$$

Using this, we find for every $\alpha > 0$ that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X^n \in F) \\
&\quad\leqslant \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X^{n,\alpha} \in F \cap C([0,1]; K_\alpha) + \mathbb{P}(T^{n,\alpha} \leqslant 1) \\
&\quad= \max\left\{\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X^{n,\alpha} \in F \cap C([0,1]; K_\alpha), \limsup_{n\to\infty} \frac{1}{n} \mathbb{P}(T^{n,\alpha} \leqslant 1)\right\} \\
&\quad\leqslant \max\{-\inf_{f\in F\cap C([0,1];K_\alpha)} I_M^\alpha(f), -\alpha\} \\
&\quad\leqslant \max\{-\inf_{f\in F} I_M(f), -\alpha\}.
\end{aligned}$$

Here, the last line follows from the fact that on $C([0,1]; K_\alpha)$, the rate function I_M^α coincides with I_M. Letting α tend to infinity proves the upper bound.

It remains to prove the lower bound. Let $G \subseteq C([0,1]; M)$ be open. Fix $g \in G$ and take $\delta > 0$ such that $B(g,\delta) \subseteq G$. Furthermore, since the sets K_α are increasing with $\bigcup_\alpha K_\alpha = M$, there exists an $\alpha > 0$ such that $g([0,1])$ is contained in the interior of K_α. By possibly shrinking δ, we then have for all $h \in B(g,\delta)$ that $h([0,1])$ is contained in the interior of K_α. From this it follows that

$$\mathbb{P}(X^n \in B(g,\delta)) = \mathbb{P}(X^{n,\alpha} \in B(g,\delta)).$$

Indeed, by continuity, $X^{n,\alpha}$ and X^n can only be different if X^n hits the boundary of K_α. However, since $h([0,1])$ is contained in the interior of $K_\alpha$for every $h \in B(g,\delta)$, this does not occur.

Using this, we find that

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X^n \in G) &\geqslant \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X^n \in B(g,\delta)) \\
&= \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X^{n,\alpha} \in B(g,\delta)) \\
&\geqslant -I_M^\alpha(g) \\
&= -I_M(g).
\end{aligned}$$

Here, the third line follows from the large deviation principle for $\{X^{n,\alpha}\}_{n\geqslant 1}$, while the last line follows from the fact that $I_M^\alpha(g) = I_M(g)$ for $g \in C([0,1]; M)$ with $g([0,1]) \subseteq K_\alpha$. As the above holds for all $g \in G$, this proves the lower bound.

Finally, to see that I_M is a good rate function, note that $I_M = \inf_\alpha I_M^\alpha$ and that the I_M^α are good rate functions. □

Remark 5.4.2. If the Ricci curvature is bounded from below, we can replace the compact containment argument by a more explicit estimate of the exit probability $\mathbb{P}(T^{n,\alpha} \leqslant 1)$. More precisely, one can use [63, Proposition 3.7] to obtain that

$$\mathbb{P}(T^{n,\alpha} \leqslant 1) \leqslant 2e^{-\frac{1}{2}n\frac{(kLn^{-1}-\frac{1}{2}\alpha^2)^2}{\alpha^2}}.$$

Here k is the dimension of the manifold and L is the lower bound on the Ricci curvature.

Remark 5.4.3. In a similar way as done in the final step of the above proof, one can also show that the large deviation principle holds for $\{U^n\}_{n\geqslant 1}$ and not only for $\{U^{n,\alpha}\}_{n\geqslant 1}$. Indeed, since horizontal lifts are unique, the fact that X_t^n and $X_t^{n,\alpha}$ agree up to time $T^{n,\alpha}$ implies that U_t^n and $U_t^{n,\alpha}$ agree up to time $T^{n,\alpha}$.

To prove the upper bound, let $F \subseteq C([0,1]; OM)$ be closed. A similar estimate as above shows that

$$\mathbb{P}(U_t^n \in F) \leqslant \mathbb{P}(U_t^{n,\alpha} \in F \cap C([0,1]; OK_\alpha)) + \mathbb{P}(T^{n,\alpha} \leqslant 1).$$

Here,

$$OK_\alpha = \{u \in OM | \pi u \in K_\alpha\}.$$

Noticing that $I_{OM}(h) = I_{OM}^\alpha(h)$ whenever $h([0,1]) \subseteq OK_\alpha$, a similar argument as above proves that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(U_t^n \in F) \leqslant -\inf_{h\in F} I_{OM}(h).$$

For the lower bound, let $G \subseteq C([0,1]; OM)$ be open. Fix $g \in G$ and $\delta > 0$ such that $B(g,\delta) \subseteq G$. Note that there exists an $\alpha > 0$ such that for all $h \in B(g,\delta)$ it holds that $\pi h([0,1]) \subseteq K_\alpha$, where we possibly have to shrink δ. By a similar argument as in the proof above, we obtain also the lower bound.

5.5. Concluding remarks

We conclude this chapter by discussing some directions in which the results from this chapter may be extended.

First of all, the discussion of the conditions of Cramér's theorem for geodesic random walks (Theorem 3.3.1) in Section 3.7 are also relevant for Mogulskii's theorem (Theorem 5.1.1). As for Cramér's theorem, it is expected to be possible to replace the boundedness of the increments with an assumption on their moment generating function.

Another interesting problem to consider is that of deducing Mogulskii's theorem for geodesic random walks from Cramér's theorem. It should be possible to follow a similar approach as in the proof of Mogulskii's theorem given in Section 5.1 of [29]. This gives two main difficulties. First of all, in comparison to Lemma 5.1.7 in [29], we need to show that the piecewise geodesic approximation of the geodesic random walk is exponentially tight in $C([0,1]; M)$. Second, we also need to show that the rate function obtained from the projective limit theorem has the desired form, see Lemma 5.1.6 in [29].

Furthermore, as already mentioned in Section 5.1, there is also reason to believe that Schilder's theorem for Riemannian Brownian motion can be obtained from Mogulskii's theorem. The only problem is that the 'increments' of Riemannian

Brownian motion are only asymptotically normal. Therefore, the main obstacle is to prove that a piecewise geodesic approximation with normal increments of a rescaled Riemannian Brownian approximates it well-enough on the exponential scale.

Finally, we can also consider Freidlin-Wentzell theory for diffusions on manifolds driven by a Euclidean Brownian motion. More precisely, we can consider processes X^n on M satisfying

$$\mathrm{d}X_t^n = b(X_t^n)\mathrm{d}t + \frac{1}{\sqrt{n}} V_i(X_t^n) \circ \mathrm{d}W_t^i.$$

This might even be pushed further, and consider stochastic differential equations driven by a Riemannian Brownian motion. We refer to [36] for the definition of such equations. For the case where the driving Brownian motion is Euclidean, both approaches we discussed in this chapter are suitable for studying the large deviations for $\{X^n\}_{n\geqslant 1}$. In particular, the approach using embedding can be applied immediately without having to lift the process to the frame bundle, as in the case for Riemannian Brownian motion.

III

Large deviations in a time-inhomogeneous setting

6 Large deviations for time-inhomogeneous processes

In this chapter we temporarily leave the geometric setting and aim to extend the classical results in large deviation theory into another direction. More precisely, we study the large deviation behaviour of random walks in Euclidean space with time-inhomogeneous increments. Furthermore, we also look at a time-inhomogeneous Schilder-type theorem by considering the process generated by a weighted Laplacian, where the weight depends on time. The main purpose of these results is to get a first look into large deviations for time-inhomogeneous processes and serve as a starting point for considering also time-inhomogeneous processes in a geometric setting. A first step in this direction is taken in Chapter 7, where we consider Schilder's theorem for Riemannian Brownian motion in a time-evolving Riemannian manifold.

This chapter is organized as follows. We first prove the large deviation principle for rescaled random walks with time-inhomogeneous increments in Section 6.1. This gives us the analogue of Cramér's theorem. It turns out that under suitable assumptions, this is a direct consequence of the Gärtner-Ellis theorem.

Next, in Section 6.2 we obtain the path space large deviations for random walks with time-dependent increments by following a similar approach as in the homogeneous case. Indeed, following the approach in [29, Section 5.1], we obtain the large deviations via the projective limit theorem of Dawson and Gärtner. However, to prove that the rate function takes on the desired form requires slightly more work than in the homogeneous case.

We conclude this chapter by considering an inhomogeneous Schilder-type theorem in Section 6.3. Since the process we consider for this is Gaussian, the result is a special case of Theorem 3.4.5 in [30]. However, we provide an alternative proof by showing how to obtain this result from the path space large deviations for time-inhomogeneous random walks. Furthermore, this result serves as a connection between this chapter and the next one, where we treat Schilder's theorem for

Riemannian Brownian motion in an evolving Riemannian manifold.

6.1. Large deviations for time-inhomogeneous random walks

Consider a collection $\{\mu_t\}_{t\in[0,1]}$ of probability measures on $\mathbb{R}^d$. Using this, we construct a time-inhomogeneous random walk. For every $n \in \mathbb{N}$, we define n independent random variables $X_1^n, \ldots, X_n^n$, where X_i^n is distributed according to $\mu_{\frac{i}{n}}$. Next, we consider the rescaled random walk

$$Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i^n. \tag{6.1.1}$$

We refer to the sequence $\{Z_n\}_{n\geqslant 1}$ as the *time-inhomogeneous random walk associated to* $\{\mu_t\}_{t\in[0,1]}$.

For a collection $\{\mu_t\}_{t\in[0,1]}$ of probability measures on $\mathbb{R}^d$, we denote by Λ_t the log-moment generating function of μ_t, i.e.,

$$\Lambda_t(\lambda) = \log \int_{\mathbb{R}^d} e^{\langle \lambda, x\rangle}\mu_t(\mathrm{d}x).$$

Using the Gärtner-Ellis theorem, we obtain the following time-inhomogeneous version of Cramér's theorem.

Theorem 6.1.1. *Let $\{\mu_t\}_{t\in[0,1]}$ be a collection of probability measures on $\mathbb{R}^d$. For every $n \in \mathbb{N}$, let Z_n be the random variable defined in* (6.1.1)*. Assume that $\Lambda_t(\lambda)$ is finite for all $\lambda \in \mathbb{R}^d$ and $t \in [0,1]$. Furthermore, assume that the map $t \mapsto \Lambda_t(\lambda)$ is continuous for every $\lambda \in \mathbb{R}^d$. Finally, assume that the map $\lambda \mapsto \int_0^1 \Lambda_t(\lambda)\,\mathrm{d}t$ is differentiable. Then the sequence $\{Z_n\}_{n\geqslant 1}$ satisfies the large deviation principle in $\mathbb{R}^d$ with good rate function*

$$I(x) = \sup_{\lambda\in\mathbb{R}^d}\left\{\langle \lambda, x\rangle - \int_0^1 \Lambda_t(\lambda)\,\mathrm{d}t\right\}. \tag{6.1.2}$$

Proof. We start with the following computation

$$\begin{aligned}\frac{1}{n}\log \mathbb{E}\left(e^{n\langle \lambda, Z_n\rangle}\right) &= \frac{1}{n}\sum_{i=1}^{n}\log \mathbb{E}\left(e^{\langle \lambda, Z_i^n\rangle}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\Lambda_{\frac{i}{n}}(\lambda).\end{aligned}$$

Since $t \mapsto \Lambda_t(\lambda)$ is continuous, we find that

$$\Lambda(\lambda) := \lim_{n\to\infty}\frac{1}{n}\log \mathbb{E}\left(e^{n\langle \lambda, Z_n\rangle}\right) = \int_0^1 \Lambda_t(\lambda)\,\mathrm{d}t.$$

Since $\Lambda(\lambda)$ is differentiable by assumption, the Gärtner-Ellis theorem (Theorem 2.1.12) implies that $\{Z_n\}_{n\geqslant 1}$ satisfies the large deviation principle in $\mathbb{R}^d$ with good rate function given by

$$I(x) = \Lambda^*(x) = \sup_{\lambda\in\mathbb{R}^d}\left\{\langle\lambda, x\rangle - \int_0^1 \Lambda_t(\lambda)\,\mathrm{d}t\right\}.$$

□

Remark 6.1.2. For the rate function I in (6.1.2) we have the upper bound

$$I(x) \leqslant \int_0^1 \Lambda_t^*(x)\,\mathrm{d}t,$$

where Λ_t^* is the Legendre transform of Λ_t, i.e.,

$$\Lambda_t^*(x) = \sup_{\lambda\in\mathbb{R}^d}\{\langle\lambda, x\rangle - \Lambda_t(\lambda)\}.$$

However, in general, equality need not hold. As an example, one can take $d = 1$ and consider $\mu_t = N(0, 1+t)$. Indeed, in this case, $\Lambda_t(\lambda) = \frac{1}{2}(1+t)\lambda^2$ and $\Lambda_t^*(x) = \frac{1}{2(1+t)}x^2$. Furthermore, we find that

$$\int_0^1 \Lambda_t(\lambda)\,\mathrm{d}t = \left[\frac{1}{4}\lambda^2(1+t)^2\right]_0^1 = \frac{3}{4}\lambda^2$$

from which we can compute that

$$I(x) = \sup_{\lambda\in\mathbb{R}}\left\{\lambda x - \frac{3}{4}\lambda^2\right\} = \frac{1}{3}x^2.$$

On the other hand,

$$\int_0^1 \Lambda_t^*(x)\,\mathrm{d}t = \left[\frac{1}{2}x^2\log(1+t)\right]_0^1 = \frac{1}{2}\log(2)x^2$$

We conclude that

$$I(x) = \frac{1}{3}x^2 < \frac{1}{2}\log(2)x^2 = \int_0^1 \Lambda_t^*(x)\,\mathrm{d}t.$$

6.2. Large deviations for trajectories of time-inhomogeneous random walks

We now turn to the path space large deviation result accompanying Theorem 6.1.1. To this end, let $Z_n(t)$ be the trajectories associated to the random variables in (6.1.1), i.e.,

$$Z_n(t) = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} X_i^n \tag{6.2.1}$$

for $t \in [0,1]$. Our aim is to prove the following theorem.

Theorem 6.2.1. *Let $\{\mu_t\}_{t\in[0,1]}$ be a collection of probability measures on $\mathbb{R}^d$. For every $n \in \mathbb{N}$ and $t \in [0,1]$, let $Z_n(t)$ be the random variable defined in (6.2.1). Assume that $\Lambda_t(\lambda) < \infty$ for all $\lambda \in \mathbb{R}^d$ and $t \in [0,1]$. Assume furthermore that the map $(t, \lambda) \mapsto \Lambda_t(\lambda)$ is continuous. Also assume that for every $r \in [0,1]$, the map $\lambda \to \int_0^r \Lambda_t(\lambda)\,\mathrm{d}t$ is differentiable. Then the sequence $\{Z_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $L^\infty([0,1];\mathbb{R}^d)$ with good rate function given by*

$$I(\gamma) = \begin{cases} \int_0^1 \Lambda_t^*(\dot\gamma(t))\,\mathrm{d}t, & \gamma \in AC_0([0,1];\mathbb{R}^d), \\ \infty, & \textit{otherwise}. \end{cases} \tag{6.2.2}$$

To prove Theorem 6.2.1, we follow the approach taken in [29, Section 5.1]. We first establish a variety of preparatory results, from which Theorem 6.2.1 will follow. Before we get to these results, we first need to make some more definitions. We define the space $\mathcal{X}$ by

$$\mathcal{X} = \left\{ f : [0,1] \to \mathbb{R}^d \;\middle|\; f(0) = 0 \right\}, \tag{6.2.3}$$

equipped with the topology of pointwise convergence, i.e., the product topology. Furthermore, let $\tilde{Z}_n$ be the piecewise linear approximation of Z_n, i.e.,

$$\tilde{Z}_n(t) = Z_n(t) + \left(t - \frac{\lfloor nt \rfloor}{n} \right) X^n_{\lfloor nt \rfloor + 1} \tag{6.2.4}$$

for $t \in [0,1]$.

We start by showing that the large deviations for $\{Z_n(\cdot)\}_{n\geqslant 1}$ are the same as for the piecewise linear approximations $\{\tilde{Z}_n(\cdot)\}_{n\geqslant 1}$. Before we can do this, we first need a technical lemma.

Lemma 6.2.2. *Let $\{\mu_t\}_{t\in[0,1]}$ be a collection of probability measures. Assume that $\Lambda_t(\lambda) < \infty$ for all $\lambda \in \mathbb{R}^d$ and $t \in [0,1]$. Assume furthermore that $t \mapsto \Lambda_t(\lambda)$ is continuous for every $\lambda \in \mathbb{R}^d$. Then*

$$\sup_{t\in[0,1]} \mathbb{E}_{\mu_t}\left(e^{\alpha |X|} \right) < \infty$$

for every $\alpha > 0$.

Proof. Since $\Lambda_t(\lambda)$ is continuous in t, so is $\mathbb{E}_{\mu_t}\left(e^{\langle \lambda, X\rangle}\right)$. Hence, the compactness of $[0,1]$ implies that

$$\sup_{t\in[0,1]} \mathbb{E}_{\mu_t}\left(e^{\langle \lambda, X\rangle} \right) < \infty \tag{6.2.5}$$

for all $\lambda \in \mathbb{R}^d$.

Furthermore, we have that

$$\mathbb{E}_{\mu_t}\left(e^{\alpha |X|} \right) \leqslant \mathbb{E}_{\mu_t}\left(\prod_{i=1}^d e^{\alpha |X_i|} \right)$$

$$\leqslant \prod_{i=1}^{d} \mathbb{E}_{\mu_t}\left(e^{\alpha d|X_i|}\right)^{\frac{1}{d}}$$
$$\leqslant \prod_{i=1}^{d} \left(\mathbb{E}_{\mu_t}\left(e^{\alpha d X_i}\right) + \mathbb{E}_{\mu_t}\left(e^{-\alpha d X_i}\right)\right)^{\frac{1}{d}}.$$

Here, the second line follows from Hölder's inequality. Using this, we find that

$$\sup_{t\in[0,1]} \mathbb{E}_{\mu_t}\left(e^{\alpha|X|}\right) \leqslant \prod_{i=1}^{d}\left(\sup_{t\in[0,1]} \mathbb{E}_{\mu_t}\left(e^{\alpha d X_i}\right) + \sup_{t\in[0,1]} \mathbb{E}_{\mu_t}\left(e^{-\alpha d X_i}\right)\right)^{\frac{1}{d}} < \infty,$$

where the latter is finite by considering the vectors $\lambda_i = \pm\alpha d e_i$ in (6.2.5), with $e_1, \ldots, e_d$ the standard basis of $\mathbb{R}^d$. □

We can now prove that the sequences $\{Z_n(\cdot)\}_{n\geqslant 1}$ and $\{\tilde{Z}_n(\cdot)\}_{n\geqslant 1}$ are exponentially equivalent.

Proposition 6.2.3. *Let the assumptions of Theorem 6.2.1 be satisfied. For every $n \geqslant 1$ and $t \in [0,1]$, let $Z_n(t)$ and $\tilde{Z}_n(t)$ be the random variables defined in* (6.2.1) *and* (6.2.4) *respectively. Then the sequences $\{Z_n(\cdot)\}_{n\geqslant 1}$ and $\{\tilde{Z}_n(\cdot)\}_{n\geqslant 1}$ are exponential equivalent in $L^\infty([0,1];\mathbb{R}^d)$.*

Proof. Note that $|Z_n(t) - \tilde{Z}_n(t)| \leqslant \frac{1}{n}|X^n_{[nt]+1}|$ for every $t \in [0,1)$, while $Z_n(1) - \tilde{Z}_n(1) = 0$. Using this, together with the union bound and Markov's inequality, we find that

$$\mathbb{P}\left(\sup_{t\in[0,1]} |Z_n(t) - \tilde{Z}_n(t)| > \delta\right) \leqslant \sum_{i=1}^{n} \mathbb{P}(|X_i^n| \geqslant n\delta)$$
$$\leqslant e^{-\lambda n\delta} \sum_{i=1}^{n} \mathbb{E}\left(e^{\lambda|X_i^n|}\right)$$
$$= e^{-\lambda n\delta} \sum_{i=1}^{n} \overline{M}_{\frac{i}{n}}(\lambda),$$

where

$$\overline{M}_t(\lambda) = \mathbb{E}_{\mu_t}\left(e^{\lambda|X|}\right).$$

This implies that

$$\mathbb{P}\left(\sup_{t\in[0,1]} |Z_n(t) - \tilde{Z}_n(t)| > \delta\right) \leqslant n e^{-\lambda n\delta} \sup_{t\in[0,1]} \overline{M}_t(\lambda),$$

where the upper bound is finite by Lemma 6.2.2. It follows that

$$\frac{1}{n}\log\mathbb{P}\left(\sup_{t\in[0,1]} |Z_n(t) - \tilde{Z}_n(t)| > \delta\right) \leqslant -\lambda\delta + \frac{1}{n}\log n + \frac{1}{n}\log\left(\sup_{t\in[0,1]} \overline{M}_t(\lambda)\right),$$

from which we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sup_{t\in[0,1]} |Z_n(t) - \tilde{Z}_n(t)| > \delta \right) \leqslant -\lambda\delta.$$

The claim follows by considering the limit $\lambda \to \infty$. □

Next, we show that the piecewise linear approximations $\{\tilde{Z}_n(\cdot)\}_{n\geqslant 1}$ satisfy the large deviation principle in $\mathcal{X}$. For this, we first use Cramér's theorem for time-inhomogeneous random walks (Theorem 6.1.1) to prove the large deviation principle for the finite dimensional distributions. We then obtain the path space large deviations by using the projective limit theorem of Dawson-Gärtner (see [28] and also [29, Theorem 4.6.1]). The most work goes into proving that the rate function has the desired form.

Proposition 6.2.4. *Let the assumptions of Theorem 6.2.1 be satisfied. For every $n \geqslant 1$ and $t \in [0,1]$, let $\tilde{Z}_n(t)$ be the random variable defined in* (6.2.4)*. Finally, let $\mathcal{X}$ be the space defined in* (6.2.3)*. Then $\{\tilde{Z}_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $\mathcal{X}$ with good rate function given by* (6.2.2)*.*

As mentioned above, before we can prove this, we first have to prove the large deviation principle for the finite-dimensional distributions.

Proposition 6.2.5. *Let the assumptions of Theorem 6.2.1 be satisfied. For every $n \geqslant 1$ and $t \in [0,1]$, let $Z_n(t)$ be the random variable defined in* (6.2.1)*. Finally, let $0 = t_0 < t_1 < \cdots < t_k \leqslant 1$ be a partition of $[0,1]$. Then the sequence $\{(Z_n(t_1), Z_n(t_2), \ldots, Z_n(t_k)\}_{n\geqslant 1}$ satisfies the large deviation principle in $(\mathbb{R}^d)^k$ with good rate function*

$$I(x_1, \ldots, x_k) = \sum_{l=1}^{k} \sup_{\lambda\in\mathbb{R}^d} \left\{ \langle \lambda, x_l - x_{l-1} \rangle - \int_{t_{l-1}}^{t_l} \Lambda_t(\lambda)\, \mathrm{d}t \right\},$$

where $x_0 = 0$.

Proof. Following the proof of Theorem 6.1.1, we find that for $s < r$, the sequence $\{Z_n(r) - Z_n(s)\}_{n\geqslant 1}$ satisfies in $\mathbb{R}^d$ the large deviation principle with rate function

$$I_{s,r}(x) = \sup_{\lambda\in\mathbb{R}^d} \left\{ \langle \lambda, x \rangle - \int_s^r \Lambda_t(\lambda)\, \mathrm{d}t \right\}.$$

Here, one uses that $\lambda \mapsto \int_s^r \Lambda_t(\lambda)\, \mathrm{d}t$ is differentiable, which follows from the assumption that for every r, the map $\lambda \mapsto \int_0^r \Lambda_t(\lambda)\, \mathrm{d}t$ is differentiable.
Now, since the increments of $Z_n(\cdot)$ are independent, we find that $\{(Z_n(t_1), Z_n(t_2) - Z_n(t_1), \ldots, Z_n(t_k) - Z_n(t_{k-1}))\}_{n\geqslant 1}$ satisfies in $(\mathbb{R}^d)^k$ the large deviation principle with rate function

$$\tilde{I}_{t_1,\ldots,t_k}(x_1, \ldots, x_k) = \sum_{l=1}^{k} I_{t_{l-1},t_l}(x_l).$$

Applying the contraction principle, we find that $\{(Z_n(t_1), Z_n(t_2), \ldots, Z_n(t_k)\}_{n\geqslant 1}$ satisfies in $(\mathbb{R}^d)^k$ the large deviation principle with good rate function

$$\begin{aligned} I_{t_1,\ldots,t_k}(x_1,\ldots,x_k) &= \tilde{I}_{t_1,\ldots,t_k}(x_1, x_2 - x_1, \ldots, x_k - x_{k-1}) \\ &= \sum_{l=1}^{k} I_{t_{l-1},t_l}(x_l - x_{l-1}) \\ &= \sum_{l=1}^{k} \sup_{\lambda\in\mathbb{R}^d} \left\{ \langle \lambda, x_l - x_{l-1}\rangle - \int_{t_{l-1}}^{t_l} \Lambda_t(\lambda)\,\mathrm{d}t \right\} \end{aligned}$$

as desired. □

In order to prove that the rate function given by the projective limit theorem of Dawson-Gärtner is of the desired form, we need the following technical lemma.

Lemma 6.2.6. *Let the assumptions of Theorem 6.2.1 be satisfied. Define $\Lambda_t = \Lambda_0$ for $t \leqslant 0$ and $\Lambda_t = \Lambda_1$ for $t \geqslant 1$. Let $H : (\mathbb{N} \cup \{\infty\}) \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ be given by*

$$H(k, \lambda, x, s) = \langle \lambda, x\rangle - k \int_{s}^{s+\frac{1}{k}} \Lambda_t(\lambda)\,\mathrm{d}t,$$

and set

$$H(\infty, \lambda, x, s) = \langle \lambda, x\rangle - \Lambda_s(\lambda).$$

Then H is continuous as function of four variables.

Proof. Let $(k_n, \lambda_n, x_n, s_n) \to (k, \lambda, x, s)$. We show that

$$\lim_{n\to\infty} H(k_n, \lambda_n, x_n, s_n) = H(k, \lambda, x, s).$$

Since the inner product is continuous, we have

$$\lim_{n\to\infty} \langle \lambda_n, x_n\rangle = \langle \lambda, x\rangle.$$

For the other term, we consider two cases. First assume that $k < \infty$. Then there exists and N such that $k_n = k$ for all $n \geqslant N$. Since $(t, \lambda) \mapsto \Lambda_t(\lambda)$ is continuous, it is bounded on compact sets and hence, we find by dominated convergence that

$$\lim_{n\to\infty} \int_{s_n}^{s_n+\frac{1}{k}} \Lambda_t(\lambda_n)\,\mathrm{d}t = \int_{s}^{s+\frac{1}{k}} \Lambda_t(\lambda)\,\mathrm{d}t.$$

Now consider the case $k = \infty$. Since $(t, \lambda) \mapsto \Lambda_t(\lambda)$ is continuous, by the mean value theorem, there exists for every $n \in \mathbb{N}$ a $\xi_n \in (s_n, s_n + \frac{1}{k_n})$ such that

$$k_n \int_{s_n}^{s_n+\frac{1}{k_n}} \Lambda_t(\lambda_n)\,\mathrm{d}t = \Lambda_{\xi_n}(\lambda_n).$$

Because $s_n \to s$ and $\frac{1}{k_n} \to 0$, we find that $\lim_{n\to\infty} \xi_n = s$. Again using the continuity of $(t, \lambda) \mapsto \Lambda_t(\lambda)$, we conclude that

$$\lim_{n\to\infty} k_n \int_{s_n}^{s_n+\frac{1}{k_n}} \Lambda_t(\lambda_n)\,\mathrm{d}t = \lim_{n\to\infty} \Lambda_{\xi_n}(\lambda_n) = \Lambda_s(\lambda)$$

as desired. □

We are now able to prove Proposition 6.2.4.

Proof of Proposition 6.2.4. By combining Propositions 6.2.3 and 6.2.5, we find that $\{(\tilde{Z}_n(t_1), \ldots, \tilde{Z}_n(t_k)\}_{n\geqslant 1}$ satisfies in $(\mathbb{R}^d)^k$ the large deviation principle with rate function

$$I_{t_1,\ldots,t_k}(x_1,\ldots,x_k) = \sum_{l=1}^{k} \sup_{\lambda\in\mathbb{R}^d} \left\{ \langle \lambda, x_l - x_{l-1}\rangle - \int_{t_{l-1}}^{t_l} \Lambda_t(\lambda)\,\mathrm{d}t \right\}.$$

Following the proof of [29, Lemma 5.1.6], the projective limit theorem of Dawson-Gärtner ([29, Theorem 4.6.1]) implies that $\{\tilde{Z}_n(\cdot)\}_{n\geqslant 1}$ satisfies in $\mathcal{X}$ the large deviation principle with good rate function given by

$$\begin{aligned} I_{\mathcal{X}}(\gamma) &= \sup_{0=t_0<t_1<\cdots<t_k\leqslant 1} I_{t_1,\ldots,t_k}(\gamma(t_1),\ldots,\gamma(t_k)) \\ &= \sup_{0=t_0<t_1<\cdots<t_k=1} \sum_{l=1}^{k} \sup_{\lambda\in\mathbb{R}^d} \left\{ \langle \lambda, \gamma(t_l) - \gamma(t_{l-1})\rangle - \int_{t_{l-1}}^{t_l} \Lambda_t(\lambda)\,\mathrm{d}t \right\}. \end{aligned}$$

Here, in the last line we can take $t_k = 1$, since the functions involved are nonnegative. We are done once we show that $I_{\mathcal{X}} = I$, where I is as in (6.2.2). We first prove that $I_{\mathcal{X}} \leqslant I$. If γ is not absolutely continuous, then $I(\gamma) = \infty$ and certainly $I_{\mathcal{X}}(\gamma) \leqslant I(\gamma)$. If γ is absolutely continuous, then

$$\begin{aligned} I_{\mathcal{X}}(\gamma) &= \sum_{l=1}^{k} \sup_{\lambda\in\mathbb{R}^d} \left\{ \langle \lambda, \gamma(t_l) - \gamma(t_{l-1})\rangle - \int_{t_{l-1}}^{t_l} \Lambda_t(\lambda)\,\mathrm{d}t \right\} \\ &= \sum_{l=1}^{k} \sup_{\lambda\in\mathbb{R}^d} \int_{t_{l-1}}^{t_l} \langle \lambda, \dot{\gamma}(t)\rangle - \Lambda_t(\lambda)\,\mathrm{d}t \\ &\leqslant \sum_{l=1}^{k} \int_{t_{l-1}}^{t_l} \sup_{\lambda\in\mathbb{R}^d} \{\langle \lambda, \dot{\gamma}(t)\rangle - \Lambda_t(\lambda)\}\,\mathrm{d}t \\ &= \int_0^1 \Lambda_t^*(\dot{\gamma}(t))\,\mathrm{d}t \\ &= I(\gamma). \end{aligned}$$

For the reverse inequality, first consider the case where γ is absolutely continuous. For $k \in \mathbb{N}$, define the points $t_l = \frac{l}{k}$. Then

$$I_{\mathcal{X}}(\gamma) \geqslant \sum_{l=1}^{k} \sup_{\lambda\in\mathbb{R}^d} \left\{ \left\langle \lambda, \gamma\left(\frac{l}{k}\right) - \gamma\left(\frac{l-1}{k}\right) \right\rangle - \int_{(l-1)/k}^{l/k} \Lambda_t(\lambda)\,\mathrm{d}t \right\}$$

$$= \frac{1}{k}\sum_{l=1}^{k} \sup_{\lambda\in\mathbb{R}^d} \left\{ \left\langle \lambda, k\left(\gamma\left(\frac{l}{k}\right) - \gamma\left(\frac{l-1}{k}\right)\right)\right\rangle - k\int_{(l-1)/k}^{l/k} \Lambda_t(\lambda)\,\mathrm{d}t \right\}.$$

Now define for $s < r$ the function

$$F_{s,r}(\lambda) = \frac{1}{r-s}\int_s^r \Lambda_t(\lambda)\,\mathrm{d}t,$$

together with its Legendre transform

$$F^*_{s,r}(x) = \sup_{\lambda\in\mathbb{R}^d} \left\{\langle\lambda, x\rangle - F_{s,r}(\lambda)\right\}.$$

With this notation, we can write the above estimate as

$$I_{\mathcal{X}}(\gamma) \geqslant \frac{1}{k}\sum_{l=1}^{k} F^*_{\frac{l-1}{k},\frac{l}{k}}\left(k\left[\gamma\left(\frac{l}{k}\right) - \gamma\left(\frac{l-1}{k}\right)\right]\right). \tag{6.2.6}$$

Now define the function

$$G_k(t) = F^*_{\frac{\lfloor kt\rfloor}{k},\frac{\lfloor kt\rfloor+1}{k}}\left(k\int_{\frac{\lfloor kt\rfloor}{k}}^{\frac{\lfloor kt\rfloor+1}{k}} \dot\gamma(u)\,\mathrm{d}u\right),$$

and set $G_k(1) = G_k\left(\frac{k-1}{k}\right)$. Then the inequality in (6.2.6) may be rewritten as

$$I_{\mathcal{X}}(\gamma) \geqslant \int_0^1 G_k(t)\,\mathrm{d}t.$$

We will show that

$$\liminf_{k\to\infty} G_k(t) \geqslant \Lambda^*_t(\dot\gamma(t)).$$

To this end, consider the function $H : (\mathbb{N}\cup\{\infty\})\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}\to\mathbb{R}$ as in Lemma 6.2.6 and define the function $H^* : (\mathbb{N}\cup\{\infty\})\times\mathbb{R}^d\times\mathbb{R}\to\mathbb{R}$ by

$$H^*(k,x,s) = \sup_{\lambda\in\mathbb{R}^d} H(k,\lambda,x,s).$$

Because H is continuous by Lemma 6.2.6, it follows that H^* is lower-semicontinuous. Now note that

$$G_k(t) = H^*\left(k, k\int_{\frac{\lfloor kt\rfloor}{k}}^{\frac{\lfloor kt\rfloor+1}{k}} \dot\gamma(u)\,\mathrm{d}u, \frac{\lfloor kt\rfloor}{k}\right).$$

Since H^* is lower-semicontinuous, we find that

$$\begin{aligned}\liminf_{k\to\infty} G_k(t) &\geqslant H^*\left(\infty, \liminf_{k\to\infty} k\int_{\frac{\lfloor kt\rfloor}{k}}^{\frac{\lfloor kt\rfloor+1}{k}} \dot\gamma(u)\,\mathrm{d}u, t\right)\\ &= H^*(\infty,\dot\gamma(t),t)\end{aligned}$$

$$= \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, \dot{\gamma}(t) \rangle - \Lambda_t(\lambda)\}$$
$$= \Lambda_t^*(\dot{\gamma}(t)).$$

Here, the second line follows from the Lebesgue differentiation theorem, because $\dot{\gamma} \in L^1([0,1]; \mathbb{R}^d)$.

It now follows from Fatou's lemma that

$$I_{\mathcal{X}}(\gamma) \geqslant \liminf_{k \to \infty} \int_0^1 G_k(t)\,\mathrm{d}t \geqslant \int_0^1 \liminf_{k \to \infty} G_k(t)\,\mathrm{d}t \geqslant \int_0^1 \Lambda_t^*(\dot{\gamma}(t))\,\mathrm{d}t,$$

which shows that $I_{\mathcal{X}}(\gamma) \geqslant I(\gamma)$ whenever γ is absolutely continuous.

It remains to prove that if γ is not absolutely continuous, then $I_{\mathcal{X}}(\gamma) = \infty$. Since γ is not absolutely continuous, given $\delta > 0$, we can find a sequence $\{0 < t_1^n < s_1^n \leqslant \cdots \leqslant t_{k(n)}^n < s_{k(n)}^n \leqslant 1\}$ of partitions, such that

$$\lim_{n \to \infty} \sum_{l=1}^{k(n)} (s_l^n - t_l^n) = 0,$$

while

$$\sum_{l=1}^{k(n)} |\gamma(s_l^n) - \gamma(t_l^n)| \geqslant \delta.$$

For these partitions we have

$$I_{\mathcal{X}}(\gamma) \geqslant \sum_{l=1}^{k(n)} \sup_{\lambda \in \mathbb{R}^d} \left\{ \langle \lambda, \gamma(s_l^n) - \gamma(t_l^n) \rangle - \int_{t_l^n}^{s_l^n} \Lambda_t(\lambda) \right\}.$$

Indeed, this follows from the fact that for every $n \geqslant 1$ and every $1 \leqslant l \leqslant k(n)$ we have that

$$\sup_{\lambda \in \mathbb{R}^d} \left\{ \langle \lambda, \gamma(t_{l+1}^n) - \gamma(s_l^n) \rangle - \int_{s_l^n}^{t_{l+1}^n} \Lambda_t(\lambda) \right\} \geqslant 0$$

by considering $\lambda = 0$.

If we now consider $\lambda_l = \rho \frac{\gamma(s_l^n) - \gamma(t_l^n)}{|\gamma(s_l^n) - \gamma(t_l^n)|}$ whenever $\gamma(s_l^n) - \gamma(t_l^n) \neq 0$, we find that

$$I_{\mathcal{X}}(\gamma) \geqslant \sum_{l=1}^{k(n)} \left\{ \langle \lambda_l, \gamma(s_l^n) - \gamma(t_l^n) \rangle - \int_{t_l^n}^{s_l^n} \Lambda_t(\lambda_l) \right\}$$
$$\geqslant \rho \sum_{l=1}^{k(n)} |\gamma(s_l^n) - \gamma(t_l^n)| - \left[\sup_{0 \leqslant t \leqslant 1, |\lambda| = \rho} \Lambda_t(\lambda) \right] \sum_{l=1}^{k(n)} (s_l^n - t_l^n).$$

Now, because $(t, \lambda) \mapsto \Lambda_t(\lambda)$ is continuous, we have that $\sup_{0 \leqslant t \leqslant 1, |\lambda| = \rho} \Lambda_t(\lambda) < \infty$. Therefore, we find that

$$I_{\mathcal{X}}(\gamma) \geqslant \limsup_{n \to \infty} \left(\rho \sum_{l=1}^{k(n)} |\gamma(s_l^n) - \gamma(t_l^n)| - \left[\sup_{0 \leqslant t \leqslant 1, |\lambda| = \rho} \Lambda_t(\lambda) \right] \sum_{l=1}^{k(n)} (s_l^n - t_l^n) \right)$$

$$\geqslant \rho\delta.$$

Since $\rho > 0$ is arbitrary, the result follows by letting ρ tend to infinity. □

We have now shown that $\{\tilde{Z}_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $\mathcal{X}$. In order to prove that the large deviation principle also holds in the supremum norm topology, we need to prove that $\{\tilde{Z}_n(\cdot)\}_{n\geqslant 1}$ is exponentially tight in $C_0([0,1];\mathbb{R}^d)$.

Proposition 6.2.7. *Let the assumptions of Theorem 6.2.1 be satisfied. For every $n \geqslant 1$ and $t \in [0,1]$, let $\tilde{Z}_n(t)$ be the random variable defined in* (6.2.4)*. Then the sequence $\{\tilde{Z}_n(\cdot)\}_{n\geqslant 1}$ is exponentially tight in $C_0([0,1];\mathbb{R}^d)$.*

Proof. If X is and $\mathbb{R}^d$-valued random variable with distribution μ_t, then we write $\mu_{t,j}$ for the distribution of X^j, the j-th coordinate of X. Furthermore, we denote by $\Lambda_{t,j}$ the log moment generating function of the distribution $\mu_{t,j}$.
Now, given $\alpha > 0$ and $n \geqslant 1$, consider the sets

$$K^j_{\alpha,n} = \left\{\gamma \in AC([0,1];\mathbb{R}^d) \middle| \gamma(0) = 0, \int_0^1 \Lambda^*_{\frac{\lfloor nt\rfloor}{n},j}(\dot{\gamma}(t))\,\mathrm{d}t \leqslant \alpha\right\}$$

for $j = 1,\ldots,d$. Furthermore, define the sets

$$K^j_\alpha = \bigcup_{n=1}^{\infty} K^j_{\alpha,n}$$

and set

$$K_\alpha = \bigcap_{j=1}^{d} K^j_\alpha.$$

Then

$$\mathbb{P}(\tilde{Z}_n(\cdot) \notin K_\alpha) \leqslant d \max_{j=1}^{d} \mathbb{P}(\tilde{Z}_n(\cdot) \notin K^j_\alpha).$$

Furthermore, since $\frac{\mathrm{d}\tilde{Z}_n(t)}{\mathrm{d}t} = X^n_{\lfloor nt\rfloor+1}$ almost everywhere, we find that

$$\begin{aligned}\mathbb{P}(\tilde{Z}_n(\cdot) \notin K^j_\alpha) &\leqslant \mathbb{P}(\tilde{Z}_n(\cdot) \notin K^j_{\alpha,n})\\ &\leqslant \mathbb{P}\left(\frac{1}{n}\sum_{l=1}^{n} \Lambda^*_{\frac{l-1}{n},j}(X^{n,j}_l) > \alpha\right)\end{aligned}$$

for every $j = 1,\ldots,d$ and every $n \geqslant 1$. Estimating further, we have

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{l=1}^{n} \Lambda^*_{\frac{l-1}{n},j}(X^{n,j}_l) > \alpha\right) &= \mathbb{P}\left(e^{\delta\sum_{l=1}^{n}\Lambda^*_{\frac{l-1}{n},j}(X^{n,j}_l)} > e^{n\delta\alpha}\right)\\ &\leqslant e^{-n\delta\alpha}\mathbb{E}\left(e^{\delta\sum_{l=1}^{n}\Lambda^*_{\frac{l-1}{n},j}(X^{n,j}_l)}\right)\end{aligned}$$

$$= e^{-n\delta\alpha} \prod_{l=1}^{n} \mathbb{E}\left(e^{\delta \Lambda^*_{\frac{l-1}{n},j}(X_l^{n,j})} \right).$$

Here, the second line follows from Markov's inequality and the last line from the independence of the increments.

Now, since $X_l^{n,j}$ has distribution $\mu_{\frac{l-1}{n},j}$, it follows from [29, Lemma 5.1.14] that for $\delta < 1$ we have

$$\mathbb{E}\left(e^{\delta \Lambda^*_{\frac{l-1}{n},j}(X_l^{n,j})} \right) \leqslant \frac{2}{1-\delta}.$$

Combining everything, we find that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{l=1}^{n} \Lambda^*_{\frac{l-1}{n},j}(X_l^{n,j}) > \alpha \right) \leqslant -\delta\alpha + \log\left(\frac{2}{1-\delta} \right),$$

from which we conclude that

$$\lim_{\alpha\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{l=1}^{n} \Lambda^*_{\frac{l-1}{n},j}(X_l^{n,j}) > \alpha \right) = -\infty.$$

Since

$$\mathbb{P}(\tilde{Z}_n(\cdot) \notin K_\alpha) \leqslant d \max_{j=1}^{d} \mathbb{P}\left(\frac{1}{n} \sum_{l=1}^{n} \Lambda^*_{\frac{l-1}{n},j}(X_l^{n,j}) > \alpha \right),$$

it follows that

$$\lim_{\alpha\to\infty} \limsup_{n\to\infty} \mathbb{P}(\tilde{Z}_n(\cdot) \notin K_\alpha) = -\infty.$$

It remains to show that K_α is relatively compact. Since $K_\alpha = \bigcap_{j=1}^{d} K_\alpha^j$, it is sufficient to show that K_α^j is compact for arbitrary j. By the Arzelà-Ascoli theorem, it suffices to show that K_α^j is bounded and equicontinuous. We first prove equicontinuity of K_α^j, since boundedness will be a consequence of one of the estimates.

Let $\gamma \in K_\alpha^j$ be arbitrary. Then there exists an $n \geqslant 1$ such that

$$\int_0^1 \Lambda^*_{\frac{\lfloor nt \rfloor}{n},j}(\dot{\gamma}(t))\,\mathrm{d}t \leqslant \alpha.$$

First, let $\frac{l-1}{n} \leqslant s < r \leqslant \frac{l}{n}$. By Jensen's inequality, we have

$$\begin{aligned} \Lambda^*_{\frac{l-1}{n},j}\left(\frac{\gamma(r)-\gamma(s)}{r-s} \right) &= \Lambda^*_{\frac{l-1}{n},j}\left(\int_s^r \dot{\gamma}(t)\,\mathrm{d}t \right) \\ &\leqslant \frac{1}{r-s} \int_s^r \Lambda^*_{\frac{l-1}{n},j}\left(\dot{\gamma}(t)\right)\,\mathrm{d}t \\ &\leqslant \frac{\alpha}{r-s}, \end{aligned}$$

where in the last line we used that fact that $\Lambda^*_{t,j}$ is nonnegative for all $t \in [0,1]$.

Now, since

$$\Lambda^*_{t,j}(x) = \sup_{\lambda\in\mathbb{R}} \{\lambda x - \Lambda_{t,j}(\lambda)\},$$

by considering $\lambda = M$ or $\lambda = -M$, we find that

$$\Lambda^*_{t,j}(x) \geqslant M|x| - \max\{\Lambda_{t,j}(M), \Lambda_{t,j}(-M)\}$$

for all $M > 0$. Using this, we have

$$\begin{aligned} &|\gamma(r) - \gamma(s)| \\ &\leqslant \frac{r-s}{M}\Lambda^*_{\frac{l-1}{n},j}\left(\frac{\gamma(r)-\gamma(s)}{r-s}\right) + \frac{r-s}{M}\max\left\{\Lambda_{\frac{l-1}{n},j}(M), \Lambda_{\frac{l-1}{n},j}(-M)\right\} \\ &\leqslant \frac{1}{M}\left(\int_s^r \Lambda^*_{\frac{l-1}{n},j}(\dot\gamma(t))\,\mathrm{d}t + (r-s)\max\left\{\Lambda_{\frac{l-1}{n},j}(M), \Lambda_{\frac{l-1}{n},j}(-M)\right\}\right). \end{aligned} \tag{6.2.7}$$

If we simply have that $0 \leqslant s < r \leqslant 1$, we find by the triangle inequality that

$$\begin{aligned} &|\gamma(r) - \gamma(s)| \\ &\leqslant \left|\gamma\left(\frac{\lfloor ns\rfloor + 1}{n}\right) - \gamma(s)\right| + \sum_{l=\lfloor ns\rfloor+1}^{\lfloor nr\rfloor}\left|\gamma\left(\frac{l+1}{n}\right) - \gamma\left(\frac{l}{n}\right)\right| + \left|\gamma(r) - \gamma\left(\frac{\lfloor nr\rfloor}{n}\right)\right|. \end{aligned}$$

Now, we wish to find a choice for M, such that we can estimate the second term in (6.2.7) independent of l. For this, we need to show that for any $\delta > 0$, there exist $M_j(\delta) > 0$ satisfying $\lim_{\delta\to 0} M_j(\delta) = \infty$ and such that

$$\Lambda_{t,j}(M_j(\delta)) \leqslant \frac{1}{\delta}, \qquad \Lambda_{t,j}(-M_j(\delta)) \leqslant \frac{1}{\delta}$$

for all $t \in [0,1]$. To this end, note that $(t,\lambda) \mapsto \Lambda_{t,j}(\lambda)$ is continuous, and hence,

$$\Lambda_j(\lambda) := \sup_{t\in[0,1]} \Lambda_{t,j}(\lambda)$$

is lower-semicontinuous. As a consequence, the sets $\{\Lambda_j \leqslant \frac{1}{\delta}\}$ are closed, and increasing to $\mathbb{R}$. From this it follows that we can find a sequence $M_j(\delta)$ with $\lim_{\delta\to 0} M_j(\delta) = \infty$ and such that

$$\Lambda_j(M_j(\delta)) \leqslant \frac{1}{\delta}, \qquad \Lambda_j(-M_j(\delta)) \leqslant \frac{1}{\delta}.$$

Since $\Lambda_{t,j} \leqslant \Lambda_j$, the sequence has the desired properties.

Now, given $r > s$ with $r - s \leqslant \delta$, using the sequence $M_j(\delta)$ constructed above, we find that

$$|\gamma(r) - \gamma(s)|$$

$$\leqslant \left|\gamma\left(\frac{\lfloor ns\rfloor+1}{n}\right)-\gamma(s)\right| + \sum_{l=\lfloor ns\rfloor+1}^{\lfloor nr\rfloor}\left|\gamma\left(\frac{l+1}{n}\right)-\gamma\left(\frac{l}{n}\right)\right| + \left|\gamma(r)-\gamma\left(\frac{\lfloor nr\rfloor}{n}\right)\right|$$

$$\begin{aligned}
&\leqslant \frac{1}{M_j(\delta)}\left(\int_s^{\frac{\lfloor ns\rfloor+1}{n}} \Lambda^*_{\frac{\lfloor ns\rfloor}{n},j}(\dot\gamma(t))\,\mathrm{d}t + \left(\frac{\lfloor ns\rfloor+1}{n}-s\right)\frac{1}{\delta}\right)\\
&\quad + \frac{1}{M_j(\delta)}\sum_{l=\lfloor ns\rfloor+1}^{\lfloor nr\rfloor-1}\left(\int_{\frac{l}{n}}^{\frac{l+1}{n}} \Lambda^*_{\frac{l}{n},j}(\dot\gamma(t))\,\mathrm{d}t + \frac{1}{n\delta}\right)\\
&\quad + \frac{1}{M_j(\delta)}\left(\int_{\frac{\lfloor nr\rfloor}{n}}^{r} \Lambda^*_{\frac{\lfloor nr\rfloor}{n},j}(\dot\gamma(t))\,\mathrm{d}t + \left(r-\frac{\lfloor nr\rfloor}{n}\right)\frac{1}{\delta}\right)\\
&= \frac{1}{M_j(\delta)}\left(\int_s^r \Lambda^*_{\frac{\lfloor nt\rfloor}{n},j}(\dot\gamma(t))\,\mathrm{d}t + \frac{r-s}{\delta}\right)\\
&\leqslant \frac{\alpha+1}{M_j(\delta)}. \qquad (6.2.8)
\end{aligned}$$

Here, in the last line we used the fact that $\gamma \in K_{\alpha,n}$ and the fact that $r-s<\delta$. Now, since $\lim_{\delta\to 0} M_j(\delta)=\infty$, given $\varepsilon>0$, we can choose $\delta>0$ independent of γ such that $\frac{\alpha+1}{M_j(\delta)}<\varepsilon$, in which case $|r-s|<\delta$ implies that $|\gamma(r)-\gamma(s)|<\varepsilon$. This proves the equicontinuity of K_α.

To prove the boundedness of K_α, we can take $s=0$ and $\delta=1$ in (6.2.8), giving us that

$$|\gamma(t)| = |\gamma(t)-\gamma(0)| \leqslant \frac{\alpha+1}{M_j(1)}.$$

Here, we used that $\gamma(0)=0$. We find that $||\gamma||_\infty \leqslant \frac{\alpha+1}{M_j(1)}$ for all $\gamma\in K_\alpha$, hence K_α is bounded. □

With all the preparations done, we can prove Theorem 6.2.1.

Proof of Theorem 6.2.1. By Proposition 6.2.4 we find that the sequence $\{\tilde Z_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $\mathcal{X}$ with good rate function I as in (6.2.2). Now observe that the rate function I is infinite outside $C_0([0,1];\mathbb{R}^d)$. Furthermore, for every n, $\tilde Z_n(\cdot)$ is almost surely contained in $C_0([0,1];\mathbb{R}^d)$. Therefore, by Lemma 4.1.5 in [29] we find that $\{\tilde Z_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $C_0([0,1];\mathbb{R}^d)$ with the topology of pointwise convergence. Because the sequence $\{\tilde Z_n(\cdot)\}_{n\geqslant 1}$ is exponentially tight in $C_0([0,1];\mathbb{R}^d)$ with the supremum norm topology by Proposition 6.2.7, we can strengthen the large deviation principle to this space. Since $C_0([0,1];\mathbb{R}^d)$ is closed in $L^\infty([0,1];\mathbb{R}^d)$, we conclude that $\{\tilde Z_n(\cdot)\}_{n\geqslant 1}$ satisfies in $L^\infty([0,1];\mathbb{R}^d)$ the large deviation principle with rate function I. Finally, by Proposition 6.2.3, $\{\tilde Z_n(\cdot)\}_{n\geqslant 1}$ is exponentially equivalent in $L^\infty([0,1];\mathbb{R}^d)$ to $\{Z_n(\cdot)\}_{n\geqslant 1}$. From this we conclude that $\{Z_n(\cdot)\}_{n\geqslant 1}$ satisfies in $L^\infty([0,1];\mathbb{R}^d)$ the large deviation principle with rate function I as desired. □

6.3. Inhomogeneous Schilder-type theorem

In [93], Schilder's theorem is extended to hold also for processes generated by weighted Laplacians, i.e., operators of the form

$$A = \frac{1}{2} \sum_{i,j=1}^{d} G_{ij} \partial_i \partial_j$$

for some positive definite matrix G. One can view this process as a Riemannian Brownian motion when we equip $\mathbb{R}^d$ with the inner product $\langle v, w \rangle_G = \langle v, G^{-1} w \rangle$. In this section we prove the time-inhomogeneous analogue of this.

To this end, consider a collection $\{G(t)\}_{t \in [0,1]}$ of symmetric, positive definite matrices depending continuously on t. Define the operators

$$A_t = \frac{1}{2} \sum_{i,j=1}^{d} G_{ij}(t) \partial_i \partial_j$$

for $t \in [0, 1]$. We say a process $W(t)$ is *generated* by the time-dependent operator A_t if for every $f \in C_c^\infty(\mathbb{R}^d)$ we have that

$$f(W(t)) - f(W(0)) - \int_0^t A_s f(W(s)) \, \mathrm{d}s$$

is a martingale. A continuous process $W(\cdot)$ generated by A_t with $W(0) = 0$ is called a $G(t)$-*Brownian motion* on $\mathbb{R}^d$. Such a process exists, since we can take

$$W(t) = \int_0^t \sqrt{G(t)} \, \mathrm{d}B(t),$$

where $B(\cdot)$ is a standard $\mathbb{R}^d$-valued Brownian motion. From this observation we obtain the following property of a $G(t)$-Brownian motion.

Proposition 6.3.1. *Let $\{G(t)\}_{t \in [0,1]}$ be a collection of symmetric, positive definite matrices depending continuously on t. Let $W(\cdot)$ be a $G(t)$-Brownian motion and let $\{\mathcal{F}_s\}_{s \in [0,1]}$ be its natural filtration. Then for every $s < r$, $W(r) - W(s)$ is independent of $\mathcal{F}_s$ and has a multivariate normal distribution with mean 0 and convariance matrix $C_{s,r} = \int_s^r G(t) \, \mathrm{d}t$.*

Now define for every $n \geqslant 1$ the process $W_n(t) := \frac{1}{\sqrt{n}} W(t)$, where $W(t)$ is a $G(t)$-Brownian motion. We show that the large deviations for $\{W_n(\cdot)\}_{n \geqslant 1}$ may be obtained from Theorem 6.2.1. Before we get to the theorem, we remark that Proposition 6.3.1 shows that $\{W_n(\cdot)\}_{n \geqslant 1}$ is a sequence of Gaussian processes with small covariance operators. Therefore, the result obtained in Theorem 6.3.2 is a special case of [30, Theorem 3.4.5], which in turn is a generalization of the results in [33]. Their proof is based on the observation that $W_n(\cdot)$ can be written as the empirical average of independent copies of $W(\cdot)$. The result then follows from Cramér's theorem for Banach spaces, see e.g. [32, 30, 29]. For our method, we do not require this representation as empirical averages. Additionally, our approach is general in that it shows how to use random walk approximations to study large deviations for continuous-time processes in the setting of time-inhomogeneous processes.

We now state a time-inhomogeneous variant of Schilder's theorem in $\mathbb{R}^d$ and prove it using Mogulskii's theorem for time-inhomogeneous random walks (Theorem 6.2.1).

Theorem 6.3.2. *Let $\{G(t)\}_{t\in[0,1]}$ be a collection of positive definite matrices, such that $t \mapsto G(t)$ is Lipschitz. Let $W(t)$ be a $G(t)$-Brownian motion and define for every $n \geqslant 1$ the process $W_n(t) := \frac{1}{\sqrt{n}}W(t)$. Then $\{W_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $C_0([0,1];\mathbb{R}^d)$ with good rate function*

$$I(\gamma) = \begin{cases} \frac{1}{2}\int_0^1 \langle G(t)^{-1}\dot{\gamma}(t), \dot{\gamma}(t)\rangle \, \mathrm{d}t, & \gamma \in AC_0([0,1];\mathbb{R}^d) \\ \infty, & \textit{otherwise.} \end{cases} \tag{6.3.1}$$

Remark 6.3.3. If we define the inner products $\langle v, w\rangle_{G(t)} := \langle v, G^{-1}(t)w\rangle$, the rate function in (6.3.1) may be written as

$$I(\gamma) = \frac{1}{2}\int_0^1 ||\dot{\gamma}(t)||^2_{G(t)} \, \mathrm{d}t$$

whenever $\gamma \in AC_0([0,1];\mathbb{R}^d)$.

Before we get to the proof of Theorem 6.3.2, we first need some preliminary results.

Proposition 6.3.4. *Let $\{G(t)\}_{t\in[0,1]}$ be a collection of symmetric, positive definite matrices depending continuously on t. Let $W(t)$ be a $G(t)$-Brownian motion and define for every $n \geqslant 1$ the process $W_n(t) := \frac{1}{\sqrt{n}}W(t)$. Furthermore, set*

$$\widetilde{W}_n(t) = W_n\left(\frac{\lfloor nt\rfloor}{n}\right) \tag{6.3.2}$$

for every $n \geqslant 1$ and $t \in [0,1]$. Then $\{W_n(\cdot)\}_{n\geqslant 1}$ and $\{\widetilde{W}_n(\cdot)\}_{n\geqslant 1}$ are exponentially equivalent in $L^\infty([0,1];\mathbb{R}^d)$.

Proof. First of all, note that by the union of events bound we have

$$\mathbb{P}\left(\sup_{t\in[0,1]} |W_n(t) - \widetilde{W}_n(t)| \geqslant \delta\right) \leqslant \sum_{l=1}^d \mathbb{P}\left(\sup_{t\in[0,1]} |W_n^l(t) - \widetilde{W}_n^l(t)| \geqslant \frac{\delta}{\sqrt{d}}\right).$$

Furthermore, for every $l = 1, \ldots, d$ we have

$$\sup_{t\in[0,1]} |W_n^l(t) - \widetilde{W}_n^l(t)| = \sup_{0\leqslant i\leqslant n-1} \sup_{t\in[\frac{i}{n},\frac{i+1}{n}]} \left|W_n^l(t) - W_n^l\left(\frac{i}{n}\right)\right|.$$

This implies that

$$\mathbb{P}\left(\sup_{t\in[0,1]}|W_n^l(t)-\widetilde{W}_n^l(t)|\geqslant\frac{\delta}{\sqrt{d}}\right)\leqslant\sum_{i=0}^{n-1}\mathbb{P}\left(\sup_{t\in[\frac{i}{n},\frac{i+1}{n}]}\left|W_n^l(t)-W_n^l\left(\frac{i}{n}\right)\right|\geqslant\frac{\delta}{\sqrt{d}}\right).$$

Now, it follows from Doob's inequality (see e.g. [37]) that

$$\begin{aligned}\mathbb{P}&\left(\sup_{t\in[\frac{i}{n},\frac{i+1}{n}]}\left|W_n^l(t)-W_n^l\left(\frac{i}{n}\right)\right|\geqslant\frac{\delta}{\sqrt{d}}\right)\\&\leqslant 2\mathbb{P}\left(\sup_{t\in[\frac{i}{n},\frac{i+1}{n}]}e^{n\lambda\left(W_n^l(t)-W_n^l\left(\frac{i}{n}\right)\right)}\geqslant e^{n\lambda\delta d^{-\frac{1}{2}}}\right)\\&\leqslant 2e^{-n\lambda\delta d^{-\frac{1}{2}}}\mathbb{E}\left(e^{n\lambda\left(W_n^l\left(\frac{i+1}{n}\right)-W_n^l\left(\frac{i}{n}\right)\right)}\right)\end{aligned}$$

for $\lambda>0$. From Proposition 6.3.1 it follows that $W_n\left(\frac{i+1}{n}\right)-W_n\left(\frac{i}{n}\right)$ has a multivariate normal distribution with mean 0 and covariance matrix $\frac{1}{n}\int_{\frac{i}{n}}^{\frac{i+1}{n}}G(t)\,\mathrm{d}t$. Therefore, we find that $W_n^l\left(\frac{i+1}{n}\right)-W_n^l\left(\frac{i}{n}\right)$ has a normal distribution with mean 0 and variance $\frac{1}{n}\int_{\frac{i}{n}}^{\frac{i+1}{n}}G_{ll}(t)\,\mathrm{d}t$. From this it follows that

$$\mathbb{E}\left(e^{n\lambda\left(W_n^l\left(\frac{i+1}{n}\right)-W_n^l\left(\frac{i}{n}\right)\right)}\right)=e^{\frac{1}{2}\lambda^2 n\int_{\frac{i}{n}}^{\frac{i+1}{n}}G_{ll}(t)\,\mathrm{d}t}.$$

Now, since $t\mapsto G(t)$ is continuous, it is bounded on $[0,1]$. Therefore, there exists a constant $C>0$ such that for all $t\in[0,1]$ and all $l=1,\dots,d$ we have $G_{ll}(t)\leqslant C$. Using this bound, we find that

$$\mathbb{E}\left(e^{n\lambda\left(W_n^l\left(\frac{i+1}{n}\right)-W_n^l\left(\frac{i}{n}\right)\right)}\right)\leqslant e^{\frac{1}{2}C\lambda^2}.$$

Putting all estimates together, we obtain

$$\mathbb{P}\left(\sup_{t\in[0,1]}|W_n(t)-\widetilde{W}_n(t)|\geqslant\delta\right)\leqslant 2nde^{-n\lambda\delta d^{-\frac{1}{2}}}e^{\frac{1}{2}C\lambda^2}.$$

From this, it follows that for every $\lambda>0$ we have

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\sup_{t\in[0,1]}|W_n(t)-\widetilde{W}_n(t)|\geqslant\delta\right)\leqslant-\frac{\lambda\delta}{\sqrt{d}}.$$

Letting λ tend to infinity now proves the claim. □

The piecewise constant process defined in (6.3.2) can be written as

$$\widetilde{W}_n(t)=\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}X_i^n,$$

where by Proposition 6.3.1, X_i^n follows a multivariate normal distribution with mean 0 and covariance matrix $n\int_{\frac{i-1}{n}}^{\frac{i}{n}} G(t)\,\mathrm{d}t$. In order to apply Theorem 6.2.1, we need to find a collection $\{\mu_t\}_{t\in[0,1]}$ of probability measures, such that X_i^n is distributed according to $\mu_{\frac{i-1}{n}}$. Unfortunately, in our case, the distributions will also depend on n. However, for n large, the covariance matrix of X_i^n can be approximated by $G\left(\frac{i-1}{n}\right)$. This inspires the following proposition.

Proposition 6.3.5. *Let the assumptions of Theorem 6.3.2 be satisfied. For every $n \geqslant 1$, let $\widetilde{W}_n(\cdot)$ be the process as defined in* (6.3.2)*. Let $\{\mu_t\}_{t\in[0,1]}$ be the collection of measures given by $\mu_t = \mathcal{N}(0, G(t))$. Denote by*

$$Z_n(t) = \sum_{i=1}^{\lfloor nt\rfloor} Y_i^n$$

the time-inhomogeneous random walk associated to the collection $\{\mu_t\}_{t\in[0,1]}$ as defined in (6.2.1)*. Then $\{\widetilde{W}_n(\cdot)\}_{n\geqslant 1}$ and $\{Z_n(\cdot)\}_{n\geqslant 1}$ are exponentially equivalent in $L^\infty([0,1];\mathbb{R}^d)$.*

Proof. For every $n \geqslant 1$, let $Y_1^n, \ldots, Y_n^n$ be independent, with Y_i^n distributed according to $\mu_{\frac{i-1}{n}}$. Define

$$\tilde{X}_i^n = \left(n\int_{\frac{i-1}{n}}^{\frac{i}{n}} G(t)\,\mathrm{d}t\right)^{\frac{1}{2}} G\left(\frac{i-1}{n}\right)^{-\frac{1}{2}} Y_i^n.$$

Then $\tilde{X}_i^n$ has a multivariate normal distribution with mean 0 and covariance matrix $n\int_{\frac{i-1}{n}}^{\frac{i}{n}} G(t)\,\mathrm{d}t$. Therefore, $\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}\tilde{X}_i^n$ equals $\widetilde{W}_n(t)$ in distribution. Now note that

$$\left|\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}\tilde{X}_i^n - \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} Y_i^n\right| \leqslant \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}|\tilde{X}_i^n - Y_i^n| \leqslant \frac{1}{n}\sum_{i=1}^{n}|\tilde{X}_i^n - Y_i^n|.$$

Plugging in the definition of $\tilde{X}_i^n$, we can estimate further to find that

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}\tilde{X}_i^n - \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} Y_i^n\right| \\
&\quad\leqslant \frac{1}{n}\sum_{i=1}^{n}\left\|\left(n\int_{\frac{i-1}{n}}^{\frac{i}{n}} G(t)\,\mathrm{d}t\right)^{\frac{1}{2}} G\left(\frac{i-1}{n}\right)^{-\frac{1}{2}} - I\right\| |Y_i^n| \\
&\quad\leqslant \frac{1}{n}\sum_{i=1}^{n}\left\|G\left(\frac{i-1}{n}\right)^{-\frac{1}{2}}\right\| \left\|\left(n\int_{\frac{i-1}{n}}^{\frac{i}{n}} G(t)\,\mathrm{d}t\right)^{\frac{1}{2}} - G\left(\frac{i-1}{n}\right)^{\frac{1}{2}}\right\| |Y_i^n|.
\end{aligned}$$

Since $G(t)$ is symmetric positive definite, we find that

$$||G(t)^{-\frac{1}{2}}|| = ||G(t)||^{-\frac{1}{2}}.$$

Because $t \mapsto G(t)$ is continuous and $||G(t)|| > 0$ for all $t \in [0,1]$, there exists an $\eta > 0$ such that $||G(t)|| \geqslant \eta$ for all $t \in [0,1]$. From this, we obtain the uniform bound

$$||G(t)^{-\frac{1}{2}}|| \leqslant \frac{1}{\sqrt{\eta}}.$$

Furthermore, the continuity of $t \mapsto G(t)$ also gives, by the mean value theorem, that there exists a $\xi \in (\frac{i-1}{n}, \frac{i}{n})$ such that

$$n \int_{\frac{i-1}{n}}^{\frac{i}{n}} G(t)\,\mathrm{d}t = G(\xi).$$

Using this, together with the fact that $t \mapsto G(t)$ is Lipschitz, say with constant $L > 0$, we find that

$$\begin{aligned}\left\| n \int_{\frac{i-1}{n}}^{\frac{i}{n}} G(t)\,\mathrm{d}t - G\left(\frac{i-1}{n}\right) \right\| &= \left\| G(\xi) - G\left(\frac{i-1}{n}\right) \right\| \\ &\leqslant L \left| \xi - \frac{i-1}{n} \right| \\ &\leqslant L\frac{1}{n},\end{aligned}$$

where we used that $\xi \in (\frac{i-1}{n}, \frac{i}{n})$. Now, since the square root is Lipschitz on the set of symmetric positive definite matrices with norm bounded away from 0, we find that there exists a possibly different constant $L > 0$ such that

$$\left\| \left(n \int_{\frac{i-1}{n}}^{\frac{i}{n}} G(t)\,\mathrm{d}t \right)^{\frac{1}{2}} - G\left(\frac{i-1}{n}\right)^{\frac{1}{2}} \right\| \leqslant L' \left\| n \int_{\frac{i-1}{n}}^{\frac{i}{n}} G(t)\,\mathrm{d}t - G\left(\frac{i-1}{n}\right) \right\| \leqslant L\frac{1}{n}.$$

Collecting everything, we find that

$$\left| \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} \tilde{X}_i^n - \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} Y_i^n \right| \leqslant \frac{L}{\sqrt{\eta}} \frac{1}{n^2} \sum_{i=1}^{n} |Y_i^n|.$$

But then we find that

$$\begin{aligned}\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} \tilde{X}_i^n - \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} Y_i^n \right| > \delta \right) &\leqslant \mathbb{P}\left(\frac{L}{\sqrt{\eta}} \frac{1}{n^2} \sum_{i=1}^{n} |Y_i^n| > \delta \right) \\ &= \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} |Y_i^n| > \frac{n\delta\sqrt{\eta}}{L} \right)\end{aligned}$$

$$\begin{aligned}
&\leqslant \sum_{i=1}^{n} \mathbb{P}\left(\frac{1}{n}|Y_i^n| > \frac{\delta\sqrt{\eta}}{L}\right) \\
&\leqslant \sum_{l=1}^{d}\sum_{i=1}^{n} \mathbb{P}\left(|(Y_i^n)^l| > \frac{n\delta\sqrt{\eta}}{L\sqrt{d}}\right) \\
&\leqslant 2e^{-n\lambda\delta\eta^{\frac{1}{2}}d^{-\frac{1}{2}}L^{-1}} \sum_{l=1}^{d}\sum_{i=1}^{n} \mathbb{E}\left(e^{\lambda(Y_i^n)^l}\right).
\end{aligned}$$

Here we used the union bound in the fourth line and Markov's inequality in the last line.

Since $(Y_i^n)^l$ has a normal distribution with mean 0 and variance $G_{ll}(\frac{i-1}{n})$, we find that

$$\mathbb{E}\left(e^{\lambda(Y_i^n)^l}\right) = e^{\frac{1}{2}\lambda^2 G_{ll}(\frac{i-1}{n})} \leqslant e^{\frac{1}{2}\lambda^2 C},$$

because $t \mapsto G(t)$ is bounded. Combining the above estimates, we find that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} \tilde{X}_i^n - \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} Y_i^n\right| > \delta\right) \leqslant 2nde^{-n\lambda\delta\eta^{\frac{1}{2}}d^{-\frac{1}{2}}L^{-1}} e^{\frac{1}{2}\lambda^2 C},$$

from which it follows that

$$\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} \tilde{X}_i^n - \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} Y_i^n\right| > \delta\right) \leqslant -\frac{\lambda\delta\sqrt{\eta}}{L\sqrt{d}}.$$

Considering the limit $\lambda \to \infty$ now proves the claim. ☐

With all preparations done, we are ready to prove Theorem 6.3.2

Proof of Theorem 6.3.2. Define the measures $\mu_t = \mathcal{N}(0, G(t))$ and let $Z_n(\cdot)$ be the associated time-inhomogeneous random walk as defined in (6.2.1). Note that for every $t \in [0,1]$ and every $\lambda \in \mathbb{R}^d$, we have

$$\Lambda_t(\lambda) = \frac{1}{2}\langle \lambda, G(t)\lambda\rangle.$$

Since $t \mapsto G(t)$ is continuous, it follows that $(t,\lambda) \mapsto \Lambda_t(\lambda)$ is continuous. Furthermore, we have

$$\int_0^r \Lambda_t(\lambda)\,\mathrm{d}t = \frac{1}{2}\left\langle \lambda, \left(\int_0^r G(t)\,\mathrm{d}t\right)\lambda\right\rangle,$$

which is differentiable with respect to λ. Therefore, by Theorem 6.2.1, we have that $\{Z_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $L^\infty([0,1];\mathbb{R}^d)$ with good rate function

$$I(\gamma) = \begin{cases} \int_0^1 \Lambda_t^*(\dot{\gamma}(t))\,\mathrm{d}t, & \gamma \in AC_0([0,1];\mathbb{R}^d), \\ \infty, & \text{otherwise.} \end{cases}$$

We can compute

$$\Lambda_t^*(\dot\gamma(t)) = \sup_{\lambda\in\mathbb{R}^d}\left\{\langle\lambda,\dot\gamma(t)\rangle - \frac{1}{2}\langle\lambda, G(t)\lambda\rangle\right\} = \frac{1}{2}\langle\dot\gamma(t), G(t)^{-1}\dot\gamma(t)\rangle.$$

It follows that the rate function reduces to

$$I(\gamma) = \begin{cases} \frac{1}{2}\int_0^1\langle\dot\gamma(t), G(t)^{-1}\dot\gamma(t)\rangle\,\mathrm{d}t, & \gamma\in AC_0([0,1];\mathbb{R}^d),\\ \infty, & \text{otherwise.}\end{cases}$$

Now, from Proposition 6.3.5 and 6.3.4 it follows that $\{Z_n(\cdot)\}_{n\geqslant 1}$ and $\{W_n(\cdot)\}_{n\geqslant 1}$ are exponentially equivalent in $L^\infty([0,1];\mathbb{R}^d)$. Therefore, by Theorem 2.1.8 it follows that $\{W_n(\cdot)\}_{n\geqslant 1}$ satisfies the large deviation principle in $L^\infty([0,1];\mathbb{R}^d)$ with the same rate function I. Finally, noticing that for every $n\geqslant 1$, $W_n(\cdot)$ is almost surely contained in $C_0([0,1];\mathbb{R}^d)$, together with the fact that the domain of the rate function is contained in $C_0([0,1];\mathbb{R}^d)$, the large deviation principle actually holds in $C_0([0,1];\mathbb{R}^d)$ as desired. □

7

Large deviations for $g(t)$-Brownian motion

In this chapter we continue the study of large deviations for time-inhomogeneous processes. We extend the results obtained in Section 6.3 to the setting of Riemannian manifolds. More precisely, we study the large deviation behaviour of Riemannian Brownian motion with small variance in evolving Riemannian manifolds. We follow the approach taken in Section 5.4 by constructing time-dependent variants of the horizontal lift and anti-development of curves. The results presented in this chapter are based on:

Rik Versendaal. "Large deviations for Brownian motion in evolving Riemannian manifolds". In: *Preprint; ArXiv:* 2004.00358 (2020). ArXiv: `2004.00358`

In the past decades, the study of evolving Riemannian manifolds has received a lot of attention. The treatment of stochastic processes in this setting was initiated in [5], where Brownian motion with respect to a collection of time-dependent metrics is defined. The existence of this process is proven, and the gradient of the associated heat-semigroup is studied when the metric evolves under the Ricci-flow. This is further developed in [23]. More generally, in [49], the theory of martingales with respect to a time-dependent connection is studied. Finally, the central limit problem for geodesic random walks in this setting is considered in [64].

In [24], the so-called Onsager-Machlup functional is studied for elliptic diffusions on manifolds with time-dependent metric. It is shown that the probability that a Brownian motion deviates from a smooth curve by at most a distance $\frac{1}{n} > 0$ decays exponentially in n. More precisely, if X_t is a Brownian motion with respect to a time-dependent metric $\{g(t)\}_{0 \leqslant t \leqslant 1}$, and $\gamma : [0,1] \to M$ is a smooth curve, it is proven that for n large

$$\mathbb{P}\left(\sup_{0 \leqslant t \leqslant 1} d_t(X_t, \gamma(t)) \leqslant \frac{1}{n}\right) \sim e^{-n^2 C} \exp\left\{\int_0^1 -\frac{1}{2}|\dot\gamma(t)|^2_{g(t)} - \frac{1}{12} R_{g(t)}(\gamma(t)) + \frac{1}{4}\mathrm{Tr}_{g(t)}(\partial_1 g(t))\,\mathrm{d}t\right\}.$$

Here, d_t is the Riemannian distance associated to the metric $g(t)$ and $R_{g(t)}$ is the scalar curvature of the metric $g(t)$. Furthermore, $\mathrm{Tr}_{g(t)}(\partial_1 g(t))$ denotes the trace of the time-derivative $\partial_1 g(t)$ with respect to the metric $g(t)$. More precisely, it is the trace of the linear map $X \mapsto \partial_1 g(t)(X, \cdot)^{\#}$. The result is an extension of the time-homogeneous case, in which the term containing the derivative $\partial_1 g(t)$ is non-existent.

A result related to this is Schilder's theorem, which is concerned with the large deviations for Brownian paths with small variance. In Chapter 5 we discussed Schilder's theorem for Riemannian Brownian motion in a (stationary) Riemannian manifold (M, g). More precisely, the result states that on the exponential scale we have

$$\mathbb{P}\left(X^n \approx \gamma\right) \approx \exp\left\{-\frac{n}{2}\int_0^1 |\dot{\gamma}(t)|_g^2 \,\mathrm{d}t\right\},$$

where $X_t^n = X_{tn^{-1}}$ with X_t a Riemannian Brownian motion. Our aim is to extend this result to the context of a Riemannian manifold with a time-dependent metric. For this, we follow the approach taken in Section 5.4, where we prove Schilder's theorem by lifting the process to the orthonormal frame bundle, and embedding this into some Euclidean space in order to use Freidlin-Wentzell theory. To carry out this procedure in the time-inhomogeneous case, we define an appropriate way of lifting a Brownian motion with respect to a time-dependent metric to the frame bundle over the manifold. Furthermore, we also adapt Freidlin-Wentzell theory to the setting where the drift and diffusion constants are time-dependent. Finally, to reduce to compact sets, we extend the compact containment argument from Proposition 5.2.10 to Markov processes with time-dependent generators.

This chapter is organized as follows. In Section 7.1 we introduce the notion of a Brownian motion with respect to a time-dependent metric and state the main result, the analogue of Schilder's theorem. Additionally, we sketch the approach to proving this result. Section 7.2 is devoted to extending the notion of horizontal lift and antidevelopment of curves to the time-inhomogeneous case. Finally, in Section 7.3 we provide all details of the proof of our main result.

7.1. Main result

Following [5, 23], we define Brownian motion with respect to a collection of metrics $\{g(t)\}_{t\in[0,1]}$. We state our main result concerning the large deviations for such processes and give a sketch of its proof.

7.1.1. $g(t)$-Brownian motion

Let M be a manifold, which in our case always means it is smooth, finite-dimensional and second countable. Let $\mathcal{G} = \{g(t)\}_{t\in[0,1]}$ be a collection of Riemannian metrics on M, smoothly depending on t. We will interchangeably use $\mathcal{G}$ and $\{g(t)\}_{t\in[0,1]}$ to refer to this collection of metrics. For $x \in M$ and $v, w \in T_xM$ we write $\langle v, w\rangle_{g(t)}$ for the inner product of v and w with respect to the metric $g(t)$. Furthermore, we denote by ∇^t the Levi-Civita connection of $g(t)$ and by Δ_M^t the associated Laplace-Beltrami operator.

We can now define a Brownian motion with respect to a collection of metrics $\{g(t)\}_{t\in[0,1]}$. We follow the definition in [23], which is equivalent to the definition in [5].

Definition 7.1.1. *Let M be a manifold and let $\{g(t)\}_{t\in[0,1]}$ be a collection of Riemannian metrics on M, smoothly depending on t. A process X_t is called a $g(t)$-* Brownian motion *if it is continuous and if for all $f \in C^\infty(M)$,*

$$f(X_t) - f(X_0) - \frac{1}{2}\int_0^t \Delta_M^s f(X_s)\,\mathrm{d}s$$

is a local martingale. In that case, we say that X_t is generated *by (the time-dependent generator) Δ_M^t.*

In general, a $g(t)$-Brownian motion only exists up to some explosion time $e(X)$. In the time-homogeneous setting we have that if the Ricci-curvature is bounded from below, then $e(X)$ is almost surely infinite, see Proposition 2.4.15. This result is extended to the time-inhomogeneous setting in [65] by requiring that $g(t)$ evolves under a backwards super Ricci flow, i.e., $g(t)$ satisfies

$$\partial_1 g(t) \leqslant \mathrm{Ric}_{g(t)}.$$

In that case, $g(t)$-Brownian motion exists up to time T for every $T > 0$.

7.1.2. Statement of the main result

Next, we state the main result, which is the analogue of Schilder's theorem for a $g(t)$-Brownian motion. Before we do this, we first need to introduce a proper rescaling of a $g(t)$-Brownian motion.

To motivate the rescaling, note that by Theorem 5.1.3, if X_t is a Riemannian Brownian motion, then

$$\mathbb{P}\left(X^n \approx \gamma\right) \approx \exp\left\{-\frac{n}{2}\int_0^1 |\dot{\gamma}(t)|_g^2\,\mathrm{d}t\right\},$$

where $X_t^n = X_{tn^{-1}}$. Since X_t is generated by $\frac{1}{2}\Delta_M$, a substitution yields that X_t^n is generated by $\frac{1}{2n}\Delta_M$.

In the time-inhomogeneous setting, we want the process X_t^n to evolve according to a collection of metrics $\{g(t)\}_{t\in[0,1]}$. Therefore, we have to consider X_t as a $g(nt)$-Brownian motion, i.e., X_t is generated by $\frac{1}{2}\Delta_M^{nt}$. In that case, substitution yields that the process X_t^n is generated by $\frac{1}{2n}\Delta_M^{n^{-1}nt} = \frac{1}{2n}\Delta_M^t$. Our main result gives the large deviations for $\{X^n\}_{n\geqslant 1}$.

Before we give the precise statement, let us relate the above construction to the result in Section 6.3. Let W_t be a $G(t)$-Brownian motion in $\mathbb{R}^d$ in the sense of

Section 6.3. This is a $G(t)$-Brownian motion in the sense of Definition 7.1.1 if we equip $\mathbb{R}^d$ with the time-dependent inner product

$$\langle v, w\rangle_{G(t)} = \langle v, G^{-1}(t)w\rangle.$$

We have

$$W_t^{G(t)} = \int_0^t \sqrt{G(s)}\,\mathrm{d}B_s,$$

where B_t is a standard, $\mathbb{R}^d$-valued Brownian motion. Similarly, a $G(nt)$-Brownian motion is given by

$$W_t^{G(nt)} = \int_0^t \sqrt{G(ns)}\,\mathrm{d}B_s.$$

But then we find that

$$W_{tn^{-1}}^{G(nt)} = \int_0^{tn^{-1}} \sqrt{G(ns)}\,\mathrm{d}B_s = \int_0^t \sqrt{G(s)}\,\mathrm{d}B_{sn^{-1}} = \frac{1}{\sqrt{n}}\int_0^t \sqrt{G(s)}\,\mathrm{d}B_s = \frac{1}{\sqrt{n}}W_t$$

in distribution. Therefore, it follows from Theorem 6.3.2 that

$$\mathbb{P}\left(W_{n^{-1}\cdot}^{G(nt)} \approx \gamma\right) \approx \exp\left\{-\frac{n}{2}\int_0^1 |\dot\gamma(u)|^2_{G(u)}\,\mathrm{d}u\right\}.$$

Our main theorem states that this happens in general. In order to write down the rate function, we define the space

$$H^{1,\mathcal{G}}([0,1];M) = \left\{\gamma : [0,1] \to M \,\middle|\, \gamma \text{ is differentiable a.e. and } \int_0^1 |\dot\gamma(t)|^2_{g(t)}\,\mathrm{d}t < \infty\right\}.$$

We have the following theorem.

Theorem 7.1.2. *Let M be a manifold and let $\{g(t)\}_{t\in[0,1]}$ be a collection of Riemannian metrics on M, smoothly depending on t. Fix $x_0 \in M$, and let X_t be a $g(t)$-Brownian motion with $X_0 = x_0$. Furthermore, for every $n \geqslant 1$, let X_t^n be the continuous process generated by $\frac{1}{2n}\Delta_M^t$ with $X_0^n = x_0$. Assume the processes X_t and X_t^n exist for all time $t \in [0,1]$. Then $\{X^n\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];M)$ with good rate function I_M given by*

$$I_M(\gamma) = \begin{cases} \frac{1}{2}\int_0^1 |\dot\gamma(t)|^2_{g(t)}\,\mathrm{d}t, & \gamma \in H^{1,\mathcal{G}}_{x_0}([0,1];M), \\ \infty & \textit{otherwise.}\end{cases}$$

7.1.3. Sketch of the proof of Theorem 7.1.2

The proof of Theorem 7.1.2 follows the same lines as the proof of Theorem 5.1.3 given in Section 5.4 for the time-homogeneous case. The main work lies in defining a good analogue of the concept of horizontal lift and anti-development in the time-inhomogeneous setting. The detailed construction is given in Section 7.2.

For X_t^n we denote by U_t^n the horizontal lift with respect to $\{g(t)\}_{t\in[0,1]}$ to the frame bundle FM. As explained in Section 7.2.4 (see also [23, 5]), this process satisfies the Stratonovich stochastic differential equation

$$\mathrm{d}U_t^n = H_i(t, U_t^n) \circ \mathrm{d}W_t^{n,i} - \frac{1}{2}(\partial_1 g(t))_{ij}(U_t^n e_i, U_t^n e_j)V^{ij}(U_t^n)\,\mathrm{d}t. \tag{7.1.1}$$

Here, $W_t^n = \frac{1}{\sqrt{n}} W_t$ with W_t a standard, $\mathbb{R}^d$-valued Brownian motion. Furthermore, the vector fields $H_i(t, \cdot)$ are the fundamental horizontal fields with respect to the metric $g(t)$ defined in (2.3.5) and V^{ij} is the canonical basis of vertical vector fields over FM defined in (2.3.1). Finally, $\{e_1, \ldots, e_d\}$ denotes the standard basis of $\mathbb{R}^d$.

By embedding FM smoothly in some Euclidean space $\mathbb{R}^N$, we can push-forward equation (7.1.1) to $\mathbb{R}^N$ to obtain a stochastic differential equation on $\mathbb{R}^N$ with a drift, and a diffusion of order $\frac{1}{\sqrt{n}}$, see Proposition 2.4.9. This shows that, at least if we restrict to compact sets, we can apply Freidlin-Wentzell theory for time-inhomogeneous diffusions (Theorem 7.3.6) to get the large deviations for the embedded process. By the contraction principle (see [29, Theorem 4.2.1], this can then be transferred to the sequence $\{X^n\}_{n\geqslant 1}$. The relation between the derivative of a curve in M and the derivative of its anti-development with respect to $\{g(t)\}_{t\in[0,1]}$ in $\mathbb{R}^d$ then assures that we obtain the correct rate function.

Finally, as shown in Section 7.3.1, we can use a general approach using Lyapunov functions to show that the process X_t^n remains in a compact set with high probability. This, together with the result obtained when restricting to compact sets, allows us to obtain the full result of Theorem 7.1.2.

7.2. Horizontal lift and anti-development

In this section we discuss how to define a horizontal lift with respect to a collection $\{g(t)\}_{t\in[0,1]}$ of Riemannian metrics on a manifold M. In order to do this, we need a suitable definition of what we mean by horizontal curves and horizontal vectors. For this, we need to incorporate time into our analysis. In order for the upcoming constructions to make sense also for $t \notin [0,1]$, we set $g(t) = g(0)$ for $t < 0$ and $g(t) = g(1)$ for $t > 1$.

7.2.1. A time-dependent connection which is metric

Denote spacetime by $\mathbb{M} := \mathbb{R} \times M$ and let $T\mathbb{M}$ be its tangent bundle. For $(t,x) \in \mathbb{M}$ we have $T_{(t,x)}\mathbb{M} = \mathbb{R} \oplus T_x M$. We denote the basis tangent vector in the time-direction by ∂_1.

Instead of considering the tangent bundle $T\mathbb{M}$, we also want to view TM as bundle over $\mathbb{M}$. More precisely, we define the bundle $\overline{TM}$ over $\mathbb{M}$ with fibres given by

$$\overline{TM}_{(t,x)} = T_x M$$

for all $t \in \mathbb{R}$ and all $x \in M$. A smooth section of $\overline{TM}$ is called a time-dependent vector field. We will often write $Z(t) \in \Gamma(\overline{TM})$ to stress that Z is a time-dependent vector field on M.

To define the desired connection on $\overline{TM}$, we first need to define what we mean by the derivative of $g(t)$ with respect to t. This is a 2-tensor $\partial_1 g(t) : TM \times TM \to C^\infty(M)$, which in coordinates is given by

$$\partial_1 g(t)(v, w) = \partial_1 g_{ij}(t) v^i w^j,$$

where $v = v^i \partial_i$, $w = w^j \partial_j$ and $g(t) = g_{ij}(t) \mathrm{d}x^i \otimes \mathrm{d}x^j$.

Furthermore, for $Y \in \Gamma(TM)$, we denote by $(\partial_1 g(t))(Y, \cdot)^{\#_t}$ the vector field obtained by 'raising an index' with respect to the metric $g(t)$. More precisely, it is the unique vector field such that for all vector fields $Z \in \Gamma(TM)$ we have

$$(\partial_1 g(t))(Y, Z) = \langle (\partial_1 g(t))(Y, \cdot)^{\#_t}, Z \rangle_{g(t)},$$

see also Section 2.2.3. Finally, recall that we denote by ∇^t the Levi-Civita connection of the metric $g(t)$.

Following the idea in [52, 53], see also Chapter 6 in [3], we equip the bundle $\overline{TM}$ over $\mathbb{M}$ with a natural connection $\nabla : \Gamma(T\mathbb{M}) \times \Gamma(\overline{TM}) \to \Gamma(\overline{TM})$ given by

$$\begin{cases} \nabla_X Y(t) = \nabla^t_X Y(t), \\ \nabla_{\partial_1} Y(t) = \partial_1 Y(t) + \frac{1}{2}(\partial_1 g(t))(Y(t), \cdot)^{\#_t}, \end{cases} \tag{7.2.1}$$

for $X \in \Gamma(TM)$ a vector field over M and $Y(t) \in \Gamma(\overline{TM})$ a time-dependent vector field over M. By C^∞-linearity, this defines $\nabla_Z Y$ for all $Z \in \Gamma(T\mathbb{M})$ and all $Y \in \Gamma(\overline{TM})$. This connection is compatible with the collection $\{g(t)\}_{t \in [0,1]}$ of Riemannian metrics on M, as we will show in the following proposition.

Proposition 7.2.1. *The connection defined in* (7.2.1) *is metric in the following sense: for all time-dependent vector fields $X(t), Y(t) \in \Gamma(\overline{TM})$ and $Z \in \Gamma(T\mathbb{M})$ we have*

$$Z \langle X(t), Y(t) \rangle_{g(t)} = \langle \nabla_Z X(t), Y(t) \rangle_{g(t)} + \langle X(t), \nabla_Z Y(t) \rangle_{g(t)}$$

for all $t \in \mathbb{R}$.

Proof. Note that $Z \in \Gamma(T\mathbb{M})$ can be written as $Z(t, x) = c_1(t, x)\partial_1 + \tilde{Z}(t)(x)$ where $c_1 : \mathbb{M} \to \mathbb{R}$ is a smooth function and $\tilde{Z}(t) \in \Gamma(\overline{TM})$ a time-dependent vector field over M. Since ∇^t is metric with respect to $g(t)$, we have

$$\begin{aligned} \tilde{Z}(t) \langle X(t), Y(t) \rangle_{g(t)} &= \langle \nabla^t_{\tilde{Z}(t)} X(t), Y(t) \rangle_{g(t)} + \langle X(t), \nabla^t_{\tilde{Z}(t)} Y(t) \rangle_{g(t)} \\ &= \langle \nabla_{\tilde{Z}(t)} X(t), Y(t) \rangle_{g(t)} + \langle X(t), \nabla_{\tilde{Z}(t)} Y(t) \rangle_{g(t)}. \end{aligned}$$

For the derivative with respect to ∂_1, if we write $X(t) = X^i(t)\partial_i$, $Y(t) = Y^j(t)\partial_j$ and $g(t) = g_{ij}(t) \mathrm{d}x^i \otimes \mathrm{d}x^j$ in coordinates, we get

$$\partial_1 \langle X(t), Y(t) \rangle_{g(t)} = \partial_1 (X^i(t) Y^j(t) g_{ij}(t))$$

$$
\begin{aligned}
&= \partial_1 X^i(t)Y^j(t)g_{ij}(t) + X^i(t)\partial_1 Y^j(t)g_{ij}(t) + X^i(t)Y^j(t)\partial_1 g_{ij}(t) \\
&= \langle \partial_1 X(t), Y(t)\rangle_{g(t)} + \langle X(t), \partial_1 Y(t)\rangle_{g(t)} + (\partial_1 g(t))(X(t), Y(t)) \\
&= \langle \nabla_{\partial_1} X(t), Y(t)\rangle_{g(t)} + \langle X(t), \nabla_{\partial_1} Y(t)\rangle_{g(t)}.
\end{aligned}
$$

Here, the last line follows by splitting $(\partial_1 g(t))(X(t), Y(t))$ in two, and raising one index.

Finally, using that ∇ is C^∞-linear in the first variable proves the claim. □

As a corollary, we obtain the derivative of the inner product between two time-dependent vector fields along a curve in M.

Corollary 7.2.2. *Let $X(t), Y(t) \in \Gamma(\overline{TM})$ be time-dependent vector fields and let $\gamma : [0,1] \to M$ be a curve. Then*

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\langle X(t,\gamma(t)), Y(t,\gamma(t))\rangle_{g(t)} \\
&\qquad = \langle \nabla_{\partial_1+\dot\gamma(t)} X(t), Y(t)\rangle_{g(t)} + \langle X(t), \nabla_{\partial_1+\dot\gamma(t)} Y(t)\rangle_{g(t)}.
\end{aligned}
$$

Proof. Consider the curve $\varphi : [0,1] \to \mathbb{M}$ given by $\varphi(t) = (t, \gamma(t))$. From Proposition 7.2.1 it follows that

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\langle X(t,\gamma(t)), Y(t,\gamma(t))\rangle_{g(t)} \\
&\quad = \dot\varphi(t)\langle X(t,\gamma(t)), Y(t,\gamma(t))\rangle_{g(t)} \\
&\quad = \langle \nabla_{\dot\varphi(t)} X(t,\gamma(t)), Y(t,\gamma(t))\rangle_{g(t)} + \langle X(t,\gamma(t)), \nabla_{\dot\varphi(t)} Y(t,\gamma(t))\rangle_{g(t)} \\
&\quad = \langle \nabla_{\partial_1+\dot\gamma(t)} X(t), Y(t)\rangle_{g(t)} + \langle X(t), \nabla_{\partial_1+\dot\gamma(t)} Y(t)\rangle_{g(t)}.
\end{aligned}
$$

Here, the last line follows from the fact that $\dot\varphi(t) = \partial_1 + \dot\gamma(t)$. □

Remark 7.2.3. If $X(t) = X$ for some fixed vector field $X \in \Gamma(TM)$, then $\partial_1 X(t) = 0$, and we reduce to the setting in [23]. If we consider another stationary vector field $Y(t) = Y \in \Gamma(TM)$ and a curve $\gamma : [0,1] \to M$, we have

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\langle X(\gamma(t)), Y(\gamma(t))\rangle_{g(t)} &= (\partial_1 g(t))(X(\gamma(t)), Y(\gamma(t))) + \\
&\quad + \langle \nabla^t_{\dot\gamma(t)} X(\gamma(t)), Y(\gamma(t))\rangle_{g(t)} + \langle X(\gamma(t)), \nabla^t_{\dot\gamma(t)} Y(\gamma(t))\rangle_{g(t)}.
\end{aligned}
$$

Corollary 7.2.2 inspires us to define a notion of a time-dependent vector field being parallel along a curve in M with respect to a collection $\{g(t)\}_{t\in[0,1]}$ of Riemannian metrics. We have the following definition.

Definition 7.2.4. *Let $\gamma : [0,1] \to M$ be a curve. A time-dependent vector field $X(t)$ along γ is said to be* parallel along γ with respect to $\{g(t)\}_{t\in[0,1]}$ *if it is parallel along the curve $(t, \gamma(t))$ in $\mathbb{M}$ with respect to the connection ∇. More precisely, $X(t)$ is parallel along γ if and only if for all $t \in [0,1]$ we have*

$$
\nabla_{\partial_1+\dot\gamma(t)} X(t)(\gamma(t)) = 0.
$$

Remark 7.2.5. If $X(t)$ and $Y(t)$ are time-dependent vector fields which are parallel along γ with respect to $\{g(t)\}_{t\in[0,1]}$, then by Corollary 7.2.2 we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle X(t,\gamma(t)), Y(t,\gamma(t))\rangle_{g(t)} = 0.$$

This shows that the inner product between parallel vector fields is constant. In particular, by taking $Y(t) = X(t)$, we find that $|X(t,\gamma(t))|_{g(t)}$ is constant.

7.2.2. Horizontal lift with respect to a family of metrics

In Section 2.3, a thorough explanation is given of the frame bundle and the horizontal lift of curves and vectors to this frame bundle. It is also explained that these notions extend to general principal bundles. The aim of this section is to extend these notions to the time-dependent setting.

Instead of performing horizontal lift with respect to a fixed connection, we wish to define the horizontal lift with respect to a time-dependent family of connections. More precisely, we wish to define the horizontal lift with respect to the family of Levi-Civita connections associated to the collection $\mathcal{G} = \{g(t)\}_{t\in[0,1]}$ of Riemannian metrics on M. To do this, we use the parallel transport given in Definition 7.2.4.

Definition 7.2.6. *Let $\gamma : [0,1] \to M$ be a curve in M. A curve $u(t) \in FM$ is called a* horizontal lift of γ with respect to $\{g(t)\}_{t\in[0,1]}$ *if for all $a \in \mathbb{R}^d$ we have that $u(t)a$ is parallel along γ with respect to $\{g(t)\}_{t\in[0,1]}$, i.e., for all $a \in \mathbb{R}^d$ we have*

$$\nabla_{\partial_1+\dot\gamma(t)}(u(t)a) = 0$$

for all $t \in [0,1]$.

If $u(t)$ is the horizontal lift with respect to $\{g(t)\}_{t\in[0,1]}$ of a curve γ, then by Corollary 7.2.2 we have for all $a \in \mathbb{R}^d$ that

$$\frac{\mathrm{d}}{\mathrm{d}t}|u(t)a|_{g(t)} = 0, \tag{7.2.2}$$

i.e., $|u(t)a|_{g(t)} = |u(0)a|_{g(0)}$ for all $t \in [0,1]$. Therefore, if the initial frame $u(0) : \mathbb{R}^d \to (T_{\gamma(0)}M, g(0))$ is an isometry, then $u(t) : \mathbb{R}^d \to (T_{\gamma(t)}M, g(t))$ is an isometry for all $t \in [0,1]$.

We use this observation to show that the horizontal lift with respect to $\{g(t)\}_{t\in[0,1]}$ exists for all time, and is unique once an initial (orthonormal) frame is given. We do this by showing that the horizontal lift defined in 7.2.6 is a special instance of a horizontal lift from the manifold $\mathbb{M} = \mathbb{R} \times M$ to a principal fibre bundle over $\mathbb{M}$. To this end, consider the bundle $\mathcal{O}$ over $\mathbb{M}$ with fibres given by

$$\mathcal{O}_{(t,x)} = \{u : \mathbb{R}^d \to (T_xM, g(t)) | u \text{ isometry}\}, \tag{7.2.3}$$

i.e., $\mathcal{O}_{(t,x)}$ consists of the orthonormal frames for T_xM with respect to the metric $g(t)$. The bundle $\mathcal{O}$ is a principal bundle with structure group $G = O(d)$, the orthogonal group.

Now, let $\gamma(t)$ be a curve in M with horizontal lift $u(t)$ as in Definition 7.2.6, such that $u(0)$ is an orthonormal frame for $T_{\gamma(0)}M$ with respect to $g(0)$. From (7.2.2) it follows that for all $t \in [0,1]$, $u(t)$ is orthonormal with respect to $g(t)$ and hence $u(t) \in \mathcal{O}_{(t,\gamma(t))}$ for all $t \in [0,1]$. If we now define $\varphi(t) = (t, \gamma(t)) \in \mathbb{M}$, then $\dot{\varphi}(t) = \partial_1 + \dot{\gamma}(t)$. This implies that the curve $u(t)$ can also be interpreted as the horizontal lift of $\varphi(t)$ with respect to the connection ∇ as in (7.2.1) to the bundle $\mathcal{O}$. Because $\mathcal{O}$ is a principal bundle, it follows that a horizontal lift of $\varphi(t) = (t, \gamma(t))$ exists for all time $t \in [0,1]$ and is unique if the initial condition $u(0) = u_0 \in \mathcal{O}_{(0,\gamma(0))}$ is fixed. For this, we refer to Section 2.3. We conclude that the horizontal lift defined in Definition 7.2.6 always exists and is unique if an initial orthonormal frame with respect to $g(0)$ is given.

As explained in Section 2.3, if a horizontal lift for curves is defined, we can use this to define the horizontal lift of tangent vectors. In particular, the horizontal lift of curves in $\mathbb{M}$ to the bundle $\mathcal{O}$ with respect to the connection ∇ in (7.2.1) allows us to lift tangent vectors $X \in T_{(t,x)}\mathbb{M}$ to $T\mathcal{O}$. In what follows, we denote this lift by X^*.

Since we also have a notion of horizontal lifts of curves in M with respect to $\{g(t)\}_{t\in[0,1]}$, we can use this to define the horizontal lift of a tangent vector in TM with respect to $\{g(t)\}_{t\in[0,1]}$.

Definition 7.2.7. *Let $X \in T_pM$ and $u \in \mathcal{O}_{(s,p)}$. Let $\gamma : [0,1] \to M$ be a curve with $\gamma(s) = p$ and $\dot{\gamma}(s) = X$. Denote by $u(t)$ the horizontal lift of γ with respect to $\mathcal{G} = \{g(t)\}_{t\in[0,1]}$, satisfying $u(s) = u$. We define the* horizontal lift of X via u with respect to $\{g(t)\}_{t\in[0,1]}$ *by $X^{*\mathcal{G}}(u) = \dot{u}(s)$.*

Remark 7.2.8. If γ is a curve in M, we can identify its horizontal lift with respect to $\{g(t)\}_{t\in[0,1]}$ with the horizontal lift of the curve $\varphi(t) = (t, \gamma(t))$ in $\mathbb{M}$ with respect to the connection ∇ defined in (7.2.1). This implies that $\dot{u}(s)$ is the horizontal lift of $\dot{\varphi}(s) = \partial_1 + \dot{\gamma}(s)$ to $T_{u(s)}\mathcal{O}_{(s,\gamma(s))}$ via $u(s)$. Therefore, we have that $X^{*\mathcal{G}}(u) = (\partial_1 + X)^*(u)$.

Next, we relate the horizontal lift of X via $u \in \mathcal{O}_{(s,p)}$ with respect to $\{g(t)\}_{t\in[0,1]}$ to the horizontal lift of X via u with respect to the metric $g(s)$. Before we can do this, we first need the following result, the proof of which is inspired by the proof of [23, Proposition 1.2].

Proposition 7.2.9. *Let $u \in \mathcal{O}_{(s,p)}$. Then the horizontal lift of ∂_1 via u with respect to the connection ∇ in* (7.2.1) *is given by*

$$\partial_1^*(u) = -\frac{1}{2}(\partial_1 g(s))(ue_i, ue_j)V_{ij}(u).$$

Here, $\{e_1, \ldots, e_d\}$ is the canonical basis of $\mathbb{R}^d$ and $V_{ij}(u)$ are the canonical vertical basis vectors of V_uFM defined in (2.3.1).

Proof. Consider the curve $\eta(t) = (s+t, p)$. Then $\dot{\eta}(0) = \partial_1$. Let $u(t)$ be the horizontal lift of $\eta(t)$ with $u(0) = u$. Then $\partial_1^*(u) = \dot{u}(0)$. Since $\dot{\eta}(t) = \partial_1$, we have for all $a \in \mathbb{R}^d$ that

$$\nabla_{\partial_1}(u(t)a) = 0,$$

which gives by (7.2.1) that

$$\partial_1(u(t)a) + \frac{1}{2}(\partial_1 g(s+t))(u(t)a, \cdot)^{\#_{s+t}} = 0. \tag{7.2.4}$$

Since $u(t) \in F_pM$ for all t, we have that $\dot{u}(t) \in V_{u(t)}FM$. As a consequence, we can write

$$\dot{u}(t) = c_{\alpha\beta}(t, u(t))V_{\alpha\beta}(u(t)),$$

where $V_{\alpha\beta}$ are the canonical vertical basis vector fields defined in (2.3.1). Note that $u(t)a = \mathrm{ev}_a(u(t))$, where $\mathrm{ev}_a : FM \to TM$ denotes evaluation in a. From this it follows that $\partial_1(u(t)a) = \mathrm{d}(\mathrm{ev}_a)(u(t))(\dot{u}(t))$. Furthermore, note that

$$\begin{aligned}\mathrm{d}(\mathrm{ev}_a)(u(t))(V_{\alpha\beta}(u(t)) &= \mathrm{d}(\mathrm{ev}_a \circ L_{u(t)})(I)(E_{\alpha\beta}) \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} u(t)(I + sE_{\alpha\beta})a \\ &= u(t)(a_\beta e_\alpha),\end{aligned}$$

where we write $a = a_\beta e_\beta$.
By linearity, we find for every $i = 1, \dots, d$ that

$$\begin{aligned}\partial_1(u(t)e_i) &= c_{\alpha\beta}(t, u(t))\mathrm{d}(\mathrm{ev}_{e_i})(u(t))(V_{\alpha\beta}(u(t))) \\ &= c_{\alpha\beta}(t, u(t))u(t)\delta_{i\beta}e_\alpha \\ &= c_{\alpha i}(t, u(t))u(t)e_\alpha.\end{aligned}$$

Furthermore, since $\partial_1(u(t)e_i) = -\frac{1}{2}(\partial_1 g(s+t))(u(t)e_i, \cdot)^{\#_{s+t}}$ by (7.2.4), we have

$$\langle \partial_1(u(t)e_i), u(t)e_j \rangle_{g(s+t)} = -\frac{1}{2}(\partial_1 g(s+t))(u(t)e_i, u(t)e_j)$$

for every $j = 1, \dots, d$. Now, the left hand side is given by

$$\begin{aligned}\langle \partial_1(u(t)e_i), u(t)e_j \rangle_{g(s+t)} &= c_{\alpha i}(t, u(t))\langle u(t)e_\alpha, u(t)e_j \rangle_{g(s+t)} \\ &= c_{\alpha i}(t, u(t))\langle e_\alpha, e_j \rangle_{\mathbb{R}^d} \\ &= c_{ji}(t, u(t)).\end{aligned}$$

Here we used in the second line that $u(t)$ is an isometry from $\mathbb{R}^d$ to $(T_pM, g(s+t))$. Combining the two equalities above, we find for every $i, j = 1, \dots, d$ that

$$c_{ji}(t, u(t)) = -\frac{1}{2}(\partial_1 g(s+t))(u(t)e_i, u(t)e_j).$$

Because $\partial_1 g(s+t)$ is symmetric, it follows that $c_{ij} = c_{ji}$. Therefore, we can write

$$\dot{u}(t) = -\frac{1}{2}(\partial_1 g(s+t))(u(t)e_i, u(t)e_j)V_{ij}(u(t)),$$

so that

$$\partial_1^*(u) = \dot{u}(0) = -\frac{1}{2}(\partial_1 g(s))(ue_i, ue_j)V_{ij}(u),$$

where we used that $u(0) = u$. □

From Proposition 7.2.9 we deduce the relation between the horizontal lift of $X \in T_pM$ with respect to $\{g(t)\}_{t\in[0,1]}$ and with respect to the metric $g(s)$ at a specific time $s \in [0,1]$.

Corollary 7.2.10. *For $X \in T_pM$ and $u \in \mathcal{O}_{(s,p)}$ we have*

$$X^{*\mathcal{G}}(u) = X^{*_s}(u) - \frac{1}{2}(\partial_1 g(s))(ue_i, ue_j)V_{ij}(u),$$

where $X^{*_s}(u)$ denotes the horizontal lift of X via u with respect to the metric $g(s)$, and the e_i and V_{ij} are as in Proposition 7.2.9.

Proof. From Remark 7.2.8 it follows that $X^{*\mathcal{G}}(u) = (\partial_1 + X)^*(u)$. Since it holds that $(\partial_1 + X)^*(u) = \partial_1^*(u) + X^*(u)$ (see e.g. [87]), it follows from Proposition 7.2.9 that we are done once we show that $X^*(u) = X^{*_s}(u)$. To see the latter, consider a curve $\gamma : (-\varepsilon, \varepsilon) \to M$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = X$ and define $\varphi : (-\varepsilon, \varepsilon) \to \mathbb{M}$ by $\varphi(t) = (s, \gamma(t))$. Then $\varphi(0) = (s, p)$ and $\dot{\varphi}(0) = X$. Let $u(t)$ be the horizontal lift of φ with $u(0) = u$. Since $\dot{\varphi}(t) = X$, we have

$$\nabla_X^s(u(t)a) = 0$$

for every $a \in \mathbb{R}^d$. This implies that $u(t)$ is the horizontal lift of $\gamma(t)$ with respect to ∇^s, i.e., the Levi-Civita connection of $g(s)$. It follows that $X^*(u) = X^{*_s}(u)$ as desired. □

7.2.3. Development and anti-development of curves

In Section 2.3 we explained how we can use the notion of a horizontal lift to associate to a curve in M a curve in $\mathbb{R}^d$ and vice versa. In the time-inhomogeneous case we take the same approach, but now using the horizontal lift with respect to $\{g(t)\}_{t\in[0,1]}$.

Definition 7.2.11. *Let $\gamma : [0,1] \to M$ be a curve in M and let $u(t)$ be a horizontal lift of γ with respect to $\{g(t)\}_{t\in[0,1]}$. We define the* anti-development of γ with respect to $\{g(t)\}_{t\in[0,1]}$ *as the curve $w : [0,1] \to \mathbb{R}^d$ given by*

$$w(t) = \int_0^t u(s)^{-1}\dot{\gamma}(s)\,\mathrm{d}s. \tag{7.2.5}$$

If we fix a frame $u \in \mathcal{O}_{(0,\gamma(0))}$ (see (7.2.3)), we can speak about the anti-development of γ via u with respect to $\{g(t)\}_{t\in[0,1]}$ since the horizontal lift with respect to $\{g(t)\}_{t\in[0,1]}$ satisfying $u(0) = u$ is unique.

If $w(t)$ is the anti-development of $\gamma(t)$ with respect to $\{g(t)\}_{t\in[0,1]}$ via the horizontal lift $u(t)$, then (7.2.5) implies that

$$\dot{w}(t) = u(t)^{-1}\dot{\gamma}(t),$$

which rewrites to

$$\dot{\gamma}(t) = u(t)\dot{w}(t).$$

Since both sides are elements of $T_{\gamma(t)}M$, we can consider their horizontal lifts with respect to the metric $g(t)$, which must be equal:

$$H(t, u(t))\dot{w}(t) := (u(t)\dot{w}(t))^{*_t} = (\dot{\gamma}(t))^{*_t}. \tag{7.2.6}$$

Here $H(t, u(t))$ is as defined in (2.3.4) with respect to the Levi-Civita connection ∇^t of the metric $g(t)$. Furthermore, since $u(t)$ is the horizontal lift of γ with respect to $\{g(t)\}_{t\in[0,1]}$, we have that $\dot{u}(t) = \dot{\gamma}(t)^{*\mathcal{G}}$. Therefore, by applying Corollary 7.2.10 and using (7.2.6) we obtain

$$\begin{aligned}\dot{u}(t) &= \dot{\gamma}(t)^{*\mathcal{G}}\\ &= (\dot{\gamma}(t))^{*_t} - \frac{1}{2}(\partial_1 g(t))(u(t)e_i, u(t)e_j)V^{ij}(u(t))\\ &= H(t, u(t))\dot{w}(t) - \frac{1}{2}(\partial_1 g(t))(u(t)e_i, u(t)e_j)V^{ij}(u(t)).\end{aligned}$$

We thus obtained a differential equation for the horizontal lift u with respect to $\{g(t)\}_{t\in[0,1]}$ in terms of the anti-development w. This shows how to invert the operation of taking the anti-development of a curve. We make the following definition.

Definition 7.2.12. *Let $w : [0,1] \to \mathbb{R}^d$ be a curve in $\mathbb{R}^d$ and fix $u_0 \in \mathcal{O}_{(0,p)}$. Let $u : [0,1] \to FM$ be the solution of*

$$\dot{u}(t) = H(t, u(t))\dot{w}(t) - \frac{1}{2}(\partial_1 g(t))(u(t)e_i, u(t)e_j)V^{ij}(u(t)) \tag{7.2.7}$$

with $u(0) = u_0$, where $H(t, u(t))$ is as defined in (2.3.4) *for the Levi-Civita connection ∇^t of the metric $g(t)$. Then the curve $\gamma(t) = \pi u(t)$ is called the* development of w onto M with respect to $\{g(t)\}_{t\in[0,1]}$.

Sometimes, the curve u is referred to as the development of w, rather than the projection of u onto M.

7.2.4. Horizontal lift of $g(t)$-Brownian motion

In this section we explain how a $g(t)$-Brownian motion may be obtained by solving a stochastic differential equation on FM, and projecting the solution down to the manifold. The approach is similar to time-homogeneous case, see Section 2.4.4.

Malliavin's transfer principle (see e.g. [72]) suggests that constructions for manifold-valued curves can be extended to manifold-valued processes by replacing differential equations by Stratonovich stochastic differential equations. This is because

Stratonovich integrals satisfy the ordinary fundamental theorem of calculus. This suggests that we can obtain a $g(t)$-Brownian motion as the development with respect to $\{g(t)\}_{t\in[0,1]}$ of a standard Brownian motion in $\mathbb{R}^d$.

More precisely, we replace the curve w in (7.2.7) by a standard $\mathbb{R}^d$-valued Brownian motion W_t and interpret the so obtained stochastic differential equation in Stratonovich sense. In symbols this means that for $x_0 \in M$ fixed, we consider the solution U_t of the Stratonovich stochastic differential equation

$$\mathrm{d}U_t = H_i(t, U_t) \circ \mathrm{d}W_t^i - \frac{1}{2}(\partial_1 g(t))_{ij}(U_t e_i, U_t e_j)V^{ij}(U_t)\,\mathrm{d}t, \tag{7.2.8}$$

with $U_0 \in \mathcal{O}_{(0,x_0)}$ (see (7.2.3)). Here, $H_i(t,u)$ are the canonical horizontal vector fields with respect to the Levi-Civita connection ∇^t of the metric $g(t)$ as defined in (2.3.5). Furthermore, $\{e_1, \ldots, e_d\}$ denotes the standard basis of $\mathbb{R}^d$ and V^{ij} is the standard basis of vertical vectors, see (2.3.1). The following is [23, Proposition 1.4], see also [5, Proposition 1.3].

Proposition 7.2.13. *Let U_t be the process on FM solving equation* (7.2.8). *Then $X_t = \pi U_t$ is a $g(t)$-Brownian motion on M starting in $x_0 \in M$.*

7.3. Proof of Theorem 7.1.2 using embeddings

In this section we prove Theorem 7.1.2, the analogue of Schilder's theorem for $g(t)$-Brownian motion. Let us recall the statement of the theorem.

Theorem 7.3.1. *Let M be a manifold and let $\{g(t)\}_{t\in[0,1]}$ be a collection of Riemannian metrics on M, smoothly depending on t. Fix $x_0 \in M$, and let X_t be a $g(t)$-Brownian motion with $X_0 = x_0$. Furthermore, for every $n \geqslant 1$, let X_t^n be the continuous process generated by $\frac{1}{2n}\Delta_M^t$ with $X_0^n = x_0$. Assume the processes X_t and X_t^n exist for all time $t \in [0,1]$. Then $\{X^n\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];M)$ with good rate function I_M given by*

$$I_M(\gamma) = \begin{cases} \frac{1}{2}\int_0^1 |\dot\gamma(t)|^2_{g(t)}\,\mathrm{d}t, & \gamma \in H^{1,\mathcal{G}}_{x_0}([0,1];M), \\ \infty & \textit{otherwise.} \end{cases}$$

As we have seen in Proposition 7.2.13, the horizontal lift U_t with respect to $\{g(t)\}_{t\in[0,1]}$ of a $g(t)$-Brownian motion satisfies the Stratonovich stochastic differential equation

$$\mathrm{d}U_t = H_i(t, U_t) \circ \mathrm{d}W_t^i - \frac{1}{2}(\partial_1 g(t))_{ij}(U_t e_i, U_t e_j)V^{ij}(U_t)\,\mathrm{d}t,$$

with $U_0 = u_0 \in \mathcal{O}_{(0,x_0)}$, where $\mathcal{O}_{(0,x_0)}$ is defined in (7.2.3). Similarly, if $\tilde X_t^n$ is a $g(nt)$-Brownian motion, then its horizontal lift $\tilde U_t^n$ with respect to $\{g(nt)\}_{t\in[0,n^{-1}]}$ satisfies

$$\mathrm{d}\tilde U_t^n = H_i(nt, \tilde U_t^n) \circ \mathrm{d}W_t^i - \frac{n}{2}(\partial_1 g(nt))_{ij}(\tilde U_t^n e_i, \tilde U_t^n e_j)V^{ij}(\tilde U_t^n)\,\mathrm{d}t,$$

with $\tilde{U}_0^n = u_0 \in \mathcal{O}(0, x_0)$. Finally, the horizontal lift of $X_t^n = \tilde{X}_{tn^{-1}}^n$ with respect to $\{g(nt)\}_{t\in[0,n-1]}$ is given by $U_t^n = \tilde{U}_{tn^{-1}}^n$. This process satisfies

$$\mathrm{d}U_t^n = H_i(t, U_t^n) \circ \mathrm{d}W_t^{n,i} - \frac{1}{2}(\partial_t g(t))_{ij}(U_t^n e_i, U_t^n e_j)V^{ij}(U_t^n)\,\mathrm{d}t, \tag{7.3.1}$$

with $U_0 = u_0 \in \mathcal{O}_{(0,x_0)}$. Here, $W_t^n = W_{tn^{-1}} = \frac{1}{\sqrt{n}}W_t$. As explained above Theorem 7.1.2, X_t^n is the process generated by $\frac{1}{2n}\Delta_M^t$ that we are studying.
The stochastic differential equation for the horizontal lift of X_t^n obtained in (7.3.1) is an important tool for proving Theorem 7.1.2. However, before we can get to this, we first need to make some preparations.

7.3.1. Compact containment

As part of the proof of Theorem 7.1.2, we need to show that the process X_t^n generated by $\frac{1}{2n}\Delta_M^t$ stays within a compact set with high enough probability when n tends to infinity. In this section we adapt the general approach using Lyapunov functions discussed in Section 5.2.2 to the time-inhomogeneous case.

Definition 7.3.2. *Let $\mathcal{H}_t : T^*M \to \mathbb{R}$ be a collection of maps. We call a function $\Upsilon : M \to \mathbb{R}$ a* good containment function *for the collection $\{\mathcal{H}_t\}_t$ if Υ is a good containment function for $\mathcal{H}_t$ for every t in the sense of Definition 5.2.4 and if additionally we have*

$$\sup_{t,x} \mathcal{H}_t(x, \mathrm{d}\Upsilon(x)) < \infty.$$

In what follows, we use the notion of operator convergence defined in Defintion 5.2.8. The following assumption is a (simplified) version of Assumption 5.2.9.

Assumption 7.3.3. *For every $n \geqslant 1$, let $A_n^t \subset C_b(M) \times C_b(M)$ be the (time-inhomogeneous) generator of a Markov process X_n. Assume that for every $x \in M$, the process X_n started in x is right-continuous and exists for all t. Define the operator*

$$H_n^t f = \frac{1}{n} e^{-nf} A_n^t e^{nf}, \qquad e^{nf} \in \mathcal{D}(A_n^t).$$

Suppose that for every t, there is an operator $H^t : \mathcal{D}(H^t) \subset C_b(M) \to C_b(M)$ with $\mathcal{D}(H^t) = C_c^\infty(M)$ and such that $H^t \subset ex - \lim_{n\to\infty} H_n^t$. Finally, assume that H^t can be written as $H^t f(x) = \mathcal{H}^t(x, \mathrm{d}f(x))$ for some map $\mathcal{H}^t : T^*M \to \mathbb{R}$.

The following is an extension of Proposition 5.2.10 to the time-homogeneous case. The proof is a straightforward adaptation of the proof of Proposition A.15 in [22], which is based on Lemma 4.22 in [39].

Proposition 7.3.4. *Let Assumption 7.3.3 be satisfied and assume that $X_n(0) = x \in M$ for all $n \geqslant 1$. Assume that Υ is a good containment function for the collection $\mathcal{H}_t$. Assume furthermore that for every $f \in C_c^\infty(M)$ and every $n \geqslant 1$ the map $t \mapsto H_n^t f$ is continuous. Then for every $\alpha > 0$, there exists a compact set $K_\alpha \subset M$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(X_n(t) \notin K_\alpha \text{ for some } t \in [0,1]\right) \leqslant -\alpha.$$

Moreover, the sequence K_α can be chosen to be increasing with $\bigcup_\alpha K_\alpha = M$.

Remark 7.3.5. The continuous dependence of $H_n^t f$ on t in Proposition 7.3.4 is used to assure that $\int_0^s H_n^t f(X_n(t))\,\mathrm{d}t$ exists. This is necessary to construct a local exponential martingale used in the proof.

7.3.2. Freidlin-Wentzell theory for time-inhomogeneous diffusions

For the proof of Theorem 7.1.2, we embed the frame bundle FM into some Euclidean space $\mathbb{R}^N$. Using this embedding, we push forward the stochastic differential equation in (7.3.1) to obtain a stochastic differential equation in $\mathbb{R}^N$. To obtain the large deviations for such diffusions, we use Freidlin-Wentzell theory ([41]). Since the stochastic differential equations has time-inhomogeneous coefficients, we have to adapt the Freidlin-Wentzell theory to this setting. One can follow the line of proof for Freidlin-Wentzell theory for time-homogeneous diffusions, i.e., by using Euler approximations and making the drift and variance constant on small intervals of time, see e.g. [29, Theorem 5.6.7].

Theorem 7.3.6. *Let W_t be a standard Brownian motion with values in $\mathbb{R}^d$. Let $b : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,1] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be bounded and Lipschitz. For every $n \geqslant 1$, let X_t^n be the process satisfying*

$$\mathrm{d}X_t^n = b(t, X_t^n)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\sigma(t, X_t^n)\,\mathrm{d}W_t,$$

with $X_0^n = x_0$. Then $\{X^n\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];\mathbb{R}^d)$ with good rate function

$$I(\gamma) = \inf\left\{\frac{1}{2}\int_0^1 |\dot{\varphi}(t)|^2\,\mathrm{d}t \,\middle|\, \varphi \in H^1([0,1];\mathbb{R}^d), \right.$$
$$\left. \gamma(t) = y + \int_0^t b(s,\gamma(s))\,\mathrm{d}s + \int_0^t \sigma(s,\gamma(s))\dot{\varphi}(s))\,\mathrm{d}s\right\}. \quad (7.3.2)$$

The same result also holds when we consider Stratonovich stochastic differential equations instead of the Itô ones. Following the same reasoning as in the proof of [63, Theorem 2.5], we have the following corollary.

Corollary 7.3.7. *Let the assumptions of Theorem 7.3.6 be satisfied. Assume furthermore that σ has bounded derivative. For every $n \geqslant 1$, let X_t^n be the process satisfying the Stratonovich stochastic differential equation*

$$\mathrm{d}X_t^n = b(t, X_t^n)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\sigma(t, X_t^n) \circ \mathrm{d}W_t,$$

with $X_0^\varepsilon = x_0$. Then $\{X^n\}_{n\geqslant 1}$ satisfies in $C([0,1];\mathbb{R}^d)$ the large deviation principle with good rate function I given in (7.3.2)*.*

7.3.3. Proof of Theorem 7.1.2

Before we prove Theorem 7.1.2, we first need some preliminary results. In the following proposition we prove that given a collection of metrics $\{g(t)\}_{t\in[0,1]}$ depending continuously on t, we can find another metric that dominates all of these metrics.

Proposition 7.3.8. *Let $\{g(t)\}_{t\in[0,1]}$ be a collection of Riemannian metrics on M, depending smoothly on t. There exists a Riemannian metric $\overline{g}$ such that for all $x \in M$ and all $v \in T_xM$ we have*

$$g_t(v,v) \leqslant \overline{g}(v,v)$$

for all $t \in [0,1]$.

Proof. Let $\{U_n\}_{n\in\mathbb{N}}$ be a countable collection of relatively compact charts covering M. Furthermore, let $\{\varphi_n\}_{n\in\mathbb{N}}$ be a partition of unity for the collection $\{U_n\}_{n\in\mathbb{N}}$. Writing $G_t(x)$ for the matrix of coordinates of the metric $g(t)$ in a chart U_n, we have

$$\begin{aligned} g(t)(v,v) &= \left\langle G_t^{\frac{1}{2}}(x)v, G_t^{\frac{1}{2}}(x)v \right\rangle_2 \\ &= \left\langle G_t^{\frac{1}{2}}(x)G_0^{-\frac{1}{2}}(x)G_0^{\frac{1}{2}}(x)v, G_t^{\frac{1}{2}}(x)G_0^{-\frac{1}{2}}(x)G_0^{\frac{1}{2}}(x)v \right\rangle_2 \end{aligned}$$

for all $v \in T_xM$. Here, the Euclidean inner product has to be understood as the Euclidean inner product of the vector of coefficients of v. Using the Cauchy-Schwarz inequality, we find

$$g(t)(v,v) \leqslant \left|\left|G_t^{\frac{1}{2}}(x)G_0^{-\frac{1}{2}}(x)\right|\right|_2^2 \left|\left|G_0^{\frac{1}{2}}(x)v\right|\right|_2^2 = \left|\left|G_t^{\frac{1}{2}}(x)G_0^{-\frac{1}{2}}(x)\right|\right|_2^2 g(0)(v,v). \quad (7.3.3)$$

Note that $G_t(x)$ depends continuously on t and x, and hence so does $G_t^{\frac{1}{2}}(x)$. Similarly, $G_0^{-\frac{1}{2}}(x)$ depends continuously on x. Since $[0,1]$ is compact and U_n is relatively compact, the continuity implies that $||G_t^{\frac{1}{2}}(x)G_0^{-\frac{1}{2}}(x)||_2$ is bounded on $[0,1]\times U_n$. If we write

$$C = \sup_{t\in[0,1],x\in U_n} \left|\left|G_t^{\frac{1}{2}}(x)G_0^{-\frac{1}{2}}(x)\right|\right|_2 < \infty,$$

then we can define the Riemannian metric $\overline{g}_n$ on U_n by

$$\overline{g}_n = Cg(0).$$

From (7.3.3) it follows that

$$g_t(v,v) \leqslant \overline{g}_n(v,v)$$

for all $v \in T_xM$ and all $x \in U_n$.
We now define on M the metric

$$\overline{g} = \sum_{n=1}^{\infty} \varphi_n \overline{g}_n,$$

which has the desired property by construction. ∎

Let us denote by $\overline{d}$ the Riemannian distance function associated to the metric $\overline{g}$ from Proposition 7.3.8. Fix $x_0 \in M$ and consider the radial function $\overline{r}(x) = \overline{d}(x, x_0)$. Since $\overline{r}$ is not everywhere smooth, it is not suitable for constructing a good containment function as in Definition 7.3.2. However, since $\overline{r}$ is 1-Lipschitz (with respect to the metric $\overline{g}$), we can find a smooth function $\tilde{r}$ with $\tilde{r}(x_0) = \overline{r}(x_0) = 0$ and such that $||\tilde{r} - \overline{r}|| \leqslant 1$ and $|\mathrm{d}\tilde{r}|_{\overline{g}} \leqslant 2$, see also Proposition 5.3.1. Using this, we define Υ by

$$\Upsilon(x) = \log(1 + \tilde{r}(x)^2). \tag{7.3.4}$$

We now show that Υ can be used as a good containment function for the operators arising from the generator of a $g(t)$-Brownian motion.

Proposition 7.3.9. *Assume M is complete and let $\{g(t)\}_{t \in [0,1]}$ be a collection of Riemannian metrics on M, smoothly depending on t. For every $t \in [0, 1]$, define $\mathcal{H}_t : T^*M \to \mathbb{R}$ by $\mathcal{H}_t(x, p) = \frac{1}{2}|p|^2_{g(t)(x)}$. Let $\overline{g}$ be a metric as in Proposition 7.3.8 and define Υ as in* (7.3.4)*. Then Υ is a good containment function for the collection $\{\mathcal{H}_t\}_{t \in [0,1]}$.*

Proof. Following the proof of Theorem 5.1.3 in Section 5.3.3, it follows from Lemma 5.3.2 that Υ is a good containment function for each $\mathcal{H}_t$ individually. Hence, we are done once we show that

$$\sup_{t,x} \mathcal{H}_t(x, \mathrm{d}\Upsilon(x)) < \infty.$$

Observe that

$$\mathrm{d}\Upsilon(x) = \frac{2\tilde{r}(x)}{1 + \tilde{r}(x)^2} \mathrm{d}\tilde{r}(x),$$

so that

$$|\mathrm{d}\Upsilon(x)|_{\overline{g}(x)} \leqslant 2|\mathrm{d}\tilde{r}(x)|_{\overline{g}(x)} \leqslant 4.$$

From this, it follows that

$$\mathcal{H}_t(x, \mathrm{d}\Upsilon(x)) = \frac{1}{2}|\mathrm{d}\Upsilon(x)|^2_{g(t)(x)} \leqslant \frac{1}{2}|\mathrm{d}\Upsilon(x)|^2_{\overline{g}(x)} \leqslant 8.$$

for all t and x. Hence, we find that $\sup_{t,x} \mathcal{H}_t(x, \mathrm{d}\Upsilon(x)) < \infty$. □

We can now show that X^n_t remains in compact sets with high enough probability.

Proposition 7.3.10. *Let M be a complete manifold and let $\{g(t)\}_{t \in [0,1]}$ be a collection of Riemannian metrics on M, smoothly depending on t. Assume that for every $n > 1$, the continuous process X^n_t generated by $\frac{1}{2n}\Delta^t_M$ exists for all $t \in [0, 1]$. Then for every $\alpha \geqslant 0$, there exists a compact set $K_\alpha \subset M$ such that*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(X^n_t \notin K_\alpha \textit{ for some } t \in [0, 1]\right) \leqslant -\alpha.$$

Moreover, the sets K_α can be chosen to be increasing with $\bigcup_\alpha K_\alpha = M$.

Proof. We verify the conditions of Proposition 7.3.4. Let $f \in C_c^\infty(M)$ and define

$$H_n^t f = \frac{1}{n} e^{-nf} \frac{1}{2n} \Delta_M^t e^{nf}.$$

Then

$$H_n^t f = \frac{1}{n} e^{-nf} e^{nf} \frac{1}{2}(\Delta_M^t f + n|\mathrm{d}f|_{g(t)}^2) = \frac{1}{2n} \Delta_M^t f + \frac{1}{2}|\mathrm{d}f|_{g(t)}^2.$$

Now define $H^t \subset C_b(M) \times C_b(M)$ with domain $\mathcal{D}(H^t) = C_c^\infty(M)$ and $H^t f = \frac{1}{2}|\mathrm{d}f|_{g(t)}^2$. Then for all $f \in C_c^\infty(M)$ we have

$$\lim_{n\to\infty} ||H_n^t f - H^t f|| = 0,$$

so that $H \subset ex - \lim_{n\to\infty} H_n$. Furthermore, note that $H^t f(x) = \mathcal{H}^t(x, \mathrm{d}f(x))$ for $\mathcal{H}^t(x,p) = \frac{1}{2}|p|_{g(t)(x)}^2$. It follows that Assumption 7.3.3 is fulfilled, and by Proposition 7.3.9, the function Υ given in (7.3.4) is a good containment function for the collection $\{\mathcal{H}_t\}_{t\in[0,1]}$. Since $g(t)$ depends continuously on t, we find that $t \mapsto H_n^t f$ is continuous for every $f \in C_c^\infty(M)$, so that the claim follows from Proposition 7.3.4. □

Finally, we also need the following technical lemma.

Lemma 7.3.11. *Let M be a manifold, and let $\{g(t)\}_{t\in[0,1]}$ be a collection of metrics on M, smoothly depending on t. For every $t \in [0,1]$ and $x \in M$, define*

$$\mathcal{O}_{(t,x)} = \{u : \mathbb{R}^d \to (T_x M, g(t)) | u \text{ isometry}\}.$$

Let $K \subset M$ be compact. Then the set

$$\bigcup \{\mathcal{O}_{(t,x)} \big| t \in [0,1], x \in K\}$$

is a compact subset of FM.

Proof. Consider the bundle $\mathcal{O}$ over $\mathbb{R} \times M$ with fibres $\mathcal{O}_{(t,x)}$. For every $(t,x) \in [0,1] \times K$, let $U_{(t,x)} \subset [0,1] \times M$ be open and relatively compact such that there exists a smooth section $u_{(t,x)}$ of $\mathcal{O}$ on $\overline{U}_{(t,x)}$. Since $[0,1] \times K$ is compact, we can find finitely many $(t_1,x_1),\ldots,(t_k,x_k)$ such that

$$[0,1] \times K \subset \bigcup_{i=1}^k U_{(t_i,x_i)} \subset \bigcup_{i=1}^k \overline{U}_{(t_i,x_i)}.$$

As a consequence, we have

$$\bigcup \{\mathcal{O}_{(t,x)} \big| t \in [0,1], x \in K\} \subset \bigcup_{i=1}^k \bigcup \{\mathcal{O}_{(t,x)} \big| t \in [0,1], x \in \overline{U}_{(t_i,x_i)}\}$$

Since

$$\bigcup \{\mathcal{O}_{(t,x)} \big| t \in [0,1], x \in K\}$$

is closed, it suffices to show that

$$\bigcup \left\{ \mathcal{O}_{(t,x)} \middle| t \in [0,1], x \in \overline{U}_{(t_i,x_i)} \right\}$$

is compact for all $i = 1, \ldots, k$.
For this, consider the map $\Phi_i : [0,1] \times \overline{U}_{(t_i,x_i)} \times O(d) \to FM$ given by

$$\Phi_i(t,x,g) = u_{(t_i,x_i)}(t,x)g.$$

Then Φ_i is continuous as composition of continuous maps. Furthermore, we have that

$$\Phi_i([0,1] \times \overline{U}_{(t_i,x_i)} \times O(d)) = \bigcup \left\{ \mathcal{O}_{t,x} \middle| t \in [0,1], x \in \overline{U}_{(t_i,x_i)} \right\}.$$

Since $[0,1] \times \overline{U}_{(t_i,x_i)} \times O(d)$ is compact, the above, together with the continuity of Φ_i proves the claim. ☐

With all the preparations done, we are ready to prove Theorem 7.1.2. The proof is similar to the one given in Section 5.4 for the time-homogeneous case. In order to improve readability of certain equations in the upcoming proof, we define

$$(\partial_1 g(t))_{ij}(u) := \partial_1 g(t)(ue_i, ue_j) \tag{7.3.5}$$

for $i, j = 1, \ldots d$ and $u \in FM$, where $\{e_1, \ldots, e_d\}$ denotes the standard basis of $\mathbb{R}^d$.

Proof of Theorem 7.1.2. Let W_t be a standard, $\mathbb{R}^d$-valued Brownian motion and define $W_t^n = \frac{1}{\sqrt{n}}$ for every $n \geqslant 1$. Consider the process U_t^n in FM with $U_0^n = u_0 \in \mathcal{O}_{(0,x_0)}$ and satisfying

$$\mathrm{d}U_t^n = H_i(t, U_t^n) \circ \mathrm{d}W_t^{n,i} - \frac{1}{2}(\partial_1 g(t))_{ij}(U_t^n) V^{ij}(U_t^n)\,\mathrm{d}t,$$

where we used the notation introduced in (7.3.5).
Now, let $\{K_\alpha\}_{\alpha>0}$ be an increasing sequence of compact sets with $\bigcup_\alpha K_\alpha = M$ as in Proposition 7.3.10. By Lemma 7.3.11 we have that

$$\tilde{K}_\alpha := \bigcup \left\{ \mathcal{O}_{(t,x)} \middle| x \in K_\alpha, t \in [0,1] \right\} \subset FM$$

is compact.
Let $\varphi_\alpha : FM \to \mathbb{R}$ be a smooth function with compact support and $\varphi \equiv 1$ on $\tilde{K}_\alpha$. Since FM is locally compact, such a function exists. Consider the process $U_t^{n,\alpha}$ in FM given by

$$\mathrm{d}U_t^{n,\alpha} = \varphi_\alpha(U_t^{n,\alpha}) H_i(t, U_t^{n,\alpha}) \circ \mathrm{d}W_t^{n,i} - \frac{1}{2}\varphi_\alpha(U_t^{n,\alpha})(\partial_1 g(t))_{ij}(U_t^{n,\alpha}) V^{ij}(U_t^{n,\alpha})\,\mathrm{d}t,$$

with $U_0^{n,\alpha} = u_0$.
By Whitney's embedding theorem, there exists an $N \in \mathbb{N}$ and a smooth embedding $\iota : FM \to \mathbb{R}^N$. It follows from Proposition 2.4.9 that the $\mathbb{R}^N$-valued process $\tilde{U}_t^{\varepsilon,\alpha} := \iota(U_t^{\varepsilon,\alpha})$ satisfies

$$\mathrm{d}(\iota(U_t^{n,\alpha}))$$

$$= \varphi_\alpha(U_t^{n,\alpha}) H_i(t,\cdot)\iota(U_t^{n,\alpha}) \circ \mathrm{d}W_t^{n,i} - \frac{1}{2}\varphi_\alpha(U_t^{n,\alpha}) \left[(\partial_1 g(t))_{ij} V^{ij}\right] \iota(U_t^{n,\alpha})\,\mathrm{d}t$$
$$= \varphi_\alpha(U_t^{n,\alpha}) \iota^* H_i(t,\iota(U_t^{n,\alpha})) \circ \mathrm{d}W_t^{n,i} - \frac{1}{2}\varphi_\alpha(U_t^{n,\alpha})\iota^* \left[(\partial_1 g(t))_{ij} V^{ij}\right] (\iota(U_t^{n,\alpha}))\,\mathrm{d}t.$$

In terms of $\tilde{U}_t^{n,\alpha}$, this can be written as

$$\mathrm{d}\tilde{U}_t^{n,\alpha} = \varphi_\alpha(\iota^{-1}(\tilde{U}_t^{n,\alpha}))\iota^* H_i(t,\tilde{U}_t^{n,\alpha}) \circ \mathrm{d}W_t^{n,i} - \frac{1}{2}\varphi_\alpha(\iota^{-1}(U_t^{n,\alpha}))\iota^* \left[(\partial_1 g(t))_{ij} V^{ij}\right] (\tilde{U}_t^{n,\alpha}))\,\mathrm{d}t. \quad (7.3.6)$$

Since ι and ι^{-1} are smooth, the vector fields

$$\varphi_\alpha(\iota^{-1}(\tilde{U}_t^{n,\alpha}))\iota^* H_i(t,\tilde{U}_t^{n,\alpha})$$

and

$$\frac{1}{2}\varphi_\alpha(\iota^{-1}(\tilde{U}_t^{n,\alpha}))\iota^* \left[(\partial_1 g(t))_{ij} V^{ij}\right] (\tilde{U}_t^{n,\alpha})$$

are smooth and compactly supported inside $\iota(FM)$. By putting them equal to zero outside $\iota(FM)$, we obtain smooth, compactly supported vector fields on $\mathbb{R}^N$. With slight abuse of notation, we denote these vector fields by the same symbol. This observation allows us to consider (7.3.6) as equation on $\mathbb{R}^N$. Since the drift and diffusion are smooth and compactly supported, we can apply Corollary 7.3.7 to obtain that $\{\tilde{U}^{n,\alpha}\}_{n\geqslant 1}$ satisfies in $C([0,1];\mathbb{R}^N)$ the large deviation principle with good rate function $\tilde{I}^\alpha_{\mathbb{R}^N}$ given by

$$\tilde{I}^\alpha_{\mathbb{R}^N}(\gamma) = \inf\left\{ \int_0^1 |\dot{\phi}(t)|^2_{\mathbb{R}^d}\,\mathrm{d}t \middle| \gamma(0) = \iota(u_0),\ \dot{\gamma}(t) = \varphi_\alpha(\iota^{-1}(\gamma(t)))\iota^* H_i(t,\gamma(t))\dot{\phi}^i(t) - \frac{1}{2}\varphi_\alpha(\iota^{-1}(\gamma(t)))\iota^* \left[(\partial_1 g(t))_{ij} V^{ij}\right] (\gamma(t)) \right\}.$$

Now note that $\iota(FM)$ is closed, and by construction it holds that $\tilde{U}^{n,\alpha}$ is almost surely contained in $C([0,1];\iota(FM))$. Furthermore, suppose that $\gamma(0) \in \iota(FM)$ and that there exists a curve $\phi \in H^1([0,1];\mathbb{R}^d)$ such that

$$\dot{\gamma}(t) = \varphi_\alpha(\iota^{-1}(\gamma(t)))\iota^* H_i(t,\gamma(t))\dot{\phi}^i(t) - \frac{1}{2}\varphi_\alpha(\iota^{-1}(\gamma(t)))\iota^* \left[(\partial_1 g(t))_{ij} V^{ij}\right] (\gamma(t)).$$

Then, since the vector fields

$$(\varphi_\alpha \circ \iota^{-1})\iota^* H_i(t,\cdot)$$

and

$$\frac{1}{2}(\varphi_\alpha \circ \iota^{-1})\iota^* \left[(\partial_1 g(t))_{ij} V^{ij}\right]$$

are tangent to $\iota(FM)$ at points of $\iota(FM)$, we find that $\gamma(t) \in \iota(FM)$ for all $t \in [0,1]$ so that $\gamma \in C([0,1];\iota(FM))$. Therefore, if $\gamma \notin C([0,1];\iota(FM))$, then no such

ϕ exists, and $\tilde{I}^{\alpha}_{\mathbb{R}^N}(\gamma) = \infty$. It now follows from [29, Lemma 4.1.5] that $\{\tilde{U}^{n,\alpha}\}_{n\geqslant 1}$ satisfies the large deviation principle in $\iota(FM)$ with good rate function $\tilde{I}^{\alpha}_{\iota(FM)}$ given as the restriction of $\tilde{I}^{\alpha}_{\mathbb{R}^N}$ to $C([0,1];\iota(FM))$.

Since ι is a homeomorphism and $U^{n,\alpha}_t = \iota^{-1}(\tilde{U}^{n,\alpha}_t)$, the contraction principle (Theorem 2.1.6) implies that $\{U^{n,\alpha}\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];FM)$ with good rate function I^{α}_{FM} given by

$$\begin{aligned}
I^{\alpha}_{FM}(\eta) &= \tilde{I}^{\alpha}_{\iota(FM)}(\iota\circ\eta)\\
&= \inf\left\{\int_0^1 |\dot{\phi}(t)|^2_{\mathbb{R}^d}\,\mathrm{d}t \,\middle|\, \iota(\eta(0)) = \iota(u_0),\ \frac{\mathrm{d}}{\mathrm{d}t}(\iota\circ\eta)(t) = \varphi_\alpha(\eta(t))\iota^* H_i(t,\iota(\eta(t)))\dot{\phi}^i(t)\right.\\
&\qquad\left. -\frac{1}{2}\varphi_\alpha(\eta(t))\iota^*\left[(\partial_1 g(t))_{ij}V^{ij}\right](\iota(\eta(t)))\right\}\\
&= \inf\left\{\int_0^1 |\dot{\phi}(t)|^2_{\mathbb{R}^d}\,\mathrm{d}t \,\middle|\, \eta(0) = u_0,\ \dot{\eta}(t) = \varphi_\alpha(\eta(t))H_i(t,\eta(t))\dot{\phi}^i(t)\right.\\
&\qquad\left. -\frac{1}{2}\varphi_\alpha(\eta(t))\left[(\partial_1 g(t))_{ij}V^{ij}\right](\eta(t))\right\}
\end{aligned}$$

Now, if we set $X^{n,\alpha}_t := \pi(U^{n,\alpha}_t)$, it follows from the continuity of the projection $\pi : FM \to M$ and the contraction principle that $\{X^{n,\alpha}\}_{n\geqslant 1}$ satisfies the large deviation principle in $C([0,1];M)$ with good rate function I^{α}_M given by

$$I^{\alpha}_M(\zeta) = \inf\{I^{\alpha}_{FM}(\eta) | \pi(\eta) = \zeta\}.$$

We show how to obtain the desired expression for I^{α}_M, at least when $\zeta \in C([0,1];K_\alpha)$. Consider such a curve ζ and suppose that $\eta : [0,1] \to FM$ is such that $\pi\eta = \zeta$ and $I^{\alpha}_{FM}(\eta) < \infty$. Then $\eta(0) = u_0$, η (and hence also ζ) is almost everywhere differentiable and there exists a $\phi : [0,1] \to \mathbb{R}^d$ such that

$$\dot{\eta}(t) = \varphi_\alpha(\eta(t))H_i(t,\eta(t))\dot{\phi}^i(t) - \frac{1}{2}\varphi_\alpha(\eta(t))\left[(\partial_1 g(t))_{ij}V^{ij}\right](\eta(t)). \tag{7.3.7}$$

Since $\eta(0) = u_0 \in \mathcal{O}_{(0,x_0)}$, the solution $\tilde{\eta}$ of the equation

$$\dot{\tilde{\eta}}(t) = H_i(t,\tilde{\eta}(t))\dot{\phi}^i(t) - \frac{1}{2}\left[(\partial_1 g(t))_{ij}V^{ij}\right](\eta(t)),$$

with $\tilde{\eta}(0) = u_0$ satisfies $\tilde{\eta}(t) \in \mathcal{O}_{(t,\zeta(t))}$ for all $t \in [0,1]$. Since $\zeta(t) \in K_\alpha$, we find that $\tilde{\eta}(t) \in \tilde{K}_\alpha$ and hence $\varphi_\alpha(\tilde{\eta}(t)) = 1$ for all $t \in [0,1]$. But then $\tilde{\eta}(t)$ is also the solution of (7.3.7). We conclude that η is the unique horizontal lift with respect to $\{g(t)\}_{t\in[0,1]}$ of ζ with $\eta(0) = u_0$. In that case, ϕ is the anti-development with respect to $\{g(t)\}_{t\in[0,1]}$ of ζ (see Section 7.2.3), and we have

$$|\dot{\phi}(t)|_{\mathbb{R}^d} = |\eta(t)\dot{\zeta}(t)|_{\mathbb{R}^d} = |\dot{\zeta}(t)|_{g(t)}.$$

Therefore, if ζ is contained in K_α and almost everywhere differentiable, then the rate function reduces to

$$I^{\alpha}_M(\zeta) = \frac{1}{2}\int_0^1 |\dot{\zeta}(t)|^2_{g(t)}\,\mathrm{d}t.$$

If ζ is not almost everywhere differentiable, the above argument shows that $I_M^\alpha(\zeta) = \infty$.

Finally, to deduce the large deviations for $\{X^n\}_{n\geqslant 1}$ from the large deviations for $\{X^{n,\alpha}\}_{n\geqslant 1}$ is done in exactly the same way as in the proof of Theorem 5.1.3 given in Section 5.4. □

7.4. Concluding remarks

We conclude this chapter by discussing some further directions which can be investigated which are related to the results in this chapter.

First of all, when comparing this chapter to Chapter 6, we have only extended the result of the time-inhomogeneous Schilder-type theorem from Section 6.3 to the geometric setting. It is natural to ask if the result concerning time-inhomogeneous random walks can also be extended to time-inhomogeneous geodesic random walks. We can use a similar approach as in Section 7.2 to define geodesics and parallel transport with respect to a family of connections. This allows us to define time-inhomogeneous geodesic random walks with indepedent, identically distributed increments in the sense of Chapter 3. Furthermore, it also allows us to define a time-inhomogeneous Riemannian exponential map. If it is possible to study this exponential map in a similar way as done in Section 3.4, then one should be able to obtain the analogue of Cramér's theorem for time-inhomogeneous random walks.

Another direction we can think about, is to consider randomly evolving Riemannian manifolds. In this case, we will not have a time-dependent Riemannian metric, but we consider a random process in the space of Riemannian metrics. All the classical large deviation theorems can then be studied in this setting. In this case, there are two sources of randomness which can give rise to large deviations. The contribution of both has to be understood, and it is interesting to see if both contribute on the same scale, or if there occur different scales of large deviations.

Bibliography

[1] Andrei Agrachev et al. “Intrinsic random walks in Riemannian and sub-Riemannian geometry *via* volume sampling”. In: *ESAIM Control Optim. Calc. Var.* 24.3 (2018), pp. 1075–1105. ISSN: 1292-8119. DOI: 10.1051/cocv/2017037. URL: https://doi.org/10.1051/cocv/2017037.

[2] Noga Alon and Joel H. Spencer. *The probabilistic method*. Fourth. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 2016.

[3] Ben Andrews and Christopher Hopper. *The Ricci flow in Riemannian geometry*. Vol. 2011. Lecture Notes in Mathematics. A complete proof of the differentiable 1/4-pinching sphere theorem. Springer, Heidelberg, 2011.

[4] Alexis Arnaudon et al. “Diffusion bridges for stochastic Hamiltonian systems with applications to shape analysis”. In: *Preprint; ArXiv: 2002.00885* 2002.00885 (2020). ArXiv: 2002.00885.

[5] Marc Arnaudon, Kolehe Abdoulaye Coulibaly, and Anton Thalmaier. “Brownian motion with respect to a metric depending on time: definition, existence and applications to Ricci flow”. In: *C. R. Math. Acad. Sci. Paris* 346.13-14 (2008), pp. 773–778. ISSN: 1631-073X. DOI: 10.1016/j.crma.2008.05.004. URL: https://doi.org/10.1016/j.crma.2008.05.004.

[6] Marc Arnaudon, Anton Thalmaier, and Feng-Yu Wang. “Harnack inequality and heat kernel estimates on manifolds with curvature unbounded below”. In: *Bull. Sci. Math.* 130.3 (2006), pp. 223–233. ISSN: 0007-4497. URL: https://doi.org/10.1016/j.bulsci.2005.10.001.

[7] V. Arnold. “Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l’hydrodynamique des fluides parfaits”. In: *Ann. Inst. Fourier (Grenoble)* 16.fasc., fasc. 1 (1966), pp. 319–361. ISSN: 0373-0956. URL: http://www.numdam.org/item?id=AIF_1966__16_1_319_0.

[8] V. I. Arnold. *Mathematical methods of classical mechanics*. Second. Vol. 60. Graduate Texts in Mathematics. Translated from the Russian by K. Vogtmann and A. Weinstein. Springer-Verlag, New York, 1989.

[9] R. Azencott. “Grandes déviations et applications”. In: *Eighth Saint Flour Probability Summer School—1978 (Saint Flour, 1978)*. Vol. 774. Lecture Notes in Math. Springer, Berlin, 1980, pp. 1–176.

[10] Paolo Baldi and Lucia Caramellino. “Large and moderate deviations for random walks on nilpotent groups”. In: *J. Theoret. Probab.* 12.3 (1999), pp. 779–809. ISSN: 0894-9840. DOI: 10.1023/A:1021684000713. URL: https://doi.org/10.1023/A:1021684000713.

[11] Richard Bellman. "Limit theorems for non-commutative operations. I". In: *Duke Math. J.* 21 (1954), pp. 491–500. ISSN: 0012-7094. URL: http://projecteuclid.org/euclid.dmj/1077465878.

[12] Patrick Billingsley. *Convergence of probability measures.* Second. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999, pp. x+277. ISBN: 0-471-19745-9. DOI: 10.1002/9780470316962. URL: https://doi.org/10.1002/9780470316962.

[13] Jean-Michel Bismut. *Large deviations and the Malliavin calculus.* Vol. 45. Progress in Mathematics. Birkhäuser Boston, Inc., Boston, MA, 1984.

[14] Philippe Bougerol and Jean Lacroix. *Products of random matrices with applications to Schrödinger operators.* Vol. 8. Progress in Probability and Statistics. Birkhäuser Boston, Inc., Boston, MA, 1985.

[15] E. Breuillard. "Random walks on Lie groups". In: 2004.

[16] Marcus Brubaker, Mathieu Salzmann, and Raquel Urtasun. "A Family of MCMC Methods on Implicitly Defined Manifolds". In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics.* Ed. by Neil D. Lawrence and Mark Girolami. Vol. 22. Proceedings of Machine Learning Research. La Palma, Canary Islands: PMLR, 2012, pp. 161–172. URL: http://proceedings.mlr.press/v22/brubaker12.html.

[17] Hernán Cendra et al. "Variational principles for Lie-Poisson and Hamilton-Poincaré equations". In: vol. 3. 3. {Dedicated to Vladimir Igorevich Arnold on the occasion of his 65th birthday}. 2003, pp. 833–867, 1197–1198. DOI: 10.17323/1609-4514-2003-3-3-833-867. URL: https://doi.org/10.17323/1609-4514-2003-3-3-833-867.

[18] Li-Juan Cheng and Anton Thalmaier. "Characterization of pinched Ricci curvature by functional inequalities". In: *J. Geom. Anal.* 28.3 (2018), pp. 2312–2345. ISSN: 1050-6926. DOI: 10.1007/s12220-017-9905-1. URL: https://doi.org/10.1007/s12220-017-9905-1.

[19] Li-Juan Cheng and Anton Thalmaier. "Spectral gap on Riemannian path space over static and evolving manifolds". In: *J. Funct. Anal.* 274.4 (2018), pp. 959–984. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2017.12.004. URL: https://doi.org/10.1016/j.jfa.2017.12.004.

[20] Herman Chernoff. "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations". In: *Ann. Math. Statistics* 23 (1952), pp. 493–507. ISSN: 0003-4851. DOI: 10.1214/aoms/1177729330. URL: https://doi.org/10.1214/aoms/1177729330.

[21] Claude Chevalley. *Theory of Lie Groups. I.* Princeton Mathematical Series, vol. 8. Princeton University Press, Princeton, N. J., 1946.

[22] Francesca Collet and Richard C. Kraaij. “Dynamical moderate deviations for the Curie–Weiss model”. In: *Stochastic Process. Appl.* 127.9 (2017), pp. 2900–2925. ISSN: 0304-4149. DOI: 10.1016/j.spa.2017.01.002. URL: http://dx.doi.org/10.1016/j.spa.2017.01.002.

[23] Koléhè A. Coulibaly-Pasquier. “Brownian motion with respect to time-changing Riemannian metrics, applications to Ricci flow”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 47.2 (2011), pp. 515–538. ISSN: 0246-0203. DOI: 10.1214/10-AIHP364. URL: https://doi.org/10.1214/10-AIHP364.

[24] Koléhè A. Coulibaly-Pasquier. “Onsager-Machlup functional for uniformly elliptic time-inhomogeneous diffusion”. In: *Séminaire de Probabilités XLVI*. Vol. 2123. Lecture Notes in Math. Springer, Cham, 2014, pp. 105–123. DOI: 10.1007/978-3-319-11970-0_5. URL: https://doi.org/10.1007/978-3-319-11970-0_5.

[25] H. Cramér. “Sur un nouveau théorème limite dans la théorie des probabilités”. In: *Colloque consacré à la théoriedes probabilités* 736 (1938), pp. 2–23.

[26] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. “User’s guide to viscosity solutions of second order partial differential equations”. In: *Bull. Amer. Math. Soc. (N.S.)* 27.1 (1992), pp. 1–67. ISSN: 0273-0979. DOI: 10.1090/S0273-0979-1992-00266-5. URL: http://dx.doi.org/10.1090/S0273-0979-1992-00266-5.

[27] A. Crisanti et al. “Products of random matrices for disordered systems”. In: *Phys. Rev. E* 49 (2 1994), R953–R955. DOI: 10.1103/PhysRevE.49.R953. URL: https://link.aps.org/doi/10.1103/PhysRevE.49.R953.

[28] Donald A. Dawson and Jürgen Gärtner. “Large deviations from the McKean-Vlasov limit for weakly interacting diffusions”. In: *Stochastics* 20.4 (1987), pp. 247–308. ISSN: 0090-9491. DOI: 10.1080/17442508708833446. URL: https://doi.org/10.1080/17442508708833446.

[29] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*. Second. Vol. 38. Applications of Mathematics (New York). Springer-Verlag, New York, 1998.

[30] Jean-Dominique Deuschel and Daniel W. Stroock. *Large deviations*. Vol. 137. Pure and Applied Mathematics. Academic Press, Inc., Boston, MA, 1989.

[31] M. D. Donsker and S. R. S. Varadhan. “Asymptotic evaluation of certain Markov process expectations for large time. III”. In: *Comm. Pure Appl. Math.* 29.4 (1976), pp. 389–461. ISSN: 0010-3640. DOI: 10.1002/cpa.3160290405. URL: https://doi.org/10.1002/cpa.3160290405.

[32] M. D. Donsker and S. R. S. Varadhan. “Asymptotic evaluation of certain Markov process expectations for large time. IV”. In: *Comm. Pure Appl. Math.* 36.2 (1983), pp. 183–212. ISSN: 0010-3640. DOI: 10.1002/cpa.3160360204. URL: https://doi.org/10.1002/cpa.3160360204.

[33] M. D. Donsker and S. R. S. Varadhan. “Large deviations for stationary Gaussian processes”. In: *Comm. Math. Phys.* 97.1-2 (1985), pp. 187–210. ISSN: 0010-3616. URL: http://projecteuclid.org/euclid.cmp/1103941986.

[34] Richard S. Ellis. “Large deviations for a general class of random vectors”. In: *Ann. Probab.* 12.1 (1984), pp. 1–12. ISSN: 0091-1798. URL: http://links.jstor.org/sici?sici=0091-1798(198402)12:1<1:LDFAGC>2.0.CO;2-D&origin=MSN.

[35] K. D. Elworthy. *Stochastic differential equations on manifolds*. Vol. 70. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge-New York, 1982.

[36] Michel Émery. *Stochastic calculus in manifolds*. Universitext. With an appendix by P.-A. Meyer. Springer-Verlag, Berlin, 1989.

[37] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Characterization and convergence. John Wiley & Sons, Inc., New York, 1986.

[38] Lawrence C. Evans. *An introduction to stochastic differential equations*. American Mathematical Society, Providence, RI, 2013, pp. viii+151. ISBN: 978-1-4704-1054-4. DOI: 10.1090/mbk/082. URL: https://doi.org/10.1090/mbk/082.

[39] Jin Feng and Thomas G. Kurtz. *Large deviations for stochastic processes*. Vol. 131. Mathematical Surveys and Monographs. American Mathematical Society, 2006.

[40] Theodore Frankel. *The geometry of physics*. Second. An introduction. Cambridge University Press, Cambridge, 2004.

[41] Mark I. Freidlin and Alexander D. Wentzell. *Random perturbations of dynamical systems*. Third. Vol. 260. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Translated from the 1979 Russian original by Joseph Szücs. Springer, Heidelberg, 2012.

[42] H. Furstenberg and H. Kesten. “Products of random matrices”. In: *Ann. Math. Statist.* 31 (1960), pp. 457–469. ISSN: 0003-4851. DOI: 10.1214/aoms/1177705909. URL: https://doi.org/10.1214/aoms/1177705909.

[43] Robert G. Gallager. *Information Theory and Reliable Communication*. USA: John Wiley Sons, Inc., 1968. ISBN: 0471290483.

[44] Bernard Gaveau. “Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents”. In: *Acta Math.* 139.1-2 (1977), pp. 95–153. ISSN: 0001-5962. DOI: 10.1007/BF02392235. URL: https://doi.org/10.1007/BF02392235.

[45] Bart van Ginkel and Frank Redig. “Hydrodynamic limit of the symmetric exclusion process on a compact Riemannian manifold”. In: *J. Stat. Phys.* 178.1 (2020), pp. 75–116. ISSN: 0022-4715. DOI: 10.1007/s10955-019-02420-2. URL: https://doi.org/10.1007/s10955-019-02420-2.

[46] Maria Gordina. "Hilbert-Schmidt groups as infinite-dimensional Lie groups and their Riemannian geometry". In: *J. Funct. Anal.* 227.2 (2005), pp. 245–272. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2005.05.011. URL: https://doi.org/10.1016/j.jfa.2005.05.011.

[47] M. Guerra and A. Sarychev. "On the stochastic Lie algebra". In: *Preprint; ArXiv: 1805.07299* 1805.07299 (2018). ArXiv: 1805.07299.

[48] Y. Guivarc'h. "Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire". In: *Conference on Random Walks (Kleebach, 1979) (French)*. Vol. 74. Astérisque. Soc. Math. France, Paris, 1980, pp. 47–98, 3.

[49] Hongxin Guo, Robert Philipowski, and Anton Thalmaier. "Martingales on manifolds with time-dependent connection". In: *J. Theoret. Probab.* 28.3 (2015), pp. 1038–1062. ISSN: 0894-9840. DOI: 10.1007/s10959-013-0536-6. URL: https://doi.org/10.1007/s10959-013-0536-6.

[50] Jurgen Gärtner. "On large deviations from an invariant measure". In: *Teor. Verojatnost. i Primenen.* 22.1 (1977), pp. 27–42. ISSN: 0040-361x.

[51] Brian Hall. *Lie groups, Lie algebras, and representations*. Second. Vol. 222. Graduate Texts in Mathematics. An elementary introduction. Springer, Cham, 2015.

[52] Richard S. Hamilton. "Four-manifolds with positive curvature operator". In: *J. Differential Geom.* 24.2 (1986), pp. 153–179. ISSN: 0022-040X. URL: http://projecteuclid.org/euclid.jdg/1214440433.

[53] Richard S. Hamilton. "The Harnack estimate for the Ricci flow". In: *J. Differential Geom.* 37.1 (1993), pp. 225–243. ISSN: 0022-040X. URL: http://projecteuclid.org/euclid.jdg/1214453430.

[54] Robert Haslhofer and Aaron Naber. "Characterizations of the Ricci flow". In: *J. Eur. Math. Soc. (JEMS)* 20.5 (2018), pp. 1269–1302. ISSN: 1435-9855. DOI: 10.4171/JEMS/787. URL: https://doi.org/10.4171/JEMS/787.

[55] Sigurdur Helgason. *Differential geometry, Lie groups, and symmetric spaces*. Vol. 80. Pure and Applied Mathematics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1978.

[56] Frank den Hollander. *Large deviations*. Vol. 14. Fields Institute Monographs. American Mathematical Society, Providence, RI, 2000.

[57] Elton P. Hsu. *Stochastic analysis on manifolds*. Vol. 38. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2002.

[58] Erik Jørgensen. "The central limit problem for geodesic random walks". In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 32 (1975), pp. 1–64. DOI: 10.1007/BF00533088. URL: http://dx.doi.org/10.1007/BF00533088.

[59] Wilhelm Klingenberg. *Riemannian geometry*. Vol. 1. de Gruyter Studies in Mathematics. Walter de Gruyter & Co., Berlin-New York, 1982.

[60] Anthony W. Knapp. *Lie groups beyond an introduction*. Vol. 140. Progress in Mathematics. Birkhäuser Boston, Inc., Boston, MA, 1996.

[61] Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of differential geometry. Vol I.* Interscience Publishers, a division of John Wiley & Sons, New York-London, 1963.

[62] Richard Kraaij. "Large deviations for finite state Markov jump processes with mean-field interaction via the comparison principle for an associated Hamilton-Jacobi equation". In: *J. Stat. Phys.* 164.2 (2016), pp. 321–345. ISSN: 0022-4715. DOI: 10.1007/s10955-016-1542-8. URL: http://dx.doi.org/10.1007/s10955-016-1542-8.

[63] Richard C. Kraaij, Frank Redig, and Rik Versendaal. "Classical large deviation theorems on complete Riemannian manifolds". In: *Stochastic Process. Appl.* 129.11 (2019), pp. 4294–4334. ISSN: 0304-4149. DOI: 10.1016/j.spa.2018.11.019. URL: https://doi.org/10.1016/j.spa.2018.11.019.

[64] Kazumasa Kuwada. "Convergence of time-inhomogeneous geodesic random walks and its application to coupling methods". In: *Ann. Probab.* 40.5 (2012), pp. 1945–1979. ISSN: 0091-1798. DOI: 10.1214/11-AOP676. URL: https://doi.org/10.1214/11-AOP676.

[65] Kazumasa Kuwada and Robert Philipowski. "Non-explosion of diffusion processes on manifolds with time-dependent metric". In: *Math. Z.* 268.3-4 (2011), pp. 979–991. ISSN: 0025-5874. DOI: 10.1007/s00209-010-0704-7. URL: https://doi.org/10.1007/s00209-010-0704-7.

[66] Émile Le Page. "Théorèmes limites pour les produits de matrices aléatoires". In: *Probability measures on groups (Oberwolfach, 1981).* Vol. 928. Lecture Notes in Math. Springer, Berlin-New York, 1982, pp. 258–303.

[67] Michel Ledoux. *The concentration of measure phenomenon.* Vol. 89. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001.

[68] John M. Lee. *Introduction to smooth manifolds.* Vol. 218. Graduate Texts in Mathematics. Springer-Verlag, New York, 2003.

[69] John M. Lee. *Riemannian manifolds.* Vol. 176. Graduate Texts in Mathematics. An introduction to curvature. Springer-Verlag, New York, 1997.

[70] Marco Lorenzi and Xavier Pennec. "Geodesics, parallel transport & one-parameter subgroups for diffeomorphic image registration". In: *Int. J. Comput. Vis.* 105.2 (2013), pp. 111–127. ISSN: 0920-5691. DOI: 10.1007/s11263-012-0598-4. URL: https://doi.org/10.1007/s11263-012-0598-4.

[71] Paul Malliavin. *Géométrie différentielle stochastique.* Vol. 64. Séminaire de Mathématiques Supérieures [Seminar on Higher Mathematics]. Notes prepared by Danièle Dehen and Dominique Michel. Presses de l'Université de Montréal, Montreal, Que., 1978, p. 144. ISBN: 0-8405-0425-X.

[72] Paul Malliavin. *Stochastic analysis.* Vol. 313. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1997, pp. xii+343. ISBN: 3-540-57024-1. DOI: 10.1007/978-3-642-15074-6. URL: https://doi.org/10.1007/978-3-642-15074-6.

[73] John Milnor. "Curvatures of left invariant metrics on Lie groups". In: *Advances in Math.* 21.3 (1976), pp. 293–329. ISSN: 0001-8708. DOI: 10.1016/S0001-8708(76)80002-3. URL: https://doi.org/10.1016/S0001-8708(76)80002-3.

[74] A. A. Mogulskii. "Large deviations for the trajectories of multidimensional random walks". In: *Theor. Probability Appl.* 21.2 (1976), pp. 300–315. ISSN: 0040-361x.

[76] Daniel Neuenschwander. *Probabilities on the Heisenberg group.* Vol. 1630. Lecture Notes in Mathematics. Limit theorems and Brownian motion. Springer-Verlag, Berlin, 1996.

[77] Peter Petersen. *Riemannian geometry.* Second. Vol. 171. Graduate Texts in Mathematics. Springer, New York, 2006.

[78] Huyên Pham. "Some applications and methods of large deviations in finance and insurance". In: *Paris-Princeton Lectures on Mathematical Finance 2004.* Vol. 1919. Lecture Notes in Math. Springer, Berlin, 2007, pp. 191–244. DOI: 10.1007/978-3-540-73327-0_5. URL: https://doi.org/10.1007/978-3-540-73327-0_5.

[79] David G. Poole. "The stochastic group". In: *Amer. Math. Monthly* 102.9 (1995), pp. 798–801. ISSN: 0002-9890. DOI: 10.2307/2974507. URL: https://doi.org/10.2307/2974507.

[80] M. M. Postnikov. *Geometry VI.* Vol. 91. Encyclopaedia of Mathematical Sciences. Riemannian geometry, Translated from the 1998 Russian edition by S. A. Vakhrameev. Springer-Verlag, Berlin, 2001.

[81] P. H. Roberts and H. D. Ursell. "Random walk on a sphere and on a Riemannian manifold". In: *Philos. Trans. Roy. Soc. London Ser. A* 252 (1960), pp. 317–356. ISSN: 0080-4614. DOI: 10.1098/rsta.1960.0008. URL: https://doi.org/10.1098/rsta.1960.0008.

[82] R. Tyrrell Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[83] M. Schilder. "Some asymptotic formulas for Wiener integrals". In: *Trans. Amer. Math. Soc.* 125 (1966), pp. 63–85. ISSN: 0002-9947. DOI: 10.2307/1994588. URL: https://doi.org/10.2307/1994588.

[84] C. E. Shannon. "A mathematical theory of communication". In: *Bell System Tech. J.* 27 (1948), pp. 379–423, 623–656. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1948.tb01338.x. URL: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[85] A. Singer. "From graph to manifold Laplacian: the convergence rate". In: *Appl. Comput. Harmon. Anal.* 21.1 (2006), pp. 128–134. ISSN: 1063-5203. DOI: 10.1016/j.acha.2006.03.004. URL: https://doi.org/10.1016/j.acha.2006.03.004.

[86] Michael Spivak. *A comprehensive introduction to differential geometry. Vol. I.* Second. Publish or Perish, Inc., Wilmington, Del., 1979.

[87] Michael Spivak. *A comprehensive introduction to differential geometry. Vol. II.* Second. Publish or Perish, Inc., Wilmington, Del., 1979.

[88] D. W. Stroock. *An introduction to the theory of large deviations.* Universitext. Springer-Verlag, New York, 1984.

[89] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. "A Global Geometric Framework for Nonlinear Dimensionality Reduction". In: *Science* 290.5500 (2000), pp. 2319–2323. ISSN: 0036-8075. DOI: 10.1126/science.290.5500.2319. eprint: https://science.sciencemag.org/content/290/5500/2319.full.pdf. URL: https://science.sciencemag.org/content/290/5500/2319.

[90] Hugo Touchette. "The large deviation approach to statistical mechanics". In: *Phys. Rep.* 478.1-3 (2009), pp. 1–69. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2009.05.002. URL: https://doi.org/10.1016/j.physrep.2009.05.002.

[91] V. Tutubalin. "On Limit Theorems for the Product of Random Matrices". In: *Theory of Probability & Its Applications* 10.1 (1965), pp. 15–27. DOI: 10.1137/1110002. URL: https://doi.org/10.1137/1110002.

[92] V. S. Varadarajan. *Lie groups, Lie algebras, and their representations.* Vol. 102. Graduate Texts in Mathematics. Reprint of the 1974 edition. Springer-Verlag, New York, 1984.

[93] S. R. S. Varadhan. "Diffusion processes in a small time interval". In: *Comm. Pure Appl. Math.* 20 (1967), pp. 659–685. ISSN: 0010-3640. DOI: 10.1002/cpa.3160200404. URL: http://dx.doi.org/10.1002/cpa.3160200404.

[94] S. R. S. Varadhan. "Random walks in a random environment". In: *Proc. Indian Acad. Sci. Math. Sci.* 114.4 (2004), pp. 309–318. ISSN: 0253-4142. DOI: 10.1007/BF02829438. URL: https://doi.org/10.1007/BF02829438.

[95] Rik Versendaal. "Large deviations for Brownian motion in evolving Riemannian manifolds". In: *Preprint; ArXiv:* 2004.00358 (2020). ArXiv: 2004.00358.

[96] Rik Versendaal. "Large deviations for geodesic random walks". In: *Electron. J. Probab.* 24 (2019), Paper No. 93, 39 pp.

[97] Rik Versendaal. "Large deviations for random walks on Lie groups". In: *Preprint; ArXiv:* 1909.05065 (2019). ArXiv: 1909.05065.

[98] Stefan Waldmann. "Geometric wave equations". In: *Lecture notes; ArXiv:* 1208.4706v1 (2012). ArXiv: 1208.4706v1.

[99] Frank W. Warner. *Foundations of differentiable manifolds and Lie groups.* Vol. 94. Graduate Texts in Mathematics. Corrected reprint of the 1971 edition. Springer-Verlag, New York-Berlin, 1983.

[100] Atilla Yilmaz. *Large deviations for random walk in a random environment.* Thesis (Ph.D.)–New York University. ProQuest LLC, Ann Arbor, MI, 2008, p. 82. isbn: 978-0549-82338-4.

[101] Y. Zhao, S. Tasoulis, and T. Roos. "Manifold Visualization via Short Walks". In: *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization: Short Papers.* EuroVis '16. Groningen, The Netherlands: Eurographics Association, 2016, 85–89.

Summary

This thesis is concerned with large deviations for processes in Riemannian manifolds. In particular, we study the extensions of large deviations for random walks and Brownian motion to the geometric setting.

In the first chapters, we study large deviations for random walks in various geometric settings. First, in Chapter 3 we consider geodesic random walks in Riemannian manifolds. Since geodesic random walks are not simply sums of random variables, we discuss a notion of independent, identically distributed increments for such random walks. We then prove the large deviation principle for geodesic random walks with independent, identically distributed increments which are bounded en centered. The idea of the proof is to relate the large deviations for the geodesic random walk to the large deviations for a random walk in a tangent space. Since the tangent space is a vector space, these large deviations follow from the original Cramér's theorem. The desired comparison is made by carefully analyzing the geometric properties of geodesic random walks, and in particular the short-time behaviour of the Riemannian exponential map.

In Chapter 4 we study random walks in special Riemannian manifolds, namely Lie groups. The additional group structure allows us to identify each tangent space with the Lie algebra. This lets us define a random walk in a Lie group as product of group elements which are the exponential of an element of the Lie algebra. We explain that such random walks are geodesic random walks for the Levi-Civita connection if and only if the Riemannian metric is bi-invariant. We then prove the large deviation principle for such random walks. The proof follows similar steps as the proof of the large deviation principle for geodesic random walks. However, the estimates are obtained differently and make use of the Baker-Campbell-Hausdorff formula, rather than properties of the Riemannian exponential map.

In chapter 5 we move on to path-space large deviations for processes in Riemannian manifolds. In particular, we prove path-space large deviation principles for geodesic random walks and Riemannian Brownian motion. Although the result for Riemannian Brownian motion is already known, we provide two novel approaches to obtain this result. We prove the path-space large deviation principle for geodesic random walks via a general method to study large deviations for Markov processes. This method relies on the convergence of non-linear semigroups and viscosity solutions for Hamiltonian-Jacobi equations. Furthermore, we show how this method can be used to study the large deviations for Riemannian Brownian motion. For the latter, we also provide a proof by horizontally lifting the Brownian motion to the frame bundle over the manifold. The horizontal lift satisfies a stochastic differential equation driven by a Euclidean Brownian motion. To prove the large deviation principle, we embed the frame bundle into Euclidean space, push-forward the stochastic

differential equation and apply Freidlin-Wentzell theory. The large deviations for Riemannian Brownian motion then follow from the contraction principle.

The final chapters are concerned with large deviations for time-inhomogeneous processes, both in the Euclidean and geometric context. First, in Chapter 6 we restrict to the Euclidean setting. We prove the large deviation principle for random walks with time-inhomogeneous increments. Furthermore, we show how to obtain from this the associated path-space large deviations. We conclude the chapter by studying large deviations for a diffusion generated by a weighted Laplacian, where the weights depend on time. Since such a diffusion is a Gaussian process, this result is already known. However, we provide an alternative proof that shows how this result can be obtained from the path-space large deviations for time-inhomogeneous random walks. The results in this chapter serve as motivation for obtaining similar results for time-inhomogeneous processes in a geometric setting. In the final chapter, we initiate this direction.

More precisely, the final chapter, Chapter 7, is concerned with large deviations for Riemannian Brownian motion in a time-evolving Riemannian manifold. For this, we consider a manifold equipped with a Riemannian metric which depends on time. First of all, we explain how to define Riemannian Brownian motion in this setting. Then, to prove the large deviation principle, we follow the lifting approach taken in Chapter 5. In order make this work, we define the notion of horizontal lift to the frame bundle with respect to a time-dependent collection of connections. By also considering the associated anti-development to Euclidean space, one obtains a stochastic differential equation driven by a Euclidean Brownian motion for the horizontally lifted process. By embedding into Euclidean space and applying Freidlin-Wentzell theory (adapted to work for time-dependent drift and diffusion), we obtain the large deviations for the embedded process. The contraction principle then gives us the large deviations for the Riemannian Brownian motion in the evolving Riemannian manifold.

Samenvatting

Dit proefschrift behandelt grote afwijkingen voor processen in Riemannse manifolds. In het bijzonder bestuderen we uitbreidingen van grote afwijkingen voor random walks en Brownse beweging naar de meetkundige context.

In de eerste hoofdstukken bestuderen we de grote afwijkingen voor random walks in verschillende meetkundige omgevingen. Allereerst behandelen we in Hoofdstuk 3 geodetische random walks in Riemannse manifolds. Aangezien geodetische random walks niet simpelweg geschreven kunnen worden als som van kansvariabelen, bespreken we een notie van onafhankelijk, identiek verdeelde incrementen voor dit soort random walks. Vervolgens bewijzen we het grote afwijkingen principe voor geodetische random walks met onafhanlijk, identiek verdeelde incrementen die begrensd en gecentreerd zijn. Het idee van het bewijs is om de grote afwijkingen voor de geodetische random walk te relateren aan grote afwijkingen voor een random walk in een raakruimte. Aangezien de raakruimte een vectorruimte is, volgen deze grote afwijkingen uit de originele versie van Cramérs stelling. De gewenste vergelijking wordt verkregen door een zorgvuldige analyse van de meetkundige eigenschappen van geodetische random walks en in het bijzonder van het korte-tijd gedrag van de Riemannse exponentiële afbeelding.

In Hoofdstuk 4 bestuderen we random walks in speciale Riemannse manifolds, namelijk Lie groepen. De extra groepsstructuur zorgt ervoor dat we elke raakruimte kunnen identificeren met de Lie algebra. Hierdoor kunnen we een random walk in een Lie groep definiëren als product van groepselementen die het beeld zijn van een element van de Lie algebra onder de exponentiële afbeelding. We leggen uit dat dit soort random walks geodetische random walks voor de Levi-Civita connectie zijn dan en slechts dan als de Riemannse metriek bi-invariant is. Vervolgens bewijzen we het grote afwijkingen principe voor dit soort random walks. Het bewijs is vergelijkbaar met het bewijs van het grote afwijkingen principe voor geodetische random walks. Echter, de afschattingen worden op een andere manier verkregen en maken gebruik van de Baker-Campbell-Hausdorff formule in plaats van eigenschappen van de Riemannse exponentiële afbeelding.

In Hoofdstuk 5 gaan we over naar padsgewijze grote afwijkingen voor processen in Riemannse manifolds. In het bijzonder bewijzen we padsgewijze grote afwijkingen principes voor geodetische random walks en Riemannse Brownse beweging. Hoewel het resultaat voor Riemannse Brownse beweging reeds bekend is, geven wij twee nieuwe aanpakken om dit resultaat te verkrijgen. We bewijzen het padsgewijze grote afwijkingen principe voor geodetische random walks via een algemene methode om grote afwijkingen voor Markovprocessen te bestuderen. Deze methode is gebaseerd op de convergentie van niet-lineaire halfgroepen en viscositeitsoplossingen voor Hamilton-Jacobi vergelijkingen. Verder laten we zien hoe deze methode

gebruikt kan worden om de grote afwijkingen voor Riemannse Brownse beweging te bestuderen. Voor laatstgenoemde geven we ook een bewijs door de Brownse beweging horizontaal te liften naar de frame bundel over de manifold. De horizontale lift voldoet aan een stochastische differentiaalvergelijking gedreven door een Euclidische Brownse beweging. Om het grote afwijkingen principe te bewijzen, embedden we de frame bundel in een Euclidische ruimte, zetten we de stochastische differentiaalvergelijking over en passen we Freidlin-Wentzel theorie toe. De grote afwijkingen voor Riemannse Brownse beweging volgen daarna uit het contractieprincipe.

De laatste hoodstukken gaan over grote afwijkingen voor tijdsinhomogene processen, zowel in een Euclidische als meetkundige context. Allereerst behandelen we in Hoofdstuk 6 de Euclidische omgeving. We bewijzen het grote afwijkingen principe voor random walks met tijdsinhomogene incrementen. Verder laten we zien hoe hieruit de bijbehorende padsgewijze grote afwijkingen afgeleid kunnen worden. We sluiten het hoofdstuk af met het bestuderen van grote afwijkingen voor diffusies gegenereerd door een gewogen Laplaciaan, waarbij de gewichten tijdsafhankelijk zijn. Sinds dit soort diffusies Gaussische processen zijn, is dit resultaat al bekend. Echter geven wij een alternatief bewijs dat laat zien hoe dit resultaat volgt uit de padsgewijze grote afwijkingen voor tijdsinhomogene random walks. De resultaten in dit hoofdstuk dienen als motivatie voor het verkrijgen van vergelijkbare resultaten voor tijdsinhomogene processen in een meetkundige context. In het laatste hoofdstuk zetten we de eerste stappen in deze richting.

Preciezer gezegd, het laatste hoofdstuk, Hoofdstuk 7, behandelt grote afwijkingen voor Riemannse Brownse beweging in een evoluerende Riemannse manifold. Hiervoor rusten we de manifold uit met een Riemannse metriek die afhangt van de tijd. Allereerst leggen we uit hoe we in deze context een Riemannse Brownse beweging kunnen definiëren. Daarna bestuderen we grote afwijkingen hiervoor, waarbij we de aanpak volgen met de horizontale lift uit Hoofdstuk 5. Om dit te laten slagen, definiëren we een notie van horizontale lift met betrekking tot een tijdsafhankelijke familie van connecties. Als we vervolgens ook de anti-ontwikkeling naar een Euclidische ruimte beschouwen, verkrijgen we een stochastische differentiaalvergelijking gedreven door een Euclidische Brownse beweging voor het horizontaal gelifte proces. Door nu weer te embedden in een Euclidische ruimte en Freidlin-Wentzell theorie (aangepast voor tijdsafhankelijke drift en diffusie) toe te passen, verkrijgen we de grote afwijkingen voor het geëmbedde proces. Het contractieprincipe geeft vervolgens de grote afwijkingen voor de Riemannse Brownse beweging in de evoluerende Riemannse manifold.

Acknowledgements

First of all, I want to express my thanks to Frank en Jan for supporting me during the past years, already from far before the start of my PhD. Frank, thank you for all the probability theory you taught me, especially on large deviations. I enjoyed our discussions and the interesting background stories you told with great joy. However, I am the most grateful for your support in letting me to choose my own path and develop freely. Jan, thank you for preparing me for this adventure and for your useful comments that helped me improve my writing and presentation of mathematics. Furthermore, I want to thank the Peter Paul Peterich foundation, and especially Jasper Peterich, for their financial support and giving me the opportunity to carry out this PhD research.

I want to thank Anton Thalmaier for hosting me during my visit in Luxembourg. Our discussions on stochastic processes in manifolds, especially in time-evolving manifolds, have helped me a great deal. It improved my understanding of the underlying constructions and allowed me to continue my research in that direction. I also want to thank his group member, James, Li-Juan and Robert for discussing related topics and providing new insights, as well as for the interesting discussions during and after lunch with you and your fellow colleagues.

Back in Delft, I want to thank my PhD-colleagues, but also the other scientific staff of the probability, statistics and analysis groups for helpful and interesting discussions, both about math related topics and practically everything else. I also want to thank the supporting staff. Carl, for helping me out with my computer related problems. Stefanie, Cindy and the other secretaries for helping out with paper work and other organizatory issues.

I want to thank Bart, Elisa and Mario for the many discussion we had in our office about a diverse range of topics. I also want to thank you for listening to my complaints at times when I was stuck or when my motivation was dropping during the writing of this thesis.

Thanks to Bas Janssens, whose comments and suggestion as part of the doctoral committee lead to some improvements in the presentation of the work.

Thanks also to Richard Kraaij, to whom I feel like a successor and who showed me around the probability and statistics group the first day I arrived at the office, introducing me to everyone. I also want to thank you for our collaboration that formed the starting point of my research, helping me out with all kind of problems and questions regarding large deviations and giving me new ideas to continue.

Jasper, it is great to have a friend close by who follows the same road at the same time. It made it easier to fit in and it has given a lot of support.

Pa en ma, bedankt voor de steun die jullie mij gegeven hebben, en het vertrouwen dat dit voor mij de juiste weg is. Bedankt ook voor het aanhoren van de frustraties die er soms bij kwamen kijken. Bryan, dank je wel voor de discussies over wiskunde en het scherp houden van mijn bredere wiskundige kennis met je verschillende vragen.

Ik wil ook Poes en Wikky bedanken, die tijdens het schrijven van dit proefschrift en het verdere thuiswerken voor de nodige afleiding hebben gezorgd.

Als laatste wil ik mijn steun en toeverlaat bedanken, die al mijn problemen en frustraties aan wil horen, en er vervolgens voor zorgt dat ik weer tot rust kom. Die mij eraan herinnert om ook te ontspannen en ervoor zorgt dat ik ook aan mezelf denk. Dank je wel, Melissa.

Curriculum Vitæ

Rik Versendaal was born on the 29^{th} of July 1993 in Dordrecht, the Netherlands. He completed his high-school education in 2011 at 'C.S.G de Willem van Oranje' in Oud-Beijerland. He acquired both his Bachelor and Master degree in Applied Mathematics cum laude at TU Delft in 2014 and 2016 respectively. He wrote his Master thesis titled '*The Riesz transform on a complete Riemannian manifold with Ricci curvature bounded from below*' under the supervision of Prof. dr. J.M.A.M. van Neerven.

In the end of 2015, he worked as an intern at TBA, studying deadlock avoidance algorithms in the context of multi-agent routing.

In December 2015, he won the Peter Paul Peterich scholarship allowing him to carry out PhD research at TU Delft. In October 2016 he started his PhD research under the supervision of Prof. dr. F.H.J. Redig and Prof. dr. J.M.A.M. van Neerven at TU Delft. As mentioned, the research is funded by the Peter Paul Peterich foundation via the TU Delft University fund.

List of Publications

Publications:

Rik Versendaal. “Large deviations for geodesic random walks”. In: *Electron. J. Probab.* 24 (2019), Paper No. 93, 39 pp.

Richard C. Kraaij, Frank Redig, and Rik Versendaal. “Classical large deviation theorems on complete Riemannian manifolds”. In: *Stochastic Process. Appl.* 129.11 (2019), pp. 4294–4334. ISSN: 0304-4149. DOI: 10.1016/j.spa.2018.11.019. URL: https://doi.org/10.1016/j.spa.2018.11.019

Jan van Neerven and Rik Versendaal. “L^p-analysis of the Hodge-Dirac operator associated with Witten Laplacians on complete Riemannian manifolds”. In: *J. Geom. Anal.* 28.4 (2018), pp. 3109–3138. ISSN: 1050-6926. DOI: 10.1007/s12220-017-9947-4. URL: https://doi.org/10.1007/s12220-017-9947-4

Preprints:

Rik Versendaal. “Large deviations for Brownian motion in evolving Riemannian manifolds”. In: *Preprint; ArXiv:* 2004.00358 (2020). ArXiv: 2004.00358

Rik Versendaal. “Large deviations for random walks on Lie groups”. In: *Preprint; ArXiv:* 1909.05065 (2019). ArXiv: 1909.05065