Large deviations for slow–fast processes on connected complete Riemannian manifolds

Hu, Yanyan; Kraaij, Richard C.; Xi, Fubao

DOI
10.1016/j.spa.2024.104478

Publication date
2024

Document Version
Final published version

Published in
Stochastic Processes and their Applications

Stochastic Processes and their Applications 178 (2024) 104478

Contents lists available at ScienceDirect

Stochastic Processes and their Applications

journal homepage: www.elsevier.com/locate/spa

Large deviations for slow–fast processes on connected complete Riemannian manifolds

Yanyan Hu [a,*], Richard C. Kraaij [a], Fubao Xi [b]

[a] Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands
[b] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, P.R. China

ARTICLE INFO

MSC:
primary 60F10
60J25
secondary 60J35
49L25

Keywords:
Riemannian manifold
Large deviation principle
Hamilton–Jacobi–Bellman equations
Comparison principle
Action-integral representation

ABSTRACT

We consider a class of slow–fast processes on a connected complete Riemannian manifold M. The limiting dynamics as the scale separation goes to ∞ is governed by the averaging principle. Around this limit, we prove large deviation principles with an action-integral rate function for the slow process by nonlinear semigroup methods together with Hamilton–Jacobi–Bellman (HJB) equation techniques. Our main innovation is solving the comparison principle for viscosity solutions for the HJB equation on M and the construction of a variational viscosity solution for the non-smooth Hamiltonian, which lies at the heart of deriving the action integral representation for the rate function.

1. Introduction

In this paper, let M be a d-dimensional connected complete Riemannian manifold and $S = \{1, 2, \dots, N\}$, $N < \infty$. We consider a stochastic differential equation consisting of Riemannian Brownian motion with a switching drift on $M \times S$ with an initial value (x_0, k_0):

$$\mathrm{d}X_n^\varepsilon(t) = \frac{1}{\sqrt{n}} U_n^\varepsilon(t) \circ \mathrm{d}W(t) + b(X_n^\varepsilon(t), \Lambda_n^\varepsilon(t))\mathrm{d}t, \tag{1.1}$$

where $\Lambda_n^\varepsilon(t)$ is a switching process with transition rate on the set S,

$$\mathbb{P}(\Lambda_n^\varepsilon(t+\Delta) = j \mid \Lambda_n^\varepsilon(t) = i, X_n^\varepsilon(t) = x) = \frac{1}{\varepsilon} q_{ij}(x)\Delta + o(\Delta), \quad \text{if } j \neq i, \tag{1.2}$$

for small $\Delta > 0$, $i, j \in S$, $x \in M$, and $\varepsilon > 0$ is a small parameter. $U_n^\varepsilon(\cdot)$ is a unique element such that $X_n^\varepsilon(t) = \mathbf{p} U_n^\varepsilon(t)$, where $\mathbf{p} : O(M) \to M$ is the canonical projection map from the orthonormal frame bundle on $O(M)$ to M. Precise details and conditions on this system will be specified later. Obviously, (1.1) and (1.2) together is a slow–fast system.

It is not too difficult to see that under some conditions, the effective behavior of the slow process (1.1) can be accurately described by the averaged system as $\varepsilon \to 0$ and $n \to \infty$, utilizing the averaging principle. To be more specific, if $X_n^\varepsilon(t) \approx x$, if the jump coefficient $x \mapsto q_{ij}(x)$ is continuous and the jump-matrix is uniformly ergodic, one expects that the fast process $\Lambda_n^\varepsilon(t)$ equilibrates in the stationary measure corresponding to the jump kernel.

* Corresponding author.
E-mail address: y.hu-2@tudelft.nl (Y. Hu).

https://doi.org/10.1016/j.spa.2024.104478
Received 8 March 2024; Received in revised form 23 July 2024; Accepted 28 August 2024
Available online 31 August 2024

This observation implies that, as $\varepsilon \to 0$ and $n \to \infty$, the slow process converges to an averaged process defined as follows

$$\begin{cases} \mathrm{d}\bar{X}(t) = \bar{b}(\bar{X}(t))\mathrm{d}t, \\ \bar{X}(0) = x_0, \end{cases} \tag{1.3}$$

where $\bar{b}(x) = \sum_{i\in S} b(x,i)\pi_i^x(t)$ and $\pi^x(t) = (\pi_i^x(t))_{i\in S}$ is the unique invariant probability measure of the fast process with the slow variable being "frozen" at a deterministic point $x \in M$. The application of this averaging principle provides an effective method to reduce computational complexity. It can be viewed as a variant of the law of large numbers.

In contrast to the averaging principle, the large deviation principle (LDP) excels in providing a more precise description of the dynamic behavior, it specifically addresses the characterization of the exponential decay rate associated with probabilities of rare events. Informally, LDP is the estimate of the form

$$\mathbb{P}(X_n(t) \approx \gamma(t)) \sim e^{-nI(\gamma)}, \qquad \text{as } n \to \infty,$$

for $\gamma : [0,\infty) \to M$. I takes the form

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s),\dot{\gamma}(s))\,\mathrm{d}s, & \text{if } \gamma \in \mathcal{AC}(M), \\ \infty, & \text{otherwise}, \end{cases}$$

where $\mathcal{AC}$ denotes the set of absolutely continuous trajectories. I_0 quantifies the large deviations for $X_n(0)$ alone, and the map $\mathcal{L} : TM \to [0,\infty]$ is called the Lagrangian. The large deviation principle indeed quantifies the decay of probabilities for trajectories away from the solution of the averaging principle (1.3), as the solution of (1.3) is the unique trajectory for which $I(\bar{X}) = 0$.

The main purpose of this paper is to prove a LDP around such averaged process on M. The theory of LDP is one of the classical topics in probability theory, see [5,8,9], which has widespread applications in different areas such as information theory, thermodynamics, statistics, and engineering.

Let us mention some works related to our purposes. Huang, Mandjes and Spreij [12], studied large deviations for Markov-modulated diffusion processes with rapid switching. In [18], Peletier and Schlottke proved pathwise LDP of switching Markov processes by exploiting the connection between Hamilton–Jacobi (HJ) equations and Hamilton–Jacobi–Bellman (HJB) equations. In [15], Kraaij and Schlottke studied the LDP for the slow–fast system under regular conditions, where the fast process is a switching process. For the proof, they used the Bootstrapping procedure, which is a technology for comparison principle of the HJB equation. Later, Della Corte and Kraaij [4] continued to explore LDP in the context of molecular motors modeled by a diffusion process driven by the gradient of a weakly periodic potential that depends on an internal degree of freedom. The switch of the internal state, which can freely be interpreted as a molecular switch, is modeled as a Markov jump process that depends on the location of the motor. Subsequently, Hu, Kraaij, and Xi [11] considered the Cox–Ingersoll–Ross processes with state-dependent fast switching in the case of the degenerate diffusion coefficient.

Although there are extensive results on LDPs for slow–fast systems on Euclidean space, there is not much work in the context of Riemannian manifolds. Röckner and Zhang [19] studied sample path large deviations for diffusion processes on configuration spaces over a Riemannian manifold. Kraaij, Redig and Versendaal [14] generalized classical large deviation theorems on complete, smooth Riemannian manifolds, and also considered Riemannian Brownian motion in the single time-scale context. Furthermore, Versendaal [20] studied large deviations for $g(t)$-Brownian motion in a complete, evolving Riemannian manifold with respect to a collection $\{g(t)\}_{t\in\mathbb{R}}$ of Riemannian metrics, smoothly depending on t again in the single time-scale context.

Motivated by the aforementioned papers about LDP for slow–fast processes on Euclidean space and simple LDP on Riemannian manifold, it is a natural question to ask how to generalize the above large deviation results for slow–fast processes to Riemannian manifolds. In this paper, we address this question. That is we prove LDPs with an action-integral rate function for the slow process by nonlinear semigroup methods together with the HJB equation techniques. Note that our drift coefficient of slow process only satisfies locally one-sided Lipschitz continuity, which is weaker than the bounded condition. Moreover, the rate functions are related to the Hamiltonian $\mathcal{H} : T^*M \to [0,\infty]$ obtained by taking the Legendre transform of Lagrangian $\mathcal{L} : TM \to [0,\infty]$. One formally defines that

$$\mathcal{H}(x, \mathrm{d}f(x)) = \sup_{\pi\in\mathcal{P}(S)} \left\{ \int B_{x,\mathrm{d}f(x)}(z)\pi(\mathrm{d}z) - \mathcal{I}(x,\pi) \right\},$$

where

$$B_{x,\mathrm{d}f(x)}(z) = b(x,z)\mathrm{d}f(x) + \frac{1}{2}\,|\mathrm{d}f(x)|^2$$

coming from the slow process $X_n(t)$ and Donsker–Varadhan function

$$\mathcal{I}(x,\pi) = -\inf_{g>0} \int \frac{R_x g(z)}{g(z)} \pi(\mathrm{d}z),$$

where R_x is the generator corresponding to the fast process $\Lambda_n(t)$ defined by

$$R_x g(z) = \sum_{j\in S} q_{zj}(x)\,(g(j) - g(z))\,.$$

Although following the proof ideas from Feng and Kurtz's book [8], considering the comparison principle and the existence of solutions of HJB equations, we need to put forward some new ideas to show those owing to the special properties of the Riemannian manifold.

We first discover special properties on M, which have caused difficulties but also is the key innovation in our proof:

(i) The first one, to ensure the exponential tightness, we find a good containment function:

$$\Upsilon(x) = \frac{1}{2}\log(1 + f^2(x)),$$

where the smooth function $f(x)$ approximates $d(x_0, x)$ for some $x_0 \in M$ and satisfying formally $\sup_z \mathcal{H}(z, \mathrm{d}\Upsilon(z)) < C < \infty$ which plays the role of a relaxed Lyapunov function.

(ii) The second one, the distance function $d(x, y)$, x, $y \in M$ is never smooth. More specifically, $d(x, y)$ is not smooth on the cut-locus of x or y. This happens because the shortest path (geodesic) between two points may not be unique, for example, a spherical surface. Compared with $d(x, y)$, $d^2(x, y)$ is smooth when x closed to y. We use $d^2(x, y)$ in the proof of comparison principle.

(iii) The third one, we need to prove the global existence of solutions for a HJB equation on M to obtain an action-integral rate function. To establish existence we need to solve an appropriate control problem. A key obstacle is the construction from local solutions to global solutions.

Organization: The organization of our paper is as follows: in Section 2, we introduce fundamental concepts related to the large deviation principle. In Section 3, we construct a diffusion process with fast switching on the Riemannian manifold, and state our main results. Subsequently, in Section 4, we articulate the strategy employed in proving the large deviation. Based on the strategy in Section 4, a detailed proof of the main theorem is provided in Sections 5 and 6.

2. Preliminaries

The following convention will be used throughout the paper: C and c with or without indices will denote different positive constants whose values may change from one place to another.

We begin with the necessary definitions for introducing the large deviation principle on Riemannian manifold. Riemannian manifold is placed in Appendix A to highlight our main results. There is nothing special about defining a large deviation principle on Riemannian manifold. In the following, the definition is established on a Polish space $\mathcal{X}$.

Definition 2.1. Consider a sequence of $X_1, X_2, \ldots$ on Polish space $\mathcal{X}$. Furthermore let $I : \mathcal{X} \to [0, \infty]$.

(a) We say that I is a good rate function if for every $c \geqslant 0$, the set $\{x \in \mathcal{X} \mid I(x) \leqslant c\}$ is compact.

(b) We say that the sequence $\{X_n\}_{n\geqslant 1}$ is *exponentially tight* if for all $\alpha > 0$ there exists a compact set $K_\alpha \subseteq \mathcal{X}$ such that

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

(c) We say that the sequence $\{X_n\}_{n\geqslant 1}$ satisfies the *large deviation principle* with rate n and good rate function I, denoted by

$$\mathbb{P}[X_n \approx a] \sim \mathrm{e}^{-nI(a)}, \tag{2.1}$$

(i) if we have for every closed set $A \subseteq \mathcal{X}$ the upper bound,

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

(ii) and for every open set $U \subseteq \mathcal{X}$ the lower bound,

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

Definition 2.2 (*Absolutely Continuous Curves*)**.** We denote by $\mathcal{AC}(M)$ the space of absolutely continuous curves in M. A curve $\gamma : [0, T] \to M$ is absolutely continuous if there exists a function $g \in L^1[0, T]$ such that for $t \in [0, T]$ we have $\gamma(t) = \gamma(0) + \int_0^t g(s)\mathrm{d}s$. We write $g = \dot{\gamma}$.

A curve $\gamma : [0, \infty) \to M$ is absolutely continuous, i.e. $\gamma \in \mathcal{AC}(M)$, if the restriction to $[0, T]$ is absolutely continuous for every $T > 0$.

3. Constructing a diffusion process with fast switching on Riemannian manifold

In the above section, we only gave the basic knowledge about the large deviation principle. We next state the definition of the orthonormal frame bundle and horizontal lift to construct a diffusion process with switching on $M \times S$, for which we want to study the large deviation behavior,

$$\mathrm{d}X_n^\varepsilon(t) = \frac{1}{\sqrt{n}} U_n^\varepsilon(t) \circ \mathrm{d}W(t) + b(X_n^\varepsilon(t), \Lambda_n^\varepsilon(t))\mathrm{d}t, \tag{3.1}$$

where $\Lambda_n^\varepsilon(t)$ is a switching process with transition rate on the set S,

$$\mathbb{P}(\Lambda_n^\varepsilon(t + \Delta) = j \mid \Lambda_n^\varepsilon(t) = i, X_n^\varepsilon(t) = x) = \frac{1}{\varepsilon} q_{ij}(x)\Delta + o(\Delta), \quad \text{if } j \neq i, \tag{3.2}$$

for small $\Delta > 0$, $i, j \in S$, $x \in M$, and $\varepsilon > 0$ is a small parameter.

We start by establishing that the above process exists. As the switch is taking place on the finite set S, the key issue to be resolved is the non-explosiveness of the diffusion process (3.1). In the context without switching, non-explosiveness is implied by a lower bound on the curvature and gradient of the drift. We will also assume this for our result.

Assumption 3.1. For each $i \in S$, $b(\cdot, i)$ in (3.1) is a C^1-smooth vector field on M. There is a constant $\rho(n)$ such that the $CD(\rho(n), \infty)$ curvature condition

$$\inf_{i \in S} \mathcal{R}_g - \nabla b(\cdot, i) \geqslant \rho(n) g$$

holds where $\mathcal{R}_g$ is the Ricci tensor of the (co)-metric g.

Theorem 3.2. *Under Assumption 3.1, the system, (3.1) and (3.2), has a unique non-explosive strong solution $(X_n(t), \Lambda_n(t))$ with initial value $(X_n(0), \Lambda_n(0)) = (x_0, k_0)$.*

The proof follows the same method as Proposition 2.4 in [11]. We extend it to the context of Riemannian manifolds.

We next turn to present the definition of the orthonormal frame bundle and the horizontal lift.

Let $O_x(M)$ be the space of all orthonormal bases of $T_x M$. Denote $O(M) := \sqcup_{x \in M} O_x(M)$, which is called the *orthonormal frame bundle* over M. Obviously, $O_x(M)$ is isometric to $O(d)$, the group of orthogonal $(d \times d)$-matrices.

Let $\mathbf{p} : O(M) \to M$ with $\mathbf{p}u := x$ if $u \in O_x(M)$, which is called the *canonical projection* from $O(M)$ onto M. Now, given $e \in \mathbb{R}^d$, our goal is to define the corresponding horizontal vector field on $O(M)$. On the one hand, for any $u \in O(M)$ we have $ue \in T_{\mathbf{p}u} M$. Let u_s be the parallel transportation of u along the geodesic $\exp_{\mathbf{p}u}(sue)$, $s \geqslant 0$. We obtain a vector

$$H_e(u) := \frac{\mathrm{d}}{\mathrm{d}s} u_s|_{s=0} \in T_u O(M).$$

Thus, we have defined a vector field H_e on $O(M)$ which is indeed C^∞-smooth. In particular, let $\{e_i\}_{i=1}^d$ be an orthonormal basis on $\mathbb{R}^d$, define

$$\Delta_{O(M)} := \sum_{i=1}^d H_{e_i}^2.$$

This operator is independent of the choice of the basis $\{e_i\}$. We call $\Delta_{O(M)}$ *the horizontal Laplace operator*. On the other hand, for any vector field Z on M, we define its *horizontal lift* by $\mathbf{H}_Z(u) := H_{u^{-1}Z}(u)$, $u \in O(M)$, where $u^{-1}Z$ is the unique vector $e \in \mathbb{R}^d$ such that $Z_{\mathbf{p}u} = ue$.

Let Δ_M be the Laplace–Beltrami operator,

$$\Delta_M f = \frac{1}{\sqrt{G}} \frac{\partial}{\partial x^i} \left(\sqrt{G} g^{ij} \frac{\partial f}{\partial x^j} \right), \quad f \in C^2(M). \tag{3.3}$$

We have the conclusion below, the horizontal Laplacian $\Delta_{O(M)}$ is the lift of the Laplace–Beltrami operator Δ_M to the orthonormal frame bundle $O(M)$.

Lemma 3.3 (*Proposition 3.1.2 of [10]*). *Let $f \in C^\infty(M)$, and $\tilde{f} = f \circ \mathbf{p}$ its lift to $O(M)$. Then for any $u \in O(M)$,*

$$\Delta_M f(x) = \Delta_{O(M)} \tilde{f}(u),$$

where $x = \mathbf{p}u$.

Having the preparations of orthonormal frame bundle and horizontal lift, we can establish a diffusion process and (3.1) with switching (3.2) in detail.

Proof of Theorem 3.2. We divide the proof into two steps.

Step 1: A SDE with a fixed switching state.

Let $b : \mathbb{R}^d \to TM$ be a C^1-smooth vector field on M. According to the idea of [21, Section 2.1], we study a diffusion process generated by $A_n^M := \frac{1}{2n} \Delta_M + b$, where Δ_M is a Laplace–Beltrami operator in (3.3).

To this end, we first construct the corresponding *Horizontal diffusion process* generator by $A_n^{O(M)} := \frac{1}{n} \Delta_{O(M)} + \mathrm{H}_b$ on $O(M)$ by solving the Stratonovich stochastic differential equation

$$\mathrm{d}U_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^d H_{e_j}(U_n(t)) \circ \mathrm{d}W^j(t) + H_b(U_n(t)) \mathrm{d}t, \quad U_n(0) = u_0 \in O(M),$$

where $W(t) := (W^1(t), \ldots, W^d(t))$ is the d-dimensional Brownian motion on a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0}, \mathbb{P})$. Since H_b is C^1, it is well known that (see e.g. [7, Chapter IV, Section 6]) the equation has a unique solution up to the lifetime $\zeta := \lim_{j \to \infty} \zeta_j$, where

$$\zeta_j := \inf\{t \geqslant 0 : \mathrm{d}(\mathbf{p}U, \ \mathbf{p}U_n(t)) \geqslant j\}, \quad j \geqslant 1.$$

Using Assumption 3.1, we further get that

$$\mathbb{P}(\zeta = \infty) = 1,$$

which means that ξ is the infinite lifetime, see [10, Section 4.2].

Let $X_n(t) = \mathbf{p}U_n(t)$. Then $X_n(t)$ solves the equation

$$\mathrm{d}X_n(t) = \frac{1}{\sqrt{n}} U_n(t) \circ \mathrm{d}W(t) + b(X_n(t))\mathrm{d}t, \quad X_n(0) = x_0 := \mathbf{p}u_0 \tag{3.4}$$

up to the infinite lifetime ζ. By the Itô formula, for any $f(\cdot) \in C_0^2(M)$,

$$f(X_n(t)) - f(x_0) - \int_0^t A_n^M f(X_n(s))\mathrm{d}s = \frac{1}{\sqrt{n}} \int_0^t \langle (U_n(s))^{-1} \mathrm{grad} f(X_n(s)), \mathrm{d}W(s) \rangle$$

is a martingale up to the infinite lifetime ζ; that is $X_n(t)$ is the diffusion process generated by A_n^M, and we call it the A_n^M-diffusion process. When $b = 0$, then $X_n(t)$ is generated by $\frac{1}{2n}\Delta_M$ and is called the Brownian motion on M.

Step 2: the SDE with switching for any states. Here, we are going to introduce SDE with switching in (3.4). To achieve this, for $S = \{1, 2 \dots, N\}$ with $N < \infty$, we let the drift coefficient of the slow process depend on $i \in S$, where i represents the state of the switching process.

We construct the joint process as follows. Initialize the process from (x_0, k_0) and run the diffusion process $X_n(t)$ with $b = b(\cdot, k_0)$ in (3.4) as in Step 1. As this process has infinite lifetime, we can wait until the first switch as indicated by the jump kernel

$$\mathbb{P}(\Lambda_n^\varepsilon(t+\Delta) = j \mid \Lambda_n^\varepsilon(t) = i, X_n^\varepsilon(t) = x) = \frac{1}{\varepsilon} q_{ij}(x)\Delta + o(\Delta), \quad \text{if } j \neq i, \tag{3.5}$$

for small $\Delta > 0$, $i, j \in S$, $x \in M$, and $\varepsilon > 0$ is a parameter.

We then run (3.4) with the state to which the jump kernel points us to jump. As S is finite, we can repeat this process and obtain our desired switching process with infinite lifetime. □

3.1. The main results

In this paper, we consider the slow–fast systems (3.1) and (3.2). We first collect all the assumptions that are needed before giving the main results.

Assumption 3.4. Let $\varepsilon = \frac{1}{n}$, this shows that small disturbance and fast switching have the same rate.

This assumption means that the slow–fast system (3.1) and (3.2) becomes

$$\mathrm{d}X_n(t) = \frac{1}{\sqrt{n}} U_n(t) \circ \mathrm{d}W(t) + b(X_n(t), \Lambda_n(t))\mathrm{d}t, \tag{3.6}$$

and

$$\mathbb{P}(\Lambda_n^\varepsilon(t+\Delta) = j \mid \Lambda_n^\varepsilon(t) = i, X_n^\varepsilon(t) = x) = \frac{1}{\varepsilon} q_{ij}(x)\Delta + o(\Delta), \quad \text{if } j \neq i, \tag{3.7}$$

In the following, we will focus on (3.6) and (3.7).

Assumption 3.5. Fix $x_0 \in M$ and define $r(x) = d(x, x_0)$. We say that b is linear growth if there exists a constant $C > 0$ such that, for all $x \in M$,

$$|b(x,i)| \leqslant C(1 + r(x)), \quad \forall\, i \in S.$$

Assumption 3.6. We say that b is a locally one-sided Lipschitz function if for any compact set $K \subseteq M$, there exists a constant $C_K > 0$ such that, for all x, $y \in K$, it holds that

$$\mathrm{d}_x \left(\frac{1}{2} d^2(\cdot, y) \right)(x) b(x,i) - \mathrm{d}_y \left(-\frac{1}{2} d^2(x, \cdot) \right)(y) b(y,i) \leqslant C_K d^2(x,y), \quad \forall\, i \in S,$$

where $d(x,y) < i(K)$ and $i(K)$ is the injectivity radius of K defined in Appendix A.

Assumption 3.7. For any $x \in M$, $(q_{ij}(x))_{i,j \in S}$ is a conservative, irreducible transition rate matrix, and $\sup_{i \in S} \sum_{j \in S, j \neq i} q_{ij}(x) < \infty$.

Assumption 3.8. For any compact sets $K \subseteq M$, there exists a constant $C_K > 0$ such that

$$|q_{ij}(x) - q_{ij}(y)| \leqslant C_K d(x,y), \quad x,\ y \in K,\ i,\ j \in S.$$

Then, we give some remarks on these assumptions.

- Assumption 3.5 controls the rate at which the process may deviate to prove exponential tightness.
- Assumption 3.6 is set for proving the comparison principle.

- Assumptions 3.7 and 3.8 of a fast switching process for any given x ensures the existence of an invariant probability measure that satisfies the averaging principle.

In the following, we give the main result.

Theorem 3.9 (*Large Deviations for Slow Processes*). *Let $(X_n(t), \Lambda_n(t))$ be the Markov processes on $M \times S$. Consider the setting of Assumptions 3.4, 3.5, 3.6, 3.7 and 3.8. Suppose that the large deviation principle holds for $X_n(0)$ on M with speed n and a good rate function I_0.*

Then, the large deviation principle is satisfied with speed n for the processes $X_n(t)$ with a good rate function I having action-integral representation,

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s))\,\mathrm{d}s, & \text{if } \gamma \in \mathcal{AC}(M), \\ \infty, & \text{otherwise.} \end{cases}$$

where $\mathcal{L} : TM \to [0, \infty]$ is the Legendre transform of $\mathcal{H}$ given by $\mathcal{L}(x, v) = \sup_{p \in T_x^* M}\{\langle v, p\rangle - \mathcal{H}(x, p)\}$, and

$$\mathcal{H}(x, \mathrm{d}f(x)) = \sup_{\pi \in \mathcal{P}(S)} \left\{ \int B_{x,\mathrm{d}f(x)}(z)\pi(dz) - \mathcal{I}(x, \pi) \right\} \tag{3.8}$$

where

$$B_{x,\mathrm{d}f(x)}(z) = b(x, z)\mathrm{d}f(x) + \frac{1}{2}\,|\mathrm{d}f(x)|^2$$

coming from the slow process $X_n(t)$ and Donsker–Varadhan function

$$\mathcal{I}(x, \pi) = -\inf_{g>0} \int \frac{R_x g(z)}{g(z)} \pi(dz),$$

where R_x is the generator corresponding to the fast process $\Lambda_n(t)$ defined by

$$R_x g(z) = \sum_{j \in S} q_{zj}(x)\,(g(j) - g(z))\,.$$

4. The strategy of proof Theorem 3.9

In this section, we begin with the necessary preparation for the coming strategy. The semigroups of log-Laplace transforms of the conditional probabilities

$$V_n(t)f(x_0, k_0) = \frac{1}{n}\log\mathbb{E}[\mathrm{e}^{nf(X_n(t),\Lambda_n(t))} \mid (X_n(0), \Lambda_n(0)) = (x_0, k_0)] \tag{4.1}$$

have generators

$$H_n f = \frac{\mathrm{d}}{\mathrm{d}t} V_n(t) f\Big|_{t=0} = \frac{1}{n}\mathrm{e}^{-nf} A_n^M \mathrm{e}^{nf}. \tag{4.2}$$

Recall that $\mathcal{L} : TM \to [0, \infty]$, defined as

$$\mathcal{L}(x, v) = \sup_{p \in T_x^* M} \{\langle p, v\rangle - \mathcal{H}(x,\ p)\}.$$

This Lagrangian keeps track of the cost along a trajectory that will play a central role in the form of the rate function of the large deviation principle.

Then in terms of $\mathcal{L}$, we define a *variational semigroup* $\mathbf{V}(t)$, $t \geqslant 0$,

$$\mathbf{V}(t)f(x) := \sup_{\substack{\gamma \in \mathcal{AC} \\ \gamma(0)=x}} \left\{ f(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s \right\} \tag{4.3}$$

and *resolvent* $\mathbf{R}(\lambda)$, $\lambda > 0$,

$$\mathbf{R}(\lambda)h(x) := \sup_{\substack{\gamma \in \mathcal{AC} \\ \gamma(0)=x}} \left\{ \int_0^\infty \lambda^{-1} e^{-\lambda^{-1}t} \left(h(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(r), \dot{\gamma}(r))\mathrm{d}r \right) \mathrm{d}s \right\}. \tag{4.4}$$

Definition 4.1 (*Exponential Compact Containment*). We say that a process $(X_n(t), \Lambda_n(t))$ satisfies the *exponential compact containment condition* at speed n, for every all compact $K_0 \subseteq M$, $T > 0$ and $a \geqslant 0$, there exists a compact set $K_{a,T} \subseteq M$ such that

$$\limsup_{n\to\infty} \sup_{(x_0,k_0)\in K_0\times S} \frac{1}{n} \log \mathbb{P}\left[(X_n(t), \Lambda_n(t)) \notin K_{a,T} \times S \text{ for some } t \leqslant T \,\Big|\, (X_n(0), \Lambda_n(0)) = (x_0, k_0)\right] \leqslant -a.$$

Definition 4.2 (*Extended Limit, Definition A.12 in [8]*). For every $n \geqslant 1$, let $H_n \subset C_b(M) \times C_b(M)$ be an operator. The *extended limit* $\mathrm{ex} - \lim_{n\to\infty} H_n$ is defined as the collection $(f, g) \in C_b(M) \times C_b(M)$ for which there exists a sequence $\{f_n\}_{n\geqslant 1}$ with $f_n \in \mathcal{D}(H_n)$ such that

$$\lim_{n\to\infty} (\|f_n - f\| + \|H_n f_n - g\|) = 0.$$

An operator H is said to be contained in $\text{ex}-\lim_{n\to\infty} H_n$ if the graph $\{(f, Hf) | f \in \mathcal{D}(H)\}$ is a subset of $\text{ex}-\lim_{n\to\infty} H_n$.

We sketch our proof strategy that follows the projective limit approach:

(1) If the processes are exponentially tight, it suffices to establish the large deviation principle for finite dimensional distributions.
(2) The large deviation principle for finite dimensional distributions can be established by proving that the semigroup of log Laplace-transforms of the conditional probabilities converges to a limiting semigroup.
(3) One can often rewrite the limiting semigroup as a variational semigroup, which allows to rewrite the rate function on the Skorohod space in Lagrangian form.

In detail, the strategy to prove a path space large deviation principle for a sequence of the Markov processes $(X_n(t), \Lambda_n(t))$ on $M \times S$, formally works as follows:

(i) *Identification of a multi-valued limiting Hamiltonian*. The semigroups of log-Laplace transforms of the conditional probabilities $V_n(t)$ in (4.1) have generators H_n. We will verify that there is a operator $H \subseteq \text{ex}-\lim_{n\to\infty} H_n$. Due to our two-scale nature, this limiting operator will be multi-valued.
(ii) *Identification of a single value Hamiltonian*. Solving an eigenvalue problem for the fast dynamics at the level of the multi-valued operator H, we can construct a single valued operator $\mathcal{H}$ in terms of the variational object of (3.8).
(iii) *Exponential tightness on Riemannian manifold*. Provided one can verify the exponential compact containment condition in Definition 4.1, the convergence of the sequence H_n gives exponential tightness.
(iv) *Comparison principle on Riemannian manifold*. The theory of viscosity solutions gives applicable conditions for proving that the limiting Hamiltonian generates a semigroup. If for all $\lambda > 0$ and $h \in C_b(M)$, the Hamilton–Jacobi equation $f - \lambda H f = h$ admits a unique solution, one can extend the generator H so that the extension satisfies the conditions of Crandall–Liggett theorem [2] and thus generates a semigroup $V(t)$. Additionally, it follows that the semigroups $V_n(t)$ converge to $V(t)$, giving the large deviation principle. Uniqueness of the solution of the Hamilton–Jacobi–Bellman equation can be established via the comparison principle for sub- and super-solutions. The definitions of the comparison principle and viscosity solution can be found in Appendix B.
(v) *Variational representation of the Hamiltonian on Riemannian manifold*. By Legendre transforming the limiting Hamiltonian H, one can define a Lagrangian which can be used to define a variational semigroup and a variational resolvent. It can be shown that the variational resolvent provides a solution of the Hamilton–Jacobi–Bellman equation and therefore, by uniqueness of the solution, identifies the resolvent of H. As a consequence, an approximation procedure yields that the variational semigroup and the limiting semigroup $V(t)$ agree. A standard argument is then sufficient to give a Lagrangian form of the path-space rate function.

Feng and Kurtz summarized this in their book [8]. We modified it to fit our content.

Proposition 4.3 (*Adaptation of Theorem 5.15, Theorem 8.27 and Corollary 8.28 in [8] to Our Context*). *Let $(X_n(t), \Lambda_n(t))$ be Markov processes on $M \times S$. Suppose that*

(a) *$X_n(0)$ satisfies large deviation principle;*
(b) *there exists an operator $H \subset \text{ex}-\lim_{n\to\infty} H_n$ in the sense Definition 4.2;*
(c) *we have exponential compact containment of the process $(X_n(t), \Lambda_n(t))$;*
(d) *for all $\lambda > 0$ and $h \in C_b(M)$, the comparison principle holds for $f - \lambda H f = h$.*

Then the following hold:

(i) (Limit of nonlinear semigroup) There exists a unique operator semigroup $V(t)$ such that

$$\lim_{n\to\infty} \|V_n(t) f_n - V(t) f\| = 0 \tag{4.5}$$

and there exists a unique $R(\lambda) f$ such that

$$\lim_{m\to\infty} \|R(t/m)^m f - V(t) f\| = 0, \tag{4.6}$$

whenever $f \in \overline{\mathcal{D}(H)}$, $f_n \in C_b(M \times S)$, and $\|f_n - f\| \to 0$.

(ii) (Large deviation principle) $X_n(t)$ satisfies the large deviation principle with good rate function I given by

$$I(x) = I_0(x(t_0)) + \sup_{k\in\mathbb{N}} \sup_{0=t_0<t_1<\cdots<t_k<\infty} \sum_{i=0}^{k} I^V_{t_{i+1}-t_i}(x(t_{i+1}) \mid x(t_i)), \tag{4.7}$$

where for $\Delta t = t_{i+1} - t_i > 0$ and $x(t_{i+1}), x(t_i) \in M$, the conditional rate functions $I^V_{\Delta t}(x(t_{i+1}) \mid x(t_i))$ are

$$I^V_{\Delta t}(x(t_{i+1}) \mid x(t_i)) = \sup_{f\in C_b(M)} [f(x(t_{i+1})) - V(\Delta t) f(x(t_i))].$$

Suppose in addition that

(e) $V(t) = \mathbf{V}(t)$ *with* $\mathbf{V}$ *as in* (4.3).

Then the rate function (4.7) *can be represented in the following action-integral form:*

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s))ds, & \text{if } \gamma \in \mathcal{AC}(M), \\ \infty, & \text{otherwise.} \end{cases} \tag{4.8}$$

The proof of Theorem 3.9 is thus immediate upon checking Proposition 4.3(a) to (e) for our switching diffusion. We will verify (a) to (d) in Section 5 and (e) in Section 6.

5. The proof of Proposition 4.3(a) to (d)

In this section, will establish Proposition 4.3(a) to (d) for our switching diffusion:

Using the discussion from the previous section, we can prove items (i) and (ii) of Theorem 3.9 once the following four facts are established:

- Item (b): we obtaining a limiting multi-valued Hamiltonian $H \subseteq \text{ex} - \lim H_n$ in Section 5.1;
- Part of item (d): We identify a single valued Hamiltonian $\mathbf{H}$ via a suitable eigenvalue problem in Section 5.1.1;
- Item (c): we obtain the compact containment condition in Section 5.2;
- Part of (d): we prove the comparison principle for H and $\mathbf{H}$ in Section 5.3.

5.1. Identification of a multi-valued Hamiltonian

Our first goal is to obtain a multi-valued Hamiltonian $H \subseteq \text{ex} - \lim H_n$. We consider the solution $(X_n(t), \Lambda_n(t))$ of the system (3.6) and (3.7) with the generator A_n^M:

$$A_n^M f(x,i) = \frac{1}{2n}\Delta_M f(x,i) + b(x,i)\mathrm{d}f(x,i) + n\sum_{i\in S} q_{ij}(x)(f(x,j) - f(x,i)). \tag{5.1}$$

We give a multi-valued limit Hamiltonian by the generator A_n^M. Denote by $C_c^2(M)$ the set of twice continuously differentiable functions that are constant outside of a compact set.

Proposition 5.1 (*Multi-valued Limit Hamiltonian*). *Let* $(X_n(t), \Lambda_n(t))$ *be a Markov process on* $M \times S$ *with generator* A_n^M *in* (5.1). *Set* $H_n = \frac{1}{n} e^{nf} A_n^M e^{nf}$ *as in* (4.2) *and*

$$H := \left\{ (f, H_{f,\phi}) \mid f \in C_b^2(M), H_{f,\phi} \in C_b(M \times S) \text{ and } \phi \in C_b^2(M \times S) \right\}, \tag{5.2}$$

where

$$H_{f,\phi}(x,i) = b(x,i)\mathrm{d}f(x) + \frac{1}{2}|\mathrm{d}f(x)|^2 + \sum_{j\in S} q_{ij}(x)[e^{\phi(x,j)-\phi(x,i)} - 1]. \tag{5.3}$$

Then, $H \subset ex - \lim_{n\to\infty} H_n$.

Proof. By the generator A_n^M in (5.1), for $e^{nf} \in \mathcal{D}(A_n^M)$ we get a nonlinear generator

$$\begin{aligned} H_n f(x,i) &= \frac{1}{n} e^{-nf} A_n^M e^{nf}(x,i) \\ &= b(x,i)\mathrm{d}f(x,i) + \frac{1}{2}|\mathrm{d}f(x,i)|^2 + \frac{1}{2n}\Delta_M f(x,i) \\ &\quad + \sum_{j\in S} q_{ij}(x)[e^{n(f(x,j)-f(x,i))} - 1]. \end{aligned} \tag{5.4}$$

When $n \to \infty$, (5.4) is not convergent due to the divergence of the fourth term. To proceed, instead of using f in (5.4), we take a sequence

$$f_n(x,i) = f(x) + \frac{1}{n}\phi(x,i), \quad \forall\, f \in C_b^2(M) \text{ and } \phi \in C_b^2(M \times S).$$

As $\mathrm{d}f_n(x,i) = \mathrm{d}f(x) + \frac{1}{n}\mathrm{d}\phi(x,i)$, (5.4) implies

$$\begin{aligned} H_n f_n(x,i) &= b(x,i)\left(\mathrm{d}f(x) + \frac{1}{n}\mathrm{d}\phi(x,i)\right) + \frac{1}{2}\left|\mathrm{d}f(x) + \frac{1}{n}\mathrm{d}\phi(x,i)\right|^2 \\ &\quad + \frac{1}{2n}\Delta_M\left(f(x) + \frac{1}{n}\phi(x,i)\right) + \sum_{j\in S} q_{ij}(x)[e^{\phi(x,j)-\phi(x,i)} - 1]. \end{aligned}$$

Taking $n \to \infty$ gives the following uniform limit:

$$H_{f,\phi}(x,i) = b(x,i)\mathrm{d}f(x) + \frac{1}{2}|\mathrm{d}f(x)|^2 + \sum_{j\in S} q_{ij}(x)[e^{\phi(x,j)-\phi(x,i)} - 1],$$

establishing the claim. □

5.1.1. A single valued Hamiltonian via the eigenvalue problem

In the multi-valued operator H, we seek a single-valued operator that we will use to establish the comparison principle in Section 5.3 below. In particular, we aim to find for any $f \in \mathcal{D}(H)$ a unique g such that $(f,g) \in H$ and g does not depend on $i \in S$. This unique g will then be the basis to define $\mathbf{H}f$.

Consider (5.3) of Proposition 5.1:

$$H_{f,\phi}(x,i) = b(x,i)\mathrm{d}f(x) + \frac{1}{2}|\mathrm{d}f(x)|^2 + \sum_{j\in S} q_{ij}(x)[e^{\phi(x,j)-\phi(x,i)} - 1]. \tag{5.5}$$

As the eigenvalue problem is one in terms of the fast process, we decompose (5.5) into a function depending on i

$$B_{x,\mathrm{d}f(x)}(i) = b(x,i)\mathrm{d}f(x) + \frac{1}{2}|\mathrm{d}f(x)|^2 \tag{5.6}$$

and the jump operator acting on the state i:

$$R_x e^{\phi(x,i)} = \sum_{j\in S} q_{ij}(x)[e^{\phi(x,i)} - e^{\phi(x,j)}]$$

We thus seek a $\overline{\phi}$ such that there is a constant $\mathcal{H}(x,\mathrm{d}f(x))$ such that

$$\mathcal{H}(x,\mathrm{d}f(x)) := B_{x,\mathrm{d}f(x)}(i) + e^{-\overline{\phi}(x,i)} R_x e^{\overline{\phi}(x,i)} \tag{5.7}$$

is independent of i. Rewriting this equation in terms of $\overline{g} = e^{\overline{\phi}}$, we thus aim to find $\overline{g}$ and $\mathcal{H}(x,\mathrm{d}f(x))$ such that

$$\left(R_x + B_{x,\mathrm{d}f(x)}\right)\overline{g}(i) = \mathcal{H}(x,\mathrm{d}f(x))\overline{g}(i).$$

In other words, we aim to find the principal eigenfunction and eigenvalue for the operator $R_x + B_{x,\mathrm{d}f(x)}$ in terms of i, which can be carried out using the Perron–Frobenius theorem and leads to the representation (3.8).

Proposition 5.2 (*Principal-eigenvalue Problem*). *Let Assumption 3.7 be satisfied.*

For each $(x,\mathrm{d}f(x))$, there exist $\overline{g} > 0$ and a unique eigenvalue $\mathcal{H}(x,\mathrm{d}f(x)) \in \mathbb{R}$ such that

$$\left(R_x + B_{x,\mathrm{d}f(x)}\right)\overline{g} = \mathcal{H}(x,\mathrm{d}f(x))\overline{g}, \tag{5.8}$$

with $\mathcal{H}(x,\mathrm{d}f(x))$ given by

$$\begin{aligned}\mathcal{H}(x,\mathrm{d}f(x)) &= \sup_{\pi\in\mathcal{P}(S)} \inf_{g>0} \int \frac{\left(R_x + B_{x,\mathrm{d}f(x)}\right)g(i)}{g(i)}\pi(\mathrm{d}i) \\ &= \sup_{\pi\in\mathcal{P}(S)} \left\{ \int B_{x,\mathrm{d}f(x)}(i)\pi(\mathrm{d}i) - \mathcal{I}(x,\pi) \right\}\end{aligned} \tag{5.9}$$

where

$$\mathcal{I}(x,\pi) = -\inf_{g>0} \int \frac{R_x g(i)}{g(i)}\pi(\mathrm{d}i). \tag{5.10}$$

Proof. Using Assumption 3.7, from the Perron–Frobenius theorem in [6], we can obtain there exists a unique eigenvalue with associated eigenfunction which have the representation (5.9). □

We next aim to define a new operator in terms of $\mathcal{H}$. We first note the following result that can be obtained as in [4, Propositions 4.7 and 4.8].

Lemma 5.3. *The map $\mathcal{H}$ in (5.9) is continuous in (x,p) and convex in p for fixed x.*

As a direct consequence, we are able the introduce our single valued operator $\mathbf{H}$. Recall that $C_c^2(M)$ is the set of twice continuously differentiable functions that are constant outside of a compact set.

Definition 5.4. Set $\mathbf{H} \subseteq C_b(M)\times C_b(M)$ with domain $\mathcal{D}(\mathbf{H}) = C_c^2(M)$ by

$$\mathbf{H}f(x) := \mathcal{H}(x,\mathrm{d}f(x)) \tag{5.11}$$

with $\mathcal{H}$ as in (5.9).

5.2. Exponential compact containment

In this section, the key step, in obtaining exponential tightness on a Riemannian manifold, is to find a good containment function that can limit our analysis to a compact set.

Definition 5.5 (*Good Containment Function*). We say that $\Upsilon : M \to \mathbb{R}$ is a good containment function (for H) if

(Υa) $\Upsilon \geqslant 0$ and there exists a point x_0 such that $\Upsilon(x_0) = 0$,
(Υb) Υ is twice continuously differentiable,
(Υc) for every $c \geqslant 0$, the set $\{x \in M \mid \Upsilon(x) \leqslant c\}$ is compact,
(Υd) we have $\sup_x \mathcal{H}(x, \mathrm{d}\Upsilon(x)) < \infty$.

Let us denote by d the Riemannian distance function associated to the metric g. Fix $x_0 \in M$ and consider the radial function $r(x) = d(x, x_0)$. Since r is not everywhere smooth, it is not suitable for constructing a good containment function as in Definition 5.5. However, since r is 1-Lipschitz (with respect to the metric g), we can find a smooth function f with $f(x_0) = r(x_0) = 0$ and such that $\|f - r\| \leqslant 1$ and $|\mathrm{d}f| \leqslant 2$. Using this, we define Υ by

$$\Upsilon(x) = \frac{1}{2}\log(1 + f^2(x)). \tag{5.12}$$

We now show that Υ can be used as a good containment function. The following is an adaptation of [20, Proposition 4.11].

Lemma 5.6. *Let M be a complete Riemannian manifold. Under Assumption 3.5, Υ defined in (5.12) is a good containment function for the Hamiltonian $\mathcal{H}$ in (5.11).*

Proof. This proof is inspired by [14,20], and is therefore only different in checking property d. We prefer to spell out the proof of (Υa)–(Υc) as it will be used to prove (Υd).

Clearly $\Upsilon \geqslant 0$ and $\Upsilon(x_0) = 0$, and $\Upsilon \in C^\infty(M)$.

Furthermore, since r is smooth, it follows that Υ is smooth. Now, for $c > 0$, the continuity of Υ implies that $\{x \in M \mid \Upsilon(x) \leqslant c\}$ is closed. Furthermore, $\Upsilon(x) \leqslant c$ implies that $f(x) \leqslant \sqrt{e^{2c} - 1}$. It follows that $d(x, x_0) \leqslant 1 + f(x) \leqslant 1 + \sqrt{e^{2c} - 1}$. Hence, $\{x \in M \mid \Upsilon(x) \leqslant c\}$ is bounded. Since M is complete, we conclude that $\{x \in M \mid \Upsilon(x) \leqslant c\}$ is compact.

Note that for all $x \in M$,

$$\mathrm{d}\Upsilon(x) = \frac{f(x)}{1 + f^2(x)}\mathrm{d}f(x). \tag{5.13}$$

This, together with Assumption 3.5 and $|\mathrm{d}f| \leqslant 2$, for $z \in S$, we first estimate that

$$\begin{aligned} b(x,z)\mathrm{d}\Upsilon(x) &= b(x,z)\mathrm{d}f(x)\frac{f(x)}{1 + f^2(x)} \\ &\leqslant |b(x,z)| \cdot |\mathrm{d}f(x)| \cdot \frac{f(x)}{1 + f^2(x)} \\ &\leqslant C(2 + f(x))\frac{f(x)}{1 + f^2(x)}. \end{aligned} \tag{5.14}$$

Hence, $\sup_{x,z} b(x,z)\mathrm{d}\Upsilon(x) < \infty$. Now recall the Hamiltonian $\mathcal{H}$ in (5.11), from (5.14), we obtain

$$\begin{aligned} \mathcal{H}(x, \mathrm{d}\Upsilon(x)) &= \sup_{\pi \in \mathcal{P}(S)} \left\{ \int B_{x,\mathrm{d}\Upsilon(x)}(z)\pi(\mathrm{d}z) - \mathcal{I}(x,\pi) \right\} \\ &\leqslant \int B_{x,\mathrm{d}\Upsilon(x)}(z)\pi(\mathrm{d}z) \\ &= \int \left(b(x,z)\mathrm{d}\Upsilon(x) + \frac{1}{2}|\mathrm{d}\Upsilon(x)|^2 \right)\pi(\mathrm{d}z) \\ &\leqslant C \int \left(\frac{f(x)}{1 + f^2(x)} + \frac{f^2(x)}{1 + f^2(x)} + \frac{f^2(x)}{(1 + f^2(x))^2} \right)\pi(\mathrm{d}z), \end{aligned}$$

where the first inequality uses the definition of supremum and $\mathcal{I}(x,\pi)$ is nonnegative. We conclude that $\sup_x \mathcal{H}(x, \mathrm{d}\Upsilon(x)) < \infty$, which implies that Υ is a good containment function. □

Applying the good containment function (5.12), we proceed to consider the exponential compact containment of the system $(X_n(t), \Lambda_n(t))$.

Proposition 5.7 (*Exponential Compact Containment Condition*). *Let $(X_n(t), \Lambda_n(t))$ be a Markov process corresponding to A_n^M. Then the exponential compact containment condition as in Definition 4.1.*

The result follows using martingale control techniques as in [11] using Υ from Lemma 5.4.

5.3. Comparison principle

One of the key steps in the modern doubling of variables procedure in the comparison principle proofs is the estimate

$$\mathcal{H}\left(x_\alpha, \mathrm{d}_x \frac{\alpha}{2} d^2(\cdot, y_\alpha)\right)(x_\alpha) - \mathcal{H}\left(y, -\mathrm{d}_y \frac{\alpha}{2} d^2(x_\alpha, \cdot)\right)(y_\alpha) \leqslant \alpha C d^2(x_\alpha, y_\alpha), \tag{5.15}$$

for suitable x_α, y_α satisfying $\alpha d^2(x_\alpha, y_\alpha) \to 0$ as $\alpha \to \infty$.

In our case, the Hamiltonian is that

$$\mathcal{H}(x, \mathrm{d}f(x)) = \sup_{\pi \in \mathcal{P}(S)} \left\{ \int B_{x,\mathrm{d}f(x)}(z)\pi(\mathrm{d}z) - \mathcal{I}(x, \pi) \right\},$$

where

$$B_{x,\mathrm{d}f(x)}(z) = b(x, z)\mathrm{d}f(x) + \frac{1}{2} |\mathrm{d}f(x)|^2 . \tag{5.16}$$

In the proof of Lemma 5.18 below, we will pick an optimizer π^* for $\mathcal{H}\left(y, -\mathrm{d}_y \frac{\alpha}{2} d^2(x_\alpha, \cdot)\right)(y_\alpha)$, so that the estimate (5.15) reduces to

- the use of Assumption 3.6 to control the difference of the two terms that include the drift b;
- properties of the Riemannian metric d to threat the quadratic part $|df|^2$, see Lemma 5.8 below;
- estimates on objects of the type $|\mathcal{I}(x_\alpha, \pi^*) - \mathcal{I}(y_\alpha, \pi^*)|$, see Lemma 5.19.

A final issue arises from the fact that the metric d^2 is non-differentiable on the cut-locus, which we will treat by using that as $\alpha d^2(x_\alpha, y_\alpha) \to 0$, we will always work outside of the cut-locus.

5.3.1. Smooth distance functions

We first present the differential property of the distance function to deal with the quadratic part in (5.16), the proof is shown in [14, Appendix C.1].

Lemma 5.8. *Let $x, y \in M$ and assume that $x \notin cut(y)$ (or equivalently, $y \notin cut(x)$), where $cut(\cdot)$ is a cut-locus. Then for all $V \in T_y M$ we have*

$$\mathrm{d}_y(d^2(x, \cdot))(y)(V) = 2\langle \dot\gamma(1),\ V\rangle_{g(y)},$$

where $\gamma : [0, 1] \to M$ is the unique geodesic of minimal length connecting x and y. Consequently, we obtain

$$\tau_{x,y}\mathrm{d}_x(d^2(\cdot, y))(x) = -\mathrm{d}_y(d^2(x, \cdot))(y). \tag{5.17}$$

Remark 5.9. Note that (5.17) implies that if $x \notin \mathrm{cut}(y)$ (or equivalently, $y \notin \mathrm{cut}(x)$), we have

$$|\mathrm{d}_x(d^2(\cdot, y))(x)|^2_{g(x)} = |\mathrm{d}_y(d^2(x, \cdot))(y)|^2_{g(y)}$$

useful for estimating the quadratic part in the estimate (5.15).

Our approach to proving the comparison principle is to double variables, as in the classical setting of viscosity solutions in Euclidean spaces, using the distance function as a penalizing function.

Lemma 5.10 (*Lemma A.10 in [1]*). *Let u be bounded and upper semicontinuous, let v be bounded and lower semicontinuous, and let Υ be a good containment function as defined in (5.12).*

Fix $\delta > 0$. For every $m > 0$ there exist points $x_{\delta,m},\ y_{\delta,m} \in M$, such that

$$\begin{aligned} &\frac{u(x_{\delta,m})}{1-\delta} - \frac{v(y_{\delta,m})}{1+\delta} - \frac{m}{2} d^2(x_{\delta,m}, y_{\delta,m}) - \frac{\delta}{1-\delta}\Upsilon(x_{\delta,m}) - \frac{\delta}{1+\delta}\Upsilon(y_{\delta,m}) \\ &= \sup_{x,y\in M} \left\{ \frac{u(x)}{1-\delta} - \frac{v(y)}{1+\delta} - \frac{m}{2} d^2(x, y) - \frac{\delta}{1-\delta}\Upsilon(x) - \frac{\delta}{1+\delta}\Upsilon(y) \right\}. \end{aligned} \tag{5.18}$$

Additionally, for every $\delta > 0$ we have that

(a) *The set $\{x_{\delta,m},\ y_{\delta,m} \mid m > 0\}$ is relatively compact in M.*

(b) *All limit points of $\{(x_{\delta,m},\ y_{\delta,m})\}_{m>0}$ are of the form $(z,\ z)$ and for these limit points we have*

$$u(z) - v(z) = \sup_{x\in M} u(x) - v(x).$$

(c) *We have*

$$\lim_{m\to\infty} m d^2(x_{\delta,m},\ y_{\delta,m}) = 0.$$

In the proof of the comparison principle, Lemma 5.18, below, we will have to work with smooth test functions that are derived from the optimization procedure in (5.18) above. Due to the presence of the cut-locus, smoothness of $\frac{m}{2}d^2$ is, however, not guaranteed. For any fixed δ, we see that the injectivity radius $i(K) = \inf_{x\in K} i_x$ is bounded away from 0 on the compact K obtained in (a). Thus by (c) our optimizing values must lie in the complement of the cut-locus for large m. The next lemma allows us to replace d^2 by a smooth function behaving similarly outside the cut-locus.

Lemma 5.11. *For any compact set $K \subseteq M$, there is smooth function $\Psi_K : M^2 \to [0,\infty)$ satisfying*

$$\Psi_K(x,y) = \frac{1}{2}d^2(x,y) \qquad \text{if } d(x,y) \leqslant \frac{i(K)}{2},$$
$$\Psi_K(x,y) > \frac{1}{8}i(K)^2 \qquad \text{if } d(x,y) > \frac{i(K)}{2}.$$

The proof is similar to Lemma 7.7 of [14].

5.3.2. The necessary operators for proving comparison principle

To prove the comparison principle for the Hamilton–Jacobi equation in terms of H and relate it to the variational Hamiltonian $\mathbf{H}$ of Definition 5.4, we introduce two new pairs of Hamiltonians (H_1, H_2) and $(H_\dagger, H_\ddagger)$ that serve as natural upper and lower bounds for H and $\mathbf{H}$ respectively. These new Hamiltonians are both defined in terms of the containment function Υ of (5.12), which introduces unboundedness in our test functions, allowing us to work with optimizing points in the definition of viscosity sub and supersolutions.

Denote by $C_l^\infty(M)$ the set of smooth functions on M that have a lower bound and by $C_u^\infty(M)$ the set of smooth functions on M that have an upper bound. Denote $C_\Upsilon := \sup_{(x,i)\in M\times S} B_{x,\mathrm{d}\Upsilon(x)}(i) < \infty$.

Definition 5.12 (*Multi-valued Operators*)**.** Recall the definition of $H_{f,\phi}$ in (5.3).

- For $f \in C_l^\infty(M)$, $\delta \in (0,1)$ and $\phi \in C_b^2(M\times S)$. Set

$$f_1^\delta(x) := (1-\delta)f(x) + \delta\Upsilon(x),$$
$$H_{1,f,\phi}^\delta(x,i) := (1-\delta)H_{f,\phi}(x,i) + \delta C_\Upsilon,$$

  and set

$$H_1 := \left\{\left(f_1^\delta, H_{1,f,\phi}^\delta\right) \,\middle|\, f \in C_l^\infty(M), \delta \in (0,1),\ \phi \in C_b^2(M\times S)\right\}.$$

- For $f \in C_u^\infty(M)$, $\delta \in (0,1)$ and $\phi \in C_b^2(M\times S)$. Set

$$f_2^\delta(x) := (1+\delta)f(x) - \delta\Upsilon(x),$$
$$H_{2,f,\phi}^\delta(x,i) := (1+\delta)H_{f,\phi}(x,i) - \delta C_\Upsilon,$$

  and set

$$H_2 := \left\{\left(f_2^\delta, H_{2,f,\phi}^\delta\right) \,\middle|\, f \in C_u^\infty(M),\ \delta \in (0,1),\ \phi \in C_b^2(M\times S)\right\}.$$

We use the single valued Hamiltonian $\mathbf{H}$ to define two new single valued operators.

Definition 5.13 (*Single Valued Operators*)**.** Recall the definition of $\mathcal{H}(x, \mathrm{d}f(x))$ of (5.9).

- For $f \in C_l^\infty(M)$ and $\delta \in (0,1)$ set

$$f_\dagger^\delta(x) := (1-\delta)f(x) + \delta\Upsilon(x),$$
$$H_{\dagger,f}^\delta(x) := (1-\delta)\mathcal{H}(x,\mathrm{d}f(x)) + \delta C_\Upsilon,$$

  and set

$$H_\dagger := \left\{\left(f_\dagger^\delta, H_{\dagger,f}^\delta\right) \,\middle|\, f \in C_l^\infty(M),\ \delta \in (0,1)\right\}.$$

- For $f \in C_u^\infty(M)$ and $\delta \in (0,1)$ set

$$f_\ddagger^\delta(x) := (1+\delta)f(x) - \delta\Upsilon(x),$$
$$H_{\ddagger,f}^\delta(x) := (1+\delta)\mathcal{H}(x,\mathrm{d}f(x)) - \delta C_\Upsilon,$$

  and set

$$H_\ddagger := \left\{\left(f_\ddagger^\delta,\ H_{\ddagger,f}^\delta\right) \,\middle|\, f \in C_u^\infty(M),\ \delta \in (0,1)\right\}.$$

We collect H, $\mathbf{H}$, $H_\dagger$, $H_\ddagger$, H_1 and H_2 in Fig. 1, which intuitively provides the proof strategy for the comparison principle in the following subsection. Note that to obtain the comparison principle for H only the left-hand side of the figure is necessary. We aim to establish a variational expression for the rate function, however, by showing that $V(t) = \mathbf{V}(t)$. This we will carry out in Section 6 on which we will show that the variational resolvent will give viscosity solutions in terms of the Hamilton–Jacobi equation in terms of $\mathbf{H}$. The right-hand side of the figure will show that all viscosity solutions under consideration must be the same.

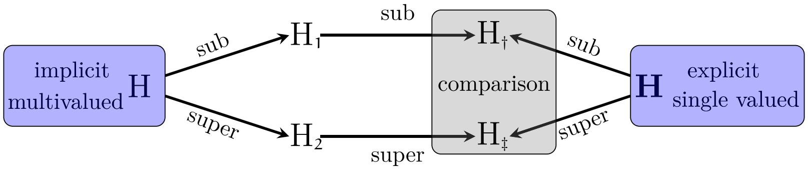

Fig. 1. An arrow connecting an operator A with operator B with subscript 'sub' means that viscosity subsolutions of $f - \lambda Af = h$ are also viscosity subsolutions of $f - \lambda Bf = h$. Similarly, we get the description for arrows with a subscript 'super'. The middle gray box around the operators $H_\dagger$ and $H_\ddagger$ indicates that the comparison principle holds for subsolutions of $f - \lambda H_\dagger f = h$ and supersolutions of $f - \lambda H_\ddagger f = h$. The left blue box indicates that H is an implicit and multi-valued operator. The right blue box indicates $\mathbf{H}$ is an explicit single valued operator.

5.3.3. Main propositions: comparison principle

Based on the above preparations, we are ready to state the proposition of this subsection.

Proposition 5.14 (*Comparison Principle*)**.** *Let Assumption 3.5, 3.6, 3.7 and 3.8 be satisfied. Let $h_1, h_2 \in C_b(M)$ and $\lambda > 0$. Let u be any subsolution to $f - \lambda Hf = h_1$ and let v be any supersolution to $f - \lambda Hf = h_2$. Then we have that*

$$\sup_{x\in M} u(x) - v(x) \leqslant \sup_x h_1(x) - h_2(x).$$

Proof. The result is immediate from Lemmas 5.15, 5.17, and 5.18 below. □

The proofs of the next three lemma's are analogous to those in [11].

Lemma 5.15. *Let Assumption 3.7 be satisfied. Fix $\lambda > 0$ and $h \in C_b(M)$.*

(a) *Every subsolution to $f - \lambda H_1 f = h$ is also a subsolution to $f - \lambda H_\dagger f = h$.*
(b) *Every supersolution to $f - \lambda H_1 f = h$ is also a supersolution to $f - \lambda H_\ddagger f = h$.*

Lemma 5.16. *Let Assumption 3.5 be satisfied. Fix $\lambda > 0$ and $h \in C_b(M)$.*

(a) *Every subsolution to $f - \lambda \mathbf{H} f = h$ is also a subsolution to $f - \lambda H_\dagger f = h$.*
(b) *Every supersolution to $f - \lambda \mathbf{H} f = h$ is also a supersolution to $f - \lambda H_\ddagger f = h$.*

Lemma 5.17. *Let Assumption 3.5 be satisfied. Fix $\lambda > 0$ and $h \in C_b(M)$.*

(a) *Every subsolution to $f - \lambda H f = h$ is also a subsolution to $f - \lambda H_1 f = h$.*
(b) *Every supersolution to $f - \lambda H f = h$ is also a supersolution to $f - \lambda H_2 f = h$.*

In addition to the lemmas above, we still need to verify the comparison principle for $f - \lambda H_\dagger f = h_1$ and $f - \lambda H_\ddagger f = h_2$ on M from Fig. 1.

Lemma 5.18. *Suppose Assumption 3.5, 3.6, 3.7 and 3.8 hold. Let $h_1, h_2 \in C_b(M)$ and $\lambda > 0$. Let u be any subsolution to $f - \lambda H_\dagger f = h_1$ and let v be any supersolution to $f - \lambda H_\ddagger f = h_2$. Then we have*

$$\sup_{x\in M} u(x) - v(x) \leqslant \sup_{x\in M} h_1(x) - h_2(x). \tag{5.19}$$

Proof. For a sub and supersolution u and v, $\delta \in (0,1)$ and $m \geqslant 1$, we follow (5.18) and set

$$\Phi_{\delta,m}(x,y) := \frac{u(x)}{1-\delta} - \frac{v(y)}{1+\delta} - \frac{m}{2} d^2(x,y) - \frac{\delta}{1-\delta}\Upsilon(x) - \frac{\delta}{1+\delta}\Upsilon(y), \tag{5.20}$$

By Lemma 5.10, we find a compact set K and $(x_{\delta,m}, y_{\delta,m}) \in K \times K$ satisfying

$$\Phi_{\delta,m}(x_{\delta,m},\ y_{\delta,m}) = \sup_{(x,y)\in M\times M} \Phi_{\delta,m}(x,\ y). \tag{5.21}$$

By Lemma 5.11, we can replace $\frac{m}{2}d^2$ by Ψ_K and consider

$$\widehat{\Phi}_{\delta,m}(x,y) := \frac{u(x)}{1-\delta} - \frac{v(y)}{1+\delta} - m\Psi_K(x,y) - \frac{\delta}{1-\delta}\Upsilon(x) - \frac{\delta}{1+\delta}\Upsilon(y). \tag{5.22}$$

It follows from (5.21) that for large m we have

$$\widehat{\Phi}_{\delta,m}(x_{\delta,m},\ y_{\delta,m}) = \sup_{(x,y)\in M\times M} \widehat{\Phi}_{\delta,m}(x,\ y). \tag{5.23}$$

In view of (5.23), it follows that $x_{\delta,m}$ is the unique maximizing point for

$$\sup_{x\in M} u(x) - \varphi_1^{\delta,m}(x) = u(x_{\delta,m}) - \varphi_1^{\delta,m}(x_{\delta,m})$$

where $\varphi_1^{\delta,m}$ is constructed by taking the appropriate remaining terms of (5.20), with an additional penalization $(1-\delta)d^2(x,x_{\delta,m})$ to turn $x_{\delta,m}$ into the unique optimizer:

$$\begin{aligned}
\varphi_1^{\delta,m}(x) :&= -(1-\delta)\Phi_{\delta,m}(x,y_{\delta,m}) + u(x) + (1-\delta)d^2(x,x_{\delta,m}) \\
&= (1-\delta)\left(-\frac{u(x)}{1-\delta} + \frac{v(y_{\delta,m})}{1+\delta} + m\Psi_K(x,y_{\delta,m}) + \ frac\delta 1-\delta \Upsilon(x) + \frac{\delta}{1+\delta}\Upsilon(y_{\delta,m})\right) \\
&\quad + u(x) + (1-\delta)d^2(x,x_{\delta,m}) \\
&= (1-\delta)\left(\frac{v(y_{\delta,m})}{1+\delta} + m\Psi_K(x,y_{\delta,m}) + \frac{\delta}{1-\delta}\Upsilon(x) + \frac{\delta}{1+\delta}\Upsilon(y_{\delta,m})\right) \\
&\quad + (1-\delta)d^2(x,x_{\delta,m}) \\
&= (1-\delta)\left(m\Psi_K(x,y_{\delta,m}) + d^2(x,x_{\delta,m}) + \frac{\delta}{1+\delta}\Upsilon(y_{\delta,m}) + \frac{v(y_{\delta,m})}{1+\delta}\right) + \delta\Upsilon(x).
\end{aligned}$$

Since u is a viscosity subsolution of $f - \lambda H_\dagger f = h_1$ we conclude that

$$u(x_{\delta,m}) - \lambda\left[(1-\delta)\mathcal{H}(x_{\delta,m},p^1_{\delta,m}) + \delta C_\Upsilon\right] \leqslant h_1(x_{\delta,m}), \tag{5.24}$$

where for large m

$$p^1_{\delta,m} := m d_x\Psi_K(\cdot,y_{\delta,m})(x_{\delta,m}) = m\ \mathrm{d}_x\left(\frac{1}{2}d^2(\cdot,y_{\delta,m})\right)(x_{\delta,m}). \tag{5.25}$$

Similarly, we obtain that $y_{\delta,m}$ it the unique optimizer for

$$\inf_{y\in M} v(x) - \varphi_2^{\delta,m}(y) = v(y_{\delta,m}) - \varphi_2^{\delta,m}(y_{\delta,m}),$$

where

$$\varphi_2^{\delta,m}(y) := (1+\delta)\left(m\Psi_K(x_{\delta,m},y) - d^2(y,y_{\delta,m}) - \frac{\delta}{1-\delta}\Upsilon(x_{\delta,m}) + \frac{u(x_{\delta,m})}{1-\delta}\right) - \delta\Upsilon(y).$$

As v is a viscosity supersolution of $f - \lambda H_\ddagger f = h_2$, we then know that

$$v(x_{\delta,m}) - \lambda\left[(1+\delta)\mathcal{H}(y_{\delta,m},p^2_{\delta,m}) - \delta C_\Upsilon\right] \geqslant h_2(y_{\delta,m}), \tag{5.26}$$

where for large m

$$p^2_{\delta,m} := -m\ \mathrm{d}_y\left(\frac{1}{2}d^2(x_{\delta,m},\cdot)\right)(y_{\delta,m}). \tag{5.27}$$

By item (c) of Lemma 5.10, we have

$$\lim_{m\to\infty} md^2(x_{\delta,m},y_{\delta,m}) = 0. \tag{5.28}$$

Taking (5.24), (5.26) and (5.28) into account, we obtain that

$$\begin{aligned}
\sup_{x\in M} u(x) - v(x) &\leqslant \liminf_{\delta\to 0}\liminf_{m\to\infty}\left(\frac{u(x_{\delta,m})}{1-\delta} - \frac{v(y_{\delta,m})}{1+\delta}\right) \\
&\leqslant \liminf_{\delta\to 0}\liminf_{m\to\infty}\left\{\frac{h_1(x_{\delta,m})}{1-\delta} - \frac{h_2(y_{\delta,m})}{1+\delta}\right. &(5.29)\\
&\quad + \frac{\delta}{1-\delta}C_\Upsilon + \frac{\delta}{1+\delta}C_\Upsilon &(5.30)\\
&\quad \left. + \lambda\big(\mathcal{H}(x_{\delta,m},p^1_{\delta,m}) - \mathcal{H}(y_{\delta,m},p^2_{\delta,m})\big)\right\}, &(5.31)
\end{aligned}$$

where in the first inequality we use (5.21) and drop the nonnegative functions $d^2(\cdot,\cdot)$ and $\Upsilon(\cdot)$.

The term (5.30) vanishes as $\delta\to 0$. For the term (5.29), the sequence $(x_{\delta,m},y_{\delta,m})$ takes its values in a compact set and, hence, admits converging subsequences as $m\to\infty$. By (b) of Lemma 5.10, these subsequences converge to points of the form (x,x). Hence, by the above analysis, we get

$$\begin{aligned}
\sup_{x\in M} u(x) - v(x) &\leqslant \lambda\liminf_{\delta\to 0}\liminf_{m\to\infty}\big(\mathcal{H}(x_{\delta,m},p^1_{\delta,m}) - \mathcal{H}(y_{\delta,m},p^2_{\delta,m})\big) \\
&\quad + \sup_{x\in M} h_1(x) - h_2(x).
\end{aligned}$$

It follows that the comparison principle holds for $f - \lambda H_\dagger f = h_1$ and $f - \lambda H_\ddagger f = h_2$ whenever for any $\delta > 0$

$$\liminf_{m\to\infty}\big(\mathcal{H}(x_{\delta,m},p^1_{\delta,m}) - \mathcal{H}(y_{\delta,m},p^2_{\delta,m})\big) \leqslant 0. \tag{5.32}$$

To that end, recall $\mathcal{H}(x, \mathrm{d}f(x))$ in (5.11):

$$\mathcal{H}(x, \mathrm{d}f(x)) = \sup_{\pi \in \mathcal{P}(S)} \left\{ \int B_{x,\mathrm{d}f(x)}(z)\pi(\mathrm{d}z) - \mathcal{I}(x,\pi) \right\},$$

where $\pi \mapsto \int B_{x,\mathrm{d}f(x)}(z)\pi(\mathrm{d}z)$ is bounded and continuous, and $\mathcal{I}(x,\cdot)$ has compact sub-level sets in $\mathcal{P}(S)$. Thus, there exists an optimizer $\pi_{\delta,m} \in \mathcal{P}(S)$ such that

$$\mathcal{H}(x_{\delta,m}, p^1_{\delta,m}) = \int B_{x_{\delta,m},p^1_{\delta,m}}(z)\pi_{\delta,m}(\mathrm{d}z) - \mathcal{I}(x_{\delta,m}, \pi_{\delta,m}) \tag{5.33}$$

and

$$\mathcal{H}(y_{\delta,m}, p^2_{\delta,m}) \geqslant \int B_{y_{\delta,m},p^2_{\delta,m}}(z)\pi_{\delta,m}(\mathrm{d}z) - \mathcal{I}(y_{\delta,m}, \pi_{\delta,m}). \tag{5.34}$$

Combining (5.33) and (5.34), we obtain

$$\begin{aligned} &\mathcal{H}(x_{\delta,m}, p^1_{\delta,m}) - \mathcal{H}(y_{\delta,m}, p^2_{\delta,m}) \\ &\leqslant \int \big(B_{x_{\delta,m},p^1_{\delta,m}}(z) - B_{y_{\delta,m},p^2_{\delta,m}}(z)\big)\pi_{\delta,m}(\mathrm{d}z) \qquad (5.35)\\ &\quad + \mathcal{I}(y_{\delta,m}, \pi_{\delta,m}) - \mathcal{I}(x_{\delta,m}, \pi_{\delta,m}). \qquad (5.36)\end{aligned}$$

It is enough to prove that (5.35) and (5.36) go to 0 as $m \to \infty$. For (5.35), by calculating the difference of integrand $B_{x,p}$ in detail, for any $z \in S$, from (5.6), (5.25), (5.27) and Remark 5.9, one has

$$\begin{aligned} &B_{x_{\delta,m},p^1_{\delta,m}}(z) - B_{y_{\delta,m},p^2_{\delta,m}}(z) \\ &= m\mathrm{d}_x\left(\frac{1}{2}d^2(\cdot, y_{\delta,m})\right)(x_{\delta,m})b(x_{\delta,m}, z) + \frac{1}{2}\left|m\ \mathrm{d}_x\left(\frac{1}{2}d^2(\cdot, y_{\delta,m})\right)(x_{\delta,m})\right|^2 \\ &- \left[-m\ \mathrm{d}_y\left(\frac{1}{2}d^2(x_{\delta,m}, \cdot)\right)(y_{\delta,m})b(y_{\delta,m}, z) + \frac{1}{2}\left|-m\ \mathrm{d}_y\left(\frac{1}{2}d^2(x_{\delta,m}, \cdot)\right)(y_{\delta,m})\right|^2\right] \\ &= m\mathrm{d}_x\left(\frac{1}{2}d^2(\cdot, y_{\delta,m})\right)(x_{\delta,m})b(x_{\delta,m}, z) + m\ \mathrm{d}_y\left(\frac{1}{2}d^2(x_{\delta,m}, \cdot)\right)(y_{\delta,m})b(y_{\delta,m}, z) \\ &+ \frac{m^2}{2}\left(\left|\mathrm{d}_x\left(\frac{1}{2}d^2(\cdot, y_{\delta,m})\right)(x_{\delta,m})\right|^2 - \left|-\mathrm{d}_y\left(\frac{1}{2}d^2(x_{\delta,m}, \cdot)\right)(y_{\delta,m})\right|^2\right) \\ &= m\mathrm{d}_x\left(\frac{1}{2}d^2(\cdot, y_{\delta,m})\right)(x_{\delta,m})b(x_{\delta,m}, z) + m\ \mathrm{d}_y\left(\frac{1}{2}d^2(x_{\delta,m}, \cdot)\right)(y_{\delta,m})b(y_{\delta,m}, z) \\ &\leqslant Cd^2(x_{\delta,m}, y_{\delta,m}), \end{aligned} \tag{5.37}$$

where in the last inequality, we use Assumption 3.6. Noting that the last term in line 5 vanishes. This is happened because, fix $\delta > 0$, there is a compact $K^\delta \subseteq M$ such that $\{x_{m,\delta},\ y_{m,\delta} \mid m > 0\}$ is contained in K^δ by item (a) of Lemma 5.10. By the continuity of the injectivity radius and the compactness of K^δ, we can find a $\Delta > 0$ such that $i(K^\delta) \geqslant \Delta > 0$. Then there exists a unique geodesic of minimal length connecting $x_{m,\delta}$ and $y_{m,\delta}$. Furthermore, by Lemma 5.8 we have

$$\mathrm{d}_x d^2(\cdot, y_{m,\delta})(x_{m,\delta}) = -\tau_{x_{m,\delta},\ y_{m,\delta}}\mathrm{d}_y d^2(x_{m,\delta},\ \cdot)(y_{m,\delta}), \tag{5.38}$$

where $\tau_{x_{m,\delta},y_{m,\delta}}$ denotes parallel transport along the unique geodesic of minimal length connecting $x_{m,\delta}$ and $y_{m,\delta}$. As parallel transport is an isometry, we find as in Remark 5.9 that

$$\left|\mathrm{d}_x\left(\frac{1}{2}d^2(\cdot, y_{\delta,m})\right)(x_{\delta,m})\right|^2_{g(x_{\delta,m})} = \left|-\mathrm{d}_y\left(\frac{1}{2}d^2(x_{\delta,m}, \cdot)\right)(y_{\delta,m})\right|^2_{g(y_{\delta,m})}$$

Hence, (5.35) is sufficiently small, as $m \to \infty$, using (5.28) and (5.37). To obtain that (5.36) is sufficiently small, we utilize the equi-continuity of $\mathcal{I}(\cdot,\pi)$ established in Lemma 5.19 below for the spatial variable. This finishes the proof of (5.32) and the comparison principle for $H_\dagger$ and $H_\ddagger$. □

Here, we state the equi-continuity of $\mathcal{I}(\cdot,\pi)$ to finish the proof of the comparison principle of $H_\dagger$ and $H_\ddagger$ in Lemma 5.18. The proof is analogous to Lemma 6.11 in [11].

Lemma 5.19. *Let Assumption 3.8 be satisfied. Recall (5.10):*

$$\mathcal{I}(x,\pi) = -\inf_{g>0}\int \frac{R_x g(z)}{g(z)}\pi(\mathrm{d}z).$$

For any compact set $K \subseteq M$ and for all $\pi \in \mathcal{P}(S)$, then $\{x \mapsto J(x,\pi)\}_{x\in K,\pi\in\mathcal{P}(S)}$ is equi-continuous.

6. The proof of Proposition 4.3(e)

In this final chapter, we will establish (e) of Proposition 4.3, which is the key statement to obtain the variational representation of the rate function in Theorem 3.9.

The proof is based on the analysis of variational semigroups and resolvents of Chapter 8 in [8] and are based on their main Conditions 8.9, 8.10 and 8.11 of [8], which we adapt to the Riemannian context below as Condition 6.1 and 6.2.

We will then carry out two main steps.

- We will show in Section 6.1 which key results of [8, Chapter 8] are used to obtain $V(t) = \mathbf{V}(t)$, and how this relates to our set-up in Section 5.
- We verify in Sections 6.2 and 6.3 Conditions 6.1 and 6.2 respectively in our context.

Condition 6.1. Suppose that the map $T^*M \ni (x,p) \to \mathcal{H}(x,p) \in \mathbb{R}$ is continuous, and is convex in the second variable p. Define $\mathcal{L}$ as its Legendre transform. Suppose that there is a good containment function Υ for $\mathcal{H}$. Then

(a) the function $\mathcal{L} : TM \to [0,\infty]$ is lower semi-continuous and for each compact set $K \subseteq M$ and $c \in \mathbb{R}$ the set

$$\{(x,v) \in TM \mid x \in K,\ \mathcal{L}(x,v) \leqslant c\}$$

is compact in TM.

(b) for each compact $K \subseteq M$, any finite time $T > 0$ and finite bound $C \geqslant 0$, there exists a compact set $\hat{K} = \hat{K}(K,T,C) \subseteq M$ such that $x \in \mathcal{AC}(M)$ and $x(0) \in K$, if

$$\int_0^T \mathcal{L}(x(s),\dot{x}(s))\mathrm{d}s \leqslant C,$$

then $x(t) \in \hat{K}$ for all $0 \leqslant t \leqslant T$.

(c) for each compact set $K \subseteq M$ and $c \in \mathbb{R}$, there exists a right-continuous non-decreasing function $\psi_{K,c} : \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\lim_{r\to\infty} r^{-1}\psi_{K,c}(r) = 0. \tag{6.1}$$

and

$$|\mathrm{d}f(x)v| \leqslant \psi_{K,c}(\mathcal{L}(x,v)), \qquad \forall (x,v) \in TM,\ x \in K,$$

where $f \in C_{K,c}$ and

$$C_{K,c} := \left\{ f \in C_c^2(M) \,\middle|\, \forall x \in K,\ |df(x)| \leqslant c \right\}. \tag{6.2}$$

Condition 6.2. For any initial point $x(0) \in M$, $T > 0$ and $f \in \mathcal{D}(\mathbf{H})$, there exists an absolutely continuous curve $x : [0,T] \to M$ such that for all $0 < t \leqslant T$

$$\int_0^t \mathcal{H}(x(s),\mathrm{d}f(x(s)))\mathrm{d}s + \int_0^t \mathcal{L}(x(s),\dot{x}(s))\mathrm{d}s = \int_0^t \mathrm{d}f(x(s))\dot{x}(s)\mathrm{d}s. \tag{6.3}$$

6.1. Connecting Conditions 6.1 and 6.2 to Section 5

In this section, we state two results of [8] and show how these can be used to obtain $V(t) = \mathbf{V}(t)$. For readability, we repeat the definitions of $\mathbf{V}$ and $\mathbf{R}$ of (4.3) and (4.4):

$$\mathbf{V}(t)f(x) := \sup_{\substack{\gamma\in\mathcal{AC}\\ \gamma(0)=x}} \left\{ f(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(s),\dot{\gamma}(s))\mathrm{d}s \right\} \tag{6.4}$$

and

$$\mathbf{R}(\lambda)h(x) := \sup_{\substack{\gamma\in\mathcal{AC}\\ \gamma(0)=x}} \left\{ \int_0^\infty \lambda^{-1}e^{-\lambda^{-1}t} \left(h(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(r),\dot{\gamma}(r))\mathrm{d}r \right) \mathrm{d}s \right\}. \tag{6.5}$$

Proposition 6.3 (*Lemma 8.18 of [8]*). *Let Conditions 6.1 and 6.2 be satisfied.*

For any $f \in C_b(M)$, $t \geqslant 0$ and $x \in M$, we have

$$\lim_{m\to\infty} |\mathbf{R}(t/m)^m f(x) - \mathbf{V}(t)f(x)| = 0.$$

The next result of [8], obtainable from the proof of Theorem 8.27 in [8], establishes that the variational resolvent gives viscosity solutions for the operator $\mathbf{H}$.

Proposition 6.4. *Let Conditions 6.1 and 6.2 be satisfied. Then we have*

- *$\mathbf{R}(\lambda)h$ is a viscosity subsolution to $f - \lambda\mathbf{H}f = h$,*
- *the lower semi-continuous regularization $(\mathbf{R}(\lambda)h)_*$ of $\mathbf{R}(\lambda)h$ is a viscosity supersolution to $f - \lambda\mathbf{H}f = h$.*

Combining these two statements with Proposition 4.3(i), it follows that Proposition 4.3(e), namely that $V(t) = \mathbf{V}(t)$, is satisfied if $R(\lambda) = \mathbf{R}(\lambda)$. Using the results of Section 5, we thus obtain the following result:

Proposition 6.5. *Let Conditions 6.1 and 6.2 be satisfied. Then $V(t)f = \mathbf{V}(t)$.*

Proof. By Proposition 6.4, Lemmas 5.16 and 5.18, $\mathbf{R}(\lambda)h$ equals the unique viscosity solution to the pair of equations

$$f - \lambda H_{\dagger} f = h, \qquad f - \lambda H_{\ddagger} f = h,$$

and thus equals $R(\lambda)h$ from Proposition 4.3. By Propositions 6.3 and 4.3(i), it follows that $V(t) = \mathbf{V}(t)$ establishing the claim. □

We are left to prove Conditions 6.1 and 6.2 in the following two sections.

6.2. Verification of Condition 6.1

In this section, we verify Condition 6.1.

Proposition 6.6. *Let Assumption 3.5 be satisfied. Then Condition 6.1 holds.*

Proof. To obtain Item (a), observe that $\mathcal{L} \geqslant 0$ follows from $\mathcal{H}(x,0) = 0$. The Lagrangian $\mathcal{L}$ is convex, and lower semicontinuous as it is the Legendre transform of $\mathcal{H}$. For $C \geqslant 0$, we prove that the set $\{(x,v) \in TM : x \in K, \mathcal{L}(x,v) \leqslant C\}$ is bounded, and hence is relatively compact. For any $p \in T_x^* M$ and $v \in T_x M$, we have

$$pv \leqslant \mathcal{L}(x,v) + \mathcal{H}(x,p) \quad x \in K.$$

Thereby, if $\mathcal{L}(x,v) \leqslant C$, then

$$|v| = \sup_{|p|=1} pv \leqslant \sup_{|p|=1} [\mathcal{L}(x,v) + \mathcal{H}(x,p)] \leqslant C + C_1,$$

where C_1 exists due to continuity of $\mathcal{H}$ obtained in Lemma 5.3 and $x \in K$. Then for $R := C + C_1$,

$$\{(x,v) \in TM : \mathcal{L}(x,v) \leqslant C\} \subseteq \{v : |v| \leqslant R\},$$

thus $\{\mathcal{L} \leqslant C\}$ is a bounded subset in TM.

For (b), recalling that by Assumption 3.5 and Lemma 5.6 the level sets of Υ are compact and we control the growth of Υ. For $K \subseteq M$, $T > 0$, $C \geqslant 0$ and $x \in \mathcal{AC}(M)$ as above, this follows by noting that

$$\begin{aligned}\Upsilon(x(t)) &= \Upsilon(x(0)) + \int_0^t \mathrm{d}\Upsilon(x(s))\dot{x}(s)\mathrm{d}s \\ &\leqslant \Upsilon(x(0)) + \int_0^t \left[\mathcal{L}(x(s),\dot{x}(s)) + \mathcal{H}(x(s),\mathrm{d}\Upsilon(x(s)))\right] \mathrm{d}s \\ &\leqslant \sup_{y\in K} \Upsilon(y) + C_1 + T \sup_{z \in M} \mathcal{H}(z, \mathrm{d}\Upsilon(z)) = C < \infty,\end{aligned}$$

for any $0 \leqslant t \leqslant T$, so that the compact set $\hat{K} = \{z \in M : \Upsilon(x) \leqslant C\}$ satisfies the condition.

Proof of (c) is inspired by that of Lemma 10.21 of [8]. We first prove that $\mathcal{L}(x,v)$ is superlinear. Recall that by Lemma 5.3 $\mathcal{H}$ is continuous, which implies

$$\overline{H}_K(c) := \sup_{x\in K} \sup_{p \in T_x^* M, |p|\leqslant c} \mathcal{H}(x,p) < \infty.$$

Using the definition of $\mathcal{L}$, it thus follows for any $(x,v) \in TM$, $x \in K$ with $|v| > 0$ that

$$\frac{\mathcal{L}(x,v)}{|v|} \geqslant \sup_{p \in T_x^* M, |p| \leqslant c} \frac{pv}{|v|} - \frac{\overline{H}_K(c)}{|v|} = c - \frac{\overline{H}_K(c)}{|v|}$$

It follows that

$$\lim_{N\uparrow\infty} \inf_{x\in K} \inf_{v\in T_x M : |v|=N} \frac{\mathcal{L}(x,v)}{|v|} = \infty.$$

Secondly, for $s \geqslant 0$, define the map $\vartheta(s)$ by

$$\vartheta(s) := s \inf_{x\in K} \inf_{v \in T_x M : |v| \geqslant s} \frac{\mathcal{L}(x,v)}{|v|}. \tag{6.6}$$

It thus follows that ϑ is a strictly increasing function satisfying

$$\lim_{s\uparrow\infty} \frac{\vartheta(s)}{s} = \infty. \tag{6.7}$$

Next, define $\Psi_{K,c}(r) =: C_{K,c}\vartheta^{-1}(r)$ with $\vartheta^{-1}(r) = \inf\{\omega : \vartheta(\omega) \geqslant r\}$. By monotonicity of ϑ, we have for any $x \in K$ that

$$\vartheta(C_{K,c}^{-1}|\mathrm{d}f(x)v|) \overset{(6.2)}{\leqslant} \vartheta(|v|) \overset{(6.6)}{\leqslant} \mathcal{L}(x,v).$$

Hence by monotonicity of $\Psi_{K,c}$, we find $|\mathrm{d}f(x)v| \leqslant \Psi_{K,c}(\mathcal{L}(x,v))$ for any $f \in C_{K,c}$, and $(x,v) \in TM$ with $x \in K$. Finally (6.1) follows by (6.7) and the definition of $\vartheta^{-1}(r)$. □

6.3. Verification of Condition 6.2

In this section, we verify Condition 6.2: the construction of curve with arbitrary lifetime, starting point and $f \in \mathcal{D}(\mathbf{H}) = C_c^2(M)$ satisfying

$$\int_0^t \mathcal{H}(x(s), \mathrm{d}f(x(s)))\mathrm{d}s + \int_0^t \mathcal{L}(x(s), \dot{x}(s))\mathrm{d}s = \int_0^t \mathrm{d}f(x(s))\dot{x}(s)\mathrm{d}s. \tag{6.8}$$

Proposition 6.7. *Let Assumption 3.5 be satisfied. Then Condition 6.2 holds.*

The key insight in Lemma 6.8 below is that, when working on local coordinate charts, the problem can be transferred to Euclidean space. Solutions can then be found via convex analysis and differential inclusion theory, see e.g. [3, Lemma 5.1]. Transferring back the solution to the manifold leads to locally defined solutions of (6.8). We perform this analysis in Section 6.3.1 below.

As usual, the problem thus resides into patching these curves together to form a curve of arbitrary length. For this we need to control the time of existence for our local solutions. We do so in multiple steps. Fix some time interval $[0, T]$ for which we aim to construct our curve.

- In Lemma 6.11, we will show that for any $T > 0$ and any curve satisfying (6.8) the curve remains in a compact set $\widehat{K}$ up to time T. We can thus construct curves locally on sets that have a radius that is lower bounded by the injectivity radius $i(\widehat{K}) = \inf_{x\in\widehat{K}} i_x > 0$.
- Given any such locally constructed curve, we control the Lagrangian linearly in time in Lemma 6.12
- Using this linear Lagrangian growth, we show in Lemma 6.13 that the squared distance to the starting point of the curve grows at most linearly. As the size of the ball is controlled by the injectivity radius on $\widehat{K}$ it follows that there is a lower bound on the interval of existence of the locally constructed curve.

Based on these three steps, we conclude that we can construct solutions to (6.8) on arbitrarily sized intervals $[0, T]$.

6.3.1. Local construction of solutions

In the first result, we show how the various quantities in (6.8) transfer from M to a local coordinate chart. This result is essentially a write-up of basic Riemannian coordinate transformations acting on $\mathcal{L}$ and $\mathcal{H}$. We write it down for an arbitrary smooth invertible map from a subset of a manifold M to a subset of a manifold N.

Lemma 6.8. *Let M be a Riemannian manifold. For an invertible smooth map $\varphi : \mathcal{O} \subseteq M \to \varphi(\mathcal{O}) := N$, via push-forward and pullback in Appendix A define*

$$\mathcal{H}_\varphi := \mathcal{H} \circ \varphi^* : T^*N \to \mathbb{R}$$

and

$$\mathcal{L}_\varphi := \mathcal{L} \circ \varphi_*^{-1} : TN \to \mathbb{R},$$

where $\mathcal{H} : T^*M \to \mathbb{R}$ and $\mathcal{L} : TM \to \mathbb{R}$. Define $f_\varphi = f \circ \varphi^{-1}$. Let $x : [0, T] \to \mathcal{O}$, suppose that $y(s) = \varphi(x(s)) : [0, T] \to \varphi(\mathcal{O})$, then we have that

(a) $\mathrm{d}f_\varphi(y(s))\dot{y}(s) = \mathrm{d}f(x(s))\dot{x}(s)$,
(b) $\mathcal{H}_\varphi(y(s), \mathrm{d}f_\varphi(y(s))) = \mathcal{H}(x(s), \mathrm{d}f(x(s)))$,
(c) $\mathcal{L}_\varphi(y(s), \dot{y}(s)) = \mathcal{L}(x(s), \dot{x}(s))$.
(d) *$\mathcal{L}_\varphi$ is the Legendre transform of $\mathcal{H}_\varphi$, i.e., $\mathcal{H}_\varphi(\eta, \xi) = \sup_{w\in T_\eta N}\left\{\xi(w) - \mathcal{L}_\varphi(\eta, w)\right\}$, for any $(\eta, \xi) \in T^*N$.*

Proof. We start to prove item (a). By Lemma A.6, there exists a curve $x(s)$ on $\mathcal{O}$ such that

$$\begin{aligned} \mathrm{d}f_\varphi(y(s))\dot{y}(s) &= \mathrm{d}f_\varphi(\varphi(x(s)))\varphi_*(\dot{x}(s)) \\ &= \mathrm{d}f(x(s))\dot{x}(s), \end{aligned}$$

where in the last show we used the chain rule (A.3) such that

$$\begin{aligned} \mathrm{d}f_\varphi(\varphi(x(s))) &= \mathrm{d}(f \circ \varphi^{-1})(\varphi(x(s))) \\ &= \mathrm{d}f(\varphi^{-1}(\varphi(x(s))))\mathrm{d}(\varphi^{-1}(\varphi(x(s))))\phi_*(\dot{x}(s)) \\ &= \mathrm{d}f(x(s))\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=s}\varphi^{-1}(\varphi(x(t))) \\ &= \mathrm{d}f(x(s))\dot{x}(s). \end{aligned} \tag{6.9}$$

We then prove item (b) based on the ideas when we obtain item (a). By calculating, we have

$$\begin{aligned}\mathcal{H}_\varphi(y(s), \mathrm{d}f_\varphi((y(s)))) &= \mathcal{H}\circ\varphi^*(\varphi(x(s)), \mathrm{d}f_\varphi(\varphi(x(s))))\\ &= \mathcal{H}(x(s), \varphi^*(\mathrm{d}f_\varphi(\varphi(x(s)))))\\ &= \mathcal{H}(x(s), \mathrm{d}f(x(s))),\end{aligned}$$

where in the last equality we use (6.9). Therefore, item (b) is obtained. We continue to prove item (c) by simple calculating, and get

$$\begin{aligned}\mathcal{L}_\varphi(y(s), \dot{y}(s)) &= \mathcal{L}\circ\varphi_*^{-1}(\varphi(x(s)), \varphi_*(\dot{x}(s)))\\ &= \mathcal{L}(x(s), \dot{x}(s)).\end{aligned}$$

To prove item (d), for any $(\eta, \xi) \in T^*N$, we have

$$\begin{aligned}\mathcal{H}_\varphi(\eta, \xi) &= \mathcal{H}(\varphi^{-1}(\eta), \varphi^*(\xi))\\ &= \sup_{v\in T_\eta M}\left\{\varphi^*(\xi)(v) - \mathcal{L}(\varphi^{-1}(\eta), v)\right\}\\ &\overset{(A.2)}{=} \sup_{\varphi_*(v)\in T_\eta N}\left\{\xi(\varphi_*(v)) - \mathcal{L}_\varphi(\eta, \varphi_*(v))\right\}\\ &= \sup_{w\in T_\eta N}\left\{\xi(w) - \mathcal{L}_\varphi(\eta, w)\right\},\end{aligned}$$

where the second equality is the fact that $\mathcal{L}$ is the Legendre transform of $\mathcal{H}$. The proof is completed. □

Using a transfer of M to a coordinate chart, we can work on Euclidian space. We will there construct local solutions using convex analysis and differential inclusion theory. Below, we will use the notion of a subdifferential.

Definition 6.9. For a general convex functional $p \mapsto \Phi(p)$ we denote the subdifferential at $p_0 \in \mathbb{R}^d$ as the set

$$\partial_p\Phi(p_0) := \{\xi \in \mathbb{R}^d : \Phi(p) \geqslant \Phi(p_0) + \xi(p - p_0), \forall p \in \mathbb{R}^d\}.$$

In the next result, we obtain a local solution to (6.8) by transferring to a chart. We follow the notation of Lemma 6.8

Lemma 6.10. *Let M be a Riemannian manifold and let $x_0 \in M$. Let $\varphi : \mathcal{O} \subseteq M \to \varphi(\mathcal{O}) \subseteq \mathbb{R}^d$ be a coordinate chart. Consider the open ball $\mathcal{O} := B_R(x_0)$ around x_0 with the radius $R > 0$ strictly smaller than the injectivity radius i_{x_0} at x_0. Fix $f \in C^1(M)$. Then the following content holds.*

(a) *There exists a solution $y(t) : [0, T_0(x)) \to \varphi(\mathcal{O}) \subseteq \mathbb{R}^d$ to the differential inclusion*

$$\begin{cases}\dot{y}(t) \in \partial_p\mathcal{H}_\varphi(y(t), \mathrm{d}f_\varphi(y(t))),\\ y(0) = 0 = \varphi(x_0)\end{cases} \tag{6.10}$$

with

$$T_0(x) = \inf\left\{t > 0 \,\middle|\, y(t) \notin \varphi(B_{R/2}(x_0))\right\}. \tag{6.11}$$

(b) *Set $x(t) = \varphi^{-1}(y(t))$. Then the curve $x : [0, T_0(x)) \to B_{R/2}(x_0) \subseteq M$ satisfies $x(0) = x_0$ and*

$$\int_0^t \mathcal{H}(x(s), \mathrm{d}f(x(s)))\mathrm{d}s + \int_0^t \mathcal{L}(x(s), \dot{x}(s))\mathrm{d}s = \int_0^t \mathrm{d}f(x(s))\dot{x}(s)\mathrm{d}s. \tag{6.12}$$

for any $t < T_0(x)$.

Proof. We first prove the existence of a solution to the differential inclusion (6.10). By taking $\mathcal{O} = B_R(x_0)$ in Lemma 6.8 and define $T_0(x) = \inf\left\{t > 0 \,\middle|\, y(t) \notin \varphi(B_{R/2}(x_0))\right\}$, the subdifferential $\partial_p\mathcal{H}_\varphi(y(t)), \mathrm{d}f_\varphi(y(t))$ satisfies all the conditions of Lemma 5.1 of [3]. Note that for this statement, we use that the convexity of $\mathcal{H}$ in p obtained in Lemma 5.3 transfers to $\mathcal{H}_\varphi$. Hence, there exists a solution $y(t)$ such that (6.10) holds.

Next, we turn to prove that there exists a solution such that (6.12) holds by local construction. To do it, for the initial point $x_0 \in M$, there exists a ball $B_R(x_0)$ with R strictly smaller than i_{x_0}. We claim that

$$\int_0^t \mathcal{H}_\varphi(y(s), \mathrm{d}f_\varphi(y(s)))\mathrm{d}s + \int_0^t \mathcal{L}_\varphi(y(s), \dot{y}(s))\mathrm{d}s = \int_0^t \mathrm{d}f_\varphi(y(s))\dot{y}(s)\mathrm{d}s. \tag{6.13}$$

on $\varphi(\mathcal{O})$. Then (6.12) follows from (6.13) and Lemma 6.8.

We are left to prove that (6.13) holds. On the one hand, we have that

$$\mathcal{H}_\varphi(y(s), \mathrm{d}f_\varphi(y(s))) \geqslant \mathrm{d}f_\varphi(y(s))\dot{y}(s) - \mathcal{L}_\varphi(y(s), \dot{y}(s)),$$

for all $y(s) \in \varphi(\mathcal{O})$, via convex duality. Then, integrating the above inequality gives one inequality in (6.13).

Regarding the other inequality, via (6.10) we obtain for all $p \in \varphi(\mathcal{O})$,

$$\mathcal{H}_\varphi(y(s), p) \geqslant \mathcal{H}_\varphi(y(s), \mathrm{d}f_\varphi(y(s))) + \dot{y}(s)\left(p - \mathrm{d}f_\varphi(y(s))\right),$$

and as a consequence

$$\mathcal{H}_\varphi(y(s), \mathrm{d}f_\varphi(y(s))) \leqslant \mathrm{d}f_\varphi(y(s))\dot{y}(s) - \mathcal{L}_\varphi(y(s), \dot{y}(s)),$$

and integrating gives the other inequality. □

6.3.2. Lower bounding the time of existence of local solutions

The first step in lower bounding the time of existence of local solutions is a a-priori control of any curve satisfying (6.8). The next result follows as a by-product of Condition 6.1.

Lemma 6.11. *Let Assumption 3.5 be satisfied. Let $K_0 \subseteq M$ a compact set and $T > 0$. For any $f \in \mathcal{D}(\mathbf{H})$, then there is a compact set $\hat{K} \subseteq M$ such that any curve $x : [0, T_0) \to M$ with $T_0 \leqslant T$ satisfying $x(0) \in K_0$ and for all $t < T_0$*

$$\int_0^t \mathcal{H}(x(s), \mathrm{d}f(x(s)))\mathrm{d}s + \int_0^t \mathcal{L}(x(s), \dot{x}(s))\mathrm{d}s = \int_0^t \mathrm{d}f(x(s))\dot{x}(s)\mathrm{d}s \tag{6.14}$$

it holds that $x(t) \in \hat{K}$ for any $t < T_0$.

Recall that by Definition 5.4 of $\mathbf{H}$ we have $\mathcal{D}(\mathbf{H}) = C_c^2(M)$.

Proof. First of all, note that

$$c_f := \sup_z \{-\mathbf{H}f(z)\} < \infty$$

as $\mathcal{H}$ is continuous by Lemma 5.3 and $f \in C_c^2(M)$. Furthermore, write $\|f\| = \sup_x |f(x)|$. Note that by (6.14), we have for any curve

$$\int_0^t \mathcal{L}(x(s), \dot{x}(s))\mathrm{d}s = f(x(t)) - f(x(0)) - \int_0^t \mathcal{H}(x(s), \mathrm{d}f(x(s)))\mathrm{d}s \leqslant 2\|f\| + tc_f \leqslant 2\|f\| + Tc_f$$

the result thus follows by Condition 6.1(b). □

For the next three Lemma's and the proof of Condition 6.2, we first provide a short sketch of the approach before giving a rigorous proof. The approach involves proof by contradiction. Specifically, we assume that there does not exist a global curve on $[0, T]$ satisfying (6.3). We first find the maximum time interval $[0, T_{\max}]$, $T_{\max} < T$, in which the curve satisfies (6.3). However, by extending the existing curve through patching in a chart at a new point and lower bound on the time length of the extension, we obtain a new curve that operates over a longer time interval, which leads to a contradiction.

Next we show the curve as in Lemma 6.10 has Lagrangian cost that grows linearly in time uniformly in their starting point in a compact set.

Lemma 6.12. *Let M be a Riemannian manifold and $K \subseteq M$ a compact set. Fix $R \in [i(K)/2, i(K))$. Then there is a constant C such that for any curve $x(t) : [0, T_0(x)) \to B_{R/2}(x_0)$ with $x(0) = x_0 \in K$ as in Lemma 6.10, we have*

$$\int_0^t \mathcal{L}(x(s), \dot{x}(s))\mathrm{d}s \leqslant Ct$$

for any $t < T_0(x)$.

Proof. First of all, denote by $\hat{K}$ the compact set obtained by covering K by balls of radius $R/2$. No considered curve can leave $\hat{K}$ by construction.

Denote $c_{f,\hat{K}} = \sup_{z \in \hat{K}}\{-\mathbf{H}f(z)\}$. As x satisfies (6.12), by Condition 6.1(c), there exists a function $\psi_{\hat{K},R}$, R is independent of x, such that for $t < T_0(x)$

$$\begin{aligned}\int_0^t \mathcal{L}(x(s), \dot{x}(s))\mathrm{d}s &= \int_0^t \mathrm{d}f(x(s))\dot{x}(s)\mathrm{d}s - \int_0^t \mathbf{H}f(x(s))\mathrm{d}s \\ &\leqslant \int_0^t \psi_{\hat{K},R}\left(\mathcal{L}(x(s), \dot{x}(s))\right)\mathrm{d}s + tc_{f,\hat{K}}.\end{aligned}$$

Furthermore, as $\psi_{\hat{K},R}$ is non-decreasing and the fact that $\frac{\psi_{\hat{K},R}(r)}{r}$ converges to 0 for $r \to \infty$, there exist $0 < m < 1$ and $r^* \geqslant 1$ such that $\frac{\psi_{\hat{K},R}(r)}{r} \leqslant m$ for $r \geqslant r^*$. Proceeding our estimate, by splitting the integral into regions $[0, t] = I_1 \cup I_2$ with

$$I_1 := \left\{ s \in [0, t] \,\middle|\, \mathcal{L}(x(s), \dot{x}(s)) \geqslant r^* \right\},$$
$$I_2 := \left\{ s \in [0, t] \,\middle|\, \mathcal{L}(x(s), \dot{x}(s)) < r^* \right\},$$

we get

$$\int_0^t \mathcal{L}(x(s),\dot{x}(s))\mathrm{d}s \leqslant \int_{I_1} \frac{\psi_{\hat{K},R}(\mathcal{L}(x(s),\dot{x}(s)))}{\mathcal{L}(x(s),\dot{x}(s))}\mathcal{L}(x(s),\dot{x}(s))\mathrm{d}s + \int_{I_2} \psi_{\hat{K},R}(\mathcal{L}(x(s),\dot{x}(s)))\mathrm{d}s + tc_{f,\hat{K}}$$
$$\leqslant m\int_0^t \mathcal{L}(x(s),\dot{x}(s))\mathrm{d}s + t\left(\psi_{\hat{K},R}(r^*) + c_{f,\hat{K}}\right).$$

Rearranging terms leads to

$$\int_0^t \mathcal{L}(x(s),\dot{x}(s))\mathrm{d}s \leqslant t\frac{\psi_{\hat{K},R}(r^*) + c_{f,\hat{K}}}{1-m}$$

establishing the claim with $C = \frac{\psi_{\hat{K},R}(r^*)+c_{f,\hat{K}}}{1-m}$. □

Next we control the speed at which curves as in Lemma 6.10 move away from their starting point.

Lemma 6.13. *Let M be a Riemannian manifold and $K \subseteq M$ a compact set. Fix $R \in [i(K)/2, i(K))$.*

Then there is a $C > 0$ such that for any $x_0 \in K$ and any curve $x(t) : [0, T_0(x)) \to B_{R/2}(x_0)$ with $x(0) = x_0$ as in Lemma 6.10, we have

$$\frac{1}{2}d^2(x(t), x_0) \leqslant tC$$

for any $t < T_0(x)$. In particular $T_0(x) \geqslant \frac{R^2}{8C}$.

The proof of Lemma 6.13 relies on Lemma 6.10 and the following preliminary lemma. We first state the preliminary lemma before proceeding to prove Lemma 6.13.

Lemma 6.14. *Let $K \subseteq M$ be a compact set in M. For any $x_0 \in K$ and radius $R < i_{x_0}$, set*

$$g_{x_0,R}(x) = \theta_R\left(\frac{1}{2}d^2(x,x_0)\right)$$

where $\theta_R : [0,\infty) \to [0, \frac{3}{4}R]$ is a smooth non-decreasing function, satisfying $\theta'_R(r) \leqslant 1$ where $\theta_R(r) = r$ for $r \leqslant R/2$ and $\theta_R(r)$ is constant for $r \geqslant \frac{3}{4}R$.

For any such R, we have $g_{x_0,R} \in C_{K,R}$ where $C_{K,R}$ was defined in Condition 6.1(c) Eq. (6.2). Moreover, $g_{x_0,R} \in \mathcal{D}(\mathbf{H})$.

Proof. By construction, we have

$$\mathrm{d}g_{x_0,R}(x) = \theta'_R\left(\frac{1}{2}d^2(x,x_0)\right)d(x,x_0)$$

which by the properties of θ_R satisfies

$$|\mathrm{d}g_{x_0,R}(x)| \leqslant d(x,x_0) \leqslant R$$

for any $x \in B(x_0,R)$. In particular, we have $g_{x_0,R} \in C_{K,R}$. Moreover, since $g_{x_0,R}$ is twice continuously differentiable and constant outside of a compact set, we conclude that $g_{x_0,R} \in \mathcal{D}(\mathbf{H})$. □

Proof of Lemma 6.13. Fix $x_0 \in K$ and any curve $x(t) : [0, T_0(x)) \to B_{R/2}(x_0)$ with $x(0) = x_0$ as in Lemma 6.10. Let $g_{x_0,R} \in \mathcal{D}(\mathbf{H})$ be any smooth bounded function as in Lemma 6.14 and $\hat{K}$ be the compact set obtained by covering K by balls of radius $R/2$.

It thus follows by the proof strategy of Lemma 6.12 that for any $t < T_0(x)$, we have

$$\frac{1}{2}d^2(x(t),x_0) \leqslant t\left(mC_1 + \psi_{\hat{K},R}(r^*)\right).$$

The result thus follows for $C = mC_1 + \psi_{\hat{K},R}(r^*)$. □

We are ready to verify Condition 6.2.

Proof of Condition 6.2. We argue by contradiction. Fix $x_0 \in M$ and $T > 0$. Suppose there does not exist an absolutely continuous curve $x(t)$, $t \in [0,T]$ started at $x_0 \in M$ such that (6.3) holds.

In other words,

$$T_{\max} = \sup\left\{T_0(x)\,\middle|\,\exists\, x : [0,T_0(x)) \to M \text{ satisfying (6.3)}, x(0) = x_0\right\} < T. \tag{6.15}$$

By Lemma 6.11 there is a compact set $K \subseteq M$ such that any curve considered in (6.15) stays in K. Fix $\varepsilon < \frac{R^2}{8C} \leqslant T_0(x)$ as in Lemma 6.13.

Fix the curve x satisfying (6.3) with $x(0) = x_0$, $T_0(x) > T_{\max} - \varepsilon$. Patching the curve $\tilde{x} : [0, T_0(\tilde{x}))$ started from $x(T_0(x)-\varepsilon)$ obtained from Lemma 6.10 to the curve x at time $T_0(x) - \varepsilon$, we obtain from Lemma 6.13 that this curve, is a solution to (6.3) on the time interval $[0, T_0(x) - \varepsilon + T_0(\tilde{x}))$, which contradicts (6.15).

This establishes the claim. □

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author thanks Rik Versendaal for numerous valuable comments. We sincerely thank our referees for their detailed remarks and suggestions, which have greatly enhanced the content and presentation of our paper.

The research of Y. Hu was supported by the China Scholarship Council (CSC). The research of F. Xi was supported by the National Natural Science Foundation of China (Grant No. 12071031).

Appendix A. Riemannian manifold

In this section, we introduce some of the definitions, properties, and symbols that were mentioned earlier. This can be found in any textbook on Riemannian manifold, for example, [17,21].

Throughout the paper, (M, g) is a d-dimensional connected complete Riemannian manifold. We start with the definition of *chart*, which is used to prove Condition 6.2. A coordinate chart (or just a chart) on M is a pair $(\mathcal{O}, \varphi)$, where $\mathcal{O}$ is a homeomorphism from $\mathcal{O}$ to an open subset $\tilde{\mathcal{O}} = \varphi(\mathcal{O}) \subset \mathbb{R}^d$.

The tangent space of M at $x \in M$ is denoted by $T_x M$. We denote by $\langle \cdot, \cdot \rangle_x = g(\cdot, \cdot)$ the scalar product on $T_x M$ with the associated norm $| \cdot |_x$, where the subscript x is sometimes omitted. The tangent bundle of M is denoted by $TM := \sqcup_{x\in M} T_x M$, which is naturally a manifold. Let $T_x^* M = (T_x M)^*$ be the *cotangent space* at $x \in M$, namely the dual space of the tangent space $T_x M$ (the space of linear functions on $T_x M$). Let $T^* M = \sqcup_{x\in M} T_x^* M$, which is called the *cotangent bundle* on M.

Given a piecewise smooth curve $\gamma : [a, b] \to M$ joining x to y, i.e. $\gamma(a) = x$ and $\gamma(b) = y$, we can define the length of γ by $l(\gamma) = \int_a^b |\dot{\gamma}(t)| dt$. Then the Riemannian distance $d(x, y)$, which induces the original topology on M, is defined by minimizing this length over the set of all such curves joining x to y.

Let ∇ be the Levi-Civita connection associated with the Riemannian metric. Let γ be a smooth curve in M. A vector field X is said to be parallel along γ if and only if $\nabla_{\dot{\gamma}_t} X = 0$. If $\dot{\gamma}$ itself is parallel along γ, we say that γ is a geodesic, and in this case $|\dot{\gamma}|$ is constant. When $|\dot{\gamma}| = 1$, γ is said to be normalized. A geodesic joining x to y in M is said to be minimal if its length equals $d(x, y)$.

A Riemannian manifold is complete if for any $x \in M$ all geodesics emanating from x are defined for all $-\infty < t < \infty$. By the Hopf–Rinow Theorem [16, Theorem 6.13], we know that if M is complete then any pair of points in M can be joined by a minimal geodesic. Moreover, (M, d) is a complete metric space and bounded closed subsets are compact.

Given a (piecewise) smooth curve $\gamma : [a, b] \to M$, we denote parallel transport along γ from $\gamma(t_0)$ to $\gamma(t_1)$ by $\tau_{\gamma,t_0 t_1}$, or simply $\tau_{t_0 t_1}$ whenever the meant curve is clear. If points $x, y \in M$ can be connected by a unique geodesic of minimal length, we will also write τ_{xy} meaning parallel transport from x to y along this specific geodesic.

The exponential map $\exp_x$: $T_x M \to M$ at x is defined by $\exp_x v = \gamma_v(1, x)$ for each $v \in T_x M$, where $\gamma(\cdot) = \gamma_v(\cdot, x)$ is the geodesic starting at x with velocity v. Then $\exp_x(tv) = \gamma_v(t, x)$ for each real number t. Note that the mapping $\exp_x$ is differentiable on $T_x M$ for any $x \in M$.

In many cases, the minimal geodesic is not unique. For instance, for the unit sphere $\mathbb{S}^d$, each half circle linking the highest and the lowest points is a minimal geodesic. This fact leads to the notion of cut-locus.

Definition A.1. Let $x \in M$. For any $X \in \mathbb{S}_x := \{X \in T_x M \ : \ |X| = 1\}$, let

$$r(X) := \sup\{t > 0 : d(x, \ \exp_x(tX)) = t\}.$$

If $r(X) < \infty$ then we call $\exp_x(r(X)X)$ a cut-point of x. The set

$$\mathrm{cut}(x) := \{\exp_x(r(X)X) \ : \ X \in \mathbb{S}_x, \ r(X) < \infty\}$$

is called the *cut-locus* of the point of x. Moreover, the quantity

$$i_x := \inf\{r(X) : X \in \mathbb{S}_x\}$$

is called the *injectivity radius* of x. For any set $A \subseteq M$ we write $i(A) := \inf_{x\in A} i_x$ the injectivity radius of A.

Lemma A.2 ([13]). *The injectivity radius i_x depends continuously on x. In particular, if $K \subseteq M$ is compact we have $i(K) > 0$.*

Note that $i(K) > 0$ is used to find a smooth distance on M.

Definition A.3. Let $\mathcal{T}(M)$ be the space of smooth vector fields on M and let ∇ be any connection on M. The formula

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

for $X, Y, Z \in \mathcal{T}(M)$, defines a function $\mathcal{R} : \mathcal{T}(M) \times \mathcal{T}(M) \times \mathcal{T}(M)$ called the Riemannian curvature of M, where $[X, Y] = XY - YX$ is the commutator of X and Y.

By taking the trace of the curvature tensor with respect to the first and the last entry, we obtain a 2-tensor which we will call the Ricci tensor of the (co)-metric g, denoted by $\mathcal{R}_g$.

To prove the existence solutions of HJB equations on M, we need the definitions of push-forward and pullback.

Definition A.4 (*Push-forward*). If M and N are smooth manifolds and $\varphi : M \to N$ is a smooth map, for each $p \in M$ we define a map

$$\varphi_{*p} : T_pM \to T_{\varphi(p)}N, \tag{A.1}$$

called the *push-forward* associated with φ, by

$$(\varphi_{*p}(v))(f) = v(f \circ \varphi), \quad v \in T_pM,\ f \in C^\infty(M).$$

Definition A.5 (*Pullback*). If M and N are smooth manifolds and $\varphi : M \to N$ be an invertible smooth map, for each $p \in M$ we define a map

$$\varphi_p^* : T^*_{\varphi(p)}N \to T^*_pM$$

by *pullback* associated with φ

$$(\varphi_p^*\xi)(v) = \xi(\varphi_{*p}(v)), \quad \xi \in T^*_{\varphi(p)}N,\ v \in T_pM. \tag{A.2}$$

We can put all φ_{*p} and φ_p^* together to obtain $\varphi_* : TM \to TN$ and $\varphi^* : T^*N \to T^*M$, respectively.

The next lemma shows that tangent vectors to curves behave well under composition with smooth maps.

Lemma A.6 (*Proposition 3.11 in [17]*). *Let $\varphi : M \to N$ be a smooth map, and let $\gamma : J \to M$ be a smooth curve, where $J \in \mathbb{R}$ is an interval. For any $t \in J$, the tangent vector to the composite curve $\varphi\circ\gamma$ at $t = t_0$ is given by*

$$(\varphi\dot{\circ}\gamma)(t_0) = (\varphi\circ\gamma)_* \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0} = \varphi_*\dot{\gamma}(t_0).$$

The chain rule for total derivatives is important in Riemannian manifolds because it allows us to compute the derivative of a composite function.

Lemma A.7 (*The Chain Rule for Total Derivatives, Proposition A.24 in [17]*). *Suppose V, W, X are finite-dimensional vector spaces, $U \subset V$ and $\tilde{U} \subset W$ are open sets, and $F : U \to \tilde{U}$ and $G : \tilde{U} \to X$ are maps. If F is differentiable at $a \in U$ and G is differentiable at $F(a) \in U$, then $G\circ F$ is differentiable at a, and*

$$D(G\circ F)(a) = DG(F(a))\circ DF(a). \tag{A.3}$$

Appendix B. Viscosity solutions

Now we define viscosity sub and supersolutions, which is often used in the proof.

Definition B.1 (*Viscosity Solutions*). Let $H \subseteq C_b(M) \times C_b(M \times S)$ be a multi-valued operator. We denote $\mathcal{D}(H)$ for the domain of H and $\mathcal{R}(H)$ for the range of H. Let $\lambda > 0$ and $h \in C_b(M)$. Consider the Hamilton–Jacobi equations

$$f - \lambda H f = h. \tag{B.1}$$

Classical solutions We say that u is a classical subsolution of (B.1) if there is a g such that $(u, g) \in H$ and $u - \lambda g \leqslant h$. We say that v is a classical supersolution of (B.1) if there is a function g such that $(v, g) \in H$ and $v - \lambda g \geqslant h$. We say that u is a classical solution if it is both a subsolution and a supersolution.

Viscosity subsolutions We say that u is a (viscosity) subsolution of (B.1) if u is bounded, upper semicontinuous, and if for every $(f, g) \in H$ there exists a sequence $(x_n, z_n) \in M \times S$ such that

$$\lim_{n\to\infty} u(x_n) - f(x_n) = \sup_x u(x) - f(x),$$

$$\limsup_{n\to\infty} u(x_n) - \lambda g(x_n, z_n) - h(x_n) \leqslant 0.$$

Viscosity supersolutions We say that v is a (viscosity) supersolution of (B.1) if v is bounded, lower semicontinuous, and if for every $(f, g) \in H$ there exists a sequence $(x_n, z_n) \in M \times S$ such that

$$\lim_{n\to\infty} v(x_n) - f(x_n) = \inf_x v(x) - f(x),$$

$$\liminf_{n\to\infty} v(x_n) - \lambda g(x_n, z_n) - h(x_n) \geqslant 0.$$

Viscosity solutions We say that u is a (viscosity) solution of (B.1) if it is both a subsolution and a supersolution to (B.1).

Remark B.2. Consider the definition of subsolutions. Suppose that the test function $(f, g) \in H$ has compact sublevel sets, then instead of working with a sequence (x_n, z_n), we can pick (x_0, z_0) such that

$$u(x_0) - f(x_0) = \sup_x u(x) - f(x),$$

$$u(x_0) - \lambda g(x_0, z_0) - h(x_0) \leqslant 0.$$

Similarly, a simplification holds in the case of supersolutions. This is used in the proof Lemma 5.15.

Definition B.3 (*Comparison Principle*)**.** We say that (B.1) satisfies the comparison principle if for every viscosity subsolutions u and viscosity supersolutions v to (B.1), we have $u \leqslant v$.

Remark B.4 (*Uniqueness*)**.** The comparison principle implies uniqueness of viscosity solutions. Suppose that u and v are both viscosity solutions, then the comparison principle yields that $u \leqslant v$ and $v \leqslant u$, implying that $u = v$.

References

[1] F. Collet, R.C. Kraaij, Dynamical moderate deviations for the Curie–Weiss model, Stochastic Process. Appl. 127 (9) (2017) 2900–2925.
[2] M.G. Crandall, T.M. Liggett, Generation of semi-groups of nonlinear transformations on general Banach spaces, Amer. J. Math. 93 (2) (1971) 265–298.
[3] K. Deimling, Multivalued Differential Equations, Walter de Gruyter & Co., Berlin, 1992.
[4] S. Della Corte, R.C. Kraaij, Large deviations for Markov processes with switching and homogenisation via Hamilton–Jacobi–Bellman equations, Stochastic Process. Appl. 170 (2024) 104301.
[5] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Springer Berlin, Heidelberg, 1998.
[6] M.D. Donsker, S.R.S. Varadhan, On a variational formula for the principal eigenvalue for operators with maximum principle, Proc. Natl. Acad. Sci. USA 72 (1975) 780–783.
[7] K.D. Elworthy, Stochastic Differential Equations on Manifolds, Cambridge University Press, Cambridge-New York, 1982.
[8] J. Feng, T.G. Kurtz, Large Deviations for Stochastic Processes, Springer Berlin, Heidelberg, 2006.
[9] F. den Hollander, Large Deviations, American Mathematical Society, New York, 2008.
[10] E.P. Hsu, Stochastic Analysis on Manifolds, American Mathematical Society, Providence, RI, 2002.
[11] Y. Hu, R.C. Kraaij, F. Xi, Large deviations for Cox-Ingersoll-Ross processes with state-dependent fast switching, 2023, arXiv:2307.11936.
[12] G. Huang, M. Mandjes, P. Spreij, Large deviations for Markov-modulated diffusion processes with rapid switching, Stochastic Process. Appl. 126 (6) (2016) 1785–1818.
[13] W. Klingenberg, Riemannian Geometry, Walter de Gruyter & Co., Berlin-New York, 1982.
[14] R. Kraaij, F. Redig, R. Versendaal, Classical large deviation theorems on complete Riemannian manifolds, Stochastic Process. Appl. 129 (11) (2019) 4294–4334.
[15] R.C. Kraaij, M.C. Schlottke, A large deviation principle for Markovian slow-fast systems, 2020, arXiv:2011.05686.
[16] J.M. Lee, An introduction to curvature, in: Riemannian Manifolds, Springer-Verlag, New York, 1997.
[17] J.M. Lee, Introduction To Smooth Manifolds, Springer-Verlag, New York, 2003.
[18] M.A. Peletier, M.C. Schlottke, Large-deviation principles of switching Markov processes via Hamilton-Jacobi equations, 2021, arXiv:1901.08478.
[19] M. Röckner, T.S. Zhang, Sample path large deviations for diffusion processes on configuration spaces over a Riemannian manifold, Publ. Res. Inst. Math. Sci. 40 (2) (2004) 385–427.
[20] R. Versendaal, Large deviations for Brownian motion in evolving Riemannian manifolds, 2020, arXiv:2004.00358.
[21] F.-Y. Wang, Analysis For Diffusion Processes on Riemannian Manifolds, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2014.