



DELFT UNIVERSITY OF TECHNOLOGY

FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE

Large Deviations for Drifted Geodesic Random Walks

MASTER THESIS IN APPLIED MATHEMATICS

Author:
Niek VAN HEERDE

Supervisor:
dr. Rik VERSENDAAL

Academic year 2025–2026

Abstract

This thesis proves a Cramér-type large deviation principle for *drifted* geodesic random walks on a complete Riemannian manifold. A geodesic random walk is the manifold analogue of a sum of independent increments: at each step a random tangent vector is drawn and the walk moves along the geodesic that this vector defines. We add a deterministic drift, given by a smooth vector field V , so that each step follows the geodesic determined by the random increment together with the value of V at the current point. For the empirical average of such a walk we establish a large deviation principle at speed n , with an explicit good rate function obtained as the Legendre transform of the increment law, corrected by the drift. The proof rests on a single construction, the *drift map*, which sends a noise path to the path obtained by carrying it on top of the flow of V . We show that this map is continuous, and the large deviation principle then follows from the Riemannian version of Mogulskii's theorem by two applications of the contraction principle. Setting the drift to zero recovers the undrifted theorem of Kraaij, Redig and Versendaal. A final chapter removes the bounded-increment hypothesis. It is replaced by Cramér's condition on the increment law together with a bounded-geometry assumption on the manifold, and the same rate function is obtained.

Contents

1	Introduction	4
2	Riemannian geometry	8
2.1	Riemannian metrics	8
2.2	Connections	9
2.3	Covariant differentiation along curves, geodesics, and parallel transport	10
2.4	The exponential map and normal coordinates	12
2.5	Riemannian distance and completeness	13
2.5.1	Length and distance	13
2.5.2	Minimising curves and the first variation of length	14
2.5.3	Geodesic balls	17
2.5.4	Completeness and the injectivity radius	17
2.5.5	Absolutely continuous paths	18
2.6	Smooth dependence of parallel transport on its endpoints	18
2.7	Curvature	19
2.8	Jacobi fields and curvature bounds	20
3	Analytic and probabilistic prerequisites	24
3.1	Large deviations	24
3.1.1	The large deviation principle	24
3.1.2	Cramér and Mogulskii in $\mathbb{R}^d$	25
3.1.3	The contraction principle	25
3.1.4	Exponential equivalence	25
3.1.5	Exponential tightness	26
3.1.6	Exponentially good approximations	26
3.2	Functional analysis	26
3.2.1	Compactness	27
3.2.2	Weak convergence of velocities	27
3.3	Convex analysis	28
3.4	Ordinary differential equations	28
3.4.1	Grönwall’s inequality	29
3.4.2	The Carathéodory–Picard theorem	29
4	Convergence of the geodesic Euler scheme	30
4.1	Warm-up: the flat case $\mathbb{R}^d$	30
4.1.1	Setup	30
4.1.2	Local truncation error	30
4.1.3	Error recursion and discrete Grönwall	30
4.1.4	Uniform convergence	31
4.2	Extension to complete Riemannian manifolds	31
4.2.1	One-step consistency	32
4.2.2	Stability of the flow	32
4.2.3	Global error and uniform convergence	33

5	Geodesic random walks	34
5.1	Consistent increment measures and geodesic random walks	34
5.2	Drifted geodesic random walks	35
6	Main results	37
6.1	The extension to unbounded increments	38
6.2	Examples	39
7	LDP for drifted random walks in $\mathbb{R}^d$	43
7.1	Setup	43
7.2	The undrifted walk and Mogulskii's theorem	43
7.3	The drift map and its continuity	43
7.4	Contraction principle: path-space LDP	44
7.5	Exponential equivalence	44
7.6	Endpoint rate function	45
8	Proof of Theorem 6.3	46
8.1	Proof strategy	46
8.2	Path-space LDP for the undrifted walk	46
8.2.1	Well-posedness of the drift ODE	47
8.2.2	Compactness of the admissible path space	51
8.2.3	Continuity of the drift map	52
8.3	The path-space LDP for the drifted flow	53
8.4	Exponential equivalence	55
8.4.1	Coupling	55
8.4.2	The noisy one-step consistency lemma	55
8.4.3	Global error and uniform convergence on a compact event	57
8.4.4	Exponential equivalence of the walk and the flow	58
8.5	Contraction to the endpoint	58
9	Removing the bounded-increments assumption	59
9.1	Truncation preliminaries	59
9.2	Convex-analytic preliminaries	61
9.3	Path-space exponential tightness	63
9.4	Layer 0: the theorem at a fixed truncation level	65
9.5	Layer 1: exponentially good approximation	66
9.6	Layer 2: identification of the rate function	70
10	Outlook: position-dependent noise	76
10.1	The expected theorems	76
10.2	Freidlin–Wentzell	77
10.3	What survives, and what does not	77

1 Introduction

Probability theory grew up on the real line and in Euclidean space. Many of the objects we now want to model with it do not live there. A diffusion tensor in medical imaging is a symmetric positive definite matrix. The orientation of a rigid body is a point on the rotation group $SO(3)$. A wind direction is a point on a circle, and a protein backbone is a sequence of angles on a torus. Data of this kind are called *manifold-valued*, and they occur across directional statistics, shape analysis, computer vision, robotics, and diffusion imaging [11]. To do statistics on such spaces, and to model dynamics on them, we need random processes that respect the geometry. The *geodesic random walk* is the natural analogue, on a curved space, of a sum of independent steps. In $\mathbb{R}^d$ a random walk adds a random vector at each step. On a manifold there is no global addition, so instead we draw a random tangent vector and follow the geodesic it points along for one step. The exponential map takes the place of the sum. Jørgensen studied these walks already in 1975 and showed that, when suitably rescaled, they converge to Brownian motion on the manifold [5]. The same construction is today a standard computational tool: geodesic walks are used to sample from distributions on manifolds and from convex bodies, and their mixing behaviour is an active topic [10, 9].

Convergence to Brownian motion is a law-of-large-numbers and central-limit statement. It describes the *typical* behaviour of the walk. Large deviation theory describes the opposite regime: the small probability of an *atypical* outcome, and the exponential rate at which that probability decays. In $\mathbb{R}^d$ the model result is Cramér’s theorem. The empirical mean of n independent, identically distributed steps satisfies a large deviation principle at speed n , with a rate function equal to the Legendre transform of the log-moment generating function of a single step [3]. Rate functions of this kind are not only of theoretical interest. They quantify the cost of rare events, they drive the importance-sampling schemes that make rare events computable, and, through the Freidlin–Wentzell theory, they identify the most likely path of a rare transition in a noisy dynamical system [4]. Carrying this theory from Euclidean space to manifolds lets these tools reach the manifold-valued setting described above.

Such an extension was carried out for the *undrifted* geodesic random walk by Versendaal [13, 14] and by Kraaij, Redig and Versendaal [6], who proved Cramér- and Mogulskii-type theorems on complete Riemannian manifolds. Their walk is driven by noise alone. Most processes that appear in applications, however, combine noise with a deterministic *drift*: a gradient descent seeking a minimum, a physical system pulled by a force, a stochastic differential equation with a drift term. This thesis adds that drift.

We study the drifted geodesic random walk, in which each step follows the geodesic determined by a random increment plus the value of a fixed vector field V at the current point (Definition 5.3). The main result is a Cramér-type large deviation principle for the empirical average of this walk (Theorem 6.3): the endpoints satisfy a large deviation principle in M at speed n , with an explicit good rate function built from the Legendre transform of the increment law and the drift. When V vanishes the result reduces to the undrifted theorem of [6]. The proof is organised around a single geometric object, the *drift map*, which carries a noise path onto the trajectory obtained by riding it on top of the flow of V . We show that this map is continuous, and the large deviation principle then passes through it by the contraction principle, starting from the Riemannian Mogulskii theorem for the undrifted walk. A final extension removes the assumption of bounded increments. It is replaced by Cramér’s condition on the increment law, together with a bounded-geometry hypothesis on the manifold, and reaches the same rate function (Theorem 6.5).

Outline

Sections 2–4 are preliminary, Section 5 introduces the objects and Section 6 states the main theorem, and the remaining sections carry out the proof.

Section 2 collects the necessary background on Riemannian geometry, recalling the definitions and results that will be used later. Readers familiar with connections, geodesics, parallel transport, the exponential map, and Jacobi fields may wish to skip ahead. Section 3 recalls the large deviation prerequisites: the large deviation principle, the classical Cramér and Mogulskii theorems, the contraction principle, and exponential equivalence.

In Section 4 we prove, as a geometric tool used later in the proof, the convergence of the geodesic Euler scheme: the piecewise-geodesic approximation to the integral curve of a smooth vector field converges uniformly at rate $O(h)$, where $h = 1/n$ is the step size. We first develop the Euclidean case in detail, as we will use a similar argumentation structure to extend to compact Riemannian manifolds.

Section 5 introduces geodesic random walks and their drifted variant, and Section 6 states the main theorem of the thesis together with its extension to unbounded increments and illustrative examples.

The proof follows a contraction-principle strategy. In Section 7 we carry it out first in $\mathbb{R}^d$ as a warm-up, exhibiting the strategy without geometric complications; Section 8 then executes the same four steps on the manifold, with the Riemannian Mogulskii theorem replacing the classical one and Jacobi-field estimates replacing the flat Grönwall bound.

Finally, Section 9 removes the bounded-increment assumption of the main theorem, replacing it with Cramér’s condition (finiteness of all exponential moments) at the cost of a bounded-geometry hypothesis on the manifold. The argument is a truncation tower built on the main theorem: it is applied verbatim at each truncation level, the truncated walks are shown to approximate the original one exponentially well, and the limiting rate function is identified by a Γ -convergence argument. This recovers, for the drifted walk, the generality of the undrifted theorem of [6].

Conventions and Notation

All manifolds are smooth, connected, and without boundary unless stated otherwise. We write ∇ for the Levi-Civita connection and d for the Riemannian distance. For a curve $\gamma : [a, b] \rightarrow M$, two notations for parallel transport are used: τ_{xy}^γ along γ itself, and, when x and y are joined by a unique minimising geodesic, $P_{x \rightarrow y}$ along that geodesic, which is the case that occurs throughout the proof. The two letters distinguish an arbitrary given curve from the canonical minimising geodesic, and are used consistently for this distinction; both symbols are listed in the table below. The flow of a vector field V is denoted φ_t , so $\dot{\varphi}_t(x) = V(\varphi_t(x))$ and $\varphi_0 = \text{id}$. Throughout we use the *Einstein summation convention*: an index that appears once as a subscript and once as a superscript in a single term is summed over, so that e.g. $X = X^i \partial_i$ abbreviates $\sum_i X^i \partial_i$ and $X_j = g_{ij} X^i$ abbreviates $\sum_i g_{ij} X^i$.

The symbols and function spaces used repeatedly are collected below for reference.

Geometry.

(M, g) complete Riemannian manifold; d its distance, ∇ its Levi-Civita connection.

$B(x, r)$, $\overline{B}(x, r)$ open and closed ball, centre x , radius r .

Exp_p , $\log_p$ exponential map at p and its local inverse.

$\text{inj}(A)$ injectivity radius of $A \subseteq M$; $\iota(p) := \text{inj}(\{p\})$ pointwise, and $\iota_0 := \text{inj}(K_0)$.

Rm , $\|\text{Rm}\|_\infty$ Riemann curvature tensor and its uniform operator-norm bound.

τ_{xy}^γ parallel transport $T_x M \rightarrow T_y M$ along the curve γ .

$P_{x \rightarrow y}$ parallel transport along the unique minimising geodesic from x to y .

V , φ_t drift vector field and its flow.

x_0 , K_0 base point and the tube $K_0 := \overline{B}(x_0, r + \|V\|_\infty)$.

Probability and large deviations.

μ_x , $\{\mu_x\}$ increment law on $T_x M$ and the consistent family of such laws.

Λ_x , Λ_x^* log-moment generating function $\Lambda_x(\lambda) = \log \int_{T_x M} e^{\langle \lambda, u \rangle} \mu_x(du)$ and its Legendre transform (the rate integrand).

Z^n , $\hat{Y}^n$, S_n^n undrifted interpolated walk, drifted interpolated walk, and drifted empirical average (endpoint).

I_M^V , I_{path}^V endpoint and path-space rate functions; the large deviation principle is always at speed n .

Function spaces (on $[0, 1]$, valued in M or $\mathbb{R}^d$).

C , C_{x_0} continuous paths, and those with $\phi(0) = x_0$, under the uniform metric $d_\infty(\phi_1, \phi_2) = \sup_t d(\phi_1(t), \phi_2(t))$.

AC absolutely continuous paths.

$\text{AC}_{x_0, r}$ absolutely continuous paths from x_0 with speed $\|\dot{\psi}\|_\infty \leq r$.

Acknowledgements

My greatest thanks go to my daily supervisor, Rik, who was always available when I needed him during this project. He was understanding whenever the work went slowly, he helped me to collect my own thoughts at times, and, by simply being a nice person to work with throughout, he gave me a warm welcome into doing research-level mathematics.

I also thank my responsible supervisor, Richard, for meeting Rik and me during the project and for letting me spar some ideas off him; those ideas did not end up in the thesis, but I enjoyed thinking them through with him. Thanks go to Havva as well, for taking the time to sit on my defence committee in the middle of the summer.

Coming to Delft meant a new city and a different direction of mathematics from the one I had done in my bachelor's, and my friends here made that transition far easier than it could have been. The material became easier to understand because I could work through it with them, spar with them, and complain to them when it refused to make sense. Most important of all, Delft would have been a lot more boring without them to hang out with.

My friends from my bachelor's and from Nijmegen in general put up with the stressed version of me whenever things were not going easily, during this thesis and during the master's as a whole, and I thank them for it. I thank my parents too, for their support throughout it all. Special thanks go to my girlfriend, Julie, for always being there for me, and for finding the words for the feelings I had at times when I could not completely express them myself.

2 Riemannian geometry

This section collects the Riemannian geometry needed for the rest of the thesis. We recall the main definitions and state the results that will be used, giving proofs only when they are short or illuminate a point relevant to later sections. Acquaintance with smooth manifolds, tangent spaces, smooth maps, the differential, and vector bundles is assumed; see Lee [8] for a thorough treatment. For a reference on Riemannian geometry, we recommend Lee [7] as well.

The material is organised building up from the definition of the Riemannian metric. This metric equips each tangent space with an inner product, giving lengths and angles pointwise. To then compare vectors at different points, and so to differentiate along curves, we introduce a connection; asking it to be compatible with the metric and torsion-free singles out the Levi-Civita connection and, with it, geodesics. The exponential map Exp_p assembles all geodesics leaving a point into a single map and lets us define normal coordinates, linearising the geometry locally. Lengths of curves then turn M into a metric space, and completeness guarantees that nearby points are joined by minimising geodesics. Once that minimising geodesic is available, parallel transport along it becomes a map of the two endpoints, and we prove smooth dependence on in a subsection of itself as we will use this in later sections. Curvature then arises as the effect on tangent vectors when moving them along curves and Jacobi fields capture how this curvature governs nearby geodesics. We turn that control into a growth bound for Jacobi fields which, together with a Lipschitz bound for Exp_p in its vector argument, gives the two estimates on the exponential map that the later sections quote by name. The first variation of length and the Jacobi equation are two instances of the same idea: differentiating a family of curves in the family parameter. It is this idea, rather than any single formula, that the later sections use again and again, so the section treats it as an object in its own right and not only as a step inside two proofs.

2.1 Riemannian metrics

Definition 2.1 (Riemannian metric). A *Riemannian metric* on a smooth manifold M is a smooth symmetric $(0, 2)$ -tensor field g that is positive definite at each point. A smooth manifold equipped with a Riemannian metric is called a *Riemannian manifold* and is denoted (M, g) .

In other words, a Riemannian metric is nothing more than a smoothly varying inner product on the tangent spaces. We write $\langle X, Y \rangle = g(X, Y)$ and $\|X\| = \langle X, X \rangle^{1/2}$. We also get natural notions of angles (via $\cos \theta = \langle X, Y \rangle / (\|X\| \|Y\|)$) and orthogonality ($\langle X, Y \rangle = 0$).

Any smooth manifold can be equipped with a Riemannian metric: one pulls back the Euclidean inner product via coordinate charts and glues using a partition of unity. This is a standard construction; see [8, Proposition 13.3].

As the Riemannian metric adds a layer of structure to the manifold, it comes with natural structure-preserving maps:

Definition 2.2 (Isometry). A diffeomorphism $\varphi : (M, g) \rightarrow (\widetilde{M}, \widetilde{g})$ between Riemannian manifolds is an *isometry* if $\varphi^* \widetilde{g} = g$, i.e., for all $p \in M$ and $X_p, Y_p \in T_p M$,

$$\widetilde{g}_{\varphi(p)}(d\varphi_p(X_p), d\varphi_p(Y_p)) = g_p(X_p, Y_p).$$

A smooth map $\varphi : M \rightarrow \widetilde{M}$ is a *local isometry* if for each $p \in M$ there is a neighbourhood U of p such that $\varphi|_U$ is an isometry onto its image.

So an isometry is a diffeomorphism whose differential is a linear isometry on each fibre.

2.2 Connections

Geodesic random walks rely on geodesics. To define them, we need a way to differentiate vector fields along curves and compare vectors in different tangent spaces. This motivates introducing a connection.

Definition 2.3 (Connection). A (linear) connection on TM is a map $\nabla : \Gamma(TM) \times \Gamma(TM) \rightarrow \Gamma(TM)$ (where $\Gamma(TM)$ denotes the space of smooth vector fields on M), written $(X, Y) \mapsto \nabla_X Y$, satisfying:

- (a) $\nabla_X Y$ is $C^\infty(M)$ -linear in X ;
- (b) $\nabla_X Y$ is $\mathbb{R}$ -linear in Y ;
- (c) $\nabla_X(fY) = f\nabla_X Y + (Xf)Y$ for $f \in C^\infty(M)$ (the Leibniz rule).

Right now, the connection looks like a global abstract operator on smooth vector fields. The fact that it has something to do with derivatives is motivated by locality: one can show that $\nabla_X Y|_p$ depends only on the value of X at p and the values of Y in an arbitrarily small neighbourhood of p . This means we can evaluate $\nabla_X Y$ at a point without knowing X and Y globally.

In local coordinates with frame $\{\partial_i\}$, a connection is completely determined by its *Christoffel symbols* Γ_{ij}^k , defined by $\nabla_{\partial_i} \partial_j = \Gamma_{ij}^k \partial_k$. The covariant derivative of $Y = Y^j \partial_j$ in the direction $X = X^i \partial_i$ is then

$$\nabla_X Y = (X(Y^k) + X^i Y^j \Gamma_{ij}^k) \partial_k.$$

An important special case is the *Euclidean connection* $\bar{\nabla}$ on $\mathbb{R}^d$, defined by $\bar{\nabla}_X(Y^j \partial_j) = (X(Y^j)) \partial_j$, i.e., one simply differentiates the component functions. In standard coordinates, its Christoffel symbols are all zero.

To make the connection feel less abstract, we make two remarks. First, on a single chart, choosing a connection amounts to nothing more than choosing the d^3 smooth functions Γ_{ij}^k freely: any assignment of Christoffel symbols defines a connection through the formula above, and conversely every connection arises this way. Second, by choosing the Christoffel symbols locally and using a partition of unity, we can get a global connection. So every smooth manifold admits a linear connection (in fact, infinitely many). The rest of this subsection will show that a Riemannian metric singles out one canonical choice among them.

Definition 2.4 (Torsion and symmetry). The *torsion tensor* of a connection ∇ is $T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$, where $[X, Y]$ denotes the Lie bracket of the vector fields X, Y . A connection is *symmetric* (or torsion-free) if $T \equiv 0$, which in coordinates amounts to $\Gamma_{ij}^k = \Gamma_{ji}^k$.

Definition 2.5 (Metric compatibility). A connection ∇ on a Riemannian manifold (M, g) is *compatible with g* if

$$X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$$

for all vector fields X, Y, Z .

Metric compatibility has several equivalent reformulations. We pause at one of them, as it might give a more intuitive idea of metric compatibility: A connection ∇ is compatible with g if and only if parallel transport (defined below) along any curve is a linear isometry between tangent spaces. Thus metric compatibility says that transporting vectors parallelly never distorts their lengths or the angles between them. A proof of this equivalence and the others can be found in [7, Proposition 5.5].

Metric compatibility alone does not determine a unique connection: there are many connections with $\nabla g = 0$, differing by “twisting” terms that still preserve inner products. A natural way to remove this freedom is to also impose symmetry. Geometrically, we are after the connection that makes geodesics “as straight as possible”. A concrete model comes from the embedded case: for a submanifold $M \subseteq \mathbb{R}^n$ carrying the induced metric (the Euclidean inner product of $\mathbb{R}^n$ restricted to each tangent space $T_p M \subseteq T_p \mathbb{R}^n \cong \mathbb{R}^n$), one differentiates a tangent vector field in the ambient Euclidean space and projects the result back onto TM . This *tangential connection* $\nabla_X^T Y = \pi^T(\tilde{\nabla}_X Y)$ is both compatible with the induced metric and torsion-free, and by Nash’s embedding theorem [2] every Riemannian manifold arises this way. Compatibility and symmetry are thus exactly the intrinsic shadow of the naive “differentiate and project” derivative. We will now also see that together they pin the connection down completely.

Theorem 2.6 (Fundamental Lemma of Riemannian Geometry). *Let (M, g) be a Riemannian manifold. There exists a unique linear connection ∇ on TM that is compatible with g and symmetric. This connection is called the Levi-Civita connection.*

In local coordinates, the Christoffel symbols of the Levi-Civita connection are given by

$$\Gamma_{ij}^k = \frac{1}{2} g^{k\ell} (\partial_i g_{\ell j} + \partial_j g_{\ell i} - \partial_\ell g_{ij}).$$

Remark 2.7 (Isometries and the Levi-Civita connection). The uniqueness of the Levi-Civita connection means that isometries automatically preserve it: if $\varphi : (M, g) \rightarrow (\tilde{M}, \tilde{g})$ is an isometry, then $\varphi_*(\nabla_X Y) = \tilde{\nabla}_{\varphi_* X}(\varphi_* Y)$, because $\varphi_* \nabla$ is again metric-compatible and torsion-free and so equals $\tilde{\nabla}$ by uniqueness.

We will always use the Levi-Civita connection unless stated otherwise.

One consequence of property (a) of a connection is worth recording separately, because the constants of the later sections are built from it. For a fixed smooth vector field Y , the map $X \mapsto \nabla_X Y$ is $C^\infty(M)$ -linear, so at each point it is a linear map $\nabla Y|_p : T_p M \rightarrow T_p M$, $v \mapsto \nabla_v Y$. We write

$$\|\nabla Y(p)\|_{\text{op}} := \sup\{\|\nabla_v Y\| : v \in T_p M, \|v\| \leq 1\}$$

for its operator norm. A bound $\|\nabla Y\|_{\text{op}} \leq L$ is the Riemannian form of “ Y is L -Lipschitz”, and it is what we use in the Grönwall later.

2.3 Covariant differentiation along curves, geodesics, and parallel transport

Curves on a manifold need not be embedded submanifolds: they can self-intersect, assigning two different tangent vectors to the same point. This is why we need a more general notion of vector field along a curve:

Definition 2.8 (Vector field along a curve). Let $\gamma : I \rightarrow M$ be a smooth curve. A *vector field along γ* is a smooth map $V : I \rightarrow TM$ such that $V(t) \in T_{\gamma(t)} M$ for every $t \in I$.

Lemma 2.9 (Covariant derivative along a curve). *Let ∇ be a linear connection on M and $\gamma : I \rightarrow M$ a smooth curve. There is a unique operator D_t on vector fields along γ satisfying: (i) D_t is $\mathbb{R}$ -linear; (ii) $D_t(fV) = \dot{f}V + f D_t V$ for $f : I \rightarrow \mathbb{R}$ smooth; (iii) if V extends to a vector field $\tilde{V}$ near $\gamma(t)$, then $D_t V(t) = \nabla_{\dot{\gamma}(t)} \tilde{V}$. In local coordinates $V(t) = V^j(t) \partial_j$, this takes the form*

$$D_t V = (\dot{V}^k + \Gamma_{ij}^k(\gamma(t)) \dot{\gamma}^i V^j) \partial_k.$$

We have now collected enough machinery to define geodesics.

Definition 2.10 (Geodesic). A curve $\gamma : I \rightarrow M$ is a *geodesic* if its acceleration vanishes: $D_t \dot{\gamma} \equiv 0$.

In local coordinates, the geodesic equation reads $\ddot{x}^k(t) + \Gamma_{ij}^k(\gamma(t)) \dot{x}^i(t) \dot{x}^j(t) = 0$, which is a second-order ODE. By standard existence and uniqueness theory, for any $p \in M$ and $v \in T_p M$ there is a unique maximal geodesic γ_v with $\gamma_v(0) = p$ and $\dot{\gamma}_v(0) = v$. On a compact manifold, geodesics exist for all time; this is due to the geodesic completeness of Theorem 2.25, proved there in the wider generality of a complete manifold.

Because a connection lets us compare tangent spaces and differentiate fields along curves, it also lets us transport tangent vectors along a curve without “twisting”:

Definition 2.11 (Parallel transport). Let $\gamma : I \rightarrow M$ be a smooth curve. A vector field V along γ is *parallel* if $D_t V \equiv 0$. For any $v \in T_{\gamma(a)} M$ there exists a unique parallel vector field V along γ with $V(a) = v$; the map

$$\tau_{ab}^\gamma : T_{\gamma(a)} M \rightarrow T_{\gamma(b)} M, \quad v \mapsto V(b),$$

is called *parallel transport* along γ . The subscripts here are the two curve parameters. When the curve is understood and only its endpoints matter we write τ_{xy}^γ for transport from $x = \gamma(a)$ to $y = \gamma(b)$, as in the table of notation.

Since the parallel transport equation is a linear ODE, τ_{ab}^γ is a linear isomorphism with inverse τ_{ba}^γ . Moreover, metric compatibility of the Levi-Civita connection ensures that parallel transport is an isometry: if V, W are parallel along γ , then $\frac{d}{dt} \langle V, W \rangle = 0$, so inner products are preserved.

One useful way to think of parallel transport is that it provides the “correct” way to compare tangent vectors at different points along a curve. In $\mathbb{R}^d$ this is trivial (just translate the vector), but on a curved manifold it depends on the path taken. This path-dependence is measured by the curvature tensor.

Remark 2.12 (Parallel transport recovers ∇). Conversely, the connection can be recovered from parallel transport. For a vector field Y and a curve γ with $\dot{\gamma}(0) = X_p$,

$$(\nabla_X Y)_p = \lim_{h \rightarrow 0} \frac{\tau_{h,0}^\gamma Y_{\gamma(h)} - Y_p}{h}.$$

This shows that the covariant derivative really is a “derivative” in the direction of the curve, with parallel transport playing the role of identifying nearby tangent spaces.

We close this subsection with the one extension of D_t that the rest of the thesis really runs on. Almost every estimate below compares two curves, and the way to compare them is to join them by a family of curves and differentiate in the family parameter. Such a family is a smooth map $\Gamma : A \times I \rightarrow M$ on a product of intervals, and it carries two velocity fields,

$$\partial_t \Gamma(s, t) \in T_{\Gamma(s, t)} M \quad (\text{along the curve } \Gamma(s, \cdot)), \quad \partial_s \Gamma(s, t) \in T_{\Gamma(s, t)} M \quad (\text{across the family}).$$

For fixed s the operator D_t of Lemma 2.9 differentiates vector fields along the curve $t \mapsto \Gamma(s, t)$, and symmetrically D_s differentiates along $s \mapsto \Gamma(s, t)$; both are applied to fields defined on the whole rectangle. The two mixed derivatives then agree.

Lemma 2.13 (Symmetry lemma). *Let $\Gamma : A \times I \rightarrow M$ be smooth, where $A, I \subseteq \mathbb{R}$ are intervals, and let ∇ be the Levi-Civita connection. Then*

$$D_s \partial_t \Gamma = D_t \partial_s \Gamma.$$

Proof. The claim is local, so work in a chart. Writing $\partial_t \Gamma = \partial_t \Gamma^k \partial_k$ and using the coordinate formula for D_s from Lemma 2.9,

$$D_s \partial_t \Gamma = (\partial_s \partial_t \Gamma^k + \Gamma_{ij}^k \partial_s \Gamma^i \partial_t \Gamma^j) \partial_k.$$

The first term is symmetric in s and t because ordinary partial derivatives of the smooth functions Γ^k commute, and the second is symmetric because $\Gamma_{ij}^k = \Gamma_{ji}^k$, which is exactly the torsion-freeness of the Levi-Civita connection. See also Lee [7, Ch. 6]. $\square$

Torsion-freeness enters the thesis only through this lemma, and the lemma itself is used constantly: the symmetry lemma turns the derivative of a distance, or of a separation between two solutions of an ODE, into something the geodesic equation, the flow of the vector field or the curvature tensor can control.

2.4 The exponential map and normal coordinates

Parallel transport captures the behaviour of the velocity vector along a single geodesic, but we would like a map that carries information about all geodesics emanating from a point. Since for any $p \in M$ and $v \in T_p M$ there is a unique maximal geodesic γ_v with $\gamma_v(0) = p$ and $\dot{\gamma}_v(0) = v$, we can define:

Definition 2.14 (Exponential map). Let $\mathcal{E}_p = \{v \in T_p M : \gamma_v \text{ is defined on } [0, 1]\}$. The *exponential map* at p is

$$\text{Exp}_p : \mathcal{E}_p \rightarrow M, \quad \text{Exp}_p(v) = \gamma_v(1).$$

The set $\mathcal{E}_p$ is open and star-shaped, and Exp_p is smooth on it. A key property is that for any $v \in T_p M$ and $c, t \in \mathbb{R}$, $\gamma_{cv}(t) = \gamma_v(ct)$ whenever either side is defined; in particular $\text{Exp}_p(tv) = \gamma_v(t)$, so the geodesic from p with initial velocity v is simply $t \mapsto \text{Exp}_p(tv)$ (Figure 1).

Lemma 2.15 (Normal Neighbourhood Lemma). *For any $p \in M$, there is a neighbourhood $\mathcal{V}$ of 0 in $T_p M$ and a neighbourhood $\mathcal{U}$ of p in M such that $\text{Exp}_p : \mathcal{V} \rightarrow \mathcal{U}$ is a diffeomorphism.*

Proof. By the inverse function theorem, it suffices to check that $(d\text{Exp}_p)_0 = \text{id}_{T_p M}$. For $v \in T_p M$, take the curve $\tau(t) = tv$ in $T_p M$. Then $(d\text{Exp}_p)_0(v) = \frac{d}{dt} \big|_{t=0} \text{Exp}_p(tv) = \frac{d}{dt} \big|_{t=0} \gamma_v(t) = v$. $\square$

On such a neighbourhood Exp_p is invertible; its inverse is the *logarithm map* $\log_p := \text{Exp}_p^{-1}$, so that $\log_p q \in T_p M$ is the initial velocity of the unit-time geodesic from p to q . We use this notation, together with the fact that $(q, p) \mapsto \log_q p$ is smooth, in Lemma 2.29.

Definition 2.16 (Normal coordinates). Fix a g -orthonormal basis $\{E_i\}$ of $T_p M$, defining an isomorphism $E : \mathbb{R}^d \rightarrow T_p M$ by $E(x^1, \dots, x^d) = x^i E_i$. If $\mathcal{U}$ is a normal neighbourhood of p , the *normal coordinate chart* centred at p is $\varphi = E^{-1} \circ \text{Exp}_p^{-1} : \mathcal{U} \rightarrow \mathbb{R}^d$.

Normal coordinates are convenient because several quantities simplify at the centre point p : $g_{ij}(p) = \delta_{ij}$, the Christoffel symbols vanish ($\Gamma_{ij}^k(p) = 0$), and geodesics through p appear as straight lines ($\gamma_v(t) = (tv^1, \dots, tv^d)$).

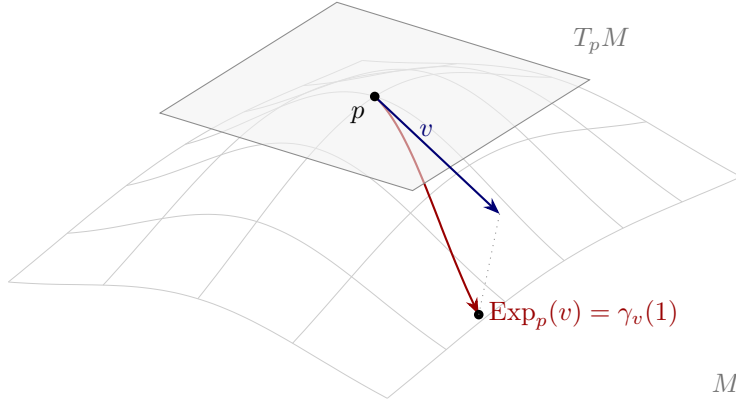


Figure 1: The exponential map. A tangent vector $v \in T_p M$ determines the geodesic $t \mapsto \gamma_v(t)$ leaving p with velocity v ; following it for unit time gives $\text{Exp}_p(v) = \gamma_v(1) \in M$. Near the origin Exp_p is a diffeomorphism; reading M through this inverse is exactly the normal coordinate chart.

Definition 2.17 (Injectivity radius). The *injectivity radius* at p is

$$\iota(p) = \sup\{r > 0 : \text{Exp}_p|_{B(0,r)} \text{ is a diffeomorphism onto its image}\}.$$

For a subset $A \subseteq M$ we write $\text{inj}(A) := \inf_{p \in A} \iota(p)$, and $\text{inj}(M)$ when $A = M$. The image $B(p, \varepsilon) = \text{Exp}_p(B(0, \varepsilon))$ of a ball of radius $\varepsilon \leq \iota(p)$ is called a *geodesic ball* around p .

We will see the number $\iota(p)$ return in almost every subsequent part of this thesis: it is the radius within which $\log_p$ exists, within which two points are joined by exactly one minimising geodesic, and therefore within which the parallel transport $P_{q \rightarrow p}$ of Subsection 2.6 is defined at all. The facts we need about it are collected in Lemma 2.24 and Lemma 2.26 below, once distance is available.

2.5 Riemannian distance and completeness

So far, our Riemannian metric has only given us a notion of distance in each tangent space. We now define distance on the manifold itself.

2.5.1 Length and distance

Definition 2.18 (Admissible curve and length). A curve $\gamma : [a, b] \rightarrow M$ is *admissible* if it is continuous and there is a partition $a = t_0 < t_1 < \dots < t_k = b$ such that each restriction $\gamma|_{[t_{i-1}, t_i]}$ is smooth with nowhere vanishing velocity. Its *length* is

$$L_g(\gamma) = \int_a^b \|\dot{\gamma}(t)\| dt.$$

The nonvanishing velocity does not change the infimum defining d , and it lets every admissible curve be reparametrised by arc length. Length is invariant under every reparametrisation, so minimising is a property of the image of a curve. The geodesic equation is not invariant: it forces constant speed, and only affine reparametrisations preserve it. A minimiser can therefore be a geodesic only after its parametrisation is normalised.

Definition 2.19 (Riemannian distance). For $p, q \in M$,

$$d(p, q) = \inf\{L_g(\gamma) : \gamma \text{ admissible from } p \text{ to } q\}.$$

This turns any connected Riemannian manifold into a metric space whose topology agrees with the manifold topology. Now that we have a distance, one might hope that the length-minimising curves between two points are precisely the geodesics, but globally this fails: a geodesic can run past the point where it stops being shortest, as the long way around a great circle shows (Figure 2). What does hold is a two-sided statement. Every minimising curve is a geodesic once reparametrised by arc length, and every geodesic minimises length between sufficiently close points on it.

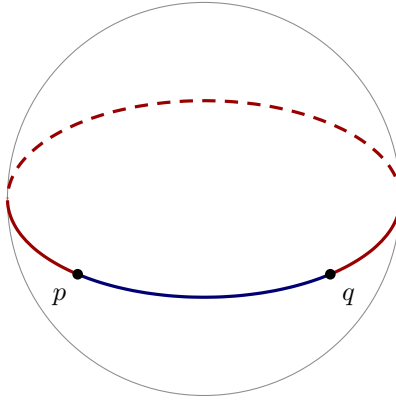


Figure 2: Geodesics need not minimise. The points p and q on the sphere lie on a common great circle, which splits into two geodesic arcs between them. The short arc (blue) is the minimising geodesic realising $d(p, q)$; the “long way round” (red) is still a geodesic but is far from length-minimising. We will see geodesics are only *locally* minimising.

2.5.2 Minimising curves and the first variation of length

The first half of this statement comes from differentiating length along a family of competing curves. Two parameters appear: t moves along a single curve, while s selects which curve of the family is being looked at.

Definition 2.20 (Variation and variation field). Let $\gamma : [a, b] \rightarrow M$ be admissible with breakpoints $a = t_0 < \dots < t_k = b$. A *variation* of γ is a continuous map

$$\Gamma : (-\varepsilon, \varepsilon) \times [a, b] \rightarrow M, \quad \Gamma(0, t) = \gamma(t),$$

which is smooth on each rectangle $(-\varepsilon, \varepsilon) \times [t_{i-1}, t_i]$. Writing $\Gamma_s := \Gamma(s, \cdot)$, each Γ_s is required to be admissible, and $\Gamma_0 = \gamma$. The variation is *proper* if all the curves Γ_s share the endpoints of γ , that is $\Gamma(s, a) = \gamma(a)$ and $\Gamma(s, b) = \gamma(b)$ for every s . Its *variation field* is the vector field along γ given by

$$V(t) := \partial_s \Gamma(0, t) \in T_{\gamma(t)} M,$$

the initial velocity with which the deformation moves the point $\gamma(t)$.

So $\partial_t \Gamma$ is the velocity of the curve Γ_s and $\partial_s \Gamma$ is the velocity of the deformation; along γ itself these are $\dot{\gamma}$ and V . The field V is continuous on $[a, b]$ and smooth on each $[t_{i-1}, t_i]$, and a variation is proper precisely when $V(a) = V(b) = 0$. Every piecewise smooth field along γ vanishing at the endpoints is therefore an admissible direction of perturbation and quantifying over proper variations is the same as quantifying over such fields. Figure 3 shows the two parameters at work.

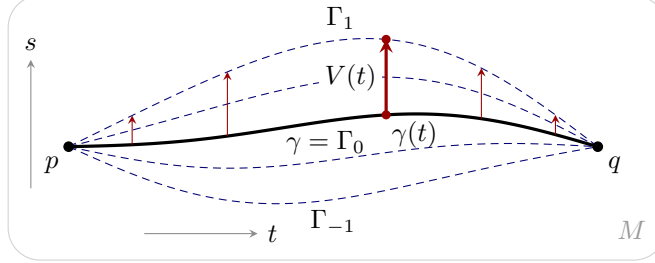


Figure 3: A proper variation of γ . Moving along a curve means varying t ; passing from one curve of the family to another means varying s , and $\Gamma_0 = \gamma$. The variation is proper because every member runs from p to q , which forces the variation field $V(t) = \partial_s \Gamma(0, t)$ to vanish at the two endpoints. At the marked parameter t the field $V(t)$ is the velocity, in s , at which the deformation pushes the point $\gamma(t)$ off the curve; here the family is affine in s , so the arrow reaches $\Gamma_1(t)$ exactly.

Proposition 2.21 (First variation of length). *Let $\gamma : [a, b] \rightarrow M$ be admissible of unit speed, with breakpoints $t_1, \dots, t_{k-1}$ in (a, b) , and let Γ be a proper variation of γ with variation field V . Then $s \mapsto L_g(\Gamma_s)$ is differentiable near $s = 0$ and*

$$\left. \frac{d}{ds} \right|_{s=0} L_g(\Gamma_s) = - \int_a^b \langle V, D_t \dot{\gamma} \rangle dt - \sum_{i=1}^{k-1} \langle V(t_i), \Delta_i \dot{\gamma} \rangle, \quad \Delta_i \dot{\gamma} := \dot{\gamma}(t_i^+) - \dot{\gamma}(t_i^-),$$

where D_t denotes covariant differentiation along the curve.

Proof. Write $T := \partial_t \Gamma$ and $S := \partial_s \Gamma$. Since $\|T(0, t)\| = 1$ for all t and $[a, b]$ is compact, $\|T\| \geq \frac{1}{2}$ on each rectangle for $|s|$ small; there $\|T\| = \langle T, T \rangle^{1/2}$ is smooth and differentiation under the integral sign is justified. The Levi-Civita connection is symmetric, so $D_s T = D_t S$ by Lemma 2.13, and hence

$$\partial_s \|T\| = \frac{\langle D_s T, T \rangle}{\|T\|} = \frac{\langle D_t S, T \rangle}{\|T\|},$$

which at $s = 0$ equals $\langle D_t V, \dot{\gamma} \rangle$. On each piece, metric compatibility gives $\langle D_t V, \dot{\gamma} \rangle = \partial_t \langle V, \dot{\gamma} \rangle - \langle V, D_t \dot{\gamma} \rangle$, so

$$\left. \frac{d}{ds} \right|_{s=0} L_g(\Gamma_s) = \sum_{i=1}^k \left(\left[\langle V, \dot{\gamma} \rangle \right]_{t_{i-1}}^{t_i} - \int_{t_{i-1}}^{t_i} \langle V, D_t \dot{\gamma} \rangle dt \right).$$

Because V is continuous while $\dot{\gamma}$ may jump, the boundary terms telescope to $\langle V(b), \dot{\gamma}(b) \rangle - \langle V(a), \dot{\gamma}(a) \rangle - \sum_{i=1}^{k-1} \langle V(t_i), \Delta_i \dot{\gamma} \rangle$, and the two endpoint terms vanish since the variation is proper. $\square$

Corollary 2.22. *An admissible unit-speed curve $\gamma : [a, b] \rightarrow M$ has vanishing first variation of length for every proper variation if and only if γ is a geodesic. In particular, every minimising admissible curve is a geodesic once reparametrised by arc length.*

Sketch. A geodesic is smooth with $D_t\dot{\gamma} \equiv 0$, so both terms of Proposition 2.21 vanish. Conversely, the two terms are killed by two choices of variation field. Taking $V := \varphi D_t\dot{\gamma}$ for a bump function φ supported in one smooth piece leaves $\int \varphi \|D_t\dot{\gamma}\|^2 = 0$, so each piece is a geodesic; taking V equal to $\Delta_i\dot{\gamma}$ at the breakpoint t_i and vanishing at the others leaves $\|\Delta_i\dot{\gamma}\|^2 = 0$, so the velocity does not jump and the pieces fit together into one geodesic. For the last claim, a minimiser makes $s = 0$ a minimum of the differentiable function $s \mapsto L_g(\Gamma_s)$ for every proper variation, so its derivative there vanishes. Details are in Lee [7, Ch. 6]. $\square$

Geodesics are thus exactly the critical points of L_g among unit-speed admissible curves with fixed endpoints, and minimisers are critical. This settles the first half of the two-sided statement. The second half, that a geodesic minimises between sufficiently close points on it, is not a first-order statement and is proved below using normal neighbourhoods and the Gauss lemma.

The proper variations of Proposition 2.21 keep the endpoints fixed, which is what makes the boundary terms disappear. The proofs of this thesis need a more general setting. There the two curves being compared are the noise path and the drifted path, both of which move, and the whole point is to bound the rate at which the distance between them grows. Since the connecting geodesics can then be taken as the family, the interior term disappears instead of the boundary terms, and what remains is the following.

Proposition 2.23 (First variation with moving endpoints). *Let $\Gamma : I \times [0, 1] \rightarrow M$ be smooth and suppose that for each t the curve $s \mapsto \Gamma(t, s)$ is a geodesic of constant speed. Write $T := \partial_s \Gamma$ and $W := \partial_t \Gamma$, and let*

$$\delta(t) := \|T(t, s)\| = L_g(\Gamma(t, \cdot)),$$

the two being equal because the speed is constant and s runs over $[0, 1]$. So $\delta(t)$ is exactly the separation of the two ends of the family at time t . If $\delta(t) > 0$, then δ is smooth near t and

$$\delta(t) \delta'(t) = \langle W(t, 1), T(t, 1) \rangle - \langle W(t, 0), T(t, 0) \rangle.$$

So δ measures how far apart the two ends are, while $\|W\|$ measures how fast they move. The identity says that the separation changes only through the components of the two end velocities along the connecting geodesic, and by Cauchy–Schwarz with $\|T(t, 0)\| = \|T(t, 1)\| = \delta(t)$ it gives $|\delta'| \leq \|W(t, 0)\| + \|W(t, 1)\|$.

Proof. Since $\|T(t, s)\| = \delta(t)$ does not depend on s , we have $\delta(t)^2 = \int_0^1 \|T(t, s)\|^2 ds$. Differentiating under the integral sign and using metric compatibility,

$$\delta(t) \delta'(t) = \frac{1}{2} \frac{d}{dt} \delta(t)^2 = \int_0^1 \langle D_t T, T \rangle ds.$$

By Lemma 2.13, $D_t T = D_t \partial_s \Gamma = D_s \partial_t \Gamma = D_s W$, so metric compatibility applied once more in the s direction gives

$$\langle D_s W, T \rangle = \partial_s \langle W, T \rangle - \langle W, D_s T \rangle = \partial_s \langle W, T \rangle,$$

because each slice $\Gamma(t, \cdot)$ is a geodesic and hence $D_s T = 0$. Integrating $\partial_s \langle W, T \rangle$ over $s \in [0, 1]$ gives the two boundary terms. $\square$

2.5.3 Geodesic balls

For the converse, that geodesics are minimising at least locally, the key tool is the Gauss Lemma: radial geodesics from p meet the geodesic spheres around p orthogonally, the geodesic spheres being the images under Exp_p of the Euclidean spheres in T_pM . See Lee [7, Ch. 6] for the statement in terms of $d\text{Exp}_p$ and for its proof. We record only the consequences, in the form in which the later sections use them: inside a geodesic ball the manifold behaves like its own tangent space, in that distance from the centre is read off from the norm and the chart $\log_p$ carries metric balls to Euclidean balls.

Lemma 2.24 (Geodesic balls). *Let $p \in M$ and $0 < \varepsilon \leq \iota(p)$. Then for every $v \in T_pM$ with $\|v\| < \varepsilon$,*

$$d(p, \text{Exp}_p(v)) = \|v\|,$$

and the radial geodesic $s \mapsto \text{Exp}_p(sv)$, $s \in [0, 1]$, is the unique minimising curve from p to $\text{Exp}_p(v)$ up to reparametrisation. Consequently:

- (i) $B(p, \varepsilon) = \text{Exp}_p(B(0, \varepsilon))$, and in the normal coordinates of p the metric ball $B(p, \varepsilon)$ is the Euclidean ball of radius ε ; in particular it is convex in that chart.
- (ii) $\log_p$ is defined on $B(p, \iota(p))$ and satisfies $\|\log_p q\| = d(p, q)$.
- (iii) If $d(q, p) < \iota(q)$ then q and p are joined by a unique minimising geodesic, so parallel transport along it is well defined.

Sketch. Both claims are the standard consequence of the Gauss lemma quoted above. For a competing curve σ from p , let $r(t) := \|\log_p \sigma(t)\|$ be its radial coordinate in the normal chart. The Gauss lemma splits $\dot{\sigma}$ into orthogonal radial and angular parts, so $\|\dot{\sigma}\| \geq |r'|$ and hence $L_g(\sigma) \geq \int |r'| \geq \|v\|$, with equality only when the angular part vanishes and r is nondecreasing. Item (i) restates $d(p, \text{Exp}_p v) = \|v\|$ together with the definition of the normal chart, (ii) is the same identity read through $\log_p = \text{Exp}_p^{-1}$, and (iii) applies the uniqueness statement at the centre q . $\square$

2.5.4 Completeness and the injectivity radius

Local minimisation says nothing about whether a minimising geodesic between two given points exists at all, or whether geodesics can be continued indefinitely. These are global questions, and both are governed by completeness through the Hopf–Rinow theorem.

Theorem 2.25 (Hopf–Rinow). *Let (M, g) be a connected Riemannian manifold. The following are equivalent:*

1. M is geodesically complete: every maximal geodesic is defined on all of $\mathbb{R}$.
2. (M, d) is metrically complete: every Cauchy sequence converges.
3. Closed and bounded subsets of (M, d) are compact.

Moreover, if any of these hold, then any two points can be joined by a length-minimising geodesic.

On a compact manifold, all three conditions are automatic. For a proof see Lee [7, Ch. 6], where the theorem is stated in this equivalent-conditions form.

Hopf–Rinow is what lets us work on compact subsets of a complete manifold: a closed metric ball is then compact, and every uniform constant of the later sections is a supremum or infimum over such a ball. The injectivity radius is the first of those constants.

Lemma 2.26 (Uniform injectivity radius). *Let (M, g) be complete. The function $p \mapsto \iota(p) \in (0, \infty]$ is continuous. Consequently $\text{inj}(K) = \inf_{p \in K} \iota(p) > 0$ for every compact $K \subseteq M$, and $\text{inj}(M) > 0$ when M is compact.*

Proof. See Sakai [15, Ch. III, Prop. 4.13]. □

2.5.5 Absolutely continuous paths

Finally, the paths that carry the large deviation principle are not piecewise smooth, so the class of admissible curves above is too small for the rate functions of Section 6. We record the class that replaces it.

Definition 2.27 (Absolutely continuous curve in M). A curve $\psi : [0, 1] \rightarrow M$ is *absolutely continuous*, written $\psi \in \text{AC}([0, 1]; M)$, if for every chart (U, φ) of M and every interval $J \subseteq \psi^{-1}(U)$ the composition $\varphi \circ \psi|_J$ is absolutely continuous as an $\mathbb{R}^d$ -valued curve.

Lemma 2.28 (Velocity, length and speed). *Let $\psi \in \text{AC}([0, 1]; M)$. Then:*

- (i) ψ is differentiable at almost every t , with velocity $\dot{\psi}(t) \in T_{\psi(t)}M$, and $t \mapsto \|\dot{\psi}(t)\|$ is measurable;
- (ii) $d(\psi(t_1), \psi(t_2)) \leq \int_{t_1}^{t_2} \|\dot{\psi}(t)\| dt$ for $t_1 \leq t_2$, and the length $L_g(\psi) := \int_0^1 \|\dot{\psi}\| dt$ extends L_g from the admissible curves;
- (iii) if $\|\dot{\psi}\|_\infty \leq r$ then ψ is r -Lipschitz from $([0, 1], |\cdot|)$ to (M, d) ; in particular $\psi([0, 1]) \subseteq \overline{B}(\psi(0), r)$.

We write $\text{AC}_{x_0, r}([0, 1]; M)$ for the absolutely continuous paths with $\psi(0) = x_0$ and $\|\dot{\psi}\|_\infty \leq r$. By (iii) this class is uniformly Lipschitz and confined to $\overline{B}(x_0, r)$, which by Hopf–Rinow is compact. This pair of properties is what makes the class the natural domain of the drift map in Section 8.

2.6 Smooth dependence of parallel transport on its endpoints

Parallel transport was defined along a fixed curve. For the rest of the thesis we need parallel transport along the *unique minimising geodesic* between two nearby points, together with the fact that it varies smoothly as the endpoints move. For $q, p \in M$ with p in a normal neighbourhood of q (e.g. $d(q, p) < \iota(q)$), let $\gamma_{q, p}(s) := \text{Exp}_q(s \log_q p)$, $s \in [0, 1]$, be that geodesic and write

$$P_{q \rightarrow p} := \tau_{0, 1}^{\gamma_{q, p}} : T_q M \rightarrow T_p M$$

for parallel transport along it. This is defined on the open set

$$\mathcal{U} := \{(q, p) \in M \times M : p \text{ lies in a normal neighbourhood of } q\},$$

on which $(q, p) \mapsto \log_q p$ is smooth, being the inverse of the diffeomorphism $(q, v) \mapsto (q, \text{Exp}_q v)$.

Lemma 2.29 (Smoothness of parallel transport in the endpoints). *For any smooth local frames (E_i) near q_0 and (F_a) near p_0 , with $(q_0, p_0) \in \mathcal{U}$, the component functions P^a_i determined by*

$$P_{q \rightarrow p} E_i(q) = P^a_i(q, p) F_a(p)$$

are smooth in (q, p) on a neighbourhood of (q_0, p_0) . In particular, for fixed q the map $p \mapsto P_{q \rightarrow p} w$ is smooth for each $w \in T_q M$, and for each $f \in C^1(M)$ the map $(q, p) \mapsto df_p \circ P_{q \rightarrow p}$ is continuous on $\mathcal{U}$.

Proof. The map $(s, q, p) \mapsto \gamma_{q,p}(s)$ is smooth on $[0, 1] \times \mathcal{U}$, since Exp is smooth and $(q, p) \mapsto \log_q p$ is smooth on $\mathcal{U}$. Fix $(q_0, p_0) \in \mathcal{U}$ and write $\sigma_0 := \gamma_{q_0, p_0}$, which has compact image. Choose a partition $0 = s_0 < s_1 < \dots < s_N = 1$ and charts (Ω_k, φ_k) with $\sigma_0([s_{k-1}, s_k]) \subset \Omega_k$. By continuity of $(s, q, p) \mapsto \gamma_{q,p}(s)$ there is a neighbourhood $\mathcal{N} \ni (q_0, p_0)$ with $\gamma_{q,p}([s_{k-1}, s_k]) \subset \Omega_k$ for all $(q, p) \in \mathcal{N}$ and all k ; the partition is now fixed. Parallel transport composes over the partition, $P_{q \rightarrow p} = P^{(N)} \circ \dots \circ P^{(1)}$, where $P^{(k)}$ transports along $\gamma_{q,p}|_{[s_{k-1}, s_k]}$. In the chart φ_k the components $X(s) = (X^l(s))$ of a parallel field solve

$$\dot{X}^l(s) = -\Gamma_{ij}^{(k)l}(\gamma_{q,p}(s)) \dot{\gamma}_{q,p}^i(s) X^j(s) =: A_{(k)j}^l(s; q, p) X^j(s),$$

with $A_{(k)}(\cdot; q, p)$ smooth in (s, q, p) on $[s_{k-1}, s_k] \times \mathcal{N}$ (smoothness of the Christoffel symbols and of $\gamma_{q,p}$). Each column of the fundamental matrix solves a linear system $\dot{X} = A_{(k)}(s; q, p) X$ with parameter (q, p) ; being linear, it has a solution on all of $[s_{k-1}, s_k]$. To obtain smooth dependence on (q, p) , adjoin the parameter as a constant variable μ : the augmented system

$$\dot{X} = A_{(k)}(s; \mu) X, \quad \dot{\mu} = 0, \quad \mu(s_{k-1}) = (q, p),$$

has a smooth right-hand side, so by the smooth dependence on initial data of Teschl [17, Thm. 2.10] its solution is smooth in the initial value $(X(s_{k-1}), \mu(s_{k-1}))$, in particular in (q, p) . Hence the coordinate matrix of $P^{(k)}$ is smooth in $(q, p) \in \mathcal{N}$. Finally $P_{q \rightarrow p}$ is the product of the matrices of the $P^{(k)}$, conjugated by the change-of-frame Jacobians at the junctions $\gamma_{q,p}(s_k)$ and at q, p ; these are smooth in position, evaluated at points depending smoothly on (q, p) . A product of (q, p) -smooth matrices is smooth, so P^a_i is smooth on $\mathcal{N}$. As (q_0, p_0) was arbitrary, the claim follows; the two consequences come from fixing q , resp. contracting with the continuous coefficients $df_p(F_a(p))$ for $f \in C^1$. $\square$

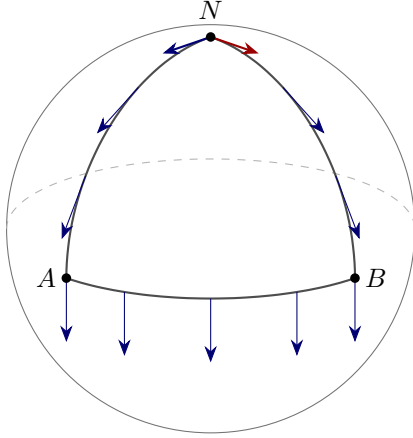
2.7 Curvature

The curvature of a Riemannian manifold measures the extent to which parallel transport depends on the path. It is encoded by the Riemann curvature tensor. Concretely, transporting a vector around a small closed loop (Figure 4) and measuring how far it returns rotated is, to leading order, the same as measuring the failure of the second covariant derivatives $\nabla_X \nabla_Y$ and $\nabla_Y \nabla_X$ to commute. The definition below turns this observation into a tensor.

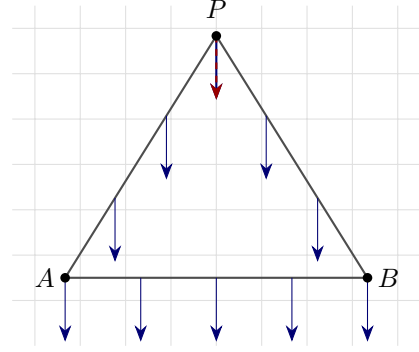
Definition 2.30 (Riemann curvature tensor). The *Riemann curvature endomorphism* is the $(1, 3)$ -tensor

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

The associated $(0, 4)$ -tensor is $\text{Rm}(X, Y, Z, W) = \langle R(X, Y)Z, W \rangle$.



(a) Curved: the sphere.



(b) Flat: a plane patch.

Figure 4: Curvature as the path-dependence of parallel transport. Carrying a tangent vector once around the geodesic triangle $N \rightarrow A \rightarrow B \rightarrow N$ on the sphere returns it rotated relative to the starting vector: the intermediate arrows show it carried along each edge, staying at a constant angle to that edge, so that “start” (blue) and “returned” (red) differ by the angle of the triangle at N . On the right the same experiment in a flat patch, around the plane triangle $P \rightarrow A \rightarrow B \rightarrow P$: parallel transport there keeps the vector constant, so every arrow is the same arrow and the returned vector (red, dashed) lies exactly on top of the starting one.

In local coordinates, $R(\partial_i, \partial_j)\partial_k = R^\ell_{kij} \partial_\ell$ with $R^\ell_{kij} = \partial_i \Gamma^\ell_{jk} - \partial_j \Gamma^\ell_{ik} + \Gamma^m_{jk} \Gamma^\ell_{im} - \Gamma^m_{ik} \Gamma^\ell_{jm}$.

A Riemannian manifold is *flat* if $R \equiv 0$, which is equivalent to being locally isometric to Euclidean space. The curvature tensor satisfies a number of symmetries (skew-symmetry in the first and last pair of indices, the pair-swap symmetry, and the algebraic Bianchi identity); see [7, Proposition 7.12].

The quantity the estimates below use is the size of R as a trilinear map. We therefore set

$$\|\text{Rm}(p)\|_{\text{op}} := \sup\{\|R(u, v)w\| : u, v, w \in T_p M, \|u\|, \|v\|, \|w\| \leq 1\}, \quad \|\text{Rm}\|_\infty := \sup_{p \in M} \|\text{Rm}(p)\|_{\text{op}},$$

so that by homogeneity

$$\|R(u, v)w\| \leq \|\text{Rm}\|_\infty \|u\| \|v\| \|w\| \quad \text{for all } u, v, w \in T_p M, p \in M. \quad (1)$$

On a compact manifold $\|\text{Rm}\|_\infty < \infty$. We write $\kappa := \|\text{Rm}\|_\infty$ when this constant appears in an estimate. The hypothesis $\|\text{Rm}\|_\infty < \infty$ on a non-compact manifold is one half of the bounded-geometry assumption of Section 9, the other half being $\text{inj}(M) > 0$.

2.8 Jacobi fields and curvature bounds

The last piece of geometry we need records how curvature bends families of nearby geodesics. Positive curvature makes geodesics that leave a point with almost the same initial velocity bend back towards one another, while negative curvature makes them spread apart (Figure 5).

Jacobi fields linearise this effect. Fix $v, w \in T_p M$ and consider the one-parameter family of geodesics leaving p ,

$$\gamma_s(t) = \text{Exp}_p(t(v + sw)),$$

with reference geodesic $\gamma := \gamma_0$. Differentiating in the family parameter gives a vector field along γ ,

$$J(t) = \frac{\partial}{\partial s} \Big|_{s=0} \gamma_s(t),$$

which records the first-order drift of the neighbouring geodesic away from γ : at time t the two are at distance $s\|J(t)\| + O(s^2)$, and $J(t)$ is based at $\gamma(t)$ and points towards γ_s at that same time. Differentiating the geodesic equation in s , and exchanging the two derivatives with Lemma 2.13, shows that J satisfies a linear second-order equation in which the curvature enters as the coefficient. We take that equation as the definition.

Definition 2.31 (Jacobi field). A vector field J along a geodesic γ is a *Jacobi field* if it satisfies the Jacobi equation

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

Conversely, every Jacobi field arises as above from a family of geodesics, so the definition captures exactly the infinitesimal behaviour of nearby geodesic families. Since the equation is linear and of second order, a Jacobi field is determined by $J(0)$ and $D_t J(0)$, and the Jacobi fields along a fixed geodesic form a $2d$ -dimensional space.

For constant sectional curvature K , that is $R(X, Y)Y = KX$ for orthonormal X, Y , a unit-speed γ and $J \perp \dot{\gamma}$ the equation reduces to $\|J\|'' + K\|J\| = 0$, whose solutions with $J(0) = 0$ and $\|D_t J(0)\| = 1$ are $\|J(t)\| = \sin t$ for $K \equiv 1$ and $\|J(t)\| = \sinh t$ for $K \equiv -1$, the two panels of Figure 5. At speed $\|\dot{\gamma}\| = \sigma$ the coefficient becomes $K\sigma^2$, which is the form the constant $\sqrt{\kappa}\sigma$ of Lemma 2.32 takes.

The family above starts with $J(0) = 0$ and $D_t J(0) = w$, and the same derivative computes the differential of the exponential map,

$$J(t) = d(\text{Exp}_p)_{tv}(tw).$$

We use this identity only to see what limits the injectivity radius. $d(\text{Exp}_p)_{tv}$ is singular exactly when some nontrivial Jacobi field vanishing at p vanishes again at $\gamma(t)$; such a $\gamma(t)$ is called *conjugate* to p . It is a point where Exp_p ceases to be a local diffeomorphism, and the point $q = \gamma(\pi)$ in the left panel of Figure 5 is of this kind.

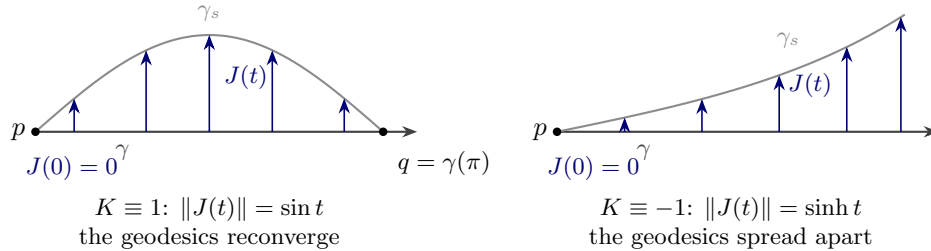


Figure 5: Jacobi fields and the spreading of geodesics. In each panel the straight line is a geodesic γ leaving p and the second curve is a neighbouring geodesic γ_s of the family, with almost the same initial velocity. The blue arrow at time t is the Jacobi field $J(t)$, whose length is the separation of the two geodesics.

The later sections need a bound on how fast a Jacobi field can grow when the curvature is bounded.

Lemma 2.32 (Growth of Jacobi fields). *Let (M, g) be complete with $\kappa := \|\text{Rm}\|_\infty < \infty$, let $\gamma : [0, 1] \rightarrow M$ be a geodesic of constant speed $\|\dot{\gamma}\| = \sigma$, and let J be a Jacobi field along γ with $D_t J(0) = 0$. Then*

$$\|J(t)\| \leq e^{\sqrt{\kappa}\sigma t} \|J(0)\| \quad (t \in [0, 1]).$$

Proof. Set $a(t) := \|J(t)\|$ and $b(t) := \|D_t J(t)\|$. Both are Lipschitz, being norms of smooth fields, so both are differentiable almost everywhere, and where the field inside is nonzero, Cauchy–Schwarz gives $|a'| \leq b$ and $|b'| \leq \|D_t^2 J\|$. By the Jacobi equation and (1),

$$\|D_t^2 J\| = \|R(J, \dot{\gamma})\dot{\gamma}\| \leq \kappa \sigma^2 a.$$

If $\kappa\sigma = 0$ then $D_t J \equiv 0$, so a is constant and the claim is trivial. Otherwise put $\varphi := \sqrt{\kappa}\sigma a + b$, so that $\varphi(0) = \sqrt{\kappa}\sigma a(0)$ and, almost everywhere,

$$\varphi' \leq \sqrt{\kappa}\sigma b + \kappa\sigma^2 a = \sqrt{\kappa}\sigma \varphi.$$

Grönwall gives $\varphi(t) \leq \varphi(0)e^{\sqrt{\kappa}\sigma t}$, and $\sqrt{\kappa}\sigma a(t) \leq \varphi(t)$ finishes the proof. $\square$

The geometric reading of the lemma is a stability statement for the exponential map in its base point, which the truncation argument of Section 9 uses: if $d(q, p) < \iota(q)$, if $\tau := P_{q \rightarrow p}$ and if $w \in T_q M$, then applying Lemma 2.32 to the family $\Gamma(s, t) := \text{Exp}_{\gamma(s)}(tW(s))$, where γ is the minimising geodesic from q to p and W is the parallel field along it with $W(0) = w$, gives

$$d(\text{Exp}_q(w), \text{Exp}_p(\tau w)) \leq e^{\sqrt{\kappa}\|w\|} d(q, p).$$

Two geodesic steps launched from nearby points with parallel initial velocities therefore separate by at most a factor exponential in the length of the step. This is Lemma 9.11, which is where the bound is used and where the family above is written out.

The remaining estimate is one that needs no curvature bound at all, only compactness. It controls Exp_p in its *vector* argument, with the base point held fixed. The two together are the division of labour that the later sections rely on: Lemma 2.32 controls Exp in the base point, at a cost exponential in the curvature bound, and Lemma 2.33 controls it in the vector argument, at a cost that is a single constant on a compact set.

Lemma 2.33 (Uniform Lipschitz bound for Exp). *Let (M, g) be complete, let $K \subseteq M$ be compact, and let $\rho > 0$. There is a constant $C_{\text{exp}}(K, \rho) > 0$ such that for all $p \in K$ and all $u, v \in T_p M$ with $\|u\|, \|v\| \leq \rho$,*

$$d(\text{Exp}_p(u), \text{Exp}_p(v)) \leq C_{\text{exp}}(K, \rho) \|u - v\|.$$

Proof. Consider the closed disk bundle of radius ρ over K ,

$$D := \{(p, u) \in TM : p \in K, \|u\| \leq \rho\}.$$

This is compact, as a closed and fibrewise bounded subset of TM over the compact base K . Since M is complete, geodesics exist for all time, so the global exponential map $\text{Exp} : TM \rightarrow M$, $(p, u) \mapsto \text{Exp}_p(u)$, is smooth on all of TM . Working in local trivialisations of TM , the assignment $(p, u) \mapsto \|(d\text{Exp}_p)_u\|_{\text{op}}$ is continuous (it is the operator norm of a smoothly varying linear map between fibres of TM , both equipped with the smoothly varying Riemannian inner product). By compactness of D ,

$$C_{\text{exp}}(K, \rho) := \sup_{(p, u) \in D} \|(d\text{Exp}_p)_u\|_{\text{op}} < \infty.$$

Fix $p \in K$ and $u, v \in T_p M$ with $\|u\|, \|v\| \leq \rho$. The ball $\overline{B}(0, \rho) \subset T_p M$ is convex, so the segment $\alpha(t) := u + t(v - u)$ satisfies $\|\alpha(t)\| \leq \rho$ for all $t \in [0, 1]$. The curve $\beta(t) := \text{Exp}_p(\alpha(t))$ in M has velocity $\dot{\beta}(t) = (d\text{Exp}_p)_{\alpha(t)}(v - u)$, so

$$\|\dot{\beta}(t)\| \leq \|(d\text{Exp}_p)_{\alpha(t)}\|_{\text{op}} \|v - u\| \leq C_{\text{exp}}(K, \rho) \|v - u\|,$$

since $(p, \alpha(t)) \in D$. The length of β is therefore at most $C_{\text{exp}}(K, \rho) \|v - u\|$, and Riemannian distance is at most the length of any connecting curve (Lemma 2.28(ii), which needs no nonvanishing-velocity hypothesis). $\square$

The compact set K and the radius ρ can be chosen freely, so the bound is available wherever the walk or the flow is known to stay in a compact set. On compact M one takes $K = M$ and $\rho = \text{inj}(M)/2$, which is the form used in Section 8; Section 9, where the walk is not confined to a fixed tube, uses it with K the tightness ball of the level in question.

This closes the geometry. Most of what precedes is background, and a reader who wants only the results that the proofs actually invoke needs four of them. Lemma 2.29 makes $P_{q \rightarrow p}$ smooth in both endpoints, which is what lets the drift equation (9) be read as an ODE and its solution map be differentiated. Lemma 2.33 controls Exp_p in its vector argument by a single constant on a compact set, and carries the one-step consistency estimates of Section 4 and Subsection 8.4. Lemma 2.32 controls Exp in its base point instead, at a cost exponential in the curvature bound, and becomes the one-step stability estimate Lemma 9.11 on which Section 9 runs. Proposition 2.23 differentiates the distance between two moving curves, and is the mechanism behind every separation bound in Section 8, beginning with Lemma 8.2. Behind all four sit the symmetry of the Levi Civita connection (Lemma 2.13), and Hopf–Rinow (Theorem 2.25), the latter making the closed balls compact and hence the constants uniform.

3 Analytic and probabilistic prerequisites

This section collects the analytic and probabilistic tools that the later sections use. Subsection 3.1 is about large deviation theory: what a large deviation principle is, the two classical theorems in $\mathbb{R}^d$ that the Riemannian results generalise, and the devices that carry a principle from one sequence to another. Subsection 3.2 collects the compactness and weak-convergence results behind the continuity of the drift map. Subsection 3.4 collects the facts about ordinary differential equations on which the drift map and the Euler scheme rest. Subsection 3.3 fixes the convex-analytic notation that the rate-function identification of Section 9 uses.

Standard results are quoted with a reference and no proof; Dembo and Zeitouni [3] is the reference for all of Subsection 3.1. The exception is Subsection 3.2.2, where two statements about absolutely continuous paths are proved, because they are assembled there from standard pieces into a combined form.

3.1 Large deviations

There are several results that help understand a stochastic process. Probably the most well known of these is the law of large numbers. While it locates the typical value of an empirical mean, it says nothing about the cost of an atypical one. Under suitable independence and moment assumptions, the probability of such a deviation decays exponentially in the number of terms, and a large deviation principle is the statement that this exponential decay has a rate, together with the identification of that rate as a single function on the state space.

One function covers every event: the cost of a Borel set is obtained by minimising over it, and where the minimum is attained the minimiser shows *how* the rare event happens. In path space that minimiser is the trajectory the process follows when it deviates, the Freidlin–Wentzell picture of the introduction. Theorem 6.3 computes such a function for the drifted walk on a manifold.

3.1.1 The large deviation principle

Let $\mathcal{X}$ be a Polish space (i.e. a separable, completely metrisable topological space) with Borel σ -algebra. We will only use $\mathcal{X} = \mathbb{R}^d$, $\mathcal{X} = M$ a Riemannian manifold, or $\mathcal{X} = C([0, 1]; \mathcal{Y})$ with the sup-norm.

Definition 3.1 (Rate function). A *rate function* on $\mathcal{X}$ is a lower semicontinuous map $I : \mathcal{X} \rightarrow [0, +\infty]$. It is called a *good rate function* if its sublevel sets $\{x \in \mathcal{X} : I(x) \leq \alpha\}$ are compact for every $\alpha \in [0, \infty)$.

Definition 3.2 (Large deviation principle). A sequence $\{\mu_n\}_{n \geq 1}$ of Borel probability measures on $\mathcal{X}$ satisfies the *large deviation principle* (LDP) with *speed* $a_n \rightarrow \infty$ and rate function I if for every Borel set $A \subseteq \mathcal{X}$,

$$-\inf_{x \in A^\circ} I(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{a_n} \log \mu_n(A) \leq \limsup_{n \rightarrow \infty} \frac{1}{a_n} \log \mu_n(A) \leq -\inf_{x \in \bar{A}} I(x).$$

A sequence of $\mathcal{X}$ -valued random variables $\{Y_n\}_{n \geq 1}$ is said to satisfy the LDP if their laws do.

Throughout this thesis the speed is always $a_n = n$.

3.1.2 Cramér and Mogulskii in $\mathbb{R}^d$

Let $X_1, X_2, \dots$ be i.i.d. random vectors in $\mathbb{R}^d$ and write $\Lambda(\lambda) = \log \mathbb{E}[e^{\langle \lambda, X_1 \rangle}]$ for their cumulant generating function. We say that X_1 satisfies *Cramér's condition* if $\Lambda(\lambda) < \infty$ for all $\lambda \in \mathbb{R}^d$. The *Fenchel–Legendre transform* of Λ is

$$\Lambda^*(w) = \sup_{\lambda \in \mathbb{R}^d} [\langle \lambda, w \rangle - \Lambda(\lambda)], \quad w \in \mathbb{R}^d.$$

Under Cramér's condition, Λ^* is a convex good rate function and vanishes precisely at $w = m := \mathbb{E}[X_1]$.

Theorem 3.3 (Cramér, [3, Theorem 2.2.30]). *Under Cramér's condition, the empirical means $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ satisfy the LDP in $\mathbb{R}^d$ with speed n and good rate function Λ^* .*

For the path-space LDP we interpolate. Let $Z^n(t)$ denote the piecewise-linear path on $[0, 1]$ obtained from $Z^n(k/n) = \frac{1}{n} \sum_{j=1}^k X_j$ by linear interpolation.

Theorem 3.4 (Mogulskii, [3, Theorem 5.1.2]). *Under Cramér's condition, the random paths Z^n satisfy the LDP on $C([0, 1]; \mathbb{R}^d)$ (with the sup-norm) at speed n with good rate function*

$$J(\psi) = \begin{cases} \int_0^1 \Lambda^*(\dot{\psi}(t)) dt & \text{if } \psi(0) = 0 \text{ and } \psi \in \text{AC}([0, 1]; \mathbb{R}^d), \\ +\infty & \text{otherwise.} \end{cases}$$

3.1.3 The contraction principle

The contraction principle transports an LDP across a continuous map. It is the workhorse of both the flat and the Riemannian proofs: applied once through the drift map and once through the time-1 evaluation.

Theorem 3.5 (Contraction principle, [3, Theorem 4.2.1]). *Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces and $f : \mathcal{X} \rightarrow \mathcal{Y}$ continuous. If $\{\mu_n\}_{n \geq 1}$ satisfies the LDP on $\mathcal{X}$ with speed a_n and good rate function I , then the pushforwards $\{\mu_n \circ f^{-1}\}$ satisfy the LDP on $\mathcal{Y}$ with the same speed and good rate function*

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

with the convention $\inf \emptyset = +\infty$.

3.1.4 Exponential equivalence

Two sequences are *exponentially equivalent* if they remain super-exponentially close in probability; in that case any LDP carries from one to the other. We use this to replace the discrete drifted walk by the continuous flow of the drift map, where the discretisation error is controlled by the geodesic Euler scheme of Section 4.

Definition 3.6 (Exponential equivalence). Two sequences $\{Y_n\}, \{\tilde{Y}_n\}$ of $\mathcal{X}$ -valued random variables (defined on a common probability space, with $\mathcal{X}$ metrised by d) are *exponentially equivalent at speed a_n* if for every $\delta > 0$,

$$\limsup_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}(d(Y_n, \tilde{Y}_n) > \delta) = -\infty.$$

Theorem 3.7 (Principle of exponential equivalence, [3, Theorem 4.2.13]). *If $\{Y_n\}$ satisfies the LDP at speed a_n with good rate function I , and $\{\tilde{Y}_n\}$ is exponentially equivalent to $\{Y_n\}$ at the same speed, then $\{\tilde{Y}_n\}$ also satisfies the LDP at speed a_n with the same good rate function I .*

3.1.5 Exponential tightness

Exponential tightness confines a family to a compact set, up to a probability that decays at the large-deviation rate. It is what upgrades a weak large deviation principle, in which the upper bound is only known on compacta, to a full one. Under bounded increments the support bound makes it unnecessary (Remark 8.17); once the support bound goes, it is what replaces the deterministic tube (Proposition 9.7).

Definition 3.8 (Exponential tightness, [3, Section 1.2]). A sequence $\{\xi^n\}_{n \geq 1}$ of random elements of a Polish space $(\mathcal{X}, d)$ is *exponentially tight* at speed n if for every $\alpha < \infty$ there is a compact set $K_\alpha \subseteq \mathcal{X}$ such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\xi^n \notin K_\alpha) \leq -\alpha.$$

Theorem 3.9 (Weak to full LDP, [3, Lemma 1.2.18]). *If $\{\xi^n\}$ is exponentially tight and satisfies a weak large deviation principle at speed n with rate function I , then it satisfies the full large deviation principle at speed n , and I is a good rate function.*

3.1.6 Exponentially good approximations

The last transfer tool compares a family with a whole sequence of approximating families rather than with a single one. It is weaker than exponential equivalence in what it assumes, since the approximation need only improve in the limit, and correspondingly weaker in what it gives: the rate function comes out as an abstract limit of the approximating rates, which then has to be identified. This is the mechanism behind the truncation tower of Section 9, where the approximating families are the walks with increments truncated at level c .

Definition 3.10 (Exponentially good approximation, [3, Definition 4.2.14]). Let $\{\xi^n\}$ and, for each $c > 0$, $\{\xi^{n,(c)}\}$ be random elements of a metric space $(\mathcal{X}, d)$, defined on a common probability space for each fixed (n, c) . The families $\{\xi^{n,(c)}\}$ are *exponentially good approximations* of $\{\xi^n\}$ as $c \rightarrow \infty$ if for every $\delta > 0$

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(d(\xi^{n,(c)}, \xi^n) > \delta) = -\infty.$$

In [3] the approximating family is indexed by $\varepsilon \downarrow 0$; the index here is $c \rightarrow \infty$.

Theorem 3.11 (Approximation theorem, [3, Theorem 4.2.16(a)]). *Let $\{\xi^{n,(c)}\}$ be exponentially good approximations of $\{\xi^n\}$ as $c \rightarrow \infty$, and suppose that for each c the family $\{\xi^{n,(c)}\}$ satisfies the large deviation principle at speed n with good rate function I_c . Then $\{\xi^n\}$ satisfies a weak large deviation principle at speed n with rate function*

$$\widehat{I}(x) = \sup_{\delta > 0} \liminf_{c \rightarrow \infty} \inf_{d(y,x) \leq \delta} I_c(y).$$

3.2 Functional analysis

Two kinds of statement are needed. Compactness results turn a bounded family of paths or velocities into a convergent subsequence, which is how the admissible path space is shown to be compact and how the drift map is shown to be continuous. Weak-convergence results say what survives in the limit once the subsequence is extracted. All of them are applied in normal coordinate charts, so the ambient space is always $\mathbb{R}^d$ or a space of $\mathbb{R}^d$ -valued functions, even when the objects of interest live on M .

3.2.1 Compactness

The first result is used for the positions of a family of paths. We state it for maps into a metric space, since the paths of interest take values in M .

Theorem 3.12 (Arzelà–Ascoli, [1, Theorem 4.25]). *Let K be a compact metric space, (Y, ρ) a complete metric space, and $\mathcal{F} \subseteq C(K; Y)$ a family that is equicontinuous and for which $\bigcup_{f \in \mathcal{F}} f(K)$ is relatively compact in Y . Then $\mathcal{F}$ is relatively compact in $C(K; Y)$ with the uniform metric. In particular every sequence in $\mathcal{F}$ has a uniformly convergent subsequence.*

An equi-Lipschitz family of paths $[0, 1] \rightarrow M$ with a common starting point satisfies both hypotheses on a complete manifold, the second by Hopf–Rinow (Theorem 2.25). This is how $\text{AC}_{x_0, r}([0, 1]; M)$ is shown to be compact in Subsection 8.2.2.

The second result is used for the velocities. A sequence of velocities bounded in L^∞ need not converge in any strong sense, but the unit ball of a dual space is compact in the weak-* topology.

Theorem 3.13 (Weak and weak-* compactness). *Let E be a Banach space.*

- (i) (Banach–Alaoglu) *The closed unit ball of E^* is compact in the weak-* topology [1, Theorem 3.16].*
- (ii) (Dunford–Pettis) *For $A \subseteq \mathbb{R}^m$ of finite measure, a subset of $L^1(A)$ is relatively weakly compact if and only if it is bounded and uniformly integrable [1, Theorem 4.30].*
- (iii) (Eberlein–Šmulian) *A subset of E is relatively weakly compact if and only if it is relatively weakly sequentially compact [1, Ch. 3].¹*

Applied to $E = L^1$, whose dual is L^∞ [1, Theorem 4.14], Banach–Alaoglu is what makes a uniformly bounded family of velocities weak-* precompact. Dunford–Pettis is the replacement when the velocities are only uniformly integrable, which is the situation in Section 9. The next subsubsection packages both in the single form the later arguments call on.

3.2.2 Weak convergence of velocities

Two elementary facts about absolutely continuous $\mathbb{R}^d$ -valued paths recur in the path-space arguments. They appear first in the continuity proof for the drift map (Subsection 8.2.3), where the velocities are uniformly bounded, and again in the identification of the rate function for unbounded increments (Subsection 9.6), where they are only uniformly integrable. We record them once, in the weak- L^1 form that covers both cases; the bounded case is recovered as an upgrade to weak-*. Both are applied in normal coordinate charts.

Lemma 3.14 (Uniform convergence forces weak- L^1 velocity convergence). *Let $x_n, x : [a, b] \rightarrow \mathbb{R}^d$ be absolutely continuous with $x_n \rightarrow x$ uniformly and $\{\dot{x}_n\}$ uniformly integrable. Then x is absolutely continuous and $\dot{x}_n \rightharpoonup \dot{x}$ weakly in $L^1([a, b]; \mathbb{R}^d)$. If in addition the x_n are uniformly Lipschitz, that is $\sup_n \|\dot{x}_n\|_{L^\infty} < \infty$, the convergence holds weak-* in $L^\infty([a, b]; \mathbb{R}^d)$.*

¹Brezis states a companion result, [1, Theorem 3.19]: a Banach space in which every bounded sequence has a weakly convergent subsequence is reflexive. The equivalence quoted here is discussed in the comments to that chapter, which is why the citation stays at chapter level.

Proof. Uniform integrability gives a concave modulus ω with $\omega(0+) = 0$ and $\sup_n \int_A \|\dot{x}_n\| \leq \omega(|A|)$ for measurable A ; then $\|x(t) - x(s)\| = \lim_n \|\int_s^t \dot{x}_n\| \leq \omega(t - s)$ and more generally over finite disjoint unions, so x is absolutely continuous. By the Dunford–Pettis theorem [1, Theorem 4.30] (with the Eberlein–Šmulian theorem for sequential extraction), every subsequence of $\{\dot{x}_n\}$ has a further subsequence converging weakly in L^1 to some v . For $g \in C^1$, integration by parts gives

$$\int_a^b \langle g, \dot{x}_n \rangle dt = \langle g(b), x_n(b) \rangle - \langle g(a), x_n(a) \rangle - \int_a^b \langle \dot{g}, x_n \rangle dt \longrightarrow \int_a^b \langle g, \dot{x} \rangle dt,$$

so $\int \langle g, v \rangle = \int \langle g, \dot{x} \rangle$ for all $g \in C^1$, whence $v = \dot{x}$ almost everywhere. All weak subsequential limits coincide, so the full sequence converges weakly to $\dot{x}$.

For the final claim, suppose the x_n are uniformly Lipschitz, so $L := \sup_n \|\dot{x}_n\|_{L^\infty} < \infty$. Weak-* convergence in $L^\infty = (L^1)^*$ [1, Thm. 4.14] is tested against $g \in L^1$. Given such g and $\varepsilon > 0$, choose $g_\varepsilon \in L^\infty$ with $\|g - g_\varepsilon\|_{L^1} < \varepsilon$ (truncation). Then

$$\left| \int_a^b \langle g, \dot{x}_n - \dot{x} \rangle \right| \leq \left| \int_a^b \langle g_\varepsilon, \dot{x}_n - \dot{x} \rangle \right| + 2L \|g - g_\varepsilon\|_{L^1},$$

and the first term vanishes as $n \rightarrow \infty$ (as $g_\varepsilon \in L^\infty$, by the weak- L^1 convergence already proved), so $\limsup_n \left| \int_a^b \langle g, \dot{x}_n - \dot{x} \rangle \right| \leq 2L\varepsilon$; let $\varepsilon \rightarrow 0$. $\square$

Lemma 3.15 (Weak- $L^1 \times$ strong pairing). *Let $v_n \rightharpoonup v$ weakly in $L^1([a, b]; \mathbb{R}^d)$ and let $b_n \rightarrow b$ in $L^\infty([a, b]; \mathbb{R}^{d' \times d})$. Then, for every $t \in [a, b]$,*

$$\int_a^t b_n v_n ds \longrightarrow \int_a^t b v ds.$$

Proof. Split $\int_a^t b_n v_n - \int_a^t b v = \int_a^t (b_n - b) v_n + \int_a^t b (v_n - v)$. The first term is bounded in norm by $\|b_n - b\|_{L^\infty} \sup_n \|v_n\|_{L^1} \rightarrow 0$ uniformly in t , because weakly convergent sequences are L^1 -bounded [1, Prop. 3.5(iii)]. The second tends to zero for each fixed t , since the rows of $\mathbf{1}_{[a, t]} b$ lie in L^∞ , the dual of L^1 , against which $v_n \rightharpoonup v$. $\square$

3.3 Convex analysis

Four pieces of convex-analytic notation are used in Section 9, all standard and all in the sense of Rockafellar [12]. For $C \subseteq \mathbb{R}^d$ convex, $\text{aff } C$ is its affine hull, the smallest affine subspace containing it, and $\text{ri } C$ is its *relative interior*, the interior of C taken inside $\text{aff } C$ rather than inside $\mathbb{R}^d$. The distinction matters whenever C is a proper lower-dimensional subset of $\mathbb{R}^d$, since the ordinary interior is then empty while the relative interior is not. For arbitrary C we write $\overline{\text{conv}} C$ for the closed convex hull. For $f : \mathbb{R}^d \rightarrow [-\infty, +\infty]$ we write $\text{dom } f := \{\lambda : f(\lambda) < +\infty\}$ for its domain of finiteness. Finally, the *support function* of a set K is $\hat{e} \mapsto \sup_{y \in K} \langle \hat{e}, y \rangle$, which describes K by how far it reaches in each direction.

3.4 Ordinary differential equations

Two tools are used repeatedly. Grönwall’s inequality converts a differential or recursive bound on a nonnegative quantity into a closed-form one, and is often used in the stability and error estimates in this thesis. The Carathéodory–Picard theorem is the existence and uniqueness result for ODEs whose right-hand side is only measurable in time, which is the regularity the drift equation has.

3.4.1 Grönwall's inequality

Lemma 3.16 (Grönwall, differential form). *Let $u : [0, T] \rightarrow [0, \infty)$ be absolutely continuous and let $a, b \in L^1([0, T])$ with $b \geq 0$. If $u'(t) \leq a(t) + b(t)u(t)$ for almost every t , then for all $t \in [0, T]$,*

$$u(t) \leq u(0) e^{\int_0^t b} + \int_0^t a(s) e^{\int_s^t b} ds.$$

In particular, if $a \equiv 0$ then $u(t) \leq u(0) e^{\int_0^t b}$, and if $b \equiv L$ is constant then $u(t) \leq u(0) e^{Lt}$.

The discrete counterpart is what the Euler analysis needs. The two growth factors it allows, $1+Lh$ and e^{Lh} , are the ones produced by the flat and the Riemannian one-step estimates respectively.

Lemma 3.17 (Grönwall, discrete form). *Let $e_0, \dots, e_N \geq 0$, let $G \geq 1$ and $\beta \geq 0$, and suppose*

$$e_{k+1} \leq G e_k + \beta \quad (0 \leq k < N).$$

Then $e_k \leq G^k e_0 + \beta \frac{G^k - 1}{G - 1}$ for $G > 1$, and $e_k \leq e_0 + k\beta$ for $G = 1$. If in addition $e_0 = 0$, $\beta = Ch^2$ with $h > 0$, $kh \leq T$, and either $G = 1 + Lh$ or $G = e^{Lh}$ for some $L > 0$, then

$$e_k \leq \frac{C}{L} (e^{LT} - 1) h.$$

Proof. Iterating the recursion gives $e_k \leq G^k e_0 + \beta \sum_{j=0}^{k-1} G^j$, and the geometric sum is $(G^k - 1)/(G - 1)$ for $G > 1$ and k for $G = 1$. For the last claim, both choices of G satisfy $G^k \leq e^{Lkh} \leq e^{LT}$ and $G - 1 \geq Lh$, so $e_k \leq Ch^2(e^{LT} - 1)/(Lh)$. $\square$

3.4.2 The Carathéodory–Picard theorem

The drift equation (9) has a right-hand side that is continuous in the state but only measurable in time, because the noise path ψ contributing $\dot{\psi}$ is merely absolutely continuous. The Picard–Lindelöf theorem does not apply. The following version, due to Carathéodory, does. Throughout, a *solution* means an absolutely continuous function satisfying the ODE almost everywhere, which is the natural notion at this regularity.

Theorem 3.18 (Carathéodory–Picard). *Let $I = [t_0, t_0 + a]$ and $F : I \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfy*

(C1) *$F(\cdot, y)$ is Lebesgue measurable for each $y \in \mathbb{R}^d$,*

(C2) *$F(t, \cdot)$ is continuous on $\mathbb{R}^d$ for almost every $t \in I$,*

(C3) *$\|F(t, y)\| \leq h(t)$ for some $h \in L^1(I)$, for a.e. $t \in I$ and all $y \in \mathbb{R}^d$,*

(C4) *$\|F(t, y_1) - F(t, y_2)\| \leq \ell(t) \|y_1 - y_2\|$ for some $\ell \in L^1(I)$, for a.e. $t \in I$ and all $y_1, y_2 \in \mathbb{R}^d$,*

Then for each $y_0 \in \mathbb{R}^d$ the initial value problem $\dot{y} = F(t, y)$, $y(t_0) = y_0$, has a unique absolutely continuous solution on I satisfying the ODE a.e.

For a textbook treatment see Walter [16, §10, Supplement II, Theorem XVIII(b)]. Verifying (C1) to (C4) for the chart form of (9) is the work of Subsection 8.2.1.

4 Convergence of the geodesic Euler scheme

A key ingredient in the proof of the main theorem is that the discrete drifted walk (which freezes the vector field at the start of each substep) and the continuous flow differ by at most $O(1/n)$ in the sup-norm. This is the convergence of a geodesic analogue of the explicit Euler method.

We develop this in full detail, starting with the classical Euclidean case as a warm-up and then extending to compact Riemannian manifolds.

4.1 Warm-up: the flat case $\mathbb{R}^d$

4.1.1 Setup

Let $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a C^1 vector field that is globally Lipschitz with constant L and bounded by B :

$$\|f(a) - f(b)\| \leq L\|a - b\|, \quad \|f(z)\| \leq B \quad \text{for all } a, b, z \in \mathbb{R}^d.$$

Let $y : [0, 1] \rightarrow \mathbb{R}^d$ be the unique solution of $\dot{y}(t) = f(y(t))$, $y(0) = x$. The Euler iterates with step size $h = 1/n$ are

$$x_0 = x, \quad x_{k+1} = x_k + h f(x_k),$$

and we write $t_k = kh$ for the grid points. The piecewise linear interpolant is $\gamma_n(t) = x_k + (t - t_k)f(x_k)$ for $t \in [t_k, t_{k+1}]$.

The idea is straightforward: on each step, the true solution moves and so f changes slightly along the trajectory, whereas the Euler method freezes f at the start of the step. This causes a local error of $O(h^2)$ per step; after $n \sim 1/h$ steps, a discrete Grönwall estimate accumulates this into a global error of $O(h)$.

4.1.2 Local truncation error

Lemma 4.1 (One-step consistency). *There exists $C_0 \geq 0$ such that for all k ,*

$$\|y(t_{k+1}) - (y(t_k) + h f(y(t_k)))\| \leq C_0 h^2.$$

Proof. Using the integral form of the ODE, we can write

$$y(t_{k+1}) = y(t_k) + h f(y(t_k)) + \tau_k, \quad \tau_k = \int_{t_k}^{t_{k+1}} (f(y(s)) - f(y(t_k))) \, ds.$$

To estimate the remainder, note that $\|y(s) - y(t_k)\| \leq \int_{t_k}^s \|f(y(u))\| \, du \leq B(s - t_k)$. Then the Lipschitz bound gives

$$\|\tau_k\| \leq \int_{t_k}^{t_{k+1}} L \cdot B(s - t_k) \, ds = \frac{LB}{2} h^2,$$

so we can take $C_0 = LB/2$. □

4.1.3 Error recursion and discrete Grönwall

Let $e_k = \|x_k - y(t_k)\|$ be the error at the grid points.

Lemma 4.2. *The errors satisfy $e_{k+1} \leq (1 + Lh)e_k + C_0 h^2$, and for all k with $t_k \leq 1$,*

$$e_k \leq \frac{C_0}{L}(e^L - 1)h.$$

Proof. Subtracting the expressions for x_{k+1} and $y(t_{k+1})$ and taking norms:

$$\begin{aligned} e_{k+1} &= \|x_{k+1} - y(t_{k+1})\| \\ &\leq \|x_k - y(t_k) + h(f(x_k) - f(y(t_k)))\| + \|\tau_k\| \\ &\leq (1 + Lh)e_k + C_0 h^2. \end{aligned}$$

This is the recursion of Lemma 3.17 with $G = 1 + Lh$, $\beta = C_0 h^2$ and $e_0 = 0$; that lemma with $T = 1$ gives the stated bound. $\square$

4.1.4 Uniform convergence

Theorem 4.3 (Euler convergence in $\mathbb{R}^d$). *There exists $C > 0$ such that $\sup_{t \in [0,1]} \|\gamma_n(t) - y(t)\| \leq Ch$.*

Proof. For $t \in [t_k, t_{k+1}]$, $\|\gamma_n(t) - y(t)\| \leq \underbrace{\|\gamma_n(t) - x_k\|}_{\leq Bh} + e_k + \underbrace{\|y(t_k) - y(t)\|}_{\leq Bh}$. Taking the supremum and inserting the result from Lemma 4.2 gives the result with $C = 2B + \frac{C_0}{L}(e^L - 1)$. $\square$

4.2 Extension to complete Riemannian manifolds

Now let (M, g) be a complete Riemannian manifold and $V \in \Gamma(TM)$ a smooth vector field. Let $\gamma(t) = \varphi_t(x_0)$ be the true integral curve.

The estimates of this subsection are quantitative, so they need a set on which to take suprema. We fix one:

$$\begin{aligned} K &\subseteq M \text{ compact, containing } \gamma(t) \text{ and } \gamma_h(t) \text{ for all } t \in [0, 1] \text{ and all small } h, \\ &\text{together with the flow lines } \varphi_t(c(s)) \text{ of the minimising geodesics } c \text{ joining } \gamma(t_k) \text{ to } \bar{p}_k. \end{aligned} \quad (2)$$

On compact M one takes $K = M$ and the condition is vacuous. Every later application is of the second kind: the trajectories are confined to a tube by an a priori bound, and K is that tube. Stating the hypothesis here, rather than adding a localisation remark afterwards, is what lets Subsection 8.4 and Section 9 cite these results directly.

Definition 4.4 (Geodesic Euler scheme). For step size $h = 1/n$, define the numerical iterates by $\bar{p}_0 = x_0$, $\bar{p}_{k+1} = \text{Exp}_{\bar{p}_k}(hV(\bar{p}_k))$, and the piecewise-geodesic interpolant by $\gamma_h(t) = \text{Exp}_{\bar{p}_k}((t - t_k)V(\bar{p}_k))$ for $t \in [t_k, t_{k+1}]$.

We write $\Psi_h(p) = \text{Exp}_p(hV(p))$ for the one-step map. Compactness of K gives us uniform constants:

$$B := \sup_{p \in K} \|V(p)\| < \infty, \quad L := \sup_{p \in K} \|\nabla V(p)\|_{\text{op}} < \infty, \quad \iota := \text{inj}(K) > 0.$$

All three are finite for the same reason as before, that a continuous function on a compact set is bounded; $\text{inj}(K) > 0$ is Lemma 2.26. The constants now carry a K , which the statements below suppress.

The proof of convergence follows the same three-step pattern as in $\mathbb{R}^d$.

4.2.1 One-step consistency

Lemma 4.5. *There exist $h_0 > 0$ and $C_{\text{loc}} > 0$ such that for all $0 < h \leq h_0$ and all $p \in K$,*

$$d(\varphi_h(p), \Psi_h(p)) \leq C_{\text{loc}} h^2.$$

Proof. Fix $p \in K$ and let $\gamma : [0, 1] \rightarrow M$ be the unique integral curve of V with $\gamma(0) = p$, i.e. $\gamma(t) = \varphi_t(p)$. Smoothness of V guarantees local existence and uniqueness of the integral curve through arbitrary points; the curve stays in the compact set K of (2) on $[0, 1]$, so it does not escape in finite time, so in particular $\gamma := \varphi \cdot (p) : [0, 1] \rightarrow M$ is well-defined.

Set $h_0 := \iota/(4B)$. The bound $d(\gamma(t), p) \leq \int_0^t \|\dot{\gamma}(s)\| ds \leq Bt \leq Bh$ ensures that $\gamma(t)$ remains in the closed ball $\overline{B}(p, \iota/4)$ for all $t \in [0, h]$, provided $h \leq h_0$. We may therefore define

$$y(t) = \log_p(\gamma(t)) \in \overline{B}(0, \iota/4) \subset T_p M, \quad t \in [0, h],$$

and work in the chart $(B(p, \iota/2), \log_p)$. By the chain rule, y satisfies $\dot{y}(t) = F_p(y(t))$ with $y(0) = 0$, where $F_p(u) = (d\log_p)_{\text{Exp}_p(u)} V(\text{Exp}_p(u))$. Since $(d\log_p)_p = \text{id}$, we have $F_p(0) = V(p)$. The map F_p is smooth on $B(0, \iota/2)$ as a composition of smooth maps. By compactness of K and joint smoothness of $(p, u) \mapsto F_p(u)$ on the closed disk bundle of radius $\iota/4$ over K , there exist constants A and B_F , independent of p , such that F_p is A -Lipschitz and bounded by B_F on $\overline{B}(0, \iota/4) \subset T_p M$.

The same local truncation argument as in Lemma 4.1, applied to the ODE $\dot{y} = F_p(y)$ with $F_p(0) = V(p)$, yields

$$\|y(h) - hV(p)\| \leq \frac{AB_F}{2} h^2.$$

To get the local truncation error, we push back to M via Exp_p : Lemma 2.33 with this K and $\rho = \iota/2$ (applicable since $\|y(h)\|, \|hV(p)\| \leq Bh \leq \iota/4 \leq \iota/2$) gives

$$d(\varphi_h(p), \Psi_h(p)) = d(\text{Exp}_p(y(h)), \text{Exp}_p(hV(p))) \leq C_{\text{exp}}(K, \iota/2) \cdot \frac{AB_F}{2} h^2 =: C_{\text{loc}} h^2.$$

□

4.2.2 Stability of the flow

Lemma 4.6 (Lipschitz continuity of the flow). *Let $p, q \in K$, let c be a minimising geodesic from p to q , and suppose $\varphi_t(c(s)) \in K$ for all $s, t \in [0, 1]$. Then for all $t \in [0, 1]$, $d(\varphi_t(p), \varphi_t(q)) \leq e^{Lt} d(p, q)$.*

The containment is what the constant L is taken over, and it is the only place where the flow of a point other than the two endpoints is used. On compact M , with $K = M$, it holds for free and the statement is the usual one.

Proof. Let $c : [0, 1] \rightarrow M$ be a minimising geodesic from p to q with constant speed $\|c'(s)\| = d(p, q)$. Consider the two-parameter family $F(t, s) = \varphi_t(c(s))$ and set $W(t, s) = \partial_s F(t, s)$. Since $\partial_t F = V(F)$, the symmetry lemma (Lemma 2.13) gives $D_t \partial_s F = D_s \partial_t F$, and hence $\nabla_t W = \nabla_s(V \circ F) = (\nabla_{\partial_s F(t, s)} V)(F(t, s)) = (\nabla_W V)(F)$.

The squared norm $\psi(t) = \|W(t, s)\|^2$ then satisfies $\psi'(t) = 2\langle \nabla_t W, W \rangle \leq 2\|\nabla_W V(F)\| \|W\| \leq 2L\|W\|^2 = 2L\psi(t)$. By the continuous Grönwall lemma, $\psi(t) \leq e^{2Lt}\psi(0)$, so $\|W(t, s)\| \leq e^{Lt}\|W(0, s)\| = e^{Lt} d(p, q)$.

Since $s \mapsto F(t, s)$ is a curve from $\varphi_t(p)$ to $\varphi_t(q)$, its length is $\int_0^1 \|W(t, s)\| ds \leq e^{Lt} d(p, q)$, and Riemannian distance is at most the length of any connecting curve (Lemma 2.28(ii)). □

4.2.3 Global error and uniform convergence

Lemma 4.7 (Mesh-point error). *Let $e_k = d(\gamma(t_k), \bar{p}_k)$. Then $e_{k+1} \leq C_{\text{loc}} h^2 + e^{Lh} e_k$, and there exists $C_1 > 0$ such that $e_k \leq C_1 h$ for all k with $t_k \leq 1$.*

Proof. By the triangle inequality (inserting $\varphi_h(\bar{p}_k)$ as the intermediate point),

$$\begin{aligned} e_{k+1} &= d(\varphi_h(\gamma(t_k)), \Psi_h(\bar{p}_k)) \\ &\leq d(\Psi_h(\bar{p}_k), \varphi_h(\bar{p}_k)) + d(\varphi_h(\bar{p}_k), \varphi_h(\gamma(t_k))) \\ &\leq C_{\text{loc}} h^2 + e^{Lh} e_k, \end{aligned}$$

using Lemma 4.5 for the first term and Lemma 4.6 for the second, both applicable because $\bar{p}_k, \gamma(t_k)$ and the flow lines between them lie in K by (2). Lemma 3.17, with $G = e^{Lh}$, $\beta = C_{\text{loc}} h^2$ and $T = 1$, then gives $e_k \leq \frac{C_{\text{loc}}}{L} (e^L - 1) h$. Thus

$$\max_{0 \leq k \leq n} d(\gamma_h(t_k), \gamma(t_k)) \leq C_1 h,$$

with $C_1 := \frac{C_{\text{loc}}}{L} (e^L - 1)$. □

Theorem 4.8 (Convergence of the geodesic Euler scheme). *Let (M, g) be complete and V smooth. Let γ be the integral curve of V with $\gamma(0) = x_0$. Let $\gamma_h(t)$ be the piecewise-geodesic interpolant defined in Definition 4.4, and let K be as in (2). Then there exists $C = C(K, V) > 0$ such that for all sufficiently small $h = 1/n$,*

$$\sup_{t \in [0, 1]} d(\gamma_h(t), \gamma(t)) \leq C h.$$

In particular, $\gamma_h \rightarrow \gamma$ uniformly on $[0, 1]$.

Proof. Fix $t \in [0, 1]$ and let k be such that $t \in [t_k, t_{k+1}]$. Inserting the mesh point $\bar{p}_k$ as an intermediate, the triangle inequality gives

$$d(\gamma_h(t), \gamma(t)) \leq d(\gamma_h(t), \bar{p}_k) + d(\bar{p}_k, \gamma(t_k)) + d(\gamma(t_k), \gamma(t)).$$

The first term measures how far the numerical interpolant γ_h has travelled within the substep $[t_k, t_{k+1}]$. By construction, γ_h restricted to this interval is the geodesic $u \mapsto \text{Exp}_{\bar{p}_k}((u - t_k)V(\bar{p}_k))$, which has constant speed $\|V(\bar{p}_k)\| \leq B$. Bounding distance by length,

$$d(\gamma_h(t), \bar{p}_k) \leq \int_{t_k}^t \|\dot{\gamma}_h(u)\| du \leq B(t - t_k) \leq Bh.$$

The second term is the mesh-point error, which by Lemma 4.7 satisfies $d(\bar{p}_k, \gamma(t_k)) = e_k \leq C_1 h$. The third term is bounded analogously to the first: since γ is the integral curve of V , its velocity is $\dot{\gamma}(u) = V(\gamma(u))$ and so $\|\dot{\gamma}(u)\| \leq B$, which gives

$$d(\gamma(t_k), \gamma(t)) \leq \int_{t_k}^t \|\dot{\gamma}(u)\| du \leq B(t - t_k) \leq Bh.$$

Combining the three estimates yields $d(\gamma_h(t), \gamma(t)) \leq (2B + C_1) h$, with the right-hand side independent of t . Taking the supremum over $t \in [0, 1]$ gives the result with $C := 2B + C_1$. □

5 Geodesic random walks

This section introduces the objects the thesis is about. We first recall, from [13], the geodesic random walk and the consistent families of increment measures that drive it, and we then add a drift vector field to obtain the *drifted* geodesic random walk (Definition 5.3), whose large deviations are the subject of Theorem 6.3.

5.1 Consistent increment measures and geodesic random walks

We recall the essential definitions from [13]. The construction reads best as a translation of the classical random walk, so we set it out that way before stating it. A walk in $\mathbb{R}^d$ is a sum $\sum_{k \leq n} X_k$ of increments, and each increment is a vector whose addition to the current position means moving along the straight line in that direction, for unit time and at speed $\|X_k\|$, which uses the whole increment. Neither half of that description survives on a general manifold, which has no linear structure in which a point and a vector can be added, so each half needs a replacement.

The geometry supplies both. A direction at a point becomes a tangent vector at that point, and the straight line becomes the geodesic that this vector generates; since the straight lines of $\mathbb{R}^d$ are exactly its geodesics, the new description agrees with the old one in the flat case. One step of the walk is therefore to sample a tangent vector at the current point and then follow its geodesic for unit time, which the exponential map does in one symbol. A step is an application of Exp , and a walk is a concatenation of such steps.

Two features of the manifold walk have no counterpart in $\mathbb{R}^d$. The first is that an increment must be sampled where the walk currently is, because tangent spaces at different points are different vector spaces. One law on one space is therefore not enough, and the walk needs a family $\{\mu_x\}$ with one law for every point. What makes such a family behave like a single distribution is Definition 5.1, which asks that the laws agree once they are compared by parallel transport. That is as close as a manifold comes to saying that the increments are identically distributed, and it is why consistency appears in every hypothesis of this thesis.

The second is that the walk cannot be rescaled after the event. In $\mathbb{R}^d$ one divides the sum by n and studies $\frac{1}{n} \sum_{k \leq n} X_k$, whereas in M the expression $\frac{1}{n} S_n$ has no meaning, since S_n is a point and points on general manifolds do not scale. What is rescaled instead is each increment, and because $\text{Exp}_x(\alpha X)$ travels along the geodesic of X for time α , scaling the increments amounts to shortening the time for which each geodesic is followed.

Definition 5.1 (Consistent family of increment measures). A collection $\{\mu_x\}_{x \in M}$ with $\mu_x \in \mathcal{P}(T_x M)$ satisfies the *consistency property* if for all $x, y \in M$ and every smooth curve γ from x to y ,

$$\mu_y = \mu_x \circ (\tau_{xy}^\gamma)^{-1}.$$

Equivalently, the log-moment generating functions satisfy $\Lambda_x(\lambda) = \Lambda_y(\tau_{xy}^\gamma \lambda)$ for all $\lambda \in T_x M$.

Definition 5.2 (Rescaled geodesic random walk). Fix $x_0 \in M$ and $\alpha > 0$. An α -rescaled geodesic random walk started at x_0 with increments $\{X_n\}_{n \geq 1}$ compatible with $\{\mu_x\}$ is defined by $(\alpha * S)_0 = x_0$ and

$$(\alpha * S)_{k+1} = \text{Exp}_{(\alpha * S)_k}(\alpha X_{k+1}), \quad X_{k+1} \sim \mu_{(\alpha * S)_k}.$$

Here $X_{k+1} \in T_{(\alpha * S)_k} M$ and the recursion runs over all $k \geq 0$.

The *empirical average process* is $Z_n := (\frac{1}{n} * S)_n$.

The sphere makes this concrete. Let each μ_x be the uniform law on a circle of fixed radius in $T_x M$. Geodesics of the sphere are the great circles, so the walk moves along great-circle arcs of one fixed length in uniformly random directions, and rescaling shortens those arcs by the factor α while it leaves their directions alone. Figure 6 draws such a walk, with a drift added as in Definition 5.3, and Figure 8 shows longer sample paths of the same kind.

5.2 Drifted geodesic random walks

We now define the central object of this thesis. In $\mathbb{R}^d$ a walk with drift adds a deterministic displacement to each step, and if that displacement comes from a vector field V , the step reads $x \mapsto x + \frac{1}{n}X + \frac{1}{n}V(x)$, where the two summands are vectors at the same point. A vector field on a manifold supplies exactly one tangent vector at each point, so $V(S_k)$ and the sampled increment X_{k+1}^n again live in the same tangent space $T_{S_k} M$ and can be added there. Their sum is a single tangent vector, and one application of Exp then takes the step.

This is worth pausing on, because it is the only place where the two ingredients meet: noise and drift are added *before* the exponential map, and not composed after it. A reader might instead expect two half-steps, one along the flow of V and one along the noise geodesic, but adding first is the definition we take.

The factor $\frac{1}{n}$ appears twice, and for two different reasons. On the noise it is the rescaling of the previous subsection, which turns the sum of n increments into an average, while on the drift it is a time step, so that after n steps the drift has acted for total time 1 and contributes the time-1 flow of V rather than something that vanishes or diverges. The remark below makes both statements explicit in the flat case, where the drift part is visibly a Riemann sum for $\int_0^1 V$ along the trajectory.

Definition 5.3 (Drifted geodesic random walk). Fix $x_0 \in M$, a smooth vector field $V \in \Gamma(TM)$, and a consistent family $\{\mu_x\}_{x \in M}$ with $\text{supp}(\mu_x) \subseteq \overline{B}(0, r)$ for some $r > 0$, the same for every x .

For $n \geq 1$, define $S_0^n = x_0$ and recursively

$$S_{k+1}^n = \text{Exp}_{S_k^n} \left(\frac{1}{n} X_{k+1}^n + \frac{1}{n} V(S_k^n) \right), \quad k = 0, \dots, n-1,$$

where $X_{k+1}^n \in T_{S_k^n} M$ is distributed according to $\mu_{S_k^n}$ with independent increments (in the sense of [13, Definition 3.2.3]). The *drifted empirical average* is the endpoint S_n^n . (We do not abbreviate it further, to avoid confusion with the undrifted walk Z^n below.)

Remark 5.4. To see what this looks like in the flat case: in $\mathbb{R}^d$ we have $\text{Exp}_x(v) = x + v$, so

$$S_n^n = x_0 + \frac{1}{n} \sum_{k=1}^n X_k^n + \frac{1}{n} \sum_{k=1}^n V(S_{k-1}^n).$$

The first sum is the noise, and the second is a Riemann-sum approximation to the integral of V along the trajectory, which converges to the flow $\varphi_1(x_0)$ as $n \rightarrow \infty$.

Figure 6 shows one step of the recursion on the sphere, where the two vectors being added are visibly in the same tangent space and Exp visibly bends the sum back onto the manifold.

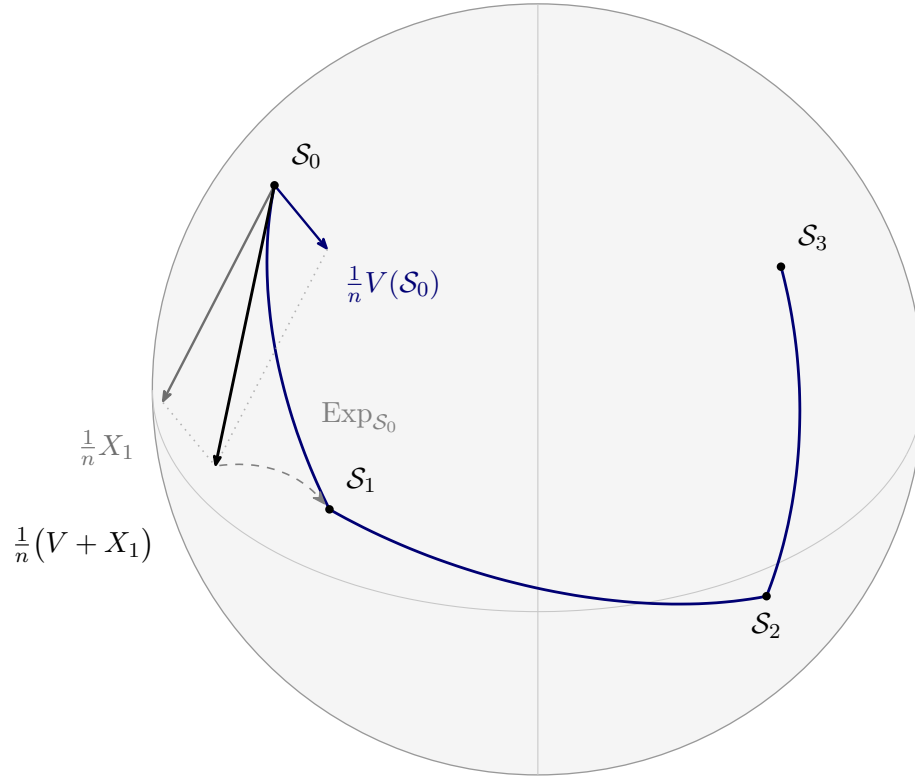


Figure 6: One step of Definition 5.3, on the sphere. At $\mathcal{S}_0$ the drift $\frac{1}{n}V(\mathcal{S}_0)$ and the increment $\frac{1}{n}X_1$ lie in the same tangent space, so they can be added; the dotted parallelogram adds them, and $\text{Exp}_{\mathcal{S}_0}$ carries the sum, which points off the sphere, to the next position $\mathcal{S}_1$. The later steps repeat the construction, and only their arcs are drawn. Geodesics of the sphere are great circles, so each step is a great-circle arc.

6 Main results

With the geodesic random walk and its drifted variant in hand (Section 5), we can state the main results. The drifted empirical averages satisfy a Cramér-type large deviation principle in M , with a rate function built from the Legendre transform of the increment law; we reach it in two steps, each an application of the contraction principle. The proof occupies the rest of the thesis, following a strategy that is easiest to see in the Euclidean warm-up (Section 7) before it is carried out on the manifold (Section 8).

The proof rests on a single construction, the drift map. Given an absolutely continuous noise path ψ with $\psi(0) = x_0$, let $\Phi(\psi)$ be the path obtained by carrying ψ on top of the drift V : the solution ϕ of

$$\phi(0) = x_0, \quad \dot{\phi}(t) = V(\phi(t)) + P_{\psi(t) \rightarrow \phi(t)} \dot{\psi}(t),$$

where $P_{\psi(t) \rightarrow \phi(t)}$ is parallel transport along the minimising geodesic from $\psi(t)$ to $\phi(t)$. We need the transport here to make the sum meaningful: $\dot{\psi}(t) \in T_{\psi(t)}M$ and $V(\phi(t)) \in T_{\phi(t)}M$ live in different tangent spaces, and $P_{\psi(t) \rightarrow \phi(t)}$ carries the noise velocity to the tangent space of the output before adding the drift. In Section 8 we construct Φ and its domain $\text{AC}_{x_0, r}$, the noise paths of speed at most r ; here we record the one property the argument uses. Figure 7 illustrates the construction.

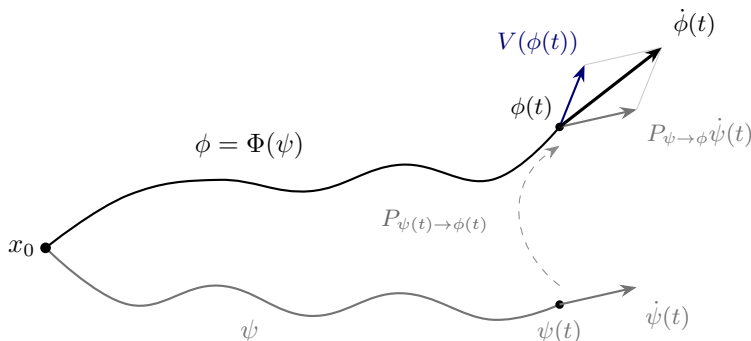


Figure 7: The drift map Φ . The noise path ψ (grey) and the output $\phi = \Phi(\psi)$ (black) leave from the same point x_0 ; both are drawn up to time t . There the noise velocity $\dot{\psi}(t) \in T_{\psi(t)}M$ is parallel-transported along the minimising geodesic from $\psi(t)$ to $\phi(t)$ and added to the drift $V(\phi(t))$; blue is the drift, grey the transported noise, black their sum. By Lemma 8.2 the two points never separate by more than $\|V\|_\infty t$, so both stay in $K_0 = \overline{B}(x_0, r + \|V\|_\infty)$, and $\|V\|_\infty < \text{inj}(K_0)$ makes the transporting geodesic unique.

Proposition 6.1 (Continuity of the drift map). *The map $\Phi : \text{AC}_{x_0, r} \rightarrow \text{AC}_{x_0, r + \|V\|_\infty}$ is (uniformly) continuous with respect to d_∞ (on both sides).*

Continuity is what lets a large deviation principle pass through Φ . The undrifted rescaled walk Z^n already satisfies one on the path space, the Riemannian Mogulskii theorem of Kraaij, Redig and Versendaal (Theorem 8.1). Since Z^n is the noise path driving the drifted walk and Φ is continuous, the contraction principle carries that principle over to the drifted flow.

The principle holds for the discrete walk as well, and we state the two together. Write $\hat{Y}^n$ for the piecewise-geodesic interpolation of the drifted walk $S_0^n, \dots, S_n^n$ of Definition 5.3, at the times $t_k = k/n$. It is $\hat{Y}^n$, not the flow $\Phi(Z^n)$, that the walk actually produces; the two differ only by the Euler discretisation error, and Subsection 8.4 shows that the error is superexponentially small.

Theorem 6.2 (Path-space LDP for the drifted flow and the drifted walk). *Let (M, g) be a complete Riemannian manifold and let $V \in \Gamma(TM)$ be smooth and bounded. Let $\{\mu_x\}_{x \in M}$ be a consistent family of probability measures with $\text{supp}(\mu_x) \subseteq \overline{B}(0, r)$ for some $r > 0$, the same for every x . Set $K_0 := \overline{B}(x_0, r + \|V\|_\infty)$ and assume $\|V\|_\infty < \text{inj}(K_0)$, the uniform injectivity radius over K_0 . Each of the processes*

- (i) *the drifted flow $\Phi(Z^n)$, and*
- (ii) *the interpolated drifted walk $\hat{Y}^n$,*

satisfies the LDP on $(C_{x_0}([0, 1]; M), d_\infty)$ with speed n and the same good rate function

$$I_{\text{path}}^V(\phi) = \begin{cases} \int_0^1 \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) dt & \text{if } \phi \in \text{AC}([0, 1]; M) \text{ and } \phi(0) = x_0, \\ +\infty & \text{otherwise.} \end{cases} \quad (3)$$

The last step is evaluation at the final time. The map $\text{ev}_1 : C([0, 1]; M) \rightarrow M$, $\phi \mapsto \phi(1)$, is continuous for the uniform metric, so the contraction principle applies once more and reduces the path-space principle to the endpoint. The rate becomes the infimum of I_{path}^V over paths ending at x . This is the main theorem.

Theorem 6.3 (Cramér's theorem for drifted geodesic random walks). *Let (M, g) be a complete Riemannian manifold and let $V \in \Gamma(TM)$ be smooth and bounded. Let $\{\mu_x\}_{x \in M}$ be a consistent family of probability measures with $\text{supp}(\mu_x) \subseteq \overline{B}(0, r)$ for some $r > 0$, the same for every x . Set $K_0 := \overline{B}(x_0, r + \|V\|_\infty)$ and assume $\|V\|_\infty < \text{inj}(K_0)$, the uniform injectivity radius over K_0 . These are the hypotheses of Theorem 6.2.*

For $n \geq 1$, let S_n^n be the drifted empirical average from Definition 5.3. Then $\{S_n^n\}_{n \geq 1}$ satisfies the large deviation principle in M with speed n and good rate function

$$I_M^V(x) = \inf_{\substack{\phi \in \text{AC}([0, 1]; M) \\ \phi(0) = x_0, \phi(1) = x}} \int_0^1 \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) dt.$$

The absolutely continuous paths $\text{AC}([0, 1]; M)$ are those of Definition 2.27. Here Λ_x^ denotes the Legendre transform of $\Lambda_x(\lambda) = \log \int_{T_x M} e^{\langle \lambda, u \rangle} \mu_x(du)$.*

6.1 The extension to unbounded increments

The bounded-support hypothesis $\text{supp}(\mu_x) \subseteq \overline{B}(0, r)$ can be dropped, at the cost of two mild geometric hypotheses, and replaced by *Cramér's condition*: finiteness of all exponential moments. This is the classical hypothesis under which the flat Cramér and Mogulskii theorems (Theorem 3.3, Theorem 3.4) hold. The conclusion is the same large deviation principle, with the same rate function; only the class of admissible increment laws grows, now including the isotropic Gaussian family of Example 6.6.

As in the bounded case we state the path-space principle first and contract it to the endpoint. Only the walk $\hat{Y}^n$ appears here. The flow $\Phi(Z^n)$ has no counterpart once the increments are unbounded, because Φ is defined on the class $\text{AC}_{x_0, r}$ of paths of bounded speed and the noise path Z^n leaves every such class; this is the same obstruction that forces the truncation tower, and it is discussed after Theorem 6.5.

Theorem 6.4 (Path-space LDP with unbounded increments). *Under the hypotheses of Theorem 6.5 below, the interpolated drifted walks $\hat{Y}^n$ satisfy the large deviation principle on $(C_{x_0}([0, 1]; M), d_\infty)$ with speed n and good rate function I_{path}^V of (3), the same rate function as in Theorem 6.2.*

Evaluating at time 1 and contracting gives the endpoint form, which is the theorem the chapter is aimed at.

Theorem 6.5 (Cramér’s theorem for drifted geodesic random walks with unbounded increments). *Let (M, g) be a complete Riemannian manifold with bounded geometry:*

$$\text{inj}(M) > 0 \quad \text{and} \quad \| \text{Rm} \|_\infty := \sup_{p \in M} \| \text{Rm}(p) \|_{\text{op}} < \infty.$$

Let $\{\mu_x\}_{x \in M}$ be a consistent family of probability measures satisfying Cramér’s condition,

$$\Lambda_{x_0}(\lambda) < \infty \quad \text{for all } \lambda \in T_{x_0}M,$$

and let $V \in \Gamma(TM)$ be smooth and bounded with $\|V\|_\infty < \text{inj}(M)$. Let S_n^n be the drifted empirical average of Definition 5.3, with the support bound on the increments dropped. Then $\{S_n^n\}_{n \geq 1}$ satisfies the large deviation principle in M with speed n and good rate function

$$I_M^V(x) = \inf_{\substack{\phi \in \text{AC}([0, 1]; M) \\ \phi(0) = x_0, \phi(1) = x}} \int_0^1 \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) \, dt,$$

the same rate function as in Theorem 6.3.

The proof is carried out in Section 9. It does not rerun the argument above with the support bound relaxed: without a speed bound one cannot control the drift map uniformly over the whole path space: Proposition 6.1 is stated on $\text{AC}_{x_0, r}$, and without a speed bound that class isn’t compact, so the weak-* argument has no domain. Instead it keeps Theorem 6.3 intact and wraps it in a *truncation tower*. At each level c the increments are truncated to $\bar{B}(0, c)$, a bounded family to which Theorem 6.3 applies and yields a path-space LDP with some rate I_c . Cramér’s condition makes the c -truncated walks exponentially good approximations of the original walk as $c \rightarrow \infty$, and the rates I_c converge to I_{path}^V ; the approximation theorem of Dembo and Zeitouni then transfers the principle to the original walk. The two bounded-geometry hypotheses take over the role of the deterministic tube K_0 , which no longer exists once the increments are unbounded.

6.2 Examples

We record two worked examples: isotropic Gaussian increments and, building on them, an explicit rate function for a Killing drift. We also illustrate the walk and its concentration in Figure 8.

Example 6.6 (Isotropic Gaussian increments). Take $\mu_x = \mathcal{N}(0, \sigma^2 \text{id})$, the isotropic Gaussian on each tangent space $T_x M$. We verify the hypotheses and compute the resulting rate function explicitly.

Consistency. The compatibility condition $\mu_y = (\tau_{xy}^\gamma)_* \mu_x$ requires that transporting the increment law from x to y reproduce the increment law at y . Parallel transport $\tau_{xy}^\gamma: T_x M \rightarrow T_y M$ is a linear isometry, and the pushforward of a centred Gaussian by a linear map τ transforms its covariance by $\Sigma \mapsto \tau \Sigma \tau^*$. With $\Sigma = \sigma^2 \text{id}$ and τ an isometry,

$$\tau(\sigma^2 \text{id})\tau^* = \sigma^2 \tau \tau^* = \sigma^2 \text{id},$$

so $(\tau_{xy}^\gamma)_* \mu_x = \mathcal{N}(0, \sigma^2 \text{id}) = \mu_y$ and consistency holds.

Cumulant generating function. For $X \sim \mathcal{N}(0, \sigma^2 \text{id})$ on $T_x M$ one has $\mathbb{E} e^{\langle \lambda, X \rangle} = e^{\frac{1}{2} \sigma^2 \|\lambda\|^2}$, hence

$$\Lambda_x(\lambda) = \frac{1}{2} \sigma^2 \|\lambda\|^2.$$

This is finite for every $\lambda \in T_x M$ and independent of the base point x : the former is the global Cramér integrability, the latter makes the pathwise functional below autonomous.

Legendre transform. Maximising $\langle \lambda, v \rangle - \frac{1}{2} \sigma^2 \|\lambda\|^2$ over λ gives the stationarity condition $v = \sigma^2 \lambda$, i.e. $\lambda^* = v/\sigma^2$; substituting back,

$$\Lambda_x^*(v) = \langle \lambda^*, v \rangle - \frac{1}{2} \sigma^2 \|\lambda^*\|^2 = \frac{1}{2\sigma^2} \|v\|^2,$$

again independent of x . The local rate is thus the quadratic form dual to Λ_x under Legendre duality, with reciprocal coefficient.

Pathwise rate function. Inserting the constant integrand $\Lambda_{\phi(t)}^* = \frac{1}{2\sigma^2} \|\cdot\|^2$ into the drifted rate function $I_M^V(x) = \inf_{\phi} \int_0^1 \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) dt$, in which the drift enters only as a shift of the realised velocity, collapses it to

$$I_M^V(x) = \inf_{\substack{\phi \in \text{AC}([0,1]; M) \\ \phi(0)=x_0, \phi(1)=x}} \frac{1}{2\sigma^2} \int_0^1 \|\dot{\phi}(t) - V(\phi(t))\|^2 dt.$$

This is the geodesic-walk analogue of the Freidlin–Wentzell rate for the small-noise diffusion

$$dX = V(X) dt + n^{-1/2} \sigma dB^M.$$

Example 6.7 (Explicit rate function for a Killing drift). We continue with the isotropic Gaussian family of Example 6.6 and take the drift V to be a *Killing field*: a vector field whose flow $(\psi_t)_{t \in \mathbb{R}}$ consists of isometries. Recall from Theorem 6.5 that the drifted rate function is

$$I_M^V(x) = \inf_{\substack{\phi \in \text{AC}([0,1]; M) \\ \phi(0)=x_0, \phi(1)=x}} \int_0^1 \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) dt. \quad (4)$$

For a Killing drift this reduces to the undrifted rate of [6], evaluated at the endpoint carried back by the drift flow:

$$I_M^V(x) = I_M^0(\psi_{-1}(x)), \quad I_M^0(y) := \inf_{\substack{\eta \in \text{AC}([0,1]; M) \\ \eta(0)=x_0, \eta(1)=y}} \int_0^1 \Lambda_{\eta(t)}^*(\dot{\eta}(t)) dt. \quad (5)$$

To see this, let ϕ be admissible for (4) and set $\eta(t) := \psi_{-t}(\phi(t))$. The map $(s, p) \mapsto \psi_{-s}(p)$ is smooth, hence locally Lipschitz in both arguments, so η is again absolutely continuous, and at every t at which ϕ is differentiable the chain rule gives

$$\dot{\eta}(t) = -V(\eta(t)) + (d\psi_{-t})_{\phi(t)} \dot{\phi}(t) = (d\psi_{-t})_{\phi(t)} (\dot{\phi}(t) - V(\phi(t))),$$

where the second equality uses $(d\psi_{-t})_{\phi(t)} V(\phi(t)) = V(\eta(t))$, the invariance of a vector field under its own flow. Only now does the Killing property enter: $(d\psi_{-t})_{\phi(t)}$ is a linear isometry, so $\|\dot{\eta}(t)\| = \|\dot{\phi}(t) - V(\phi(t))\|$, and since Λ_x^* depends only on the norm of its argument and not on the base point

(Example 6.6), the integrands in (4) and (5) agree pointwise. The map $\phi \mapsto \eta$ is a bijection of the absolutely continuous paths started at x_0 (with inverse $\phi(t) = \psi_t(\eta(t))$, absolutely continuous by the same argument) and carries the constraint $\phi(1) = x$ to $\eta(1) = \psi_{-1}(x)$. The two infima therefore coincide, which is (5).

For the isotropic Gaussian family, $\Lambda_x^*(v) = \|v\|^2/(2\sigma^2)$, and $I_M^0(y) = d(x_0, y)^2/(2\sigma^2)$: by Cauchy–Schwarz the infimum of $\int_0^1 \|\dot{\eta}\|^2 dt$ over paths from x_0 to y equals $d(x_0, y)^2$, attained by a constant-speed minimising geodesic. Since ψ_1 is an isometry, $d(x_0, \psi_{-1}(x)) = d(\psi_1(x_0), x)$, so

$$I_M^V(x) = \frac{1}{2\sigma^2} d(\psi_1(x_0), x)^2.$$

A Killing drift therefore transports the undrifted rate function along its own flow and does nothing else: the level sets of I_M^V are the geodesic spheres centred at the time-1 point $\psi_1(x_0)$, where the rate vanishes and the walk concentrates. This is what Figure 8 shows.

On the unit round sphere S^2 , take the rotation field $V(p) = a \times p$, whose flow is rotation about the axis a by angle $t\|a\|$; with $d(p, q) = \arccos\langle p, q \rangle$ this gives

$$I_{S^2}^V(x) = \frac{1}{2\sigma^2} (\arccos\langle R_{a, \|a\|} x_0, x \rangle)^2,$$

where $R_{a, \theta}$ denotes rotation about a by angle θ . Here $\|V(p)\| = \|a\| \sin \theta(p)$ with $\theta(p)$ the angle between a and p , so $\|V\|_\infty = \|a\|$, attained on the great circle orthogonal to a . Bounded geometry is immediate, with $\text{inj}(S^2) = \pi$ and $\|\text{Rm}\|_\infty = 1$, and the standing hypothesis reads $\|V\|_\infty = \|a\| < \text{inj}(S^2) = \pi$: the time-1 flow rotates by less than a half turn.

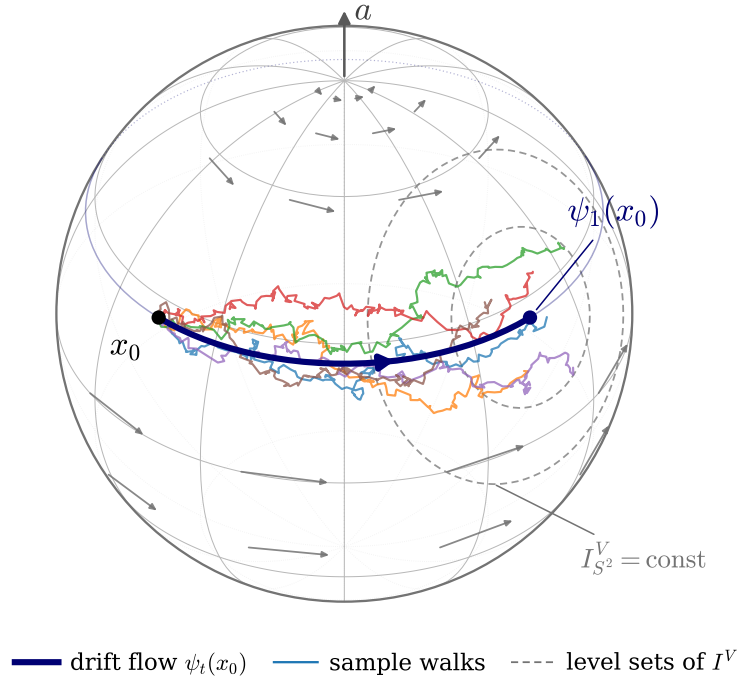


Figure 8: A Killing drift on S^2 , the walk it carries, and the level sets of its rate function. The drift is the rotation field $V(p) = a \times p$, whose flow ψ_t turns the sphere about a through the angle $t\|a\|$. The thin coloured curves are sample walks of Definition 5.3 started at x_0 , with the isotropic Gaussian increments of Example 6.6 in each tangent plane, whose unbounded support puts them under Theorem 6.5; they concentrate on the thick arc $t \mapsto \psi_t(x_0)$, which the drift traverses in unit time. Here $I_{S^2}^V(x) = d(\psi_1(x_0), x)^2/(2\sigma^2)$, so its level sets are the dashed geodesic circles about $\psi_1(x_0)$, and the endpoints scatter across them.

7 LDP for drifted random walks in $\mathbb{R}^d$

Having stated the main result (Theorem 6.3), we now turn to its proof. This section is a warm-up: we prove the drifted large deviation principle in $\mathbb{R}^d$, where the geometric complications are absent, so that the four-step contraction-principle strategy stands out clearly. The Riemannian proof of the next section (Section 8) follows the same four steps, and we shall refer back to the present one as the blueprint.

Before moving to manifolds, we prove the drifted LDP in $\mathbb{R}^d$. This illustrates the contraction principle strategy in a setting where the geometric complications are absent.

7.1 Setup

Let $X_1, X_2, \dots$ be i.i.d. random vectors in $\mathbb{R}^d$ with cumulant generating function $\Lambda(\lambda) = \log \mathbb{E}[e^{\langle \lambda, X_1 \rangle}]$ satisfying Cramér's condition: $\Lambda(\lambda) < \infty$ for all $\lambda \in \mathbb{R}^d$. Let $m = \mathbb{E}[X_1]$ and let $\Lambda^*(w) = \sup_{\lambda} [\langle \lambda, w \rangle - \Lambda(\lambda)]$ be the Fenchel–Legendre transform.

Let $V : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a Lipschitz vector field with constant L , and fix a starting point $x_0 \in \mathbb{R}^d$. We define the *drifted random walk* with step size $h = 1/n$ by

$$Y_0^n = x_0, \quad Y_{k+1}^n = Y_k^n + h V(Y_k^n) + h X_{k+1}.$$

7.2 The undrifted walk and Mogulskii's theorem

We define the *undrifted walk* by $Z_0^n = x_0$, $Z_{k+1}^n = Z_k^n + h X_{k+1}$, with piecewise-linear interpolation $Z^n(t)$ that has slope X_{k+1} on $[t_k, t_{k+1}]$.

Mogulskii's theorem ([3, Section 5.1]) then tells us that Z^n satisfies an LDP on $C([0, 1]; \mathbb{R}^d)$ with speed n and good rate function

$$J(\psi) = \begin{cases} \int_0^1 \Lambda^*(\dot{\psi}(t)) dt & \text{if } \psi(0) = x_0 \text{ and } \psi \text{ is absolutely continuous,} \\ +\infty & \text{otherwise.} \end{cases}$$

7.3 The drift map and its continuity

We now define the *drift map* $\Phi : C_{x_0}([0, 1]; \mathbb{R}^d) \rightarrow C_{x_0}([0, 1]; \mathbb{R}^d)$, where $C_{x_0}([0, 1]; \mathbb{R}^d)$ denotes the space of continuous paths $\psi : [0, 1] \rightarrow \mathbb{R}^d$ with $\psi(0) = x_0$, equipped with the sup-norm. Given an input path ψ , let $\phi = \Phi(\psi)$ be the solution of the ODE

$$\dot{\phi}(t) = V(\phi(t)) + \dot{\psi}(t), \quad \phi(0) = x_0. \quad (6)$$

In words: Φ takes the “noise path” ψ and produces the trajectory obtained by adding the vector field drift on top of it.

Proposition 7.1 (Lipschitz continuity of Φ in $\mathbb{R}^d$). *The map Φ is Lipschitz with respect to the sup-norm: $\|\Phi(\psi_1) - \Phi(\psi_2)\|_\infty \leq e^L \|\psi_1 - \psi_2\|_\infty$.*

Proof. Let $\phi_i = \Phi(\psi_i)$ and $e(t) = \|\phi_1(t) - \phi_2(t)\|$. From (6) we get $\dot{\phi}_1 - \dot{\phi}_2 = V(\phi_1) - V(\phi_2) + \dot{\psi}_1 - \dot{\psi}_2$. Integrating and using the Lipschitz bound on V gives $e(t) \leq \int_0^t L e(s) ds + \|\psi_1 - \psi_2\|_\infty$. The Grönwall inequality then gives $e(t) \leq e^{Lt} \|\psi_1 - \psi_2\|_\infty$. $\square$

7.4 Contraction principle: path-space LDP

Since Z^n satisfies the LDP with rate J and Φ is continuous, the contraction principle gives us that $\Phi(Z^n)$ satisfies an LDP on $C([0, 1]; \mathbb{R}^d)$ with rate function $I_{\text{path}}^V(\phi) = \inf\{J(\psi) : \Phi(\psi) = \phi\}$.

The condition $\Phi(\psi) = \phi$ determines ψ uniquely: from (6), $\dot{\psi}(t) = \dot{\phi}(t) - V(\phi(t))$. This means the infimum in the contraction principle reduces to evaluation.

$$I_{\text{path}}^V(\phi) = \int_0^1 \Lambda^*(\dot{\phi}(t) - V(\phi(t))) dt. \quad (7)$$

7.5 Exponential equivalence

The drifted walk $\hat{Y}^n$ (the piecewise-linear interpolation of $Y_0^n, \dots, Y_n^n$) differs from $\Phi(Z^n)$ only by the Euler discretisation error. The following proposition makes this precise and shows that the two are exponentially equivalent at speed n .

Proposition 7.2 (Exponential equivalence of $\hat{Y}^n$ and $\Phi(Z^n)$). *For every $\delta > 0$,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\|\hat{Y}^n - \Phi(Z^n)\|_\infty > \delta) = -\infty.$$

In particular, by the principle of exponential equivalence [3, Theorem 4.2.13], $\hat{Y}^n$ satisfies the same LDP as $\Phi(Z^n)$.

Proof. Let $T_n := \frac{1}{n} \sum_{k=1}^n \|X_k\|$ denote the total variation of Z^n on $[0, 1]$. For $R > 0$, decompose

$$\mathbb{P}(\|\hat{Y}^n - \Phi(Z^n)\|_\infty > \delta) \leq \mathbb{P}(\|\hat{Y}^n - \Phi(Z^n)\|_\infty > \delta, T_n \leq R) + \mathbb{P}(T_n > R). \quad (8)$$

We treat the two terms separately.

First term: deterministic Euler bound on the bounded-variation event. Write $\phi := \Phi(Z^n)$, let $Y_k := Y_k^n$ be the drifted iterates, and set $e_k := \|Y_k - \phi(t_k)\|$. We first confine both processes on the event $\{T_n \leq R\}$. Since V is Lipschitz, $\|V(z)\| \leq \|V(x_0)\| + L\|z - x_0\|$. From the integrated ODE $\phi(t) = x_0 + \int_0^t V(\phi(s)) ds + (Z^n(t) - x_0)$ and $\|Z^n(t) - x_0\| \leq T_n \leq R$, Grönwall gives $\|\phi - x_0\|_\infty \leq (\|V(x_0)\| + R)e^L =: R'$; the discrete Grönwall lemma gives the same bound for $\max_k \|Y_k - x_0\|$. Both processes therefore stay in $\bar{B}(x_0, R')$, on which $\|V\| \leq B_R := \|V(x_0)\| + LR'$.

For the one-step comparison, note that

$$\phi(t_{k+1}) = \phi(t_k) + \int_{t_k}^{t_{k+1}} V(\phi(s)) ds + h X_{k+1}, \quad Y_{k+1} = Y_k + h V(Y_k) + h X_{k+1},$$

so the common noise increment cancels in the difference and

$$e_{k+1} \leq (1 + Lh) e_k + L \int_{t_k}^{t_{k+1}} \|\phi(s) - \phi(t_k)\| ds.$$

On the substep, $\dot{\phi}(s) = V(\phi(s)) + X_{k+1}$, so $\|\phi(s) - \phi(t_k)\| \leq (B_R + \|X_{k+1}\|)(s - t_k)$ and the integral is at most $(B_R + \|X_{k+1}\|)h^2/2$. Unrolling the recursion with $e_0 = 0$ and $(1 + Lh)^k \leq e^L$,

$$e_k \leq \frac{Lh^2}{2} e^L \sum_{j=1}^n (B_R + \|X_j\|) = \frac{Le^L}{2} (B_R + T_n) h \leq \frac{Le^L}{2} (B_R + R) h.$$

Between grid points, for $t \in [t_k, t_{k+1}]$ both paths carry the same noise term: $\hat{Y}^n(t) = Y_k + (t - t_k)(V(Y_k) + X_{k+1})$ and $\phi(t) = \phi(t_k) + \int_{t_k}^t V(\phi(s)) ds + (t - t_k)X_{k+1}$, giving

$$\|\hat{Y}^n(t) - \phi(t)\| \leq e_k + (t - t_k)\|V(Y_k)\| + \int_{t_k}^t \|V(\phi(s))\| ds \leq e_k + 2B_R h.$$

Combining with our bound for e_k on $\{T_n \leq R\}$ we have $\|\hat{Y}^n - \Phi(Z^n)\|_\infty \leq C(R)/n$ with $C(R) := \frac{Le^L}{2}(B_R + R) + 2B_R$, and for $n > C(R)/\delta$ the first term in (8) vanishes.

Second term: a Chernoff bound. Cramér's condition implies $\Lambda_{\text{abs}}(\lambda) := \log \mathbb{E}[e^{\lambda\|X_1\|}] < \infty$ for every $\lambda \geq 0$: indeed $e^{\lambda\|x\|} \leq \prod_{i=1}^d e^{\lambda|x_i|} \leq \sum_{\sigma \in \{-1,1\}^d} e^{\lambda\langle \sigma, x \rangle}$, so $\mathbb{E}[e^{\lambda\|X_1\|}] \leq \sum_{\sigma \in \{-1,1\}^d} e^{\Lambda(\lambda\sigma)} < \infty$. By Markov's inequality and independence, for any $\lambda > 0$,

$$\mathbb{P}(T_n > R) \leq e^{-n\lambda R} \mathbb{E}\left[e^{\lambda \sum_{k=1}^n \|X_k\|}\right] = e^{-n(\lambda R - \Lambda_{\text{abs}}(\lambda))}.$$

Combining. Given $\ell > 0$, take $\lambda = 1$ and $R := \ell + \Lambda_{\text{abs}}(1)$. Then for $n > C(R)/\delta$ the first term in (8) vanishes, and the second term gives

$$\frac{1}{n} \log \mathbb{P}(\|\hat{Y}^n - \Phi(Z^n)\|_\infty > \delta) \leq -\ell.$$

Since $\ell > 0$ was arbitrary, the limsup equals $-\infty$. □

7.6 Endpoint rate function

Applying the contraction principle to the evaluation map $\text{ev}_1(\phi) = \phi(1)$ yields:

Theorem 7.3 (Drifted Cramér theorem in $\mathbb{R}^d$). *The drifted empirical averages Y_n^n satisfy the LDP in $\mathbb{R}^d$ with speed n and good rate function*

$$I^V(y) = \inf_{\substack{\phi \in \text{AC}([0,1]; \mathbb{R}^d) \\ \phi(0)=x_0, \phi(1)=y}} \int_0^1 \Lambda^*(\dot{\phi}(t) - V(\phi(t))) dt.$$

Remark 7.4 (Zero of the rate function). Consider the path $\phi(t) = \gamma(t)$, the integral curve of $\dot{\gamma} = V(\gamma) + m$ with $\gamma(0) = x_0$. Its noise velocity is $\dot{\phi}(t) - V(\phi(t)) = m$, so the cost is $\int_0^1 \Lambda^*(m) dt = 0$ (since Λ^* vanishes at the mean). This gives us $I^V(\gamma(1)) = 0$, confirming that the walk concentrates around the time-1 flow of $V + m$.

8 Proof of Theorem 6.3

8.1 Proof strategy

Before the details, we outline the argument. It follows the same four-step contraction-principle strategy as the Euclidean warm-up of Section 7, with the Riemannian Mogulskii theorem of [6] replacing the classical one and Jacobi-field estimates replacing the flat Grönwall bound.

Step 1 (undrifted path LDP). The rescaled undrifted walk Z^n satisfies a path-space LDP on $C([0, 1]; M)$ by the Riemannian Mogulskii theorem (Theorem 8.1).

Step 2 (the drift map). We introduce the drift map Φ , which turns a noise path into the trajectory obtained by adding the vector-field drift on top of it, and show it is well-defined and continuous on the admissible path space (Subsection 8.2.1–8.2.3).

Step 3 (contraction). Applying the contraction principle to Φ transports the undrifted path LDP to a path-space LDP for the drifted flow (Theorem 6.2).

Step 4 (exponential equivalence and endpoint). We show the interpolated drifted walk is exponentially equivalent to $\Phi(Z^n)$ (Subsection 8.4), so it satisfies the same path-space LDP; contraction with the evaluation map $\phi \mapsto \phi(1)$ then yields the endpoint LDP of Theorem 6.3.

8.2 Path-space LDP for the undrifted walk

The $\frac{1}{n}$ -rescaled geodesic random walk Z^n (without drift) satisfies the LDP on $C([0, 1]; M)$ by the Riemannian Mogulskii theorem:

Theorem 8.1 (Riemannian Mogulskii, [6, Theorem 5.2]). *The random paths Z^n satisfy the LDP on $C([0, 1]; M)$ with speed n and good rate function*

$$J_M(\psi) = \begin{cases} \int_0^1 \Lambda_{\psi(t)}^*(\dot{\psi}(t)) dt & \text{if } \psi \in \text{AC}([0, 1]; M) \text{ and } \psi(0) = x_0, \\ +\infty & \text{otherwise.} \end{cases}$$

This is the Riemannian analogue of the classical Mogulskii theorem used in Section 7, proven in [6].

We define the *drift map* $\Phi : C_{x_0}([0, 1]; M) \rightarrow C_{x_0}([0, 1]; M)$ as follows; here $C_{x_0}([0, 1]; M)$ is the space of continuous paths $\phi : [0, 1] \rightarrow M$ with $\phi(0) = x_0$, equipped with the uniform distance $d_\infty(\phi_1, \phi_2) := \sup_{t \in [0, 1]} d(\phi_1(t), \phi_2(t))$. Given an absolutely continuous input path ψ with $\psi(0) = x_0$, the image $\phi = \Phi(\psi)$ is the solution of

$$\phi(0) = x_0, \quad \dot{\phi}(t) = V(\phi(t)) + P_{\psi(t) \rightarrow \phi(t)} \dot{\psi}(t), \quad (9)$$

where $P_{\psi(t) \rightarrow \phi(t)} : T_{\psi(t)}M \rightarrow T_{\phi(t)}M$ is parallel transport along the minimising geodesic from $\psi(t)$ to $\phi(t)$.

For this transport map to be well defined, the distance between $\psi(t)$ and $\phi(t)$ must stay below the local injectivity radius. We verify this below through a single a priori estimate: every solution of (9) satisfies

$$d(\psi(t), \phi(t)) \leq \|V\|_\infty t \leq \|V\|_\infty \quad (t \in [0, 1]). \quad (10)$$

Since ψ stays in $\overline{B}(x_0, r)$, estimate (10) confines ϕ to the compact tube

$$K_0 := \overline{B}(x_0, r + \|V\|_\infty) \subset M$$

(compact by Hopf–Rinow, Theorem 2.25), on which the injectivity radius is uniformly positive:

$$\iota_0 := \text{inj}(K_0) = \inf_{p \in K_0} \iota(p) > 0.$$

Combined with the standing assumption of our main theorem,

$$\|V\|_\infty < \iota_0, \tag{11}$$

the estimate gives $d(\psi(t), \phi(t)) \leq \|V\|_\infty < \iota_0$ for all t , so the two paths never leave each other's injectivity radius and $P_{\psi(t) \rightarrow \phi(t)}$ is well defined throughout.

8.2.1 Well-posedness of the drift ODE

Before discussing the continuity properties of Φ we should verify that Φ is well-defined: for every admissible input ψ , the ODE (9) has a unique absolutely continuous solution on $[0, 1]$. The argument is a chart reduction to the classical Carathéodory–Picard theorem on $\mathbb{R}^d$. The key inputs are an a priori bound that keeps $\phi(t)$ within the injectivity radius of $\psi(t)$, and a finite covering of $[0, 1]$ by time intervals on which ϕ stays inside a single normal coordinate chart.

The class of admissible inputs is, in the notation of Definition 2.27 and Lemma 2.28,

$$\text{AC}_{x_0, r}([0, 1]; M) := \{\psi \in \text{AC}([0, 1]; M) : \psi(0) = x_0, \|\dot{\psi}\|_\infty \leq r\},$$

which is the natural one for the LDP application: the increment bound $\text{supp}(\mu_x) \subseteq \overline{B}(0, r)$ forces all interpolated walks into this class, and so do all paths of finite J_M cost (both claims are verified in Corollary 8.10 below).

The tool is the Carathéodory–Picard theorem (Theorem 3.18), whose hypotheses (C1) to (C4) the chart form of (9) is checked against below. Recall from Subsection 3.4 that a solution means an absolutely continuous function satisfying the ODE almost everywhere.

We show well-posedness in three steps. The first decouples Φ from the parallel-transport subtleties: any solution of (9) has uniformly bounded speed and stays close to ψ . The bound is $d(\psi(t), \phi(t)) < \iota_0 \leq \iota(\psi(t))$, and by Lemma 2.24(iii) that makes the minimising geodesic from $\psi(t)$ to $\phi(t)$ unique, so $P_{\psi(t) \rightarrow \phi(t)}$ is well-defined. The radius we compare against is the one at $\psi(t)$, the point the transport starts from.

Lemma 8.2 (A priori bound). *Let $\psi \in \text{AC}_{x_0, r}([0, 1]; M)$ and suppose $\phi : [0, T] \rightarrow M$ is an absolutely continuous solution of (9) on $[0, T] \subseteq [0, 1]$. Then for all $t \in [0, T]$,*

$$\|\dot{\phi}(t)\| \leq r + \|V\|_\infty, \quad d(\phi(t), \psi(t)) \leq \|V\|_\infty \cdot t < \iota_0.$$

Proof. The speed bound is immediate from (9) since parallel transport is an isometry: $\|\dot{\phi}(t)\| \leq \|V(\phi(t))\| + \|\dot{\psi}(t)\| \leq \|V\|_\infty + r$.

For the separation bound, set $\delta(t) := d(\phi(t), \psi(t))$. We argue by a variation-field calculation in the spirit of Lemma 4.6, with the variation now interpolating between ψ and ϕ rather than along a single flow. Throughout the proof we work on the open set

$$\mathcal{O} := \{t \in [0, T] : \delta(t) < \iota_0\},$$

which contains $t = 0$. Since $\psi(t) \in \overline{B}(x_0, r) \subseteq K_0$ we have $\iota(\psi(t)) \geq \iota_0$, so on $\mathcal{O}$ the minimising geodesic from $\psi(t)$ to $\phi(t)$ is unique, and we define $v(t) := \log_{\psi(t)} \phi(t) \in T_{\psi(t)}M$ with $\|v(t)\| = \delta(t)$.

The fact that $d : M \times M \rightarrow \mathbb{R}$ is 1-Lipschitz combined with the absolute continuity of $t \mapsto (\phi(t), \psi(t))$ shows that δ is absolutely continuous on $\mathcal{O}$ (concretely, $|\delta(t) - \delta(s)| \leq d(\phi(t), \phi(s)) + d(\psi(t), \psi(s))$ by the triangle inequality, and each term on the right is the increment of an absolutely continuous function, so δ inherits absolute continuity).

Consider the geodesic variation

$$\Gamma : \mathcal{O} \times [0, 1] \rightarrow M, \quad \Gamma(t, s) := \text{Exp}_{\psi(t)}(s v(t)),$$

so $\Gamma(t, 0) = \psi(t)$ and $\Gamma(t, 1) = \phi(t)$. For each fixed t , $s \mapsto \Gamma(t, s)$ is a constant-speed geodesic of length exactly $\delta(t)$. Write $T(t, s) := \partial_s \Gamma$ and $W(t, s) := \partial_t \Gamma$ for the longitudinal and transverse variation fields. At the endpoints,

$$T(t, 0) = v(t), \quad T(t, 1) = P_{\psi(t) \rightarrow \phi(t)} v(t),$$

both of norm $\delta(t)$, and

$$W(t, 0) = \dot{\psi}(t), \quad W(t, 1) = \dot{\phi}(t).$$

The family Γ is exactly the configuration of Proposition 2.23: each slice $\Gamma(t, \cdot)$ is a geodesic of constant speed, and $\delta(t) = \|T(t, s)\| = L_g(\Gamma(t, \cdot))$ is the separation of its two ends. That proposition gives

$$\delta(t) \delta'(t) = \langle W(t, 1), T(t, 1) \rangle - \langle W(t, 0), T(t, 0) \rangle. \quad (12)$$

Proposition 2.23 is stated for a family that is smooth in t , whereas Γ inherits from ψ and ϕ only absolute continuity in that variable. This is enough, for three reasons. The separation δ is absolutely continuous on $\mathcal{O}$, as shown above. The map $(q, p) \mapsto \log_q p$ is smooth on the unique-minimising regime, so $t \mapsto \Gamma(t, s)$ is differentiable wherever ψ and ϕ are, and the transverse field $W = \partial_t \Gamma$ therefore exists at almost every t . And the computation behind the proposition differentiates δ^2 in t and integrates $\partial_s \langle W, T \rangle$ in s , both of which are valid at any such t with $\delta(t) > 0$. The identity (12) thus holds for a.e. $t \in \mathcal{O}$ with $\delta(t) > 0$, and since only its integrated form is used below, nothing is lost.

Using that parallel transport is an isometry, write $\langle W(t, 0), T(t, 0) \rangle = \langle P W(t, 0), P T(t, 0) \rangle = \langle P \dot{\psi}(t), T(t, 1) \rangle$, where $P := P_{\psi(t) \rightarrow \phi(t)}$. Substituting in (12):

$$\delta(t) \delta'(t) = \langle \dot{\phi}(t) - P \dot{\psi}(t), T(t, 1) \rangle.$$

By (9), $\dot{\phi}(t) - P \dot{\psi}(t) = V(\phi(t))$. Applying Cauchy–Schwarz and dividing by $\delta(t)$ (where positive):

$$\delta'(t) \leq \|V(\phi(t))\| \leq \|V\|_\infty \quad \text{for a.e. } t \in \mathcal{O}.$$

Since δ is absolutely continuous on $\mathcal{O}$ with $\delta(0) = 0$, integration gives $\delta(t) \leq \|V\|_\infty \cdot t$ on $\mathcal{O}$. By the standing assumption (11), this is strictly less than ι_0 , so $\mathcal{O}$ is in fact all of $[0, T]$ and the bound holds throughout. $\square$

Remark 8.3. If the set $E := \{t \in \mathcal{O} : \delta(t) = 0\}$ has positive Lebesgue measure, the needed bound $\delta'(t) \leq \|V\|_\infty$ on E follows from the fact that an absolutely continuous function constant on a set has derivative zero almost everywhere on that set.

The a priori bound confirms the set-up above: since $\psi([0, 1]) \subseteq \overline{B}(x_0, r)$ and $d(\phi(t), \psi(t)) \leq \|V\|_\infty$, every solution stays in $K_0 = \overline{B}(x_0, r + \|V\|_\infty)$, and the joint trajectory (ψ, ϕ) remains in the unique-minimising-geodesic regime by (11).

Single-chart reduction. Set the chart radius

$$\rho := \frac{1}{4}(\iota_0 - \|V\|_\infty) > 0,$$

positive by (11). Fix $p_\star \in K_0$; since $\rho < \iota_0 \leq \iota(p_\star)$, the normal coordinate chart $\varphi_{p_\star} := \log_{p_\star}$ is a diffeomorphism on a neighbourhood of $\overline{B}(p_\star, \rho)$, and we set

$$K' := \varphi_{p_\star}(B(p_\star, \rho)) \subset \mathbb{R}^d,$$

a bounded open neighbourhood of the origin. If $\phi(t)$ stays inside $B(p_\star, \rho)$ on a subinterval $J = [t_0, t_1]$, then $y := \varphi_{p_\star} \circ \phi : J \rightarrow K'$ is absolutely continuous and satisfies

$$\dot{y}(t) = F^\star(t, y(t)) \quad \text{a.e. } t \in J, \quad (13)$$

where

$$F^\star(t, y) := (d\varphi_{p_\star})_{\varphi_{p_\star}^{-1}(y)} \left[V(\varphi_{p_\star}^{-1}(y)) + P_{\psi(t) \rightarrow \varphi_{p_\star}^{-1}(y)} \dot{\psi}(t) \right].$$

The radius ρ is chosen precisely so that whenever the source $\psi(t)$ lies within $\iota_0 - \rho$ of the chart centre $p_\star$, every target $\varphi_{p_\star}^{-1}(y)$ with $y \in K'$ lies within ι_0 of $\psi(t)$, which keeps the parallel transport in $F^\star$ well-defined. This is the content of the next lemma.

Lemma 8.4 (The chart ODE is Carathéodory). *Let $J = [t_0, t_1] \subseteq [0, 1]$ be such that $d(\psi(t), p_\star) < \iota_0 - \rho$ for all $t \in J$. Then $F^\star$ is defined on $J \times K'$ and satisfies hypotheses (C1)–(C4) of Theorem 3.18 there.*

Proof. Well-definedness. For $y \in K'$ we have $d(\varphi_{p_\star}^{-1}(y), p_\star) < \rho$, so for $t \in J$

$$d(\psi(t), \varphi_{p_\star}^{-1}(y)) \leq d(\psi(t), p_\star) + d(p_\star, \varphi_{p_\star}^{-1}(y)) < (\iota_0 - \rho) + \rho = \iota_0 \leq \iota(\psi(t)),$$

so $(\psi(t), \varphi_{p_\star}^{-1}(y)) \in \mathcal{U}$ and the parallel transport $P_{\psi(t) \rightarrow \varphi_{p_\star}^{-1}(y)}$ along the unique minimising geodesic is well-defined. Hence $F^\star$ is defined on $J \times K'$.

(C1) For fixed y , $F^\star(t, y) = \Theta_y(\psi(t), \dot{\psi}(t))$ for a smooth map Θ_y , depending on y . The curve $\psi(t)$ is absolutely continuous, and thus $\dot{\psi}(t)$ exists a.e. and is Lebesgue measurable. The composition of a measurable function with a smooth one is measurable again.

(C2) Fix t . The maps $y \mapsto \varphi_{p_\star}^{-1}(y)$, $p \mapsto V(p)$ and $p \mapsto (d\varphi_{p_\star})_p$ are smooth, and by Lemma 2.29 so is $p \mapsto P_{\psi(t) \rightarrow p}$ on the unique-minimising regime, which by the well-definedness step contains $\{\psi(t)\} \times \varphi_{p_\star}^{-1}(K')$. Since $\dot{\psi}(t) \in T_{\psi(t)}M$ is a fixed vector for fixed t , the map $y \mapsto P_{\psi(t) \rightarrow \varphi_{p_\star}^{-1}(y)} \dot{\psi}(t)$ is smooth; hence $F^\star(t, \cdot)$ is a composition of smooth maps on K' and is in particular continuous.

(C3) Since $\|\dot{\psi}(t)\| \leq r$ and parallel transport is an isometry, $\|F^\star(t, y)\| \leq \|(d\varphi_{p_\star})_{\varphi_{p_\star}^{-1}(y)}\|_{\text{op}} (r + \|V\|_\infty)$. The operator norm of $(d\varphi_{p_\star})_{\varphi_{p_\star}^{-1}(y)}$ is continuous in y and attains a finite maximum $C_{K'}$ on the compact set $\overline{K'}$, giving $\|F^\star(t, y)\| \leq C_{K'} (r + \|V\|_\infty)$ on K' .

(C4) Note that $F^\star$ need not be continuous in t , since $\dot{\psi}$ is merely measurable; the Lipschitz constant must therefore be obtained uniformly in t rather than by compactness in (t, y) . For a.e. fixed t write $\partial_y F^\star(t, y) = G(y, \psi(t), \dot{\psi}(t))$, where G is smooth in the first two arguments and linear in the velocity argument (the noise enters $F^\star$ linearly through $P_{\psi(t) \rightarrow \varphi_{p_\star}^{-1}(y)} \dot{\psi}(t)$). The arguments of G range over the compact set $\overline{K'} \times \psi(J) \times \overline{B}(0, r)$, which is contained in the open unique-minimising regime since $\max_{t \in J} d(\psi(t), p_\star) < \iota_0 - \rho$; hence G is bounded there. Since K' is convex (the $\log_{p_\star}$ -image of a metric ball is a Euclidean ball), the mean value theorem gives a constant $L_{p_\star}$ (independent of t , since the bound on G is) with $\|F^\star(t, y_1) - F^\star(t, y_2)\| \leq L_{p_\star} \|y_1 - y_2\|$ for a.e. $t \in J$ and all $y_1, y_2 \in K'$. $\square$

Remark 8.5. Note that Theorem 3.18 is stated for a function $F : I \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. However, multiplication by a smooth cutoff $\chi : \mathbb{R}^d \rightarrow [0, 1]$ supported in a relatively compact neighbourhood of $\bar{K}'$ and equal to 1 on K' yields a global extension $\tilde{F}^* := \chi \cdot F^* : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfying (C1)–(C4) globally. The cutoff extension preserves (C1)–(C4) globally with constants depending only on p_* and K' (and not on F^* outside $\bar{K}'$).

The next corollary will be the building block of the patching argument below: it gives existence and uniqueness of the drift ODE on any subinterval short enough that the trajectory is confined to a single normal coordinate chart.

Corollary 8.6. *Let $J = [t_0, t_1] \subseteq [0, 1]$ and $q_0 \in B(p_*, \rho)$. Suppose $d(\psi(t), p_*) < \iota_0 - \rho$ for all $t \in J$, and that any absolutely continuous solution of (9) starting from q_0 at time t_0 remains inside $B(p_*, \rho)$ throughout J (both verified in the patching argument below). Then (9) has a unique absolutely continuous solution on J with initial value q_0 .*

Proof. By the source-closeness hypothesis and Lemma 8.4, F^* is defined and Carathéodory on $J \times K'$, so its cutoff extension $\tilde{F}^*$ (Remark 8.5) satisfies (C1)–(C4) on $[0, 1] \times \mathbb{R}^d$. Apply Theorem 3.18 to $\dot{y} = \tilde{F}^*(t, y)$ with initial condition $y(t_0) = \varphi_{p_*}(q_0)$. This yields a unique absolutely continuous $\tilde{y} : J \rightarrow \mathbb{R}^d$. By the confinement hypothesis, $\tilde{y}$ takes values in K' , where $\tilde{F}^* = F^*$, so $\tilde{y}$ is also the unique solution of the original chart ODE (13). Pushing forward via $\varphi_{p_*}^{-1}$ gives the unique solution $\phi = \varphi_{p_*}^{-1} \circ \tilde{y}$ of (9) on J . $\square$

Patching and conclusion. We now choose a time-step size $\tau > 0$ small enough that, on each substep, the candidate solution stays in the chart ball $B(p_*, \rho)$ and the source ψ stays close enough to the chart centre for Corollary 8.6 to apply. Set

$$\tau := \frac{\rho}{r + \|V\|_\infty + 1} > 0.$$

Consider a substep $[t_0, t_0 + \tau]$ with chart centred at $p_* := \phi(t_0) \in K_0$, where ϕ is the partial solution built so far (so the a priori bound of Lemma 8.2 holds, giving $\|\dot{\phi}\|_\infty \leq r + \|V\|_\infty$ and $d(\psi(t_0), p_*) \leq \|V\|_\infty$). Then:

- *Confinement.* For $t \in [t_0, t_0 + \tau]$, by the speed bound $\|\dot{\phi}\|_\infty \leq r + \|V\|_\infty$ of Lemma 8.2, $d(\phi(t), p_*) \leq (r + \|V\|_\infty)(t - t_0) \leq (r + \|V\|_\infty)\tau < \rho$, so ϕ stays in $B(p_*, \rho)$.
- *Source-closeness.* Using $\|\dot{\psi}\|_\infty \leq r$,

$$d(\psi(t), p_*) \leq d(\psi(t), \psi(t_0)) + d(\psi(t_0), p_*) \leq r\tau + \|V\|_\infty < \rho + \|V\|_\infty < \iota_0 - \rho,$$

the last step because $\|V\|_\infty + 2\rho = \frac{1}{2}(\iota_0 + \|V\|_\infty) < \iota_0$.

Both hypotheses of Corollary 8.6 hold, so it gives existence and uniqueness on the substep.

Theorem 8.7 (Well-posedness of the drift ODE). *Under the standing assumption (11), for every $\psi \in \text{AC}_{x_0, r}([0, 1]; M)$ the ODE (9) has a unique absolutely continuous solution $\phi = \Phi(\psi) : [0, 1] \rightarrow M$, which satisfies the a priori bounds of Lemma 8.2.*

Proof. Partition $[0, 1]$ as $0 = t_0 < t_1 < \dots < t_N = 1$ with $t_{k+1} - t_k \leq \tau$, where $N := \lceil 1/\tau \rceil$ and $\tau = \rho/(r + \|V\|_\infty + 1)$ is the step size fixed above. Build ϕ inductively: starting from $\phi(t_0) = x_0$ and the chart centred at $p_\star^{(0)} := x_0$, Corollary 8.6 applies on $[t_0, t_1]$ because the substep has length at most τ , which is what the confinement and source-closeness checks above need, producing a unique $\phi^{(0)} : [t_0, t_1] \rightarrow M$. Given $\phi^{(k)}$ on $[t_0, t_{k+1}]$, take the chart centred at $p_\star^{(k+1)} := \phi^{(k)}(t_{k+1}) \in K_0$ (still in K_0 by Lemma 8.2) and apply Corollary 8.6 on $[t_{k+1}, t_{k+2}]$ to extend.

For uniqueness, if ϕ_1, ϕ_2 are two solutions on $[0, 1]$ with the same initial condition, the set $S := \{t : \phi_1|_{[0,t]} = \phi_2|_{[0,t]}\}$ contains 0, is closed by continuity, and is open by the local uniqueness of Corollary 8.6 applied at any $t \in S$. Connectedness gives $S = [0, 1]$. $\square$

This justifies the definition of the drift map $\Phi : \text{AC}_{x_0, r}([0, 1]; M) \rightarrow \text{AC}_{x_0, r + \|V\|_\infty}([0, 1]; M)$. We now turn to its continuity properties.

8.2.2 Compactness of the admissible path space

Continuity of Φ will follow from two properties of the admissible path space, both consequences of the bounded-increment assumption that defines the model. All interpolated walks have speed at most r , so their laws are carried by $\text{AC}_{x_0, r}([0, 1]; M)$, and

$$\text{AC}_{x_0, r}([0, 1]; M) = \{\psi : [0, 1] \rightarrow M : \psi(0) = x_0, \psi \text{ is } r\text{-Lipschitz with respect to } d\} :$$

indeed $d(\psi(s), \psi(t)) \leq \int_s^t \|\dot{\psi}\| du \leq r|t - s|$, and conversely an r -Lipschitz curve is absolutely continuous with $\|\dot{\psi}\| \leq r$ almost everywhere (working in charts and using that Lipschitz functions of one variable are differentiable almost everywhere). On this set the sup-norm topology is much better behaved than on the full path space:

Proposition 8.8 (Compactness of the admissible paths). *For every $R > 0$, the set $\text{AC}_{x_0, R}([0, 1]; M)$ is compact in $(C_{x_0}([0, 1]; M), d_\infty)$.*

Proof. All paths in the set take values in $\overline{B}(x_0, R)$, which is compact by Hopf–Rinow (Theorem 2.25), meaning that the set $\{\psi(t) : \psi \in \text{AC}_{x_0, R}([0, 1]; M), t \in [0, 1]\}$ is relatively compact. Furthermore the paths are uniformly R -Lipschitz, hence equicontinuous. By the Arzelà–Ascoli theorem the set is relatively compact in $C([0, 1]; M)$. It is also closed: suppose $\psi_n \rightarrow \psi$ uniformly. Then:

- $\psi(0) = \lim_n \psi_n(0) = x_0$
- $d(\psi(s), \psi(t)) = \lim_n d(\psi_n(s), \psi_n(t)) \leq \lim_n R|s - t| = R|s - t|$, using the continuity of d .

This shows that the uniform limit of a sequence of R -Lipschitz paths started at x_0 is again R -Lipschitz and started at x_0 . $\square$

The second ingredient is that, within $\text{AC}_{x_0, R}$, uniform convergence automatically carries the velocity information, in the weak sense. The two facts we need are the general velocity lemmas of Subsection 3.2.2: Lemma 3.14 (uniform convergence forces weak- L^1 velocity convergence) and Lemma 3.15 (weak- $L^1 \times$ strong pairing). They are stated there in the weak- L^1 form that also serves the unbounded-increment argument of Subsection 9.6; here the paths are uniformly Lipschitz, so the velocities are bounded and the convergence is in fact weak-*. Both concern $\mathbb{R}^d$ -valued paths and are applied below in normal coordinate charts.

8.2.3 Continuity of the drift map

We prove that Φ is continuous on $(AC_{x_0,r}, d_\infty)$. The proof uses the compactness of $AC_{x_0,r}$ (Proposition 8.8) and the two elementary velocity lemmas of Subsection 8.2.2, passing to the limit in the chart form of the drift ODE by weak-* convergence of the velocities.

We record the chart apparatus the proof below refers to. Fix the normal charts of Subsection 8.2.1: for a chart centre $p_\star \in K_0$ the chart is $\varphi_{p_\star} = \log_{p_\star}$, of radius ρ , and τ is the associated time step. In such a chart, writing $y := \varphi_{p_\star} \circ \phi$ and $x := \varphi_{p_\star} \circ \psi$ for an output/input pair, the drift ODE (9) becomes $\dot{y} = b(y) + A(y, x) \dot{x}$ almost everywhere, where

$$b(y) := (d\varphi_{p_\star})_{\varphi_{p_\star}^{-1}(y)} V(\varphi_{p_\star}^{-1}(y)), \quad A(y, x) := (d\varphi_{p_\star})_{\varphi_{p_\star}^{-1}(y)} \circ P_{\varphi_{p_\star}^{-1}(x) \rightarrow \varphi_{p_\star}^{-1}(y)} \circ ((d\varphi_{p_\star})_{\varphi_{p_\star}^{-1}(x)})^{-1}. \quad (14)$$

This is the chart field $F^\star$ of Lemma 8.4, linear in the velocity slot (parallel transport is a linear isometry) and smooth on the configurations that occur (Lemma 2.29). On the compact region in play the chart is bi-Lipschitz with a constant $c \geq 1$ uniform over $p_\star \in K_0$. The differential of $\varphi_{p_\star} = \log_{p_\star}$ and its inverse have operator norms bounded on the compact disk bundle $\{(p, v) \in TM : p \in K_0, \|v\| \leq \|V\|_\infty + 3\rho\}$, which lies below the injectivity radius, by the compactness argument of Lemma 2.33. So an r -Lipschitz input has chart speed $|\dot{x}| \leq cr$. We write $S := \overline{B}(p_\star, \|V\|_\infty + 2\rho)$ for the source region reached from a chart centred at an output point; the choices of Subsection 8.2.1 keep S inside the chart domain and all parallel transports below well-defined and, by Lemma 2.29, smooth.

Proof of Proposition 6.1, via weak- compactness.* Let $\psi_n \rightarrow \psi$ in $(AC_{x_0,r}, d_\infty)$ and set $\phi_n := \Phi(\psi_n)$, $\phi := \Phi(\psi)$. Since $AC_{x_0,r+\|V\|_\infty}$ is d_∞ -compact (Proposition 8.8), it suffices to show that every subsequence of (ϕ_n) has a further subsequence converging uniformly to ϕ . So relabel and assume $\phi_n \rightarrow \phi^\dagger$ uniformly for some $\phi^\dagger \in AC_{x_0,r+\|V\|_\infty}$ (such a limit exists, and lies in this set, by the d_∞ -compactness just invoked); we must prove $\phi^\dagger = \phi$.

Partition $[0, 1]$ with mesh at most τ and centre the k -th chart at $p_\star := \phi^\dagger(t_k)$. Over a substep, $\phi^\dagger$ moves at most $\rho - \tau$, so for all n large enough that $d_\infty(\phi_n, \phi^\dagger) < \tau$ and $d_\infty(\psi_n, \psi) < \tau$, the outputs $\phi_n, \phi^\dagger$ are confined to $B(p_\star, \rho)$ on I_k , the sources ψ_n stay within $\|V\|_\infty + \rho$ of $p_\star$ (by Lemma 8.2, $d(\psi_n, \phi_n) \leq \|V\|_\infty$, while $\phi_n \in B(p_\star, \rho)$), and hence ψ stays within $\|V\|_\infty + \rho + \tau \leq \|V\|_\infty + 2\rho$ of $p_\star$, i.e. inside the source region S . In the chart, with $y_n := \varphi_{p_\star} \circ \phi_n$, $x_n := \varphi_{p_\star} \circ \psi_n$ and $w_n := \dot{x}_n$, the integrated drift ODE reads

$$y_n(t) = y_n(t_k) + \int_{t_k}^t [b(y_n(s)) + A(y_n(s), x_n(s)) w_n(s)] ds, \quad t \in I_k. \quad (15)$$

We pass to the limit term by term. Uniform convergence of ϕ_n and ψ_n and smoothness of the chart give $y_n \rightarrow y^\dagger := \varphi_{p_\star} \circ \phi^\dagger$ and $x_n \rightarrow x := \varphi_{p_\star} \circ \psi$ uniformly on I_k ; hence $b(y_n) \rightarrow b(y^\dagger)$ and $A(y_n, x_n) \rightarrow A(y^\dagger, x)$ uniformly, by continuity of b and A on the compact chart region. The x_n are uniformly Lipschitz ($|w_n| \leq cr$), so Lemma 3.14 gives $w_n \rightharpoonup w := \dot{x}$ weakly in $L^1(I_k)$, and Lemma 3.15 (with $b_n := A(y_n, x_n) \rightarrow A(y^\dagger, x)$ in L^∞ , $v_n := w_n$) lets us pass to the limit in (15): for every $t \in I_k$,

$$y^\dagger(t) = y^\dagger(t_k) + \int_{t_k}^t [b(y^\dagger(s)) + A(y^\dagger(s), x(s)) w(s)] ds.$$

Translating back to M : $\phi^\dagger$ satisfies (9) with input ψ almost everywhere on I_k . Patching over the finitely many substeps and using $\phi^\dagger(0) = \lim_n \phi_n(0) = x_0$, the uniqueness part of Theorem 8.7 gives

$\phi^\dagger = \Phi(\psi) = \phi$. This proves continuity. Since the domain $(\text{AC}_{x_0, r}, d_\infty)$ is compact (Proposition 8.8), the continuity is automatically uniform, as stated. $\square$

8.3 The path-space LDP for the drifted flow

We now combine Theorem 8.1 with Proposition 6.1 through the contraction principle. Two reductions are needed: the Mogulskii LDP must be restricted to the compact set $\text{AC}_{x_0, r}$ on which Φ is defined and continuous, and the contracted rate function must be identified with the drifted integrand of Theorem 6.3.

Lemma 8.9 (Restriction of an LDP to a supporting closed set). *Let $\mathcal{X}$ be a Polish space, let $\{\mu_n\}$ satisfy the LDP on $\mathcal{X}$ at speed n with good rate function I , and let $F \subseteq \mathcal{X}$ be closed with $\mu_n(F) = 1$ for all n and $\{I < \infty\} \subseteq F$. Then $\{\mu_n\}$ satisfies the LDP on F (with the subspace topology) at speed n with good rate function $I|_F$; cf. [3, Lemma 4.1.5].*

Proof. A relatively closed set $C = \tilde{C} \cap F$, with $\tilde{C} \subseteq \mathcal{X}$ closed, satisfies $\mu_n(C) \leq \mu_n(\tilde{C})$ and $\inf_{\tilde{C}} I = \inf_C I$ because $I \equiv +\infty$ off F ; the upper bound follows. A relatively open set $A = \tilde{A} \cap F$, with $\tilde{A}$ open, satisfies $\mu_n(A) = \mu_n(\tilde{A})$ because $\mu_n(F) = 1$, and again $\inf_{\tilde{A}} I = \inf_A I$; the lower bound follows. The sublevel sets of $I|_F$ coincide with those of I , are compact in $\mathcal{X}$, and are contained in the closed set F ; hence they are compact in F . $\square$

Corollary 8.10 (Mogulskii on the admissible paths). *The rescaled undrifted walks Z^n satisfy the LDP on the compact space $(\text{AC}_{x_0, r}, d_\infty)$ with speed n and good rate function J_M .*

Proof. The interpolated walk is piecewise geodesic with speed $\|X_{k+1}^n\| \leq r$ on each substep, so $Z^n \in \text{AC}_{x_0, r}$ almost surely; and $\text{AC}_{x_0, r}$ is closed in $C([0, 1]; M)$ by Proposition 8.8. It remains to check $\{J_M < \infty\} \subseteq \text{AC}_{x_0, r}$. If $J_M(\psi) < \infty$, then ψ is absolutely continuous, $\psi(0) = x_0$, and $\Lambda_{\psi(t)}^*(\dot{\psi}(t)) < \infty$ almost everywhere. Since $\text{supp}(\mu_x) \subseteq \overline{B}(0, r)$ implies $\Lambda_x(\lambda) \leq \log \int \exp(\|\lambda\| \|u\|) \mu_x du \leq r \|\lambda\|$, we have for $\|w\| > r$

$$\Lambda_x^*(w) \geq \sup_{s>0} [s\|w\| - \Lambda_x(s w / \|w\|)] \geq \sup_{s>0} s(\|w\| - r) = +\infty,$$

so $\|\dot{\psi}\| \leq r$ almost everywhere, i.e. $\psi \in \text{AC}_{x_0, r}$. The claim now follows from Lemma 8.9 together with Theorem 8.1. $\square$

Next we identify the contracted rate function. The key observation is that Φ is a *bijection* onto the set of paths with admissible drifted velocity, with an explicit inverse.

Lemma 8.11 (The inverse of the drift map). *Let*

$$\mathcal{D} := \{\phi \in \text{AC}([0, 1]; M) : \phi(0) = x_0, \|\dot{\phi}(t) - V(\phi(t))\| \leq r \text{ a.e.}\}.$$

Then Φ maps $\text{AC}_{x_0, r}$ bijectively onto $\mathcal{D}$, and for $\phi \in \mathcal{D}$ the preimage $\psi = \Phi^{-1}(\phi)$ is the unique absolutely continuous solution of

$$\psi(0) = x_0, \quad \dot{\psi}(t) = P_{\phi(t) \rightarrow \psi(t)}(\dot{\phi}(t) - V(\phi(t))). \quad (16)$$

Proof. That $\Phi(\text{AC}_{x_0,r}) \subseteq \mathcal{D}$ is immediate from (9): $\dot{\phi} - V(\phi) = P_{\psi \rightarrow \phi} \dot{\psi}$ has norm at most r , parallel transport being an isometry.

Fix $\phi \in \mathcal{D}$ and write $w(t) := \dot{\phi}(t) - V(\phi(t)) \in T_{\phi(t)}M$, a measurable vector field along ϕ with $\|w\| \leq r$ almost everywhere; note $\|\dot{\phi}\| \leq r + \|V\|_\infty$, so $\phi([0,1]) \subseteq K_0$. The ODE (16) has the same structure as (9) (bounded measurable velocity read off along a moving source curve and parallel-transported to the solution point) and the well-posedness argument of Subsection 8.2.1 applies verbatim, with the following bookkeeping. *A priori bounds:* any solution has $\|\dot{\psi}\| \leq r$, hence stays in $\overline{B}(x_0, r)$; and the first-variation computation in the proof of Lemma 8.2, run along the geodesic variation from the base curve ϕ to ψ , gives, with $v(t) := \log_{\phi(t)} \psi(t)$ and $\delta(t) := d(\phi(t), \psi(t))$,

$$\delta(t) \delta'(t) = \langle \dot{\psi}, P_{\phi \rightarrow \psi} v \rangle - \langle \dot{\phi}, v \rangle = \langle w - \dot{\phi}, v \rangle = -\langle V(\phi(t)), v(t) \rangle \leq \|V\|_\infty \delta(t),$$

so again $\delta(t) \leq \|V\|_\infty t < \iota_0$.

Chart reduction and patching: since the source ϕ has speed at most $r + \|V\|_\infty$ and the solution ψ has speed at most r , the chart radius $\rho = \frac{1}{4}(\iota_0 - \|V\|_\infty)$ and step size $\tau = \rho/(r + \|V\|_\infty + 1)$ of Subsection 8.2.1 again confine each substep to a single normal chart whose centre stays within $\iota_0 - \rho$ of the source, and the chart right-hand side satisfies (C1)–(C4) exactly as in Lemma 8.4. Theorem 3.18 and the patching argument of Theorem 8.7 then produce a unique absolutely continuous global solution $\psi \in \text{AC}_{x_0,r}$.

Finally, the two ODEs invert each other. If ψ solves (16), then applying the isometry $P_{\psi(t) \rightarrow \phi(t)} = (P_{\phi(t) \rightarrow \psi(t)})^{-1}$ shows that ϕ satisfies (9) with input ψ , so $\Phi(\psi) = \phi$ by Theorem 8.7. Conversely, if $\Phi(\psi) = \phi$, the same manipulation shows that ψ solves (16); by the uniqueness just established, ψ is its only solution. Hence $\Phi : \text{AC}_{x_0,r} \rightarrow \mathcal{D}$ is onto with singleton preimages. $\square$

Lemma 8.12 (Transport invariance of the rate integrand). *Let $x, y \in M$ be joined by a smooth curve γ . Then for every $w \in T_x M$,*

$$\Lambda_y^*(\tau_{xy}^\gamma w) = \Lambda_x^*(w).$$

Proof. By the consistency property, $\Lambda_y(\tau_{xy}^\gamma \mu) = \Lambda_x(\mu)$ for all $\mu \in T_x M$. Substituting $\lambda = \tau_{xy}^\gamma \mu$ in the definition of the Legendre transform and using that τ_{xy}^γ is a surjective isometry,

$$\Lambda_y^*(\tau_{xy}^\gamma w) = \sup_{\mu \in T_x M} [\langle \tau_{xy}^\gamma \mu, \tau_{xy}^\gamma w \rangle - \Lambda_y(\tau_{xy}^\gamma \mu)] = \sup_{\mu \in T_x M} [\langle \mu, w \rangle - \Lambda_x(\mu)] = \Lambda_x^*(w). \quad \square$$

Proof of Theorem 6.2(i) (path-space LDP for the drifted flow). By Corollary 8.10, Z^n satisfies the LDP on the compact Polish space $(\text{AC}_{x_0,r}, d_\infty)$ with good rate function J_M ; by Proposition 6.1, Φ is continuous from this space to $C([0,1]; M)$. The contraction principle (Theorem 3.5) yields the LDP for $\Phi(Z^n)$ with good rate function

$$I_{\text{path}}^V(\phi) = \inf \{ J_M(\psi) : \psi \in \text{AC}_{x_0,r}, \Phi(\psi) = \phi \}.$$

It remains to identify this infimum with (3). If $\phi \notin \mathcal{D}$ (notation of Lemma 8.11), the preimage is empty and the infimum is $+\infty$; and the right-hand side of (3) is $+\infty$ as well, since either ϕ fails to be absolutely continuous with $\phi(0) = x_0$, or $\|\dot{\phi} - V(\phi)\| > r$ on a set of positive measure, where the integrand is infinite (as shown in the proof of Corollary 8.10). If $\phi \in \mathcal{D}$, the preimage is the singleton $\{\psi\}$ with $\psi = \Phi^{-1}(\phi)$ solving (16), and Lemma 8.12, applied along the minimising geodesic from $\phi(t)$ to $\psi(t)$ with $w = \dot{\phi}(t) - V(\phi(t))$, gives

$$\Lambda_{\psi(t)}^*(\dot{\psi}(t)) = \Lambda_{\psi(t)}^*(P_{\phi(t) \rightarrow \psi(t)}(\dot{\phi}(t) - V(\phi(t)))) = \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) \quad \text{a.e.},$$

so $J_M(\psi) = \int_0^1 \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) dt$, as claimed. $\square$

8.4 Exponential equivalence

In this subsection we show that the drifted random walk

$$\hat{Y}^n(t) := \text{Exp}_{S_k^n}((t - t_k)(V(S_k^n) + X_{k+1}^n)), \quad t \in [t_k, t_{k+1}],$$

(the piecewise-geodesic interpolation of Definition 5.3) and the deterministic image $\Phi(Z^n)$ of the undrifted walk under the drift map are exponentially equivalent at speed n .

The argument follows the same one-step-consistency-plus-discrete-Grönwall pattern as the geodesic Euler scheme of Section 4. The novelty is that both processes are now noise-driven: $\hat{Y}^n$ freezes the drift at the start of each substep but combines noise and drift inside a single exponential map, while $\Phi(Z^n)$ integrates a continuously-varying drift on top of the noise path. The one-step error is the gap between these two ways of advancing by $h = 1/n$ on a single substep.

8.4.1 Coupling

To compare $\hat{Y}^n$ and $\Phi(Z^n)$ pathwise we need them to be driven by a common noise. We construct the coupling step-by-step. Let $\{Z_k^n\}$ be the undrifted walk with increments $X_{k+1}^n \in T_{Z_k^n}M$ distributed according to $\mu_{Z_k^n}$. Set $S_0^n = Z_0^n = x_0$ and recursively define

$$\tilde{X}_{k+1}^n := P_{Z_k^n \rightarrow S_k^n}(X_{k+1}^n), \quad S_{k+1}^n := \text{Exp}_{S_k^n}(\frac{1}{n}(V(S_k^n) + \tilde{X}_{k+1}^n)), \quad (17)$$

where $P_{Z_k^n \rightarrow S_k^n}$ denotes parallel transport along the unique minimising geodesic from Z_k^n to S_k^n (well-defined as long as $d(Z_k^n, S_k^n) < \iota_0$; this is verified in the proof of Proposition 8.14 below). By the consistency property of $\{\mu_x\}$, $\tilde{X}_{k+1}^n \sim \mu_{S_k^n}$, so the marginal law of $\{S_k^n\}$ coincides with that of Definition 5.3.

8.4.2 The noisy one-step consistency lemma

All processes in this subsection remain in the compact tube $K_0 = \overline{B}(x_0, r + \|V\|_\infty)$ almost surely: the interpolated undrifted walk Z^n has speed at most r , so $Z^n([0, 1]) \subseteq \overline{B}(x_0, r)$; the drifted walk moves a distance of at most $(r + \|V\|_\infty)/n$ per step, so $S_k^n \in K_0$ for all $k \leq n$, and the interpolant $\hat{Y}^n$ stays in K_0 as well; finally $\Phi(Z^n)([0, 1]) \subseteq K_0$ by Lemma 8.2. Recall $\iota_0 = \text{inj}(K_0) > 0$, the standing assumption (11), and the chart radius $\rho = \frac{1}{4}(\iota_0 - \|V\|_\infty)$ from Subsection 8.2.1.

The constants below depend only on K_0 , V , and r . Lemma 2.33 and Lemma 4.5 are applied with $K = K_0$ and $\rho = \iota_0/2$, which is legitimate because the paragraph above puts the walk, its interpolant and the drift flow inside K_0 .

Lemma 8.13 (Noisy one-step consistency). *There exist $h_0 > 0$ and $C_{\text{loc}} > 0$ such that the following holds for all $0 < h \leq h_0$. Let $p, q \in K_0$ with $d(p, q) \leq \|V\|_\infty + \rho$, let $X \in T_qM$ with $\|X\| \leq r$, and let $\zeta : [0, h] \rightarrow M$ be the geodesic $\zeta(s) = \text{Exp}_q(sX)$. Let $\tilde{\phi} : [0, h] \rightarrow M$ be the unique solution of*

$$\dot{\tilde{\phi}}(s) = V(\tilde{\phi}(s)) + P_{\zeta(s) \rightarrow \tilde{\phi}(s)} \dot{\zeta}(s), \quad \tilde{\phi}(0) = p, \quad (18)$$

and write $\tilde{X} := P_{q \rightarrow p}(X)$, $p^+ := \text{Exp}_p(h(V(p) + \tilde{X}))$. Then

$$d(\tilde{\phi}(h), p^+) \leq C_{\text{loc}} h^2.$$

Proof sketch. This is the noise-driven analogue of Lemma 4.5; the new feature is that the source geodesic ζ starts at $q \neq p$, so in the chart centred at p the source is *not* at the origin.

Take h_0 small enough that $(r + \|V\|_\infty) h_0 < \rho$ (which also forces $rh_0 < \rho$). The source moves at speed $\|X\| \leq r$, and since parallel transport is an isometry and ζ has constant speed, $\|\dot{\phi}\| \leq \|V(\tilde{\phi})\| + \|\dot{\zeta}\| \leq \|V\|_\infty + r$. Hence for $s \in [0, h]$, $h \leq h_0$,

$$d(\zeta(s), p) \leq d(q, p) + rs \leq \|V\|_\infty + \rho + rh < \|V\|_\infty + 2\rho, \quad d(\tilde{\phi}(s), p) \leq (\|V\|_\infty + r)s < \rho,$$

both below $\iota_0 = \|V\|_\infty + 4\rho \leq \iota(p)$, so ζ and $\tilde{\phi}$ stay in the normal neighbourhood of p and the chart $\log_p$ is available. The parallel transport in (18) is well-defined too, since

$$d(\zeta(s), \tilde{\phi}(s)) \leq d(\zeta(s), q) + d(q, p) + d(p, \tilde{\phi}(s)) < \rho + (\|V\|_\infty + \rho) + \rho = \|V\|_\infty + 3\rho < \iota_0 :$$

$\zeta(s)$ and $\tilde{\phi}(s)$ are joined by a unique minimising geodesic, and $P_{\zeta(s) \rightarrow \tilde{\phi}(s)}$ depends smoothly on its endpoints (Lemma 2.29).

Work in normal coordinates centred at p and set $y := \log_p \circ \tilde{\phi}$, $\tilde{\zeta} := \log_p \circ \zeta$, both valued in $T_p M \cong \mathbb{R}^d$ and smooth on $[0, h]$ by the previous paragraph. In these coordinates (18) becomes

$$\dot{y}(s) = F_p(y(s), \tilde{\zeta}(s), \tilde{\zeta}(s)), \quad y(0) = \log_p(p) = 0,$$

where F_p encodes the drift and the parallel transport and is smooth in all three slots on the regime. Note $\tilde{\zeta}(0) = \log_p(q) \neq 0$: the source enters away from the chart centre, so F_p is *not* evaluated at the origin in its last two slots.

For the linear term only the *output* matters, and it does start at the centre: $\tilde{\phi}(0) = p$ gives $(d\log_p)_p = \text{id}$, whence

$$\dot{y}(0) = (d\log_p)_{\tilde{\phi}(0)} [V(\tilde{\phi}(0)) + P_{\zeta(0) \rightarrow \tilde{\phi}(0)} \dot{\zeta}(0)] = V(p) + P_{q \rightarrow p}(X) = V(p) + \tilde{X},$$

which is exactly the velocity of the Euler step $p^+ = \text{Exp}_p(h(V(p) + \tilde{X}))$. The data range over

$$\mathcal{T} := \{(p, q, X) : p, q \in K_0, d(p, q) \leq \|V\|_\infty + \rho, X \in T_q M, \|X\| \leq r\},$$

a compact set (closed, fibrewise bounded over the compact base $K_0 \times K_0$), and F_p together with $\tilde{\zeta}, \dot{\tilde{\zeta}}$ depend smoothly on these data with bounds uniform over $\mathcal{T}$. Hence $s \mapsto \dot{y}(s)$ is C^1 with $\|\dot{y}\|$ bounded by a constant C uniform over $\mathcal{T}$, and Taylor's theorem at $s = 0$ gives

$$\|y(h) - h(V(p) + \tilde{X})\| = \|y(h) - y(0) - h\dot{y}(0)\| \leq \frac{C}{2} h^2.$$

Finally, both $y(h)$ and $h(V(p) + \tilde{X})$ have norm $\leq (\|V\|_\infty + r)h < \rho < \iota_0/2$, within the range of the uniform Lipschitz bound for Exp_p (Lemma 2.33), so

$$d(\tilde{\phi}(h), p^+) = d(\text{Exp}_p(y(h)), \text{Exp}_p(h(V(p) + \tilde{X}))) \leq C_{\text{exp}}(K_0, \iota_0/2) \|y(h) - h(V(p) + \tilde{X})\| \leq C_{\text{exp}}(K_0, \iota_0/2) \frac{C}{2} h^2 =: \epsilon. \quad \square$$

8.4.3 Global error and uniform convergence on a compact event

Proposition 8.14 (Deterministic Euler bound). *There exist $n_0 \in \mathbb{N}$ and $C_{\text{glob}} > 0$ such that, almost surely,*

$$d_\infty(\hat{Y}^n, \Phi(Z^n)) \leq \frac{C_{\text{glob}}}{n} \quad \text{for all } n \geq n_0.$$

Proof. Set $e_k := d(S_k^n, \Phi(Z^n)(t_k))$ and $h = 1/n$. Let C_{loc}, h_0 be the constants of Lemma 8.13 and let L_ζ be a Lipschitz constant for the dependence of (18) on its initial condition over K_0 , determined by $\sup_{K_0} \|\nabla V\|_{\text{op}}$ and the curvature bound on K_0 ; Remark 8.15 below says where it comes from, and fix n_0 so large that $h \leq h_0$ and $(C_{\text{loc}}/L_\zeta)(e^{L_\zeta} - 1)h \leq \rho$ for all $n \geq n_0$. We prove by induction on k that the coupling (17) is well-defined at step k and that $e_k \leq (C_{\text{loc}}/L_\zeta)(e^{L_\zeta} - 1)h$; the supremum bound on $[0, 1]$ then follows from the same triangle-inequality argument as in the proof of Theorem 4.8.

For $k = 0$ there is nothing to prove: $e_0 = 0$ and $P_{Z_0^n \rightarrow S_0^n}$ is the identity. Suppose the claim holds up to step k . By Lemma 8.2, $d(Z^n(t_k), \Phi(Z^n)(t_k)) \leq \|V\|_\infty t_k$, so

$$d(Z_k^n, S_k^n) \leq d(Z^n(t_k), \Phi(Z^n)(t_k)) + e_k \leq \|V\|_\infty + \rho.$$

In particular $d(Z_k^n, S_k^n) < \iota_0$, so the parallel transport in (17), and with it S_{k+1}^n , is well-defined, and the hypothesis of Lemma 8.13 holds for the pair $q = Z_k^n, p = S_k^n$.

Apply Lemma 8.13 on the substep $[t_k, t_{k+1}]$ with $q = Z_k^n, p = S_k^n, X = X_{k+1}^n$. Note that $\zeta = Z^n$ on this substep by construction, so the solution $\tilde{\phi}$ of (18) starting at $p = S_k^n$ is exactly the restriction of $\Phi(Z^n)$ to $[t_k, t_{k+1}]$ shifted to start at S_k^n instead of $\Phi(Z^n)(t_k)$. The lemma gives

$$d(\tilde{\phi}(h), S_{k+1}^n) \leq C_{\text{loc}} h^2.$$

Let ϕ_k^* denote the actual restriction $\Phi(Z^n)|_{[t_k, t_{k+1}]}$, which starts at $\Phi(Z^n)(t_k)$ rather than S_k^n . By Lipschitz dependence of (18) on its initial condition (Remark 8.15, which is where this estimate is justified),

$$d(\tilde{\phi}(h), \phi_k^*(t_{k+1})) \leq e^{L_\zeta h} e_k.$$

Combining via the triangle inequality,

$$e_{k+1} = d(S_{k+1}^n, \phi_k^*(t_{k+1})) \leq C_{\text{loc}} h^2 + e^{L_\zeta h} e_k.$$

This is exactly the recursion of Lemma 4.7, so the same argument gives $e_{k+1} \leq C_{\text{loc}} h^2 \sum_{j=0}^k e^{L_\zeta j h} \leq (C_{\text{loc}}/L_\zeta)(e^{L_\zeta} - 1)h$, closing the induction. The piecewise-geodesic interpolation step is identical to that in the proof of Theorem 4.8: within each substep both $\hat{Y}^n$ and $\Phi(Z^n)$ have speed at most $r + \|V\|_\infty$, which yields the claim with $C_{\text{glob}} := (C_{\text{loc}}/L_\zeta)(e^{L_\zeta} - 1) + 2(r + \|V\|_\infty)$. $\square$

Remark 8.15 (Lipschitz dependence on initial condition). The Lipschitz estimate $d(\tilde{\phi}(h), \phi_k^*(t_{k+1})) \leq e^{L_\zeta h} e_k$ used above is the noise-driven analogue of Lemma 4.6, and this remark is what justifies it. The constant is not the L of Lemma 4.6: that one is $\sup_K \|\nabla V\|_{\text{op}}$, while transporting between two solutions brings in curvature as well. The proof runs identically: build a one-parameter family of solutions of (18) with varying initial conditions, differentiate in the parameter to get the variational equation, and apply Grönwall to the squared norm of the variation field. The driving signal $\dot{\zeta}$ is the same for both initial conditions and so cancels in the variational equation, leaving only the curvature-and- ∇V terms.

8.4.4 Exponential equivalence of the walk and the flow

Proposition 8.16 (Exponential equivalence). *For every $\delta > 0$,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(d_\infty(\hat{Y}^n, \Phi(Z^n)) > \delta) = -\infty.$$

Proof. By the confinement established at the beginning of this subsection, Z^n , $\hat{Y}^n$ and $\Phi(Z^n)$ remain in the compact tube K_0 almost surely, so Proposition 8.14 applies: $d_\infty(\hat{Y}^n, \Phi(Z^n)) \leq C_{\text{glob}}/n$ almost surely for all $n \geq n_0$. For $n > \max(n_0, C_{\text{glob}}/\delta)$ the event $\{d_\infty(\hat{Y}^n, \Phi(Z^n)) > \delta\}$ is therefore a null event, and the limsup equals $-\infty$. $\square$

Remark 8.17 (Where exponential tightness enters). In contrast with the flat case (Proposition 7.2), no large-deviation estimate is needed at this step *under the standing bounded-increment assumption*: the support bound $\text{supp}(\mu_x) \subseteq \overline{B}(0, r)$ confines all processes to the deterministic compact tube K_0 , so the exponential compact containment of [6, Proposition 7.15] is not required here. This is an artifact of bounded increments rather than a feature of the argument: once the support assumption is removed (Section 9), no deterministic tube exists and the logic reverses: exponential tightness, established from Cramér's condition alone in Proposition 9.7, is precisely what replaces K_0 .

Proof of Theorem 6.2(ii) (path-space LDP for the drifted walk). By Theorem 6.2(i) the processes $\Phi(Z^n)$ satisfy this principle; by Proposition 8.16 the interpolated walk $\hat{Y}^n$ is exponentially equivalent to $\Phi(Z^n)$, so the principle of exponential equivalence (Theorem 3.7) transfers it to $\hat{Y}^n$. $\square$

8.5 Contraction to the endpoint

The evaluation map $\text{ev}_1 : C([0, 1]; M) \rightarrow M$, $\phi \mapsto \phi(1)$, is continuous. Applying the contraction principle to the path-space LDP of $\hat{Y}^n$ (Theorem 6.2(ii)), the drifted empirical averages $S_n^n = \hat{Y}^n(1)$ satisfy the LDP in M with rate function

$$I_M^V(x) = \inf_{\substack{\phi \in \text{AC}([0, 1]; M) \\ \phi(0) = x_0, \phi(1) = x}} \int_0^1 \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) \, dt.$$

This completes the proof of Theorem 6.3. $\square$

9 Removing the bounded-increments assumption

The goal of the chapter is to prove Theorem 6.5, which we restate for reference. Assume bounded geometry, that is $\text{inj}(M) > 0$ and $\|\text{Rm}\|_\infty < \infty$. Assume Cramér's condition, $\Lambda_{x_0}(\lambda) < \infty$ for all $\lambda \in T_{x_0}M$. Let $V \in \Gamma(TM)$ be smooth and bounded with $\|V\|_\infty < \text{inj}(M)$. Then the drifted empirical averages S_n^n satisfy the large deviation principle in M at speed n , with the same rate function I_M^V as in Theorem 6.3. The boundedness of the increments is dropped. *Cramér's condition* takes the place of the bound on the increments, and it is the classical hypothesis of Cramér's and Mogulskii's theorems in $\mathbb{R}^d$ (Theorem 3.3, Theorem 3.4). The conclusion is the same drifted large deviation principle as Theorem 6.3, with the same rate function; what changes is the class of increment laws to which it applies, now including every distribution with finite exponential moments.

A new strategy is needed, because without a speed bound the paths are not guaranteed to stay in a compact set. We do not replace the continuity argument for Φ on the full path space. Instead we keep Section 8 intact, apply it unchanged at each fixed scale, and handle the unbounded increments in an outer approximation layer.

Concretely, the strategy is a *truncation tower* with three layers. For each truncation level $c > 0$ we modify the increment measures, by keeping the mass inside $\overline{B}(0, c)$ and collapsing the rest to the origin. This way, Theorem 6.3 applies, yielding a path-space LDP at every fixed level with rate function I_c (Layer 0, Subsection 9.4). Here $\text{inj}(M) > 0$ enters: it makes the injectivity-radius hypothesis of Theorem 6.3 hold at every level at the same time, although the level- c tube grows without bound. We then couple the original walk to the c -truncated walk and show, using only Cramér's condition, that the latter are exponentially good approximations of the former as $c \rightarrow \infty$ (Layer 1, Subsection 9.5). The comparison of the two walks uses the curvature bound $\|\text{Rm}\|_\infty < \infty$, through a one-step stability estimate, and uses $\text{inj}(M) > 0$ a second time, to keep the coupling on minimising geodesics. Finally we identify the limiting rate function. We show that I_c converges to I_{path}^V in the sense the approximation theorem of Dembo and Zeitouni requires, a Γ -convergence of integral functionals. The two one-sided bounds are proved in Subsection 9.6. Assembling the three layers proves the path-space principle Theorem 6.4, of which the endpoint theorem is the contraction. We first deal with two preparatory statements: the structural facts about truncation and consistency (Subsection 9.1), and the path-space exponential tightness that replaces the deterministic tube of the bounded theory (Subsection 9.3).

9.1 Truncation preliminaries

The tower relies on two structural facts. Consistency makes every radial quantity base-point free; and the truncated families inherit consistency, and their log-moment generating functions converge locally uniformly. Throughout, $\{\mu_x\}_{x \in M}$ is a consistent family of probability measures on the tangent spaces, and X denotes a generic increment, i.e. a random vector with law μ_x in some tangent space $T_x M$; expectations of functions of X are taken with respect to μ_x .

Lemma 9.1 (The radial law is base-point free). *Under consistency, the pushforward of μ_x under $v \mapsto \|v\| \in [0, \infty)$ is the same for every $x \in M$. In particular, any moment or tail condition on $\|X\|$ verified at one point holds at every point, with identical constants.*

Proof. Let $x, y \in M$. Recall we work with M connected, so that there is a smooth curve γ from x to y , and consistency gives $\mu_y = \mu_x \circ (\tau_{xy}^\gamma)^{-1}$. Parallel transport is a linear isometry, so $\{v \in T_x M :$

$\|\tau_{xy}^\gamma v\| \in B\} = \{v \in T_x M : \|v\| \in B\}$ for every Borel $B \subseteq [0, \infty)$, and hence

$$\mu_y(\|w\| \in B) = \mu_x(\|\tau_{xy}^\gamma v\| \in B) = \mu_x(\|v\| \in B). \quad \square$$

We write $\nu \in \mathcal{P}([0, \infty))$ for the common law of $\|X\|$ and

$$\Lambda_{\text{abs}}(\lambda) := \log \mathbb{E}[e^{\lambda\|X\|}] = \log \int_0^\infty e^{\lambda s} \nu(ds) \in (-\infty, +\infty], \quad \lambda \geq 0,$$

which by Lemma 9.1 does not depend on the base point; it is the manifold version of the quantity of the same name in the proof of Proposition 7.2. The 2^d -corner bound used there also settles the relation between Cramér's condition and Λ_{abs} :

Lemma 9.2 (Cramér's condition and absolute exponential moments). *The following are equivalent:*

- (i) $\Lambda_{x_0}(\lambda) < \infty$ for all $\lambda \in T_{x_0} M$ (Cramér's condition);
- (ii) $\Lambda_x(\lambda) < \infty$ for all $x \in M$ and all $\lambda \in T_x M$;
- (iii) $\Lambda_{\text{abs}}(\lambda) < \infty$ for all $\lambda \geq 0$.

Proof. (i) $\Rightarrow$ (iii): fix an orthonormal basis of $T_{x_0} M$, identifying it with $\mathbb{R}^d$. Exactly as in the proof of Proposition 7.2, $\|v\| \leq \sum_{i=1}^d |v_i|$ gives $e^{\lambda\|v\|} \leq \prod_{i=1}^d e^{\lambda|v_i|} \leq \sum_{\sigma \in \{-1,1\}^d} e^{\lambda\langle \sigma, v \rangle}$, so $\mathbb{E}[e^{\lambda\|X\|}] \leq \sum_{\sigma \in \{-1,1\}^d} e^{\Lambda_{x_0}(\lambda\sigma)} < \infty$. (iii) $\Rightarrow$ (ii): by Cauchy-Schwarz, $\langle \lambda, v \rangle \leq \|\lambda\|\|v\|$, so $\Lambda_x(\lambda) \leq \Lambda_{\text{abs}}(\|\lambda\|) < \infty$, where Lemma 9.1 lets us evaluate Λ_{abs} at the base point x . (ii) $\Rightarrow$ (i) is trivial. $\square$

We now introduce the truncated increment measures. The truncation is defined radially, so that it commutes with parallel transport; the mass removed from outside the ball is moved to the origin, keeping the total mass equal to one.

Concretely, fix $c > 0$ and define, for each $x \in M$, the *level- c truncated increment measure* $\mu_x^{(c)}$ on $T_x M$ by

$$\mu_x^{(c)}(A) := \mu_x(A \cap \{\|v\| \leq c\}) + \mu_x(\|v\| > c) \delta_0(A), \quad A \subseteq T_x M \text{ Borel},$$

so that the mass $\mu_x(\|v\| > c)$ lying outside the ball of radius c is collapsed onto the origin.

Lemma 9.3 (Truncation preserves consistency). *Fix $c > 0$. The level- c truncated family $\{\mu_x^{(c)}\}_{x \in M}$ just defined satisfies:*

- (i) $\{\mu_x^{(c)}\}_{x \in M}$ is a consistent family of probability measures with $\text{supp}(\mu_x^{(c)}) \subseteq \overline{B}(0, c)$;
- (ii) its log-moment generating function is

$$\Lambda_x^{(c)}(\lambda) = \log \left(\mathbb{E}[e^{\langle \lambda, X \rangle} \mathbf{1}_{\{\|X\| \leq c\}}] + \mathbb{P}(\|X\| > c) \right), \quad \lambda \in T_x M;$$

- (iii) if Cramér's condition holds, then $\Lambda_{x_0}^{(c)} \rightarrow \Lambda_{x_0}$ pointwise on $T_{x_0} M$ as $c \rightarrow \infty$, and the convergence is uniform on compact subsets of $T_{x_0} M$.

Proof. (i) Total mass: $\mu_x^{(c)}(T_x M) = \mu_x(\|v\| \leq c) + \mu_x(\|v\| > c) = 1$, and both summands of $\mu_x^{(c)}$ are supported in $\overline{B}(0, c)$. For consistency, let γ be a smooth curve from x to y and write $\tau := \tau_{xy}^\gamma$. For Borel $A \subseteq T_y M$,

$$\mu_x^{(c)}(\tau^{-1}A) = \mu_x(\tau^{-1}A \cap \{\|v\| \leq c\}) + \mu_x(\|v\| > c) \mathbf{1}_{\{0 \in \tau^{-1}A\}}.$$

Since τ is a linear isometry, $\tau^{-1}A \cap \{\|v\| \leq c\} = \tau^{-1}(A \cap \{\|w\| \leq c\})$ and $0 \in \tau^{-1}A \Leftrightarrow 0 \in A$; consistency of the original family gives $\mu_x(\tau^{-1}(A \cap \{\|w\| \leq c\})) = \mu_y(A \cap \{\|w\| \leq c\})$, and Lemma 9.1 gives $\mu_x(\|v\| > c) = \mu_y(\|w\| > c)$. Hence $\mu_x^{(c)} \circ \tau^{-1} = \mu_y^{(c)}$.

(ii) Immediate: $\int e^{\langle \lambda, u \rangle} \mu_x^{(c)}(du) = \mathbb{E}[e^{\langle \lambda, X \rangle} \mathbf{1}_{\{\|X\| \leq c\}}] + \mathbb{P}(\|X\| > c) e^{\langle \lambda, 0 \rangle}$.

(iii) As $c \uparrow \infty$, the integrands $e^{\langle \lambda, X \rangle} \mathbf{1}_{\{\|X\| \leq c\}}$ increase pointwise to $e^{\langle \lambda, X \rangle}$, so by monotone convergence the first term inside the logarithm increases to $e^{\Lambda_{x_0}(\lambda)} < \infty$, while $\mathbb{P}(\|X\| > c) \downarrow 0$; continuity of log gives the pointwise convergence. Each $\Lambda_{x_0}^{(c)}$ is finite on all of $T_{x_0}M$ (the argument of the logarithm lies between $\max(e^{-c\|\lambda\|} \mathbb{P}(\|X\| \leq c), \mathbb{P}(\|X\| > c)) > 0$ and $e^{c\|\lambda\|}$) and convex, being a log-moment generating function (Hölder's inequality); Λ_{x_0} is finite by Cramér's condition. A sequence of finite convex functions on $\mathbb{R}^d$ converging pointwise to a finite function converges uniformly on compact sets [12, Theorem 10.8], which gives the local uniformity. $\square$

9.2 Convex-analytic preliminaries

The rate-function identification of Subsection 9.6 rests on a group of convex-analytic facts about the truncated log-moment generating functions and their conjugates. We collect them here so that the argument itself reads uninterrupted.

The notation is that of Subsection 3.3: aff , ri , $\overline{\text{conv}}$, dom , and the support function. The relative interior is the one to watch here, because $\mathcal{S}$ below may be a proper lower-dimensional subset of $T_{x_0}M$, in which case its ordinary interior is empty while its relative interior is not.

Fix the base point x_0 and abbreviate $\Lambda := \Lambda_{x_0}$, $\Lambda^{(c)} := \Lambda_{x_0}^{(c)}$, both finite convex functions on $T_{x_0}M \cong \mathbb{R}^d$ (Cramér's condition and Lemma 9.3). Write $X \sim \mu_{x_0}$ and introduce

$$\mathcal{S} := \overline{\text{conv}}(\text{supp } \mu_{x_0}), \quad \mathcal{A} := \text{aff } \mathcal{S} = a + E, \quad m := \mathbb{E}[X],$$

the closed convex support, its affine hull, and the mean, finite by Lemma 9.2. Here $a \in \mathcal{S}$ is any base point of the affine hull and $E \subseteq T_{x_0}M$ is the linear subspace parallel to it, so that $\mathcal{A} = a + E$; equivalently $E = \text{span}(\mathcal{S} - \mathcal{S})$. Only directions in E ever matter: a λ orthogonal to E changes $\langle \lambda, v \rangle$ by the same constant for every $v \in \mathcal{S}$, so every supremum over λ below may be restricted to E , which is Lemma 9.4(iii). Since parallel transport is a linear isometry and the family is consistent, the corresponding objects at any other base point are the isometric images of these; every statement below is therefore transport-covariant, and we work at x_0 without loss of generality. Alongside $\Lambda^{(c)}$ we use the *missing-mass* log-moment generating function

$$\Lambda_c^-(\lambda) := \log \mathbb{E}[e^{\langle \lambda, X \rangle} \mathbf{1}_{\{\|X\| \leq c\}}],$$

finite and convex for every $c \geq c_0$, where c_0 is any radius with $\mathbb{P}(\|X\| \leq c_0) > 0$; by monotone convergence and [12, Theorem 10.8] (as in Lemma 9.3), $\Lambda_c^- \uparrow \Lambda$ pointwise and locally uniformly. Its significance: $\Lambda^{(c)} \geq \Lambda_c^-$ pointwise (dropping the atom's contribution), so that

$$\Lambda^{(c)*} \leq (\Lambda_c^-)^*, \tag{19}$$

and, unlike $\Lambda^{(c)}$ itself, the family Λ_c^- is *monotone* in c .

Lemma 9.4 (Support, domain, and the mean).

(i) $\Lambda^*(m) = 0$ and $m \in \text{ri } \mathcal{S}$.

(ii) $\text{dom } \Lambda^* \subseteq \mathcal{S}$.

(iii) For $y \in \mathcal{S}$ and $c \in [c_0, \infty]$ (reading $\Lambda_\infty^- := \Lambda$), the map $\lambda \mapsto \langle \lambda, y \rangle - \Lambda_c^-(\lambda)$ depends on λ only through its orthogonal projection onto E ; in particular $(\Lambda_c^-)^*(y) = \sup_{\lambda \in E} [\langle \lambda, y \rangle - \Lambda_c^-(\lambda)]$.

Proof. (i) Jensen's inequality gives $\Lambda(\lambda) \geq \langle \lambda, m \rangle$, so $\Lambda^*(m) \leq 0$; and $\Lambda(0) = 0$ gives $\Lambda^* \geq 0$. If $m \notin \text{ri } \mathcal{S}$ (Subsection 3.3), then by proper separation [12, Theorem 11.6] there is λ , not orthogonal to E , with $\langle \lambda, v \rangle \leq \langle \lambda, m \rangle$ for all $v \in \mathcal{S}$; then $\langle \lambda, X - m \rangle \leq 0$ almost surely with mean zero, so $\langle \lambda, X \rangle = \langle \lambda, m \rangle$ almost surely, placing $\text{supp } \mu_{x_0}$, hence $\mathcal{A}$, inside a hyperplane to which λ is normal. This contradicts that λ is not orthogonal to E .

(ii) If $y \notin \mathcal{S}$, separate: there are λ and α with $\langle \lambda, y \rangle > \alpha \geq \langle \lambda, u \rangle$ for all $u \in \text{supp } \mu_{x_0}$; then $\Lambda(t\lambda) \leq t\alpha$ for $t > 0$ and $\Lambda^*(y) \geq t(\langle \lambda, y \rangle - \alpha) \rightarrow \infty$.

(iii) $X \in \mathcal{A}$ almost surely, so for the decomposition $\lambda = \lambda_E + \lambda_\perp$ one has $\langle \lambda_\perp, X \rangle = \langle \lambda_\perp, a \rangle$ almost surely, whence $\Lambda_c^-(\lambda) = \langle \lambda_\perp, a \rangle + \Lambda_c^-(\lambda_E)$; and $\langle \lambda_\perp, y \rangle = \langle \lambda_\perp, a \rangle$ since $y - a \in E$. The two $\lambda_\perp$ -terms cancel. $\square$

Call a compact set $K \subseteq T_{x_0}M$ an s -margin compactum ($s > 0$) if $K + 2s(\overline{B}(0, 1) \cap E) \subseteq \mathcal{S}$; every compact subset of $\text{ri } \mathcal{S}$ is an s -margin compactum for some $s > 0$. The next lemma is the workhorse of the recovery bound.

Lemma 9.5 (Margin bound and convergence of the conjugates). *Let K be an s -margin compactum. There exist $c_1 < \infty$ and $m_0 \in (0, 1]$ such that*

$$\langle \lambda, y \rangle - \Lambda_c^-(\lambda) \leq -\frac{s}{2} \|\lambda\| - \log m_0 \quad \text{for all } y \in K, \lambda \in E, c \in [c_1, \infty]. \quad (20)$$

Consequently:

(i) $\Lambda^* \leq -\log m_0$ on K ; in particular $\text{ri } \mathcal{S} \subseteq \text{dom } \Lambda^*$;

(ii) $\lim_{c \rightarrow \infty} \sup_{y \in K} [(\Lambda_c^-)^*(y) - \Lambda^*(y)] = 0$, and hence by (19) also $\limsup_{c \rightarrow \infty} \sup_{y \in K} [\Lambda^{(c)*}(y) - \Lambda^*(y)] \leq 0$.

The two limits in (ii) are stated differently on purpose. Since $\Lambda_c^- \leq \Lambda$, the conjugates satisfy $(\Lambda_c^-)^* \geq \Lambda^*$, so the first bracket is nonnegative and its supremum decreases to 0. The atom at the origin makes $\Lambda^{(c)}$ non-monotone in c and not comparable with Λ , so the second limit has no matching lower bound. Only the one-sided form is used, in the proof of Proposition 9.16.

Proof. The master bound. Write $h_K(\hat{e}) := \sup_{y \in K} \langle \hat{e}, y \rangle$ for the support function of K (Subsection 3.3), continuous on the unit sphere of E . For each unit vector $\hat{e} \in E$ and $y \in K$ the margin gives $y + 2s\hat{e} \in \mathcal{S}$, so the support function of $\text{supp } \mu_{x_0}$ (which coincides with that of $\mathcal{S}$) satisfies $\sup_{u \in \text{supp } \mu} \langle \hat{e}, u \rangle \geq h_K(\hat{e}) + 2s$; pick $u_{\hat{e}} \in \text{supp } \mu_{x_0}$ with $\langle \hat{e}, u_{\hat{e}} \rangle \geq h_K(\hat{e}) + \frac{3}{2}s$. Both sides being continuous in the direction, each $\hat{e}$ has a neighbourhood on the unit sphere of E where $\langle \cdot, u_{\hat{e}} \rangle \geq h_K(\cdot) + s$; by compactness finitely many such neighbourhoods cover, yielding a finite set $F \subset \text{supp } \mu_{x_0}$ with

$$\max_{u \in F} \langle \hat{e}, u \rangle \geq h_K(\hat{e}) + s \quad \text{for every unit } \hat{e} \in E.$$

Set $c_1 := \max_{u \in F} \|u\| + s$ and $m_0 := \min_{u \in F} \mu_{x_0}(B(u, s/2)) > 0$ (positive since $F \subset \text{supp } \mu_{x_0}$; also $m_0 \leq 1$). For $\lambda \in E \setminus \{0\}$, $\hat{e} := \lambda/\|\lambda\|$, and $u_* \in F$ attaining the maximum above, the ball $B(u_*, s/2)$ is contained in $\{\|v\| \leq c_1\}$, so for every $c \in [c_1, \infty]$,

$$\Lambda_c^-(\lambda) \geq \log \mathbb{E}[e^{\langle \lambda, X \rangle} \mathbf{1}_{\{X \in B(u_*, s/2)\}}] \geq \langle \lambda, u_* \rangle - \frac{s}{2} \|\lambda\| + \log m_0 \geq \langle \lambda, y \rangle + \frac{s}{2} \|\lambda\| + \log m_0$$

for every $y \in K$, which is (20); the case $\lambda = 0$ reads $-\Lambda_c^-(0) \leq -\log m_0$, true since $\mathbb{P}(\|X\| \leq c_1) \geq m_0$.

(i) Take $c = \infty$ in (20) and use Lemma 9.4(iii): $\Lambda^*(y) \leq \sup_{\lambda \in E} [-\frac{s}{2} \|\lambda\| - \log m_0] = -\log m_0$.

(ii) Fix $\ell_0 := \frac{2}{s} (-\log m_0) \geq 0$. For $y \in K$ and $c \geq c_1$, split the supremum in Lemma 9.4(iii): for $\|\lambda\| > \ell_0$, (20) gives $\langle \lambda, y \rangle - \Lambda_c^-(\lambda) \leq -\frac{s}{2} \ell_0 - \log m_0 \leq 0 \leq \Lambda^*(y)$; for $\|\lambda\| \leq \ell_0$,

$$\langle \lambda, y \rangle - \Lambda_c^-(\lambda) \leq \langle \lambda, y \rangle - \Lambda(\lambda) + \eta_c \leq \Lambda^*(y) + \eta_c, \quad \eta_c := \sup_{\|\lambda\| \leq \ell_0} (\Lambda - \Lambda_c^-)(\lambda) \xrightarrow{c \rightarrow \infty} 0$$

by the locally uniform convergence $\Lambda_c^- \uparrow \Lambda$. Hence $(\Lambda_c^-)^*(y) \leq \Lambda^*(y) + \eta_c$ uniformly on K . $\square$

Lemma 9.6 (Liminf and uniform superlinearity).

(i) For every sequence $y_c \rightarrow y$ in $T_{x_0}M$, $\liminf_{c \rightarrow \infty} \Lambda^{(c)*}(y_c) \geq \Lambda^*(y)$.

(ii) For every $\ell \geq 0$ there is $A_\ell < \infty$, independent of c , with $\Lambda^{(c)*}(y) \geq \ell\|y\| - A_\ell$ and $\Lambda^*(y) \geq \ell\|y\| - A_\ell$ for all y ; one may take $A_\ell := \sup_{\|\lambda\| \leq \ell} \Lambda(\lambda)^+ + \log 2$.

Proof. (i) For each fixed λ , $\Lambda^{(c)*}(y_c) \geq \langle \lambda, y_c \rangle - \Lambda^{(c)}(\lambda) \rightarrow \langle \lambda, y \rangle - \Lambda(\lambda)$ by Lemma 9.3(iii); take the supremum over λ . (ii) From Lemma 9.3(ii), $e^{\Lambda^{(c)}(\lambda)} \leq e^{\Lambda(\lambda)} + 1 \leq 2e^{\Lambda(\lambda)^+}$, so $\sup_{\|\lambda\| \leq \ell} \Lambda^{(c)}(\lambda) \leq A_\ell - \log 2 + \log 2 = A_\ell$; hence $\Lambda^{(c)*}(y) \geq \sup_{\|\lambda\| = \ell} \langle \lambda, y \rangle - A_\ell = \ell\|y\| - A_\ell$. The same bound for Λ is immediate. $\square$

Part (ii) yields uniform integrability on sublevel sets. Take a family of measurable $u : [0, 1] \rightarrow T_{x_0}M$ with $\int_0^1 \Lambda^{(c)*}(u) dt \leq \alpha$, where $c = c(u)$ may vary with u . Then $\int \|u\| \leq \alpha + A_1$, and for every ℓ and $\varrho > 0$,

$$\int_{\{\|u\| > \varrho\}} \|u\| dt \leq \frac{1}{\ell} \int_{\{\|u\| > \varrho\}} (\Lambda^{(c)*}(u) + A_\ell) dt \leq \frac{\alpha}{\ell} + \frac{A_\ell(\alpha + A_1)}{\ell \varrho}, \quad (21)$$

so the family is uniformly integrable (ℓ large, then ϱ large).

9.3 Path-space exponential tightness

Under bounded increments, every process in the proof of Theorem 6.3 is confined to the deterministic compact tube K_0 , and the exponential-equivalence step needed no large-deviation estimate at all (Remark 8.17). With unbounded increments no deterministic confinement is available, and its role is taken over by *exponential tightness* (Definition 3.8), which now Cramér's condition supplies. The proposition below is the manifold analogue of the $\{T_n \leq R\}$ decomposition from the flat exponential-equivalence proof (Proposition 7.2), and we follow that proof's structure here. A Chernoff bound on T_n confines the paths to a compact ball, with the probability of leaving it superexponentially small; windowed Chernoff bounds then control the modulus of continuity; and Arzelà–Ascoli combines the two into the compactness we need.

Throughout this subsection, $\hat{Y}^n$ denotes the piecewise-geodesic interpolation of the drifted walk of Definition 5.3, with the boundedness assumption on the increments dropped, as in Subsection 8.4. Three pieces of notation recur. We write X_k^n for the increment applied on the substep $[t_{k-1}, t_k]$ and $T_n := \frac{1}{n} \sum_{k=1}^n \|X_k^n\|$ for the total-variation clock of the noise. For $\gamma \in C([0, 1]; M)$ and $\eta > 0$ we write

$$\text{osc}_\eta(\gamma) := \sup\{d(\gamma(s), \gamma(t)) : s, t \in [0, 1], |s - t| \leq \eta\}$$

for the modulus of continuity.

Proposition 9.7 (Path-space exponential tightness under Cramér's condition). *Let (M, g) be complete, let $\{\mu_x\}_{x \in M}$ be a consistent family satisfying Cramér's condition, and let $V \in \Gamma(TM)$ be smooth and bounded. Then $\{\hat{Y}^n\}_{n \geq 1}$ is exponentially tight in $(C_{x_0}([0, 1]; M), d_\infty)$.*

Proof. Step 0: length bound. On the substep $[t_{k-1}, t_k]$ the interpolant is a geodesic of constant speed $\|V(S_{k-1}^n) + X_k^n\| \leq \|V\|_\infty + \|X_k^n\|$. Bounding distance by length, for $0 \leq s \leq t \leq 1$,

$$d(\hat{Y}^n(s), \hat{Y}^n(t)) \leq (t - s) \|V\|_\infty + \frac{1}{n} \sum_{k \in \mathcal{K}_n(s, t)} \|X_k^n\|, \quad (22)$$

where $\mathcal{K}_n(s, t) := \{1 \leq k \leq n : [t_{k-1}, t_k] \cap [s, t] \neq \emptyset\}$ contains at most $n(t - s) + 2$ indices (each substep contributes its speed times the length of $[t_{k-1}, t_k] \cap [s, t] \leq 1/n$).

Step 1: Chernoff bound for windowed sums. Let $\mathcal{K} \subseteq \{1, \dots, n\}$ be any deterministic index set and $\lambda \geq 0$. Peeling off the largest index by conditioning on the walk up to that step and using that each $\|X_k^n\|$ has conditional law ν (Lemma 9.1), induction gives $\mathbb{E}[e^{\lambda \sum_{k \in \mathcal{K}} \|X_k^n\|}] \leq e^{|\mathcal{K}| \Lambda_{\text{abs}}(\lambda)}$, so by Markov's inequality, for every $a > 0$,

$$\mathbb{P}\left(\frac{1}{n} \sum_{k \in \mathcal{K}} \|X_k^n\| > a\right) \leq \exp(-n\lambda a + |\mathcal{K}| \Lambda_{\text{abs}}(\lambda)), \quad (23)$$

with $\Lambda_{\text{abs}}(\lambda) < \infty$ by Lemma 9.2. This is the exact analogue of the Chernoff bound in the proof of Proposition 7.2, with independence replaced by the tower property and Lemma 9.1.

Step 2: confinement. Taking $s = 0$ and $\mathcal{K}_n(0, t) = \{1, \dots, n\}$ in (22), $d(x_0, \hat{Y}^n(t)) \leq \|V\|_\infty + T_n$ for all t , so on the event $\{T_n \leq R\}$ the path stays in $K_R := \overline{B}(x_0, R + \|V\|_\infty)$, compact by Hopf–Rinow (Theorem 2.25). By (23) with $\mathcal{K} = \{1, \dots, n\}$ and $\lambda = 1$: for $R_\alpha := \alpha + \Lambda_{\text{abs}}(1)$,

$$\mathbb{P}(T_n > R_\alpha) \leq e^{-n\alpha} \quad \text{for all } n \geq 1. \quad (24)$$

Step 3: oscillation control. Fix $\varepsilon > 0$ and $\eta \in (0, 1]$, and cover $[0, 1]$ by the $m := \lceil 1/\eta \rceil$ windows $J_i := [(i-1)\eta, i\eta] \cap [0, 1]$. Every $[s, t]$ with $t - s \leq \eta$ is contained in $J_i \cup J_{i+1}$ for some i , and the index set $\mathcal{K}_n(J_i \cup J_{i+1})$ of substeps meeting two adjacent windows has at most $2n\eta + 4$ elements. Hence, by (22), if $\text{osc}_\eta(\hat{Y}^n) > \varepsilon$ then some pair of adjacent windows carries noise mass $\frac{1}{n} \sum_k \|X_k^n\| > \varepsilon - \eta\|V\|_\infty$, and the union bound with (23) gives, for every $\lambda \geq 0$,

$$\mathbb{P}(\text{osc}_\eta(\hat{Y}^n) > \varepsilon) \leq \left\lceil \frac{1}{\eta} \right\rceil \exp(-n\lambda(\varepsilon - \eta\|V\|_\infty) + (2n\eta + 4) \Lambda_{\text{abs}}(\lambda)),$$

so that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\text{osc}_\eta(\hat{Y}^n) > \varepsilon) \leq -\lambda(\varepsilon - \eta\|V\|_\infty) + 2\eta \Lambda_{\text{abs}}(\lambda).$$

Since osc_η is nondecreasing in η , the left-hand side is nondecreasing in η and its limit as $\eta \downarrow 0$ exists; the right-hand side tends to $-\lambda\varepsilon$ as $\eta \downarrow 0$, and λ was arbitrary, so

$$\lim_{\eta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\text{osc}_\eta(\hat{Y}^n) > \varepsilon) = -\infty \quad \text{for every } \varepsilon > 0. \quad (25)$$

Step 4: assembly. Fix $\alpha < \infty$ and let R_α be as in Step 2. For each $j \in \mathbb{N}$, apply (25) with $\varepsilon = 1/j$ to obtain $\eta'_j > 0$ and $N_j \in \mathbb{N}$ such that

$$\mathbb{P}(\text{osc}_{\eta'_j}(\hat{Y}^n) > 1/j) \leq 2^{-j} e^{-n\alpha} \quad \text{for all } n \geq N_j.$$

The finitely many remaining indices $n < N_j$ are handled by shrinking η'_j : for fixed n , $\text{osc}_\eta(\hat{Y}^n) \leq \eta(\|V\|_\infty + \max_{k \leq n} \|X_k^n\|) \rightarrow 0$ almost surely as $\eta \downarrow 0$, so $\mathbb{P}(\text{osc}_\eta(\hat{Y}^n) > 1/j) \rightarrow 0$ as $\eta \downarrow 0$, and we may pick $\eta_j \in (0, \eta'_j]$ with $\mathbb{P}(\text{osc}_{\eta_j}(\hat{Y}^n) > 1/j) \leq 2^{-j} e^{-n\alpha}$ also for $n = 1, \dots, N_j - 1$; by monotonicity of osc_η in η , the bound for $n \geq N_j$ survives the shrinking. Now set

$$K_\alpha := \{\gamma \in C([0, 1]; M) : \gamma(0) = x_0, \gamma([0, 1]) \subseteq K_{R_\alpha}, \text{osc}_{\eta_j}(\gamma) \leq 1/j \ \forall j \in \mathbb{N}\}.$$

K_α is closed under uniform convergence (each defining condition passes to uniform limits), its members take values in the compact set K_{R_α} , and they are equicontinuous ($|s - t| \leq \eta_j$ implies $d(\gamma(s), \gamma(t)) \leq 1/j$); by the Arzelà–Ascoli theorem, K_α is compact in $(C([0, 1]; M), d_\infty)$. Finally, for every $n \geq 1$, combining (24) with the choices above,

$$\mathbb{P}(\hat{Y}^n \notin K_\alpha) \leq \mathbb{P}(T_n > R_\alpha) + \sum_{j=1}^{\infty} \mathbb{P}(\text{osc}_{\eta_j}(\hat{Y}^n) > 1/j) \leq e^{-n\alpha} + \sum_{j=1}^{\infty} 2^{-j} e^{-n\alpha} = 2e^{-n\alpha},$$

whence $\limsup_n \frac{1}{n} \log \mathbb{P}(\hat{Y}^n \notin K_\alpha) \leq -\alpha$. □

Proposition 9.7 is used twice in this section. First, it replaces the deterministic tube. In Layer 1 (Subsection 9.5) all Grönwall-type comparisons between the original and truncated walks are run on the event $\{T_n \leq R\}$, on which both walks are confined to K_R . That is exactly how the flat proof of Proposition 7.2 runs its Euler estimate: on $\{T_n \leq R\}$ first, and then disposing of the complement by (24). Second, exponential tightness guarantees that the abstract limit rate function produced by the approximation machinery of Layer 2 (Subsection 9.6) is a *good* rate function [3, Theorem 4.2.16 and Lemma 1.2.18].

9.4 Layer 0: the theorem at a fixed truncation level

The entire bounded-increment machinery of Section 8 applies verbatim to the truncated families of Lemma 9.3, at every fixed truncation level $c > 0$. That machinery is: well-posedness of the drift ODE, the a priori bounds, the continuity of the drift map on $AC_{x_0, r}$, the restricted Mogulskii theorem, the contraction step, and the exponential equivalence. Two pieces of bookkeeping are needed: a parametrised restatement of the bounded-increment theorem, so that it can be cited with $r = c$; and the observation that bounded geometry makes the injectivity-radius hypothesis hold at every level simultaneously. The exponential-map Lipschitz bound quoted here and in Subsection 9.5 is Lemma 2.33, whose compact set and radius are free.

The bounded-increment path-space LDP (Theorem 6.2(ii)) holds for any consistent bounded family at any radius, with constants depending only on (K_0, r, V) . Layer 0 of the tower is now a matter of instantiating it at $r = c$ and checking hypotheses.

Corollary 9.8 (Level- c path-space LDP). *Let the assumptions of Theorem 6.5 be satisfied. For $c > 0$, let $\hat{Y}^{n,(c)}$ denote the interpolated drifted walk driven by the truncated family $\{\mu_x^{(c)}\}$ of Lemma 9.3. Then for every $c > 0$, the processes $\hat{Y}^{n,(c)}$ satisfy the LDP on $(C_{x_0}([0, 1]; M), d_\infty)$ with speed n and good rate function*

$$I_c(\phi) := \begin{cases} \int_0^1 \Lambda_{\phi(t)}^{(c)*}(\dot{\phi}(t) - V(\phi(t))) dt & \text{if } \phi \in \text{AC}([0, 1]; M) \text{ and } \phi(0) = x_0, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\Lambda_x^{(c)*}$ is the Legendre transform of the truncated log-moment generating function $\Lambda_x^{(c)}$.

Proof. We verify that the family $\{\mu_x^{(c)}\}$ satisfies the hypotheses of Theorem 6.2(ii) with $r = c$. Consistency and $\text{supp}(\mu_x^{(c)}) \subseteq \overline{B}(0, c)$ are Lemma 9.3(i). For the injectivity-radius hypothesis, write $K_0(c) := \overline{B}(x_0, c + \|V\|_\infty)$ for the level- c tube. For *any* compact $K \subset M$ one has $\text{inj}(K) = \inf_{p \in K} \iota(p) \geq \text{inj}(M)$, so bounded geometry gives, *uniformly in c ,*

$$\text{inj}(K_0(c)) \geq \text{inj}(M) > \|V\|_\infty,$$

the last inequality being the hypothesis of Theorem 6.5. The two remaining hypotheses, completeness of M and smoothness and boundedness of V , are assumptions of Theorem 6.5. All hypotheses of Theorem 6.2(ii) therefore hold at level c , and its conclusion is the stated LDP. $\square$

9.5 Layer 1: exponentially good approximation

The probabilistic core of the tower is a coupling. The original walk and the level- c walks are realised on a common probability space in such a way that they differ, up to a Grönwall factor, by the *removed mass*

$$D_n^{(c)} := \frac{1}{n} \sum_{k=1}^n \|X_k^n\| \mathbf{1}_{\{\|X_k^n\| > c\}},$$

the rescaled total noise carried by the truncated-away increments. We use a Chernoff bound to show that $D_n^{(c)}$ is superexponentially negligible in the limit ($n \rightarrow \infty$, then $c \rightarrow \infty$), and a pathwise estimate on the tightness event $\{T_n \leq R\}$ converts this into uniform closeness of the walks. The argument again follows the flat $\{T_n \leq R\}$ decomposition of Proposition 7.2. The curvature half of the bounded-geometry hypothesis, $\|\text{Rm}\|_\infty =: \kappa < \infty$, enters through the one-step stability estimate of Lemma 9.11. Throughout, we also write $\iota_M := \text{inj}(M) > 0$. The notion of approximation in play is Definition 3.10, and the theorem that consumes it is Theorem 3.11.

The coupling. Fix n and $c > 0$, write $h = 1/n$, and let $T_c : T_x M \rightarrow T_x M$, $T_c(v) := v \mathbf{1}_{\{\|v\| \leq c\}}$, denote the radial truncation map, so that $(T_c)_* \mu_x = \mu_x^{(c)}$ for every x (the computation is that of Lemma 9.3(i)). Let $\{S_k^n\}$ be the drifted walk of Definition 5.3 with increments $X_{k+1}^n \sim \mu_{S_k^n}$. On the same probability space define $S_0^{n,c} := x_0$ and, recursively,

$$X_{k+1}^{n,c} := \tau_k(T_c(X_{k+1}^n)), \quad S_{k+1}^{n,c} := \text{Exp}_{S_k^{n,c}}\left(\frac{1}{n}(V(S_k^{n,c}) + X_{k+1}^{n,c})\right), \quad (26)$$

where $\tau_k : T_{S_k^n} M \rightarrow T_{S_k^{n,c}} M$ is parallel transport from S_k^n to $S_k^{n,c}$

- along the unique minimising geodesic, if $d(S_k^n, S_k^{n,c}) < \iota_M$;
- along the concatenation of the reversed original walk path $\hat{Y}^n|_{[0,t_k]}$ and the truncated walk path $\hat{Y}^{n,(c)}|_{[0,t_k]}$, otherwise.

The second branch is always defined (both paths are piecewise geodesic, hence piecewise smooth) and makes the coupling well-posed with no smallness condition; the estimates below will show that on the relevant events only the first branch is ever active. We write $\hat{Y}^{n,(c)}$ for the piecewise-geodesic interpolation of $\{S_k^{n,c}\}$, in the same way as $\hat{Y}^n$ interpolates $\{S_k^n\}$. The construction is the truncation analogue of the coupling (17) used in Subsection 8.4.1.

Lemma 9.9 (The coupled walk has the level- c law). *The process $\{S_k^{n,c}\}$ defined by (26) is a drifted geodesic random walk driven by the truncated family $\{\mu_x^{(c)}\}$, in the sense of Definition 5.3 (without the boundedness). In particular, $\hat{Y}^{n,(c)}$ has the law of the level- c interpolated walk, and Corollary 9.8 applies to it.*

Proof. Let $\mathcal{F}_k$ be the σ -algebra generated by $X_1^n, \dots, X_k^n$; then $S_k^n, S_k^{n,c}$ and the transport τ_k are $\mathcal{F}_k$ -measurable (τ_k depends measurably on the data: on the first branch by the smoothness of Lemma 2.29, on the second because it is a finite composition of transports along the first k substep geodesics, each depending smoothly on the increments). Conditionally on $\mathcal{F}_k$, the increment X_{k+1}^n has law $\mu_{S_k^n}$, so

$$\text{Law}(X_{k+1}^{n,c} \mid \mathcal{F}_k) = (\tau_k)_*(T_c)_* \mu_{S_k^n} = (\tau_k)_* \mu_{S_k^n}^{(c)} = \mu_{S_k^{n,c}}^{(c)},$$

where the last equality is the consistency of the truncated family (Lemma 9.3(i)): τ_k is parallel transport along a piecewise-smooth curve from S_k^n to $S_k^{n,c}$, and consistency along piecewise-smooth curves follows from consistency along smooth ones by composing the defining relation over the pieces. Since the conditional law of $X_{k+1}^{n,c}$ given the entire past depends only on the current position $S_k^{n,c}$ and equals $\mu^{(c)}$ there, $\{S_k^{n,c}\}$ is a drifted geodesic random walk driven by $\{\mu_x^{(c)}\}$. $\square$

Lemma 9.10 (Tail estimate on the removed mass). *Assume Cramér's condition and set $\varepsilon_c(\lambda) := \mathbb{E}[e^{\lambda\|X\|} \mathbf{1}_{\{\|X\| > c\}}]$ for $\lambda \geq 0$. Then for every $\delta > 0$ and $\lambda \geq 0$,*

$$\mathbb{P}(D_n^{(c)} > \delta) \leq e^{-n\lambda\delta} (1 + \varepsilon_c(\lambda))^n,$$

and $\varepsilon_c(\lambda) \rightarrow 0$ as $c \rightarrow \infty$ for each fixed λ . Consequently,

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(D_n^{(c)} > \delta) = -\infty \quad \text{for every } \delta > 0.$$

Proof. For a single increment, by Lemma 9.1,

$$\mathbb{E}[e^{\lambda\|X\|} \mathbf{1}_{\{\|X\| > c\}} \mid \mathcal{F}_{k-1}] = \mathbb{P}(\|X\| \leq c) + \mathbb{E}[e^{\lambda\|X\|} \mathbf{1}_{\{\|X\| > c\}}] \leq 1 + \varepsilon_c(\lambda),$$

regardless of the (random) base point. Applying the tower property to the increments, as in the proof of Proposition 9.7, $\mathbb{E}[e^{n\lambda D_n^{(c)}}] \leq (1 + \varepsilon_c(\lambda))^n$, and Markov's inequality gives the display. Since $e^{\lambda\|X\|} \mathbf{1}_{\{\|X\| > c\}} \leq e^{\lambda\|X\|}$, which is integrable by Lemma 9.2, dominated convergence gives $\varepsilon_c(\lambda) \downarrow 0$. Hence

$$\limsup_n \frac{1}{n} \log \mathbb{P}(D_n^{(c)} > \delta) \leq -\lambda\delta + \log(1 + \varepsilon_c(\lambda)) \xrightarrow{c \rightarrow \infty} -\lambda\delta$$

for each fixed λ (the left-hand side is nonincreasing in c , since $D_n^{(c)}$ is, so the c -limit exists). Since this bound holds for every $\lambda \geq 0$, letting $\lambda \rightarrow \infty$ forces the c -limit to $-\infty$, which is the claim. $\square$

The geometric ingredient is a one-step stability estimate: two geodesic steps launched from nearby points with parallel initial velocities separate by at most a factor exponential in the *length* of the step.

Lemma 9.11 (One-step stability). *Let (M, g) be a complete manifold satisfying $\kappa = \|\text{Rm}\|_\infty < \infty$, the operator-norm bound of (1). Let $q, p \in M$ with $d(q, p) < \iota(q)$, let $\tau := P_{q \rightarrow p}$ be parallel transport along the unique minimising geodesic, and let $w \in T_q M$. Then*

$$d(\text{Exp}_q(w), \text{Exp}_p(\tau w)) \leq e^{\sqrt{\kappa} \|w\|} d(q, p).$$

Proof. Let $\gamma : [0, 1] \rightarrow M$ be the minimising geodesic from q to p , of constant speed $\|\dot{\gamma}\| = d(q, p) =: \delta$, let $W(s) := \tau_{q \rightarrow \gamma(s)}^\gamma w$ be the parallel field along γ with $W(0) = w$, $W(1) = \tau w$, and consider the variation

$$\Gamma(s, t) := \text{Exp}_{\gamma(s)}(t W(s)), \quad (s, t) \in [0, 1]^2,$$

smooth and globally defined by completeness. For fixed s , the curve $t \mapsto \Gamma(s, t)$ is a geodesic with speed $\|W(s)\| = \|w\|$, and $J_s(t) := \partial_s \Gamma(s, t)$ is a Jacobi field along it (it is the variation field of a family of geodesics), with

$$J_s(0) = \dot{\gamma}(s), \quad D_t J_s(0) = D_s W(s) = 0,$$

using Lemma 2.13 and that W is parallel.

These are the hypotheses of Lemma 2.32, applied to the geodesic $t \mapsto \Gamma(s, t)$ of constant speed $\|w\|$ and to the Jacobi field J_s along it, whose initial value has norm $\|J_s(0)\| = \|\dot{\gamma}(s)\| = \delta$. That lemma gives

$$\|J_s(1)\| \leq e^{\sqrt{\kappa} \|w\|} \delta \quad (s \in [0, 1]).$$

The curve $s \mapsto \Gamma(s, 1)$ joins $\text{Exp}_q(w)$ to $\text{Exp}_p(\tau w)$, and distance is at most length (Lemma 2.28(ii)), so

$$d(\text{Exp}_q(w), \text{Exp}_p(\tau w)) \leq \int_0^1 \|J_s(1)\| \, ds \leq \delta e^{\sqrt{\kappa} \|w\|}.$$

□

We can now run the discrete Grönwall argument. Recall $K_R = \overline{B}(x_0, R + \|V\|_\infty)$ from Subsection 9.3, and set $K_R^+ := \overline{B}(x_0, R + \|V\|_\infty + \iota_M)$, $C_V(R) := \sup_{K_R^+} \|\nabla V\|_{\text{op}}$, both finite by compactness and smoothness of V .

Proposition 9.12 (Grönwall stability of the coupling). *Assume the hypotheses of Theorem 6.5. For $R > 0$ put*

$$L(R) := \sqrt{\kappa} (\|V\|_\infty + R) + C_{\text{exp}}(K_R, 1) C_V(R), \quad \delta_{\max}(R) := \frac{1}{8} \iota_M e^{-L(R)}.$$

Let $c \geq 2\|V\|_\infty$, $n \geq c + \|V\|_\infty$, and $0 < \delta \leq \delta_{\max}(R)$. Then on the event $\{T_n \leq R\} \cap \{D_n^{(c)} \leq \delta\}$, the coupling (26) uses the minimising-geodesic branch at every step, and

$$d_\infty(\hat{Y}^n, \hat{Y}^{n, (c)}) \leq 3 e^{L(R)} D_n^{(c)} + \frac{2(c + \|V\|_\infty)}{n}.$$

Proof. Work on the event $\{T_n \leq R\} \cap \{D_n^{(c)} \leq \delta\}$ throughout, and abbreviate $q_k := S_k^n$, $p_k := S_k^{n,c}$, $e_k := d(q_k, p_k)$, $X_k := X_k^n$, $D_n := D_n^{(c)}$. *Confinement:* since $\|X_k^{n,c}\| \leq \|X_k\|$, both walks (and their interpolants) move total noise distance at most $T_n \leq R$, so they remain in K_R ; and any minimising geodesic between q_k and p_k with $e_k \leq \iota_M$ stays in K_R^+ .

One-step recursion. Suppose $e_j < \iota_M$ for all $j \leq k$ (true for $k = 0$), so that the transports $\tau_0, \dots, \tau_k$ are all along minimising geodesics. Write $u := h(V(q_k) + X_{k+1}) \in T_{q_k}M$, so $q_{k+1} = \text{Exp}_{q_k}(u)$ and $p_{k+1} = \text{Exp}_{p_k}(h(V(p_k) + \tau_k T_c X_{k+1}))$.

If $\|X_{k+1}\| \leq c$ (a *kept* step), then $\tau_k T_c X_{k+1} = \tau_k X_{k+1}$ and we compare through $\text{Exp}_{p_k}(\tau_k u)$:

$$e_{k+1} \leq d(\text{Exp}_{q_k}(u), \text{Exp}_{p_k}(\tau_k u)) + d(\text{Exp}_{p_k}(\tau_k u), \text{Exp}_{p_k}(h(V(p_k) + \tau_k X_{k+1}))).$$

The first term is at most $e^{\sqrt{\kappa}\|u\|} e_k$ by Lemma 9.11 ($e_k < \iota_M \leq \iota(q_k)$). For the second, the two vectors at p_k differ by $h\|\tau_k V(q_k) - V(p_k)\|$; parametrising the minimising geodesic γ from q_k to p_k and differentiating $s \mapsto \tau_{\gamma(s) \rightarrow p_k}^\gamma V(\gamma(s))$, whose covariant description gives derivative $\tau_{\gamma(s) \rightarrow p_k}^\gamma(\nabla_{\dot{\gamma}(s)} V)$, we get

$$\|V(p_k) - \tau_k V(q_k)\| \leq \int_0^1 \|\nabla V\|_{\text{op}} \|\dot{\gamma}\| ds \leq C_V(R) e_k,$$

since γ stays in K_R^+ . Both vectors have norm at most $h(\|V\|_\infty + c) \leq 1$ (here $n \geq c + \|V\|_\infty$ is used), so Lemma 2.33 with $K = K_R$, $\rho = 1$ applies and the second term is at most $C_{\text{exp}}(K_R, 1) h C_V(R) e_k$. Combining, and using $e^x + yh \leq e^{x+yh}$ for $x, y \geq 0$ (as $e^x \geq 1$),

$$e_{k+1} \leq e^{\theta_{k+1}} e_k, \quad \theta_{k+1} := \sqrt{\kappa} h(\|V\|_\infty + \|X_{k+1}\|) + C_{\text{exp}}(K_R, 1) C_V(R) h.$$

If $\|X_{k+1}\| > c$ (a *removed* step), then $X_{k+1}^{n,c} = 0$ and we bound crudely: each walk moves at most the length of its step geodesic, so

$$e_{k+1} \leq e_k + h(\|V\|_\infty + \|X_{k+1}\|) + h\|V\|_\infty \leq e^{\theta_{k+1}} e_k + h\beta_{k+1},$$

with $\beta_{k+1} := (\|X_{k+1}\| + 2\|V\|_\infty) \mathbf{1}_{\{\|X_{k+1}\| > c\}}$. The unified recursion $e_{k+1} \leq e^{\theta_{k+1}} e_k + h\beta_{k+1}$ (with $\beta_{k+1} = 0$ for kept steps) unrolls, with $e_0 = 0$, to

$$e_m \leq e^{\sum_{j \leq n} \theta_j} \sum_{k \leq n} h\beta_k \leq e^{L(R)} \left(D_n + \frac{2\|V\|_\infty}{c} D_n \right) \leq 2e^{L(R)} D_n, \quad m \leq n,$$

where we used $\sum_j \theta_j = \sqrt{\kappa}(\|V\|_\infty + T_n) + C_{\text{exp}}(K_R, 1) C_V \leq L(R)$; that the number of removed steps is at most nD_n/c (each of them contributes more than c to the sum nD_n), so that the drift terms in the β_k contribute at most $2\|V\|_\infty D_n/c$ to $\sum_k h\beta_k$; and $c \geq 2\|V\|_\infty$. Since $2e^{L(R)} D_n \leq 2e^{L(R)} \delta_{\text{max}}(R) = \iota_M/4 < \iota_M$, the induction hypothesis $e_{k+1} < \iota_M$ is restored, and the minimising-geodesic branch is active at every step.

Between grid points. For $t \in [t_k, t_{k+1}]$, both interpolants are geodesics; bounding each by its speed,

$$d(\hat{Y}^n(t), \hat{Y}^{n,(c)}(t)) \leq e_k + h(\|V\|_\infty + \|X_{k+1}\|) + h(\|V\|_\infty + \|X_{k+1}^{n,c}\|).$$

For a kept step the last two terms total at most $2h(\|V\|_\infty + c) \leq 2(c + \|V\|_\infty)/n$; for a removed step they total at most $2h\|V\|_\infty + h\|X_{k+1}\| \leq 2h\|V\|_\infty + D_n$ (a single removed term of D_n 's defining sum is at most D_n). In both cases,

$$d_\infty(\hat{Y}^n, \hat{Y}^{n,(c)}) \leq \max_k e_k + D_n + \frac{2(c + \|V\|_\infty)}{n} \leq 3e^{L(R)} D_n + \frac{2(c + \|V\|_\infty)}{n}. \quad \square$$

Theorem 9.13 (The truncated walks are exponentially good approximations). *Assume the hypotheses of Theorem 6.5. For every n and c , the coupling (26) realises $\hat{Y}^n$ and the level- c walk $\hat{Y}^{n,(c)}$ on a common probability space with the correct marginal laws (Lemma 9.9), and the families $\{\hat{Y}^{n,(c)}\}$ are exponentially good approximations of $\{\hat{Y}^n\}$ in $(C_{x_0}([0, 1]; M), d_\infty)$ as $c \rightarrow \infty$: for every $\delta' > 0$,*

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(d_\infty(\hat{Y}^n, \hat{Y}^{n,(c)}) > \delta') = -\infty.$$

Proof. Fix $\delta' > 0$ and $\alpha < \infty$; it suffices to show that the left-hand side is at most $-\alpha$ (the c -limit exists because $D_n^{(c)}$, hence the coupling discrepancy bound, improves as c grows; alternatively read the claim with $\limsup_c$). Let $R := R_\alpha = \alpha + \Lambda_{\text{abs}}(1)$ as in (24), and set

$$\delta_R := \min\left(\frac{\delta'}{6} e^{-L(R)}, \delta_{\max}(R)\right) > 0,$$

which depends on (α, δ') but not on c . For $c \geq 2\|V\|_\infty$ and $n \geq \max(c + \|V\|_\infty, 4(c + \|V\|_\infty)/\delta')$, Proposition 9.12 gives, on the event $\{T_n \leq R\} \cap \{D_n^{(c)} \leq \delta_R\}$,

$$d_\infty(\hat{Y}^n, \hat{Y}^{n,(c)}) \leq 3e^{L(R)}\delta_R + \frac{\delta'}{2} \leq \frac{\delta'}{2} + \frac{\delta'}{2} = \delta'.$$

Hence $\{d_\infty > \delta'\} \subseteq \{T_n > R\} \cup \{D_n^{(c)} > \delta_R\}$ for such n , and by (24) and Lemma 9.10,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(d_\infty > \delta') \leq \max\left(-\alpha, -\lambda\delta_R + \log(1 + \varepsilon_c(\lambda))\right) \quad \text{for every } \lambda \geq 0.$$

Letting $c \rightarrow \infty$ (at fixed λ) and then $\lambda \rightarrow \infty$, the second entry tends to $-\lambda\delta_R$ and then to $-\infty$, so $\lim_c \limsup_n \frac{1}{n} \log \mathbb{P}(d_\infty > \delta') \leq -\alpha$. Since α was arbitrary, the limit is $-\infty$. $\square$

9.6 Layer 2: identification of the rate function

Layers 0 and 1 combine, through the approximation theorem of Dembo and Zeitouni [3, Theorem 4.2.16], into an LDP for the original walk. That theorem does not give the rate function in closed form. It gives the abstract limit

$$\hat{I}(\phi) := \sup_{\delta > 0} \liminf_{c \rightarrow \infty} \inf_{d_\infty(\chi, \phi) \leq \delta} I_c(\chi)$$

of the level- c rates, which is the Γ -liminf of the I_c in d_∞ . Theorem 6.5 claims a specific rate function, the drifted rate functional

$$I_{\text{path}}^V(\phi) = \int_0^1 \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) dt \quad (\phi \in \text{AC}, \phi(0) = x_0; +\infty \text{ otherwise})$$

of (3). What remains is therefore an identification, $\hat{I} = I_{\text{path}}^V$. Pointwise convergence of I_c to I_{path}^V would be the natural route, but it fails: $\Lambda_x^{(c)*}$ is infinite outside $\overline{B}(0, c)$, so a finite-cost path with unbounded noise velocity has $I_c = +\infty$ at every level. The comparison must therefore allow a nearby path, and two inequalities of Γ -convergence type supply it. Proposition 9.14 is the lower bound: along *every* sequence $\phi_j \rightarrow \phi$ with $c_j \rightarrow \infty$, the cost cannot drop below $I_{\text{path}}^V(\phi)$ in the limit. Proposition 9.16 is the matching upper bound, and it need hold along only *one* sequence: for each

ϕ of finite cost it exhibits a nearby path whose level- c costs do not exceed $I_{\text{path}}^V(\phi)$ asymptotically. Together they pin the limit down from both sides. Theorem 6.4 assembles them, and the endpoint theorem follows by contraction. The liminf bound is quantified over all approximating sequences and so carries the compactness work, while the recovery bound is a construction.

The convex-analytic facts underlying both bounds are collected in Subsection 9.2. The velocity tools are the general weak- L^1 lemmas of Subsection 3.2.2: Lemma 3.14 (uniform convergence forces weak- L^1 velocity convergence) and Lemma 3.15 (weak- $L^1 \times$ strong pairing), which here replace the weak- $*$ lemmas used under bounded increments, since the uniform Lipschitz hypothesis is no longer available.

The Γ -liminf bound

Proposition 9.14 (Liminf bound). *Let $c_j \rightarrow \infty$ and let $\phi_j \rightarrow \phi$ in $(C_{x_0}([0, 1]; M), d_\infty)$. Then*

$$\liminf_{j \rightarrow \infty} I_{c_j}(\phi_j) \geq I_{\text{path}}^V(\phi).$$

The proof runs in three steps. First, cut $[0, 1]$ into finitely many windows so short that all the paths concerned stay in one normal chart. On such a window, parallel transport to the centre p_i moves every noise velocity into the single tangent space $T_{p_i}M$, where the convergence $\Lambda^{(c)} \rightarrow \Lambda$ of Lemma 9.3 is available. There the Fenchel inequality $\langle \lambda, u \rangle - \Lambda^{(c)}(\lambda) \leq \Lambda^{(c)*}(u)$ converts that convergence of the generating functions into a lower bound on the conjugates, once the velocities are known to converge weakly, which is where uniform integrability enters; taking the supremum over λ recovers Λ^* in the limit. Summing the window bounds gives the proposition.

Proof. Passing to a subsequence, assume $\lim_j I_{c_j}(\phi_j) = \alpha := \liminf_j I_{c_j}(\phi_j) < \infty$ (otherwise there is nothing to prove) and $\sup_j I_{c_j}(\phi_j) \leq \alpha + 1$; then every ϕ_j is absolutely continuous with $\phi_j(0) = x_0$, and $\phi(0) = x_0$.

Charts. The set $\bar{K} := \{z : d(z, \phi([0, 1])) \leq 1\}$ is compact; put $\rho_\star := \frac{1}{4} \min(\text{inj}(\bar{K}), 1) > 0$. By uniform continuity of ϕ , partition $[0, 1]$ into finitely many windows $U_i = [s_{i-1}, s_i]$ with $\phi(U_i) \subseteq B(p_i, \rho_\star/2)$, $p_i := \phi(s_{i-1})$; for all large j , also $\phi_j(U_i) \subseteq B(p_i, \rho_\star)$. On $B(p_i, 2\rho_\star)$ the normal chart $\log_{p_i}$ and the transports $P_{z \rightarrow p_i}$ along the (unique minimising) radial geodesics are defined and smooth (Lemma 2.29). Write, on U_i , $x_j := \log_{p_i} \circ \phi_j$, $x := \log_{p_i} \circ \phi$, and

$$u_j(t) := P_{\phi_j(t) \rightarrow p_i}(\dot{\phi}_j(t) - V(\phi_j(t))) = \mathcal{A}(x_j(t)) \dot{x}_j(t) - \mathbf{b}(x_j(t)),$$

where $\mathcal{A}(x) := P_{z \rightarrow p_i} \circ ((d \log_{p_i})_z)^{-1}$ and $\mathbf{b}(x) := P_{z \rightarrow p_i} V(z)$, $z := \text{Exp}_{p_i}(x)$, are smooth in x with $\mathcal{A}$ invertible, and all four of $\mathcal{A}, \mathcal{A}^{-1}, \mathbf{b}$ and their derivatives bounded on the closed chart ball. By the transport invariance of the rate integrand (Lemma 8.12, whose proof uses only consistency and so applies verbatim to the family $\{\mu_x^{(c)}\}$),

$$\int_{U_i} \Lambda_{\phi_j(t)}^{(c_j)*}(\dot{\phi}_j - V(\phi_j)) dt = \int_{U_i} \Lambda_{p_i}^{(c_j)*}(u_j(t)) dt.$$

Compactness. By Lemma 9.6(ii) and (21) (all applied at base point p_i ; the constants A_ℓ are base-point free by consistency), the families $\{u_j\}$, and hence $\{\dot{x}_j\} = \{\mathcal{A}^{-1}(x_j)(u_j + \mathbf{b}(x_j))\}$, are uniformly integrable on U_i . By Lemma 3.14, x is absolutely continuous on U_i and $\dot{x}_j \rightharpoonup \dot{x}$ weakly

in $L^1(U_i)$; gluing windows, $\phi \in \text{AC}([0, 1]; M)$. Moreover, by Lemma 3.15 (with $b_j := \mathcal{A}(x_j) \rightarrow \mathcal{A}(x)$ in L^∞ , using uniform convergence of x_j and smoothness of $\mathcal{A}$) and uniform convergence of $\mathfrak{b}(x_j)$,

$$u_j \rightharpoonup u := \mathcal{A}(x)\dot{x} - \mathfrak{b}(x) = P_{\phi(\cdot) \rightarrow p_i}(\dot{\phi} - V(\phi)) \quad \text{weakly in } L^1(U_i).$$

Duality. Let $\lambda : U_i \rightarrow T_{p_i}M$ be a simple function. Then

$$\int_{U_i} [\langle \lambda(t), u_j(t) \rangle - \Lambda_{p_i}^{(c_j)}(\lambda(t))] dt \leq \int_{U_i} \Lambda_{p_i}^{(c_j)*}(u_j) dt,$$

and the left-hand side converges to $\int_{U_i} [\langle \lambda, u \rangle - \Lambda_{p_i}(\lambda)]$: the pairing term by weak- L^1 convergence (test each of the finitely many level sets), and the Λ -term because $\Lambda_{p_i}^{(c_j)} \rightarrow \Lambda_{p_i}$ pointwise, since Lemma 9.3(iii) transports to the base point p_i by consistency. Hence

$$\int_{U_i} [\langle \lambda, u \rangle - \Lambda_{p_i}(\lambda)] dt \leq \beta_i := \liminf_j \int_{U_i} \Lambda_{p_i}^{(c_j)*}(u_j) dt$$

for every simple λ . Taking λ ranging over simple functions with values in a countable dense subset $\{\lambda_r\}$ of $\overline{B}(0, N) \subseteq T_{p_i}M$, choosing on each of finitely many measurable sets the index $r \leq R$ maximising $\langle \lambda_r, u(t) \rangle - \Lambda(\lambda_r)$, and letting $R \rightarrow \infty$ and then $N \rightarrow \infty$ (monotone convergence twice, using continuity in λ), the left-hand side increases to $\int_{U_i} \Lambda_{p_i}^*(u) dt$. By transport invariance again, $\int_{U_i} \Lambda_{p_i}^*(u) = \int_{U_i} \Lambda_{\phi(t)}^*(\dot{\phi} - V(\phi)) dt$, and summing over the windows,

$$I_{\text{path}}^V(\phi) = \sum_i \int_{U_i} \Lambda_{\phi(t)}^*(\dot{\phi} - V(\phi)) dt \leq \sum_i \beta_i \leq \liminf_j I_{c_j}(\phi_j). \quad \square$$

The recovery bound

The goal is now the reverse inequality, and unlike the liminf bound it asks for one path rather than all of them: given ϕ with $I_{\text{path}}^V(\phi) < \infty$ and given $\delta > 0$, we must exhibit a single *recovery path* ϕ^* with $d_\infty(\phi^*, \phi) \leq \delta$ and $\limsup_c I_c(\phi^*) \leq I_{\text{path}}^V(\phi)$. That is Proposition 9.16. The path is built by modifying the noise velocity of ϕ and then integrating the modified velocity back up. Lemma 9.15 performs the second step. It turns a prescribed noise velocity into an actual path, controls how far that path strays from ϕ , and shows the cost is unchanged by the transport.

The recovery construction modifies a finite-cost path in two steps, both acting on its *noise velocity* $w := \dot{\phi} - V(\phi)$. Values of norm larger than ϱ are replaced by the mean m , which costs nothing by Lemma 9.4(i). The result is then contracted towards m by a factor $1 - \theta$, which decreases the cost by convexity. The modified velocity is bounded and takes values in a margin compactum, where Lemma 9.5(ii) applies; and the modified velocity field integrates to a nearby path by the following well-posedness lemma, the one place where the drift-ODE theory of Subsection 8.2.1 is invoked beyond bounded speeds.

Lemma 9.15 (Prescribing the noise velocity along a finite-cost path). *Let $\phi \in \text{AC}([0, 1]; M)$ with $\phi(0) = x_0$ and $w := \dot{\phi} - V(\phi) \in L^1$, let $K'' \subset M$ be a compact set containing the closed 1-neighbourhood of $\phi([0, 1])$, and set $C_V'' := \sup_{K''} \|\nabla V\|_{\text{op}}$ and $\rho'' := \frac{1}{4} \min(\text{inj}(K''), 1)$. Let $\tilde{w}(t) \in T_{\phi(t)}M$ be measurable and bounded with*

$$\Delta := \int_0^1 \|\tilde{w}(t) - w(t)\| dt \leq \rho'' e^{-C_V''}.$$

Then the initial value problem

$$\dot{\phi}'(0) = x_0, \quad \dot{\phi}'(t) = V(\phi'(t)) + P_{\phi(t) \rightarrow \phi'(t)} \tilde{w}(t) \quad (27)$$

(parallel transport along minimising geodesics) has a unique absolutely continuous solution ϕ' on $[0, 1]$; it satisfies $d_\infty(\phi', \phi) \leq e^{C''_V} \Delta$, and its noise velocity is $\dot{\phi}' - V(\phi') = P_{\phi \rightarrow \phi'} \tilde{w}$, so that for every consistent family of increment measures (in particular $\{\mu_x\}$ and each $\{\mu_x^{(c)}\}$), the associated cost satisfies

$$\int_0^1 \tilde{\Lambda}_{\phi'(t)}^* (\dot{\phi}'(t) - V(\phi'(t))) dt = \int_0^1 \tilde{\Lambda}_{\phi(t)}^* (\tilde{w}(t)) dt.$$

Proof. A priori separation. Suppose ϕ' solves (27) on $[0, T]$ with $\delta(t) := d(\phi(t), \phi'(t)) < \rho''$ there, so all transports are along unique minimising geodesics inside K'' . Since ϕ itself satisfies (27) with $\tilde{w}$ replaced by w (trivially: $P_{\phi \rightarrow \phi} = \text{id}$), the first-variation computation of Lemma 8.2, run along the geodesic variation from the base curve ϕ to ϕ' exactly as in the proof of Lemma 8.11, gives for almost every t

$$\delta'(t) \leq \|\dot{\phi}'(t) - P_{\phi(t) \rightarrow \phi'(t)} \dot{\phi}(t)\| = \|V(\phi') - P V(\phi) + P(\tilde{w} - w)\| \leq C''_V \delta(t) + \|\tilde{w}(t) - w(t)\|,$$

using the transport-derivative bound $\|V(p) - P_{q \rightarrow p} V(q)\| \leq C''_V d(q, p)$ established in the proof of Proposition 9.12 (the connecting geodesics stay in K''). Grönwall yields $\delta(t) \leq e^{C''_V} \Delta \leq \rho''$ on $[0, T]$; the usual bootstrap (a continuity argument: were the separation to reach ρ'' it would first cross every smaller level, contradicting the Grönwall bound just obtained) then shows that the separation never reaches ρ'' , so the regime is never exited.

Existence and uniqueness. By absolute continuity of the integral, partition $[0, 1]$ into finitely many intervals on each of which $\int (\|\dot{\phi}\| + \|V\|_\infty + \|\tilde{w}\|_\infty) dt \leq \rho''/4$. On each interval, centre a normal chart at the current solution point; the source ϕ moves at most $\rho''/4$, stays within $\rho''/4 + \delta \leq \rho''$ of the chart centre, and the candidate solution is confined to the chart ball, so the chart representation of (27) satisfies hypotheses (C1)–(C4) of Theorem 3.18 exactly as in Lemma 8.4. The majorant is $h(t) = C(\|V\|_\infty + \|\tilde{w}\|_\infty)$ and the Lipschitz rate is $\ell(t) = C(1 + \|\tilde{w}\|_\infty)$. Both are integrable, which is all Theorem 3.18 requires. (The uniform substep length of Subsection 8.2.1 is replaced by the non-uniform partition above; nothing else changes.) Patching and the open-closed argument of Theorem 8.7 give global existence and uniqueness.

Cost identity. The identity $\dot{\phi}' - V(\phi') = P_{\phi \rightarrow \phi'} \tilde{w}$ holds almost everywhere by (27), and Lemma 8.12 (applied to the given consistent family, along the minimising geodesic from $\phi(t)$ to $\phi'(t)$) converts $\tilde{\Lambda}_{\phi'(t)}^*(P\tilde{w}(t))$ into $\tilde{\Lambda}_{\phi(t)}^*(\tilde{w}(t))$ pointwise. $\square$

Proposition 9.16 (Recovery bound). *Let ϕ satisfy $I_{\text{path}}^V(\phi) < \infty$ and let $\delta > 0$. There exists ϕ^* with $d_\infty(\phi^*, \phi) \leq \delta$ and*

$$\limsup_{c \rightarrow \infty} I_c(\phi^*) \leq I_{\text{path}}^V(\phi).$$

Proof. Write $w := \dot{\phi} - V(\phi)$; by Lemma 9.6(ii) (with $c = \infty$), $\int \|w\| \leq I_{\text{path}}^V(\phi) + A_1 < \infty$. Let $m_{\phi(t)} \in T_{\phi(t)} M$ denote the mean vector field, which is transport-covariant by consistency, of constant norm $\|m\|$. For $\theta \in (0, 1)$ and $\varrho > 0$ define, in each $T_{\phi(t)} M$,

$$w^{[\varrho]}(t) := w(t) \mathbf{1}_{\{\|w(t)\| \leq \varrho\}} + m_{\phi(t)} \mathbf{1}_{\{\|w(t)\| > \varrho\}}, \quad \tilde{w}_{\theta, \varrho}(t) := (1 - \theta) w^{[\varrho]}(t) + \theta m_{\phi(t)}.$$

Then $\tilde{w}_{\theta,\varrho}$ is measurable and bounded by $(1-\theta)\varrho + \|m\|$, and

$$\Lambda_{\phi(t)}^*(\tilde{w}_{\theta,\varrho}(t)) \leq (1-\theta)\Lambda_{\phi(t)}^*(w^{[\varrho]}(t)) \leq \Lambda_{\phi(t)}^*(w(t)) \quad \text{a.e.}, \quad (28)$$

by convexity with $\Lambda_{\phi(t)}^*(m_{\phi(t)}) = 0$ (Lemma 9.4(i)) and, on $\{\|w\| > \varrho\}$, $\Lambda^* \geq 0$. Moreover

$$\Delta(\theta, \varrho) := \int_0^1 \|\tilde{w}_{\theta,\varrho} - w\| dt \leq \theta \int_0^1 \|m_{\phi(t)} - w\| dt + \int_{\{\|w\| > \varrho\}} \|m_{\phi(t)} - w\| dt \xrightarrow[\theta \rightarrow 0]{\varrho \rightarrow \infty} 0.$$

Fix θ and ϱ such that $e^{C''_V} \Delta(\theta, \varrho) \leq \min(\delta, \rho'')$ (notation of Lemma 9.15, with K'' the closed 1-neighbourhood of $\phi([0, 1])$), and let ϕ^* be the solution provided by Lemma 9.15 for $\tilde{w} := \tilde{w}_{\theta,\varrho}$; then $d_\infty(\phi^*, \phi) \leq \delta$, and for every c ,

$$I_c(\phi^*) = \int_0^1 \Lambda_{\phi(t)}^{(c)*}(\tilde{w}_{\theta,\varrho}(t)) dt.$$

It remains to let $c \rightarrow \infty$. Transport all values to the base point x_0 along the path $\phi|_{[0,t]}$ itself: writing $\bar{w}(t) \in T_{x_0}M$ for the transported $\tilde{w}_{\theta,\varrho}(t)$, consistency gives $\Lambda_{\phi(t)}^{(c)*}(\tilde{w}_{\theta,\varrho}(t)) = \Lambda^{(c)*}(\bar{w}(t))$ and $\Lambda_{\phi(t)}^*(\tilde{w}_{\theta,\varrho}(t)) = \Lambda^*(\bar{w}(t))$, and maps the support data $(\mathcal{S}, m)$ of each $T_{\phi(t)}M$ isometrically onto those of $T_{x_0}M$. Almost everywhere, $w^{[\varrho]}(t) \in \text{dom } \Lambda_{\phi(t)}^* \subseteq \mathcal{S}_{\phi(t)}$ (Lemma 9.4(ii); on $\{\|w\| > \varrho\}$ use $m \in \mathcal{S}$), so $\bar{w}(t)$ lies in

$$\mathcal{K}_{\theta,\varrho} := ((1-\theta)\mathcal{S} + \theta m) \cap \overline{B}(0, (1-\theta)\varrho + \|m\|) \subseteq T_{x_0}M,$$

a compact convex set which is an s -margin compactum with $s := \theta r_0/2$, where $r_0 > 0$ is any radius with $(m + r_0 \overline{B}(0, 1)) \cap \mathcal{A} \subseteq \mathcal{S}$ (such r_0 exists since $m \in \text{ri } \mathcal{S}$): indeed, for $y = (1-\theta)u + \theta m$ with $u \in \mathcal{S}$,

$$y + \theta r_0 (\overline{B}(0, 1) \cap E) = (1-\theta)u + \theta(m + r_0 \overline{B}(0, 1) \cap E) \subseteq (1-\theta)\mathcal{S} + \theta\mathcal{S} = \mathcal{S}.$$

With $K := \mathcal{K}_{\theta,\varrho}$,

$$I_c(\phi^*) = \int_0^1 \Lambda^{(c)*}(\bar{w}(t)) dt \leq \int_0^1 \Lambda^*(\bar{w}(t)) dt + \sup_K (\Lambda^{(c)*} - \Lambda^*)^+,$$

and by Lemma 9.5(ii) the supremum vanishes as $c \rightarrow \infty$, so that by (28)

$$\limsup_{c \rightarrow \infty} I_c(\phi^*) \leq \int_0^1 \Lambda_{\phi(t)}^*(\tilde{w}_{\theta,\varrho}(t)) dt \leq I_{\text{path}}^V(\phi).$$

(Measurability of $\bar{w}$: the transport along $\phi|_{[0,t]}$ depends continuously on t .) □

Identification and proof of the main theorem

Proof of Theorem 6.4. By Corollary 9.8, Lemma 9.9 and Theorem 9.13, the walks $\hat{Y}^{n,(c)}$ satisfy, for each c , the LDP with good rate function I_c and are exponentially good approximations of $\hat{Y}^n$ as $c \rightarrow \infty$. By Theorem 3.11, $\hat{Y}^n$ satisfies a weak LDP with the rate function

$$\hat{I}(\phi) := \sup_{\delta > 0} \liminf_{c \rightarrow \infty} \inf_{d_\infty(\chi, \phi) \leq \delta} I_c(\chi).$$

Combined with the exponential tightness of $\{\hat{Y}^n\}$ (Proposition 9.7), the weak LDP upgrades to a full LDP and $\hat{I}$ is a good rate function (Theorem 3.9). It remains to identify $\hat{I} = I_{\text{path}}^V$; we show the chain

$$I_{\text{path}}^V \leq \hat{I} \leq \hat{I}^+ := \sup_{\delta > 0} \limsup_{c \rightarrow \infty} \inf_{d_\infty(\chi, \phi) \leq \delta} I_c(\chi) \leq I_{\text{path}}^V,$$

which also shows that the liminf and limsup variants of the abstract rate coincide here.

Left inequality. Let $\beta := \hat{I}(\phi) < \infty$ (otherwise nothing to prove). For each $j \in \mathbb{N}$, taking $\delta = 1/j$, there are $c_j \geq j$ and ϕ_j with $d_\infty(\phi_j, \phi) \leq 1/j$ and $I_{c_j}(\phi_j) \leq \beta + 1/j$. Then $\phi_j \rightarrow \phi$ and Proposition 9.14 gives $I_{\text{path}}^V(\phi) \leq \liminf_j I_{c_j}(\phi_j) \leq \beta$.

Right inequality. Let $I_{\text{path}}^V(\phi) < \infty$ and $\delta > 0$. With ϕ^* from Proposition 9.16,

$$\limsup_{c \rightarrow \infty} \inf_{d_\infty(\chi, \phi) \leq \delta} I_c(\chi) \leq \limsup_{c \rightarrow \infty} I_c(\phi^*) \leq I_{\text{path}}^V(\phi);$$

take the supremum over δ . The middle inequality $\hat{I} \leq \hat{I}^+$ is trivial. $\square$

The chapter's main theorem now follows.

Proof of Theorem 6.5. The evaluation map $\text{ev}_1 : C_{x_0}([0, 1]; M) \rightarrow M$, $\phi \mapsto \phi(1)$, is continuous, and $S_n^n = \hat{Y}^n(1)$. By Theorem 6.4 and the contraction principle (Theorem 3.5), $\{S_n^n\}$ satisfies the LDP on M with speed n and good rate function

$$I_M^V(x) = \inf\{I_{\text{path}}^V(\phi) : \phi(1) = x\} = \inf_{\substack{\phi \in \text{AC}([0, 1]; M) \\ \phi(0) = x_0, \phi(1) = x}} \int_0^1 \Lambda_{\phi(t)}^*(\dot{\phi}(t) - V(\phi(t))) \, dt. \quad \square$$

10 Outlook: position-dependent noise

Everything in this thesis rests on consistency: the increment laws at different points are the same law, moved around by parallel transport. A next question would be what happens when they are not. This chapter states what we expect in the mildest case where they are not, and where we give a rough estimate of how much of the machinery survives. The claims are labelled conjectures, and the estimates below are quoted at the level of which quantity controls which, without the constants.

The mildest case is to keep a consistent family underneath and modulate it by one smooth linear factor.

Definition 10.1 (σ -modulated drifted walk). Let $\{\mu_x\}$ be a consistent family, $V \in \Gamma(TM)$ smooth and bounded, and $\sigma \in \Gamma(\text{End}(TM))$ a smooth field of endomorphisms of TM that is bounded together with $\nabla\sigma$ and uniformly elliptic, so that $\sigma(x)^{-1}$ exists with a bound uniform in x . Put

$$S_{k+1}^n := \text{Exp}_{S_k^n} \left(\frac{1}{n} (V(S_k^n) + \sigma(S_k^n) X_{k+1}^n) \right), \quad X_{k+1}^n \sim \mu_{S_k^n}, \quad (29)$$

started at x_0 , and write $\hat{Y}^{n,\sigma}$ for the piecewise-geodesic interpolation, as in Definition 5.3.

The effective increment law at x is $(\sigma(x))_* \mu_x$, with log-moment generating function $\Lambda_x(\sigma(x)^* \cdot)$. It is the tractable half of the position-dependent regime: the position dependence is one smooth, linear, invertible factor, and the family underneath keeps every transport identity. Note that the undrifted case $V \equiv 0$ with $\sigma \not\equiv \text{id}$ is already new.

The drift map generalises by inserting σ at the image point,

$$\dot{\phi}(t) = V(\phi(t)) + \sigma(\phi(t)) P_{\psi(t) \rightarrow \phi(t)} \dot{\psi}(t), \quad \phi(0) = x_0, \quad (30)$$

with P the transport along the minimising geodesic as in (9). Call this Φ_σ ; for $\sigma \equiv \text{id}$ it is the drift map Φ of Section 8. Inputs of speed at most r now produce outputs of speed at most $\|V\|_\infty + \|\sigma\|_\infty r$, so the tube inflates accordingly.

10.1 The expected theorems

Conjecture 10.2 (Bounded increments). *Let (M, g) be complete, let $\{\mu_x\}$ be consistent with $\text{supp}(\mu_x) \subseteq \overline{B}(0, r)$, let V and σ be as in Definition 10.1, and assume a separation condition of the form*

$$\|V\|_\infty + \|\sigma - \text{id}\|_\infty r < \text{inj}(\text{tube}). \quad (31)$$

Then $\hat{Y}^{n,\sigma}$ satisfies the large deviation principle on $(C_{x_0}([0, 1]; M), d_\infty)$ at speed n with good rate function

$$I_{\text{path}}^{V,\sigma}(\phi) = \int_0^1 \Lambda_{\phi(t)}^* \left(\sigma(\phi(t))^{-1} (\dot{\phi}(t) - V(\phi(t))) \right) dt \quad (32)$$

for absolutely continuous ϕ started at x_0 and $+\infty$ otherwise, and the endpoint satisfies the large deviation principle in M with the contracted rate.

For $\sigma \equiv \text{id}$, (31) is the standing assumption (11) and the conjecture is Theorem 6.2(ii). The separation condition is not decoration: it is what keeps the transport in (30) well defined, for the reason given in Subsection 10.3. It is also a real restriction, and one could look into a Cramér-level statement of this type.

Conjecture 10.3 (Cramér increments). *Let (M, g) be complete with bounded geometry, let $\{\mu_x\}$ be consistent and satisfy Cramér’s condition, and let V and σ be as in Definition 10.1. Then the conclusions of Conjecture 10.2 hold with no support bound and no separation condition.*

No smallness condition on $\|V\|_\infty$ appears, because any such condition inflates along the truncation tower and becomes unusable; the conjecture is stated in the form a transport without an injectivity constraint would deliver, which would also remove $\|V\|_\infty < \text{inj}(M)$ from Theorem 6.5. Whether there is such a transport is open, and Subsection 10.3 says where it is needed. And the bounds on σ , $\nabla\sigma$ and σ^{-1} are global, playing the role that boundedness of V plays in Theorem 6.5: the walk explores arbitrarily large compacta, so coefficient bounds have to be uniform. There is no intermediate statement between the two conjectures. The guarded version with minimising-geodesic transport exists only at bounded increments.

10.2 Freidlin–Wentzell

The integrand of (32) is a Lagrangian $\mathcal{L}_\sigma(x, v) = \Lambda_x^*(\sigma(x)^{-1}(v - V(x)))$, whose Legendre transform in v is

$$\mathcal{H}_\sigma(x, p) = \langle V(x), p \rangle + \Lambda_x(\sigma(x)^* p)$$

by the substitution $v = V(x) + \sigma(x)u$. So the conjectures say, in words, that Cramér’s theorem holds for the effective non-consistent increment law, with the rate integrand one would have guessed. Where the drift perturbs $\Lambda_x(p)$ linearly in p , the coefficient σ deforms it inside the argument: convexity in p survives, genuine x -dependence appears.

For isotropic Gaussian increments $\mu_x = \mathcal{N}(0, s^2 \text{id})$, writing $a := s^2 \sigma \sigma^*$ for the effective diffusion tensor, (32) becomes

$$I_{\text{path}}^{V, \sigma}(\phi) = \frac{1}{2} \int_0^1 \langle \dot{\phi} - V(\phi), a(\phi)^{-1}(\dot{\phi} - V(\phi)) \rangle dt. \quad (33)$$

This is the Freidlin–Wentzell action of the small-noise diffusion $dX = V(X) dt + \sqrt{1/n} s \sigma(X) \circ dB^M$ [4, Chapter 5, eq. (3.5)]. Conjecture 10.3 is therefore a Freidlin–Wentzell theorem for the geodesic-walk discretisation, with the quadratic form replaced by an arbitrary Legendre transform.

10.3 What survives, and what does not

The intended proof is the proof of Theorem 6.5: contraction through the drift map at bounded increments, then the truncation tower. Going through the components, almost all of them are indifferent to σ , in the sense that the statement and proof survive after substituting constants. The exceptions are two.

Well-posedness of Φ_σ . Unchanged. In a chart the right-hand side of (30) is still measurable in t , still linear in the input velocity and still smooth in the state, with an integrable majorant and an integrable Lipschitz rate, which is all the Carathéodory–Picard theorem (Theorem 3.18) asks for.

Continuity of Φ_σ . The mechanism survives, and it asks less of σ than the other components do. Continuity of Φ is proved in Subsection 8.2.3 by compactness rather than by an estimate: the outputs lie in a d_∞ -compact class (Proposition 8.8), so it is enough to identify every uniform limit point of $\Phi(\psi_n)$, and that is done by passing to the limit in the chart form of the drift ODE. Only two features of the chart coefficient $A(y, x)$ are used there. It is linear in the velocity slot, so the input needs no more than the weak- L^1 velocity convergence that uniform convergence already supplies

(Lemma 3.14); and it is continuous in the state slots, so it converges uniformly and Lemma 3.15 pairs the two against each other. Inserting σ replaces A by $(d\varphi)_y \circ \sigma \circ P_{x \rightarrow y} \circ ((d\varphi)_x)^{-1}$ with σ at the output point, which is still linear in the velocity slot, since σ takes values in $\text{End}(TM)$, and still continuous in the state slots, since σ is smooth. Both features are intact, so the argument goes through with the output class inflated to $\text{AC}_{x_0, \|V\|_\infty + \|\sigma\|_\infty r}$ and the source region governed by the separation bound below. Two things are worth recording. Continuity of σ is all that is used, not $\|\nabla\sigma\|_\infty$, because the weak-convergence route never differentiates the coefficient in time, which is what an estimate-based proof would need. And the argument still lives on a fixed speed class, since $\text{AC}_{x_0, r}$ is compact only for finite r , so σ buys no shortcut past the tower. The explicit inverse of the drift map becomes $\dot{\psi} = P_{\phi \rightarrow \psi} \sigma(\phi)^{-1}(\dot{\phi} - V(\phi))$, and uniform ellipticity is what keeps its image in a bounded-speed class; transport invariance of the integrand, which uses only consistency of $\{\mu_x\}$ and so does not see σ , then makes (32) the right contraction.

Exponential equivalence, and Layers 1 and 2 of the tower. Unchanged up to constants. The new discrepancies all have the shape “ σ at one point against σ at another”, controlled by $\|\nabla\sigma\|_\infty$ times a distance, and they enter the Grönwall bookkeeping exactly where the corresponding drift and curvature terms already do, telescoping into the same total-variation clock. Layer 1 never mentions the drift map at all. In Layer 2 the level- c integrand becomes $\Lambda^{(c)*}(A(x)v - b(x))$ with $A = \sigma^{-1}$ smooth and uniformly invertible and $b = \sigma^{-1}V$ smooth and bounded, which is the shape the convex analysis there already handles; uniform ellipticity supplies the coercivity in v that the superlinearity argument needs.

The a priori separation bound. Here a new hypothesis appears. Under $\sigma \equiv \text{id}$ the transported input velocity moves source and image in parallel, so only the drift separates them and $d(\phi(t), \psi(t)) \leq \|V\|_\infty t$. With σ the first-variation estimate leaves an extra term $\|\sigma - \text{id}\|_\infty \|\dot{\psi}\|$, so on $\text{AC}_{x_0, r}$ the separation is only bounded by $(\|V\|_\infty + \|\sigma - \text{id}\|_\infty r) t$. Minimising-geodesic transport in (30) therefore needs the inflated condition (31). At bounded increments this is a hypothesis one can simply assume, and it is the only new restriction there.

Layer 0 of the tower, where that hypothesis breaks. Layer 0 applies the fixed-scale theorem at truncation level c , that is with $r = c \rightarrow \infty$, and (31) then fails for every large c unless $\sigma \equiv \text{id}$. The tower cannot be set up at all. This is the one structural obstruction, and it sits entirely in the choice of transport in (30).

Past this point we have no proof strategy. A fix has to replace the minimising-geodesic transport by one that carries no injectivity constraint, and transport along the path history is the obvious candidate; the thesis already uses such a device once, as the fallback branch of the Layer 1 coupling (26). Whether the drift map stays continuous under it we do not know. That is what Conjecture 10.3 rests on, and we leave it open.

References

- [1] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Universitext, Springer, New York, 2011.
- [2] J. Nash, *The imbedding problem for Riemannian manifolds*, Ann. Math. **63** (1956), 20–63.
- [3] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer, 1998.
- [4] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, 3rd ed., Springer, 2012.
- [5] E. Jørgensen, *The central limit problem for geodesic random walks*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **32** (1975), 1–64.
- [6] R. C. Kraaij, F. Redig, and R. Versendaal, *Classical large deviation theorems on complete Riemannian manifolds*, Stochastic Process. Appl. **129** (2019), 4294–4334.
- [7] J. M. Lee, *Introduction to Riemannian Manifolds*, 2nd ed., Graduate Texts in Mathematics, vol. 176, Springer, 2018.
- [8] J. M. Lee, *Introduction to Smooth Manifolds*, 2nd ed., Graduate Texts in Mathematics, vol. 218, Springer, 2013.
- [9] Y. T. Lee and S. S. Vempala, *Geodesic walks in polytopes*, in Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2017), ACM, 2017, pp. 927–940.
- [10] O. Mangoubi and A. Smith, *Rapid mixing of geodesic walks on manifolds with positive curvature*, Ann. Appl. Probab. **28** (2018), 2501–2543.
- [11] X. Pennec, P. Fillard, and N. Ayache, *A Riemannian framework for tensor computing*, Int. J. Comput. Vis. **66** (2006), 41–66.
- [12] R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series, vol. 28, Princeton University Press, Princeton, 1970.
- [13] R. Versendaal, *Large deviations for geodesic random walks*, Electron. J. Probab. **24** (2019), Paper No. 93, 39 pp.
- [14] R. Versendaal, *Large deviations for stochastic processes on Riemannian manifolds*, PhD thesis, Delft University of Technology, 2020.
- [15] T. Sakai, *Riemannian Geometry*, AMS Translations of Mathematical Monographs 149, 1996.
- [16] W. Walter, *Ordinary Differential Equations*, Springer Graduate Texts in Mathematics 182, 1998.
- [17] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*, Graduate Studies in Mathematics, vol. 140, American Mathematical Society, 2012.