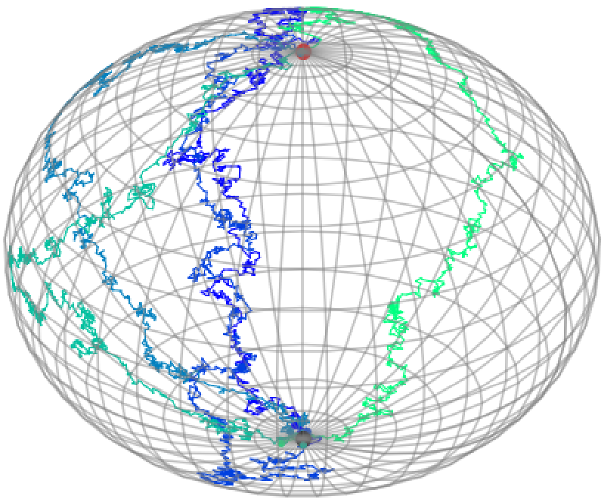
UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE



PhD Thesis

Mathias Højgaard Jensen

Stochastic Bridge Sampling on Manifolds

Department of Computer Science

Advisor: Stefan Sommer

Handed in: May 22, 2022

University of Copenhagen
Faculty of Science
Department of Computer Science

Stochastic Bridge Sampling on Manifolds

Mathias Højgaard Jensen

Advisor: Stefan Sommer

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen, on November 30, 2021

Abstract

Over the past two decades, simulation of conditioned diffusion processes has been investigated actively, while the field of geometric statistics has emerged due to the growing need for handling non-linear data within the fields like robotics, biology, and medical image analysis, among others. Simulation of conditioned diffusions is a challenging statistical problem as analytically closed-form expressions for the transition densities of the desired distribution are often unavailable. Much of the existing literature provides methods to simulate conditioned diffusion processes in a non-geometric setting, and the intersection between these two fields is relatively unexplored. This dissertation is devoted to studying the simulation of guided bridge processes on Riemannian manifolds, Lie groups, and homogeneous spaces and their applications.

We propose a simple method for simulating conditioned processes on various non-linear spaces and show its application to likelihood inference for diffusion processes in different geometric settings.

In the first paper of the thesis, we present a bridge simulation scheme on the orthonormal frame bundle of Riemannian manifolds. We derive an expression for the transition densities analytically, and we obtain approximations of the transition densities on specific manifolds using Monte Carlo techniques. In particular, for the two-dimensional sphere, we compare the approximation with the truncated exact heat kernel and show how an iterative maximum likelihood approach yields an approximation of the mean value.

In the second paper, we utilize the simulation scheme developed in the first paper to derive a simulation scheme on Lie groups and homogeneous spaces. An iterative maximum likelihood procedure applied to the approximated transition density yields an estimate of the unknown underlying metric on the Lie group and homogeneous space. Moreover, we obtain anisotropic distributions on homogeneous spaces, arising from non-invariant metrics on the total space.

In the third and last paper of the thesis, we consider bridge simulations on Riemannian product manifolds. Given a set of data points on a manifold, we estimate the diffusion mean by simulating a single diffusion bridge on the product manifold conditioned on the diagonal. We verify experimentally the computational efficiency of the forward sampled diffusion mean estimate compared to the Fréchet mean on the space of Landmarks.

Resumé

I løbet af de sidste to årtier er simulering af betingede diffusionsprocesser blevet undersøgt aktivt, mens området for geometrisk statistik er opstået på grund af det voksende behov for håndtering af ikke-lineære data inden for områder som robotik, biologi og medicinsk billedanalyse. Simulering af betinget diffusionsprocesser er et udfordrende statistisk problem, da analytisk kendte udtryk for overgangstæthederne for den ønskede fordeling ofte er utilgængelige. Meget af den eksisterende litteratur giver metoder til at simulere betingede diffusionsprocesser i et ikke-geometrisk miljø, og skæringspunktet mellem disse to felter er relativt uudforsket. Denne afhandling er viet til at studere simuleringen af guidede broprocesser på Riemannske-mangfoldigheder, Lie-grupper og homogene rum og deres anvendelser.

Vi foreslår en simpel metode til simulering af betingede processer på forskellige ikke-lineære rum og viser dens anvendelse på sandsynlighedsinferens for diffusionsprocesser i forskellige geometriske kontekster.

I afhandlingens første papir præsenterer vi et brosimuleringsskema på rummet bestående af ortonormale basisvektorer over Riemannske-mangfoldigheder. Vi udleder et udtryk for overgangstæthederne analytisk, og vi opnår tilnærmelser af overgangstæthederne på specifikke mangfoldigheder ved hjælp af Monte Carlo-teknikker. Især for den todimensionelle sfære sammenligner vi tilnærmelsen med den trunkerede eksakte varmekerne og viser, hvordan en iterativ maksimum likelihood tilgang giver en tilnærmelse af middelværdien.

I det andet papir bruger vi simuleringsskemaet udviklet i det første papir til at udlede et simuleringsskema på Lie-grupper og homogene rum. En iterativ procedure for maksimum likelihood anvendt på den approksimerede overgangstæthed giver et estimat af den ukendte underliggende metrik på Lie-gruppen og det homogene rum. Desuden opnår vi anisotrope fordelinger på homogene rum, der stammer fra ikke-invariante metrikker på det samlede rum.

I specialets tredje og sidste artikel behandler vi brosimuleringer på Riemannske produktmangfoldigheder. Givet et sæt datapunkter på en mangfoldighed estimerer vi diffusionsgennemsnittet ved at simulere en enkelt diffusionsbro på produktmangfoldigheden betinget på diagonalen. Vi verificerer eksperimentelt den beregningsmæssige effektivitet af det fremadsamplede diffusionsmiddelestimat sammenlignet med Fréchet-gennemsnittet på landmark rummet.

Acknowledgements

This Ph.D. would not have been possible had it not been for CSGB, the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum foundation. Many colleagues from CSGB I am proud to call my friends today. I want to especially thank Mads for his constant encouragement. Thank you, Anton, Aasa, and Tom. I would not have known of the existence of the image section without you guys.

I want to thank all my colleagues in the Image section. It has always been a pleasure to come to work. A special thanks go to my office mates Line, Pernille, and Lennard. I always enjoyed discussing ideas with you guys. Mainly, I want to thank Line for welcoming me to the group. Unfortunately, we only shared a few months together, but those months were some of the best. Lennard, thank you for always being available.

At last, I would like to thank all of my friends, family, and my girlfriend, for their understanding and support. I am looking forward to seeing you a lot more in the future. And by that, I mean outside of the office, Faezeh.

List of Publications

The following list of publications and manuscripts presents the work conducted over the last three years during my enrollment as a PhD. student.

Chapter 2 Mathias Højgaard Jensen, Stefan Sommer. "*Simulation of Conditioned Semimartingales on Riemannian Manifolds*" *arXiv:2105.13190*, (2022).

Chapter 3 Mathias Højgaard Jensen, Lennard Hilgendorf, Sarang Joshi, Stefan Sommer. "*Bridge Simulation and Metric Estimation on Lie Groups and Homogeneous Spaces*" *arXiv:2106.03431*, (2022).

Chapter 4 Mathias Højgaard Jensen, Stefan Sommer. "*Diffusion Mean Estimation on the Diagonal of Product Manifolds*" *Algorithms*, 2022, Vol.15 (3), p.92.

Appendix A Mathias Højgaard Jensen, Anton Mallasto, Stefan Sommer. "*Simulation of Conditioned Diffusions on the Flat Torus*" *Geometric Science of Information* (GSI) (2019).

Appendix B Mathias Højgaard Jensen, Sarang Joshi, Stefan Sommer. "*Bridge Simulation and Metric Estimation on Lie Groups*" *Geometric Science of Information* (GSI) (2021).

Contents

Chapter 1

Introduction and Background

This dissertation comprises the work conducted over the last three years during the author's enrollment as a Ph.D. student at the Department of Computer Science at the University of Copenhagen. The dissertation contains the following manuscripts

- Mathias Højgaard Jensen, Stefan Sommer. "*Simulation of Conditioned Semimartingales on Riemannian Manifolds*" *arXiv:2105.13190*, (2022).
- Mathias Højgaard Jensen, Lennard Hilgendorf, Sarang Joshi, Stefan Sommer. "*Bridge Simulation and Metric Estimation on Lie Groups and Homogeneous Spaces*" *arXiv:2106.03431*, (2022).
- Mathias Højgaard Jensen, Stefan Sommer. "*Diffusion Mean Estimation on the Diagonal of Product Manifolds*" *Algorithms*, 2022, Vol.15 (3), p.92.

presented in Chapters 2-4. The two former manuscripts are available as preprints on ArXiv, while the latter article has been published in the special issue *Stochastic Algorithms and Their Applications* of the open access journal *Algorithms*. Besides the manuscripts appearing in Chapters 2-4, the author authored two papers accepted at the international conference on Geometric Science of Information (GSI) and published in the conference proceedings of GSI'19 and GSI'21, respectively.

- Mathias Højgaard Jensen, Anton Mallasto, and Stefan Sommer, "*Simulation of Conditioned Diffusions on the Flat Torus*", Geometric Science of Information (GSI), 2019,
- Mathias Højgaard Jensen, Sarang Joshi, and Stefan Sommer, "*Bridge Simulation and Metric Estimation on Lie Groups*", Geometric Science of Information (GSI), 2021,

The two published conference articles were presented at the GSI'19 and GSI'21 conferences, respectively. The content of "*Simulation of Conditioned Diffusions on the Flat Torus*" appears implicit in the manuscript of Chapter 2. In contrast, the content of "*Bridge Simulation and Metric Estimation on Lie Groups*" explicitly features in the manuscript of Chapter 3. For these reasons, the dissertation omits the two articles, however, the articles can be found in Appendix A and B.

The Ph.D. project aims to develop statistical methodologies to infer the distributional properties of observations in geometrical spaces such as Riemannian manifolds, Lie groups, and quotient spaces. The following sections review the mathematical background theory and methodology used in the manuscripts.

1.1 Structure of Thesis

This thesis is structured as follows.

Chapter 1 briefly introduces the different fields appearing throughout the thesis. Section 1.2 describes the problem formulation and recent advances within bridge simulation on non-linear spaces. In Section 1.3, we briefly introduce geometric statistics and establish the role of bridge simulation within this field. Section 1.4 gives a preliminary introduction to the theory of Riemannian geometry. In Section 1.5, we briefly introduce stochastic processes and the theory of stochastic calculus, including definitions of semimartingales, Itô integrals, stochastic differential equations, and Stratonovich integrals. Section 1.6 presents stochastic differential equations on manifolds by using the stochastic development approach. Lastly, we introduce bridge processes through guiding in Euclidean spaces and on manifolds in Section 1.7.

Chapter 2-4 contains the central part of the thesis. Each chapter is a separate paper and can therefore be read independently of one another. As the topics in each chapter are very related, it makes the chapters very repetitive.

The last chapter summarizes the thesis and provides an overview of further investigations in future works.

1.2 Problem Formulation

This thesis is devoted to studying conditioned diffusion processes in the context of statistical inference for continuous-time stochastic processes in the presence of curvature. While the statistical inference for diffusion processes has attracted much attention over the last two decades, the presence of curvature is a new and relatively unexplored area in this field.

1.2.1 Why Bridge Processes?

Diffusion bridge simulation lies at the heart of statistical inference for discretely observed diffusion processes. It plays a fundamental role in likelihood and Bayesian inference for diffusion processes. As the name suggests, diffusion bridges link points together, thereby creating a bridge between them. The diffusion bridges serve as the missing observations of continuous diffusion processes. Given discrete observations of a continuous diffusion process, the statistical inference problem can be regarded as a missing data problem. The missing data are the continuous paths between the observations. Had the entire paths of the diffusion process been observed, Girsanov's formula would have provided the likelihood functions. Therefore, the theory of conditioned diffusion processes naturally appears in the context of statistical inference.

Simulating conditioned diffusion processes is a challenging task, even in Euclidean spaces. Conditioning the continuous diffusion process on arriving at a pre-specified point at a pre-determined time means conditioning the process on a set of probability zero. Hence, giving a precise mathematical definition to this problem is challenging, as the usual definition of conditioned probabilities, Bayes' formula, does not apply. To emphasize the challenges in simulating conditioned diffusion processes, the transition density related to a diffusion process, which describes the probability mass and its change in time and space, is often intractable. The intractable transition densities make simulation from the desired conditioned diffusion troublesome. Proposing methods for simulating conditioned diffusion processes has been one of the main concerns in the field during the last two decades and is still an active field. Not only does the proposed method need to converge to the desired point, but the method also needs to converge at the right speed and have a likelihood that is easy to handle.

1.2.2 The Geometric Challenges

The presence of curvature introduces various new challenges. For one, numerical approximations of stochastic differential equations need to take the curvature into account. Simulating stochastic differential equations on smooth non-linear spaces like smooth manifolds can be done in various ways. For any smooth curve on a smooth manifold, we expect that its derivative is tangential to the manifold along its path. If this is not the case, the curve will escape to the ambient space. The same holds for a manifold valued diffusion. Integrating vector fields along the path with respect to Euclidean valued diffusion processes should produce a manifold-valued diffusion process. That is indeed true. However, the classical formulation describes stochastic differential equations in terms of Itô integrals. Itô integrals are integrals where the integrator are stochastic processes. Such integrals do not obey the fundamental theorem of calculus. This fact follows from the limiting sum approximation of the integrals, where the integrand evaluates at the left-most endpoint. As such, Itô integrals are not well suited to handle stochastic differential equations in a manifold setting.

The Stratonovich integral is a variation of the Itô integral, designed such that it follows the fundamental theorem of calculus. The advantage of the Stratonovich integral is that it transforms consistently with respect to diffeomorphisms. Smooth manifolds can be parameterized using local coordinate charts, which map local neighborhoods of the manifold to open subsets of Euclidean space. Since diffeomorphisms patch together such local coordinate charts, the Stratonovich integral is the natural candidate for introducing stochastic differential equations on smooth manifolds. If a single chart can cover the manifold, the Itô integrals can be used as no local patching is required. However, this is rarely the case. In the classical Euclidean sense, describing stochastic differential equations in coordinate charts would require extensive use of local patching between charts.

Embedding the manifold in a high-dimensional Euclidean space offers an alternative approach to simulating stochastic processes on manifolds. Whitney's embedding theorem ensures that embedding smooth manifolds is always possible. Simulating a stochastic process in the Euclidean embedding space and projecting the infinitesimal increments of the Euclidean process orthogonally onto the tangent space results in a stochastic process on the manifold. This method depends on the embedding, and different embeddings result in different projections.

A third option exists, which depends only on the chosen starting frame for the tangent space. A frame at a point on the manifold is an ordered basis for the tangent space at that point. The method is known as the Eells-Elworthy-Malliavin construction, or in colloquial terms as "rolling without slipping," and introduces stochastic differential equations in the frame bundle. The Eells-Elworthy-Malliavin construction provides a stochastic process in the frame bundle, which carries a frame along its path in the manifold. The construction is intrinsic and gives a one-to-one map between Euclidean and manifold-valued processes.

Under suitable conditions on the drift and diffusion coefficient, e.g., Lipschitz continuity and bounded growth, Euclidean valued process has long-term existence. Introducing curvature does not ensure the long-term existence of even simple processes like the Brownian motion. In particular, the Brownian motion on a manifold may explode in finite time, which means that the process will escape all compact subsets of the manifold.

Compact manifolds ensure the non-explosion of a Brownian motion. However, compact manifolds fail to be diffeomorphic to non-compact manifolds, as a topological consequence of the diffeomorphism. They will only be locally diffeomorphic to a subset of the tangent space. The maximal subset of the manifold diffeomorphic to a subset of the tangent space covers the entire manifold except for a closed subset. This subset is called the cut locus. Any point of a compact manifold has a cut locus. The distance function, measuring the distance from a fixed point, is smooth on the subset away from the fixed point and its cut locus. Thus, one can apply Itô's fundamental theorem of stochastic calculus to this function, thereby obtaining the radial process's one-dimensional stochastic differential equation. It is possible to extend the stochastic differential equation for the radial process to the entire manifold for relatively general stochastic processes on a manifold [5, 12, 31, 32, 48, 49]. However, by doing so on a compact manifold, one introduces two new processes: the local time at zero and the geometric local time, with support at the fixed point and on its cut locus, respectively.

A characteristic property of bridge processes is that they converge almost surely to the endpoint. Showing the convergence typically requires a stochastic differential equation for the radial process of the proposed bridge. Therefore, the two local time processes need to be dealt with on any compact manifold.

1.2.3 Recent Advances in Geometric Bridge Simulation

From the viewpoint of statistical inference for diffusion processes, conditioned diffusion processes on manifolds are a little-studied field. In the context of landmark manifolds, diffusion bridge simulation were studied by Arnaudon et al. [1, 2, 4] and Sommer et al. [44], while Bui et al. [9] studied diffusion bridge simulation on the space of symmetric positive definite matrices. The methods in [1, 2, 9, 44] build on the guiding terms introduced by Delyon and Hu [13], while [4] use guided proposal as introduced by Schauer, van der Meulen, and Zanten [42] for inexact matching.

1.3 Geometric Statistics

Geometric statistics is the sub-field of statistics concerning the analyzes of geometric data. Any data residing in a space without a vector space structure but with a geometric structure are classified as geometric data. Commonly, geometric data have no vector space structure, and thus traditional statistical methodologies do not apply. Examples of geometric data such as diffusion tensor images or shapes of organs are common within the medical image community. The reader can find a detailed introduction to the field of geometric statistics in Pennec et al. [37].

The lack of vector space structure entails that operations such as addition and multiplication are not defined. In geometric statistics, these operations are generalized to non-linear spaces, for example, by exploiting the vector space structure of tangent spaces related to smooth manifolds. The addition operation in Euclidean space $y = x + v$ has the manifold equivalent $y = \mathrm{Exp}_x(v)$, and subtraction $v = x - y$ becomes $y = \mathrm{Log}_x(y)$, where Exp_x and Log_x are maps between a manifold and its tangent space at a point x (more details are presented in Section 1.4).

The most fundamental statistic is arguably the mean. In Euclidean space, the mean point has various defining properties: it is the unique point where the sum of the residuals equals zero. It is the unique point that minimizes the variance. Moreover, it is the maximum likelihood estimate of the mean given i.i.d. observations. In contrast, on manifolds, these defining properties generally lead to different points.

Fréchet generalized the Euclidean mean to metric space, exploiting the Euclidean mean's defining property as the unique point that minimizes the sum of squared distances [19]. For any random variable X on a metric space (E, d), the mean set defined by

$$\mu = \underset{p \in E}{\arg\min}\, \mathbb{E}\left[d(p, X)^2\right], \tag{1.1}$$

is called the Fréchet mean or Fréchet mean set. The estimator of the Fréchet mean is then naturally given by

$$\hat{\mu} = \underset{p \in E}{\arg\min}\, \frac{1}{n} \sum_{i=1}^{n} d(p, x_i)^2, \tag{1.2}$$

for observations $x_1, \ldots, x_n$. In general, the Fréchet mean is not unique. A simple example of a non-unique Fréchet mean is the case of the sphere $\mathbb{S}^2$ when the anti-podal point of the mean $\mathrm{Cut}(\mu)$ has positive point mass. In such cases, the Fréchet mean is a circle.

The diffusion-t-mean (diffusion mean for short) [22, 24] offers an alternative to the Fréchet mean. The diffusion mean is defined as the maximum likelihood estimate

$$\mu_t = \underset{y \in M}{\arg\min}\, \mathbb{E}\left[-\log p_t(X, y)\right], \tag{1.3}$$

for $t > 0$, where p denotes the transition density of a diffusion process. The additional time parameter has the interpretation as the variance of the diffusion mean, and the diffusion-t-mean is the most likely origin of a Brownian motion at time t. The relation between the Fréchet mean and the diffusion mean is clear from the asymptotic relation $\lim_{t \downarrow 0} 2t \log p_t(x, y) = -d(x, y)^2$, whenever x and y are not in each other's cut locus [25].

The maximum likelihood definition of the diffusion mean requires likelihood inference, unlike the Fréchet mean. Given discrete observations, the likelihood inference problem can be regarded as a missing data problem. Therefore, a natural assumption throughout the thesis is that the geometric data considered arises as missing observations of observed diffusion process endpoints. Given their fundamental role in likelihood and Bayesian inference for missing data of diffusion processes, it seems natural that simulation of diffusion bridge processes should play a significant role when inferring properties of the underlying probability distribution.

1.4 Riemannian Geometry

This section is primarily based on [10, 33, 37], but the material presented here can be found in most textbooks on differential geometry.

1.4.1 Riemannian Manifolds

A Riemannian manifold (M, g) is a pair consisting of a d-dimensional smooth manifold M endowed with a Riemannian metric $g_p \colon T_pM \times T_pM \to \mathbb{R}$, providing an inner product on each tangent space T_pM that varies smoothly over points $p \in M$. A chart (U, ϕ) around $p \in M$ is a local neighborhood around p where $\phi \colon U \subseteq M \to \mathbb{R}^d$ defines a homeomorphism. A countable collection of charts $\{(U_\alpha, \phi_\alpha)\}_\alpha$, such that the neighborhoods provides a cover of M, $\cup_\alpha U_\alpha = M$, and the compositions $\phi_\alpha \circ \phi_\beta^{-1}$ on non-empty overlaps $U_\alpha \cap U_\beta \neq \emptyset$ are diffeomorphisms, is called an atlas. The atlas parametrizes the manifold and thus M is said to be locally Euclidean. In particular, any point $x \in M$ in a coordinate chart (U, ϕ) can be described in Euclidean coordinates, i.e., $\phi(x) = (x_1, \dots, x_d) \in \mathbb{R}^d$. Throughout, there will be made no distinction between between (U, ψ), U, and ψ when referring to a chart and the distinction will be clear from the context.

1.4.2 Connection

Let $\Gamma(TM)$ denote the set of smooth vector fields on M. A manifold can be endowed with an affine connection $\nabla \colon \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$ which, loosely speaking, provides a way to differentiate vector fields along curves, see e.g. do Carmo [10, Chapter 2]. If X and Y are two smooth vector fields on M, then $(X, Y) \overset{\nabla}{\to} \nabla_X Y$, and the connection moreover satisfies

a) $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$.

b) $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$.

c) $\nabla_X(fY) = f\nabla_X Y + X(f)Y$,

where $X, Y, Z \in \Gamma(TM)$ and $f, g \in C^\infty(M)$. In a local coordinate chart $(x_1, \dots, x_d)$, we can write a vector field, using Einstein summation, as $X = x^i\partial_i := \sum_i x^i\partial_i$, where the $\partial_i = \partial/\partial x_i$ constitute a basis for the tangent space at x. If $Y = y^i\partial_i$ is another vector field, then

$$\nabla_X Y = Xb^j\partial_j + a^ib^j\nabla_{\partial_i}\partial_j = \big(Xb^k + a^ib^j\Gamma^k_{ij}\big)\partial_k,$$

where $\Gamma_{ij}^k \partial_k := \nabla_{\partial_i} \partial_j$ are the Christoffel symbols of the connection ∇, with respect to the chosen chart. Let $\gamma\colon (-\varepsilon, \varepsilon) \to M$ be a smooth curve on M, and let $Y = Y(\gamma(t))$ be a vector field along γ. The covariant derivative of Y along $\gamma(t)$ is

$$\frac{DY}{dt} = \nabla_{\dot\gamma(t)} Y = \left(\partial_{\dot\gamma} y^k + y^i \dot\gamma^j \Gamma_{ij}\right) \partial_k,$$

where $\dot\gamma(t) = d\gamma/dt$ is the time derivative of the curve. The vector field Y is said to be parallel along γ if $\nabla_{\dot\gamma(t)} Y = 0$. A geodesic is a curve whose tangent vector remains parellel to itself, i.e., $\nabla_{\dot\gamma(t)} \dot\gamma(t) = 0$. From the above, we then see that a curve γ is a geodesic if it satisfies the second order equation

$$\ddot\gamma^k + \dot\gamma^i \dot\gamma^j \Gamma_{ij}^k = 0. \tag{1.4}$$

Geodesic curves are locally length minimizing curves. They are the manifold equivalent of straight lines. Indeed, any curve satisfying the equation above have zero acceleration.

In any chart around $x \in M$, we obtain coordinate basis vectors for the tangent space given by $\partial_1, \ldots, \partial_d$. With these basis vectors at hand, the matrix coefficients of the Riemannian metric can be evaluated by considering $g_{ij}(x) = g_x(\partial_i, \partial_j) = \langle \partial_i, \partial_j \rangle$. More generally, the Riemannian metric tensor G_x can be defined as $g_x(v, w) = v^T G_x w$, for every $v, w \in T_x M$, and we can define a norm at each tangent space $T_x M$ by

$$\|v\|_x^2 = v^T G_x v, \tag{1.5}$$

identifying each tangent space with a Euclidean space. For any manifold, there are many connections, but when endowing the manifold with a specific metric, one choice of connection seems more natural. The Levi-Civita connection is the unique connection that preserves the metric and is torsion-free. The Levi-Civita connection is the natural choice in many aspects of Riemannian geometry. Note that there is a whole family of natural connections in the particular case of Lie groups [38].

1.4.3 Exponential Map

One particular choice of chart typically used within geometric statistics is the exponential normal chart related to the exponential map $\mathrm{Exp}_p\colon T_p M \to M$, defined by $\mathrm{Exp}_p(v) = \gamma_1(v)$, where γ is a geodesic. Therefore, the exponential map is intimately related to the geodesic equation (1.4). Locally the exponential map is a diffeomorphism onto its image. For every $p \in M$, there exists a neighborhood $U \subseteq M$ of p such that Exp_p defines a diffeomorphism between $A \subset T_p M$ and $U = \mathrm{Exp}_p(A)$. The largest such U is the open set $M \setminus \mathrm{Cut}(p)$, where the cut locus of p $\mathrm{Cut}(p)$ consists of points where either geodesics from p are no longer unique or points where the first point where the derivative of the exponential map fails to be invertible. The latter points are called the conjugate part of the cut locus. The invertibility of the exponential map implies the existence of an inverse map often referred to as the logarithmic map $\mathrm{Log}_p\colon M\setminus \mathrm{Cut}(p) \to T_pM$. The logarithmic map sends points in $M\setminus \mathrm{Cut}(p)$ to unique tangent vectors.

1.4.4 Riemannian Distance Function

As the Riemannian metric gives a notion of distances of tangent vectors, the logarithm map allows for a definition of the Riemannian distance function given as the norm (with respect to g) of the logarithmic map, i.e.,

$$d(p,q)^2 = \|\mathrm{Log}_p(q)\|_p^2. \tag{1.6}$$

The defintion given is symmetric, i.e., $d(p,q) = d(q,p)$, but the definition in (1.6) depends on the logarithm map at p and the Riemannian metric at p. The Riemannian distance in (1.6) fixes the point p. To compute the distance between two new points requires computing the logarithm map at one of those points and using the inner product at that tangent space. It is therefore natural to think of (1.6) as the distance from p. This is called the radial distance from p.

In the exponential normal chart centered at p, the point $q \in M\setminus \mathrm{Cut}(p)$ can be described using normal coordinates $(x^1,\dots,x^d) \in \mathbb{R}^d$ as $q = \mathrm{Exp}_p(x^i e_i)$, where $(e_1,\dots,e_d)$ is an orthonormal basis of T_pM. The radial distance function $r_p(q) := d(p,q)$ in normal coordinates then have the simple form

$$r_p(q) = \left(\sum_{i=1}^{d} \left(x^i\right)^2\right)^{1/2}. \tag{1.7}$$

The corresponding unit radial vector field is given as $\nabla r_p(q) = \left(\frac{\partial}{\partial r_p}\right)_q = \frac{x^i}{r_p(q)}(\partial_i)_q$. Here we use the Einstein summation convention - a convention that will appear throughout the dissertation.

The negative radial vector field $-(\partial/\partial r_p)_q$ at q is the unit length vector pointing in the direction of p. A stochastic process with the negative radial vector field as its drift will be pushed in the direction of p. It will therefore serve as the manifold equivalent of the guiding term used in Delyon and Hu [13] (see Section1.7). The radial process and its derivative will be the focal term in our guided bridges on manifolds.

1.4.5 Lie Groups and Homogeneous Spaces

In Chapter 2, we study bridge processes on Lie groups and homogeneous spaces. This section briefly reviews some elementary facts about Lie groups.

A more extensive, but compact, introduction is present in Chapter 2. Lie groups are a class of manifolds endowed with a group structure. For a Lie group G, the multiplication operations $G \times G \to G$, $(x, y) \overset{\mu}{\mapsto} xy$ and the inverse operation $G \to G$, $x \mapsto x^{-1}$ are smooth maps. For any $x \in G$, the left-multiplication map $L_x y$ defined by $y \mapsto \mu(x, y)$ is a diffeomorphism from G to itself. The pushforward map of L_x is a map between tangent bundles TG. In particular, the pushforward map of L_x at some point $y \in G$ determines a map $(dL_x)_y \colon T_y G \to T_{xy} G$. The pushforward of the left-multiplication map determines a class of vector fields on G called left-invariant vector fields. A vector field V is left-invariant if $(dL_x)_y V(y) = V(xy)$. The set of left-invariant vector fields are fundamental in Lie group theory. Obviously, a Lie group is a group. Hence there exists an identity element $e \in G$. The tangent space at the identity $T_e G$ equipped with the operation $(x, y) \mapsto [x, y] = xy - yx$, for $x, y \in T_e G$ (Lie bracket) is called the Lie algebra of G. The set of left-invariant vector fields is isomorphic to the tangent space at the identity. As such, every Lie group is parallelizable, i.e., for any basis $(v_1, \ldots, v_d)$ of $T_e G$ the set $(d(L_x)v_1, \ldots, d(L_x)v_d)$ is a basis for the tangent space $T_x G$, for any $x \in G$. Therefore, a vector field on G can be specified from a tangent vector in the Lie algebra using the left-pushforward map.

We denote by $\langle \cdot, \cdot \rangle$ a Riemannian metric on G. A Riemannian metric on a Lie group is said to be left-invariant (resp. right-invariant) if $\langle u, v \rangle_y = \langle (dL_x)_y u, (dL_x)_y v \rangle_{L_x(y)}$ (resp. $\langle u, v \rangle_y = \langle (dL_x)_y u, (dL_x)_y v \rangle_{L_x(y)}$), for every $u, v \in T_y G$, i.e., the left-(resp. right-)multiplication maps are isometries, for every $x \in G$. If the metric is both right- and left-invariant, then the metric is called bi-invariant.

A one-parameter subgroup of G is a continuous homomorphism $\gamma \colon (\mathbb{R}, +) \to G$. The Lie group exponential map $\exp \colon \mathfrak{g} \to G$ is defined as $\exp(v) = \gamma_v(1)$, for $v \in \mathcal{G}$, where γ_v is the unique one-parameter subgroup of G whose tangent vector at e is v.

A homogeneous space is a specific type of quotient space where the Lie group acts transitively on the homogeneous space. The homogeneous space is also referred to as the base or bottom space and the Lie group as the top or total space. All homogeneous spaces can be described as a quotient space G/H, where H is a closed subgroup of the Lie group G. Familiar Riemannian manifolds arise as homogeneous space. For example, the sphere $\mathbb{S}^2$ can be regarded as the quotient manifold $\mathrm{SO}(3)/SO(2)$, where $\mathrm{SO}(n)$ denotes the set of orthogonal rotations in $\mathbb{R}^n$. Moreover, the torus arises as the quotient space $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$.

1.4.6 Frame Bundle Geometry and Horizontal Lift

Let M be a d-dimensional manifold and $p \in M$. A frame at p is an $\mathbb{R}$-linear isomorphism $u \colon \mathbb{R}^n \to T_p M$ mapping basis vectors to basis vectors.

By FM_p we shall denote the space of all frames at p. The disjoint union $FM := \bigcup_{p \in M} FM_p$ is called the frame bundle of which elements are pairs (p, u). The frame bundle is a manifold of dimension $d(d+1)$ and it associates a natural projection map $\pi \colon FM \to M$ defined by $\pi(p, u) = p$. Throughout, we simply refer to elements of FM as u. A connection on FM corresponds to a smooth choice of horizontal subspace $H_u FM$ of the tangent space $T_u FM$, for each $u \in FM$, such that $T_u FM = H_u FM \oplus V_u FM$. The projection π induces an isomorphism $\pi_* \colon H_u FM \to T_{\pi u} M$ from which it follows for any tangent vector $X \in T_p M$ and frame u at p there exists a unique horizontal vector $X^* \in H_u FM$ such that $\pi_* X^* = p$. The horizontal tangent vector X^* is called the horizontal lift of X at u. A curve u_t in FM is said to be horizontal if the vector field $u_t e$, for some $e \in \mathbb{R}^d$, is parallel along the curve $\pi(u_t)$. For any $u \in FM$ any basis $(e_1, \ldots, e_d)$ of $\mathbb{R}$, the vectors $u(e_i)$, $i = 1, \ldots, d$ define a basis of $T_{\pi(u)} M$ resulting in basis for $H_u FM$ given by $H_i(u) := (\pi_*(u(e_i)))^{-1}$. The collection $H_1, \ldots, H_d$ provides fundamental horizontal vector fields on the horizontal part of the frame bundle. Endowing M with a Riemannian metric, we may restrict attention to the set of frames which are orthonormal. Hence, there exists a sub-bundle of the frame bundle consisting of orthonormal frames called the orthonormal frame bundle OM. In this case, elements in the fiber of $\pi \colon OM \to M$, $\pi^{-1}(p)$, differs only by a rotation.

1.5 Stochastic Differential Equations

In probability theory, one always consider a probility space $(\Omega, \mathcal{F}, P)$, which consists of a sample space Ω, a sigma-algebra $\mathcal{F}$ defined on Ω, and a probability measure $P \colon \Omega \to [0, 1]$. A real-valued random variable defined on $(\Omega, \mathcal{F}, P)$ is a measurable map $X \colon \Omega \to \mathbb{R}$. A countable collection of random variables, $(X_n)_{n \in \mathbb{N}}$, is called a sequence of random variables. If we change the indexation from at countable set, say $\mathbb{N}$, to an uncountable set, say $\mathbb{R}_{\geq 0}$, the resulting family, $X := (X_t)_{t \geq 0}$, is called a stochastic process. We will often simply refer to X_t as the stochastic process.

1.5.1 Filtered Probability Space

Throughout this dissertation, we assume that the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ satisfies the usual conditions, i.e., the probability space $(\Omega, \mathcal{F}, P)$ is equipped with a filtration, consisting of a sequence of increasing sub-σ-fields contained in $\mathcal{F}$, where

(a) The probability space $(\Omega, \mathcal{F}, P)$ is a complete measure space;

(b) The filtration is right-continuous, $\mathcal{F}_t = \mathcal{F}_{t+} := \cap_{s > t} \mathcal{F}_s$, for all $t \geq 0$;

(c) For $t \geq 0$, $\mathcal{F}_t$ contains all the P-null sets of $\mathcal{F}$.

The stochastic process X is said to be adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ if the map $X_t\colon \Omega \to \mathbb{R}$ is $\mathcal{F}_t$ measurable for every $t \geq 0$. The process is said to be continuous if each of its sample paths $X_t(\omega) = \omega_t$ are continuous. We will only be considering continuous processes throughout this dissertation.

1.5.2 Semimartingales

A martingale is an adapted stochastic process, X, which is integrable, for each $t \geq 0$, and satisfies the martingale property $X_s = E[X_t|\mathcal{F}_s]$. The class of martingales can be generalized to the class of local-martingales. A process X is a local-martingale if there exists a sequence of non-decreasing stopping times, i.e., a sequence of random times $\tau_n\colon \Omega \to \mathbb{R}_{\geq 0}$ such that $\{\tau \leq t\}$ is $\mathcal{F}_t$-measurable, for which the stopped process $X^{\tau_n} = (X_{t\wedge\tau_n})_{t\geq 0}$ is a martingale, for each n. Here $t \wedge \tau = \min(t, \tau)$. The total variation of f on $[0, t]$ is defined as

$$|f_t| = \sup \sum_{k=1}^{n} \left|f_{t_k} - f_{t_{k-1}}\right|,$$

where the supremum is taken over all finite partitions $0 = t_0 < t_1 < ... < t_n = t$ of the interval $[0, t]$. f has bounded variation on $[0, t]$ if $|f_t| < \infty$. A stochastic process A has finite variation if every path $\omega \mapsto A_t(\omega)$ has bounded variation almost surely on all compact subintervals of $[0, \infty)$. A *semimartingale* is a stochastic process that can be decomposed into a sum $X = X_0 + N + A$, where X_0 is an $\mathcal{F}_0$ measurable random variable, N a local martingale with $N_0 = 0$ and A a finite variation process with $A_0 = 0$.

1.5.3 Stochastic Integrals and Differential Equations

Semimartingales typically arise as solutions to stochastic differential equations of the form

$$dX_t^j = b^j(t, X_t)dt + \sigma_i^j(t, X_t)dB_t^i, \qquad X_0 = x_0, \tag{1.8}$$

where b and σ are suitably integrable maps. Under suitable integrability conditions, the first term on the right-hand side (rhs) is a finite variation process, while the latter term is a local-martingale. The second term on the rhs of (1.8) is an Itô integral. For a semimartingale B, the Itô integral is defined as

$$Y_t^j = \int_0^t \sigma_i^j(t, X_t)dB_t^i,$$

where σ is an adapted process such that $\int_0^t a(s, X_s)d[B]_s < \infty$, almost surely, for all $t \geq 0$, where $a = \sigma\sigma^T$ and $[B]_t$ is a finite variation process. The Itô integral may be expressed in terms of convergence in probability of Riemann sums

$$\sum_{k=1}^{K_n} \sigma_i^j(t_{k-1}, X_{t_{k-1}})(X_{t_k}^i - X_{t_{k-1}}^i) \xrightarrow{P} Y_t^j,$$

where $0 = t_0 < t_1 < \cdots < t_{K_n} = t$ (see e.g. [30] or [39]). Since the Itô integral does not satisfy the fundamental theorem of calculus, it makes the Itô integral inappropriate to use in non-linear spaces. The Stratonovich integral is a stochastic integral that satisfies the fundamental theorem of calculus. Let X and Y be two semimartingales. The Stratonovich integral is defined as

$$\int_0^t X_s \circ dY_s = \int_0^t X_s dY_s + \frac{1}{2}[X, Y]_t, \quad t \geq 0,$$

where $[X, Y]_t$ is the quadratic co-variation defined as the limit in probability

$$\sum_{i=0}^{K_n} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}) \xrightarrow{P} [X, Y]_t.$$

The Stratonovich integral satisfies the fundamental theorem of calculus, which makes the Stratonovich integral the preferred choice when performing stochastic integration on non-linear spaces.

1.6 Stochastic Development

The material presented in this section can be found in textbooks such as [18,26].

1.6.1 Brownian Motion on Manifolds

The Brownian motion is the most well-known and studied stochastic process. The Brownian motion relates to the heat equation in a fundamental way

$$\left(\frac{\partial}{\partial t} - \Delta\right) f = 0, \tag{1.9}$$

where Δ denotes the usual Laplace operator. The solution f to the heat equation (1.9) is the transition density of a Brownian motion. A natural generalization of a Brownian motion to smooth manifolds then arises as to the process with transition density solving the heat equation on M

$$\left(\frac{\partial}{\partial t} - \Delta_M\right) f = 0, \tag{1.10}$$

where Δ_M denotes the Laplace-Beltrami operator - the generalization of the Laplace operator to M. It turns out that a natural geometric definition of the Laplace-Beltrami is as the divergence of the gradient $\Delta_M = \text{div}\nabla f$. From this definition, one can deduce that the Laplace-Beltrami operator coincides with the trace of the Hessian tensor, which is the case for the usual Laplace operator. In a local coordinate chart, the Laplace-Beltrami can be expressed as

$$\Delta_M f = \det(g)^{-1/2}(\partial_j g^{ji} \det(g)^{1/2}\partial_i)f, \tag{1.11}$$

where $\det(g)$ denotes the determinant of the Riemannian metric tensor and $g^{-1} = \{g^{ij}\}_{ij}$ denotes the inverse metric tensor. A simple derivation shows that (1.11) becomes

$$\Delta_M f = g^{ij}\frac{\partial^2}{\partial x_i \partial x_j}f + g^{ik}\Gamma^j_{ik}\frac{\partial}{\partial x_j}f, \tag{1.12}$$

where Γ^k_{ij} denotes the Christoffel symbols of the connection. The Brownian motion on M is a diffusion process with generator $\frac{1}{2}\Delta_M$. Locally, the Brownian motion satisfies the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{1.13}$$

where σ is the metric square root of g^{-1}, $b^k = -\frac{1}{2}g^{ij}\Gamma^k_{ij}$, and W is a d-dimensional Euclidean Brownian motion. The dependency on the chosen chart is clear from the drift and diffusion coefficient in (1.13). From (1.12), we see that locally Δ_M is a Hörmander type second order elliptic operator of the form

$$L = \sum_{i=1}^{d} V_i^2 + V_0, \tag{1.14}$$

for smooth vector fields $V_i, i = 0, \dots, d$. However, (1.12) is not intrinsically defined, and in general, it is not possible to define Δ_M intrinsically in Hörmander form. Indeed, a global definition of a Brownian motion on M in this way is only possible if a single chart can cover M.

1.6.2 Horizontal Semimartingales

An equivalent definition of a Brownian motion on M is a solution to a stochastic differential equation independent of charts. This construction, due to Eells-Elworthy-Malliavin [17], is known as stochastic development. The construction provides a one-to-one correspondence between Euclidean valued semimartingales and M-valued semimartingales. Given the horizontal vector fields $H_1, \dots, H_d$ on FM and an $\mathbb{R}^d$-valued Brownian motion Z, the solution U to the stochastic differential equation

$$dU_t = H_i(U_t) \circ dZ_t^i, \tag{1.15}$$

is a horizontal Brownian motion in FM, where $\circ$ denotes integration in the Stratonovich sense. The construction is illustrated in Figure 1.1. More generally, if Z is any Euclidean valued semimartingale, then U determines a horizontal semimartingale in FM. Using the canonical projection π, the process $X_t := \pi(U_t)$ defines an M-valued semimartingale. The solution to (1.15) provides a stochastic parallel transport on M.

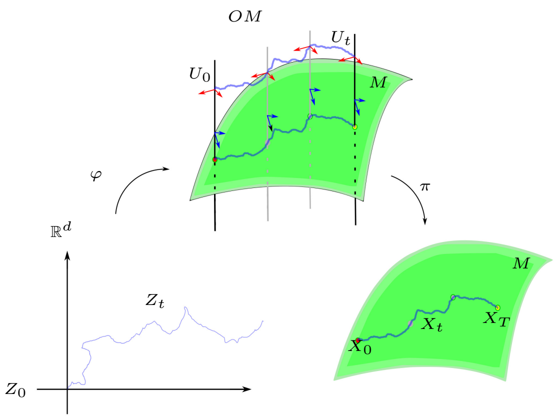


FIGURE 1.1: The stochastic development construction is defined by equation (1.15). The construction takes an $\mathbb{R}^d$-valued semimartingale Z_t which is mapped to a horizontal semimartingale $U_t := \varphi(Z_t)$. The horizontal semimartingale canonically maps to an M-valued semimartingale. After fixing the initial frame U_0, the construction is one-to-one.

On the orthonormal frame bundle OM, there is a horizontal lift of the Laplace-Beltrami operator to OM. This is sometimes denoted Bochner's horizontal Laplacian defined using the horizontal vector fields $H_i, i = 1, \dots, H_d$ by

$$\Delta_{OM} = \sum_{i=1}^{d} H_i^2. \tag{1.16}$$

From (1.16), the horizontal Laplacian is a Hörmander typer second-order elliptic operator defined intrinsically. The horizontal Laplacian defines the Brownian motion globally (up to its explosion time) on OM and hence on M.

1.7 Diffusion Bridges

In this section, we give a short introduction to the theory of diffusion bridges processes. We start by introducing diffusion bridges in the Euclidean setting before we describe certain bridges on manifolds.

1.7.1 Guiding in Euclidean Spaces

A stochastic bridge process is designed to go from point a to point b at time $T > 0$. The simplest bridge process is the standard Brownian bridge starting at zero conditioned to return at time $T > 0$. Let W denote a Brownian motion. The solution to the stochastic differential equation

$$dY_t = -\frac{Y_t}{T-t}dt + dW_t, \quad Y_0 = 0, \qquad Y_t = \int_0^t \frac{1-t}{1-s} dW_s, \tag{1.17}$$

defines a standard Brownian bridge. The standard Brownian bridge can be generalized to a Brownian bridge from a to b, by instead inserting $Y_t - b$ into the numerator of the drift term above. The transition density of a Brownian motion in $\mathbb{R}$ is given by

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{\|x-y\|^2}{2t}\right\}. \tag{1.18}$$

From the transition density of a Brownian motion (1.18), equation (1.17) can equivalently be written (as a Brownian bridge from a to b)

$$dY_t = \nabla_y \log p_{T-t}(y, b)\big|_{y=Y_t} dt + dW_t, \quad Y_0 = a. \tag{1.19}$$

We refer to the drift term in the equation above as the pulling or guiding term. The added drift term pulls the process in the direction of the pinning point. By Doob's h-transform, (1.19) holds more generally for any diffusion process X_t, which admit a transition density, in the sense that

$$dX_t = b_t(X_t)dt + \sigma_t(X_t)dW_t, \quad X_0 = a \tag{1.20}$$

$$dY_t = b_t(Y_t)dt + \left(\sigma\sigma^T\right)_t (Y_t)\nabla_y \log p(t, y; T, b)\big|_{y=Y_t} dt + \sigma_t(Y_t)dW_t, \quad Y_0 = a, \tag{1.21}$$

where $p(t, y; T, b)$ denotes the transition density of the underlying diffusion process X_t. Simulation of diffusion bridges satisfying (1.20) becomes complicated when the transition density is not explicitly available. The problem of simulating diffusion bridges has created an active research field over the last 15-20 years. More specifically, diffusion bridge simulation is the sub-field of statistics concerned with statistical inference for diffusion processes. As mentioned in [8], diffusion bridge simulation plays a fundamental role within simulation-based likelihood inference, including Bayesian inference for, e.g., discretely sampled diffusion processes.

The missing data problem dates back to Pedersen [36]. Since then, various methods have been developed to simulate diffusion bridges where the transition density is intractable [7, 8, 15, 20, 35, 40, 42, 50]. The seminal paper by Delyon and Hu [13], a generalization of the method proposed by Clark [11], suggested exchanging the drift term containing the intractable transition density with the tractable drift term in (1.19). The guiding term is easy to interpret and implement. Using a coupling method, the authors in [8] proposed a simple method to simulate conditioned diffusions going from point a to point b in time $T > 0$. Instead of introducing a guided drift term, as described above, the method relied on a forward process that started from a and backward processes that started from b to intersect at a random time in $[0, T]$. The idea of coupling using forward and backward processes was further instigated in [41, Section 4.5]. Here the transition density of the conditioned diffusion is considered as a three-point

$$p(s, y, t, x; \tau, z) = \frac{p(s, y; \tau, z)p(\tau, z; t, x)}{p(s, y; t, x)}. \tag{1.22}$$

It was conjectured that a guiding term for this coupling method should be a Delyon and Hu type guiding term, where the forward and backward processes had guiding drift terms pulling them towards each other. The idea of backward guiding was further developed in the context of geometric deep learning [43], building on the idea of [48]. In Chapter 4 in this thesis, we explore in more detail this backward guiding in the context of product manifolds and use the method to obtain estimates of the diffusion mean.

1.7.2 Guiding on Manifolds

The Brownian bridge on a manifold can be defined straightforwardly using Doob's transform. If p_t denotes the transition density of a Brownian motion on M, then the Brownian bridge on M is the stochastic process with infinitesimal generator $\frac{1}{2}\Delta_M + \nabla \log p_{T-t}(\cdot, v)$. Various properties of the Brownian bridge on manifolds have been studied by Bismut [6], Hsu [27], Driver [14], Güneysu [21], and many others. Similar to the Euclidean context, transition densities on manifolds are often intractable. Known closed-form expressions of the heat kernel on manifolds are, e.g., on hyperspheres, hyperbolic spaces, and Euclidean space. In general, approximations of the heat kernel must be made using numerical simulations.

Elworthy and Truman [16] introduced the semi-classical Riemannian bridge on manifolds with a pole, i.e., a point $p \in M$ such that the exponential map $\mathrm{Exp}_p \colon T_pM \to M$ is a global diffeomorphism. The semi-classical is the time-inhomogeneous Markov process whose infinitesimal generator is of the form $\frac{1}{2}\Delta_M + \nabla \log k_{T-t}(\cdot, v)$, with

$$k_t(x, v) = (2\pi t)^{-d/2} e^{-\frac{d(x,v)^2}{2t}} J^{-1/2}(x), \tag{1.23}$$

where $J(y) = |\det D_{\mathrm{Exp}_v^{-1}(y)} \mathrm{Exp}_v|$ denotes the Jacobian determinant of the exponential map at v. The semi-classical Riemannian bridge exhibits properties similar to the Euclidean Brownian bridge. In fact, the radial part of the semi-classical bridge is identical to the radial part of a Euclidean Brownian bridge (see, e.g. [47] and references therein). In particular, the semi-classical Riemannian bridge is identical to the Brownian bridge on $\mathbb{H}^3_\kappa$, the three-dimensional hyperbolic space with constant sectional curvature κ.

The existence of a pole is a rather restrictive condition. The generalized bridge [34] is an alternative to the semi-classical bridge. The generalized bridge process is a time-inhomogeneous Markov process with infinitesimal generator $\frac{1}{2}\Delta_M + \nabla \log q_{T-t}(\cdot, v)$, where

$$q_t(x, v) = (2\pi t)^{-d/2} e^{-\frac{d(x,v)^2}{2t}}, \tag{1.24}$$

defined for any M. When $M = \mathbb{R}^d$, the generalized bridge coincides with the Brownian bridge. It is clear from the infinitesimal generators that the semi-classical Riemannian bridge and the generalized bridge differ by a single term. The resulting process from this type of conditioning is also known as the Fermi bridge [47, 48]. Indeed, the conditioning point v can generally be any closed embedded submanifold $N \subseteq M$.

More recently, simulation of guided bridges on non-linear spaces has been done by Sommer et al. [45], Arnaudon et al. [3], and Bui et al. [9]. These papers present convincing numerical arguments for the absolute continuity to hold up until time T. The guiding terms introduced in both [45] and [9] fall in the class of generalized Brownian bridges. The validity of the scheme presented in [9] on the space of symmetric definite matrices was only presented numerically. In this thesis, we present a proof of the equivalence of the measures which holds on $[0, T]$. In fact, we show the equivalence of the measures hold on a much broader class of manifolds.

Chapter 2

Simulation of Conditioned Semimartingales on Riemannian Manifolds

The following chapter present a manuscript currently under review. The manuscript is a joint work made in collaboration with Stefan Sommer and posted to ArXiv as submission

> Mathias Højgaard Jensen, Stefan Sommer. "*Simulation of Conditioned Semimartingales on Riemannian Manifolds*" *arXiv:2105.13190*, (2021).

The paper presented in this chapter introduce a simulation scheme of guided bridge processes on Riemannian manifolds. The type of guiding term considered in the article generalizes the guiding term considered by Clark [11] and Delyon & Hu [13]. The paper appears to be the first paper to derive a bridge simulation scheme for likelihood inference on rather general manifolds. Recently, other papers have derived simulation schemes on specific manifolds [3, 9, 45].

Simulation of Conditioned Semimartingales on Riemannian Manifolds

Mathias Højgaard Jensen and Stefan Sommer*

Department of Computer Science, University of Copenhagen

Universitetsparken 1, 2100 Copenhagen, Denmark

Abstract

We present a scheme for simulating conditioned semimartingales taking values in Riemannian manifolds. The guided bridge scheme replaces the drift of the conditioned process with an approximation in terms of a scaled radial vector field, extending the guided bridge approach used for simulating Euclidean bridges. The approximation handles the fact that transition densities are generally intractable in geometric spaces, even for well-known processes. We prove the scheme's validity by a change of measure argument and show how the resulting guided processes can be used in importance sampling and approximating the unconditioned process's density. The guided bridge scheme is experimentally illustrated on two- and three-dimensional manifolds. Here, we compare density estimates using the sampling scheme to approximations using heat kernel expansions, and we use the scheme to estimate the diffusion mean of sampled data.

Keywords: bridge simulation, conditioned diffusions, diffusion mean, geometric statistics, Riemannian manifolds

1. Introduction

Techniques for simulating Euclidean diffusion bridge processes, or *bridge sampling*, have been studied in several cases over the last two decades, for instance, [8, 9, 25, 30]. Delyon and Hu [9] introduced a scheme to simulate conditioned diffusions using a stochastic differential equation (SDE) that is easy to simulate and absolutely continuous with respect to the desired distribution. The SDE proposed by Delyon and Hu interchanged a drift term in the original conditioned diffusion, depending on the possibly intractable transition density, with a drift arising from a Brownian bridge going.

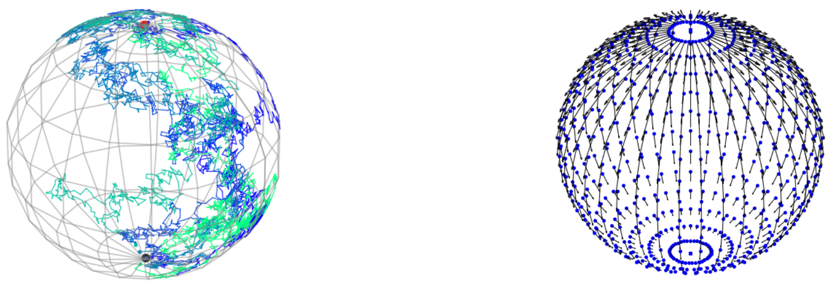
We here propose a scheme for simulating conditioned semimartingales on manifolds, which generalizes the result of Delyon and Hu [9] to the setting of

*Corresponding author. E-mail addresses: matje@di.ku.dk, sommer@di.ku.dk

smooth connected Riemannian manifolds. More precisely, we show that by adding a drift term to the semimartingale, the manifold equivalent to the drift term introduced in [9], we obtain a process that converges to the desired endpoint. The process is absolutely continuous with respect to the targeted diffusion bridge distribution.

Bridge sampling is an essential part of likelihood and Bayesian inference for discretely observed stochastic processes. For example, bridge sampling finds applications in geometric statistics, medical imaging, and shape analysis. Recent papers have introduced algorithms to simulate stochastic bridges on specific manifolds [3, 19, 32], which indicates the necessity for bridge sampling algorithms on general Riemannian manifolds. In particular, estimating the transition density of a stochastic process enables likelihood-based and Bayesian approaches to geometric statistics.

Figure 1(a) shows sample paths generated by the sampling scheme set up to approximate a conditioned Brownian motion on the sphere $\mathbb{S}^2$ starting at the north pole and conditioned to hit the south pole at $T = 1$. The scaled squared radial vector field, $\frac{\nabla r^2}{2(T-t)}$ Figure 1(b) acts as the guiding term forcing the process towards the target point.



(a) Four simulated paths of a diffusion bridge process from the north pole (red point) to the south pole (black point).

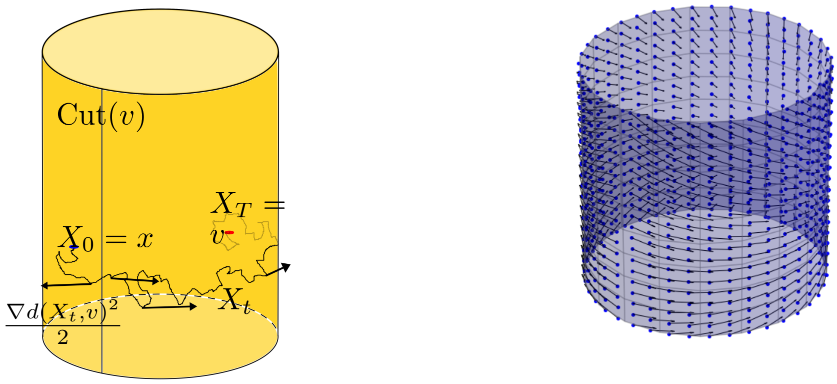
(b) Gradient vector field of the squared distance function on the 2-sphere centered at the south pole.

Figure 1: Figure 1a illustrates the simulation scheme's realizations to simulate conditioned diffusions, where the two bridging points are in each other's cut locus. Figure 1b shows the underlying radial vector field structure. It acts as a pulling term towards the point on which the process is conditioned.

1.1. Guided Bridges

Doob's h-transform is a classical way to show that the SDE for the conditioned process has a drift term that depends on the gradient of the logarithmic transition density, $\nabla_x \log p(t, x, v)$. Due to the generally intractable transition density, various methods to simulate conditioned processes have been introduced (e.g., [8, 9, 30, 25]).

Delyon & Hu [9] presented an algorithm to simulate the distribution of certain types of diffusions in $\mathbb{R}^d$ conditioned to hit a terminal point at some fixed time $T > 0$. Their main idea was to substitute the drift term $\nabla_x \log p(t, x, v)$,



(a) Illustration of one sample path of the radial bridge, X_t, from x to v, with corresponding guiding drift indicated by arrows. The drift changes sign when crossing cut locus (vertical line).

(b) A radial vector field on the cylinder, related to the radial bridge, X_t, centered at the point $X_T = v$.

Figure 2: Illustration of the effect of the cut locus on the radial bridge process. The guiding drift term, indicated by the tangent vectors along the sample path in (a) and the radial vector field (b), changes directions when crossing the cut locus.

which depends on the transition density, with the drift term appearing in the SDE for the generalized Brownian bridge. More precisely, they showed that the law of the process

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \qquad X_0 = a,$$

conditioned on $X_T = v$ (under suitable conditions on b and σ) is absolutely continuous with respect to the law of the process

$$dY_t = b(t, Y_t)dt - \frac{Y_t - v}{T - t}dt + \sigma(t, Y_t)dW_t, \qquad Y_0 = a. \tag{1}$$

Moreover, the conditional expectation given $X_T = v$ satisfies

$$\mathbb{E}\left[f(X)|X_T = v\right] = C\mathbb{E}\left[f(Y)\varphi_T\right], \tag{2}$$

for a constant $C > 0$, which depend on a, v, and T., where φ_T denotes the likelihood ratio (see [9] for more details). We present a result that generalizes the SDE in (1) and essentially also equation (2) to Riemannian manifolds, with an explicit expression for the likelihood ratio φ_T.

Switching from vector space to Riemannian manifolds, curvature removes the closed-form solution of the Brownian bridge drift. Inspired by Delyon and Hu's construction and the notion of Fermi bridges [35, 36], we instead propose to use the gradient $\nabla_x d(\cdot, v)^2/2$ in the drift term in a non-Euclidean generalization of (1). However, the existence of a non-trivial cut locus implies that this radial vector field is not continuous. Moreover, the $t \to 0$ convergence of the term to the gradient of the log-density is more intricate than the Euclidean situation. As shown by Malliavin, Stroock [24], and Turetsky [33],

$$\lim_{t\to 0} t\nabla_x^m \log p_{\mathcal{M}}(t,x,v) = -\nabla_x^m d(x,v)^2/2, \qquad \text{for all } m \geq 0, \tag{3}$$

uniformly on compact subsets of $\mathcal{M}\setminus \mathrm{Cut}(v)$, where ∇^m is the mth covariant derivative and $p_{\mathcal{M}}$ the heat kernel or the transition density of a Brownian motion on $\mathcal{M}$. The behaviour of the right hand side of (3), for $m = 1$, is illustrated in Figure 2. The cut locus adds extra difficulty because the squared distance is not differentiable on the cut locus, and the convergence in (3) is only uniform away from the cut locus.

1.2. Bridge Sampling: Parameter, Density, and Metric Estimation

Bridge processes are useful in a range of statistical and applied mathematical problems. We here list a few examples: Bridge sampling is essential for building data augmentation algorithms [25]. Treating sample points as if coming from incomplete observations, bridge sampling algorithms provides a better understanding of the underlying distribution, thereby offering a method for parameter estimation. Delyon and Hu [9] present a rather specific example of parameter estimation using bridge sampling. Furthermore, bridge sampling techniques produce accurate estimates of normalizing constants (see, e.g., [14]).

Statistical models based on diffusion processes find applications in geometric statistics, see, e.g., the book [28]. In particular, bridge sampling on manifolds yields estimates of the diffusion mean [15, 16]. The diffusion mean relies on the transition density of a Brownian motion and is a generalization of the Fréchet mean, defined as the argument that minimizes the average of distances. In this case, bridge sampling approximates the transition density or data likelihood. We demonstrate in the numerical examples below (Section 7) how bridge sampling can be used to approximate the transition density and to find diffusion means.

Estimation of transition densities by bridge sampling finds uses in the context of shape analysis [1, 2, 32]. Sommer et al. [32] show how to approximate the density and the metric structure on landmark configuration space. The latter is of particular interest since second-order elliptic diffusions describe Brownian motion with drift under a suitable change of (Riemannian) metric.

1.3. Structure of Paper

The paper is structured as follows. In Section 2, we review the background material of the frame bundle theory used to describe stochastic integration on manifolds and relevant existing work on manifold bridge processes. Section 3 describes the Radon-Nikodym derivative related to the change of measure, and we obtain a Cameron-Martin-Girsanov change of measure result. In Section 4, we present our two main results, that generalize [9, Theorem 5] to manifold-valued semimartingales. Section 5 is devoted to treating the radial process. We apply Barden and Le's result [21, Theorem 3] to the radial process and show the almost sure convergence of our guided process. We go on to rigorously prove the two propositions in Section **??** before ending with numerical examples in Section 7. We visualize the result of the simulation scheme on two- and three-dimensional manifolds, and we use the approach to approximate the heat kernel on different surfaces. Furthermore, we provide examples of how iterative bridge sampling can be used to estimate diffusion means.

1.4. Notation and Conventions

In this section, we establish the notation and conventions used in the paper. Throughout, we make the following assumptions:

- $\mathcal{M}$ is a d-dimensional smooth manifold
- W_t a d-dimensional Euclidean Brownian motion
- Z_t an d-dimensional Euclidean semimartingale, a solution to the SDE $dZ_t = b(t, Z_t)dt + \sigma(t, Z_t)dW_t$, for suitably integrable b and σ
- U_t a horizontal semimartingale in the frame bundle $\mathcal{FM}$
- X_t a $\mathcal{M}$-valued semimartingale defined as the projection $X_t := \pi(U_t)$
- $r_v(X_t) := d(X_t, v)$ is the Riemannian distance between X_t and $v \in \mathcal{M}$
- Both $\sigma_t(z) := \sigma(t, z)$ and its inverse $\sigma_t^{-1}(z)$ are bounded and differentiable
- The bracket process $[Z]$ is absolutely continuous as a random measure on $[0, \infty)$
- The set $\{t \colon X_t \in \mathrm{Cut}(v)\}$ is a Lebesgue null-set
- All SDEs admits strong solutions
- $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ is a filtered probability space satisfying the usual conditions

We will repeatedly use the Einstein summation convention of taking the sum over the index that is repeated as both a sub- and superscript, i.e., $a^i e_i = \sum_i a^i e_i$. The transpose of a matrix A will be denoted A^T. As the gradient of the distance function $d(\cdot, y)$ is non-smooth on the cut locus, we take $\nabla_x d(x, y)$ to be the usual gradient away from the cut locus and which vanishes on the cut locus. We will use $\tilde{}$ to denote lifts of functions on M to fiber bundles, i.e., with projection π, $\tilde{f} := f \circ \pi$.

2. Background

2.1. Manifold Valued Processes

In this section, we review the theory of manifolds and manifold-valued diffusion processes. In particular, we describe the main concepts of differential geometry, which are needed to define the Eells-Elworthy-Malliavin construction of manifold-valued stochastic processes through *horizontal lifts*. Two standard references for stochastic calculus on manifolds are Emery [13], and Hsu [17].

2.1.1. Riemannian Geometry

The *frame bundle* $\mathcal{FM}$ of a d-dimensional smooth manifold is a $d^2 + d$ dimensional manifold, where each point $u \in \mathcal{FM}$ correspond to a point $x \in \mathcal{M}$ together with an ordered basis (frame) of the tangent space $T_x\mathcal{M}$. It is convenient to think of u as a linear bijection $u\colon \mathbb{R}^d \to T_{\pi(u)}\mathcal{M}$, where $\pi\colon \mathcal{FM} \to \mathcal{M}$ denotes the canonical projection. In this sense we obtain an ordered basis $(u(e_1), ..., u(e_d))$ of $T_{\pi(u)}\mathcal{M}$ from the canonical basis $(e_1, ..., e_d)$ of $\mathbb{R}^d$. A *connection* in $\mathcal{FM}$ is a smooth choice of subspaces $\mathcal{H}_u\mathcal{FM} \subseteq T_u\mathcal{FM}$, for each $u \in \mathcal{FM}$, such that the tangent space at $u \in \mathcal{FM}$ splits into $T_u\mathcal{FM} = \mathcal{V}_u\mathcal{FM} \oplus \mathcal{H}_u\mathcal{FM}$, where $\mathcal{V}_u\mathcal{FM}$ denotes the tangent vectors that are tangent to the fibers. The space $\mathcal{H}_u\mathcal{FM}$ is called the horizontal tangent space. If we endow $\mathcal{M}$ with a Riemannian structure, that is, a smoothly varying inner product on the tangent spaces, the set of frames can be restricted to orthonormal frames, i.e., the map u being a linear isometry. The resulting subbundle is denoted the *orthonormal frame bundle*, $\mathcal{OM}$. We will throughout primarily work with $\mathcal{OM}$. There is a one-to-one correspondence between the horizontal tangent space at u and the tangent space at $\pi(u)$. This correspondence is described through the restriction of the pushforward map $\pi_*|_{\mathcal{H}_u\mathcal{OM}}\colon T_u\mathcal{OM} \to T_{\pi(u)}\mathcal{M}$. Furthermore, on the horizontal part of the frame bundle, there exists a set of fundamental horizontal vector fields, $H_1, ..., H_d$, defined by $H_i(u) = h_u\left(u(e_i)\right)$, where $h_u = \left(\pi_*|_{\mathcal{H}_u\mathcal{OM}}\right)^{-1}$ is the horizontal lift.

It is sometimes convenient to linearize the manifold in the following sense. Let $E\colon \mathbb{R}^d \to T_p\mathcal{M}$ be an isometric isomorphism from d-dimensional Euclidean space to the tangent space at $p \in \mathcal{M}$ and let $\mathrm{Exp}_p\colon T_p\mathcal{M} \to \mathcal{M}$ denote the exponential map at p. Define $\phi\colon D \to \mathbb{R}^d$ by $\phi := E^{-1} \circ \mathrm{Exp}_p^{-1}$, where $D \subseteq \mathcal{M}$ is the largest subset such that $\mathrm{Log}_p := \mathrm{Exp}_p^{-1}$ is well-defined, then the pair (D, ϕ) is a normal neighborhood centered at $p \in \mathcal{M}$.

2.1.2. Horizontal Semimartingales - Stochastic Development

If Z is a continuous Euclidean-valued semimartingale, i.e., a process which decomposes into a sum of a local martingale and an adapted process of locally bounded variation, the horizontal vector fields give rise to an SDE on the frame bundle, driven by the Stratonovich SDE,

$$dU_t = H_i(U_t) \circ dZ_t^i, \qquad U_0 = u \in \mathcal{FM}, \tag{4}$$

where U is a horizontal semimartingale on $\mathcal{OM}$. The canonical projection, $X := \pi(U)$, defines a process on $\mathcal{M}$. The process U is called the stochastic development of Z in $\mathcal{OM}$ and X the stochastic development of Z in $\mathcal{M}$. Similarly, Z is called the anti-development of X and U. This way of constructing stochastic processes on manifolds is contributed to Eells, Elworthy, and Malliavin, and is often referred to as the Eells-Elworthy-Malliavin construction. In colloquial terms, the construction is called "rolling without slipping." The intuition behind this terminology originates from rolling a ball along a path drawn with wet ink on a piece of paper. For a detailed description of this construction, one may consult [17]. When the starting frame $U_0 = u$ is fixed, the the construction gives a one-to-one map $\tilde{h}_u(Z_t) = U_t$. The inverse $\tilde{h}_u^{-1}(U_t) = Z_t$ is called the anti-development of U.

2.1.3. Geometric Itô Formula

The fundamental theorem of stochastic calculus is Itô's formula. The formula generalizes to manifold-valued semimartingales. If X is a continuous manifold-valued semimartingale, U its horizontal lift, and Z its anti-development, then for any smooth function $f \in C^\infty(\mathcal{M})$ the geometric Itô formula can be expressed as

$$df(X_t) = U_t(e_i)f(X_t)dZ_t^i + \tfrac{1}{2}U_t(e_j)U_t(e_k)f(X_t)d[Z^j, Z^k]_t.$$

The formula can equivalently be expressed in terms of the gradient and Hessian as

$$df(X_t) = \langle \nabla f(X_t), U_t dZ_t \rangle + \tfrac{1}{2} \operatorname{Hess}_{X_t} f(U_t, U_t)d[Z]_t.$$

2.1.4. Riemannian Brownian Bridges

Brownian bridges on manifolds also have an SDE representation. Whenever the semimartingale Z is a Brownian motion, the resulting process X on $\mathcal{M}$ (resp. U on the horizontal part of $\mathcal{OM}$) is a Brownian motion. With $p(t, x, v)$ denoting the transition density of the $\mathcal{M}$-valued Brownian motion and $\tilde{p}(t, u, v) = p(t, \pi(u), v)$ its lift, the corresponding Brownian bridge is the solution of the SDE

$$dU_t = H_i(U_t) \circ \left(dB_t^i + \left(U_t^{-1} \left(\nabla^H_{u|_{u=U_t}} \log \tilde{p}(t, u, v) \right) \right)^i dt \right), \qquad U_0 = u_0,$$

where $\nabla^H \tilde{f} = \{H_1 \tilde{f}, \ldots, H_d \tilde{f}\}$ is the horizontal gradient. As was the case in the Euclidean setting, the SDE includes the Brownian motion's transition density. However, unlike the Euclidean case for Brownian motion, the transition density is not always tractable.

Writing the transition density for the Brownian motion as $p(t, x, v) = p_t(x, v)$, a natural representation of the Brownian bridge is as a time-inhomogeneous Markov process with an infinitesimal generator

$$\tfrac{1}{2}\Delta_{\mathcal{M}} + \nabla \log p_{T-t}(\cdot, v),$$

where $\Delta_{\mathcal{M}}$ is the generalized Laplacian (Laplace-Beltrami) operator on $\mathcal{M}$. We will drop the $\mathcal{M}$ in the sequel whenever referring to the Laplace-Beltrami operator.

2.1.5. Related Constructions

As the Brownian motion's transition density is only known in a closed-form on a handful of manifolds, other types of bridge processes have been considered (see e.g. [18]). A *semi-classical Brownian bridge* on a Riemannian, also known as *Brownian Riemannian bridge*, is a time inhomogeneous strong Markov process with infinitesimal generator

$$\tfrac{1}{2}\Delta + \nabla \log k_{T-t}(\cdot, v),$$

where k_t is the function defined by

$$k_t(x,v) := (2\pi t)^{-n/2} e^{-\frac{r^2(x,v)}{2t}} J^{-1/2}(x),$$

and $J(x) = |\det D_{\exp_v^{-1}(x)} \exp_v|$ is the Jacobian determinant of the exponential map at v (See any of [10, 11, 12] for more details on semi-classical bridges.). This description typically relies on the existence of a pole, i.e. a point in $\mathcal{M}$ where the exponential map maps diffeomorphically to $\mathcal{M}$. This assumption avoids the nuissance of the cut locus. The radial part of the semi-classical bridge has the distribution of a Euclidean valued Brownian bridge.

A generalization beyond the cut locus of the heat kernel formula described by Elworthy and Truman [11] is due to Thompson [35]. Let N be a closed embedded submanifold on $\mathcal{M}$ and define the distance function to N by $r_N(\cdot) := d(\cdot, N)$, then introduce the diffusion on $\mathcal{M}$ with time-dependent infinitesimal generator

$$\frac{1}{2}\Delta_{\mathcal{M}} - \frac{\nabla r_N^2}{2(T-t)} \quad = \quad \frac{1}{2}\Delta_{\mathcal{M}} - \frac{r_N}{T-t}\frac{\partial}{\partial r_N},$$

where $\frac{\partial}{\partial r_N}$ denotes differentiation in the radial direction and Δ denotes the Laplace-Beltrami operator. This diffusion process is called a *Fermi bridge* (see [35, 36]). More generally, [22] defined *generalised Brownian bridge processes*, between x_0 and $x_T = v$ with terminal time T, as a Markov process $(x_t)_{t\geq 0}$ with infinitesimal generator

$$\tfrac{1}{2}\Delta_{\mathcal{M}} - f(t)\nabla\tfrac{r_v^2}{2},$$

where f is a suitably smooth real valued function defined on $[0,1)$ satisfying $\lim_{t\uparrow T} f(t) = \infty$ and $\lim_{t\uparrow T} x_t = x_T$ almost surely. We also refer the reader to [23] for a description of hypoelliptic bridges on manifolds.

3. A Girsanov Change of Measure

Intractable transition densities complicate the exact simulation of the desired conditioned processes. Various methods exist to approximate these processes [6, 5, 8, 9, 30, 25]. Similar to all the methods, they rely on a change of measure argument and that the changed measure respects the original measure in the sense of absolute continuity.

In this section, we recall a Cameron-Martin-Girsanov result for manifold valued processes, see [10], that we will need later on. We assume throughout that $X_t = \pi(U_t)$ is non-explosive. This is for example the case if $\mathcal{M}$ is compact. Let $dM_t = \sigma_t(Z_t)dW_t$ be the Euclidean martingale part of Z, where we have assumed that σ is invertible. Define the local martingale L_t by the equation

$$L_t := -\int_0^t \left\langle \frac{\nabla^H_{u|_{u=U_s}} \tilde{r}_v^2(u)}{2(T-s)}, U(e_i) \right\rangle dM_s^i.$$

Using that U is an isometry, the local martingale is identical to the process

$$L_t := -\int_0^t \left\langle \sigma_s^{-1}(\tilde{h}_u^{-1}(U_s))U_s^{-1}\left(\frac{\nabla^H_{u|_{u=U_s}}\tilde{r}_v^2(u)}{2(T-s)}\right), e_j\right\rangle dW_s^j.$$

The corresponding Radon-Nikodym derivative is given by

$$D_t = \exp\left[L_t - \frac{1}{2}\int_0^t \left\|\sigma_s^{-1}(\tilde{h}_u^{-1}(U_s))U_s^{-1}\left(\frac{\nabla^H_{u|_{u=U_s}}\tilde{r}_v^2(u)}{2(T-s)}\right)\right\|^2 ds\right]. \tag{5}$$

The Novikov condition ensures that the measure $\mathbb{Q}_t$, defined as $d\mathbb{Q}_t = D_t d\mathbb{P}$, is equivalent to the measure $\mathbb{P}$ on the time interval $[0, T)$. As a consequence of Girsanov's theorem, we have that the $\mathbb{P}_t$-Brownian motion, W_t, satisfies the equation

$$dW_t = d\hat{W}_t - \sigma_t^{-1}(\tilde{h}_u^{-1}(U_t))U_t^{-1}\left(\frac{\nabla^H_{u|_{u=U_s}}\tilde{r}_v^2(u)}{2(T-t)}\right)dt,$$

and we have that $\hat{W}_t$ is a $\mathbb{Q}_t$-Brownian motion. Plugging the expression for W_t into equation (4) yields the expression in (10) below.

Example 1. Consider the case where the driving semimartingale is a standard Brownian motion. The Radon-Nikodym derivative (5) simplifies to

$$D_t := \exp\left[-\int_0^t \frac{r_v(Y_s)}{T-s}\left\langle \frac{\partial}{\partial r_v}, U_s dB_s\right\rangle - \frac{1}{2}\int_0^t \frac{r_v^2(Y_s)}{(T-s)^2}ds\right], \tag{6}$$

almost surely, since the norm of the radial derivative is constantly one away from the cut locus and its starting point, $\mathrm{Cut}(x_0) \cup \{x_0\}$. In particular, as mentioned in [36], the integral

$$\int_0^t \left\langle \frac{\partial}{\partial r_N}, U_s dB_s\right\rangle = \beta_t$$

is a one-dimensional standard Brownian motion, which follows from Levy's characterization theorem of Brownian motions and that U consists of isometries.

4. Bridge Simulation on Manifolds

4.1. Guided Semimartingales

This section initiates horizontal semimartingales from Euclidean-valued semimartingales and describes the guided semimartingales considered in this paper. Throughout, we assume that $Z = (Z^1, Z^2, ..., Z^d)$ is a Euclidean-valued semimartingale given by

$$dZ_t^k = a^k(t, Z_t)dt + \sigma_m^k(t, Z_t)dW_t^m, \tag{7}$$

where W is a Euclidean-Brownian motion and a and σ are suitably integrable maps. Horizontal semimartingales are then obtained as solutions to (4) which evidentally happens if and only if for all smooth functions $\tilde{f}$ on $\mathcal{OM}$ it holds that

$$d\tilde{f}(U_t) = H_i\tilde{f}(U_t) \circ dZ_t^i.$$

By the Itô-Stratonovich conversion, we obtain the corresponding Itô equation

$$d\tilde{f}(U_t) = \tilde{\mathcal{L}}_u\tilde{f}(U_t)dt + \sigma_m^k(t, \tilde{h}_u^{-1}(U_t))H_k\tilde{f}(U_t)dW_t^m,$$

where $\tilde{\mathcal{L}}_u$ is an operator in $\mathcal{OM}$ defined by

$$\tilde{\mathcal{L}}_u\tilde{f}(U_t) := a^k(t, \tilde{h}_u^{-1}(U_t))H_k\tilde{f}(U_t) + \tfrac{1}{2}C^{ij}(t, \tilde{h}_u^{-1}(U_t))H_iH_j\tilde{f}(U_t), \qquad (8)$$

with initial frame $u \in \mathcal{OM}$, and where $C^{ij} := \left(\sigma\sigma^T\right)_{ij}$ denotes the ij'th entrance of $\sigma\sigma^T$. By definition of Bochner's horizontal Laplacian, $\Delta_{\mathcal{OM}}$, whenever σ is orthogonal (8) simplifies to $\tilde{\mathcal{L}} = \tilde{V} + \frac{1}{2}\Delta_{\mathcal{OM}}$, where $\tilde{V}$ is the horizontal lift of a vector field on $\mathcal{M}$. Adding a drift term pointing in the radial direction gives rise to a new operator on $\mathcal{OM}$

$$\tilde{\mathcal{L}}_u - \frac{\nabla^H\tilde{r}_v^2}{2(T-t)}. \qquad (9)$$

We further define $\mathcal{L}_u$ as the operator in $\mathcal{M}$ that satisfies $\tilde{\mathcal{L}}_u\tilde{f}(u) = \mathcal{L}_uf(x)$, for all $f \in C^\infty(\mathcal{M})$, where $\tilde{f} = f \circ \pi$. As we shall see below, this added drift term acts as a guiding term that forces the process towards its target.

Example 2. Whenever C is the identity matrix and a vanishes, i.e., when Z is a standard Brownian motion, $\tilde{\mathcal{L}}_u$ becomes the horizontal Laplacian, $\Delta_{\mathcal{OM}}$, identified by

$$\Delta_{\mathcal{OM}}\tilde{f}(u) = \Delta_{\mathcal{M}}f(x),$$

for all $f \in C^\infty(\mathcal{M})$, where $\tilde{f} = f \circ \pi$ is the lift of f to $C^\infty(\mathcal{OM})$.

The operator (9) on $\mathcal{OM}$ can equivalently be expressed as the SDE with the following expression

$$dV_t = H_i(V_t) \circ \left(dZ_t^i - \left(V_t^{-1}\left(\frac{\nabla^H\tilde{r}_v^2(V_t)}{2(T-t)}\right)\right)^i dt\right), \qquad V_0 = u. \qquad (10)$$

We will be considering solutions to this SDE and its projection to $\mathcal{M}$ throughout the paper.

4.2. The Main Results

We here establish a method to simulate conditioned semimartingales on a Riemannian manifold $\mathcal{M}$, starting from some fixed initial point $x_0 \in \mathcal{M}$ and conditioned to be at $v \in \mathcal{M}$ at time T, for $T > 0$. Utilizing stochastic development, we obtain manifold valued semimartingales from Euclidean valued ones by solving SDEs on the orthonormal frame bundle, $\mathcal{OM}$. Throughout, we assume a strong and non-explosive solution to (4).

We recall from Girsanov's theorem that any $\mathbb{P}$-local martingale M_t becomes a $\mathbb{Q}$-local martingale $\tilde{M}_t$ under a suitable change of measure (see e.g. [29]). Consequently, approximating the measure of a conditioned process relates naturally to Girsanov's theorem. The necessity for approximating the conditioned measure originates from likelihood inference for diffusion processes. Therefore, the typical intractable transition density and the importance of bridge sampling in likelihood and Bayesian inference motivate the necessity for approximation schemes to sample from the desired distribution.

We can now state the main result of the paper, which is a generalization of Delyon and Hu [9] to Riemannian manifolds.

Theorem 3. *Let Z be a solution of* (7)*, where σ is invertible and bounded with bounded inverse. Let U be the stochastic development of Z defined by* (4)*, with $X = \pi(U)$. Let V be the solution to* (10) *and let $Y = \pi(V)$ denote the canonical projection onto $\mathcal{M}$. The law of the process $(X_t|X_T = v)_{0\le t<T}$ is absolutely continuous with respect to the law of $(Y_t)_{0\le t<T}$ and we have that*

$$\mathbb{E}[f(X)|X_T = v] = \lim_{t\uparrow T} \frac{\mathbb{E}[f(Y)\varphi_t]}{\mathbb{E}[\varphi_t]},$$

for any non-negative measurable funtion f, where the likelihood φ_t has the form

$$\begin{aligned} -2d\log\varphi_t = {} & \frac{r_v^2(Y_t)}{T-t}\xi_t^T d\left(A(t,Z_t)\right)\xi_t + F(t)dr_v(Y_t) + J_i(t)d\xi_t^i \\ & + H_i(t)d[r_v(Y),Z^i]_t + I_j(t)d[r_v(Y),\xi^j]_t + K_{ij}(t)d[\xi^i,Z^j]_t \qquad (11) \\ & + \frac{1}{2}\left(G(t)d[r_v(Y),r_v(Y)]_t + J_{ij}(t)d[\xi^i,\xi^j]_t\right). \qquad (12) \end{aligned}$$

Proof. Here, we present the main structure of the proof. The constituent parts will be made rigorous in the sections that follow. By a proper change of measure, through Girsanov's theorem (Section 3), it is possible to obtain the SDE in equation (10) from equation (4). The solution of (10) we term the *radial bridge process*. The change of measure is valid on the interval $[0,t]$, for every $t \in [0,T)$ (Section 3). A decomposition of the Radon-Nikodym derivative leads to a term only involving the radial process and the time to the termination. By invoking an L^2 bound on the radial bridge process, we show the radial bridge process is almost sure convergence to the desired point (5.1.2). The result is concluded by an argument (Section **??**) similar to an argument made in Delyon & Hu [9, Lemma 7] □

Remark 4. The guiding drift term in (10) clearly depends on the squared radial vector field. Another way to represent the SDE (10) is by

$$dV_t = \sum_{i=1}^d H_i(V_t)\circ\left(dZ_t^i - \tfrac{1}{2(T-t)}H_i\tilde{r}_v^2(V_t)dt\right), \qquad V_0 = u.$$

In order to see this, we simply note that

$$\left(V_t^{-1}\left(\frac{\nabla^H\tilde{r}_v^2(V_t)}{2}\right)\right)^i = \tfrac{1}{2}\left\langle\nabla r_v^2(X_t), U_t(e_i)\right\rangle = \tfrac{1}{2}H_i\tilde{r}_v^2(V_t).$$

4.2.1. The Case of Brownian Motion

A special case of the above result is when the X is a Brownian motion. We state the result below as it has an independent interest. Let $\theta_x(y) = |\det D_y \exp_x|$ be the Jacobian determinant at y of the Jacobi field along the geodesic from x to y.

Theorem 5. *Let U be the solution of* (4)*, with Z a standard Brownian motion in $\mathbb{R}^d$, and let X be the canonical projection onto $\mathcal{M}$, Then X is a Brownian motion on $\mathcal{M}$. Furthermore, if $\mathcal{M}$ is a simply connected manifold and V is the solution of* (10)*, then the conditioned law of X given $X_T = v$ is absolutely continuous with respect to the law of Y on* $[0, T]$.

$$\mathbb{E}[f(X)|X_T = v] = C\mathbb{E}\left[f(Y)\varphi_T\right], \tag{13}$$

where the constant depends on the initial value y_0, the conditioning point v, the arrival time T, the curvature in the radial direction, and in certain cases the geometric local time. In particular, the likelihood φ_t in (11) *simplifies to*

$$d\log\varphi_t = \frac{r(Y_t)}{T-t}\left(d\eta_t + dL_t^{\mathring{C}}\right),$$

where

$$d\eta_t = \frac{\partial}{\partial r_v}\log\theta_v^{-\frac{1}{2}}(Y_t)dt$$

is supported on $\mathcal{M}\backslash\,\mathrm{Cut}(v)$ and $L_t^{\mathring{C}}$ is the geometric local time defined in (15)*, with support at the cut locus* $\mathrm{Cut}(v)$.

Example 6. In the situation where $\mathcal{M} = \mathbb{R}^d$, for the Brownian motion case, we obtain the result of Theorem 5 in Delyon and Hu [9] from our result. No curvature or cut locus exists in the Euclidean setting and, therefore, $\mathbb{E}[\varphi_T] \equiv 1$.

The theorem above can be used to prove analytically the results conjectured in [7], where $\mathcal{M}$ is the space of symmetric positive definite matrices SPD. The space of SPD matrices is convex and hence simply connected. The extra drift term can then be handled in a manner similar to Theorem 6 in [9].

We go on in the rest of the text to prove Theorem 3 and Theorem 5. In particular, we shall derive the expression of φ_t in (11) as well as review the behavior of the radial process of a continuous semimartingale on $\mathcal{M}$.

5. Radial Part of Semimartingale on a Manifold

To analyze the guided process behavior, we need to control the radial part of the guided process. Recall here that the radial process is defined by $r_v(x) := d(x, v)$, where d is the Riemannian distance function. Barden and Le described the behavior of the radial process for semimartingales on manifolds [4, 21], generalizing a result by Kendall [20], which describes the radial behavior of a Brownian motion. To take full advantage of Barden and Le's result, we describe the geometry of the cut locus and present their result.

Let x_0 be a given reference point on a complete Riemannian manifold $\mathcal{M}$. We let $\mathcal{Q}(x_0)$ denote the set of points in $T_{x_0}\mathcal{M}$ where the exponential map is singular, that is, points $v \in T_{x_0}\mathcal{M}$ such that $D(\exp_{x_0})(v)$ has rank $k < d$ (for a d dimensional manifold $\mathcal{M}$). Define $Q(x_0) := \exp_{x_0}(\mathcal{Q}(x_0))$ as the image under $\exp_{x_0}$. We say that $\mathcal{Q}(x_0)$ and $Q(x_0)$ are the conjugate loci of $T_{x_0}\mathcal{M}$ and $\mathcal{M}$, respectively. Let $\mathrm{Cut}(x_0)$ denote the cut locus of x_0 in $\mathcal{M}$. Then $\mathrm{Cut}(x_0) \cap Q(x_0)$ denote the conjugate part of the cut locus of x_0.

A $(d-1)$-dimensional submanifold $S \subset \mathcal{M}$ is called two-sided if its normal bundle is trivial. By Barden and Le [21, Theorem 2] the cut locus, except for a subset of Hausdorff $(d-1)$-measure zero, is described as a countable disjoint union, $\mathscr{L}$, of open two-sided $(d-1)$-dimensional submanifolds. Define E to be the union of points consisting of $\mathrm{Cut}(x_0) \cap Q(x_0)$ and the points in $\mathrm{Cut}(x_0) \backslash Q(x_0)$ where at least 3 minimal geodesics to x_0 exists. The set E has Hausdorff $(d-1)$-measure zero. A set A is said to be a polar set for X if the first hitting time of A by X is almost surely infinite.

Suppose that the process U is the solution of (4), that E is a polar set of X, and that $\mathbb{P}(X_t = x, 0 < t \leq \infty) = 0$. Furthermore, let $\mathring{C}$ be a disjoint union of $\mathrm{Cut}(v)$, which consists of countably many smooth connected two-sided $(d-1)$-dimensional submanifolds, where $\mathring{C}_i$ denote the connected components of $\mathring{C}$, and let $D \subseteq \mathcal{M}$ be a regular domain. We denote by $D_{\pm} f_x(\pm\nu)$ the one-sided Gâteaux derivatives of f, defined by

$$D_+ f_x(\nu) := \lim_{\varepsilon \downarrow 0} \tfrac{1}{\varepsilon} \left(f\left(\exp_x(\varepsilon\nu)\right) - f(x) \right), \qquad D_- f_x(\nu) := -D_+ f_x(-\nu),$$

for all $x \in \mathcal{M}$ and $\nu \in T_x\mathcal{M}$. By virtue of Barden and Le [21, Theorem 3], we get a formula for the radial process of $X = \pi(U)$ as (see also [35] for a more general formula)

$$\begin{aligned}
r_v\left(X_{t\wedge\tau_D}\right) = r_v(x) &+ \int_0^{t\wedge\tau_D} 1_{\{X_s \notin \mathring{C}\}} \langle \nabla r_v(X_s), U_s dZ_s \rangle \\
&+ \int_0^{t\wedge\tau_D} 1_{\{X_s \in \mathring{C}_i\}} \left\langle \nabla \left(r_v \circ P^i\right)(X_s), U_s dZ_s \right\rangle \\
&+ \frac{1}{2} \int_0^{t\wedge\tau_D} 1_{\{X_s \notin \mathring{C}\}} \operatorname{Hess}_{X_s} r_v(U_s, U_s) d[Z,Z]_t \\
&+ \frac{1}{2} \int_0^{t\wedge\tau_D} 1_{\{X_s \in \mathring{C}_i\}} \operatorname{Hess}_{X_s} \left(r_v \circ P^i\right)(U_s, U_s) d[Z,Z]_t \\
&+ \frac{1}{2} \int_0^{t\wedge\tau_D} \Big(D_+ r_{X_s}(\nu) dL_s^{+\nu,\mathring{C}}(X) + D_+ r_{X_s}(-\nu) dL_s^{-\nu,\mathring{C}}(X) \Big) \\
&+ L^0_{t\wedge\tau_D}\left(r_v(X)\right),
\end{aligned}$$

for all $t \geq 0$, almost surely, where τ_D is the first exit time of D by X, P^i the orthogonal projection onto the connected component $\mathring{C}_i$ of $\mathring{C}$, L^0 the local time at 0 of $r_v(X)$, $dL^{\pm\nu,\mathring{C}}$ the associated random measures of the geometric local times at $\mathring{C}$ as defined in [35], and $D_{\pm} r_{X_t}(\pm\nu)$ are the one-sided Gâteaux derivatives of r along $\pm\nu$, the unit normal vector-field of $\mathrm{Cut}(x_0) \backslash E$.

Under the assumptions made in Section 1.4, the terms which depend on the projections vanish in the equation above. In this case, the formula for the squared radial process becomes

$$\begin{aligned} r_v^2\left(X_{t\wedge\tau_D}\right) = r_v^2(x) \;&+ \int_0^{t\wedge\tau_D} \left\langle \nabla r_v^2(X_s), U_s dZ_s \right\rangle \\ &+ \frac{1}{2}\int_0^{t\wedge\tau_D} \mathrm{Hess}_{X_s}\, r_v^2(U_s, U_s) d[Z,Z]_t \\ &+ \frac{1}{2}\int_0^{t\wedge\tau_D} r_v(X_s)\left(D_+ r_{X_s}(\nu) + D_+ r_{X_s}(-\nu)\right) d\tilde{L}_s^{\mathring{C}}(X), \end{aligned} \tag{14}$$

where $d\tilde{L}^{\mathring{C}}$ is the random measure carried by $\mathring{C}$, associated to the continuous predictable non-decreasing functional process. Furthermore, as defined in [4], we have $D_+ r_{X_s}(\nu) + D_+ r_{X_s}(-\nu) \le 0$ and so we define the random measure

$$dL_t^{\mathring{C}} = -\left(D_+ r_{X_s}(\nu) + D_+ r_{X_s}(-\nu)\right) d\tilde{L}_s^{\mathring{C}}(X). \tag{15}$$

Example 7. The radial part of a Brownian motion, which is due to Kendall [20], has the following representation:

$$dr_v(X_t) = \tfrac{1}{2}\Delta_M r_v(X_t)dt + d\beta_t - dL_s^{\mathring{C}}, \tag{16}$$

where X_t is a $\mathcal{M}$-Brownian motion, $L_t^{\mathring{C}}$ is the local time at the cut locus, and β_t is a one-dimensional real valued standard Brownian motion.

5.1. Properties of the Radial Bridge

One of the most fundamental properties of any bridge proposal process is that it converges to the correct point almost surely. The radial bridge process with an infinitesimal generator given by (9) has a drift term that always points in the radial direction. This drift acts as a pulling term in the radial direction, and it ensures that the radial bridge process converges to the desired endpoint.

From (16) and its more general representation, we see that the behavior of the radial process depends on the growth of $\mathcal{L}_u r_v$ relative to r_v. By imposing growth conditions on the sectional and Ricci curvatures, the growth of $\mathcal{L}_u r_b$ can be controlled. For example, an upper bound on the sectional curvature yields a lower bound on $\Delta_M r_v$, and a lower bound on the Ricci curvature yields an upper bound on $\Delta_M r_b$ [17, Chapter 3]. We make the following assumption; Let $D \subseteq \mathcal{M}$ be a regular domain, and assume there exist constants $\nu \ge 1$ and $\lambda \in \mathbb{R}$ such that

$$\mathcal{L}_u r_v^2 \le \nu + \lambda r_v^2 \tag{17}$$

on $D \setminus \mathrm{Cut}(p)$. When $\mathcal{L}_u = \frac{1}{2}\Delta_M + b$, for some locally bounded and measurable drift b that grows linearly in r_v, geometric conditions to ensure (17) where given in Theorem 1.4.5 [35]. In particular, if $\mathcal{M}$ is compact (17) holds. The operator in defined in [7] is an example of an operator of the form $\frac{1}{2}\Delta_M + b$, where b grows linearly in r_v. The space of symmetric positive definite matrices equipped with the affine invariant metric has sectional curvatures bounded in $(-1/4, 0)$ [27], and when equipped with the log-Euclidean metric, the Ricci curvature vanished everywhere, hence the operator defined in [7] is bounded by (17) [35].

5.1.1. Bound For the Radial Bridge

Throughout this section, let U be the process generated by (9) and let $X = \pi(U)$ be the projection of U on $\mathcal{M}$. Then by Barden and Le's formula (14) together with the geometric Itô formula for semimartingales, we have

$$\begin{aligned} r_v^2(Y_{t\wedge\tau_D}) &= r_v^2(x_0) + 2\int_0^{t\wedge\tau_D} r_v(Y_s)dN_s \\ &\quad - \int_0^{t\wedge\tau_D} r_v(Y_s)dL_s^{\mathring{C}} + \int_0^{t\wedge\tau_D} \left(\mathcal{L}_u r_v^2(Y_s) - 2\frac{r_v^2(Y_s)}{T-s}\right) ds, \end{aligned}$$

where N_t is an $\mathbb{R}$-valued martingale and $\mathcal{L}_u$ is defined in (8). As the second term above is a local martingale we get

$$\begin{aligned} \mathbb{E}\left[r_v^2\left(Y_{t\wedge\tau_{D_i}}\right)\right] &= r_v^2(x_0) - \mathbb{E}\left[\int_0^{t\wedge\tau_D} r_v(Y_s)dL_s^{\mathring{C}}\right] \\ &\quad + \mathbb{E}\left[\int_0^{t\wedge\tau_D} \mathcal{L}_u r_v^2(Y_s)ds\right] - 2\mathbb{E}\left[\int_0^{t\wedge\tau_D} \frac{r_v^2(Y_s)}{T-s}ds\right], \end{aligned} \tag{18}$$

almost surely, for all $t \in [0, T)$. In particular, under the assumption of (17) (see below) the two last terms can be rewritten by Lebesgue's dominated convergence and Fubini's theorem such that

$$\mathbb{E}\left[r_v^2\left(Y_{t\wedge\tau_{D_i}}\right)\right] = r_v^2(x_0) - \mathbb{E}\left[\int_0^{t\wedge\tau_D} r_v(Y_s)dL_s^{\mathring{C}}\right] \tag{19}$$

$$+ \int_0^t \mathbb{E}\left[1_{(s<\tau_D)}\mathcal{L}_u r_v^2(Y_s)\right] ds - 2\int_0^t \mathbb{E}\left[1_{(s<\tau_D)}\frac{r_v^2(Y_s)}{T-s}\right] ds, \tag{20}$$

almost surely, for all $t \in [0, T)$.

Theorem 8. *(Adapted from theorem 3.1 in [36]) Let D be a regular domain (smooth boundary and compact closure) in $\mathcal{M}$, and τ_D be the first exit time of Y from D. Assume* (17) *holds on $D\backslash \operatorname{Cut}(v)$. Then the radial bridge process satisfies*

$$\mathbb{E}[1_{\{t<\tau_D\}} r_v^2(Y_t)] \leq \left(r_v^2(x_0) + \nu t\left(\frac{t}{T-t}\right)\right)\left(\frac{T-t}{t}\right)^2 e^{\lambda t}, \tag{21}$$

for all $t \in [0, T)$.

Proof. Define

$$f(t) := \mathbb{E}\left[1_{(t<\tau_{D_i})} r_v^2(Y_t)\right]$$

and note that

$$f(t) := \mathbb{E}\left[1_{(t<\tau_{D_i})} r_v^2(Y_t)\right] = \mathbb{E}\left[r_v^2\left(Y_{t\wedge\tau_{D_i}}\right)\right] - \mathbb{E}\left[1_{(t\geq\tau_{D_i})} r_v^2(Y_{\tau_{D_i}})\right].$$

Since the maps

$$t \mapsto \mathbb{E}\left[\int_0^{t\wedge\tau_D} r_v(Y_s) dL_s^{\mathring{C}}\right], \quad t \mapsto \mathbb{E}\left[1_{(t\geq\tau_{D_i})} r_v^2(Y_{\tau_{D_i}})\right],$$

are non-decreasing their derivatives are non-negative. Differentiating the function $f(t)$ it follows that

$$\begin{aligned}
\frac{d}{dt} f(t) \leq & \frac{d}{dt}\mathbb{E}\left[r_v^2\left(Y_{t\wedge\tau_{D_i}}\right)\right] \\
\leq & \mathbb{E}\left[1_{(t<\tau_{D_i})}\mathcal{L}_z r_v^2(Y_t)\right] - 2\mathbb{E}\left[1_{(s<\tau_{D_i})}\frac{r_v^2(Y_s)}{T-s} ds\right] \\
\leq & \nu + \left(\lambda - \frac{2}{T-t}\right) f(t),
\end{aligned}$$

where the second inequality follows by (19) and the third inequality follows from (17) and Leibniz' rule. An application of Gronwall's inequality yields the claim

$$\begin{aligned}
f(t) \leq & \left(r_v^2(x_0) + \nu \int_0^t \left(\frac{T}{T-s}\right)^2 e^{-\lambda s} ds\right) \left(\frac{T-t}{t}\right)^2 e^{\lambda t} \\
\leq & \left(r_v^2(x_0) + \nu t \left(\frac{t}{T-t}\right)\right) \left(\frac{T-t}{t}\right)^2 e^{\lambda t}.
\end{aligned}$$

□

5.1.2. Almost Sure Convergence

We show here that the bridge property is satisfied.

Proposition 9. *Let V be the solution of* (10) *and $Y_t = \pi(V_t)$. Assume* (17) *holds. The radial bridge process Y satisfies the bridge property*

$$\lim_{t\uparrow T} r_v(Y_t) = 0,$$

$\mathbb{Q}$-almost surely, where $\mathbb{Q}$ is the extension of $\mathbb{Q}_t$.

Proof. Let $\{D_n\}_{n=1}^{\infty}$ be an exhaustion of $\mathcal{M}$, that is, the sequence consists of open, relatively compact subsets of M such that $\bar{D}_n \subseteq D_{n+1}$ and $\mathcal{M} = \bigcup_{n=1}^{\infty} D_n$. Furthermore, let τ_{D_n} denote the first exit time of X from D_n, then from Theorem 8 (more precisely from (21)) we have that the sequence $\left(\mathbb{E}[1_{\{t<\tau_{D_n}\}} r_v^2(Y_t)]\right)_{n=1}^{\infty}$ is non-decreasing and bounded; hence from the monotone convergence theorem, it has a limit which is bounded by the right-hand side of (21). Applying Jensen's inequality to the left-hand side of (21)

$$\mathbb{E}[r_v(Y_t)] \leq \left(r_v^2(x_0) + \nu t\left(\frac{t}{T-t}\right)\right)^{\frac{1}{2}} \left(\frac{T-t}{t}\right) e^{\frac{\lambda t}{2}}.$$

Since obviously $\mathbb{E}[r_v(Y_T)] = r_v(Y_T)\mathbb{Q}(r_v(Y_T) \neq 0)$, by Fatou's lemma

$$\mathbb{E}[r_v(Y_T)] \leq \liminf_{t\to T} \mathbb{E}[r(Y_t)] = 0,$$

we conclude that $r_v(Y_t) \to 0$, $\mathbb{Q}$-almost surely. $\square$

6. Proof of Theorem 3 and Theorem 5

To establish the two propositions, we need a few technical lemmas. The first result provides an SDE expression for the pullback of the radial vector field.

6.1. Pullback Process of Radial Field

In this section, we follow the theory and notation in [31, Section 2.4] and [34, Section 11.5]. If s denotes a local vector field on $\mathcal{M}$, we can define a map $s^{\mathcal{FM}}\colon \mathcal{FM} \to \mathbb{R}^d$ by $s^{\mathcal{FM}}(u) = u^{-1}s(\pi(u))$. Now, if x_t is a curve on $\mathcal{M}$ and w_t is the horizontal lift of x_t from w, we can let $s_t = w_{t,i}s_t^i$. Then $s^{FM}(w_t) = (s_t^1, \ldots, s_t^d)^T$ and

$$(s^{FM})_*(h_{w_t}(\dot{x}_t)) = w_t^{-1}\nabla_{\dot{x}_t}s = \frac{d}{dt}(s_t^1, \ldots, s_t^d)^T \ , \tag{22}$$

where h_w is the horizontal lift operator. In other words, the covariant derivative takes the form of the standard derivative applied to the frame coordinates s_t^i.

Lemma 10. *Let s be a local vector field on $\mathcal{M}$, $\xi_t = U_t^{-1}(s(\pi(U_t)))$, and $dU_t = H_i(U_t) \circ dZ_t^i$. Then on $\mathcal{M}\setminus \mathrm{Cut}(p)$*

$$d\xi_t = U_t^{-1}(\nabla_{U_te_i}s(\pi(U_t))) \circ dZ_t^i.$$

The corresponding Itô equation is given by

$$d\xi_t = U_t^{-1}(\nabla_{U_te_i}s(\pi(U_t)))\,dZ_t^i + \tfrac{1}{2}U_t^{-1}\big(\nabla_{U_te_j}\nabla_{U_te_i}s(\pi(U_t))\big)\,d[Z^j, Z^i]_t.$$

Proof. We can use the first equality in (22), with $\dot{x}_0 = ue$, for some $u \in \pi^{-1}(x_0)$, and $e \in \mathbb{R}^d$, and $w_0 = u$, to get

$$(s^{FM})_*(h_u(ue)) = u^{-1}\nabla_{ue}s = u^{-1}\nabla_{ue}\frac{\nabla d(v,\cdot)^2}{2} \ .$$

Now, $dU_t = H_i(U_t) \circ dZ_t^i$ and therefore

$$d\big(u^{-1}s(\pi(U_t))\big) = d\big(s^{FM}(U_t)\big) = (s^{FM})_*\big(H_i(U_t)\big) \circ dZ_t^i = U_t^{-1}\nabla_{U_te_i}s \circ dZ_t^i \ ,$$

since s^{FM} is a local diffeomorphism.

The second claim follows by an application of the first claim on the local vector field $\nabla_{U_te_i}s$ to get an sde for $U_t^{-1}(\nabla_{U_te_i}s(\pi(U_t)))$. $\square$

Let (D, ϕ) be a normal neighborhood centered at $v \in \mathcal{M}$. Define the function $g \colon \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by

$$g(t, r_v, Z, \xi) := \frac{r_v^2(X_t)}{T-t} \|\sigma_t^{-1}(Z_t)\xi_t\|^2, \tag{23}$$

assuming invertibility of σ. On the set $\phi(D)$ the function g is smooth. The following lemma provides an Itô expansion of g.

Lemma 11. *Let $\xi_t = U_t^{-1}(\nabla_{X_t} r_v(X_t))$. On the set $\phi(D)$, g as defined in* (23) *has the Itô SDE expression*

$$\begin{aligned} dg(t, r_v, Z, \xi) &= E(t)dt + \frac{r_v^2(X_t)}{T-t}\xi_t^T d\left(A(t, Z_t)\right)\xi_t + F(t)dr_v(X_t) + J_i(t)d\xi_t^i \\ &\quad + H_i(t)d[r_v(X_t), Z^i]_t + I_j(t)d[r_v(X_t), \xi^j]_t + K_{ij}(t)d[\xi^i, Z^j]_t \\ &\quad + \frac{1}{2}\left(G(t)d[r_v(X_t), r_v(X_t)]_t + J_{ij}(t)d[\xi^i, \xi^j]_t\right), \end{aligned}$$

where

$$\begin{aligned} E(t) &:= \frac{r_v^2(X_t)}{(T-t)^2}\|\sigma_t^{-1}(Z_t)\xi_t\|^2; & F(t) &:= 2\frac{r_v(X_t)}{T-t}\left\|\sigma_t^{-1}(Z_t)\xi_t\right\|^2; \\ G(t) &:= 2\frac{1}{T-t}\left\|\sigma_t^{-1}(Z_t)\xi_t\right\|^2; & H_j(t) &:= 2\frac{r_v(X_t)}{T-t}\xi_t^T\frac{\partial}{\partial z_j}\left(A(t, Z_t)\right)\xi_t; \\ I_j(t) &:= 4\frac{r_v(X_t)}{T-t}\xi_t^T A(t, Z_t)e_j; & J_j(t) &:= 2\frac{r_v^2(X_t)}{T-t}\xi_t^T A(t, Z_t)e_i; \\ J_{ij}(t) &:= 2\frac{r_v^2(X_t)}{T-t}e_j A(t, Z_t)e_j; & K_{ij}(t) &:= 2\frac{r_v^2(X_t)}{T-t}\xi^T\frac{\partial}{\partial z_j}\left(A(t, Z_t)\right)e_i. \end{aligned}$$

Proof. A basic consequence of the multidimensional Itô's formula on open sets. □

Remark 12. In the case where the diffusion parameter σ is the identity matrix, (23) simplifies to $g(t, r_v(X_t)) = r_v^2(X_t)/(T-t)$. In this case, the Itô expression for g, by stochastic integration by parts, is simply

$$\begin{aligned} dg(t, r_v(X_t)) &= \frac{r_v^2(X_t)}{(T-t)^2}dt + 2\frac{r_v(X_t)}{(T-t)}dr_v(X_t) + \frac{1}{(T-t)}dt \\ &= \frac{r_v^2(X_t)}{(T-t)^2}dt + \frac{1}{(T-t)}dr_v^2(X_t). \end{aligned}$$

Let (D, ϕ) be the exponential normal chart centered at v, where $\phi \colon D \subseteq \mathcal{M} \to \mathbb{R}^d$ is one-to-one. Define the process

$$\psi_t := \exp\left[-\frac{1}{2}\left\|\sigma_t^{-1}(\phi(y))u^{-1}\left(\frac{\nabla_y r_v(y)^2}{2(T-t)^{1/2}}\right)\right\|^2\right], \tag{24}$$

for $y \in D$. The following result is an adaptation of Lemma 7 in [9] to Riemannian manifolds.

Lemma 13. *Let (D, ϕ) be a normal neighborhood centered at v and $0 < t_1 < t_2 < ... < t_N < T$ and $g \in C_b^\infty(\mathcal{M})$ (smooth bounded function with compact support in $\mathcal{M}$), then with ψ_t as defined in* (24) *we have*

$$\lim_{t\to T} \frac{\mathbb{E}[g(x_{t_1}, ..., x_{t_N})\psi_t]}{\mathbb{E}[\psi_t]} = \mathbb{E}[g(x_{t_1}, ..., x_{t_N})|X_T = v].$$

Proof. First, since the cut locus of any complete connected manifold has (volume) measure zero, we can integrate indifferently in any exponential chart. For any $t \in (t_N, T)$, we have

$$\frac{\mathbb{E}[g(x_{t_1}, ..., x_{t_N})\psi_t]}{\mathbb{E}[\psi_t]} = \frac{\int_{\mathcal{M}} \Phi_g(t,z) e^{-\frac{1}{2}\left\|\sigma^{-1}(t,\phi(z))u^{-1}\left(\frac{\nabla r_v(z)^2}{2(T-t)^{1/2}}\right)\right\|^2} d\,\mathrm{Vol}(z)}{\int_{\mathcal{M}} \Phi_1(t,z) e^{-\frac{1}{2}\left\|\sigma^{-1}(t,\phi(z))u^{-1}\left(\frac{\nabla r_v(z)^2}{2(T-t)^{1/2}}\right)\right\|^2} d\,\mathrm{Vol}(z)},$$

where $d\,\mathrm{Vol}(z)$ denotes the volume measure,

$$\Phi_g(t,z) = \int_{\mathcal{M}^N} g(z_1, \ldots, z_N) p_{t_1}(u, z_1) \cdots p_{t-t_N}(z_N, z) d\,\mathrm{Vol}(z_1) \cdots d\,\mathrm{Vol}(z_N),$$

with $p_{t-s}(x,y) := p(s,x;t,y)$, and of course $\Phi_1(t,z) = p_t(x_0, z)$. We can write the expectation above as

$$\int_{\mathcal{M}} \Phi_g(t,z) e^{-\frac{1}{2}\left\|\sigma^{-1}(t,\phi(z))u^{-1}\left(\frac{\nabla r_v(z)^2}{2(T-t)^{1/2}}\right)\right\|^2} d\,\mathrm{Vol}(z) =$$
$$\int_{\phi(\mathcal{M})} \Phi_g(t, \phi^{-1}(x)) e^{-\frac{1}{2}\left\|\sigma^{-1}(t,x)u^{-1}\left(\frac{\nabla r_v(\phi^{-1}(x))^2}{2(T-t)^{1/2}}\right)\right\|^2} \sqrt{\det(G(\phi^{-1}(x)))} dx,$$

where G is the matrix representation of the Riemannian metric. As we are in a normal neighborhood, and $\{e_i\}$ is an orthonormal basis of T_vM, we have in particular $r_v(\phi^{-1}(x)) := d(\phi^{-1}(x), v) = \|\mathrm{Log}_v(\phi^{-1}(x))\| = \|\mathrm{Log}_v \circ \mathrm{Exp}_v \circ E(x)\| = \|E(x)\| = \|x\|$. Thus, if we apply the change of variable $x = (T-t)^{1/2} y$ we get

$$\begin{aligned}
&C_t \int_{\phi(\mathcal{M})} \Phi_g(t, \phi^{-1}(x)) e^{-\frac{1}{2}\left\|\sigma^{-1}(t,x)u^{-1}\left(\frac{\nabla r_v(\phi^{-1}(x))^2}{2(T-t)^{1/2}}\right)\right\|^2} \sqrt{\det(G(\phi^{-1}(x)))} dx \\
&= \int_{\phi(\mathcal{M})} \Phi_g(t, \phi^{-1}((T-t)^{\frac{1}{2}} y)) H(t,y) \sqrt{\det(G(\phi^{-1}((T-t)^{\frac{1}{2}} y)))} dy \\
&\overset{t\to T}{\to} \int_{\phi(\mathcal{M})} \Phi_g(T, \phi^{-1}(0)) H(T,y) \sqrt{\det(G(\phi^{-1}(0)))} dy \\
&= \Phi_g(T, v) \sqrt{\det(G(v))} \int_{\phi(\mathcal{M})} e^{-\frac{1}{2}\left\|\sigma^{-1}(T,0)u^{-1}\left(\frac{\nabla r_v(\phi^{-1}(y))^2}{2}\right)\right\|^2} dy,
\end{aligned}$$

where $C_t = (T-t)^{-d/2}$ and

$$H(t,y) = \exp\left[-\frac{1}{2}\left\|\sigma^{-1}\left(t, (T-t)^{\frac{1}{2}} y\right) u^{-1}\left(\frac{\nabla r_v(\phi^{-1}(y))^2}{2}\right)\right\|^2\right].$$

Therefore, we see that from Bayes formula that

$$\lim_{t\to T}\frac{\mathbb{E}[g(x_{t_1},...,x_{t_N})\psi_t]}{\mathbb{E}[\psi_t]}=\frac{\Phi_g(T,v)}{\Phi_1(T,v)}=\mathbb{E}[g(x_{t_1},...,x_{t_N})|X_T=v].$$

□

6.2. The General Case - Driving Semimartingale

We here provide the remaining arguments to make the proof of Theorem 3 rigorous.

Proof of Theorem 3. The Radon-Nikodym derivative (5) together with Novikov's condition ensures the equivalence of the measures of $X|X_T=v$ and Y on $[0,t]$, for every $t<T$. By Lemma 10 and Lemma 11, we obtain the expressions for $\{\varphi_t\colon t\in[0,T)\}$ and $\{\psi_t\colon t\in[0,T)\}$. The proof is concluded by Lemma 13. □

6.3. The Case of Driving Brownian Motion

Recall that in the case of $\tilde{\mathcal{L}}_z=\frac{1}{2}\Delta_M$, the term φ_t has the particular form $\varphi_t=\exp\left[\int_0^t\frac{r(Y_s)}{T-s}(d\eta_s+dL_s)\right]$, with $d\eta_s=\frac{\partial}{\partial r}\log\theta_v^{-\frac{1}{2}}ds$.

Proof of Theorem 5. When the driving semimartingale is a standard Brownian motion, we recall that the Radon-Nikodym derivative is given by (6). In this context the function g, defined by (23), for a Brownian motion X_t, reduces to the expression

$$g(t,X)=\frac{r_v(X_t)^2}{T-t}.$$

Recall that β_t is a one-dimensional Brownian motion. The geometric Itô's formula applied to this function then yields, coming from (16), the SDE

$$\begin{aligned}-\frac{1}{2}\int_0^t\frac{r_v^2(X_s)}{(T-s)^2}ds=&-\frac{1}{2}\frac{r_v^2(X_s)}{T-t}+\int_0^t\frac{r_v(X_s)}{T-s}d\beta_s\\&+\frac{1}{2}\int_0^t\frac{r_v(X_s)\Delta_{\mathcal{M}}r_v(X_s)+1}{T-s}ds-\int_0^t\frac{r_v(X_s)}{T-s}dL_s^{\mathring{C}}.\end{aligned}\tag{25}$$

We see that (25) substituted into (6) yields, for any bounded non-negative $\mathcal{B}_t(W(\mathcal{M}))$-measurable f,

$$\begin{aligned}&\mathbb{E}^{\mathbb{Q}}[f(Y)]\\&=C_t\mathbb{E}^{\mathbb{P}}\left[f(X)\exp\left[-\frac{r_v^2(X_t)}{2(T-t)}+\frac{1}{2}\int_0^t\frac{r_v(X_s)\Delta r_v(X_s)}{T-s}ds-\int_0^t\frac{r_v(X_s)}{T-s}dL_s^{\mathring{C}}\right]\right],\end{aligned}$$

$\mathbb{Q}_t$ - a.s., $t<T$, where we have used that $C_t=(T/(T-t))^{1/2}$. We can equivalently write the above as

$$\mathbb{E}^{\mathbb{Q}}[f(Y)\varphi_t]=\tilde{C}_t\mathbb{E}^{\mathbb{P}}[f(X)\psi_t],\tag{26}$$

where

$$\varphi_t = \exp\left[\int_0^t \frac{r_v(Y_s)}{T-s}\left(d\eta_s + dL_s^{\mathring{C}}\right)\right], \qquad \psi_t = \exp\left[-\frac{r_v(X_t)^2}{2(T-t)}\right],$$

and where $\tilde{C}_t = (T/(T-t))^{d/2}$. In order to see this, we mention the following expression, (see, e.g., [35])

$$\Delta r_v = \frac{d-1}{r_t} + \frac{\partial}{\partial r_v}\log\theta_v, \tag{27}$$

which hold on $\mathcal{M}\setminus\mathrm{Cut}(v)$, where θ_v is the Jacobian determinant of the exponential map. Therefore we can rewrite D_t as

$$D_t = \exp\left[-\frac{r_v(X_t)^2}{2(T-t)} + \frac{(d-1)}{2}\int_0^t \frac{1}{T-s}ds + \int_0^t \frac{r_v(X_s)}{T-s}\left(d\eta_s + dL_s\right)\right],$$

where $d\eta_t = \frac{\partial}{\partial r_v}\log\theta_v^{-\frac{1}{2}}ds$. Letting $f \equiv 1$ in equation (26) we obtain

$$\frac{\mathbb{E}^{\mathbb{Q}}[f(Y)\varphi_t]}{\mathbb{E}^{\mathbb{Q}}[\varphi_t]} = \frac{\mathbb{E}^{\mathbb{P}}[f(X)\psi_t]}{\mathbb{E}^{\mathbb{P}}[\psi_t]}.$$

By virtue of Lemma 13 we have, as $t \uparrow T$,

$$\frac{\mathbb{E}^{\mathbb{P}}[F(X)\psi_t]}{\mathbb{E}^{\mathbb{P}}[\psi_t]} \to \mathbb{E}^{\mathbb{P}}[F(X)|X_T = v].$$

What remains to be proven in Theorem 5 follows from Lemma 14. □

Lemma 14. *Assume that $\mathcal{M}$ is simply connected. With φ_t as defined above then $\varphi_t \xrightarrow{L_1} \varphi_T$.*

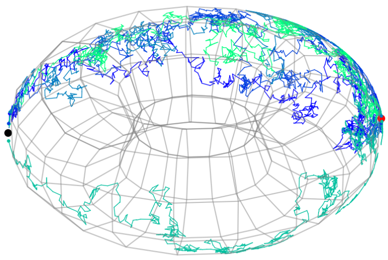
Proof. Note that for each $t \in [0,T)$ we have $\mathbb{E}^{\mathbb{Q}}[\varphi_t] < \infty$ as well as $\varphi_t \to \varphi_T$ almost surely by Proposition 9. The result then follows from the uniform integrability of $\{\varphi_t : t \in [0,T)\}$, which can be found in Appendix C.2 in [35]. □

7. Numerical Experiments

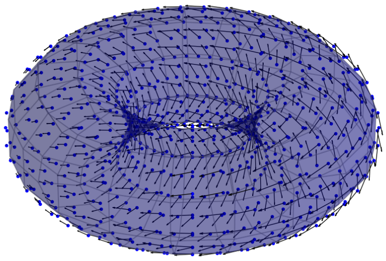
We devote this section to illustrating our simulation scheme on certain types of manifolds, including examples of two- and three-dimensional manifolds. The simulation scheme presents the behavior of the radial bridge process, which is represented in Figure 3-5. Furthermore, the radial bridge's simulation scheme is exploited for maximum likelihood estimation of diffusion mean values [15]. The code used to generate the illustrations is available at `https://bitbucket.org/stefansommer/theanogeometry`.

7.1. Simulation of Bridge Process on 2- and 3-Dimensional Manifolds

In the first example, we take our manifold $\mathcal{M}$ to be the 2-torus, i.e. $\mathcal{M} = \mathbb{T}^2 := \mathbb{S}^1 \times \mathbb{S}^1$. Our goal is to simulate a conditioned process on $\mathbb{T}^2$, where we condition a point in the cut locus of the initial point. The illustration can be found in Figure 3. Figure 3a shows four sample paths of the radial (Brownian) bridge with the initial starting value depicted by the red point. The process is conditioned to arrive at the black point - which is in the initial point's cut locus - at $T = 1$. Figure 3b presents the underlying squared radial vector field. The radial bridge's drift term follows the radial vector field multiplied by an increasing time-dependent scalar.



(a) Four sample paths from the simulation scheme of the radial bridge, X_t, from x (red point) to v (black point).



(b) Radial vector field on the torus, related to the radial bridge, X_t, centered at the point $X_T = v$.

Figure 3: Figure 3a Show four realizations of the radial bridge process, X_t, where the red point illustrates the start value, $X_0 = x$, and the black point the conditioned value, $X_T = v$. Figure 3bshows the underlying radial vector field structure.

In the second example, we consider another two-dimensional manifold, namely the cylinder $\mathcal{M} = \mathbb{S}^1 \times \mathbb{R}$. We show how our simulation scheme can be adapted to this framework. On the cylinder the cut locus of a point $v = (p, x) \in \mathbb{S}^1 \times \mathbb{R}$ is the set $\mathrm{Cut}(v) = \{q\} \times \mathbb{R}$, where $q \in \mathbb{S}^1$ is antipodal to p. This is illustrated in Figure 2.

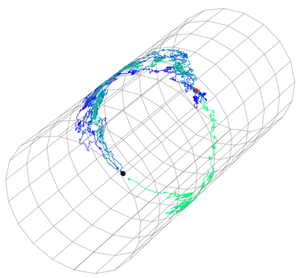


Figure 4: An example of four sample paths of the radial bridge conditioned to arrive at a point in the cut locus, $\mathrm{Cut}(x)$, of the initial point x.

Any winding around the cylinder trivially makes the process cross the cut locus. In Figure 4, we illustrate the radial bridge's sample paths to end up at the cut locus at the terminal time.

In the third example, we show how our simulation scheme works on a three-dimensional manifold. We consider the *special orthogonal group*, also known as the *rotation group*, which consists of all orthogonal matrices of determinant 1, i.e., $\mathcal{M} = \mathrm{SO}(3)$. Figure 5 shows a sample path of the radial bridge from $x \in SO(3)$ to $v \in SO(3)$, where $x = \mathrm{Id}_3$ (columns: red, blue, green vectors) and v (black) is the terminal value $X_T = v$.

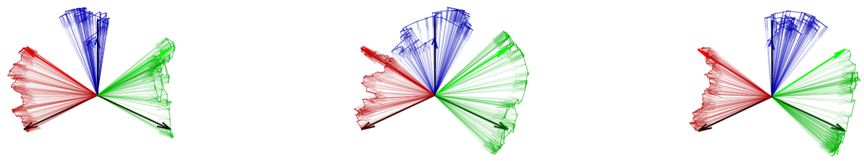


Figure 5: The figure illustrates sample paths from the radial bridge on $SO(3)$, by showing its left action on a basis of $\mathbb{R}^3$. The black arrows indicate the conditioned point.

7.2. Bridge Simulation for Density Estimation on 2-Sphere

As mentioned in [26], bridge sampling can estimate the transition density of a stochastic process. Assume that the family $\{\varphi_t \colon t \in [0, T)\}$ is uniformly integrable. As is seen from Lemma 13

$$\mathbb{E}[C_t\psi_t] \xrightarrow{t\to T} p(0,x_0;T,v)\sqrt{\det(G(v))}\int_{\phi(\mathcal{M})} e^{-\frac{1}{2}\left\|\sigma^{-1}(T,0)u^{-1}\left(\frac{\nabla d(\phi^{-1}(y),v)^2}{2}\right)\right\|^2} dy,$$

which in the case of the d-sphere can be expressed as

$$\begin{aligned}\mathbb{E}[C_t\psi_t] &\xrightarrow{t\to T} p(0,x_0;T,v)\sqrt{\det(G(v))}\int_{\mathbb{R}^d} e^{-\frac{1}{2}y^T A(T,0)y} dy\\ &= \frac{p(0,x_0;T,v)\sqrt{(2\pi)^2}\sqrt{\det\left(G(v)\right)}}{\sqrt{\det\left(A(T,0)\right)}},\end{aligned}$$

From the identity linking φ_t to ψ_t, this leads to the following expression for the transition density with respect to the Riemannian volume form

$$p(0,x_0;T,v) = \frac{\sqrt{\det\left(A(T,0)\right)}}{\sqrt{(2\pi T)^d}} e^{-\frac{1}{2}\left\|\sigma^{-1}(T,x_0)u^{-1}\left(\frac{\nabla d(x_0,v)^2}{2}\right)\right\|^2} \lim_{t\uparrow T}\mathbb{E}[\varphi_t].$$

In the Brownian motion setting, the above simplifies to

$$p(0,x_0;T,v) = \frac{1}{\sqrt{(2\pi T)^d}} e^{-\frac{r_v^2(x_0)}{2T}}\mathbb{E}[\varphi_T]. \tag{28}$$

Thus, we obtain an expression for the transition density of the Brownian motion.

Zhao and Song [37] provided an expression for the heat kernel on the hypersphere expanded as a uniformly and absolute convergent power series

$$p(0,x;t,y) = \sum_{l=0}^{\infty} e^{-l(l+d-1)t}\frac{2l+d-1}{d-1}\frac{1}{A_{\mathbb{S}^d}}C_l^{\frac{d}{2}-1}(\langle x,y\rangle_{\mathbb{R}^{d+1}}), \tag{29}$$

for $x, y \in \mathbb{S}^d$, where C_l^α are the Gegenbauer polynomials and $A_{\mathbb{S}^d} = \frac{2\pi^{(d+1)/2}}{\Gamma((d+1)/2)}$ the surface area of $\mathbb{S}^d$. In Figure 6, we have plotted the transition density estimated from (28) using sampling to approximate $\mathbb{E}[\varphi_T]$ against a truncated version ($l = 0, \ldots, 16$) of (29) as well as the density for the two-dimensional Euclidean Brownian motion. We have plotted the estimated densities along a geodesic running from the north pole to the south pole for three time points, $T = 0.5, 1, 1.5$.

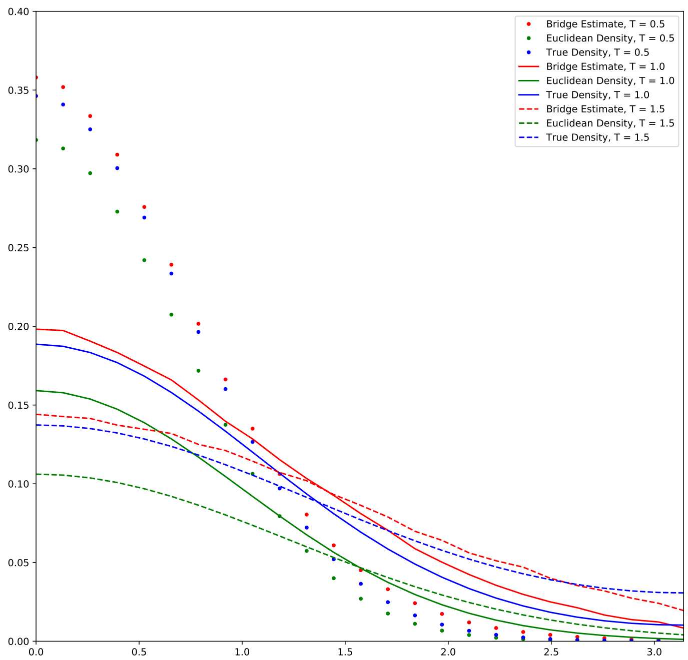


Figure 6: The figure shows estimated transition densities of a Brownian motion (red) on $\mathbb{S}^2$ using (28), run for different time length $T = 0.5, 1, 2$. The densities are computed along a geodesic from the north pole to the south pole. The estimated transition densities are compared to an approximation using (29) (blue) on $\mathbb{S}^2$ and densities of a 2-dimensional Euclidean Brownian motion (green).

When applying heat to the north pole, the heat diffuses symmetrically from the north pole (in the Riemannian sense) and the density converges to the uniform distribution on $\mathbb{S}^2$ as the time increases. The resulting estimated heat kernel on $\mathbb{S}^2$, using the sampling scheme above, is displayed in Figure 7.

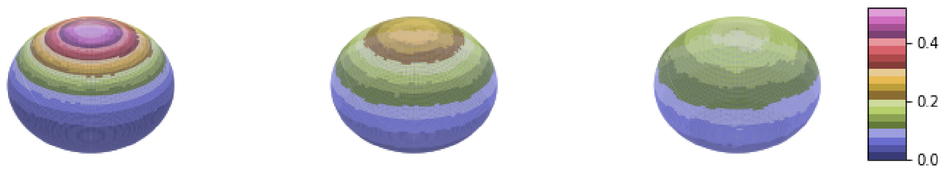


Figure 7: Estimated transition density on the 2-sphere at times $T = t$, for $t = 1/3, 2/3, 1$, respectively. We sampled 256 points as endpoints of a Brownian motion. For each observation, we sampled 50 bridges.

Figure 8 shows the estimated heat kernel on an ellipsoid where no closed-form solution is directly available.

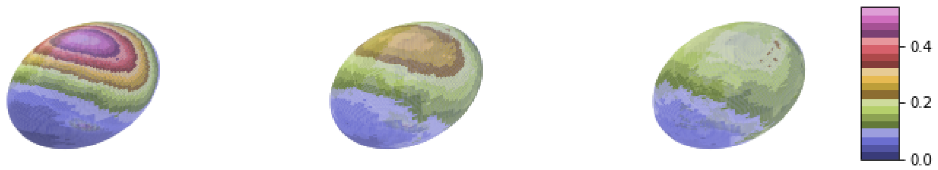
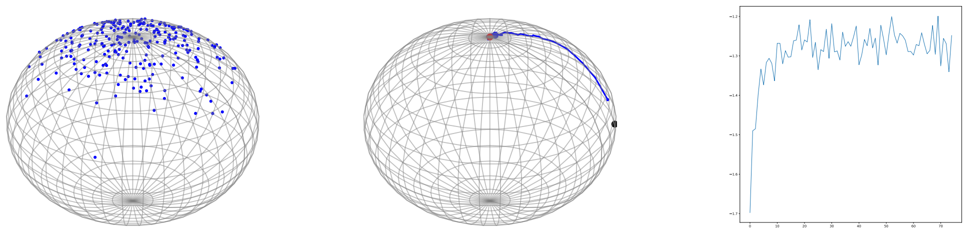


Figure 8: Estimated transition density on ellipsoid at times $T = t$, for $t = 1/3, 2/3, 1$, respectively. We sampled 256 points as endpoints of a Brownian motion. For each observation, we sampled 50 bridges.

7.3. Bridge Simulation for Diffusion Mean Estimation

The Fréchet mean is the set of points that minimizes the sum of distances, hence the Fréchet mean satisfies the geometric property of the Euclidean mean. The diffusion mean is an extension of and an alternative to the Fréchet mean (see Hansen et al. [15, 16] for an exposition on diffusion means). In contrast to the Fréchet mean, the diffusion mean satisfies the probabilistic property of the Euclidean mean, being the statistic that maximizes the likelihood of a Brownian motion. The diffusion mean-value relies explicitly on the transition density of a Brownian motion and thus contains a time parameter, which has the interpretation as the variance of the Brownian motion.

We now demonstrate how the bridge sampling scheme can be used to estimate diffusion means on the two-sphere. As described in the previous section, bridge sampling yields estimates of the underlying unknown transition density. Using (28), we approximate the transition density by bridges to each data point to obtain an approximation of the expectation over φ_T. We show how to obtain estimates of the mean of data resting on a manifold.



(a) 256 sampled data points on $\mathbb{S}^2$.

(b) Convergence to the diffusion mean (red dot) on $\mathbb{S}^2$ from initial guess (black dot).

(c) The iterated log-likelihood.

Figure 9: Our bridge sampling scheme estimates the diffusion mean on $\mathbb{S}^2$. We sampled 256 points on $\mathbb{S}^2$ as endpoints of a Brownian motion started at the north pole. Using an initial guess of the mean, we sampled one bridge per observation to estimate the likelihood and performed an iterative optimization to maximize the likelihood.

In figure 9a, we sampled 256 points on $\mathbb{S}^2$, as endpoints of a Brownian motion run until $T = 1/2$. Using our proposed bridge sampling scheme, we obtained approximations of the transition density function. Using an iterative maximum likelihood method, we obtained in each iteration an estimate of the mean as the point maximizing the log-likelihood. Figure 9b illustrates how the initial guess of the mean (black dot) converges to the true mean (red dot) using the iterative maximum likelihood optimization. The corresponding log-likelihood is illustrated in figure 9c, where we see how the iterated log-likelihood reaches an equilibrium state.

References

[1] Arnaudon, A., Holm, D.D., Pai, A., Sommer, S., 2017. A stochastic large deformation model for computational anatomy, in: International Conference on Information Processing in Medical Imaging, Springer. pp. 571–582.

[2] Arnaudon, A., Holm, D.D., Sommer, S., 2019. A geometric framework for stochastic shape analysis. Foundations of Computational Mathematics 19, 653–701.

[3] Arnaudon, A., van der Meulen, F., Schauer, M., Sommer, S., 2020. Diffusion bridges for stochastic hamiltonian systems with applications to shape analysis. arXiv preprint arXiv:2002.00885 .

[4] Barden, D., Le, H., 1997. Some consequences of the nature of the distance function on the cut locus in a riemannian manifold. Journal of the LMS , 369–383.

[5] Bladt, M., Finch, S., Sørensen, M., 2016. Simulation of multivariate diffusion bridges. Journal of the Royal Statistical Society: Series B: Statistical Methodology , 343–369.

[6] Bladt, M., Sørensen, M., et al., 2014. Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Bernoulli 20, 645–675.

[7] Bui, M.N., Pokern, Y., Dellaportas, P., 2021. Inference for partially observed riemannian ornstein–uhlenbeck diffusions of covariance matrices. arXiv preprint arXiv:2104.03193 .

[8] Clark, J., 1990. The simulation of pinned diffusions, in: 29th IEEE Conference on Decision and Control, IEEE. pp. 1418–1420.

[9] Delyon, B., Hu, Y., 2006. Simulation of conditioned diffusion and application to parameter estimation. Stochastic Processes and their Applications .

[10] Elworthy, D., 1988. Geometric aspects of diffusions on manifolds, in: École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87. Springer, pp. 277–425.

[11] Elworthy, K., Truman, A., 1982. The diffusion equation and classical mechanics: an elementary formula, in: Stochastic processes in quantum theory and statistical physics. Springer, pp. 136–146.

[12] Elworthy, K.D., 1982. Stochastic differential equations on manifolds. volume 70. Cambridge University Press.

[13] Emery, M., 1989. Stochastic calculus in manifolds. Springer.

[14] Gelman, A., Meng, X.L., 1998. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. Statistical science , 163–185.

[15] Hansen, P., Eltzner, B., Huckemann, S.F., Sommer, S., 2021a. Diffusion means in geometric spaces. arXiv preprint arXiv:2105.12061 .

[16] Hansen, P., Eltzner, B., Sommer, S., 2021b. Diffusion means and heat kernel on manifolds. arXiv preprint arXiv:2103.00588 .

[17] Hsu, E.P., 2002. Stochastic analysis on manifolds. volume 38. AMS.

[18] Hsu, P., 1990. Brownian bridges on Riemannian manifolds. Probability Theory and Related Fields 84, 103–118.

[19] Jensen, M.H., Mallasto, A., Sommer, S., 2019. Simulation of conditioned diffusions on the flat torus, in: International Conference on Geometric Science of Information, Springer. pp. 685–694.

[20] Kendall, W.S., 1987. The radial part of brownian motion on a manifold: a semimartingale property. The Annals of Probability 15, 1491–1500.

[21] Le, H., Barden, D., 1995. Itô correction terms for the radial parts of semimartingales on manifolds. Probability theory and related fields 101, 133–146.

[22] Li, X.M., 2016a. Generalised brownian bridges: examples. arXiv preprint arXiv:1612.08716 .

[23] Li, X.M., 2016b. On hypoelliptic bridge. Electron. Commun. Probab. 21, 12 pp.

[24] Malliavin, P., Stroock, D.W., 1996. Short time behavior of the heat kernel and its logarithmic derivatives. Journal of Differential Geometry 44, 550–570.

[25] van der Meulen, F., Schauer, M., et al., 2017. Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. Electronic Journal of Statistics 11, 2358–2396.

[26] Papaspiliopoulos, O., Roberts, G., 2012. Importance sampling techniques for estimation of diffusion models. Statistical methods for stochastic differential equations , 311–340.

[27] Pennec, X., 2020. Manifold-valued image processing with spd matrices, in: Riemannian geometric statistics in medical image analysis. Elsevier, pp. 75–134.

[28] Pennec, X., Sommer, S., Fletcher, T., 2019. Riemannian Geometric Statistics in Medical Image Analysis. Academic Press.

[29] Revuz, D., Yor, M., 1999. Continuous Martingales and Brownian motion. Springer.

[30] Schauer, M., Van Der Meulen, F., Van Zanten, H., et al., 2017. Guided proposals for simulating multi-dimensional diffusion bridges. Bernoulli 23, 2917–2950.

[31] Sommer, S., 2016. Anisotropically weighted and nonholonomically constrained evolutions on manifolds. Entropy 18, 425.

[32] Sommer, S., Arnaudon, A., Kuhnel, L., Joshi, S., 2017. Bridge simulation and metric estimation on landmark manifolds, in: Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics. Springer, pp. 79–91.

[33] Stroock, D.W., Turetsky, J., 1997. Short time behavior of logarithmic derivatives of the heat kernel. Asian Journal of Mathematics 1, 17–33.

[34] Taubes, C.H., 2011. Differential geometry: Bundles, connections, metrics and curvature. volume 23. OUP Oxford.

[35] Thompson, J., 2015. Submanifold bridge processes. Ph.D. thesis. University of Warwick.

[36] Thompson, J., 2018. Brownian bridges to submanifolds. Potential Analysis 49, 555–581.

[37] Zhao, C., Song, J.S., 2018. Exact heat kernel on a hypersphere and its applications in kernel svm. Frontiers in Applied Mathematics and Statistics 4, 1.

Chapter 3

Bridge Simulation and Metric Estimation on Lie Groups and Homogeneous Spaces

The following manuscript has been made in collaboration with Lennard Hilgendorf, Sarang Joshi, and Stefan Sommer and has been published on ArXiv as submission

> Mathias Højgaard Jensen, Lennard Hilgendorf, Sarang Joshi, Stefan Sommer. "*Bridge Simulation and Metric Estimation on Lie Groups and Homogeneous Spaces*" *arXiv:2106.03431*, (2022).

The manuscript is an extension of Jensen et al. [28], which was accepted and published in the conference proceedings at the international conference of Geometric Science of Information. We derive a bridge simulation scheme on Lie groups and homogeneous spaces. The guiding term used in the previous paper applies in the context of Lie groups as well. Conditioning on a fiber in the total space (Lie group) over the base space (homogeneous space), leads to a bridge simulation scheme in the base space. This is a consequence of the idea by Thompson [48] when conditioning on a submanifold. The current paper moreover generalizes the paper by Jensen and Sommer [29]. We use the simulation scheme on the total and base space to obtain metric estimation on both these spaces.

Article

Bridge simulation and metric estimation on Lie groups and homogeneous spaces

Mathias Højgaard Jensen [1,†,‡], Lennard Hilgendorf [1,†,‡], Sarang Joshi [2,‡], and Stefan Sommer [1,‡]0000-0001-6784-0328

[1] Department of Computer Science, University of Copenhagen; {matje,lenhil,sommer}@di.ku.dk
[2] Department of Biomedical Engineering, University of Utah; sjoshi@sci.utah.edu
† Current address: Universitetsparken 5, DK-2100, Copenhagen E, Denmark
‡ These authors contributed equally to this work.

Abstract: We present schemes for simulating Brownian bridges on complete and connected Lie groups and homogeneous spaces. We use this to construct an estimation scheme for recovering an unknown left- or right-invariant Riemannian metric on the Lie group from samples. We subsequently show how pushing forward the distributions generated by Brownian motions on the group results in distributions on homogeneous spaces that exhibit non-trivial covariance structure. The pushforward measure gives rise to new parametric families of distributions on commonly occurring spaces such as spheres and symmetric positive tensors. We extend the estimation scheme to fit these distributions to homogeneous space-valued data. We demonstrate both the simulation schemes and estimation procedures on Lie groups and homogenous spaces, including $\mathrm{SPD}(3) = \mathrm{GL}_+(3)/\mathrm{SO}(3)$ and $\mathbb{S}^2 = \mathrm{SO}(3)/\mathrm{SO}(2)$.

Keywords: bridge simulation, Brownian motion, Lie groups, homogeneous spaces, metric estimation, directional statistics

1. Introduction

Bridge simulation is a data augmentation technique for generating missing trajectories of continuous diffusion processes. We consider bridge simulation on Lie groups and homogeneous spaces. As an important example, we investigate the case of i.i.d. Lie group or homogeneous space valued samples that are considered discrete-time observations of a continuous diffusion process. Assuming the stochastic dynamics to be Brownian motion, we wish to estimate the underlying Riemannian metric of the Lie group or homogeneous space from the samples. To evaluate and maximize the likelihood of the data, we need to account for the diffusion process being unobserved at most time points. This requires bridge sampling, and the sampling techniques are thus the key enabler for metric estimation in this setting.

Simulation of conditioned diffusion processes is a highly non-trivial problem, even in Euclidean spaces. Because transition densities of diffusion process are only available in closed form for a narrow range of processes, simulating directly from the true bridge distribution is generally infeasible. The data augmentation used in inference for diffusion processes dates back to the seminal paper by Pedersen [1] almost three decades ago. Since then, several papers have described diffusion bridge simulation methods; see, e.g., [2–11]. The method by Delyon and Hu [5] exchanged the intractable drift term in the conditioned diffusion with a tractable drift originating from the drift of a Brownian bridge. Several papers have built on the ideas of Delyon and Hu. In particular, a manifold equivalent drift term analogous to the drift term of a Brownian bridge in Euclidean space was used in the paper [6] to describe the simulation of Brownian bridges on the flat torus, while [7] generalized this method to Riemannian manifolds. [11] used the drift to model Brownian bridges on the space of landmarks, and Bui et al. [4] used a similar drift term on the space of symmetric positive definite (covariance) matrices. The authors of the latter paper exploits the exponential map, which in the space of covariance matrices is a global diffeomorphism avoiding the cumbersome geometric local time. The details are further described in [12].

Citation: Jensen et al. Bridge simulation and metric estimation on Lie groups and homogeneous spaces. *Preprints* **2022**, *1*, 0. https://doi.org/

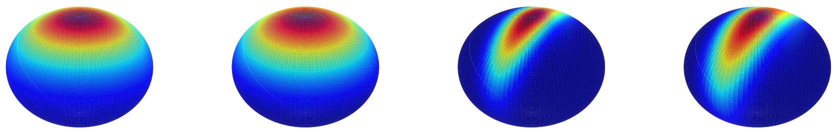


Figure 1. The two leftmost plots visualize the transition densities of a Fisher-Bingham-Kent distribution (left) and the pushforward density of a Brownian motion to $\mathbb{S}^2$ (right) with a bi-invariant metric. The Fisher-Bingham-Kent distribution has been fitted to samples from a Brownian motion using MLE. The pushforward measure of a Brownian motion on $\mathrm{SO}(3)$ to the sphere $\mathbb{S}^2$ results in anisotropic distributions on $\mathbb{S}^2$ when the metric on $\mathrm{SO}(3)$ is not bi-invariant, here displayed in the two rightmost plots, for $T = 0.5$ and $T = 1.0$, respectively. The coloring indicate the density of the pushforward at different times (different color scheme for each subfigure).

The idea of the present paper is based on the method presented in the paper by Delyon and Hu [5], and, in the geometric setting, the paper [7]. When conditioning a diffusion with transition density p_t to hit a point v at time $T > 0$, the intractable guiding drift term $\nabla_{x=X_t} \log p_{T-t}(x, v)$ in the stochastic differential equation (SDE) for the conditioned diffusion can be exchanged with the guiding drift term in the SDE for a Brownian bridge. This paper extends this idea to Lie groups and homogeneous spaces.

As an application, we consider discrete-time observations in Lie groups and homogeneous spaces regarded as incomplete observations of sample paths of Brownian motions arising from left- (or right-) invariant Riemannian metrics. The bridge simulation schemes allow to interpolate between the discrete-time observations. Furthermore, we observe how varying the metric on Lie groups affects pushforwards of the Brownian motion to homogeneous spaces being quotients of the group. These distributions encode covariance of the data resulting from the metric structure of the Lie group. We define this family of distributions and derive estimation schemes for recovering the metric structure of the group both in the case of Lie group samples and in the case of homogeneous space samples. One particular example is the two-sphere, $\mathbb{S}^2 \cong \mathrm{SO}(3)\backslash\mathrm{SO}(2)$. Changing the metric structure on $\mathrm{SO}(3)$ results in anisotropic distributions on $\mathbb{S}^2$, arising as the pushforward measure from $\mathrm{SO}(3)$. Figure 1 illustrates the isotropic and anisotropic distributions on $\mathbb{S}^2$ induced by a bi-invariant and left-invariant (non-invariant) metric on $\mathrm{SO}(3)$, respectively. The resulting distributions are analogous to the Von Mises-Fisher and Fisher-Bingham distributions [13,14]. However, unlike the Von Mises-Fisher and Fisher-Bingham distributions, the approach is independent of the chosen embedding instead resulting from the quotient structure.

For simulation on homogeneous spaces, we present three schemes. The first builds on the idea of Thompson [15] by conditioning on a submanifold in the Lie group G obtained as a fiber over the point $v \in M = G/K$, for some closed subgroup $K \subseteq G$. The second scheme assumes the homogeneous space has a discrete fiber Γ and therefore, the fiber over $v \in M$, $\pi^{-1}(v)$, is discrete. Using the k-nearest-points from the fiber $\pi^{-1}(v)$ to the initial point x_0, we obtain a truncated guiding drift term convergence to a subset of $\pi^{-1}(v)$. The last scheme assumes that the fiber is connected. By sampling a finite number of points in the fiber over v, a similar conditioning is obtained.

Statistics on Lie groups and homogeneous spaces finds applications in many diverse fields including bioinformatics, medical imaging, shape analysis, computer vision, and information geometry, see, e.g., [16–20]. Statistics in Euclidean spaces often relies on the distributional properties of the normal distribution. Here we use Brownian motions and the heat equation to generalize the normal distribution to Lie groups and homogeneous spaces as introduced by Grenander [21]. The solution to the heat equation is the transition density of a Brownian motion. Through Monte Carlo simulations of bridges, we can estimate the transition density and maximize the likelihood with respect to the Riemannian metric.

1.1. Contribution and Overview

We present simulation schemes on Lie groups and homogeneous spaces with application to parameter estimation. We outline the necessary theoretical background for the construction of bridge simulation on Lie groups and homogeneous spaces before demonstrating how the simulation scheme leads to estimates of means and underlying metric structure using maximum likelihood estimation on certain Lie groups and homogeneous spaces. The paper builds on and significantly extends the conference paper [7] that introduced bridge simulation in the Lie group setting.

The paper is organized as follows. In Section 2, we describe the relevant background theory of Lie groups, Brownian motions, and Brownian bridges in Riemannian manifolds. Section 3 presents the theory and results of bridge sampling in Lie groups, while Section 4 introduces bridge sampling on various homogeneous spaces. Numerical experiments on certain Lie groups and homogeneous spaces are presented in 6.

2. Notation and background

We here briefly describe simulating conditioned diffusions in $\mathbb{R}^n$ as developed in [5] before reviewing the theory on conditioned diffusion on Riemannian manifolds.

2.1. Euclidean diffusion bridges and simulation

Suppose a strong solution to an SDE of the form

$$dx_t = b(t, x_t)dt + \sigma(t, x_t)dw_t,$$

where b and σ satisfies certain regularity conditions and where w denote a $\mathbb{R}^n$-valued Brownian motion. In this case, x is Markov, and its transition density exists. Suppose we define the function

$$h(t, x) = \frac{p_{T-t}(x_t, v)}{p_T(x_0, v)},$$

for some $x_0, v \in \mathbb{R}^n$. Then it is easily derived that h is a martingale on $[0, T)$ with $h(0, x_0) = 1$ and Doob's h-transform implies that the SDE of the conditioned diffusion $x|x_T = v$ is given by

$$dy_t = \tilde{b}(t, y_t)dt + \sigma(t, y_t)dw_t$$

where $\tilde{b}(t, y) = b(t, y) + (\sigma\sigma^T)(t, y)\nabla_y \log p_{T-t}(y, v)$. In case that the transition density is intractable, simulation from the exact distribution is in-feasible. Delyon and Hu [5] suggested substituting the latter term in $\tilde{b}$ with a drift term of the form $-(y_t - v)/(T - t)$, which equals the drift term in a Brownian bridge. The guided process obtained by making the above substitution yields a conditioning and one obtain

$$\mathbb{E}[f(x)|x_T = v] = C\mathbb{E}[f(y)\varphi_T(y)], \tag{1}$$

where φ_T is a likelihood function that is tractable and easy to compute, y is the guided process, and the constant $C > 0$ depends on x_0, v, and T.

2.2. Riemannian manifolds and Lie groups

Let M be a finite dimensional smooth manifold of dimension d. M can be endowed with a Riemannian metric tensor, i.e., a family of inner products $\{\langle\cdot,\cdot\rangle_x\}_{x\in M}$ defined on each tangent space T_xM. The Riemannian metric tensor gives rise to a distance function between points in M. The tangent space is locally diffeomorphic with an open subset of M. The Riemannian exponential map $\mathrm{Exp}_x\colon T_xM \to M$ provides this local diffeomorphism. On the subset of M where Exp_x is a diffeomorphism the inverse Riemannian exponential map, also called the Riemannian logarithm map, $\mathrm{Log}_x\colon M \to T_xM$ is defined. The Riemannian distance function can then be defined in terms of the Riemannian inner product as $d(x,y) = \|\mathrm{Log}_x(y)\|_x$. The Riemannian logarithm map plays an important role when defining guided bridges on manifolds.

Let X be a vector field on M assigning to each point $X \in M$ a tangent vector $X(x) \in T_xM$. A connection ∇ on a manifold is an operation that allows us to compare neighboring tangent spaces and define derivatives of vector fields along other vector fields, that is, if Y is another vector field, then $\nabla_X Y$ is the derivative of Y along X (also known as the covariant derivative of Y along X). A connection also gives a notion of "straight lines" in manifolds, also known as geodesics. A curve γ is a geodesic if the vector field along γ is parallel to itself, i.e, if $\nabla_{\dot\gamma_t}\dot\gamma_t = 0$. The geodesic curves are locally length minimizing.

Generalizing the Euclidean Laplacian operator, the Laplace-Beltrami operator is defined as the divergence of the gradient, $\Delta_M f = \operatorname{div}\operatorname{grad} f$. In terms of local coordinates $(x_1, \dots, x_d)$ the expression for the Laplace-Beltrami operator becomes

$$\Delta_M f = \det(g)^{-1/2}\left(\frac{\partial}{\partial x_j} g^{ji} \det(g)^{1/2} \frac{\partial}{\partial x_i}\right) f, \tag{2}$$

where $\det(g)$ denotes the determinant of the Riemannian metric g and g^{ij} are the coefficients of the inverse of g. (2) can be written as

$$\Delta_M f = a^{ij} \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f + b^j \frac{\partial}{\partial x_j} f, \tag{3}$$

where $a^{ij} = g^{ij}$, $b^k = -g^{ij}\Gamma^k_{ij}$, and Γ denote the Christoffel symbols of the Riemannian metric.

2.3. Lie groups

Let G denote a connected Lie group of dimension d, i.e., a smooth manifold with a group structure such that the group operations $G \times G \ni (x,y) \overset{\mu}{\mapsto} xy \in G$ and $G \ni x \overset{\iota}{\mapsto} x^{-1} \in G$ are smooth maps. If $x \in G$, the left-multiplication map, $L_x y$, defined by $y \mapsto \mu(x,y)$, is a diffeomorphism from G to itself. Similarly, the right-multiplication map $R_x y$ defines a diffeomorphism from G to itself by $y \mapsto \mu(y,x)$. Let $dL_x\colon TG \to TG$ denote the pushforward map given by $(dL_x)_y\colon T_yG \to T_{xy}G$. A vector field V on G is said to be left-invariant if $(dL_x)_y V(y) = V(xy)$. The space of left-invariant vector fields is linearly isomorphic to T_eG, the tangent space at the identity element $e \in G$. By equipping the tangent space T_eG with the Lie bracket we can identify the Lie algebra $\mathfrak{g}$ with T_eG. The group structure of G makes it possible to define an action of G on its Lie algebra $\mathfrak{g}$. The conjugation map $C_x := L_x \circ R_x^{-1}\colon y \mapsto xyx^{-1}$, for $x \in G$, fixes the identity e. Its pushforward map at e, $(dC_x)_e$, is then a linear automorphism of $\mathfrak{g}$. Define $\mathrm{Ad}(x) := (dC_x)_e$, then $\mathrm{Ad}\colon x \mapsto \mathrm{Ad}(x)$ is the adjoint representation of G in $\mathfrak{g}$. The map $G \times \mathfrak{g} \ni (x,v) \mapsto \mathrm{Ad}(x)v \in \mathfrak{g}$ is the adjoint action of G on $\mathfrak{g}$. We denote by $\langle\cdot,\cdot\rangle$ a Riemannian metric on G. The metric is said to be left-invariant if $\langle u,v\rangle_y = \big\langle (dL_x)_y u, (dL_x)_y v\big\rangle_{L_x(y)}$, for every $u,v \in T_yG$, i.e., the left-multiplication maps are isometries, for every $x \in G$. The metric is $\mathrm{Ad}(G)$-invariant if $\langle u,v\rangle_e = \langle \mathrm{Ad}(x)u, \mathrm{Ad}(x)v\rangle_e$, for every $u,v \in \mathfrak{g}$. Note that an $\mathrm{Ad}(G)$-invariant metric on G is equivalent to a bi-invariant (left- and right-invariant) inner product on $\mathfrak{g}$. The differential of the Ad map at the identity yields a linear map $\mathrm{ad}(x) = \frac{d}{dt}\mathrm{Ad}(\exp(tx))|_0$. This linear map is equal to the Lie bracket $[v,w] = \mathrm{ad}(v)w$, $v,w \in \mathfrak{g}$.

A one-parameter subgroup of G is a continuous homomorphism $\gamma\colon (\mathbb{R},+) \to G$. The Lie group exponential map $\exp\colon \mathfrak{g} \to G$ is defined as $\exp(v) = \gamma_v(1)$, for $v \in \mathfrak{g}$, where γ_v is the unique one-parameter subgroup of G whose tangent vector at e is v. For matrix Lie groups the exponential map has the particular form: $\exp(A) = \sum_{k=0}^{\infty} A^k/k!$, for a square matrix A. The resulting matrix $\exp(A)$ is an invertible matrix. Given an invertible matrix B, if there exists a square matrix A such that $B = \exp(A)$, then A is said to be the logarithm of B. In general, the logarithm might not exist and if it does it may fail to be unique. However, the matrix exponential and logarithms can be computed numerically efficient (see [22, Chapter 5] and references therein). In a neighborhood sufficiently close to the identity the Lie group logarithm exist and is unique. By means of left-translation (or right-translation), the Lie group exponential map can be extended to a map $\exp_g\colon T_gG \to G$, for all $g \in G$, defined by $\exp_g(v) = g\exp(dL_{g^{-1}}v)$. Similarly, the Lie group logarithm at g becomes $\log_g(v) = dL_g \log(g^{-1}v)$.

Example 1. *A few examples of Lie groups include the Euclidean space* $(\mathbb{R}^n, +)$ *with the additive group structure,* $(\mathbb{R}_+, \cdot)$ *the positive real line with a multiplicative group structure, the space of invertible real matrices* $\mathrm{GL}(n)$ *equipped with a multiplication of matrices forms a Lie group, and the rotation group* $\mathrm{O}(n)$*, consisting of real orthogonal matrices with determinant one or minus 1 forms a subgroup of* $\mathrm{GL}(n)$*.*

The identification of the space of left-invariant vector fields with the Lie algebra $\mathfrak{g}$ allows for a global description of Δ_G. Indeed, let $\{v_1, \ldots v_d\}$ be an orthonormal basis of T_eG. Then $V_i(g) = (dL_g)_e v_i$ defines left-invariant vector fields on G and the Laplace-Beltrami operator can be written as (cf. [23, Proposition 2.5])

$$\Delta_G f(e) = \sum_{i=1}^{d} V_i^2 f(e) - V_0 f(e),$$

where $V_0 = \sum_{i,j=1}^{d} C_{ij}^{j} V_j$ and C_{ij}^{k} denote the structure coefficients given by

$$[V_i, V_j] = C_{ij}^{k} V_k. \tag{4}$$

By the left-invariance, the formula for the Laplace-Beltrami operator holds globally, i.e., $\Delta_G f(g) = \Delta_G f \circ L_g(e) = \left(dL_g\right)_e \Delta_G f(e)$.

2.4. Homogeneous spaces

A homogeneous space is a particular type of quotient manifold that arises as a smooth manifold endowed with a transitive smooth action by a Lie group G. The homogeneous space is called a G-homogeneous space to indicate the Lie group action. All G-homogeneous spaces arise as a quotient manifold G/H, for some closed subgroup $H \subseteq G$. H is a closed subgroup of the Lie group G which makes H into a Lie group. Any homogeneous space is diffeomorphic to the quotient space G/G_x, where G_x is the stabilizer for the point x. The dimension of the G-homogeneous space is equal to $\dim G - \dim H$ the quotient map $\pi\colon G \to G/H$ is a smooth submersion, i.e., the differential of π is surjective at every point. This implies that the fibers $\pi^{-1}(x)$, $x \in M$ are embedded submanifolds of G. We assume throughout that G acts on itself by left-multiplication.

Example 2. *The rotation group* $\mathrm{SO}(n)$ *acts transitively on* $\mathbb{S}^{n-1}$*, therefore* $\mathbb{S}^{n-1}$ *is a* $\mathrm{SO}(n)$*-homogeneous space. Consider a point in* $\mathbb{S}^{-1}$ *as a vector in* $\mathbb{R}^n$*. Rotations that fix the point occur precisely in the subspace orthogonal to the vector. Thus, the stabilizer or isotropy group is the rotation group* $\mathrm{SO}(n-1)$ *and* $\mathbb{S}^{n-1} = \mathrm{SO}(n)/\mathrm{SO}(n-1)$*.*

The set of invertible matrices with positive determinant $\mathrm{GL}_+(n)$ *acts on symmetric positive definite matrices* $\mathrm{SPD}(n)$*. The isotropy group is the rotation group* $\mathrm{SO}(n)$ *and thus* $\mathrm{SPD}(n) = \mathrm{GL}_+(n)/\mathrm{SO}(n)$*.*

A particular type of homogeneous space arises when the subgroup is a discrete subgroup of G*. For example, the space* $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ *defines the* n*-torus as a homogeneous space.*

2.5. Brownian motion on Riemannian manifolds

The Laplacian defines Brownian motion on M as a $\frac{1}{2}\Delta_M$-diffusion process up to its explosion time τ. The stochastic differential equation (SDE) for a Brownian motion X_t in local coordinates is

$$dX_t^k = -\frac{1}{2}g^{ij}(X_t)\Gamma_{ij}^k(X_t)dt + \sigma_j^k(X_t)dB_t^j, \tag{5}$$

where $\sigma = \sqrt{g^{-1}}$ is the matrix square root of g^{-1}.

On Lie groups, an SDE for a Brownian motion on G in terms of left-invariant vector fields takes the form

$$dg_t = -\frac{1}{2}V_0(g_t)dt + V_i(g_t) \circ dB_t^i, \qquad g_0 = e, \tag{6}$$

where $\circ$ denotes integration in the Stratonovich sense. By [23, Proposition 2.6], if the inner product is $\mathrm{Ad}(G)$ invariant, then $V_0 = 0$. The solution of (6) is conservative or non-explosive and is called the left-Brownian motion on G (see [24] and references therein).

2.6. Brownian bridges

In this section, we briefly review some facts on Brownian bridges on Riemannian manifolds, including Lie groups. On Lie groups, the existence of left-invariant (resp. right-invariant) vector fields allows identification of the Lie algebra with the vector space of left-invariant vector fields making the Lie group parallelizable. This allows constructing general semimartingales directly on the Lie groups.

Let $\mathbb{P}_x^t = \mathbb{P}_x|_{\mathcal{F}_t}$ be the measure of a Riemannian Brownian motion, X_t, at some time t started at point x. Let p_t denote the transition density of X_t so that $d\mathbb{P}_x^t = p_t(x,y)d\,\mathrm{Vol}(y)$ with $d\,\mathrm{Vol}(y)$ the Riemannian volume measure. Conditioning the Riemannian Brownian motion to hit some point $v \in M$ at time $T > 0$ defines a Riemannian Brownian bridge. We let $\mathbb{P}_{x,v}^T$ denote the corresponding probability measure. The two measures are absolutely continuous (equivalent) over the time interval $[0,T)$, however mutually singular at time $t = T$. This is an obvious consequence of the fact that $\mathbb{P}_x(X_T = v) = 0$, whereas $\mathbb{P}_{x,v}^T(X_T = v) = 1$. The corresponding Radon-Nikodym derivative is

$$\frac{d\mathbb{P}_{x,v}^T}{d\mathbb{P}_x}\Big|_{\mathcal{F}_s} = \frac{p_{T-s}(X_s,v)}{p_T(x,v)} \qquad \text{for } 0 \le s < T, \tag{7}$$

which is a martingale for $s < T$. The Radon-Nikodym derivative defines the density for the change of measure, and it provides the conditional expectation

$$\mathbb{E}[F(X_t)|X_T = v] = \frac{\mathbb{E}[p_{T-t}(X_t,v)F(X_t)]}{p_T(x,v)}, \tag{8}$$

for any bounded and $\mathcal{F}_s$-measurable random variable $F(X_s)$. The Brownian bridge is a non-homogeneous diffusion on M with infinitesimal generator

$$\mathcal{L}_s f(z) = \frac{t}{2}\Delta_M f(z) + t\nabla_z \log p_{t(1-s)}(z,v) \cdot \nabla f(z).$$

The bridge can be described by an SDE in the frame bundle FM of M. Let U_t be a lift of $X_t = \pi_{FM}(U_t)$ and, using the horizontal vector fields $H_i, \dots, H_d$, we have

$$dU_t = H_i(U_t) \circ \left(dB_t^i + \left(U_t^{-1}\left(\pi_*\left(\nabla^H_{u|u=U_t} \log \tilde{p}_{T-t}(u,v)\right)\right)\right)^i dt\right), \quad U_0 = u_0, \tag{9}$$

where $\tilde{p}_t(u,v) = p_t(\pi(u),v)$ denotes the lift of the transition density, B is an $\mathbb{R}^d$-valued Brownian motion, and $(\pi_{FM})_*\colon T\mathcal{FM} \to TM$ is the pushforward of the projection $\pi_{FM}\colon \mathcal{FM} \to M$. Her $u_0 \in FM$ is an orthonormal frame such that $\pi_{FM}(u_0) = x_0$.

It is possible to simulate from the conditioned process directly for specific homogeneous spaces. As an example, we mention the case of the flat torus $\mathbb{T}^2$ considered a homogeneous space of $\mathbb{R}^2$ with fibers the set of integers $\mathbb{Z}^2$. In this case, a Brownian motion in $\mathbb{T}^2$ conditioned at a point in $\mathbb{T}^2$ lifts to a bridge in $\mathbb{R}^2$ conditioned on a set of points isomorphic to $\mathbb{Z}^2$.

A generalization of Riemannian Brownian bridges can be found in Thompson [15]. Brownian bridges to submanifolds are here introduced by considering the transition density on a Riemannian manifold M defined by

$$p_t(x,N) := \int_N p_t(x,y) d\,\mathrm{Vol}_N(y), \tag{10}$$

where $N \subset M$ is a submanifold of M and Vol_N denotes the volume measure on N. These processes are denoted *Fermi bridges*. They have infinitesimal generator

$$\tfrac{1}{2}\Delta - \tfrac{r_N}{T-t}\tfrac{\partial}{\partial r_N}, \tag{11}$$

where $r_N(\cdot) := d(\cdot,N) = \inf_{y\in N} d(\cdot,y)$ and $\frac{\partial}{\partial r_N} = \nabla d(\cdot,N)$. The resulting conditional expectation becomes

$$\mathbb{E}[f(X_t)|X_T \in N] = \frac{\mathbb{E}[p_{T-t}(X_t,N)f(X_t)]}{p_T(x,N)}, \tag{12}$$

which holds for all bounded $\mathcal{F}_t$-measurable random variables $f(X_t)$. [25] exploited the above idea to estimate diffusion means on manifolds by conditioning on the diagonal of a product manifold. In the current paper, the fibers of homogeneous spaces are embedded submanifolds of a Lie group and a simulation scheme on homogeneous spaces are obtained by conditioning on the fibers.

2.7. One-point motions

Consider the homogeneous space $M = G/H$, where H is a Lie subgroup of the Lie group G and let $\pi\colon G \to M$ denote the canonical projection. Suppose that G acts on M on the left and that g_t is a process in G. As described in Liao [23], we obtain an induced process in M induced by the process g_t in G. For any $x \in M$, the induced process $x_t = g_t x$ defines the one-point motion of g_t in M, with initial value x.

The one-point motion, $X_t = g_t x$, of a Brownian motion g_t in G, started at $g_0 = e$, is only a Brownian motion in M under certain regularity conditions (see [23, Proposition 2.7]). In the case of a bi-invariant metric, a Brownian motion on G maps to a Brownian motion in M through its one-point motion. One-point processes might not even preserve the Markov property in the general case.

2.8. Pushforward measures

Let $\pi\colon G \to M$ be the projection to the homogeneous space $M = G/H$. Then π is a measurable map, and, if μ is a measure on G, the pushforward of μ by π, defined by $\pi_*\mu(B) = \mu\big(\pi^{-1}(B)\big)$, for all measurable subsets $B \subseteq M$, is a measure on M. A numerical example is provided in Figure 1 showing anisotropic distributions on the homogeneous space $\mathbb{S}^2$ obtained from pushing forward Brownian motions of a non-invariant metric on the top space $\mathrm{SO}(3)$.

The Riemannian volume measure Vol_G on G decomposes into a product measure consisting of the volume measure on fibers in G, e.g. $\pi^{-1}(z)$, and the volume measure on its horizontal complement, i.e., $d\,\mathrm{Vol}_G = d\,\mathrm{Vol}_{\pi^{-1}(z)}\, d\,\mathrm{Vol}_{|H}(z)$, where $d\,\mathrm{Vol}_{|H}$ is the horizontal restriction of the volume measure in G. The measure of a process g_t on G pushes forward to M, and we denote the corresponding density wrt. the volume measure on M for p_t^M. Then $p_t^M(x) = \int_{\pi^{-1}(x)} p_t^G(g_0,y) d\,\mathrm{Vol}_{\pi^{-1}(z)}(y)$.

Lemma 1. *Let g_t be a Markov process on G, started at $g_0 \in G$, with density $p_t^G(g_0, \cdot)$, and let $X_t = \pi(g_t)$. The conditional expectation on M satisfies*

$$\mathbb{E}[f(X)|X_T = v] = \mathbb{E}\left[f(X)\frac{p_{T-t}^M(X_t, v)}{p_T^M(x_0, v)}\right],$$

for all bounded, continuous, and non-negative $\mathcal{F}_t$-measurable f on M. Furthermore,

$$\mathbb{E}[\tilde{f}(g)|g_T \in N] = \mathbb{E}[f(X)|X_T = v],$$

where $\tilde{f} = f \circ \pi$.

Proof. Let f be a bounded, continuous, and non-negative measurable function on M, and let $\tilde{f} = f \circ \pi$. Then it follows directly from (7) and (12) that

$$\begin{aligned}\mathbb{E}[\tilde{f}(g)|g_T \in N] =& \mathbb{E}\left[f(\pi(g))\frac{p_{T-t}^G(g_t, N)}{p_T^G(g_0, N)}\right] = \mathbb{E}\left[f(\pi(g_t))\frac{p_{T-t}^G(g_t, \pi^{-1}(v))}{p_T^G(g_0, \pi^{-1}(v))}\right] \\ =& \mathbb{E}\left[f(\pi(g_t))\frac{\pi_* p_{T-t}^G(g_t, v)}{\pi_* p_T^G(g_0, v)}\right] = \mathbb{E}\left[f(X_t)\frac{p_{T-t}^M(X_t, v)}{p_T^M(x_0, v)}\right].\end{aligned}$$

□

3. Simulation of bridges on Lie groups

In this section, we consider the task of simulating (6) conditioned to hit $v \in G$, at time $T > 0$. The potentially intractable transition density for the solution of (6) inhibits simulation directly from (9). Instead, we propose to add a guiding term mimicking that of Delyon and Hu [5], i.e., the guiding term becomes the gradient of the distance to v divided by the time to arrival. The SDE for the guided diffusion becomes

$$dY_t = -\frac{1}{2}V_0(Y_t)dt + V_i(Y_t) \circ \left(dB_t^i - \frac{\left(\nabla_{y|_{y=Y_t}} d(y, v)^2\right)^i}{2(T-t)}dt\right), \qquad Y_0 = e, \tag{13}$$

where $d(\cdot, v)$ denotes the Riemannian distance to v. Note that we can always, for convenience, take the initial value to be the identity e. Equation (13) can equivalently be written as

$$dY_t = -\frac{1}{2}V_0(Y_t)dt + V_i(Y_t) \circ \left(dB_t^i - \frac{\mathrm{Log}_{Y_t}(v)^i}{T-t}dt\right), \qquad Y_0 = e,$$

where Log_p is the inverse of the Riemannian exponential map Exp_p. Figure 2 illustrates one sample path of the simulation scheme in (13) on the Lie group SO(3) and the corresponding axis-angle representation is visualized in Figure 3.

The guiding term in (13) is identical to the guiding term described in [7]. However, in that case, the guided processes used the frame bundle of M. In the Lie group setting, since Lie groups are parallelizable, the use of the frame bundle is not needed.

Numerical computations of the Lie group exponential map are often computationally efficient due to existence of certain algorithms (see [26] and references therein). Therefore, by a change of measures argument, the equation above can be expressed in terms of the inverse of the Lie group exponential (this process is denoted Y as well)

$$dY_t = -\frac{1}{2}V_0(Y_t)dt + V_i(Y_t) \circ \left(d\bar{B}_t^i - \frac{\log_{Y_t}(v)^i}{T-t}dt\right) \tag{14}$$

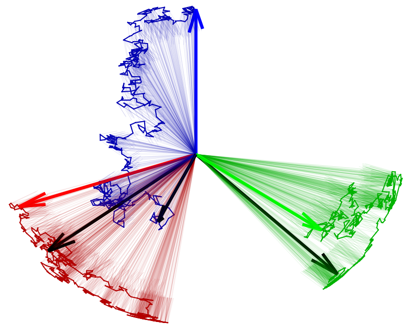


Figure 2. One sample path of the guided bridge process defined by (13) visualized by its action on basis vectors (red, blue, green) of $\mathbb{R}^3$. The bridge is conditioned on the rotation indicated by the black arrows.

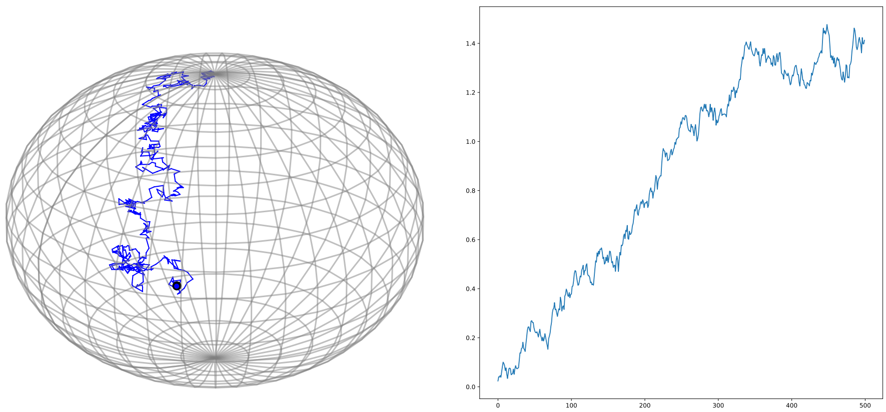


Figure 3. Angle-axis representation of the guided bridge defined by (13). (Left) The projection of the path in SO(3) to $\mathbb{S}^2$. The trajectory on $\mathbb{S}^2$ correspond to the motion of the tip of the blue vector as seen in Figure 2. (Right) The angle representation of the guided bridge in SO(3).

$Y_0 = e$, where $\bar{B}$ is a Brownian motion under a new measure, say $\bar{\mathbb{P}}$. The measure $\bar{\mathbb{P}}$ can explicitly be expressed as

$$\frac{d\bar{\mathbb{P}}}{d\mathbb{P}}|_{\mathcal{F}_t} = \exp\left[-\int_0^t H_v(s, Y_s) - \frac{1}{2}\int_0^t \frac{\|\log_{Y_s}(v) - \mathrm{Log}_{Y_s}(v)\|^2_{Y_s}}{(T-s)^2} ds\right],$$

where $\mathbb{P}$ denotes the law of the SDE in (13) and

$$H_v(t, Y_t) = \left\langle \frac{\left(\log_{Y_t}(v) - \mathrm{Log}_{Y_t}(v)\right)}{T-t}, V(Y_t) dB_t \right\rangle_{Y_t}.$$

The Radon-Nikodym derivative above is a martingale, whenever the group logarithm and the Riemannian logarithm coincide. This is for example the case when the metric is bi-invariant.

3.1. Radial process

We now aim to investigate the relation between the bridge measure and the above simulation schemes. Let $r_v(\cdot) := d(\cdot, v)$ be the distance to v such that $r_v(g_t)$ is the the radial process. Due to the singularities of the radial process on $\mathrm{Cut}(v) \cup \{v\}$, the usual Itô's formula only applies on subsets away from the cut-locus. The extension beyond the cut-locus of a Brownian motion's radial process was due to Kendall [27]. Barden and Le [28,29] generalized the result to M-valued semimartingales. The radial process of the Brownian motion (6) is given by

$$r_v(g_t) = r_v(g_0)^2 + \int_0^t \langle \nabla_{g_s} r_v(g_s), V(g_s)dB_s \rangle_{g_s} + \frac{1}{2}\int_0^t \Delta_G r_v(g_s)ds - L_s^v(g), \tag{15}$$

where L^v is the geometric local time of the cut-locus $\mathrm{Cut}(v)$, which is non-decreasing continuous random functional increasing only when g is in $\mathrm{Cut}(v)$ (see [27–29]). Let $W_t := \int_0^t \left\langle \frac{\partial}{\partial r}, V_i(g_s) \right\rangle dB_s^i$, which is the local-martingale part in the above equation. The quadratic variation of W_t satisfies $d[W,W]_t = dt$, by the orthonormality of $\{V_1, \dots, V_d\}$, thus W_t is a Brownian motion by Levy's characterization theorem. From the stochastic integration by parts formula and (15), the squared radial process of g satisfies

$$r_v(g_t)^2 = r_v(g_0)^2 + 2\int_0^t r_v(g_s)dW_s + \int_0^t r_v(g_s)\Delta_G r_v(g_s)ds - 2\int_0^t r(g_s)dL_s^v, \tag{16}$$

where dL_s^v is the random measure associated to $L_s^v(X)$.

Similarly, we obtain an expression for the squared radial process of Y. The radial process becomes

$$r_v^2(g_t) = r_v(g_0)^2 + 2\int_0^t r_v(g_s)dW_s + \int_0^t \frac{1}{2}\Delta_G r_v(g_s)^2 ds - \int_0^t \frac{r_v(g_s)^2}{T-s}ds - 2\int_0^t r_v(g_s)dL_s^v. \tag{17}$$

Imposing a growth condition on the radial process yields an L^2-bound on the radial process of the guided diffusion, [15]. So assume there exist constants $\nu \geq 1$ and $\lambda \in \mathbb{R}$ such that $\frac{1}{2}\Delta_G r_v^2 \leq \nu + \lambda r_v^2$ on $D \setminus \mathrm{Cut}(v)$, for every regular domain $D \subseteq G$. Then (17) satisfies

$$\mathbb{E}[1_{t<\tau_D} r_v(Y_t)^2] \leq \left(r_v^2(e) + \nu t\left(\frac{t}{T-t}\right)\right)\left(\frac{T-t}{t}\right)^2 e^{\lambda t}, \tag{18}$$

where τ_D is the first exit time of Y from the domain D.

3.2. Girsanov change of measure

Let B_t be a d-dimensional Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_s)_{s\geq 0}, \mathbb{P})$ and let g_t be a solution of (6). The process $\frac{\nabla r_v(g_t)^2}{2(T-t)}$ is an adapted process. As g_t is non-explosive, we see that

$$\int_0^t \left\| \frac{\nabla r_v(g_s)^2}{2(T-s)} \right\|^2 ds = \int_0^t \frac{r_v(g_s)^2}{(T-s)^2} ds \leq C, \tag{19}$$

for every $0 \leq t < T$, almost surely, and for some fixed constant $C > 0$. Define a new measure $\mathbb{Q}$ by

$$Z_t := \frac{d\mathbb{Q}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t}(g) = \exp\left[-\int_0^t \left\langle \frac{\nabla r_v(g_s)^2}{2(T-s)}, V(g_t)dB_s \right\rangle - \frac{1}{2}\int_0^t \frac{r_v(g_s)^2}{(T-s)^2}ds\right]. \tag{20}$$

From (19), the process Z_t is a martingale, for $t \in [0,T)$, and $\mathbb{Q}_t$ defines a probability measure on each $\mathcal{F}_t$ absolutely continuous with respect to $\mathbb{P}$. By Girsanov's theorem (see e.g. [30, Theorem 8.1.2]), we get a new process b_s which is a Brownian motion under the probability measure $\mathbb{Q}$. Moreover, under the probability $\mathbb{Q}$, equation (6) becomes

$$dY_t = -\frac{1}{2}V_0(Y_t)dt + V_i(Y_t) \circ \left(db_t^i - \frac{r_v(Y_t)}{T-t} \left(\frac{\partial}{\partial r_v} \right)^i dt \right), \tag{21}$$

where $\left(\frac{\partial}{\partial r}\right)^i$ is the i'th component of the unit radial vector field in the direction of v. The squared radial vector field is smooth away from $\mathrm{Cut}(v)$ and thus we set it to zero on $\mathrm{Cut}(v)$. Away from $\mathrm{Cut}(v)$, the squared radial vector field is $2\,\mathrm{Log}_v$. The added drift term acts as a guiding term, which pulls the process towards v at time $T > 0$.

From (20), we see that $\mathbb{E}[f(Y_t)] = \mathbb{E}[f(g_t)Z_t]$. Using (16) and the identity $\Delta_G r_v = \frac{d-1}{r_v} + \frac{\partial}{\partial r_v} \log \Theta_v$, θ_v being the Jacobian determinant of Exp_v (see e.g. [31]), we equivalently write $\mathbb{E}[f(Y_t)\varphi_t] = \mathbb{E}[f(X_t)\psi_t]$, with

$$\psi_{t,v} := \exp\left[\frac{-r_v^2(g_t)}{2(T-t)}\right] \qquad \varphi_{t,v} := \exp\left[\int_0^t \frac{r_v(Y_s)}{T-s}(dA_s^v + dL_s^v)\right], \tag{22}$$

where $dA_s^v = \frac{\partial}{\partial r_v} \log \theta_v^{-1/2}(Y_s)ds$ is a random measure supported on $G \setminus \mathrm{Cut}(v)$, and dL_s^v the geometric local time at $\mathrm{Cut}(v)$.

3.3. Delyon and Hu in Lie groups

We can now generalize the result of Delyon and Hu [5, Theorem 5] to the Lie group setting. The result here for Lie groups is analogous to the Riemmanian setting as covered in [7].

Theorem 1. *Let g_t be a solution of (6). The SDE (13) yields a strong solution on $[0, T)$ and satisfies $\lim_{t\uparrow T} Y_t = v$ almost surely. Moreover, the conditional expectation of g given $g_T = v$ is*

$$\mathbb{E}[f(g)|g_T = v] = \lim_{t\uparrow T} \frac{\mathbb{E}[f(Y)\varphi_{t,v}]}{\mathbb{E}[\varphi_{t,v}]}, \tag{23}$$

for every $\mathcal{F}_t$-measurable non-negative function f on G, for $t \in [0, T)$ where φ_t is given in (22).

When the geometry of G is particularly simple the equivalence of measures hold on $[0, T]$, see [7], and the above result reduces to the following.

Corollary 1. *When G is simply connected, (23) becomes*

$$\mathbb{E}[f(g)|g_T = v] = C\mathbb{E}[f(Y)\varphi_{T,v}], \tag{24}$$

where $C > 0$ is a constant, which depends on the initial point, the time $T > 0$, and the curvature in the radial direction.

4. Simulation of bridges in homogeneous spaces

We now turn to bridge simulation in homogeneous spaces by sampling bridges in G conditioned on the fiber over $v \in M = G/H$. We simulate in the top space, that is, bridge simulation schemes on the Lie group G, and subsequently project to the homogeneous space M. We will be considering two schemes. Let $v \in M$.

1. Find closest point $\bar{v}$ in the fiber $\pi^{-1}(v)v$ and iteratively update $\bar{v}$ at each time step.
2. Sample k-points $\{\bar{v}_1, \dots, \bar{v}_k\}$ in fiber above $v \in M$ and consider the bridge Y in G conditioned on $Y_T \in \{\bar{v}_1, \dots, \bar{v}_k\}$.

The motivation for considering the two schemes above derives from [6]. Here, the simulation of Brownian bridges on the flat torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ was the focal point. In this particular geometrical context of $\mathbb{T}^2$, the lift of a $\mathbb{T}^2$-valued Brownian bridge results in a

$\mathbb{R}^2$-valued Brownian bridge conditioned on a set $N \cong \mathbb{Z}^2$. The first scheme was considered in [6], where Brownian bridges on the flat torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ were lifted to bridges in $\mathbb{R}^2$. The second scheme provided a truncated version of the true bridge, for a suitable choice of k. Such a choice of k is dependent on the time to arrival $T > 0$ and the diffusivity σ.

We can sample the k points in the fiber $N = \pi^{-1}(v)$. However, we need to specify from which distribution in N we sample. One approach to defining a distribution on N uses the transition density of a Brownian motion. Thus, if there exists a point $v \in N$ closest to $g_0 \in G$, i.e., $d(v, g_0) \le d(y, g_0)$, for all $y \in N$, then recording the endpoint B_T of a Brownian motion B_t in N, started at $B_0 = v$ corresponds to sampling from a normal distribution in N. As the time T increases, the distribution tends to be uniform, and the initial starting point becomes irrelevant. Therefore, if no unique point $v \in N$ closest to g_0 exists, sampling from a uniform distribution seems more appropriate.

4.1. Guiding to closest point

Recall that the projection $\pi \colon G \to G/H$ is a submersion, hence the manifold $M = G/H$ is an embedded submanifold of G. From Lemma 1, we obtain a conditional expectation in M by conditioning on the fiber in the Lie group. The corresponding SDE for the Fermi bridge in the Lie group setting is given by

$$dY_t = -\frac{1}{2}V_0(Y_t)dt + V_i(Y_t) \circ \left(dB_t^i - \frac{\left(\nabla_{y|y=Y_t} d(y, N)^2\right)^i}{2(T-t)} dt \right), \quad Y_0 = e, \tag{25}$$

where $d(x, N) := \inf_{z \in N} d(x, z)$ and $N := \pi^{-1}(v)$, for some $v \in M$.

The one-point motion conditioned on $v \in M$ corresponds to conditioning g_t on the fiber $N := \pi^{-1}(v)$, and we can use Fermi bridges directly. Because N is an embedded submanifold of G, we get from Thompson [15] that $\varphi_{t,N}$ is of the form

$$\varphi_{t,N} := \exp\left[\int_0^t \frac{r_N(Y_s)}{T-s}\left(dA_s^N + dL_s^N\right)\right], \tag{26}$$

where $dA_s^N = \frac{\partial}{\partial r_N} \log \Theta_N^{-1/2}(Y_s)ds$ and $\Theta_N = \theta_N \circ \left(\mathrm{Exp}\,|_{\mathrm{Log}(M \setminus \mathrm{Cut}(N))}\right)^{-1}$. Similar to the single point case, we obtain

$$\mathbb{E}[f(X)|X_T \in N] = \lim_{t \uparrow T} \frac{\mathbb{E}[f(Y)\varphi_{t,N}]}{\mathbb{E}[\varphi_{t,N}]},$$

for any bounded measurable function f. There are various occasions where it can be justified to take the limit inside. See the discussion in [32, Appendix C].

4.2. Guiding to k-points in fiber

The guiding scheme presented in (25) requires an optimization step in each time step, and for practical purposes, this might be computationally inefficient. This section suggests guiding to a subset of the fiber N, reducing the guided bridge scheme in (25) to a case of guiding to a finite set of points.

For certain homogeneous spaces, e.g., $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, the fiber $N = \pi^{-1}(v)$ is a discrete subgroup in G. In this case, the volume measure Vol_N in (10) is the counting measure and we can write the density as

$$p_t^G(g, N) = \sum_{v \in N} p_t^G(g, v).$$

From a numerical perspective, when the discrete subgroup is large restricting to a smaller finite subgroup $N_k \subseteq N$ of k-nearest-points of the initial starting point may speed up computation-time.

Example 3. *To exemplify this, let p_t denote the transition density of a standard Brownian motion in $\mathbb{R}^2$. The transition density of the Brownian motion started at the origin is then given as $p_t(0,v) = (2\pi t)^{-1}\exp\left(-\frac{\|v\|^2}{2t}\right)$. Let $N_k = \{(-2,-2),\ldots,(2,2)\} \subseteq \mathbb{Z}^2$ be the subset of five-by-five grid points and let $T = 1$, then we see*

$$p_T^G(0,N) \approx \sum_{v\in N_k} p_T(0,v) = (2\pi)^{-1}\Big(1+4\Big(e^{-1/2}+e^{-1}+e^{-2}+e^{-4}+2e^{-8}\Big)\Big) \approx 0.98.$$

We recover more than 98% of the total mass when restricting to a finite set of points. If we restricted the subset to the set of three-by-three grid points $N_k = \{(-1,-1),\ldots,(1,1)\}$, the density will only describe roughly 78% of the total mass. However, if we restricted the time to arrival $T = 1/2$, then we would recover 95% of the mass. Thus, we see that both the initial point and the terminal time will affect the choice of k.

The tractable transition density on the flat torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ is given by $p_T^{\mathbb{R}}(0,N) = \sum_{v_i\in N} p_T^{\mathbb{R}}(0,v_i)$, where N is a set isomorphic to $\mathbb{Z}^2$. The corresponding SDE for the conditioned process is given by [6]

$$dY_t = \sum_{v_i\in N} f_i(t,Y_t)\frac{v_i - Y_t}{T-t}dt + dW_t, \quad \text{where} \quad f_i(t,Y_t) = \frac{\exp\left(-\frac{\|v_i - Y_t\|^2}{2t}\right)}{\sum_{v_j\in N}\exp\left(-\frac{\|v_j - Y_t\|^2}{2t}\right)}. \tag{27}$$

By the example above, good numerical approximations can be obtained by restricting to a finite set of points. We conjecture that a similar type of guided drift can be used, where the transition density above is exchanged with the transition density of the Riemannian normal distribution (see e.g. [32]). We do not pursue this approach any further in this paper. Instead, we propose to sample a point from N from a given distribution on N.

The following result is inspired by the type of conditioning found in van der Meulen and Schauer [33], Mider, Schauer, and van der Meulen [34], and Arnaudon et al. [35]. We adopt this type of inexact matching by imposing noise on the conditioning point. The proposed method alleviates the optimization procedure in each time step to finding the closest point in the fiber. One immediate application of the result below will be in the situation where we sample Brownian motions in the fiber, starting at the closest point in the fiber. Recording the endpoints after some fixed time, we obtain samples from a normal distribution in the fiber. Therefore, the simulation scheme reduces to conditioning at a point as described in Section 3, the caveat being that the endpoint is tilted to a specific distribution. The result below is a theoretical one. We note from the guided bridge scheme that $d\mathbb{P}_{x,v}^T/d\mathbb{P}_x(y) = \varphi_{T,v}/\mathbb{E}[\varphi_{T,v}]$.

Theorem 2. *Let g be a Markov process defined on $(\Omega,\mathcal{F},\mathbb{P})$ with values in G, and let $p_t^G(\cdot,\cdot)$ be its transition density defined by $\mathbb{P}(g_T\in du|g_t = g) = p_{T-t}^G(g,u)d\,\mathrm{Vol}_G(u)$. Assume $0\le t< T$ and $g_T \sim f_{u_0}\cdot \mathrm{Vol}_N$ (e.g., normal/uniform distribution on fiber over v) under the probability measure $\mathbb{P}_{g_0}$ started at g_0. The conditional law of g_t given $g_T = v$, $\mathbb{P}_{g_0,v}^T$, has density wrt. the reference measure $d\,\mathrm{Vol}_G$ given by*

$$\frac{p_{T-t}^G(z,v)p_t^G(z,g_0)}{p_T^G(g_0,v)}, \tag{28}$$

and the simultaneous distribution of (X_t,X_T) has density given by

$$\frac{\mathbb{P}(g_t\in dg, g_T\in du)}{d\,\mathrm{Vol}_N(u)d\,\mathrm{Vol}_G(g)} = f_{u_0}(u)\frac{p_{T-t}^G(z,u)p_t^G(z,g)}{p_T^G(g_0,u)}. \tag{29}$$

Furthermore, if we define the h-function as

$$h(t,g_t) = \frac{\int f_{u_0}(u)\frac{p^G_{T-t}(g_t,u)}{p^G_T(g_0,u)}d\operatorname{Vol}_N(u)}{\int_N f_{u_0}(x)d\operatorname{Vol}_N(x)}, \tag{30}$$

then for any non-negative measurable functional F we have

$$\mathbb{E}_{\mathbb{Q}}[F(g)] = \mathbb{E}_{\mathbb{P}}[h(t,g_t)F(g)] = \int \mathbb{E}_{\mathbb{P}}[F(g)|g_T = u]\frac{f_{u_0}(u)}{\int_N f_{u_0}(x)d\operatorname{Vol}_N(x)}d\operatorname{Vol}_N(u).$$

The conditional distribution of X_t given X_T has density

$$k_t(g,u) = \frac{f_{u_0}(u)}{\int_N f_{u_0}(x)d\operatorname{Vol}_N(x)}\frac{p^G_{T-t}(z,u)p^G_t(z,g)}{p^G_T(g_0,u)}, \tag{31}$$

with respect to the volume measure Vol_N. If the distribution $g_T(\mathbb{P})$ has full mass in the fiber N, i.e., $g_T(\mathbb{P})(N) = \int_N f dVol = 1$ the term above simplifies.

Proof. The fact that (28) is the conditional density wrt. $d\operatorname{Vol}_G$, the volume measure on G, follows from (7), since $d\mathbb{P}^t_g = p_t(g,z)d\operatorname{Vol}_G(z)$ and therefore

$$d\mathbb{P}^T_{g_0,v}(g_t) = \frac{p^G_{T-t}(z,v)p^G_t(g_0,z)}{p^G_T(g_0,v)}d\operatorname{Vol}_G(z).$$

Hence (29) follows.

For the second part take h as defined in (30). Without loss of generality assume that $g_T(\mathbb{P})(N) = 1$. Note that h is a martingale with $h(0,g_0) = 1$, since g is a Markov process and

$$\begin{aligned}\mathbb{E}[h(t,g_t)|g_s] &= \int p^G_{t-s}(g_s,z)h(t,z)d\operatorname{Vol}_G(z)\\ &= \int p^G_{t-s}(g_s,z)\int f_{u_0}(x)\frac{p^G_{T-t}(z,x)}{p^G_T(g_0,x)}d\operatorname{Vol}_N(x)d\operatorname{Vol}_G(z)\\ &= \int f_{u_0}(x)\frac{p^G_{T-s}(g_s,x)}{p^G_T(g_0,x)}d\operatorname{Vol}_N(x) = h(s,g_s)\end{aligned}$$

together with

$$\mathbb{E}[h(t,g_t)] = \int f_{u_0}(x)\frac{p^G_T(g_0,x)}{p^G_T(g_0,x)}d\operatorname{Vol}_N(x) = 1.$$

Since $\lim_{t\downarrow 0}\int p^G_t(g_0,z)f(z)d\operatorname{Vol}_G(z) = f(g_0)$, for any bounded continuous function f, Fatou's lemma ensures that $\mathbb{E}[h(T,g_T)] = 1$

$$\begin{aligned}1 &= \limsup_{t\uparrow T}\mathbb{E}[h(t,g_t)] \le \mathbb{E}\left[\limsup_{t\uparrow T}h(t,g_t)\right]\\ &= \int_G \limsup_{t\uparrow T}\int f_{u_0}(x)\frac{p^G_{T-t}(z,x)}{p^G_T(g_0,x)}d\operatorname{Vol}_N(x)d\operatorname{Vol}_G(z)\\ &= \int_G \frac{f_{u_0}(z)}{p^G_T(g_0,z)}d\operatorname{Vol}_G(z) = \int_G \liminf_{t\uparrow T}\int f_{u_0}(x)\frac{p^G_{T-t}(z,x)}{p^G_T(g_0,x)}d\operatorname{Vol}_N(x)d\operatorname{Vol}_G(z)\\ &\le \liminf_{t\uparrow T}\mathbb{E}[h(t,g_t)] = 1.\end{aligned}$$

Hence h is a true martingale on $[0,T]$ and thus defines a new probability measure $\mathbb{Q}$ on $\mathcal{F}$ by $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t}(g) = h(t, g_t)$.

For the second part of the proof, assume, temporarily, that F is a measurable function such that

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}[h(t,g_t)F(g_t)] &= \int p_t^G(g_0,z)h(t,z)F(z)d\,\mathrm{Vol}_G(z) \\
&= \int\int \frac{p_{T-t}^G(z,x)p_t^G(g_0,z)}{p_T^G(g_0,x)}F(z)d\,\mathrm{Vol}_G(z)f(x)d\,\mathrm{Vol}_N(x) \\
&= \int\int F(z)d\mathbb{P}^T_{g_0,x}(z)f_{u_0}(x)d\,\mathrm{Vol}_N(x) = \int \mathbb{E}_{\mathbb{P}}[F(g_t)|g_T = x]f_{u_0}(x)d\,\mathrm{Vol}_N(x).
\end{aligned}
$$

In order to conclude, we need to show that for any finite distribution $(X_{t_1},\dots,X_{t_n})$

$$
\mathbb{E}_{\mathbb{P}}[h(t,X_t)F(X_{t_1},\dots,X_{t_n})] = \int \mathbb{E}_{\mathbb{P}}[F(X_{t_1},\dots,X_{t_n})|X_T = u]f_{u_0}(u)d\,\mathrm{Vol}_N(u).
$$

Therefore, let $0 < t_1 < \cdots < t_n < T$ and $t \in (t_n, T)$. Define Φ_F similar to how it was defined in [7]. Then

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}[h(t,X_t)F(X_{t_1},\dots,X_{t_n})] &= \int_G h(t,x)\Phi_F(t,x)d\,\mathrm{Vol}_G(x) \\
&= \int f_{u_0}(u)\int F(z)P(z,x,u)d\,\mathrm{Vol}(z)d\,\mathrm{Vol}(x)d\,\mathrm{Vol}(u) \\
&= \int \mathbb{E}_{\mathbb{P}}[F(X_{t_1},\dots,X_{t_n})|X_T = u]f_{u_0}(u)d\,\mathrm{Vol}_N(u),
\end{aligned}
$$

where $z = (z_1,\dots,z_n)$ and $d\,\mathrm{Vol}(z) = d\,\mathrm{Vol}(z_1)\dots d\,\mathrm{Vol}(z_n)$ and where

$$
P(z,x,u) = \frac{p_{t_1}^G(x_0,z_1)\dots p_{t-t_n}^G(z_n,x)p_{T-t}^G(x,u)}{p_T^G(x_0,u)}.
$$

□

Assume that the constants c_i exist. We can obtain estimates of the constants c_i via simulation of sample paths of the guided process. Let $y_t^i = Y_t(\omega_i)$ be realizations of the guided bridge process. We can then use the estimator

$$
\bar{\varphi}_{T,v_i}(y) = \frac{1}{m}\sum_{n=1}^{m}\exp\left[\int_0^T \frac{r_{v_i}(y_s^n)}{T-s}\left(dA_s^{v_i} + dL_s^{v_i}\right)\right] \tag{32}
$$

to approximate the constants $c_i = \mathbb{E}[\varphi_{T,v_i}]$. Algorithm 1 provides a method for obtaining k-points in the fiber N, together with the normalizing constants, when N is connected and compact.

5. Maximum Likelihood Estimation

For a manifold-valued Brownian motion recall (5), which describes the Brownian motion locally in a chart. The diffusion coefficient is the matrix square root of the cometric tensor, i.e., the inverse metric tensor. As seen in Figure 1, the pushforward measure of a Brownian motion generated by a non-invariant metric induces anistotropic distributions on the quotient space. We aim to estimate the underlying metric by an MLE approach.

Consider observations $y^1,\dots,y^n$ on G or G/H obtained from distributions $\mathbb{P}_\theta^t$ or $\pi_*\mathbb{P}_\theta^t$, with parameters $\theta = (g, A)$, and corresponding densities $p_t(\cdot|\theta)$ and $\pi_* p_t(\cdot|\theta)$. Here $A^{-1} = \Sigma$ is the inverse covariance matrix and $\Sigma = \sigma\sigma^T$. We can define the likelihood function as

$$
\mathcal{L}(\theta|y^1,\dots,y^n;T) = \prod_{i=1}^n p_T(y^i|\theta), \tag{33}
$$

Algorithm 1: Stochastic Metropolis-Hastings Algorithm

```
// Initialization
```
Choose initial point $e \in G$ and $v_1 \in N$ closest to e. Simulate a guided bridge process to v_1 and obtain an estimate of $\mathbb{E}[\varphi_{T,v_1}]$.
```
// Main loop
```
while *k points not reached* **do**
```
// Step 1:
```
Propose u from the proposal density $f_{v_i}(u)$ (e.g. uniform density in N centered at v_i or normal density in N centered at v_i) and sample estimator for $\mathbb{E}[\varphi_{T,u}]$
```
// Step 2:
```
Calculate the acceptance ratio $g(u, v_i) = \min\left\{1, \frac{f_u(v_i) q_T(x_0,u)\bar{\varphi}_{T,u}}{f_{v_i}(u) q_T(x_0,v_i)\bar{\varphi}_{T,v_i}}\right\}$ (Note that if f is symmetric it cancels out in the acceptance probability)
```
// Step 3
```
Accept with probability $g(u, v_i)$ and set $v_{i+1} = u$ as well as $c_{i+1} = \bar{\varphi}_{T,v_{i+1}}$ otherwise do nothing.

end
```
// Output:
```
$\{(v_1, c_1), \ldots, (v_k, c_k)\}$

and similarly $\pi_* \mathcal{L} = \prod \pi_* p$. The bridge sampling scheme introduced above yields approximations of the intractable transition densities in 33. Algorithm 2 provides a detailed description for the iterative MLE approach. Visual examples of the iterative MLE can be found in Figure 4 and 6.

Algorithm 2: Parameter Estimation: Iterative MLE.

```
// Initialization
```
Given n data points $\{v_1, \ldots, v_n\}$.
// Specify initial parameters $\theta_0 = (g_0, A_0)$ and a learning rate η.

for $k = 1$ *to* K **do**

for $j = 1$ *to* n **do**

Sample m bridges from (13) conditioned on v_j to get estimate for $\mathbb{E}[\varphi_{T,v_j}] \approx \frac{1}{m} \sum_{i=1}^m \varphi^i_{T,v_j}$

end

$\ell_{\theta_{k-1}}(v_1, \ldots, v_n) = \prod_{j=1}^n \left(\frac{\det A_{k-1}(T,v_j)}{2\pi T}\right)^{3/2} e^{-\frac{\|\mathrm{Log}_{v_j}(g_{k-1})\|^2_{A_{k-1}}}{2T}} \frac{1}{m} \sum_{i=1}^m \varphi^i_{T,v_j}$
```
// Compute the gradient
```
$\xi_k = \nabla_{\theta_{k-1}} \log \ell_{\theta_{k-1}}(v_1, \ldots, v_n)$
```
// Update the parameters
```
$\theta_k = \theta_{k-1} - \eta \xi_k$

end

// Return final parameters $\theta_K = (g_K, A_K)$

6. Experiments

In this section, we present numerical results of bridge sampling on specific Lie groups and homogeneous spaces. The specific Lie groups in question are the three-dimensional rotation group SO(3) and the general linear group of invertible matrices with positive determinant $\mathrm{GL}_+(3)$. Exploiting the bridge sampling scheme described above, we show below how to estimate the true underlying metric on SO(3) with iterative maximum likelihood estimation.

The space of the symmetric positive definite matrices $\mathrm{SPD}(n)$ is an example of a non-linear space in which geometric data appear in many applications. The space $\mathrm{SPD}(3)$ can be obtained as the homogeneous space $\mathrm{GL}_+(3)/\mathrm{SO}(3)$, where GL_+ is the space of invertible matrices with a positive determinant.

Lastly, considering the two-sphere $\mathbb{S}^2$ as the homogeneous space $\mathrm{SO}(3)/\mathrm{SO}(2)$, we verify that the bridge sampling scheme on this homogeneous space yields admissible heat kernel estimates on $\mathbb{S}^2$.

6.1. Numerical simulations

The Euler-Heun scheme leads to approximation of the Stratonovich integral. With a time discretization $t_1, \dots, t_k$, $t_k - t_{k-1} = \Delta t$ and corresponding noise $\Delta B_{t_i} \sim N(0, \Delta t)$, the numerical approximation of the Brownian motion (6) takes the form

$$x_{t_{k+1}} = x_{t_k} - \frac{1}{2}\sum_{j,i} C^j_{ij} V_i(x_{t_k})\Delta t + \frac{v_{t_{k+1}} + V_i(v_{t_{k+1}} + x_{t_k})\Delta B^i_{t_k}}{2} \tag{34}$$

where $v_{t_{k+1}} = V_i(x_{t_k})\Delta B^i_{t_k}$ is only used as an intermediate value in integration. Adding the logarithmic term in (21) to (34) we obtain a numerical approximation of a guided diffusion (13).

6.2. Importance sampling and metric estimation on $\mathrm{SO}(3)$

This section takes G as the special orthogonal group $\mathrm{SO}(3)$, the space of three-dimensional rotation matrices. The special orthogonal group is a compact connected matrix Lie group. In the context of matrix Lie groups, computing left-invariant vector fields is straightforward. The Lie algebra of the rotation group $\mathrm{SO}(3)$ is the space of three-by-three skew symmetric matrices, $\mathfrak{so}(3)$. The exponential map $\exp\colon \mathfrak{so}(3) \to \mathrm{SO}(3)$ coincides with the usual matrix exponential e^A. With $a \in \mathbb{R}^3$, we can express any element $A \in \mathfrak{so}(3)$ in terms of the standard basis $\{e_1, e_2, e_3\}$ of $\mathbb{R}^3$ as

$$A = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}.$$

Let $\theta = \|a\|_2$ and assume that $\theta \neq 0$. By Rodrigues' formula the matrix Lie group exponential map $\exp\colon \mathfrak{so}(3) \to \mathrm{SO}(3)$ is given by

$$R := e^A = I + \frac{\sin(\theta)}{\theta}A + \frac{(1-\cos(\theta))}{\theta^2}A^2$$

and the corresponding inverse matrix Lie group exponential map $\log\colon \mathrm{SO}(3) \to \mathfrak{so}(3)$

$$\log(R) = \frac{\sin^{-1}(\theta)}{2\theta}(R - R^T).$$

The rotation group $\mathrm{SO}(3)$ is a semi-simple Lie group; hence, a bi-invariant inner product exists. In the case of a bi-invariant metric, the Riemannian exponential map Exp coincides with the Lie group exponential map exp and thus the Riemannian distance function $d(R, I)^2 = \|\mathrm{Log}_I(R)\|^2$, from the rotation R to the identity I, satisfies $\nabla_R d(R, I)^2 = 2\log(R)$.

The structure coefficients of $\mathfrak{so}(3)$ are particularly simple. Let $A_i = A$ with $a_j = 1$ if $i = j$ and zero otherwise. In this case, $\{A_1, A_2, A_3\}$ defines a basis of $\mathfrak{so}(3)$. The structure coefficients satisfy the relation $[A_i, A_j] = C^k_{ij}A_k = \epsilon^{ijk}A_k$, where ϵ^{ijk} denotes the Levi-Civita symbols. The Levi-Civita symbols are defined as $+1$, for (i, j, k) an even permutation of $(1, 2, 3)$, -1 for every odd permutation, and zero otherwise.

6.3. Numerical bridge sampling algorithm on SO(3)

Utilizing the simple expressions for the structure coefficients and the Lie group logarithmic map, we can explicitly write up the numerical approximation of the guided bridge processes (Brownian bridge) on SO(3) as

$$x_{t_{k+1}} = x_{t_k} - \frac{1}{2}\sum_{j,i}\epsilon^{ijj}V_i(x_{t_k})\Delta t + \frac{v_{t_{k+1}} + V_i(v_{t_{k+1}} + x_{t_k})\Big(\Delta B^i_{t_k} - \frac{\log(x_k)}{T-t_k}\Delta t\Big)}{2}, \tag{35}$$

where in this case we have $v_{t_{k+1}} = V_i(x_{t_k})\Big(\Delta B^i_{t_k} - \frac{\log(x_k)}{T-t_k}\Delta t\Big)$. Figure 2 illustrates the numerical approximation by showcasing three different sample paths from the guided diffusion conditioned to hit the rotation represented by the black vectors.

Another way of visualizing the guided bridge on the rotation group SO(3) is through the angle-axis representation. Figure 3 represents a guided process on SO(3) by presenting the axis representation on $\mathbb{S}^2$ and its corresponding angle of rotation.

6.4. Metric estimation on the three-dimensional rotation group

In the d-dimensional Euclidean case, importance sampling yields the estimate [9]

$$p_T(u,v) = \left(\frac{\det(A(T,v))}{2\pi T}\right)^{\frac{d}{2}} e^{-\frac{\|u-v\|^2_A}{2T}}\mathbb{E}[\varphi_{T,v}],$$

where $\|x\|_A = x^T A(0,u)x$. Thus, from the output of the importance sampling, we get an estimate of the transition density. Similar to the Euclidean case, we obtain an expression for the heat kernel $p_T(e,v)$ as $p_T(e,v) = q(T,e)\mathbb{E}[\varphi_{T,v}]$, where

$$q(T,g) = \left(\frac{\det A(v)}{2\pi T}\right)^{\frac{3}{2}}\exp\left(-\frac{d(g,v)^2}{2T}\right) = \left(\frac{\det A(T,v)}{2\pi T}\right)^{\frac{3}{2}}\exp\left(-\frac{\|\mathrm{Log}_g(v)\|^2_A}{2T}\right), \tag{36}$$

where the equality holds almost everywhere, and $A \in \mathrm{Sym}^+(\mathfrak{g})$ denotes the metric $A(e) := A(0,e)$. The Log_g map in (36) is the Riemannian inverse exponential map $(\mathrm{Exp}_g)^{-1}$.

Figure 4 illustrates how importance sampling on SO(3) leads to a metric estimation of the underlying unknown metric, which generated the Brownian motion. We sampled 128 points as endpoints of a Brownian motion from the metric diag(0.2, 0.2, 0.8), and used 20 time steps to sample 4 bridges per observation. An iterative MLE method using gradient descent with a learning rate of 0.2 and an initial guess of the metric being diag(1, 1, 1) yielded a convergence to the true metric. Note that in each iteration, the logarithmic map changes as can be seen from Algorithm 2.

6.5. Diffusion-mean estimation on the space of symmetric positive definite matrices

The space of symmetric positive definite (SPD) matrices is used in in a range of applied fields, one example being diffusion tensor imaging where element of SPD(3) models anisotropic diffusion of water molecules in each position of the imaged domain. The SPD matrices constitute a smooth incomplete manifold when endowed with the Euclidean metric of matrices [26]. However, endowing the space of SPD matrices with either the Log Euclidean or the Affine invariant metric makes the space geodesically complete, i.e., the exponential map is a global diffeomorphism. The space of SPD(3) matrices can be regarded as the homogeneous space $\mathrm{GL}_+(3)/\mathrm{SO}(3)$ of invertible matrices with positive determinants being rotationally invariant to three-dimensional rotations. Figure 5 illustrates the discrete time observations from three different sample paths in SPD(3) arising from the pushforward of a Fermi bridge in $\mathrm{GL}_+(3)$.

In Figure 6, the bridge sampling scheme derived above is used to obtain an estimate of the diffusion-mean [36,37] on SPD(3), by sampling guided bridge processes in the space

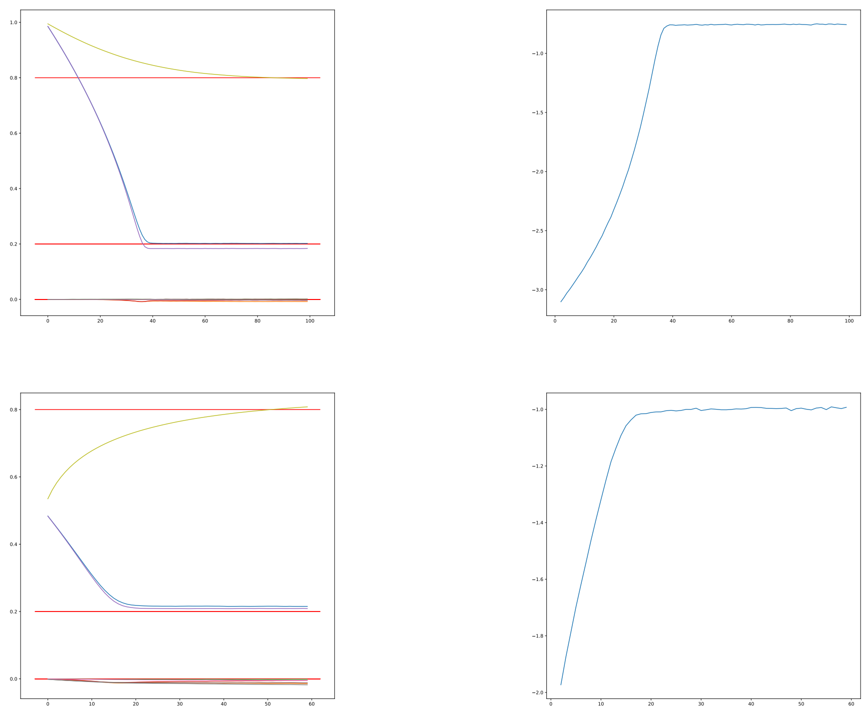


Figure 4. The importance sampling technique applies to estimate the metric on the Lie group $\mathrm{SO}(3)$. Sampling a Brownian motion from an underlying unknown metric, we obtain convergence to the true underlying metric using an iterative MLE method. Here we sampled 4 guided bridges per observation, providing a relatively smooth iterative likelihood. (Top left) Estimation of the unknown underlying metric using bridge sampling, starting from the metric $\mathrm{diag}(1,1,1)$. Here the true metric is the diagonal matrix $\mathrm{diag}(0.2, 0.2, 0.8)$ represented by the red lines. The diagonal is represented by the colors diag(purple,blue,yellow). (Top right) The correspondong iterative log-likelihood. (Bottom left) Estimation of the unknown underlying metric using bridge sampling, starting from the metric $\mathrm{diag}(0.5, 0.5, 0.5)$. (Bottom right) The corresponding iterative log-likelihood.

of invertible matrices with positive determinants $\mathrm{GL}_+(3)$. This sampling method provides an estimate of the density on $\mathrm{GL}_+(3)$, which projects to a density in $\mathrm{SPD}(3)$. Exploiting the resulting density in $\mathrm{SPD}(3)$, an iterative MLE method then yields convergence to the diffusion mean.

6.6. Density estimation on the two-sphere

As explained in Section 4.1, we introduced a simulation scheme on specific homogeneous spaces by using guided bridges in the top space conditioned to arrive in the fiber at time T. The two-sphere $\mathbb{S}^2$ can be considered the homogeneous space $\mathrm{SO}(3)/\mathrm{SO}(2)$ of three-dimensional rotations, identifying the subgroup of two-dimensional rotations as a single point. Conditioning on the fiber $\mathrm{SO}(2)$ in $\mathrm{SO}(3)$, we obtain guided bridges on $\mathbb{S}^2$.

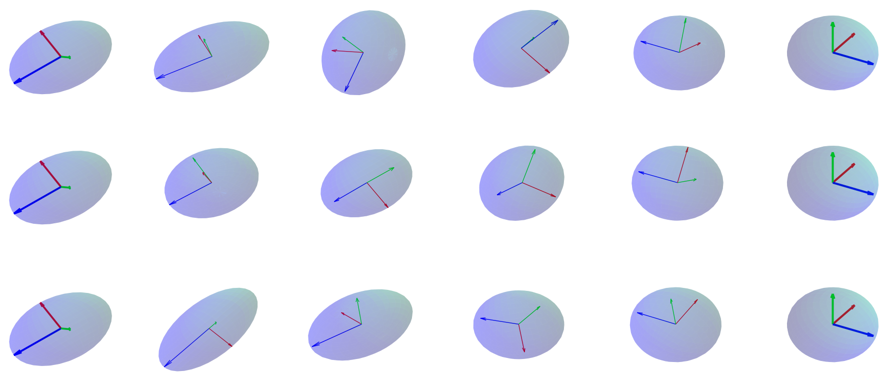


Figure 5. Discrete time observations from three sample paths on SPD(3). The sample paths are obtained as the pushforward of the Fermi bridge in $GL_+(3)$. The start and endpoint are the left- and rightmost figures, where the SPD matrices are indicated by the bold face arrows.

Code

The code used for the experiments is available in the Theano Geometry software package [1]. The implementation uses automatic differentiation libraries extensively for the geometry computations as is further described in [38].

Acknowledgement

The work presented is supported by the CSGB Centre for Stochastic Geometry and Advanced Bioimaging funded by a grant from the Villum foundation, the Villum Foundation grants 22924 and 40582, and the Novo Nordisk Foundation grant NNF18OC0052000.

1. Pedersen, A.R. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1995**, pp. 257–279.
2. Bladt, M.; Sørensen, M. Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli* **2014**, *20*, 645–675.
3. Bladt, M.; Finch, S.; Sørensen, M. Simulation of multivariate diffusion bridges. *Journal of the Royal Statistical Society: Series B: Statistical Methodology* **2016**, pp. 343–369.
4. Bui, M.N.; Pokern, Y.; Dellaportas, P. Inference for partially observed Riemannian Ornstein–Uhlenbeck diffusions of covariance matrices. *arXiv preprint arXiv:2104.03193* **2021**.
5. Delyon, B.; Hu, Y. Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Applications* **2006**.
6. Jensen, M.H.; Mallasto, A.; Sommer, S. Simulation of Conditioned Diffusions on the Flat Torus. International Conference on Geometric Science of Information. Springer, 2019, pp. 685–694.
7. Jensen, M.H.; Sommer, S. Simulation of Conditioned Semimartingales on Riemannian Manifolds. *arXiv:2105.13190* **2021**, [2105.13190].
8. van der Meulen, F.; Schauer, M.; et al. Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electronic Journal of Statistics* **2017**, *11*, 2358–2396.
9. Papaspiliopoulos, O.; Roberts, G. Importance sampling techniques for estimation of diffusion models. *Statistical methods for stochastic differential equations* **2012**.
10. Schauer, M.; Van Der Meulen, F.; Van Zanten, H.; et al. Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli* **2017**, *23*, 2917–2950.
11. Sommer, S.; Arnaudon, A.; Kuhnel, L.; Joshi, S. Bridge Simulation and Metric Estimation on Landmark Manifolds. *arXiv:1705.10943 [cs]* **2017**.

[1] http://bitbucket.org/stefansommer/theanogeometry

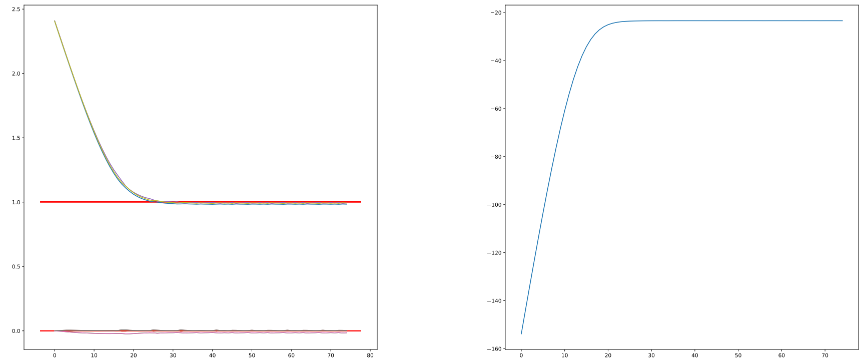


Figure 6. Given 256 data points in SPD(3), we estimated the diffusion mean on the homogeneous space by sampling bridges in the top space conditioned on the fibers. The iterative MLE in Algorithm 2 yielded the convergence of the diffusion mean parameter, using a learning rate of 0.005 and one bridge sample per observation. (Left) The purple, blue, and yellow line correspond to the diagonal, while the remaining colors represent the off diagonal. The true mean value is the identity matrix indicated by the red lines. (Right) The corresponding iterative log-likelihood.

12. Bui, M.N. Inference on Riemannian Manifolds: Regression and Stochastic Differential Equations. PhD thesis, UCL (University College London), 2022.
13. Fisher, R.A. Dispersion on a sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1953**, *217*, 295–305.
14. Kent, J.T. The Fisher-Bingham distribution on the sphere. *Journal of the Royal Statistical Society: Series B (Methodological)* **1982**, *44*, 71–80.
15. Thompson, J. Brownian bridges to submanifolds. *Potential Analysis* **2018**, *49*.
16. García-Portugués, E.; Sørensen, M.; Mardia, K.V.; Hamelryck, T. Langevin diffusions on the torus: estimation and applications. *Statistics and Computing* **2017**, pp. 1–22.
17. Hamelryck, T.; Kent, J.T.; Krogh, A. Sampling Realistic Protein Conformations Using Local Structural Bias. *PLoS Computational Biology* **2006**, *2*. doi:10.1371/journal.pcbi.0020131.
18. Pennec, X.; Fillard, P.; Ayache, N. A Riemannian Framework for Tensor Computing. *Int. J. Comput. Vision* **2006**, *66*, 41–66.
19. Vaillant, M.; Miller, M.; Younes, L.; Trouvé, A. Statistics on Diffeomorphisms via Tangent Space Representations. *NeuroImage* **2004**, *23*, S161–S169. doi:10.1016/j.neuroimage.2004.07.023.
20. Yang, L. Means of Probability Measures in Riemannian Manifolds and Applications to Radar Target Detection. PhD thesis, Poitiers University, 2011.
21. Grenander, U. *Probabilities on Algebraic Structures*; Wiley: New York and London, 1963.
22. Pennec, X.; Sommer, S.; Fletcher, T. *Riemannian Geometric Statistics in Medical Image Analysis*; Elsevier, 2020.
23. Liao, M. *Lévy processes in Lie groups*; Vol. 162, Cambridge university press, 2004.
24. Shigekawa, I. Transformations of the Brownian motion on a Riemannian symmetric space. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **1984**.
25. Jensen, M.H.; Sommer, S. Mean Estimation on the Diagonal of Product Manifolds. *Algorithms* **2022**, *15*, 92.
26. Pennec, X.; Sommer, S.; Fletcher, T. *Riemannian Geometric Statistics in Medical Image Analysis*; Academic Press, 2019.
27. Kendall, W.S. The radial part of Brownian motion on a manifold: a semimartingale property. *The Annals of Probability* **1987**, *15*, 1491–1500.
28. Barden, D.; Le, H. Some consequences of the nature of the distance function on the cut locus in a Riemannian manifold. *Journal of the LMS* **1997**.
29. Le, H.; Barden, D. Itô correction terms for the radial parts of semimartingales on manifolds. *Probability theory and related fields* **1995**, *101*, 133–146.
30. Hsu, E.P. *Stochastic analysis on manifolds*; Vol. 38, AMS, 2002.
31. Thompson, J. Submanifold bridge processes. PhD thesis, University of Warwick, 2015.

32. Thompson, J.; Li, X.M. Submanifold Bridge Processes. PhD thesis, University of Warwick, Coventry, 2015.
33. van der Meulen, F.; Schauer, M. Bayesian estimation of incompletely observed diffusions. *Stochastics* **2018**, *90*, 641–662.
34. Mider, M.; Schauer, M.; van der Meulen, F. Continuous-discrete smoothing of diffusions. *Electronic Journal of Statistics* **2021**, *15*, 4295–4342.
35. Arnaudon, A.; van der Meulen, F.; Schauer, M.; Sommer, S. Diffusion bridges for stochastic Hamiltonian systems and shape evolutions. *SIAM Journal on Imaging Sciences* **2022**, *15*, 293–323.
36. Hansen, P.; Eltzner, B.; Sommer, S. Diffusion Means and Heat Kernel on Manifolds. *Geometric Science of Information* **2021**.
37. Hansen, P.; Eltzner, B.; Huckemann, S.F.; Sommer, S. Diffusion Means in Geometric Spaces. *arXiv preprint arXiv:2105.12061* **2021**.
38. Kühnel, L.; Sommer, S.; Arnaudon, A. Differential Geometry and Stochastic Dynamics with Deep Learning Numerics. *Applied Mathematics and Computation* **2019**, *356*, 411–437. doi:10.1016/j.amc.2019.03.044.

Chapter 4

Mean Estimation on the Diagonal of Product Manifolds

The paper presented in this chapter was made in collaboration with Stefan Sommer. The chapter contains the paper

> Mathias Højgaard Jensen, Stefan Sommer. "*Mean Estimation on the Diagonal of Product Manifolds*", *Algorithms*, 2022, Vol.15 (3), p.92.

which was published in the special issue *Stochastic Algorithms and their Applications* of the journal *Algorithms.* The paper introduce a bridge sampling scheme on an n-fold product manifold conditioned on the diagonal. The paper is build on the ideas and concepts introduced by Sommer and Bronstein [46] and Thompson [48]. The Fréchet mean generalizes the Euclidean mean to non-linear metric spaces using the geometric properties. Computing the Fréchet mean is often computationally costly when closed-form expressions for geodesics are unavailable, as it leads to a nested optimization problem. The diffusion mean is an alternative to the Fréchet mean, defined using a probabilistic interpretation of the mean [23]. The aim of the paper is to derive a bridge simulation scheme on product manifolds conditioned on hitting the diagonal. A single forward simulation yields an estimate of the diffusion mean which is computationally efficient. We verify the computational efficiency on a specific data set with annotated landmarks by computing rough running times and compare them with the computation time for the Fréchet mean. In contrast, the diffusion mean estimation performed in [45] using direct optimization of the likelihood approximation with bridge sampling from the mean candidate to each data point is comparable in complexity to the Fréchet mean computation.

algorithms

MDPI

Article

Mean Estimation on the Diagonal of Product Manifolds

Mathias Højgaard Jensen † and Stefan Sommer *,†

Department of Computer Science, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark; matje@di.ku.dk
* Correspondence: sommer@di.ku.dk
† These authors contributed equally to this work.

Abstract: Computing sample means on Riemannian manifolds is typically computationally costly, as exemplified by computation of the Fréchet mean, which often requires finding minimizing geodesics to each data point for each step of an iterative optimization scheme. When closed-form expressions for geodesics are not available, this leads to a nested optimization problem that is costly to solve. The implied computational cost impacts applications in both geometric statistics and in geometric deep learning. The weighted diffusion mean offers an alternative to the weighted Fréchet mean. We show how the diffusion mean and the weighted diffusion mean can be estimated with a stochastic simulation scheme that does not require nested optimization. We achieve this by conditioning a Brownian motion in a product manifold to hit the diagonal at a predetermined time. We develop the theoretical foundation for the sampling-based mean estimation, we develop two simulation schemes, and we demonstrate the applicability of the method with examples of sampled means on two manifolds.

Keywords: diffusion mean; Fréchet mean; bridge simulation; geometric statistics; geometric deep learning

check for updates

Citation: Højgaard Jensen, M.; Sommer, S. Mean Estimation on the Diagonal of Product Manifolds. *Algorithms* **2022**, *15*, 92. https://doi.org/10.3390/a15030092

Academic Editor: Stephanie Allassonniere

Received: 5 February 2022
Accepted: 3 March 2022
Published: 10 March 2022

1. Introduction

The Euclidean expected value can be generalized to geometric spaces in several ways. Fréchet [1] generalized the notion of mean values to arbitrary metric spaces as minimizers of the sum of squared distances. Fréchet's notion of mean values therefore naturally includes means on Riemannian manifolds. On manifolds without metric, for example, affine connection spaces, a notion of the mean can be defined by exponential barycenters, see, e.g., [2,3]. Recently, Hansen et al. [4,5] introduced a probabilistic notion of a mean, the diffusion mean. The diffusion mean is defined as the most likely starting point of a Brownian motion given the observed data. The variance of the data is here modelled in the evaluation time $T > 0$ of the Brownian motion, and Varadhan's asymptotic formula relating the heat kernel with the Riemannian distance relates the diffusion mean and the Fréchet mean in the $T \to 0$ limit.

Computing sample estimators of geometric means is often difficult in practice. For example, estimating the Fréchet mean often requires evaluating the distance to each sample point at each step of an iterative optimization to find the optimal value. When closed-form solutions of geodesics are not available, the distances are themselves evaluated by minimizing over curves ending at the data points, thus leading to a nested optimization problem. This is generally a challenge in geometric statistics, the statistical analysis of geometric data. However, it can pose an even greater challenge in geometric deep learning, where a weighted version of the Fréchet mean is used to define a generalization of the Euclidean convolution taking values in a manifold [6]. As the mean appears in each layer of the network, closed-form geodesics is in practice required for its evaluation to be sufficiently efficient.

As an alternative to the weighted Fréchet mean, Ref. [7] introduced a corresponding weighted version of the diffusion mean. Estimating the diffusion mean usually requires ability to evaluate the heat kernel making it often similarly computational difficult to estimate. However, Ref. [7] also sketched a simulation-based approach for estimating the (weighted) diffusion mean that avoids numerical optimization and estimation of the heat kernel. Here, a mean candidate is generated by simulating a single forward pass of a Brownian motion on a product manifold conditioned to hit the diagonal of the product space. The idea is sketched for samples in $\mathbb{R}^2$ in Figure 1.

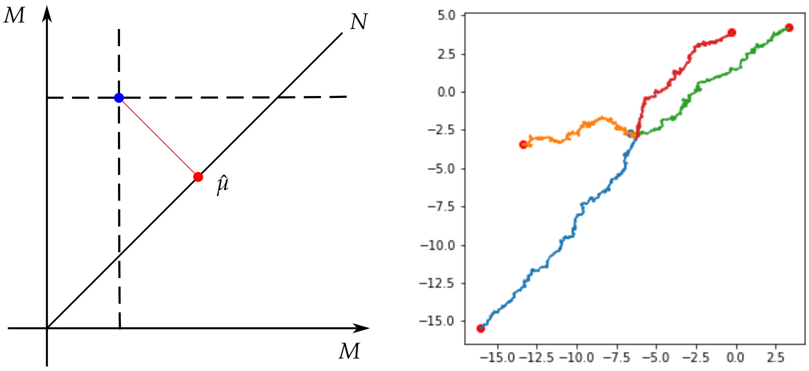


Figure 1. **(left)** The mean estimator viewed as a projection onto the diagonal of a product manifold. Given a set $x_1, \dots, x_n \in M$, the tuple $(x_1, \dots, x_n)$ (blue dot) belongs to the product manifold $M \times \cdots \times M$. The mean estimator $\hat{\mu}$ can be identified with the projection of $(x_1, \dots, x_n)$ onto the diagonal N (red dot). **(right)** Diffusion mean estimator in $\mathbb{R}^2$ using Brownian bridges conditioned on the diagonal. Here a Brownian bridge $X_t = (X_{1,t}, \dots, X_{4,t})$ in $\mathbb{R}^8$ is conditioned on hitting the diagonal $N \subseteq \mathbb{R}^8$ at time $T > 0$. The components X_j each being two-dimensional processes are shown in the plot.

Contribution

In this paper, we present a comprehensive investigation of the simulation-based mean sampling approach. We provide the necessary theoretical background and results for the construction, we present two separate simulation schemes, and we demonstrate how the schemes can be used to compute means on high-dimensional manifolds.

2. Background

We here outline the necessary concepts from Riemannian geometry, geometric statistics, stochastic analysis, and bridge sampling necessary for the sampling schemes presented later in the paper.

2.1. Riemannian Geometry

A Riemannian metric g on a d-dimensional differentiable manifold M is a family of inner products $(g_p)_{p \in M}$ on each tangent space T_pM varying smoothly in p. The Riemannian metric allows for geometric definitions of, e.g., length of curves, angles of intersections, and volumes on manifolds. A differentiable curve on M is a map $\gamma\colon [0,1] \to M$ for which the time derivative $\gamma'(t)$ belongs to $T_{\gamma_t}M$, for each $t \in (0,1)$. The length of the differentiable curve can then be determined from the Riemannian metric by $L(\gamma) := \int_0^1 \sqrt{g_{\gamma_t}(\gamma'(t), \gamma'(t))}dt = \int_0^1 \|\gamma'(t)\|_{\gamma_t} dt$. Let $p, q \in M$ and let Γ be the set of differentiable curves joining p and q, i.e., $\Gamma = \{\gamma\colon [0,1] \to M | \gamma(0) = p \text{ and } \gamma(1) = q\}$. The (Riemannian) distance between p and q is defined as $d(p,q) = \min_{\gamma \in \Gamma} L(\gamma)$. Minimizing curves are called geodesics.

A manifold can be parameterized using coordinate charts. The charts consist of open subsets of M providing a global cover of M such that each subset is diffeomorphic to an open subset of $\mathbb{R}^d$, or, equivalently, $\mathbb{R}^d$ itself. The exponential normal chart is often a convenient choice to parameterize a manifold for computational purposes. The exponential chart is related to the exponential map $\exp_p : T_pM \to M$ that for each $p \in M$ is given by $\exp_p(v) = \gamma_v(1)$, where γ_v is the unique geodesic satisfying $\gamma_v(0) = p$ and $\gamma_v'(0) = v$. For each $p \in M$, the exponential map is a diffeomorphism from a star-shaped subset V centered at the origin of T_pM to its image $\exp_p(V) \subseteq M$, covering all M except for a subset of (Riemannian) volume measure zero, $\mathrm{Cut}(p)$, the cut-locus of p. The inverse map $\log_p : M\setminus \mathrm{Cut}(p) \to T_pM$ provides a local parameterization of M due to the isomorphism between T_pM and $\mathbb{R}^d$. For submanifolds $N \subseteq M$, the cut-locus $\mathrm{Cut}(N)$ is defined in a fashion similar to $\mathrm{Cut}(p)$, see e.g., [8].

Stochastic differential equations on manifolds are often conveniently expressed using the frame bundle FM, the fiber bundle which for each point $p \in M$ assigns a frame or basis for the tangent space T_pM, i.e., FM consists of a collection of pairs (p, u), where $u : \mathbb{R}^d \to T_pM$ is a linear isomorphism. We let π denote the projection $\pi : FM \to M$. There exist a subbundle of FM consisting of orthonormal frames called the orthonormal frame bundle OM. In this case, the map $u : \mathbb{R}^d \to T_pM$ is a linear isometry.

2.2. Weighted Fréchet Mean

The Euclidean mean has three defining properties: The *algebraic property* states the uniqueness of the arithmetic mean as the mean with residuals summing to zero, the *geometric property* defines the mean as the point that minimizes the variance, and the *probabilistic property* adheres to a maximum likelihood principle given an i.i.d. assumption on the observations (see also [9] Chapter 2). Direct generalization of the arithmetic mean to non-linear spaces is not possible due to the lack of vector space structure. However, the properties above allow giving candidate definitions of mean values in non-linear spaces.

The Fréchet mean [1] uses the geometric property by generalizing the mean-squared distance minimization property to general metric spaces. Given a random variable X on a metric space (E, d), the Fréchet mean is defined by

$$\mu = \underset{p \in E}{\arg\min}\, \mathbb{E}\Big[d(p, X)^2\Big]. \tag{1}$$

In the present context, the metric space is a Riemannian manifold M with Riemannian distance function d. Given realizations $x_1, \ldots, x_n \in M$ from a distribution on M, the estimator of the weighted Fréchet mean is defined as

$$\hat{\mu} = \underset{p \in M}{\arg\min} \sum_{i=1}^{n} w_i d(p, x_i)^2, \tag{2}$$

where $w_1, \ldots, w_n$ are the corresponding weights satisfying $w_i > 0$ and $\sum_i w_i = 1$. When the weights are identical, (2) is an estimator of the Fréchet mean. Throughout, we shall make no distinction between the estimator and the Fréchet mean and will refer to both as the Fréchet mean.

In [6,10], the weighted Fréchet mean was used to define a generalization of the Euclidean convolution to manifold-valued inputs. When closed-form solutions of geodesics are available, the weighted Fréchet mean can be estimated efficiently with a recursive algorithm, also denoted an inductive estimator [6].

2.3. Weighted Diffusion Mean

The diffusion mean [4,5] was introduced as a geometric mean satisfying the probabilistic property of the Euclidean expected value, specifically as the starting point of a Brownian motion that is most likely given observed data. This results in the diffusion t-mean definition

$$\mu_t = \underset{p \in M}{\arg\min}\, \mathbb{E}[-\log p_t(p, X)], \tag{3}$$

where $p_t(\cdot,\cdot)$ denotes the transition density of a Brownian motion on M. Equivalently, p_t denotes the solution to the heat equation $\partial u/\partial t = \frac{1}{2}\Delta u$, where Δ denotes the Laplace-Beltrami operator associated with the Riemannian metric. The definition allows for an interpretation of the mean as an extension of the Fréchet mean due to Varadhan's result stating that $\lim_{t\to 0} -2t \log p_t(x,y) = d(x,y)^2$ uniformly on compact sets disjoint from the cut-locus of either x or y [11].

Just as the Fréchet mean, the diffusion mean has a weighted version, and the corresponding estimator of the weighted diffusion t-mean is given as

$$\hat{\mu}_t = \arg\min_{p\in M} \sum_{i=1}^{n} -\log p_{t/w_i}(p, x_i). \tag{4}$$

Please note that the evaluation time is here scaled by the weights. This is equivalent to scaling the variance of the steps of the Brownian motion [12].

As closed-form expressions for the heat kernel are only available on specific manifolds, evaluating the diffusion t-mean often rely on numerical methods. One example of this is using bridge sampling to numerically estimate the transition density [9,13]. If a global coordinate chart is available, the transition density can be written in the form (see [14,15])

$$p_T(z,v) = \sqrt{\frac{\det g(v)}{(2\pi T)^2}} e^{-\frac{\|a(z)(z-v)\|^2}{2T}} \mathbb{E}[\varphi], \tag{5}$$

where g is the metric matrix, a a square root of g, and φ denotes the correction factor between the law of the true diffusion bridge and the law of the simulation scheme. The expectation over the correction factor can be numerically approximated using Monte Carlo sampling. The correction factor will appear again when we discuss guided bridge proposals below.

2.4. Diffusion Bridges

The proposed sampling scheme for the (weighted) diffusion mean builds on simulation methods for conditioned diffusion processes, diffusion bridges. Here, we outline ways to simulate conditioned diffusion processes numerically in both the Euclidean and manifold context.

Euclidean Diffusion Bridges

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space, and X a d-dimensional Euclidean diffusion $[0,T]$ satisfying the stochastic differential equation (SDE)

$$dX_t = b_t(X_t)dt + \sigma_t(X_t)dW_t, \quad X_0 = x, \tag{6}$$

where W is a d-dimensional Brownian motion. Let $v \in \mathbb{R}^d$ be a fixed point. Conditioning X on reaching v at a fixed time $T > 0$ gives the bridge process $X|X_T = v$. Denoting this process Y, Doob's h-transform shows that Y is a solution of the SDE (see e.g., [16])

$$\begin{aligned} dY_t &= \tilde{b}_t(Y_t)dt + \sigma_t(Y_t)d\tilde{W}_t, \quad Y_0 = x \\ \tilde{b}_t(y) &= b_t(y) + a_t(y)\nabla_y \log p_{T-t}(y,v), \end{aligned} \tag{7}$$

where $p_t(\cdot,\cdot)$ denotes the transition density of the diffusion X, $a = \sigma\sigma^T$, and where $\tilde{W}$ is a d-dimensional Brownian motion under a changed probability law. From a numerical viewpoint, in most cases, the transition density is intractable and therefore direct simulation of (7) is not possible.

If we instead consider a Girsanov transformation of measures to obtain the system (see, e.g., [17] Theorem 1)

$$\begin{aligned} dY_t &= \tilde{b}_t(Y_t)dt + \sigma_t(Y_t)d\tilde{W}_t, \quad Y_0 = x \\ \tilde{b}_t(y) &= b_t(y) + \sigma_t(y)h(t,y), \end{aligned} \tag{8}$$

the corresponding change of measure is given by

$$\left.\frac{d\mathbb{P}^h}{d\mathbb{P}}\right|_{\mathcal{F}_t} = e^{\int_0^t h(s,X_s)^T dW_s - \frac{1}{2}\int_0^t \|h(s,X_s)\|^2 ds}. \tag{9}$$

From (7), it is evident that $h(t,x) = \sigma^T \nabla_x \log p_{T-t}(x,v)$ gives the diffusion bridge. However, different choices of the function h can yield processes which are absolutely continuous regarding the actual bridges, but which can be simulated directly.

Delyon and Hu [17] suggested to use $h(t,x) = \sigma_t^{-1}(x)\nabla_x \log q_{T-t}(x,v)$, where q denotes the transition density of a standard Brownian motion with mean v, i.e., $q_t(x,v) = (2\pi t)^{-d/2}\exp(-\|x-v\|^2/2t)$. They furthermore proposed a method that would disregard the drift term b, i.e., $h(t,x)) = \sigma_t^{-1}(x)\nabla_x \log q_{T-t}(x,v) - \sigma_t^{-1}(x)b_t(x)$. Under certain regularity assumptions on b and σ, the resulting processes converge to the target in the sense that $\lim_{t\to T} Y_t = v$ a.s. In addition, for bounded continuous functions f, the conditional expectation is given by

$$\mathbb{E}[f(X)|X_T = v] = C\mathbb{E}[f(Y)\varphi(y)], \tag{10}$$

where φ is a functional of the whole path Y on $[0,T]$ that can be computed directly. From the construction of the h-function, it can be seen that the missing drift term is accounted for in the correction factor φ.

The simulation approach of [17] can be improved by the simulation scheme introduced by Schauer et al. [18]. Here, an h-function defined by $h(t,x) = \nabla_x \log \hat{p}_{T-t}(x,v)$ is suggested, where $\hat{p}$ denotes the transition density of an auxiliary diffusion process with known transition densities. The auxiliary process can for example be linear because closed-form solutions of transition densities for linear processes are available. Under the appropriate measure $\mathbb{P}^h$, the guided proposal process is a solution to

$$dY_t = b_t(Y_t)dt + a_t(Y_t)\nabla_x \log \hat{p}_{T-t}(x,v)|_{x=Y_t}dt + \sigma_t(Y_t)dW_t. \tag{11}$$

Note the factor $a(t,y)$ in the drift in (7) which is also present in (11) but not with the scheme proposed by [17]. Moreover, the choice of a linear process grants freedom to model. For other choices of an h-functions see e.g., [19,20].

Marchand [19] extended the ideas of Delyon and Hu by conditioning a diffusion process on partial observations at a finite collection of deterministic times. Where Delyon and Hu considered the guided diffusion processes satisfying the SDE

$$dY_t = b_t(Y_t)dt - \frac{Y_t - v}{T-t}dt + \sigma_t(Y_t)dw_t, \tag{12}$$

for $v \in \mathbb{R}^d$ over the time interval $[0,T]$, Marchand proposed the guided diffusion process conditioned on partial observations $v_1,\ldots,v_N$ solving the SDE

$$dY_t = b_t(Y_t)dt - \sum_{k=1}^n P_t^k(Y_t)\frac{Y_t - u_k}{T_k - t}1_{(T_k-\varepsilon_k,T_k)}dt + \sigma_t(Y_t)dw_t, \tag{13}$$

where u_k is be any vector satisfying $L_k(x)u_k = v_k$, L_k a deterministic matrix in $M_{m_k,n}(\mathbb{R})$ whose m_k rows form a orthonormal family, P_t^k are projections to the range of L_k, and $T_k - \varepsilon_k < T_k$. The ε_k only allow the application of the guiding term on a part of the time intervals $[T_{k-1},T_k]$. We will only consider the case $k=1$. The scheme allows the sampling of bridges conditioned on $LY_T = v$.

2.5. Manifold Diffusion Processes

To work with diffusion bridges and guided proposals on manifolds, we will first need to consider the Eells–Elworthy–Malliavin construction of Brownian motion and the connected characterization of semimartingales [21]. Endowing the frame bundle FM with a connection allows splitting the tangent bundle TFM into a horizontal and vertical part. If the connection on FM is a lift of a connection on M, e.g., the Levi–Civita connection of a metric on M, the horizontal part of the frame bundle is in one-to-one correspondence with M. In addition, there exist fundamental horizontal vector fields $H_i\colon FM \to HFM$ such that for any continuous $\mathbb{R}^d$-valued semimartingale Z the process U defined by

$$dU_t = H_i(U_t) \circ dZ_t^i, \tag{14}$$

is a horizontal frame bundle semimartingale, where $\circ$ denotes integration in the Stratonovich sense. The process $X_t := \pi(U_t)$ is then a semimartingale on M. Any semimartingale X_t on M has this relation to a Euclidean semimartingale Z_t. X_t is denoted the development of Z_t, and Z_t the antidevelopment of X_t. We will use this relation when working with bridges on manifolds below.

When Z_t is a Euclidean Brownian motion, the development X_t is a Brownian motion. We can in this case also consider coordinate representations of the process. With an atlas $\{(D_\alpha, \phi_\alpha)\}_\alpha$ of M, there exists an increasing sequence of predictable stopping times $0 \leq T_k \leq T_{k+1}$ such that on each stochastic interval $[\![T_k, T_{k+1}]\!] = \{(\omega, t) \in \Omega \times \mathbb{R}_+ | T_k(\omega) \leq t \leq T_{k+1}(\omega)\}$ the process $x_t \in D_\alpha$, for some α (see [22] Lemma 3.5). Thus, the Brownian motion x on M can be described locally in a chart $D_\alpha \subset M$ as the solution to the system of SDEs, for $(\omega, t) \in [\![T_k, T_{k+1}]\!] \cap \{T_k < T_{k+1}\}$

$$dx_t^i(\omega) = b^i(x_t(\omega))dt + \sigma_j^i(x_t(\omega))dW_t^j(\omega), \tag{15}$$

where σ denotes the matrix square root of the inverse of the Riemannian metric tensor (g^{ij}) and $b^k(x) = -\frac{1}{2}g^{ij}(x)\Gamma_{ij}^k(x)$ is the contraction over the Christoffel symbols (see, e.g., [11] Chapter 3). Strictly speaking, the solution of Equation (15) is defined by $x_t^i = \phi_\alpha(x_t)^i$.

We thus have two concrete SDEs for the Brownian motion in play: The FM SDE (14) and the coordinate SDE (15).

Throughout the paper, we assume that M is stochastically complete, i.e., the Brownian motions does not explode in finite time and, consequently, $\int_M p_t(x,y) d\,\mathrm{Vol}_M(y) = 1$, for all $t > 0$ and all $x \in M$.

2.6. Manifold Bridges

The Brownian bridge process Y on M conditioned at $Y_T = v$ is a Markov process with generator $\frac{1}{2}\Delta + \nabla \log p_{T-t}(\cdot, v)$. Closed-form expressions of the transition density of a Brownian motion are available on selected manifolds including Euclidean spaces, hyperbolic spaces, and hyperspheres. Direct simulation of Brownian bridges is therefore possible in these cases. However, generally, transition densities are intractable and auxiliary processes are needed to sample from the desired conditional distributions.

To this extent, various types of bridge processes on Riemannian manifolds have been described in the literature. In the case of manifolds with a pole, i.e., the existence of a point $p \in M$ such that the exponential map $\exp_p\colon T_pM \to M$ is a diffeomorphism, the *semi-classical* (Riemannian Brownian) bridge was introduced by Elworthy and Truman [23] as the process with generator $\frac{1}{2}\Delta + \nabla \log k_{T-t}(\cdot, v)$, where

$$k_t(x, v) = (2\pi t)^{-n/2} e^{-\frac{d(x,v)^2}{2t}} J^{-1/2}(x),$$

and $J(x) = |\det D_{\exp(v)^{-1}} \exp_v|$ denotes the Jacobian determinant of the exponential map at v. Elworthy and Truman used the semi-classical bridge to obtain heat-kernel estimates, and the semi-classical bridge has been studied by various others [24,25].

By Varadhan's result (see [11] Theorem 5.2.1), as $t \to T$, we have the asymptotic relation $((T-t)\log p_{T-t}(x,y) \sim -\frac{1}{2}d(x,y)^2$. In particular, the following asymptotic relation was shown to hold by Malliavin, Stroock, and Turetsky [26,27] : $(T-t)\nabla \log p_{T-t}(x,y) \sim -\frac{1}{2}\nabla d(x,y)^2$. From these results, the generators of the Brownian bridge and the semi-classical bridge differ in the limit by a factor of $-\frac{1}{2}\nabla \log J(x)$. However, under a certain boundedness condition, the two processes can be shown to be identical under a changed probability measure [8] Theorem 4.3.1.

To generalize the heat-kernel estimates of Elworthy and Truman, Thompson [8,28] considered the *Fermi* bridge process conditioned to arrive in a submanifold $N \subseteq M$ at time $T > 0$. The Fermi bridge is defined as the diffusion process with generator $\frac{1}{2}\Delta + \nabla \log q_{T-t}(\cdot, N)$, where

$$q_t(x,N) = (2\pi t)^{-n/2} e^{-\frac{d(x,N)^2}{2t}}.$$

For both bridge processes, when $M = \mathbb{R}^d$ and N is a point, both the semi-classical bridge and the Fermi bridge agree with the Euclidean Brownian bridge.

Ref [15] introduce a numerical simulation scheme for conditioned diffusions on Riemannian manifolds, which generalize the method by Delyon and Hu [17]. The guiding term used is identical to the guiding term of the Fermi bridge when the submanifold is a single point v.

3. Diffusion Mean Estimation

The standard setup for diffusion mean estimation described in the literature (e.g., [13]) is as follows: Given a set of observations $x_1, \dots, x_n \in M$, for each observation x_i, sample a guided bridge process approximating the bridge $X_{i,t}|X_{i,T} = x_i$ with starting point x_0. The expectation over the correction factors can be computed from the samples, and the transition density can be evaluated using (5). An iterative maximum likelihood approach using gradient descent to update x_0 yielding an approximation of the diffusion mean in the final value of x_0. The computation of the diffusion mean, in the sense just described, is, similarly to the Fréchet mean, computationally expensive.

We here explore the idea first put forth in [7]: We turn the situation around to simulate n independent Brownian motions starting at each of $x_1, \dots, x_n$, and we condition the n processes to coincide at time T. We will show that the value $x_{1,T} = \cdots = x_{n,T}$ is an estimator of the diffusion mean. By introducing weights in the conditioning, we can similarly estimate the weighted diffusion mean. The construction can concisely be described as a single Brownian motion on the n-times product manifold M^n conditioned to hit the diagonal $\mathrm{diag}(M^n) = \{(x,\dots,x)|x \in M\} \subset M^n$. To shorten notation, we denote the diagonal submanifold N below. We start with examples with M Euclidean to motivate the construction.

Example 1. *Consider the two-dimensional Euclidean multivariate normal distribution*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right).$$

The conditional distribution of X given $Y = y$ follows a univariate normal distribution

$$X|Y = y \sim N\Big(\mu_1 + \sigma_{12}\sigma_{22}^{-1}(y-\mu_2), \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21}\Big).$$

This can be seen from the fact that if $X \sim N(\mu, \Sigma)$ then for any linear transformation $AX + b \sim N\big(b + A\mu, A\Sigma A^T\big)$. Defining the random variable $Z = X - \sigma_{12}\sigma_{22}^{-1}Y$, the result applied to (Z, X) gives $Z \sim N\Big(\mu_1 - \sigma_{12}\sigma_{22}^{-1}\mu_2, \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21}\Big)$. The conclusion then follows from $X = Z + \sigma_{12}\sigma_{22}^{-1}Y$. Please note that X and Y are independent if and only if $\sigma_{12} = \sigma_{21} = 0$ and the conditioned random variable is in this case identical in law to X.

Let now $x_1, \ldots, x_n \in M$ be observations and let $x = (x_1, \ldots, x_n) \in M^n$ be an element of the n-product manifold $M \times \cdots \times M$ with the product Riemannian metric. We again first consider the case $M = \mathbb{R}^d$:

Example 2. *Let $Y_i \sim N\left(x_i, \frac{T}{w_i} I_d\right)$ be independent random variables. The conditional distribution $Y_1 | Y_1 = \cdots = Y_n$ is normal $N\left(\frac{\sum_i w_i x_i}{\sum_i w_i}, \frac{T}{\sum_i w_i}\right)$. This can be seen inductively: The conditioned random variable $Y_1 | Y_1 = Y_2$ is identical to $Y_1 | Y_1 - Y_2 = 0$. Now let $X = Y_1$ and $Y = Y_1 - Y_2$ and refer to Example 1. To conclude, assume $Z_n := Y_1 | Y_1 = \cdots, = Y_{n-1}$ follows the desired normal distribution. Then $Z_n | Z_n = Y_n$ is normally distributed with the desired parameters and $Z_n | Z_n = Y_n$ is identical to $Y_1 | Y_1 = \cdots = Y_n$.*

The following example establishes the weighted average as a projection onto the diagonal.

Example 3. *Let x be a point in $(\mathbb{R}^d)^n$ and let P be the orthogonal projection to the diagonal of $(\mathbb{R}^d)^n$ such that $Px = \left(\frac{1}{nd} \sum_{i=1}^{nd} x_i \ldots \frac{1}{nd} \sum_{i=1}^{nd} x_i\right)^T$. We see that the projection yields n copies of the arithmetic mean of the coordinates. This is illustrated in Figure 2.*

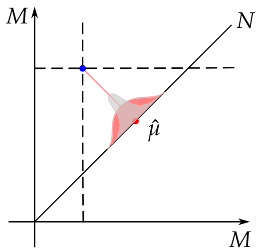


Figure 2. The mean estimator viewed as a projection onto the diagonal of a product manifold. Conditioning on the closest point in the diagonal yields a density on the diagonal depending on the time to arrival $T > 0$. As T tends to zero the density convergence to the Dirac-delta distribution (grey), whereas as T increases the variance of the distribution increases (rouge).

The idea of conditioning diffusion bridge processes on the diagonal of a product manifold originates from the facts established in Examples 1–3. We sample the mean by sampling from the conditional distribution $Y_1 | Y_1 = \cdots = Y_n$ from Example 2 using a guided proposal scheme on the product manifolds M^n and on each step of the sampling projecting to the diagonal as in Example 3.

Turning now to the manifold situation, we replace the normal distributions with mean $x_i \in \mathbb{R}^d$ and variance T/w_i with Brownian motions started at $x_i \in M$ and evaluated at time T/w_i. Please note that the Brownian motion density, the heat kernel, is symmetric in its coordinates: $p_t(x, y) = p_t(y, x)$. We will work with multiple process and indicate with superscript the density with respect to a particular process, e.g., p_T^X. Note also that change of the evaluation time T is equal to scaling the variance, i.e., $p_{\alpha T}^X(x, y) = p_T^{X^\alpha}(x, y)$ where X^α is a Brownian motion with variance of the increments scaled by $\alpha > 0$. This gives the following theorem, first stated in [7] with sketch proof:

Theorem 1. *Let $X_t = (X_{1,t}^{w_1^{-1}}, \ldots, X_{n,t}^{w_n^{-1}})$ consist of n independent Brownian motions on M with variance w_i^{-1} and $X_{i,0} = x_i$, and let $\mathbb{P}^*$ the law of the conditioned process $Y_t = X_t | X_T \in N$, $N = \mathrm{diag}(M^n)$. Let v be the random variable $Y_{1,T}$. Then v has density $p_v^Y(y) \propto \prod_{i=1}^n p_{T/w_i}(x_i; y)$ and $v = Y_{i,T}$ for all i a.s. (almost surely).*

Proof. $p_T^X((x_1,\dots,x_n),(y,\dots,y)) = \prod_{i=1}^n p_T^{X_i^{w_i^{-1}}}(x_i,y)$ because the processes $X_{i,t}$ are independent. By symmetry of the Brownian motion and the time rescaling property, $p_T^{X_i^{w_i^{-1}}}(x_i,y) = p_{T/w^i}(y,x_i)$. For elements $(y,\dots,y) \in \mathrm{diag}(M^n)$ and $x \in M^n$, $p_v(y) = p_T^Y(x,y) \propto p_T^X(x,y)$. As a result of the conditioning, $v = Y_{1,T} = \cdots = Y_{n,T}$. In combination, this establishes the result. □

Consequently, the set of modes of p_v equal the set of the maximizers for the likelihood $L(y;x_1,\dots,x_n) = \prod_{i=1}^n p_{T/w_i}(x_i;y)$ and thus the weighted diffusion mean. This result is the basis for the sampling scheme. Intuitively, if the distribution of v is relatively well behaved (e.g., close to normal), a sample from v will be close to a weighted diffusion mean with high probability.

In practice, however, we cannot sample Y_t directly. Instead, we will below use guided proposal schemes resulting in processes $\tilde{Y}_t$ with law $\tilde{\mathbb{P}}$ that we can actually sample and that under certain assumptions, will be absolutely continuous with respect to Y_t with explicitly computable likelihood ratio so that $\mathbb{P}^* = \frac{\varphi(\tilde{Y}_T)}{\mathbb{E}^{\tilde{\mathbb{P}}}[\varphi(\tilde{Y}_T)]}\tilde{\mathbb{P}}$.

Corollary 1. *Let $\tilde{\mathbb{P}}$ be the law of $\tilde{Y}_t$ and φ be the corresponding correction factor of the guiding scheme. Let $\tilde{v}$ be the random variable $\tilde{Y}_{1,T}$ with law $\frac{\varphi(\tilde{Y}_T)}{\mathbb{E}^{\tilde{\mathbb{P}}}[\varphi(\tilde{Y}_T)]}\tilde{\mathbb{P}}$. Then $\tilde{v}$ has density $p_{\tilde{v}}(y) \propto \prod_{i=1}^n p_{T/w_i}(x_i;y)$.*

We now proceed to actually construct the guided sampling schemes.

3.1. Fermi Bridges to the Diagonal

Consider a Brownian motion $X_t = (X_{1,t},\dots X_{n,t})$ in the product manifold M^n conditioned on $X_{1,T} = \cdots = X_{n,T}$ or, equivalently, $X_T \in N$, $N = \mathrm{diag}(M^n)$. Since N is a submanifold of M^n, the conditioned diffusion defined above is absolutely continuous with respect to the Fermi bridge on $[0,T)$ [8,28]. Define the FM-valued horizontal guided process

$$dU_t = H_i(U_t) \circ \left(dW_t^i - \frac{H_i \tilde{r}_N^2(U_t)}{2(T-t)} dt \right), \tag{16}$$

where $\tilde{r}$ denotes the lift of the radial distance to N defined by $\tilde{r}_N(u) := r_N(\pi(u)) = d(\pi(u), N)$. The Fermi bridge Y^F is the projection of U to M, i.e., $Y_t^F := \pi(U_t)$. Let $\mathbb{P}^F$ denotes its law.

Theorem 2. *For all continuous bounded functions f on M^n, we have*

$$\mathbb{E}[f(X)|X_{1,T} = \cdots = X_{n,T}] = \lim_{t\uparrow T} C\mathbb{E}^{\mathbb{P}^F}[f(Y)\varphi(Y)], \tag{17}$$

with a constant $C > 0$, where

$$d\log\varphi(Y_s^F) = \frac{r_N(Y_s^F)}{T-s}(d\eta_s + dL_s) \quad with \quad d\eta_s = \frac{\partial}{\partial r_N}\log\Theta_N^{-\frac{1}{2}} ds,$$

$dL_s := d\mathbb{L}_s(Y^F)$ with $\mathbb{L}$ being the geometric local time at $\mathrm{Cut}(N)$, and Θ_N the determinant of the derivative of the exponential map normal to N with support on $M^n \setminus \mathrm{Cut}(N)$ [8].

Proof. From [15] Theorem 8 and [28],

$$\mathbb{E}[f(X)|X_T \in N] = \lim_{t \uparrow T} C\mathbb{E}^{\mathbb{P}^F}\left[f(Y^F)\varphi(Y_t^F)\right].$$

□

Since N is a totally geodesic submanifold of dimension d, the results of [8] can be used to give sufficient conditions to extend the equivalence in (17) to the entire interval $[0, T]$. A set A is said to be polar for a process X_t if the first hitting time of A by X is infinity a.s.

Corollary 2. *If either of the following conditions are satisfied*

(i) *the sectional curvature of planes containing the radial direction is non-negative or the Ricci curvature in the radial direction is non-negative;*

(ii) $\mathrm{Cut}(N)$ *is polar for the Fermi bridge* Y^F *and either the sectional curvature of planes containing the radial direction is non-positive or the Ricci curvature in the radial direction is non-positive;*

then

$$\mathbb{E}[f(X)|X_{1,T} = \cdots = X_{n,T}] = C\mathbb{E}^{\mathbb{P}^F}\left[f(Y^F)\varphi(Y_T^F)\right].$$

In particular, $\frac{\varphi(Y_T^F)}{\mathbb{E}^{\mathbb{P}^F}[\varphi(Y_T^F)]} d\mathbb{P}^F \propto d\mathbb{P}^*$.

Proof. See [8] (Appendix C.2). □

For numerical purposes, the equivalence (17) in Theorem 2 is sufficient as the interval $[0, T]$ is finitely discretized. To obtain the result on the full interval, the conditions in Corollary 2 may at first seem quite restrictive. A sufficient condition for a subset of a manifold to be polar for a Brownian motion is its Hausdorff dimension being two less than the dimension of the manifold. Thus, $\mathrm{Cut}(N)$ is polar if $\dim(\mathrm{Cut}(N)) \leq nd - 2$. Verifying whether this is true requires specific investigation of the geometry of M^n.

The SDE (16) together with (17) and the correction φ gives a concrete simulation scheme that can be implemented numerically. Implementation of the geometric constructs is discussed in Section 4. The main complication of using Fermi bridges for simulation is that it involves evaluation of the radial distance r_N at each time-step of the integration. Since the radial distance finds the closest point on N to $x_1, \ldots, x_n$, it is essentially a computation of the Fréchet mean and thus hardly more computationally efficient than computing the Fréchet mean itself. For this reason, we present a coordinate-based simulation scheme below.

3.2. Simulation in Coordinates

We here develop a more efficient simulation scheme focusing on manifolds that can be covered by a single chart. The scheme follows the partial observation scheme developed [19]. Representing the product process in coordinates and using a transformation L, whose kernel is the diagonal $\mathrm{diag}(M^n)$, gives a guided bridge process converging to the diagonal. An explicit expression for the likelihood is given.

In the following, we assume that M can be covered by a chart in which the square root of the cometric tensor, denoted by σ, is C^2. Furthermore, σ and its derivatives are bounded; σ is invertible with bounded inverse. The drift b is locally Lipschitz and locally bounded.

Let $x_1, \ldots, x_n \in M$ be observations and let $X_{1,t}, \ldots, X_{n,t}$ be independent Brownian motions with $X_{1,0} = x_1, \ldots, X_{n,0} = x_n$. Using the coordinate SDE (15) for each $X_{i,t}$, we can write the entire system on M^n as

$$d\begin{pmatrix} X^1_{1,t} \\ \vdots \\ X^d_{1,t} \\ \vdots \\ X^1_{n,t} \\ \vdots \\ X^d_{n,t} \end{pmatrix} = \begin{pmatrix} b^1(X_{1,t}) \\ \vdots \\ b^d(X_{1,t}) \\ \vdots \\ b^1(X_{n,t}) \\ \vdots \\ b^d(X_{n,t}) \end{pmatrix} dt + \begin{bmatrix} \Sigma^1(X_{1,t},\dots,X_{n,t}) \\ \vdots \\ \Sigma^d(X_{1,t},\dots,X_{n,t}) \end{bmatrix} dW_t. \tag{18}$$

In the product chart, Σ and b satisfy the same assumptions as the metric and cometric tensor and drift listed above.

The conditioning $X_T \in N$ is equivalent to the requiring $X_T \in \operatorname{diag}((\mathbb{R}^d)^n)$ in coordinates. $\operatorname{diag}((\mathbb{R}^d)^n)$ is a linear subspace of $(\mathbb{R}^d)^n$, we let $L \in M^{d\times nd}$ be a matrix with orthonormal rows and $\ker L = \operatorname{diag}((\mathbb{R}^d)^n)$ so that the desired conditioning reads $LX_T = 0$. Define the following oblique projection, similar to [19],

$$P_t(x) = a(x)L^T A(x)L \tag{19}$$

where

$$a(x) = \Sigma(x)\Sigma(x)^T \quad \text{and} \quad A_t(x) = (La(x)L^T)^{-1}.$$

Set $\beta(x) = \Sigma(x)^T L^T A(x)$. The guiding scheme (13) then becomes

$$dY_t = b(Y_t)dt + \Sigma(Y_t)dW_t - \Sigma(Y_t)\beta(Y_t)\frac{LY_t}{T-t}1_{(T-\varepsilon,T)}(t)dt, \quad Y_0 = u. \tag{20}$$

We have the following result.

Lemma 1. *Equation (20) admits a unique solution on $[0,T)$. Moreover, $\|LY_t\|^2 \le C(\omega)(T-t)\log\log[(T-t)^{-1}+e]$ a.s., where C is a positive random variable.*

Proof. Since $LP = L$, the proof is similar to the proof of [19] Lemma 6. □

With the same assumptions, we also obtain the following result similar to [19] Theorem 3.

Theorem 3. *Let Y_t be a solution of (20), and assume the drift b is bounded. For any bounded function f,*

$$\mathbb{E}[f(X)|X_T \in N] = C\mathbb{E}[f(y)\varphi(Y)], \tag{21}$$

where C is a positive constant and

$$\varphi(Y_t) = \sqrt{\det(A(Y_T))}\exp\Bigg\{-\frac{\|\beta_{T-\varepsilon}(Y_{T-\varepsilon})LY_{T-\varepsilon}\|^2}{2\varepsilon} - \int_{T-\varepsilon}^{T}\frac{2(LY_s)^T Lb(Y_s)ds - (LY_s)^T d(A(Y_s))LY_s + d[A(Y_s)^{ij},(LY_s)_i(LY_s)_j]}{2(T-s)}\Bigg\}$$

Proof. A direct consequence of [19] Theorem 3, for $k = 1$, and Lemma 1. □

The theorem can also be applied for unbounded drift by replacing b with a bounded approximation and performing a Girsanov change of measure.

3.3. Accounting for φ

The sampling schemes (16) and (20) above provides samples on the diagonal and thus candidates for the diffusion mean estimates. However, the schemes sample from a distribution which is only equivalent to the bridge process distribution: We still need to handle the correction factor in the sampling to sample from the correct distribution, i.e., the rescaling $\frac{\varphi}{\mathbb{E}[\varphi]}$ of the guided law in Theorem 1. A simple way to achieve this is to do sampling importance resampling (SIR) as described in Algorithm 1. This yields an approximation of the weighted diffusion mean. For each sample y^i of the guided bridge process, we compute the corresponding correction factor $\varphi(y^i)$. The resampling step then consists of picking y_T^j with a probability determined by the correction terms, i.e., with J the number of samples we pick sample j with probability $P_j = \frac{\varphi(y_T^j)}{\sum_{i=1}^J \varphi(y_T^i)}$.

Algorithm 1: weighted Diffusion Mean

Input: Points $x_1, \ldots, x_n \in M$ **Output:** (weighted) diffusion mean sampling
for $j = 1$ *to* J **do**
  Sample path from guided process Y_t
  Record Y_T^j and compute correction factor $\varphi(Y_T^j)$
end
Sample j from $1, \ldots, J$ with probability $P_j = \frac{\varphi(Y_T^j)}{\sum_{k=1}^J \varphi(Y_T^k)}$.
`// Return` Y_T^j

It depends on the practical application if the resampling is necessary, or if direct samples from the guided process (corresponding to $J = 1$) are sufficient.

4. Experiments

We here exemplify the mean sampling scheme on the two-sphere $\mathbb{S}^2$ and on finite sets of landmark configurations endowed with the LDDMM metric [29,30]. With the experiment on $\mathbb{S}^2$, we aim to give a visual intuition of the sampling scheme and the variation in the diffusion mean estimates caused by the sampling approach. In the higher-dimensional landmark example where closed-form solutions of geodesics are not available, we compare to the Fréchet mean and include rough running times of the algorithms to give a sense of the reduced time complexity. Note, however, that the actual running times are very dependent on the details of the numerical implementation, stopping criteria for the optimization algorithm for the Fréchet mean, etc.

The code used for the experiments is available in the software package Jax Geometry http://bitbucket.org/stefansommer/jaxgeometry (accessed on 4 February 2022). The implementation uses automatic differentiation libraries extensively for the geometry computations as is further described [31].

4.1. Mean Estimation on $\mathbb{S}^2$

To illustrate the diagonal sampling scheme, Figure 3 displays a sample from a diagonally conditioned Brownian motion on $(\mathbb{S}^2)^n$, $n = 3$. The figure shows both the diagonal sample (red point) and the product process starting at the three data points and ending at the diagonal. In Figure 4, we increase the number of samples to $n = 256$ and sample 32 mean samples $(T = 0.2)$. The population mean is the north pole, and the samples can be seen to cluster closely around the population mean with little variation in the mean samples.

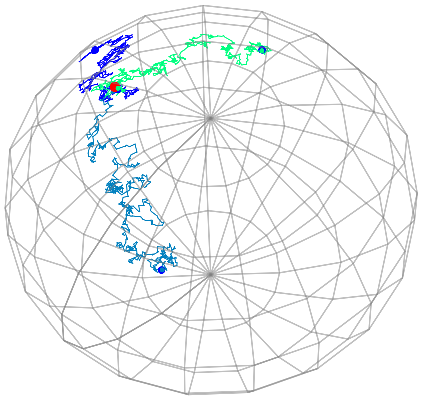


Figure 3. 3 points on $\mathbb{S}^2$ together with a sample mean (red) and the diagonal process in $(\mathbb{S}^2)^n$, $n = 3$ with $T = 0.2$ conditioned on the diagonal.

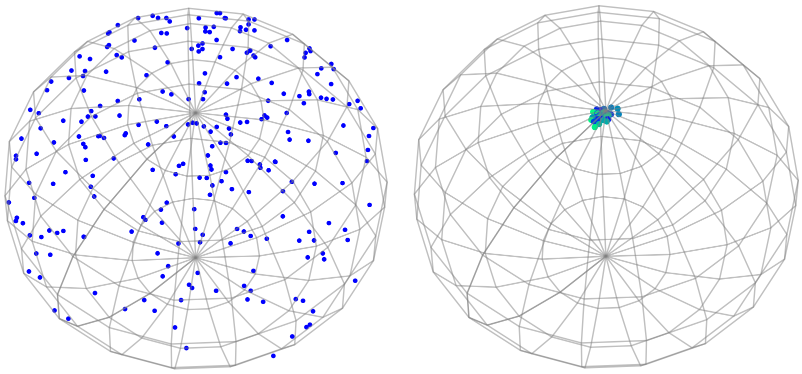


Figure 4. (**left**) 256 sampled data points on $\mathbb{S}^2$ (north pole being population mean). (**right**) 32 samples of the diffusion mean conditioned on the diagonal of $(\mathbb{S}^2)^n$, $n = 256$, $T = 0.2$. As can be seen, the variation in the mean samples is limited.

4.2. LDDMM Landmarks

We here use the same setup as in [13], where the diffusion mean is estimated by iterative optimization, to exemplify the mean estimation on a high-dimensional manifold. The data consists of annotations of left ventricles cardiac MR images [32] with 17 tlandmarks selected from the annotation set from a total of 14 images. Each configuration of 17 landmarks in $\mathbb{R}^2$ gives a point in a 34-dimensional shape manifold. We equip this manifold with the LDDMM Riemannian metric [29,30]. Please note that the configurations can be represented as points in $\mathbb{R}^{34}$, and the entire shape manifold is the subset of $\mathbb{R}^{34}$ where no two landmarks coincide. This provides a convenient Euclidean representation of the landmarks. The cometric tensor is not bounded in this representation, and we therefore cannot directly apply the results of the previous sections. We can nevertheless explore the mean simulation scheme experimentally.

Figure 5 shows one landmark configuration overlayed the MR image from which the configuration was annotated, and all 14 landmark configurations plotted together. Figure 6 displays samples from the diagonal process for two values of the Brownian motion end time T. Please note that each landmark configuration is one point on the 34-dimensional shape manifold, and each of the paths displayed is therefore a visualization of a Brownian path on this manifold. This figure and Figure 3 both show diagonal processes, but on two different manifolds.

In Figure 7, an estimated diffusion mean and Fréchet mean for the landmark configurations are plotted together. On a standard laptop, generation of one sample diffusion mean takes approximately 1 s. For comparison, estimation of the Fréchet mean with the standard nested optimization approach using the Riemannian logarithm map as implemented in Jax Geometry takes approximately 4 min. The diffusion mean estimation performed in [13] using direct optimization of the likelihood approximation with bridge sampling from the mean candidate to each data point is comparable in complexity to the Fréchet mean computation.

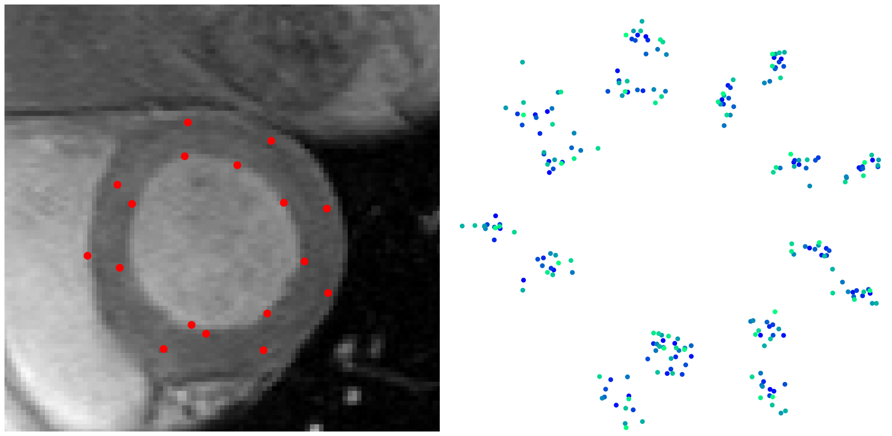


Figure 5. (**left**) One configuration of 17 landmarks overlayed the MR image from which the configuration was annotated. (**right**) All 14 landmark configurations plotted together (one color for each configuration of 17 landmarks).

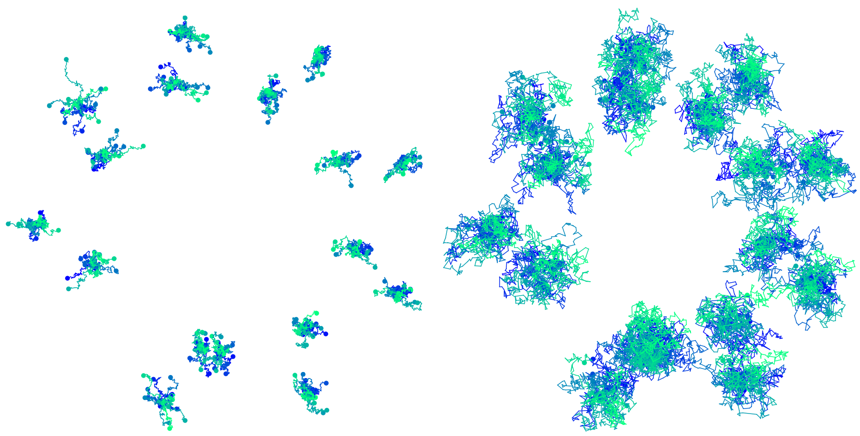


Figure 6. Samples from the diagonal process with $T = 0.2$ (**left**) and $T = 1$ (**right**). The effect of varying the Brownian motion end time T is clearly visible.

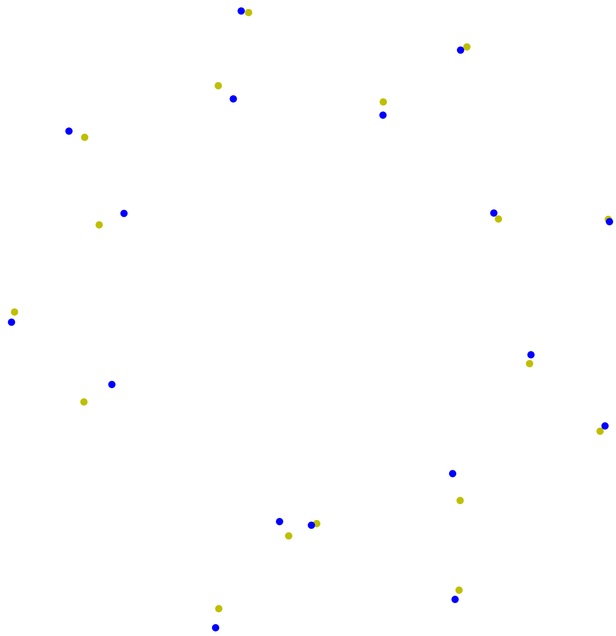


Figure 7. One sampled diffusion mean with the sampling scheme (blue configuration) together with estimated Fréchet mean (green configuration). The forward sampling scheme is significantly faster than the iterative optimization needed for the Fréchet mean on the landmark manifold where closed-form solutions of the geodesic equations are not available.

5. Conclusions

In [7], the idea of sampling means by conditioning on the diagonal of product manifolds was first described and the bridge sampling construction sketched. In the present paper, we have provided a comprehensive account of the background for the idea, including the relation between the (weighted) Fréchet and diffusion means, and the foundations in both geometry and stochastic analysis. We have constructed two simulation schemes and demonstrated the method on both low and a high-dimensional manifolds, the sphere $\mathbb{S}^2$ and the LDDMM landmark manifold, respectively. The experiments show the feasibility of the method and indicate the potential high reduction in computation time compared to computing means with iterative optimization.

Author Contributions: Conceptualization, S.S.; methodology, M.H.J. and S.S.; software, S.S.; validation, M.H.J. and S.S.; formal analysis, M.H.J. and S.S.; investigation, M.H.J. and S.S.; resources, S.S.; data curation, S.S.; writing—original draft preparation, M.H.J. and S.S.; writing—review and editing, M.H.J. and S.S.; visualization, S.S.; supervision, S.S.; project administration, M.H.J. and S.S.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Villum Fonden grant number 22924 and grant number 40582, and Novo Nordisk Fonden grant number NNF18OC0052000.

Acknowledgments: The work presented is supported by the CSGB Centre for Stochastic Geometry and Advanced Bioimaging funded by a grant from the Villum foundation, the Villum Foundation grants 22924 and 40582, and the Novo Nordisk Foundation grant NNF18OC0052000.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Fréchet, M. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. L'Institut Henri Poincaré* **1948**, *10*, 215–310.
2. Arnaudon, M.; Li, X.M. Barycenters of measures transported by stochastic flows. *Ann. Probab.* **2005**, *33*, 1509–1543. [CrossRef]
3. Pennec, X. Barycentric Subspace Analysis on Manifolds. *Ann. Stat.* **2018**, *46*, 2711–2746. [CrossRef]
4. Hansen, P.; Eltzner, B.; Sommer, S. Diffusion Means and Heat Kernel on Manifolds. In *Geometric Science of Information*; Lecture Notes in Computer Science; Springer Nature: Cham, Switzerland, 2021; pp. 111–118. [CrossRef]
5. Hansen, P.; Eltzner, B.; Huckemann, S.F.; Sommer, S. Diffusion Means in Geometric Spaces. *arXiv* **2021**, arXiv:2105.12061.
6. Chakraborty, R.; Bouza, J.; Manton, J.H.; Vemuri, B.C. ManifoldNet: A Deep Neural Network for Manifold-Valued Data With Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 799–810. [CrossRef]
7. Sommer, S.; Bronstein, A. Horizontal Flows and Manifold Stochastics in Geometric Deep Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 811–822. [CrossRef]
8. Thompson, J. Submanifold Bridge Processes. Ph.D. Thesis, University of Warwick, Coventry, UK, 2015.
9. Pennec, X.; Sommer, S.; Fletcher, T. *Riemannian Geometric Statistics in Medical Image Analysis*; Elsevier: Amsterdam, The Netherlands, 2020.
10. Pennec, X.; Fillard, P.; Ayache, N. A Riemannian Framework for Tensor Computing. *Int. J. Comput. Vis.* **2006**, *66*, 41–66. [CrossRef]
11. Hsu, E.P. *Stochastic Analysis on Manifolds*; American Mathematical Society: Providence, RI, USA, 2002; Volume 38.
12. Grong, E.; Sommer, S. Most Probable Paths for Anisotropic Brownian Motions on Manifolds. *arXiv* **2021**, arXiv:2110.15634.
13. Sommer, S.; Arnaudon, A.; Kuhnel, L.; Joshi, S. Bridge Simulation and Metric Estimation on Landmark Manifolds. In *Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2017; pp. 79–91.
14. Papaspiliopoulos, O.; Roberts, G. Importance sampling techniques for estimation of diffusion models. *Stat. Methods Stoch. Differ. Equ.* **2012**, *124*, 311–340.
15. Jensen, M.H.; Sommer, S. Simulation of Conditioned Semimartingales on Riemannian Manifolds. *arXiv* **2021**, arXiv:2105.13190.
16. Lyons, T.J.; Zheng, W.A. On conditional diffusion processes. *Proc. R. Soc. Edinb. Sect. Math.* **1990**, *115*, 243–255. [CrossRef]
17. Delyon, B.; Hu, Y. Simulation of Conditioned Diffusion and Application to Parameter Estimation. *Stoch. Process. Appl.* **2006**, *116*, 1660–1675. [CrossRef]
18. Schauer, M.; Van Der Meulen, F.; Van Zanten, H. Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli* **2017**, *23*, 2917–2950. [CrossRef]
19. Marchand, J.L. Conditioning diffusions with respect to partial observations. *arXiv* **2011**, arXiv:1105.1608.
20. van der Meulen, F.; Schauer, M. Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.* **2017**, *11*, 2358–2396. [CrossRef]
21. Elworthy, D. Geometric aspects of diffusions on manifolds. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*; Springer: Berlin/Heidelberg, Germany, 1988; pp. 277–425.
22. Emery, M. *Stochastic Calculus in Manifolds*; Springer: Berlin/Heidelberg, Germany, 1989.
23. Elworthy, K.; Truman, A. The diffusion equation and classical mechanics: An elementary formula. In *Stochastic Processes in Quantum Theory and Statistical Physics*; Springer: Berlin/Heidelberg, Germany, 1982; pp. 136–146.
24. Li, X.M. On the semi-classical Brownian bridge measure. *Electron. Commun. Probab.* **2017**, *22*, 1–15. [CrossRef]
25. Ndumu, M.N. Brownian Motion and the Heat Kernel on Riemannian Manifolds. Ph.D. Thesis, University of Warwick, Coventry, UK, 1991.
26. Malliavin, P.; Stroock, D.W. Short time behavior of the heat kernel and its logarithmic derivatives. *J. Differ. Geom.* **1996**, *44*, 550–570. [CrossRef]
27. Stroock, D.W.; Turetsky, J. Short time behavior of logarithmic derivatives of the heat kernel. *Asian J. Math.* **1997**, *1*, 17–33. [CrossRef]
28. Thompson, J. Brownian bridges to submanifolds. *Potential Anal.* **2018**, *49*, 555–581. [CrossRef]
29. Joshi, S.; Miller, M. Landmark Matching via Large Deformation Diffeomorphisms. *IEEE Trans. Image Process.* **2000**, *9*, 1357–1370. [CrossRef]
30. Younes, L. *Shapes and Diffeomorphisms*; Springer: Berlin/Heidelberg, Germany, 2010.
31. Kühnel, L.; Sommer, S.; Arnaudon, A. Differential Geometry and Stochastic Dynamics with Deep Learning Numerics. *Appl. Math. Comput.* **2019**, *356*, 411–437. [CrossRef]
32. Stegmann, M.B.; Fisker, R.; Ersbøll, B.K. Extending and Applying Active Appearance Models for Automated, High Precision Segmentation in Different Image Modalities. Available online: http://www2.imm.dtu.dk/pubdb/edoc/imm118.pdf (accessed on 4 February 2022).

Chapter 5

Summary and Future Work

Throughout this dissertation, the aim has been to generalize guided bridge simulation schemes to various non-linear spaces and present numerical experiments showcasing statistical inference on these non-linear spaces. Chapter 1 introduced the background material needed throughout Chapters 2-4 to simulate guided bridge processes in various non-linear spaces.

Riemannian Bridges In chapter 2, we considered the case of guided processes on general Riemannian manifolds through stochastic development. Using similar techniques as in [13], we showed the absolute convergence between the law of the guided process and the target law on the interval $[0, T)$ and obtained an explicit expression for the correction factor φ. The expression was shown to depend on the geometric local time at the cut locus. From a numerical perspective, automatic differentiation approximates the correction factors, and thus the geometric local does not enter explicitly. From a mathematical perspective, the geometric local is more of a nuisance that needs to be handled. We showed that the guided bridge's law and the target bridge's law, and the target bridge's law were absolutely continuous for simply connected manifolds, e.g., spheres, ellipsoids, and the space of symmetric positive definite matrices. We showed that the guided bridge's law and the target bridge's law were absolutely continuous on $[0, T]$. This is followed by the fact of uniform integrability of the correction factor. It remains an open question to prove the absolute continuity over $[0, T]$ for more general spaces, e.g., compact manifolds. Investigating the conditions needed to ensure uniform integrability of φ is a natural extension. Several approaches seem natural to explore:

1. Using an Aronson's type estimate of the heat kernel on compact manifolds as in [26, Theorem 5.3.4].
2. Simulation of geometric local times
3. Propose a guided SDE in the embedding space.

Addressing the first approach, we note that Delyon & Hu [13] used the Aronson's estimate of the transition density of a Markov process to show the uniform integrability of φ. They then concluded by using the convergence rate of the guided process to its pinning point. The convergence rate of the guided processes in Chapter 2 are in general weaker than the one in [13]. However, in the case of a Brownian motion with drift b on a compact manifolds the convergence can be improved. On a compact manifold, we have the following bound on the Laplace-Beltrami operator; $\left(\frac{1}{2}\Delta_M + b\right) r_N^2 \leq \nu + \lambda r_N^2$. By the geometric Itô formula for the radial process

$$\begin{aligned}\frac{r_N^2(Y_t(y))}{T-t} =& \frac{r_N^2(y)}{T} + \int_0^t \frac{r_N^2(Y_s(y))}{(T-s)^2} ds + 2\int_0^t \frac{r_N(Y_s(y))}{T-s} d\beta_s \\ &+ 2\int_0^t \left(\frac{1}{2}\Delta_M + b\right) \frac{r_N^2(Y_s(y))}{T-s} ds \\ &- 2\int_0^t \frac{r_N^2(Y_s(y))}{(T-s)^2} ds - 2\int_0^t \frac{r_N(Y_s(y))}{T-s} d\mathbb{L}_t^C(Y(x)).\end{aligned}$$

As the radial process is bounded above by a constant $C > 0$ and below by zero on any compact manifold, we obtain the following bound

$$\frac{r_N^2(Y_t(y))}{T-t} \leq \frac{C}{T} + 2\int_0^t \frac{C}{T-s} d\beta_s + 2\int_0^t \frac{\nu + \lambda C}{T-s} ds,$$

if $\lambda > 0$, otherwise we set $\lambda = 0$. Since β is a one-dimensional Brownian motion, it follows by Dubins-Dambin-Schwartz and the law of the iterated logarithm of Brownian motion that

$$d(Y_t(y), N) = r_N(Y_t(y)) \leq K\sqrt{(T-t)\log\log\left((T-t)^{-1} + e\right)}.$$

One could then proceed as in Delyon and Hu [13]. However, numerical experiments, for example in the case of the flat torus $\mathbb{R}^2/\mathbb{Z}^2$, suggests that a constant needs to be added to the guiding term.

Any compact manifold contains a cut locus. Therefore, for the guided bridge to be an efficient method to simulate diffusion bridges on compact manifolds, one needs to simulate geometric local times. The simulation of geometric local times is still an open research problem. We note, however, that in specific cases, the geometric local time can be avoided. For example, the geometric local times vanish if the cut locus is polar for the guided process.

This thesis has not touched upon the possibility of embedding the manifold into a higher-dimensional Euclidean space. Defining guided bridges in the embedding space and projecting them onto the manifold could be an exciting approach. One of the potential difficulties with this approach is how to choose the embedding. There are different ways to embed a manifold and dimensionality if the embedding space increases dramatically when the dimension of the manifold increases.

Bridges in Lie Groups and Homogeneous Spaces

In Chapter 3, we utilized the simulation scheme developed in Chapter 2. We obtained a modified guided bridge process by exploiting the added structure given by the Lie group. In particular, since Lie groups are parallelizable, the stochastic development approach used in Chapter 2 became redundant. Simulations could be made using the left-invariant vector fields. For matrix Lie groups, the action of left-invariant vector fields is nothing but matrix multiplication. The matrix multiplication yielded a practical algorithm for bridge simulation in Lie groups. Conditioning the guided bridges on fibers over the base space resulted in a simulation scheme on the homogeneous spaces. As homogeneous spaces themselves are manifolds, the simulation scheme in Chapter 2 applies directly to these spaces. It could be interesting to investigate the numerical advantages/disadvantages of the two simulation schemes in Chapters 2 and 3.

Moreover, we obtained estimates of the underlying unknown metric structure through the bridge sampling scheme. One possible application of our simulation scheme arises in diffusion tensor imaging (DTI). Here, each voxel is a symmetric positive definite matrix. Through a bridge simulation scheme, we can estimate the underlying metric structure on DTI manifolds. This should be of interest within the field of diffusion tensor imaging.

Bridges in Product Manifolds

In the last chapter of the dissertation, we introduced a bridge sampling scheme on product manifolds. The idea was to replace n bridges on a manifold M conditioned on observations with a single bridge process in the product manifold conditioned on the diagonal. The resulting pinning point in the product manifold corresponds to the initial value of the n bridges in M. This results in the pinning point being the most probable point of the origin and hence provides a sample estimate of the weighted diffusion mean. Since the diffusion mean is a probabilistic alternative to the Fréchet mean, the bridge sampling scheme in the product manifold conditioned on the diagonal becomes an alternative to the Fréchet estimator. A numerical example of a data set with annotated landmarks illustrated the numerical efficiency of the proposed method compared to the running times of the Fréchet mean. It seems natural to investigate the convergence properties of the proposed method and if the method provides an consistent estimator for the diffusion mean.

Appendix A

Simulation of Conditioned Diffusions on the Flat Torus

The following manuscript has been made in collaboration with Anton Mallasto and Stefan Sommer and was published in the conference proceedings at the international conference of Geometric Science of Information (GSI).

> Mathias Højgaard Jensen, Anton Mallasto, Stefan Sommer. "*Simulation of Conditioned Diffusions on the Flat Torus*" (GSI).

The manuscript was built on the idea of Delyon and Hu [13] by introducing a manifold equivalent guiding term to that of Delyon and Hu on the flat torus. To the best of our knowledge, this paper is the first paper to introduce a guiding scheme on; a compact manifold, a homogeneous space, and its lift to the Lie group. In contrast to most literature on bridge simulation, this paper does not provide a closed-form expression for the likelihood.

Simulation of Conditioned Diffusions on the Flat Torus

Mathias Højgaard Jensen, Anton Mallasto, and Stefan Sommer

Department of Computer Science, Copenhagen University, Copenhagen, Denmark
{matje,mallasto,sommer}@di.ku.dk

Abstract. Diffusion processes are fundamental in modelling stochastic dynamics in natural sciences. Recently, simulating such processes on complicated geometries has found applications for example in biology, where toroidal data arises naturally when studying the backbone of protein sequences, creating a demand for efficient sampling methods. In this paper, we propose a method for simulating diffusions on the flat torus, conditioned on hitting a terminal point after a fixed time, by considering a diffusion process in $\mathbb{R}^2$ which we project onto the torus. We contribute a convergence result for this diffusion process, translating into convergence of the projected process to the terminal point on the torus. We also show that under a suitable change of measure, the Euclidean diffusion is locally a Brownian motion.

Keywords: Simulation · Conditioned diffusion · Manifold diffusion · Flat Torus

1 Introduction

Stochastic differential equations are ubiquitous in models describing evolution of dynamical systems with, e.g. in modelling the evolution of DNA or protein structure, in pricing financial derivatives, or for modelling changes in landmark configurations which are essential in shape analysis and computational anatomy. In settings where the beginning and end values are known on some fixed time interval, the use of Brownian bridges becomes natural to evaluate the uncertainty on the intermediate time interval.

When the data elements are elements of non-linear spaces, here differentiable manifolds, methodology for simulating bridge processes is lacking. In particular, in cases where the transition probability densities are intractable, it is of interest to use simulation schemes that can numerically approximate the true densities. In this paper we propose a method for simulating diffusion bridges on the flat torus, $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, i.e. we propose a process that can easily be simulated and satisfies that the distribution of the true bridge of interest is absolutely continuous with respect to the distribution of this proposal process. This specific

MHJ, AM, and SS are supported by the CSGB Centre for Stochastic Geometry and Advanced Bioimaging funded by a grant from the Villum Foundation.

F. Nielsen and F. Barbaresco (Eds.): GSI 2019, LNCS 11712, pp. 685–694, 2019.
https://doi.org/10.1007/978-3-030-26980-7_71

case will serve as an example of the more general setting of simulating diffusion bridge processes on Riemannian manifolds. Because of the non-trivial topology of the torus $\mathbb{T}^2$, the conditioned process will be equivalent to a process in $\mathbb{R}^2$ that is conditioned on ending up in a set of points. Therefore, we will address the question of conditioning a process on infinitely many points. Secondly, we will handle the case when the process crosses the cut locus of the target point, i.e. the set of points with no unique distance minimizing geodesic.

It is a basic consequence of Doob's h-transform that the distribution of a conditioned diffusion process is the same as another diffusion process with the drift depending on the transition density. However, as mentioned in [1], using this transform directly is undesirable for simulation purposes as the transition density is often intractable. Instead, the authors introduce a diffusion process which can easily be simulated and with the property that the distribution of the true conditioned diffusion is absolutely continuous wrt. the diffusion used for simulation. We here use this approach that in [1] covers the Euclidean case as the starting point for developing a simulation scheme on the torus.

Recent papers have considered diffusion processes on the torus, for example, Langevin diffusions on the torus were studied in [3] and [4], in the latter to describe protein evolution. In this paper, we introduce a diffusion process in $\mathbb{R}^2$ which can easily be simulated and projects onto a bridge process on the torus. More generally, Brownian bridges on manifolds have been studied for example in the context of landmark manifolds [9] and used for approximating the transition density of the Brownian motion. The present paper uses bridges on the flat torus to exemplify how some of the challenges of bridge simulation on Riemannian manifolds can be addressed, here in particular non-trivial topology of the manifold.

We begin in Sect. 2 with a short introduction to Brownian bridge processes in the standard Eucliden case and how it relates to the definition of a Brownian bridge process on the flat torus. At the end we introduce the stochastic differential equation (SDE) which will be used for simulating the bridge process. In Sect. 3 we argue that a strong solution of our proposed SDE exist. We show results about convergence and absolute continuity in Sect. 4. Numerical examples are presented in Sect. 5.

2 Theoretical Setup

This section will briefly review some Brownian bridges theory and discuss the torus case. A more general theory of diffusion bridges can be found in [1], constituting the main reference for this work. At the end, we introduce our proposal process.

Consider a Brownian motion $W = (W_t)_{t\geq 0}$ in $\mathbb{R}^n$. By conditioning, it can be shown that W will end up at a given point at a given time. For example, the process given by $B_t = W_t - \frac{t}{T}W_T$ defines a Brownian bridge conditioned to return to 0 at time T. It can be shown that the diffusion process given by

$$dX_t = \frac{b - X_t}{T - t}dt + dW_t; \quad 0 \leq t < T \quad \text{and} \quad X_0 = a, \tag{1}$$

for given $a, b \in \mathbb{R}^d$ and W a d-dimensional standard Brownian motion, is a d-dimensional Brownian bridge from a to b on $[0, T]$ (see e.g. [6, sec. 5.6]). More generally, diffusion bridges can be defined through Doob's h-transform, that is, the distribution of a diffusion

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \qquad X_0 = a,$$

conditioned on $X_T = b$ is the same as that of

$$\begin{aligned} dY_t &= \tilde{b}(t, Y_t)dt + \sigma(t, Y_t)dW_t, \\ \tilde{b}(t, x) &= b(t, x) + \sigma(t, x)\sigma^T(t, x)\nabla_x \log(p(t, x; T, b)), \end{aligned}$$

where $p(t, x; T, b)$ denotes the transition density of the process X. In the usual setting where p is the transition density of a Brownian motion it has the form

$$p(s, x; t, y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(- \frac{||x-y||^2}{2(t-s)} \right), \quad s < t,$$

which yields (1).

We propose a method similar to the Euclidean scheme [1] for simulating Brownian bridges on the flat torus, which is of the form

$$dX_t = b(t, X_t)dt + \sigma dW_t; \quad 0 \le t < T \quad \text{and} \quad X_0 = a \quad \text{a.s.}, \tag{2}$$

where $\sigma > 0$, $a \in \mathbb{T}^2$ is given, and W is a two-dimensional standard Brownian motion. The exact form of $b(t, x)$ will become apparent below. It is important here to note that in the particular case of the flat torus the transition density for the Brownian motion is known and therefore it is possible to simulate from the distribution of the true Brownian bridge on $\mathbb{T}^2$, however, it requires the calculation of the distance to infinitely many points which the proposed model does not. In Fig. 4 is shown paths of the proposed model and the corresponding paths of the true bridge process.

Let $\pi\colon \mathbb{R}^2 \to \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ denote the canonical projection onto the torus. The standard two-dimensional Brownian motion $W = (W^1, W^2)$, for two independent one-dimensional Brownian motions W^1 and W^2, is mapped to a Brownian motion $B = (B_t)_{t\ge 0}$ on the flat torus $\mathbb{T}^2$ by the projection map π. Indeed, we can identify the torus $\mathbb{T}^2$ with the unit cube $Q = \{x \in \mathbb{R}^2 : -\frac{1}{2} \le x_k < \frac{1}{2}, k = 1, 2\}$. Then for $g \in C^\infty(\mathbb{T}^2)$ the Laplace-Beltrami operator, $\Delta_{\mathbb{T}^2}$, on $\mathbb{T}^2$ corresponds to the restriction to Q of the usual Euclidean Laplacian, $\Delta_{\mathbb{R}^2}\tilde{g}$, where $\tilde{g}$ denotes the periodic extension of g, i.e. $\tilde{g} = g \circ \pi$ (see [8, Sec. 3.5]). Since W is a Brownian motion in $\mathbb{R}^2$ if and only if it satisfies the diffusion equation

$$h(W_t) \overset{m}{=} h(W_0) - \frac{1}{2}\int_0^t \Delta_{\mathbb{R}^2} h(W_s)ds,$$

for all smooth functions h, where $X \overset{m}{=} Y$ means that the difference $X - Y$ is a local martingale (see e.g. [2, Sec. 1.5]), it follows that, for $h = \tilde{g}$,

$$\tilde{g}(W_t) \overset{m}{=} \tilde{g}(W_0) - \frac{1}{2}\int_0^t \Delta_{\mathbb{R}^2}\tilde{g}(W_s)ds = g(B_0) - \frac{1}{2}\int_0^t \Delta_{\mathbb{T}^2} g(B_s)ds \overset{m}{=} g(B_t).$$

As this holds for all smooth functions g on $\mathbb{T}^2$, we get that B is a Brownian motion on $\mathbb{T}^2$ in agreement with the definition of a manifold-valued Brownian motion given in [5, Sec. 3.2].

By conditioning B on $\mathbb{T}^2$ to hit a given point $a \in \mathbb{T}^2$, at some fixed time $0 \leq T < \infty$, it is seen that

$$\{\omega \in \Omega : B_T(\omega) = a\} = \{\omega \in \Omega : W_T(\omega) \in \pi^{-1}(a)\},$$

and so simulating a Brownian bridge on the flat torus $\mathbb{T}^2$ is equivalent to simulating a two-dimensional standard Brownian motion conditioned to end up in the set $\pi^{-1}(a)$ at time T. The diffusion given by (1) will not suffice as it is constructed to hit exactly one point. It will, however, provide one subset of sample paths of the Brownian bridge on $\mathbb{T}^2$, corresponding to subset of paths that will "unwrap" the same number of times that it "wraps" around the cut locus. This is illustrated in Fig. 1. To give a precise meaning to this statement we consider the h-transform

$$h(t, z) = \sum_{y \in \pi^{-1}(a)} \frac{p(t, z; T, y)}{p(0, z_0; T, y)},$$

with p denoting the transition density of the two-dimensional Brownian motion, which by Doob's h-transform implies that the distribution of W conditioned on $W_T \in \pi^{-1}(a)$ is the same as the distribution of the diffusion

$$\begin{aligned} dZ_t &= \sigma^2 \nabla_z \log \left(\sum_{y \in \pi^{-1}(a)} p(t, z; T, y) \right) \Bigg|_{x = Z_t} dt + \sigma dW_t \\ &= \sum_{y \in \pi^{-1}(a)} g_y(t, Z_t) \frac{y - Z_t}{T - t} dt + \sigma dW_t, \quad Z_0 = z_0, \end{aligned} \tag{3}$$

where $g_y(t, x) = \frac{\exp\left(-\frac{\|y-z\|^2}{2\sigma^2(T-t)}\right)}{\sum_{y \in \pi^{-1}(a)} \exp\left(-\frac{\|y-z\|^2}{2\sigma^2(T-t)}\right)}$.

Instead, we propose to consider the diffusion process on $[0, T)$, for some fixed positive T, defined by

$$dX_t = 1_{G^c}(X_t) \frac{\alpha(X_t) - X_t}{T - t} dt + \sigma dW_t, \quad X_0 = x_0 \tag{4}$$

where $\sigma > 0$ and α is defined by

$$\alpha(X_t) = \underset{y \in \pi^{-1}(a)}{\arg\min} \|y - X_t\|,$$

with $a \in \mathbb{T}^2$, and where G is the set of "straigt lines" of the form $\mathbb{R} \times \{x\}$ (resp. $\{x\} \times \mathbb{R}$) in $\mathbb{R}^2$ where $\alpha(X_t)$ is not unique (see Fig. 1). The indicator function removes the drift when the process does not have a natural attraction point.

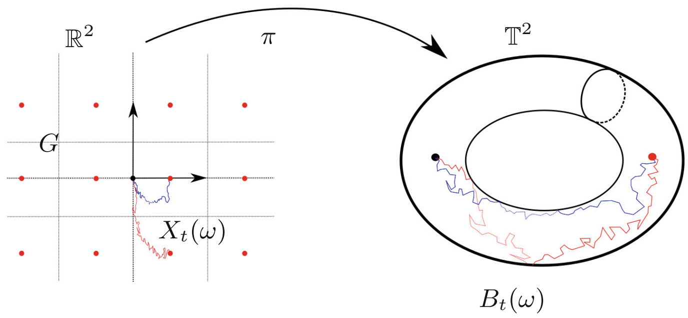


Fig. 1. The figure illustrates the possibility of the diffusion path going an arbitrary number of times around the torus, starting at the black dot and ending in the red. This is illustrated by the red path. The conditioning on single point in $\mathbb{T}^2$ therefore leads to conditioning on multiple points in $\mathbb{R}^2$. Left: Two paths from the same two-dimensional process with multiple endpoints. Right: The projection of the two paths onto the torus. (Color figure online)

3 Existence of Strong Solution

The drift term in Eq. (4) is discontinuous. However, we below show that it posses certain regularity conditions and use this to show that a strong solution to the SDE exist.

In order to ensure the existence of a solution to the diffusion in (4), we need some regularity of the drift term. The drift coefficient is given by

$$1_{G^c}(X_t)\frac{\alpha(X_t)-X_t}{T-t} = \begin{cases} \frac{\alpha(X_t)-X_t}{T-t}, & \text{if } X_t \in G^c \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

for every $0 \le t < T$, where the superscript c denotes the complement. It is a discontinuous process with the set of discontinuities being the set G consisting of the set of straight lines in $\mathbb{R}^2$ where the argmin process is non-unique. It is not even clear that the drift term is suitably measurable as the argmin map in general is not.

Lemma 1. *Let* $b\colon [0,T)\times\mathbb{R}^2 \to \mathbb{R}^2$ *be the map given by* (5)*. Then* b *is* $\mathcal{B}([0,T))\otimes\mathcal{B}(\mathbb{R}^2) - \mathcal{B}(\mathbb{R}^2)$ *measurable. Furthermore, the map* $(s,\omega) \mapsto b(s, X_s(\omega))$ *is* $\mathcal{B}([0,t])\otimes\mathcal{F}_t^0$ *measurable, for every* $0 \le t < T$*, where* $(\mathcal{F}_t^0)$ *denotes the natural filtration generated by* X*. This is called progressive measurability.*

Proof. First note that G^c is a Borel measurable set as we can write it as a countable union of open sets, i.e., for $y = (y_1, y_2)$ we have

$$G^c = \bigcup_{y\in\pi^{-1}(a)} \left(y_1 - \frac{1}{2}, y_1 + \frac{1}{2}\right) \times \left(y_2 - \frac{1}{2}, y_2 + \frac{1}{2}\right) =: \bigcup_{y\in\pi^{-1}(a)} V_y.$$

Now, we need to show that for all $A \in \mathcal{B}(\mathbb{R}^2)$, the set $b^{-1}(A)$ is an element of $\mathcal{B}([0,T))\otimes\mathcal{B}(\mathbb{R}^2)$. It is enough to consider all open subsets $U \subseteq \mathbb{R}^2$ as these sets generate the Borel algebra on $\mathbb{R}^2$. So let U be an arbitrary open subset, then we have that

$$b^{-1}(U) = b^{-1}(U) \cap \big([0,T) \times G^c\big) \cup b^{-1}(U) \cap \big([0,T) \times G\big).$$

As b is continuous on each of the sets $[0,T) \times V_y$ we have that $b^{-1}(U) \cap \big([0,T) \times G^c\big)$ is a countable union of open sets and therefore an element of $\mathcal{B}([0,T)) \otimes \mathcal{B}\big(\mathbb{R}^2\big)$. For the second part we see that

$$b^{-1}(U) \cap ([0,T) \times G) = \begin{cases} [0,T) \times G, & \text{if } (0,0) \in U \\ \emptyset, & \text{otherwise,} \end{cases}$$

where both are elements of $\mathcal{B}([0,T)) \otimes \mathcal{B}\big(\mathbb{R}^2\big)$. This shows that b is Borel measurable.

Progressive measurability follows by a very similar argument. □

Usually, global or local Lipschitz conditions are imposed on the drift and diffusion coefficients in order to secure global (resp. local) strong solutions to an SDE. This is a too strong condition for the drift term in this case, however, it is bounded in the following sense.

Lemma 2. *The drift coefficient in* (5) *is uniformly bounded in x and in t on $[0,S]$, for any $0 \le S < T$.*

Proof. The first assertion is clear. Let $S \in [0,T)$ be arbitrary and $0 \le t \le S$. For every $x \in G^c$ there exist a $y \in \pi^{-1}(a)$ such that we have

$$\left\| 1_{G^c}(x) \frac{\alpha(x) - x}{T-t} \right\|^2 = \left\| \frac{y-x}{T-t} \right\|^2 \le \frac{C}{(T-S)^2} = C_S,$$

for some positive constants $C > 0$. □

We now come to the main result of this section.

Proposition 1. *There exist a strong solution of* (4) *on* $[0,T)$*, which is strongly unique.*

Proof. The drift term is Borel measurable and bounded on $[0,S]$ by Lemmas 1 and 2. As indicated in [10, Thm. 2] and [11, Thm. 1] (4) has a strong solution which is strongly unique. □

Remark 1. The assumption in [10, Thm. 2] can be verified by using smooth bump functions.

4 Convergence and Absolute Continuity

The considerations above make the solution of (4) into a continuous semimartingale. If a semimartingale X takes its values in an open set U of $\mathbb{R}^2$ then Itô's formula holds true for any $C^{1,2}([0,T) \times U)$ functions as well.

Proposition 2. *Let X be a solution to (4) on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$. For every $\omega \in \Omega$ for which there exist an $S < T$ such that $X_t(\omega)$ stays in G^c on $[S,T)$, then X converges pointwise almost surely to $\pi^{-1}(a)$.*

Proof. Assume that for some $\omega \in \Omega$ there exist some $S < T$ such that on $[S,T)$ the process $X_t(\omega)$ takes its values in G^c. By continuity of the process it will take it its values in some open neighborhood V_y of the point $y \in \pi^{-1}$. The proof is then identical to the proof in [1, Lemma 4]. □

Remark 2. It is of course of interest to show that for almost every path the process will converge. This can be obtained by showing that the process will not intersect G infinitely many times close to T.

Consider the stochastic process $\mathcal{E}$ on $0 \leq t \leq S$ defined by

$$\mathcal{E}(L)_t = \exp\left(- \int_0^t b(s, X_s) dW_s - \frac{1}{2} \int_0^t \|b(s, X_s)\|^2 \, ds \right), \tag{6}$$

where L is the local martingale in the exponential. This is known as the Doléans-Dade exponential. From Lemma 2 it follows that, for all $t \leq S$,

$$\mathbb{E}\left[\exp\left(\int_0^t \|b(s, X_s)\|^2 \, ds \right) \right] \leq \exp\left(tC_S \right) < \infty$$

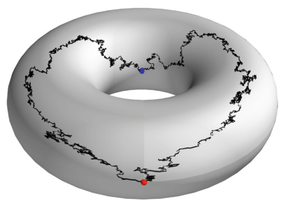
The above is known as the Novikov condition (cf. [7]) which ensures that (6) is a martingale on $[0,T)$. Girsanov's theorem ([6, Thm. 5.1 Chap. 3]) then provides that the process defined by

$$\widetilde{W}_t = W_t + \int_0^t b(s, X_s) ds$$

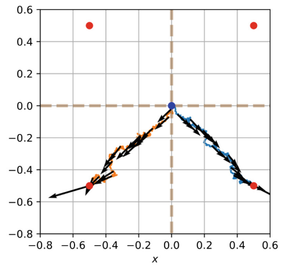
is a Brownian motion under the new measure Q introduced below.

Theorem 1. *Let X defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ be a solution of (4) on $[0,S]$ for $S < T$. The process in (6) defined on $0 \leq t \leq S$ ($S < T$) is a true martingale and so there exists a measure Q which is absolutely continuous wrt. P such that X is Q-Brownian motion.*

Proof. The martingale property of (6) on $[0,S]$ is a consequence of the Novikov condition. Then Girsanov's theorem gives us that X is a Q-Brownian motion on $[0,S]$. □



(a) Paths visualized on an embedded torus.



(b) The two Euclidean paths that are mapped onto the torus.

Fig. 2. Two different paths visualized both on the torus and in Eucliden space. The blue dot represents the starting point and the red represents the end point. (Color figure online)

From the (perhaps obvious) fact that the distribution of the true Brownian bridge is locally equivalent to the distribution of the Brownian motion up to time $t < T$, it follows that the distribution of the Brownian bridge is absolutely continuous wrt. The proposed process up to time $t < T$.

Remark 3. A bit of extra work is needed to obtain the correction term as in [1]. There are indications that it is possible to simulate from the true distribution of the Brownian bridge on the torus, however, Theorem 1 shows that (4) can approximate it.

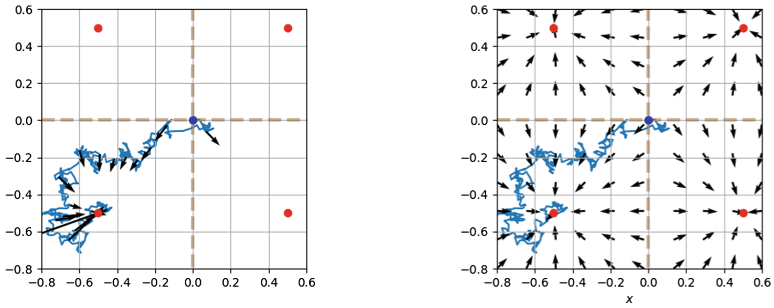
5 Numerical Experiments

For the numerical implementation of the proposed SDE in Eq. (4) we implemented the Euler-Maruyama scheme, i.e. taking n equidistant discretization points of the time interval $t_1, ..., t_n$, with $t_{i+1} - t_i = \varDelta t$, the numerical equation becomes

$$x_{t_{i+1}} = x_{t_i} + \frac{\arg\min_{y \in \pi^{-1}(a)}(\|y - x_{t_i}\|) - x_{t_i}}{T - t_i} \varDelta t + \sigma \varDelta W_{t_i},$$

where $\varDelta W_{t_{i+1}} = W_{t_{i+1}} - W_{t_i}$ is equal in distribution to a normal random variable with mean zero and variance $\varDelta t$.

Figure 2a shows the implementation of the numerical scheme on an embedded torus and Fig. 2b its Euclidean counterpart. Figure 3a shows the behaviour of the drift term along a given path, illustrating that the attraction becomes stronger as time approaches the terminal time. The vector fields in Fig. 3b shows the constant attraction to the center of the open subsets.



(a) Drift term (b) Vector field

Fig. 3. Figure a depicts the evolution of the drift term. It shows how the pull from the drift becomes stronger near the end. Figure b shows the underlying vector field.

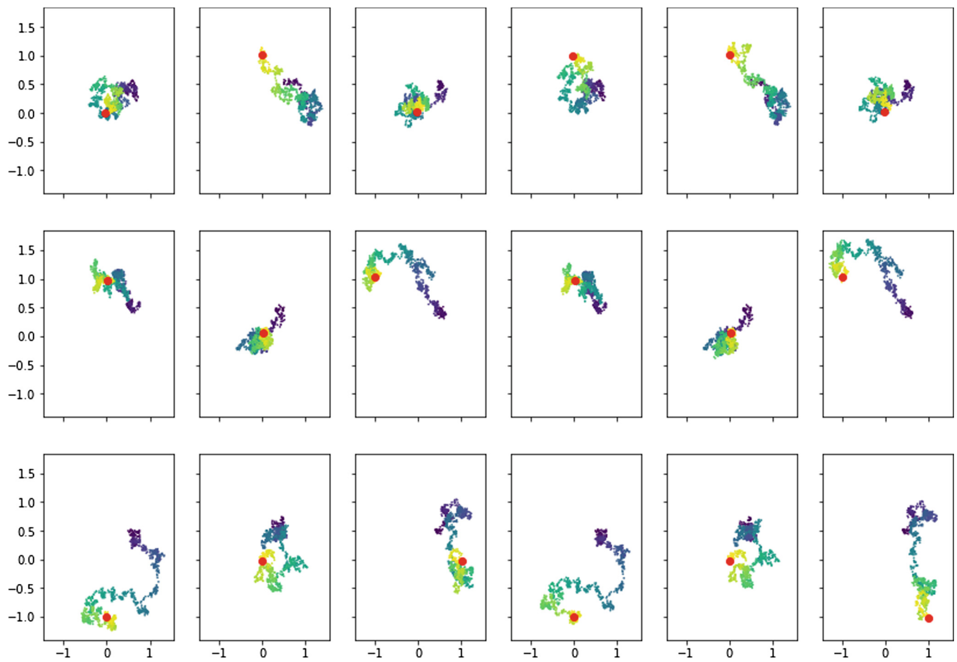


Fig. 4. Figure shows 9 paths from the proposed model (4) on the left and the corresponding paths from the true bridge (3) on the right. It is seen that the first and the last paths disagree on the limiting point, whereas the rest looks fairly similar. The picture agree with the fact that roughly four in five have the same limiting point. Here $\sigma = 0.8$ and the conditioning points being the integers in $[-2, 2] \times [-2, 2]$.

Acknowledgements. We acknowledge F. van der Meulen for discussions and insights on conditioned diffusions.

References

1. Delyon, B., Hu, Y.: Simulation of conditioned diffusion and application to parameter estimation. Stochast. Processes Appl. **116**(11), 1660–1675 (2006)
2. Emery, M.: Stochastic Calculus in Manifolds. Springer, Heidelberg (1989). https://doi.org/10.1007/978-3-642-75051-9
3. García-Portugués, E., Sørensen, M., Mardia, K.V., Hamelryck, T.: Langevin diffusions on the torus: estimation and applications. Stat. Comput. **29**, 1–22 (2017)
4. Golden, M., García-Portugués, E., Sørensen, M., Mardia, K.V., Hamelryck, T., Hein, J.: A generative angular model of protein structure evolution. Mol. Biol. Evol. **34**(8), 2085–2100 (2017)
5. Hsu, E.P.: Stochastic Analysis on Manifolds, vol. 38. American Mathematical Soc., Providence (2002)
6. Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus, 2nd edn. Springer, New York (1991)
7. Novikov, A.A.: On an identity for stochastic integrals. Theory Probab. Appl. **17**(4), 717–720 (1973)
8. Sogge, C.D.: Hangzhou Lectures on Eigenfunctions of the Laplacian (AM-188). Princeton University Press, Princeton (2014)
9. Sommer, S., Arnaudon, A., Kuhnel, L., Joshi, S.: Bridge simulation and metric estimation on landmark manifolds. In: Cardoso, M.J., et al. (eds.) GRAIL/MFCA/MICGen -2017. LNCS, vol. 10551, pp. 79–91. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67675-3_8
10. Veretennikov, A.Y.: On the strong solutions of stochastic differential equations. Theory Probab. Appl. **24**(2), 354–366 (1980)
11. Veretennikov, A.J.: On strong solutions and explicit formulas for solutions of stochastic integral equations. Sbornik: Mathematics **39**, 387–403 (1981)

Appendix B

Bridge Simulation and Metric Estimation on Lie Groups

The following manuscript has been made in collaboration with Sarang Joshi and Stefan Sommer and was published in the conference proceedings at the international conference of Geometric Science of Information (GSI).

> Mathias Højgaard Jensen, Sarang Joshi, Stefan Sommer. "*Bridge Simulation and Metric Estimation on Lie Groups*" (GSI).

The paper introduces a simulation scheme on the space of Lie groups. We exploit the simulation scheme using Monte-Carlo methods to obtain approximations of the transition density. An iterative maximum likelihood approach yields an estimation of the underlying unknown metric. The paper is an extension of Sommer et al. [45]. The content of the current paper can be found in the manuscript presented in Chapter 3.

Check for updates

Bridge Simulation and Metric Estimation on Lie Groups

Mathias Højgaard Jensen[1], Sarang Joshi[2], and Stefan Sommer[1]

[1] Department of Computer Science, University of Copenhagen, Universitetsparken 1, 2100 Copenhagen, Denmark
{matje,sommer}@di.ku.dk

[2] Department of Biomedical Engineering, University of Utah, 72 S Central Campus Drive, Salt Lake City, UT 84112, USA
sjoshi@sci.utah.edu

Abstract. We present a simulation scheme for simulating Brownian bridges on complete and connected Lie groups. We show how this simulation scheme leads to absolute continuity of the Brownian bridge measure with respect to the guided process measure. This result generalizes the Euclidean result of Delyon and Hu to Lie groups. We present numerical results of the guided process in the Lie group SO(3). In particular, we apply importance sampling to estimate the metric on SO(3) using an iterative maximum likelihood method.

Keywords: Brownian motion · Brownian bridge simulation · Importance sampling · Lie groups · Metric estimation

1 Introduction

Bridge simulation techniques are known to play a fundamental role in statistical inference for diffusion processes. Diffusion bridges in manifolds have mainly been used to provide gradient and hessian estimates. To the best of our knowledge, this paper is the first to describe a simulation technique for diffusion bridges in the context of Lie groups.

The paper is organized as follows. In Sect. 2, we describe some background theory of Lie groups, Brownian motions, and Brownian bridges in Riemannian manifolds. Section 3 presents the theory and results. Section 4 shows in practice the simulation scheme in the Lie group SO(3). Using importance sampling, we obtain an estimate of the underlying unknown metric.

2 Notation and Background

Lie Groups. Throughout, we let G denote a connected Lie Group of dimension d, i.e., a smooth manifold with a group structure such that the group operations $G \times G \ni (x, y) \overset{\mu}{\mapsto} xy \in G$ and $G \ni x \overset{\iota}{\mapsto} x^{-1} \in G$ are smooth maps. If $x \in G$, the

F. Nielsen and F. Barbaresco (Eds.): GSI 2021, LNCS 12829, pp. 430–438, 2021.
https://doi.org/10.1007/978-3-030-80209-7_47

left-multiplication map, $L_x y$, defined by $y \mapsto \mu(x, y)$, is a diffeomorphism from G to itself. Similarly, the right-multiplication map $R_x y$ defines a diffeomorphism from G to itself by $y \mapsto \mu(y, x)$. We assume throughout that G acts on itself by left-multiplication. Let $dL_x \colon TG \to TG$ denote the pushforward map given by $(dL_x)_y \colon T_y G \to T_{xy} G$. A vector field V on G is said to be left-invariant if $(dL_x)_y V(y) = V(xy)$. The space of left-invariant vector fields is linearly isomorphic to $T_e G$, the tangent space at the identity element $e \in G$. By equipping the tangent space $T_e G$ with the Lie bracket we can identify the Lie algebra $\mathcal{G}$ with $T_e G$. The group structure of G makes it possible to define an action of G on its Lie algebra $\mathcal{G}$. The conjugation map $C_x := L_x \circ R_x^{-1} \colon y \mapsto xyx^{-1}$, for $x \in G$, fixes the identity e. Its pushforward map at e, $(dC_x)_e$, is then a linear automorphism of $\mathcal{G}$. Define $\mathrm{Ad}(x) := (dC_x)_e$, then $\mathrm{Ad} \colon x \mapsto \mathrm{Ad}(x)$ is the adjoint representation of G in $\mathcal{G}$. The map $G \times \mathcal{G} \ni (x, v) \mapsto \mathrm{Ad}(x)v \in \mathcal{G}$ is the adjoint action of G on $\mathcal{G}$. We denote by $\langle \cdot, \cdot \rangle$ a Riemannian metric on G. The metric is said to be left-invariant if $\langle u, v \rangle_y = \langle (dL_x)_y u, (dL_x)_y v \rangle_{L_x(y)}$, for every $u, v \in T_y G$, i.e., the left-multiplication maps are isometries, for every $x \in G$. In particular, we say that the metric is $\mathrm{Ad}(G)$-invariant if $\langle u, v \rangle_e = \langle \mathrm{Ad}(x)u, \mathrm{Ad}(x)v \rangle_e$, for every $u, v \in \mathcal{G}$. Note that an $\mathrm{Ad}(G)$-invariant inner on $\mathcal{G}$ induces a bi-invariant (left- and right-invariant) metric on G.

Brownian Motion. Endowing a smooth manifold M with a Riemannian metric, g, allows us to define the Laplace-Beltrami operator, $\Delta_M f = \operatorname{div} \operatorname{grad} f$. This operator is the generalization of the Euclidean Laplacian operator to manifolds. In terms of local coordinates $(x_1, \ldots, x_d)$ the expression for the Laplace-Beltrami operator becomes $\Delta_M f = \det(g)^{-1/2} \left(\frac{\partial}{\partial x_j} g^{ji} \det(g)^{1/2} \frac{\partial}{\partial x_i} \right) f$, where $\det(g)$ denotes the determinant of the Riemannian metric g and g^{ij} are the coefficients of the inverse of g. An application of the product rule implies that Δ_M can be rewritten as $\Delta_M f = a^{ij} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f + b^j \frac{\partial}{\partial x_j} f$, where $a^{ij} = g^{ij}$, $b^k = -g^{ij} \Gamma^k_{ij}$, and Γ denote the Christoffel symbols related to the Riemannian metric. This diffusion operator defines a Brownian motion on the M, valid up to its first exit time of the local coordinate chart.

In the case of the Lie group G, the identification of the space of left-invariant vector fields with the Lie algebra $\mathcal{G}$ allows for a global description of Δ_G. Indeed, let $\{v_1, \ldots v_d\}$ be an orthonormal basis of $T_e G$. Then $V_i(x) = (dL_x)_e v_i$ defines left-invariant vector fields on G and the Laplace-Beltrami operator can be written as (cf. [6, Proposition 2.5]) $\Delta_G f(x) = \sum_{i=1}^d V_i^2 f(x) - V_0 f(x)$, where $V_0 = \sum_{i,j=1}^d C^j_{ij} V_j$ and C^k_{ij} denote the structure coefficients given by $[V_i, V_j] = C^k_{ij} V_k$. The corresponding stochastic differential equation (SDE) for the Brownian motion on G, in terms of left-invariant vector fields, then becomes

$$dX_t = -\frac{1}{2} V_0(X_t) dt + V_i(X_t) \circ dB^i_t, \qquad X_0 = e, \tag{1}$$

where $\circ$ denotes integration in the Stratonovich sense. By [6, Proposition 2.6], if the inner product is $\mathrm{Ad}(G)$ invariant, then $V_0 = 0$. The solution of (1) is conservative or non-explosive and is called the left-Brownian motion on G (see [8] and references therein).

Riemannian Brownian Bridges. In this section, we briefly review some classical facts on Brownian bridges on Riemannian manifolds. As Lie groups themselves are manifolds, the theory carries over *mutatis mutandis*. However, Lie groups' group structure allows the notion of left-invariant vector fields. The identification of the Lie algebra with the vector space of left-invariant vector fields makes Lie groups parallelizable. Thus, the frame bundle construction for developing stochastic processes on manifolds becomes superfluous since left-invariant vector fields ensure stochastic parallel displacement.

Let $\mathbb{P}^t_x$ be the measure of a Riemannian Brownian motion, X_t, at some time t started at point x. Suppose p denotes the transition density of the Riemannian Brownian motion. In that case, $d\mathbb{P}^t_x = p(t, x, y)d\,\mathrm{Vol}(y)$ describes the measure of the Riemannian Brownian motion, where $d\,\mathrm{Vol}(y)$ is the Riemannian volume measure. Conditioning the Riemannian Brownian motion to hit some point v at time $T > 0$ results in a Riemannian Brownian bridge. Here, $\mathbb{P}^T_{x,v}$ denotes the corresponding probability measure. The two measures are equivalent over the time interval $[0, T)$, however mutually singular at time $t = T$. The initial enlargement of the filtration remedies the singularity. The corresponding Radon-Nikodym derivative is given by

$$\frac{d\mathbb{P}^T_{x,v}}{d\mathbb{P}^T_x}\Big|_{\mathcal{F}_s} = \frac{p(T-s, X_s, v)}{p(T, x, v)} \qquad \text{for } 0 \leq s < T,$$

which is a martingale for $s < T$. The Radon-Nikodym derivative defines the density for the change of measure and provides the basis for the description of Brownian bridges. In particular, it provides the conditional expectation defined by

$$\mathbb{E}[F(X_t)|X_T = v] = \frac{\mathbb{E}[p(T-t, X_t, v)F(X_t)]}{p(T, x, v)},$$

for any bounded and $\mathcal{F}_s$-measurable random variable $F(X_s)$. As described in [3], the Brownian bridge yields an SDE in the frame bundle, $\mathcal{FM}$, given by

$$dU_t = H_i(U_t) \circ \left(dB^i_t + H_i \log \tilde{p}(T-t, U_t, v)dt\right), \qquad U_0 = u_0, \tag{2}$$

in terms of the horizontal vector fields (H_i), which is the lifted M-valued Brownian bridge, $X_t := \pi(U_t)$, where $\pi\colon \mathcal{FM} \to M$.

3 Simulation of Bridges on Lie Groups

In this section, we consider the task of simulating (1) conditioned to hit $v \in G$, at time $T > 0$. The potentially intractable transition density for the solution of (1) inhibits simulation directly from (2). Instead, we propose to add a guiding term mimicking that of Delyon and Hu [2], i.e., the guiding term becomes the gradient of the distance to v divided by the time to arrival. The SDE for the guided diffusion becomes

$$dY_t = -\frac{1}{2}V_0(Y_t)dt + V_i(Y_t) \circ \left(dB_t^i - \frac{\left(\nabla_y d(Y_t, v)^2\right)^i}{2(T-t)}dt \right), \qquad Y_0 = e, \qquad (3)$$

where d denotes the Riemannian distance function. Note that we can always, for convenience, take the initial value to be the identity e.

Radial Process. We denote by $r_v(\cdot) := d(\cdot, v)$ the radial process. Due to the radial process's singularities on $\mathrm{Cut}(v) \cup \{v\}$, the usual Itô's formula only applies on subsets away from the cut-locus. The extension beyond the cut-locus of a Brownian motion's radial process was due to Kendall [4]. Barden and Le [1,5] generalized the result to M-semimartingales. The radial process of the Brownian motion (1) is given by

$$r(X_t) = r(X_0)^2 + \int_0^t \langle \nabla r(X_s), V(X_s)dB_s \rangle + \frac{1}{2}\int_0^t \Delta_G r(X_s)ds - L_s(X), \quad (4)$$

where L is the geometric local time of the cut-locus $\mathrm{Cut}(v)$, which is non-decreasing continuous random functional increasing only when X is in $\mathrm{Cut}(v)$ (see [1,4,5]). Let $W_t := \int_0^t \left\langle \frac{\partial}{\partial r}, V_i(X_s) \right\rangle dB_s^i$, which is the local-martingale part in the above equation. The quadratic variation of W_t satisfies $d[W,W]_t = dt$, by the orthonormality of $\{V_1, \ldots, V_d\}$, thus W_t is a Brownian motion by Levy's characterization theorem. From the stochastic integration by parts formula and (4) the squared radial process of X satisfies

$$r(X_t)^2 = r(X_0)^2 + 2\int_0^t r(X_s)dW_s + \int_0^t r(X_s)\Delta_G r(X_s)ds - 2\int_0^t r(X_s)dL_s, \quad (5)$$

where dL_s is the random measure associated to $L_s(X)$.

Similarly, we obtain an expression for the squared radial process of Y. Using the shorthand notation $r_t := r_v(Y_t)$ the radial process then becomes

$$r_t^2 = r_0^2 + 2\int_0^t r_s dW_s + \int_0^t \frac{1}{2}\Delta_G r_s^2 ds - \int_0^t \frac{r_s^2}{T-s}ds - 2\int_0^t r_s dL_s. \qquad (6)$$

Imposing a growth condition on the radial process yields an L^2-bound on the radial process of the guided diffusion, [10]. So assume there exist constants $\nu \geq 1$ and $\lambda \in \mathbb{R}$ such that $\frac{1}{2}\Delta_G r_v^2 \leq \nu + \lambda r_v^2$ on $D \setminus \mathrm{Cut}(v)$, for every regular domain $D \subseteq G$. Then (6) satisfies

$$\mathbb{E}[1_{t<\tau_D} r_v(Y_t)^2] \leq \left(r_v^2(e) + \nu t \left(\frac{t}{T-t} \right) \right) \left(\frac{T-t}{t} \right)^2 e^{\lambda t}, \qquad (7)$$

where τ_D is the first exit time of Y from the domain D.

Girsanov Change of Measure. Let B be the Brownian motion in $\mathbb{R}^d$ defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_s), \mathbb{P})$ and X the solution of (1). The process $\frac{\nabla r_v(X_t)^2}{2(T-t)}$ is an adapted process. As X is non-explosive, we see that

$$\int_0^t \left\| \frac{\nabla r(X_s)^2}{2(T-s)} \right\|^2 ds = \int_0^t \frac{r(X_s)^2}{(T-s)^2} ds \leq C, \tag{8}$$

for every $0 \leq t < T$, almost surely, and for some fixed constant $C > 0$. Define a new measure $\mathbb{Q}$ by

$$Z_t := \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t}(X) = \exp\left\{ -\int_0^t \left\langle \frac{\nabla r(X_s)^2}{2(T-s)}, V(X_t)dB_s \right\rangle - \frac{1}{2}\int_0^t \frac{r(X_s)^2}{(T-s)^2} ds \right\}. \tag{9}$$

From (8), the process Z_t is a martingale, for $t \in [0, T)$, and $\mathbb{Q}_t$ defines a probability measure on each $\mathcal{F}_t$ absolutely continuous with respect to $\mathbb{P}$. By Girsanov's theorem (see e.g. [3, Theorem 8.1.2]) we get a new process b_s which is a Brownian motion under the probability measure $\mathbb{Q}$. Moreover, under the probability $\mathbb{Q}$, Eq. (1) becomes

$$dY_t = -\frac{1}{2}V_0(Y_t)dt + V_i(Y_t) \circ \left(db_t^i - \frac{r(Y_t)}{T-t}\left(\frac{\partial}{\partial r}\right)^i dt \right), \tag{10}$$

where $\left(\frac{\partial}{\partial r}\right)^i$ is the i'th component of the unit radial vector field in the direction of v. The squared radial vector field is smooth away from $\mathrm{Cut}(v)$ and thus we set it to zero on $\mathrm{Cut}(v)$. Away from $\mathrm{Cut}(v)$, the squared radial vector field is $2\,\mathrm{Log}_v$, which is the inverse exponential at v. The added drift term acts as a guiding term, which pulls the process towards v at time $T > 0$.

From (9), we see that $\mathbb{E}[f(Y_t)] = \mathbb{E}[f(X_t)Z_t]$. Using (5) and the identity $\Delta_G r_v = \frac{d-1}{r_v} + \frac{\partial}{\partial r_v}\log\Theta_v$ (see [9]), we equivalently write $\mathbb{E}[f(Y_t)\varphi_t] = \mathbb{E}[f(X_t)\psi_t]$, with

$$\psi_t := \exp\left\{ \frac{-r(X_t)^2}{2(T-t)} \right\} \qquad \varphi_t := \exp\left\{ \int_0^t \frac{r_v(Y_s)^2}{T-s}(dA_s + dL_s) \right\}, \tag{11}$$

where $dA_s = \frac{\partial}{\partial r_v}\log\Theta_v$ is a random measure supported on $G \backslash \mathrm{Cut}(v)$ and Θ_v is the Jacobian determinant of Exp_v.

Delyon and Hu in Lie Groups. This section generalizes the result of Delyon and Hu [2, Theorem 5] to the Lie group setting. The result can be modified to incorporate a generalization of [2, Theorem 6].

Theorem 1. *Let X be the solution of (1). The SDE (3) yields a strong solution on $[0, T)$ and satisfies $\lim_{t\uparrow T} Y_t = v$ almost surely. Moreover, the conditional expectation of X given $X_T = v$ is*

$$\mathbb{E}[f(X)|X_T = v] = C\mathbb{E}\left[f(Y)\varphi_T\right], \tag{12}$$

for every $\mathcal{F}_t$-measurable non-negative function f on G, $t < T$, where φ_t is given in (11).

Proof. The result is a consequence of the change of measure together with Lemma 1, Lemma 2, and Lemma 3.

Lemma 1. *The solution of SDE* (3) *satisfies* $\lim_{t\to T} Y_t = v$ *almost surely.*

Proof. Let $\{D_n\}_{n=1}^{\infty}$ be an exhaustion of G, that is, the sequence consists of open, relatively compact subsets of M such that $\bar{D}_n \subseteq D_{n+1}$ and $G = \bigcup_{n=1}^{\infty} D_n$. Furthermore, let τ_{D_n} denote the first exit time of Y from D_n, then from (7) we have that the sequence $\left(\mathbb{E}[1_{\{t<\tau_{D_n}\}} r_v^2(Y_t)]\right)_{n=1}^{\infty}$ is non-decreasing and bounded, hence from the monotone convergence theorem, it has a limit which is bounded by the right-hand side of (7). Applying Jensen's inequality to the left-hand side of (7)

$$\mathbb{E}[r_v(Y_t)] \leq \left(r_v^2(e) + \nu t \left(\frac{t}{T-t} \right) \right)^{\frac{1}{2}} \left(\frac{T-t}{t} \right) e^{\frac{\lambda t}{2}}.$$

Since obviously $\mathbb{E}[r_v(Y_T)] = r_v(Y_T)\mathbb{Q}(r_v(Y_T) \neq 0)$, by Fatou's lemma $\mathbb{E}[r_v(Y_T)] \leq \liminf_{t\to T} \mathbb{E}[r(Y_t)] = 0$, we conclude that $r(Y_t) \to 0$, $\mathbb{Q}$-almost surely.

Lemma 2. *Let* $0 < t_1 < t_2 < \cdots < t_N < T$ *and* h *be a continuous bounded function on* G^N. *With* ψ_t *as in* (11)*, then*

$$\lim_{t\to T} \frac{\mathbb{E}\left[h\left(X_{t_1}, X_{t_2}, \ldots, X_{t_N}\right)\psi_t\right]}{\mathbb{E}[\psi_t]} = \mathbb{E}\left[h\left(X_{t_1}, X_{t_2}, \ldots, X_{t_N}\right) | X_T = v\right]. \quad (13)$$

Proof. The proof is similar to that of [2, Lemma 7]. Let (U, ϕ) be a normal chart centered at $v \in G$. First, since the cut locus of any complete connected manifold has (volume) measure zero, we can integrate indifferently in any normal chart. For any $t \in (t_N, T)$ we have

$$\mathbb{E}[h(x_{t_1}, ..., x_{t_N})\psi_t] = \int_G \varPhi_h(t,z) e^{-\frac{r_v(z)^2}{2(T-t)}} d\,\mathrm{Vol}(z) \quad (14)$$

where $d\,\mathrm{Vol}(z) = \sqrt{\det(A(z))}dz$ denotes the volume measure on G, dz the Lebesgue measure, and A the metric tensor. Moreover,

$$\varPhi_h(t,z) = \int_{G^N} h(z_1, ..., z_N) p(t_1, u, z_1) \cdots p(t - t_N, z_N, z) d\,\mathrm{Vol}(z_1) \cdots d\,\mathrm{Vol}(z_N),$$

and of course $\varPhi_1(t,z) = p(t,e,z)$. Using the normal chart and applying the change of variable $x = (T-t)^{1/2}y$ we get

$$(T-t)^{-\frac{d}{2}} \mathbb{E}[h(x_{t_1}, ..., x_{t_N})\psi_t] \overset{t\to T}{\to} \varPhi_h(T,v) \det(A(v))^{\frac{d}{2}} \int_{\phi(G)} e^{-\frac{r_v(\phi^{-1}(y))^2}{2}} dy.$$

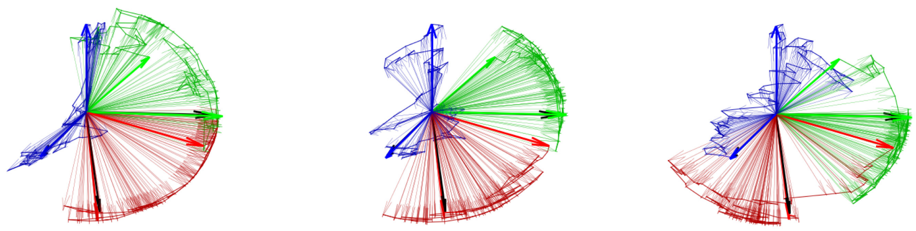
The conclusion follows from Bayes' formula.

Lemma 3. *With* φ_t *as defined above then* $\varphi_t \overset{L_1}{\to} \varphi_T$.

Proof. Note that for each $t \in [0, T)$ we have $\mathbb{E}^{\mathbb{Q}}[\varphi_t] < \infty$ as well as $\varphi_t \to \varphi_T$ almost surely by Lemma 1. The result then follows from the uniform integrability of $\{\varphi_t : t \in [0, T)\}$, which can be found in Appendix C.2 in [9].

4 Importance Sampling and Metric Estimation on SO(3)

This section takes G to be the special orthogonal group of rotation matrices, SO(3), a compact connected matrix Lie group. In the context of matrix Lie groups, computing left-invariant vector fields is straightforward.



(a) (b) (c)

Fig. 1. Three sample paths (a)–(c) of the guided diffusion process on SO(3) visualized by its action on the basis vectors $\{e_1, e_2, e_3\}$ (red, green, blue) of $\mathbb{R}^3$. The sample paths are conditioned to hit the rotation represented by the black vectors. (Color figure online)

Numerical Simulations. The Euler-Heun scheme leads to approximation of the Stratonovich integral. With a time discretization $t_1, \ldots, t_k$, $t_k - t_{k-1} = \Delta t$ and corresponding noise $\Delta B_{t_i} \sim N(0, \Delta t)$, the numerical approximation of the Brownian motion (1) takes the form

$$x_{t_{k+1}} = x_{t_k} - \frac{1}{2}\sum_{j,i} C^j{}_{ij} V_i(x_{t_k})\Delta t + \frac{v_{t_{k+1}} + V_i(v_{t_{k+1}} + x_{t_{k+1}})\Delta B^i_{t_k}}{2} \tag{15}$$

where $v_{t_{k+1}} = V_i(x_{t_k})\Delta B^i_{t_k}$ is only used as an intermediate value in integration. Adding the logarithmic term in (10) to (15) we obtain a numerical approximation of a guided diffusion (3). Figure 1 shows three different sample paths from the guided diffusion conditioned to hit the rotation represented by the black vectors.

Metric Estimation on SO(3). In the d-dimensional Euclidean case, importance sampling yields the estimate [7]

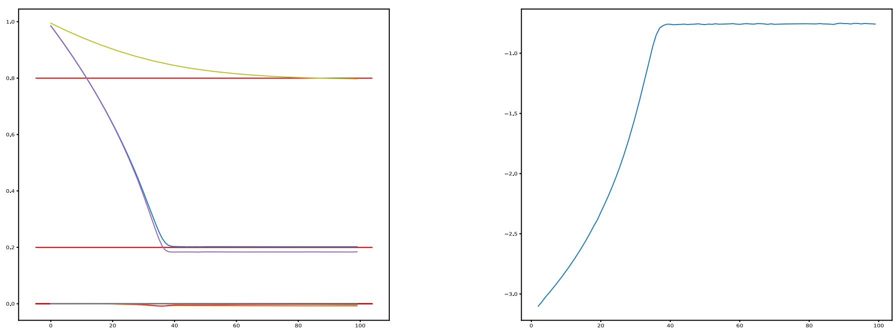
$$p(T, u, v) = \left(\frac{\det\left(A(T, v)\right)}{2\pi T}\right)^{\frac{d}{2}} e^{-\frac{\|u-v\|_A^2}{2T}}\mathbb{E}[\varphi_T],$$

where $\|x\|_A = x^T A(0, u)x$. Thus, from the output of the importance sampling we get an estimate of the transition density. Similar to the Euclidean case, we obtain an expression for the heat kernel $p(T, e, v)$ as $p(T, e, v) = q(T, e)\mathbb{E}\left[\varphi_T\right]$, where

$$q(T, e) = \left(\frac{\det A(v)}{2\pi T}\right)^{\frac{3}{2}} \exp\left(-\frac{d(e, v)^2}{2T}\right) = \left(\frac{\det A(v)}{2\pi T}\right)^{\frac{3}{2}} \exp\left(-\frac{\|\mathrm{Log}_v(e)\|_A^2}{2T}\right), \tag{16}$$

where the equality holds almost everywhere and $A \in \mathrm{Sym}^+(\mathcal{G})$ denotes the metric $A(e)$. The Log_v map in (16) is the Riemannian inverse exponential map.

Figure 2 illustrate how importance sampling on SO(3) leads to metric estimation of the underlying true metric, from which the Brownian motion was generated. We sampled 128 points as endpoints of a Brownian motion from the metric diag(0.2, 0.2, 0.8). We used 20 time steps and sampled 4 bridges per observation. An iterative maximum likelihood method using gradient descent with a learning rate of 0.2, and initial guess of the metric being diag(1, 1, 1) yielded a convergence to the true metric. Note that in iteration the logarithmic map changes.



(a) Estimation of the unknown underlying metric using bridge sampling. Here the true metric is the diagonal matrix diag(0.2, 0.2, 0.8).

(b) The iterative log-likelihood.

Fig. 2. The importance sampling technique applies to metric estimation on the Lie group SO(3). Sampling a Brownian motion from an unknown underlying metric we obtain a convergence to the true underlying metric using an iterative maximum-likelihood method. Here we sampled 4 bridge processes per observation, starting from the metric diag(1, 1, 1), providing a relatively smooth iterative likelihood in 2b.

References

1. Barden, D., Le, H.: Some consequences of the nature of the distance function on the cut locus in a Riemannian manifold. J. LMS **56**(2), 369–383 (1997)
2. Delyon, B., Hu, Y.: Simulation of conditioned diffusion and application to parameter estimation. Stochast. Processes Appl. **116**(11), 1660–1675 (2006)
3. Hsu, E.P.: Stochastic Analysis on Manifolds, vol. 38. AMS, Providence (2002)
4. Kendall, W.S.: The radial part of Brownian motion on a manifold: a semimartingale property. Ann. Probab. **15**(4), 1491–1500 (1987)

5. Le, H., Barden, D.: Itô correction terms for the radial parts of semimartingales on manifolds. Probab. Theory Relat. Fields **101**(1), 133–146 (1995). https://doi.org/10.1007/BF01192198
6. Liao, M.: Lévy Processes in Lie Groups, vol. 162. Cambridge University Press, Cambridge (2004)
7. Papaspiliopoulos, O., Roberts, G.: Importance sampling techniques for estimation of diffusion models. In: Statistical Methods for Stochastic Differential Equations (2012)
8. Shigekawa, I.: Transformations of the Brownian motion on a riemannian symmetric space. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **65**, 493–522 (1984). https://doi.org/10.1007/BF00531836
9. Thompson, J.: Submanifold bridge processes. Ph.D. thesis, University of Warwick (2015)
10. Thompson, J.: Brownian bridges to submanifolds. Potential Anal. **49**(4), 555–581 (2018). https://doi.org/10.1007/s11118-017-9667-1

Bibliography

[1] Arnaudon, A., Holm, D.D., Pai, A., Sommer, S.: A stochastic large deformation model for computational anatomy. In: International Conference on Information Processing in Medical Imaging. pp. 571–582. Springer (2017)

[2] Arnaudon, A., Holm, D.D., Sommer, S.: A geometric framework for stochastic shape analysis. Foundations of Computational Mathematics **19**(3), 653–701 (2019)

[3] Arnaudon, A., van der Meulen, F., Schauer, M., Sommer, S.: Diffusion bridges for stochastic hamiltonian systems with applications to shape analysis. arXiv preprint arXiv:2002.00885 (2020)

[4] Arnaudon, A., van der Meulen, F., Schauer, M., Sommer, S.: Diffusion bridges for stochastic hamiltonian systems and shape evolutions. SIAM Journal on Imaging Sciences **15**(1), 293–323 (2022)

[5] Barden, D., Le, H.: Some consequences of the nature of the distance function on the cut locus in a riemannian manifold. Journal of the LMS pp. 369–383 (1997)

[6] Bismut, J.M.: Large deviations and the malliavin calculus. Birkhauser Prog. Math. **45** (1984)

[7] Bladt, M., Finch, S., Sørensen, M.: Simulation of multivariate diffusion bridges. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **78**(2), 343–369 (Mar 2016). https://doi.org/10.1111/rssb.12118

[8] Bladt, M., Sørensen, M., et al.: Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Bernoulli **20**(2), 645–675 (2014)

[9] Bui, M.N., Pokern, Y., Dellaportas, P.: Inference for partially observed riemannian ornstein–uhlenbeck diffusions of covariance matrices. arXiv preprint arXiv:2104.03193 (2021)

[10] do Carmo, M.P.: Riemannian geometry. translated from the second portuguese edition by francis flaherty. mathematics: Theory & applications. Birkhauser **5052**, 5041–5052 (1992)

[11] Clark, J.: The simulation of pinned diffusions. In: 29th IEEE Conference on Decision and Control. pp. 1418–1420. IEEE (1990)

[12] Cranston, M., Kendall, W.S., March, P.: The radial part of brownian motion ii. its life and times on the cut locus. Probability theory and related fields **96**(3), 353–368 (1993)

[13] Delyon, B., Hu, Y.: Simulation of conditioned diffusion and application to parameter estimation. Stochastic Processes and their Applications (2006)

[14] Driver, B.K.: A Cameron-Martin Type Quasi-Invariance Theorem for Pinned Brownian Motion on a Compact Riemannian Manifold. Transactions of the American Mathematical Society **342**(1), 375–395 (1994)

[15] Durham, G.B., Gallant, A.R.: Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. Journal of Business & Economic Statistics **20**(3), 297–338 (2002)

[16] Elworthy, D., Truman, A.: Classical mechanics, the diffusion (heat) equation and the schrödinger equation on a riemannian manifold. Journal of mathematical physics pp. 2144–2166 (1981)

[17] Elworthy, K., Truman, A.: The diffusion equation and classical mechanics: an elementary formula. In: Stochastic processes in quantum theory and statistical physics, pp. 136–146. Springer (1982)

[18] Emery, M.: Stochastic Calculus in Manifolds. Universitext, Springer Berlin Heidelberg, Berlin, Heidelberg (1989)

[19] Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. In: Annales de l'institut Henri Poincaré. vol. 10, pp. 215–310 (1948)

[20] Golightly, A., Wilkinson, D.J.: Markov chain monte carlo algorithms for sde parameter estimation. Learning and Inference for Computational Systems Biology pp. 253–276 (2010)

[21] Güneysu, B.: On the semimartingale property of brownian bridges on complete manifolds. ALEA **16**, 15–31 (2019)

[22] Hansen, P., Eltzner, B., Huckemann, S.F., Sommer, S.: Diffusion means in geometric spaces. arXiv preprint arXiv:2105.12061 (2021)

[23] Hansen, P., Eltzner, B., Huckemann, S.F., Sommer, S.: Diffusion Means in Geometric Spaces. arXiv:2105.12061 (May 2021)

[24] Hansen, P., Eltzner, B., Sommer, S.: Diffusion means and heat kernel on manifolds. arXiv preprint arXiv:2103.00588 (2021)

[25] Hsu, E.P.: Stochastic analysis on manifolds, vol. 38. AMS (2002)

[26] Hsu, E.P.: Stochastic Analysis on Manifolds. American Mathematical Soc. (2002)

[27] Hsu, P.: Brownian bridges on Riemannian manifolds. Probability Theory and Related Fields **84**, 103–118 (Mar 1990)

[28] Jensen, M.H., Joshi, S., Sommer, S.: Bridge simulation and metric estimation on lie groups. arXiv preprint arXiv:2106.03431 (2021)

[29] Jensen, M.H., Mallasto, A., Sommer, S.: Simulation of conditioned diffusions on the flat torus. In: International Conference on Geometric Science of Information. pp. 685–694. Springer (2019)

[30] Kallenberg, O.: Foundations of modern probability. Springer (1997)

[31] Kendall, W.S.: The radial part of brownian motion on a manifold: a semimartingale property. The Annals of Probability **15**(4), 1491–1500 (1987)

[32] Le, H., Barden, D.: Itô correction terms for the radial parts of semimartingales on manifolds. Probability theory and related fields **101**(1), 133–146 (1995)

[33] Lee, J.M.: Introduction to Smooth Manifolds, vol. 218. Springer Science & Business Media (2013)

[34] Li, X.M.: Generalised brownian bridges: examples. arXiv preprint arXiv:1612.08716 (2016)

[35] Papaspiliopoulos, O., Roberts, G.O.: Importance sampling techniques for estimation of diffusion models. In: Statistical Methods for Stochastic Differential Equations. Chapman & Hall/CRC Press (2012)

[36] Pedersen, A.R.: Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. Bernoulli pp. 257–279 (1995)

[37] Pennec, X., Sommer, S., Fletcher, T.: Riemannian Geometric Statistics in Medical Image Analysis. Academic Press (2019)

[38] Pennec, X., Sommer, S., Fletcher, T.: Riemannian Geometric Statistics in Medical Image Analysis. Elsevier (2020)

[39] Revuz, D., Yor, M.: Continuous Martingales and Brownian motion. Springer (1999)

[40] Roberts, G.O., Stramer, O.: On inference for partially observed nonlinear diffusion models using the metropolis–hastings algorithm. Biometrika **88**(3), 603–621 (2001)

[41] Schauer, M.: Bayesian inference for discretely observed diffusion processes. Ph.D. thesis, Delft University of Technology (2015)

[42] Schauer, M., van der Meulen, F., van Zanten, H.: Guided proposals for simulating multi-dimensional diffusion bridges. Bernoulli **23**(4A), 2917–2950 (Nov 2017). https://doi.org/10.3150/16-BEJ833

[43] Sommer, S., Bronstein, A.M.: Horizontal Flows and Manifold Stochastics in Geometric Deep Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–1 (2020). https://doi.org/10.1109/TPAMI.2020.2994507

[44] Sommer, S., Arnaudon, A., Kuhnel, L., Joshi, S.: Bridge simulation and metric estimation on landmark manifolds. In: Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics, pp. 79–91. Springer (2017)

[45] Sommer, S., Arnaudon, A., Kuhnel, L., Joshi, S.: Bridge Simulation and Metric Estimation on Landmark Manifolds. In: Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics. pp. 79–91. Lecture Notes in Computer Science, Springer, Cham (Sep 2017), dOI: 10.1007/978-3-319-67675-3_8

[46] Sommer, S., Bronstein, A.M.: Horizontal flows and manifold stochastics in geometric deep learning. IEEE transactions on pattern analysis and machine intelligence (2020)

[47] Thompson, J.: Submanifold bridge processes. Ph.D. thesis, University of Warwick (2015)

[48] Thompson, J.: Brownian bridges to submanifolds. Potential Analysis **49**(4), 555–581 (2018)

[49] Thompson, J., Li, X.M.: Submanifold Bridge Processes. Ph.D. thesis, University of Warwick, Coventry (2015)

[50] van der Meulen, F., Schauer, M.: Continuous-discrete smoothing of diffusions. arXiv:1712.03807 [stat] (Dec 2017)