POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Schrödinger Bridges on the Manifold of SPD Matrices

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Author: **Francesco Bellotto**

Student ID: 244854
Advisor: Prof. Laura Maria Sangalli
Co-advisors: Prof. Victor Panaretos, Leonardo Santoro
Academic Year: 2025-26

Abstract

Symmetric and positive definite (SPD) matrices arise in many scientific contexts, and one is often confronted with the challenge of finding a meaningful explanation for the interaction between two sets of SPD matrix-valued data. A prominent example comes from neuroscience, where it is desirable to understand the relationship between structural connectivity and functional connectivity, two complementary ways of characterizing the brain [21].

A powerful approach to this problem is generative modeling: by viewing the two different types of data as samples from two different distributions, one can learn an interpolation between them, which can then be inspected to identify interpretable interactions.

In this work, we introduce and study, from a theoretical point of view, Schrödinger bridges over the manifold of SPD matrices endowed with the so-called *log-Cholesky* Riemannian metric [16]. The motivation for using Schrödinger bridges to study relationships between distributions is that they admit a physical interpretation, making them particularly suitable for scientific inspection. Moreover, the simple structure of this Riemannian manifold allows us to develop a simple strategy for learning Schrödinger bridges between distributions on it by leveraging $[SF]^2M$, an existing framework for fitting Schrödinger bridges in Euclidean space [20].

Achieving this goal first requires a different formulation of $[SF]^2M$, together with a proof of its correctness under suitable assumptions. We then provide a precise characterization of the SPD manifold endowed with this geometry; in particular, we show that it is globally isometric to Euclidean space through an isometric diffeomorphism. Finally, we study concepts central to the Schrödinger bridge framework, such as Brownian motion and Brownian bridges on the SPD manifold, before concluding with the proposed framework for learning Schrödinger bridges under this geometry.

Keywords: Symmetric and Positive Definite (SPD) matrices, Schrödinger bridge, log-Cholesky metric, $[SF]^2M$, Markov projection

Abstract in lingua italiana

Le matrici simmetriche e definite positive (SPD) emergono in numerosi contesti scientifici ed è spesso necessario affrontare il problema di individuare una spiegazione significativa dell'interazione tra due insiemi di dati costituiti da matrici SPD. Un esempio rilevante proviene dalle neuroscienze, dove è di particolare interesse comprendere la relazione tra connettività strutturale e connettività funzionale, due modalità complementari di caratterizzare il cervello [21].

Un approccio efficace a questo problema è rappresentato dalla modellazione generativa: considerando i due tipi di dati come campioni estratti da due distribuzioni differenti, è possibile apprendere un'interpolazione tra di esse, analizzabile per individuare interazioni interpretabili.

In questo lavoro introduciamo e studiamo i ponti di Schrödinger sulla varietà delle matrici SPD dotata della metrica riemanniana *log-Cholesky* [16]. La motivazione alla base del loro impiego nello studio delle relazioni tra distribuzioni risiede nella loro interpretazione fisica, che li rende particolarmente adatti all'analisi scientifica. Inoltre, la struttura semplice di questa varietà consente di sviluppare una strategia per apprendere ponti di Schrödinger tra distribuzioni definite su di essa, sfruttando $[SF]^2M$, un framework esistente per il caso euclideo [20].

Il raggiungimento di questo obiettivo richiede innanzitutto una riformulazione di $[SF]^2M$, accompagnata da una dimostrazione della sua correttezza sotto opportune ipotesi. Forniamo quindi una caratterizzazione della varietà delle matrici SPD dotata di questa geometria, mostrando che essa è globalmente isometrica allo spazio euclideo tramite un diffeomorfismo isometrico. Infine, studiamo alcuni concetti fondamentali del formalismo dei ponti di Schrödinger, quali il moto browniano e i ponti browniani sulla varietà SPD, e concludiamo presentando il framework proposto per l'apprendimento di ponti di Schrödinger in questa geometria.

Parole chiave: matrici Simmetriche e Definite Positive (SDP), ponte di Schrödinger, metrica di log-Cholesky, $[SF]^2M$, proiezione di Markov

Contents

Introduction

Symmetric and positive definite (SPD) matrices arise across a wide range of scientific disciplines, spanning computer vision ([4]), medical imaging ([7]), neuroscience ([9]), and finance ([3]). A prominent example comes from neuroscience, where two complementary descriptions of the brain are encoded as SPD matrix-valued data. Specifically, if the brain is parcellated into distinct regions, one can measure the strength of physical anatomical pathways connecting pairs of regions, yielding *structural connectivity* (SC), or measure the functional fluctuations across these regions, producing *functional connectivity* (FC). As these data are intrinsically graph-valued, they can be represented by means of SPD matrices.

When working with two related data modalities, such as functional and structural connectivities, it is often valuable to characterize the underlying mapping or connection between them. A powerful paradigm to achieve this is generative modeling.

Generative models, such as diffusion models ([12]) or continuous normalizing flows ([6]), aim in this context to map two distributions $\mathbb{Q}_0$ and $\mathbb{Q}_1$ through a learned transformation pushing the former forward to the latter. This way, they essentially represent an interpolation between two probability distributions. The mathematical nature of this transformation distinguishes different generative architectures; it can be either deterministic (where an initial sample $x_0 \sim \mathbb{Q}_0$ maps uniquely to a final state $x_1 \sim \mathbb{Q}_1$) or probabilistic (where x_0 evolves to some x_1 that may change across different realizations). Furthermore, instead of treating the model as a "black box" that merely pairs endpoints, it is often desirable to construct a *continuous transformation* from one datum to another. This yields a trajectory over the data space that can offer meaningful domain-specific interpretations.

This trajectory-based approach is highly relevant for uncovering interpretable connections between SCs and FCs. By viewing observed SCs and FCs as empirical samples from distinct distributions, we can train a continuous generative model to transform a structural connectome into a functional one (or vice versa). The resulting SPD-valued trajectory can provide insights into the biophysical interplay between these two modalities of brain organization.

Though conceptually clear, this approach yields the difficulty of handling trajectories over the space of SPD matrices, which in turn requires to look at the geometry of such space. It can in fact be assigned the structure of a smooth manifold, and endowed with a Riemannian metric making it a Riemannian manifold. As we will see, the choice of the metric is fundamental, and has to be balanced between its descriptive power of the space and the simplicity of the model.

The strategy of looking for interpretable interactions between SCs and FCs leveraging learned trajectories is precisely the one adopted by Zhou et al. [21], where Riemannian flow matching, introduced in [5] as an extension to Euclidean flow matching, is deployed to train a continuous normalizing flow over the SPD manifold, mapping structural connectivities directly to functional connectivities.

In Zhou et al. [21], the chosen geometry is associated with the *log-Cholesky metric*, introduced by Lin [16]. The clear advantage of it is that it *linearizes*, in some sense, the manifold geometry, turning the SPD manifold into a *flat* space similar to $\mathbb{R}^d$, where $d = \frac{n(n+1)}{2}$ and n is the size of the matrices. Specifically, the log-Cholesky metric is just the usual Euclidean scalar product on $\mathbb{R}^d$ mapped through a diffeomorphism between the SPD manifold and $\mathbb{R}^d$ (this makes of such map an *isometric diffeomorphism*). This nature implies, at least at an intuitive level, that objects residing over the log-Cholesky manifold can easily be obtained as the image through this isometry of their Euclidean counterparts.

While standard flow matching frameworks (both Euclidean and Riemannian) construct deterministic mappings between samples, the stochastic interpolant method $[SF]^2M$ (*Simulation Free Schrödinger bridges via Score and Flow Matching*) developed by Tong et al. [20] generalizes this paradigm to stochastic dynamics within Euclidean space. Specifically, this method aims to learn mixtures of processes, each solving a stochastic differential equation; processes with this form encompass the solution to the *Schrödinger bridge problem* between two endpoint distributions. Conceptually, a Schrödinger bridge identifies the stochastic process that interpolates two target marginals while minimizing the relative entropy (Kullback-Leibler divergence) with respect to a reference process, such as Brownian motion. This physical and information-theoretic grounding makes Schrödinger bridges highly appealing for generative modeling.

Unlike deterministic flow matching, stochastic interpolants can capture uncertainty in the transport by modeling a distribution over paths rather than a single trajectory. As a result, they provide greater flexibility for connecting complex or multimodal distributions and naturally recover entropy-regularized optimal transport solutions such as Schrödinger bridges.

The primary objective of this thesis is to extend the $[SF]^2M$ framework to data residing on the SPD manifold endowed with the log-Cholesky Riemannian metric. Beyond its clear applicability to modeling brain connectomes, the log-Cholesky metric simplifies this geometric extension. Because it establishes a global isometry between the SPD manifold and $\mathbb{R}^d$, generalizing $[SF]^2M$ becomes considerably more tractable than on generic Riemannian manifolds. We will demonstrate that fundamental stochastic objects—including Brownian motion and Brownian bridges—on the log-Cholesky manifold are simply the images of their Euclidean counterparts under this isometric diffeomorphism.

Although our theoretical extensions apply to general mixtures of SDE solutions, we focus primarily on parameterizing Schrödinger bridges. Consequently, significant emphasis is placed on formalizing Brownian motion and Brownian bridges within the log-Cholesky geometry, as these objects are indispensable for defining and solving the underlying Schrödinger problem.

Before moving on with the explanation of how the thesis is organized, let us look at a high-level formulation of the two models mentioned so far: Riemannian Flow Matching (for fitting Riemannian Continuous Normalizing Flows) and $[SF]^2M$.

Riemannian Flow Matching

Formally, given two manifold-valued distributions $\mathbb{Q}_0$ and $\mathbb{Q}_1$ supported on a manifold N, Riemannian flow matching aims to learn a time-dependent vector field (the *continuous flow*) $u(t, x)$ for $t \in [0, 1]$ and $x \in N$. If a trajectory $(x_t)_{t\in[0,1]} \subseteq N$ satisfies the ordinary differential equation:

$$\begin{cases} \mathrm{d}x_t = u(t, x_t)\mathrm{d}t, \quad \forall t \in (0, 1] \\ x_0 \sim \mathbb{Q}_0 \end{cases} \tag{1}$$

then the terminal state x_1 follows the distribution $\mathbb{Q}_1$. A key constraint is that the trajectory must remain on the manifold, meaning $x_t \in N$ for all $t \in [0, 1]$.

If $g(\cdot, \cdot)$ denotes the Riemannian metric on N and (X_0, X_1) is a pair of initial-final datapoint, so that $X_0 \sim \mathbb{Q}_0$ and $X_1 \sim \mathbb{Q}_1$, Riemannian flow matching typically estimates the continuous normalizing flow by minimizing the expected error:

$$\mathcal{L}(\theta) = \mathbb{E}_{t\sim\mathcal{U}[0,1],(x_0,x_1)\sim\Pi(\mathbb{Q}_0,\mathbb{Q}_1)}\left[\|\dot{x}_t^{(x_0,x_1)} - v_\theta(t,x)\|_g^2\right]$$

where $x_t^{(x_0,x_1)}$ is the geodesic (under the adopted metric) from x_0 to x_1, $\mathcal{U}[0,1]$ is the uniform distribution over $[0,1]$ and $\Pi(\mathbb{Q}_0,\mathbb{Q}_1)$ is the set of transport plans between $\mathbb{Q}_0$ and $\mathbb{Q}_1$. Here $v_\theta(t,x)$ is a family of flows parametrized by θ, which is the parameter we optimize for. In practice, this will most of the time consist of a neural network, and θ is the vector made of its weights.

It is clear that the choice of the Riemannian metric $g(\cdot,\cdot)$ is fundamental, as it entirely dictates the geometry of the geodesics and the explicit form of the target velocity $\dot{x}_t^{(x_0,x_1)}$.

Notice that $\dot{x}_t^{(x_0,x_1)}$ can be interpreted as a flow conditioned on initial and final data-points. This choice of conditional information is typical, though the framework handles general ones (i.e. conditioning on a possibly latent general random variable Z).

[SF]^{2}M

Given an initial distribution $\mathbb{Q}_0$ and a target distribution $\mathbb{Q}_1$ on $\mathbb{R}^d$, the objective is to learn a stochastic process $(X_t)_{t\in[0,1]} \subseteq \mathbb{R}^d$ such that $X_0 \sim \mathbb{Q}_0$ and $X_1 \sim \mathbb{Q}_1$, structured as a mixture of diffusion processes. Specifically, let us assume that the conditional process $X^{(x_0,x_1)} = (X|(X_0,X_1) = (x_0,x_1))$ satisfies an SDE of the form:

$$\mathrm{d}X^{(x_0,x_1)} = b(t, X^{(x_0,x_1)}, (x_0,x_1))\mathrm{d}t + g(t)\mathrm{d}W_t, \quad t \in (0,1) \tag{2}$$

where W_t is a d-dimensional standard Brownian motion. The law of the full process can be written as the mixture:

$$\mathbb{P}((X_t)_t \in \cdot) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{P}(X_t^{(x_0,x_1)} \in \cdot)q(\mathrm{d}x_0,\mathrm{d}x_1) \tag{3}$$

where $q \in \Pi(\mathbb{Q}_0,\mathbb{Q}_1)$ is a transport plan between the two distributions. As demonstrated by Tong et al. [20], the framework does not directly simulate X, but rather identifies its Markov projection ξ (i.e. a Markov process sharing the same time-marginal distributions, $\xi_t \sim X_t$ for all $t \in [0,1]$). This is achieved by fitting a drift field $b(t,x)$ such that:

$$\begin{cases} \mathrm{d}\xi_t = b(t, \xi_t)\mathrm{d}t + g(t)\mathrm{d}W_t, \quad t \in (0, 1] \\ \xi_0 \sim \mathbb{Q}_0 \end{cases} \tag{4}$$

This formulation extends standard Euclidean flow matching to stochastic regimes, reducing to the deterministic case when the diffusion coefficient $g(t) = 0$.

Once again, notice that in the general case the conditional information will not necessarily be (X_0, X_1), but rather any random variable Z. However, the former choice encompasses the form of the Schrödinger bridge.

Though clear and conceptually simple, extending $[SF]^2M$ to the SPD manifold endowed with the log-Cholesky metric reveals several subtle difficulties. On the one hand, the method introduced by Tong et al. [20] does not recover the target process itself, but rather its Markovization; consequently, we must provide an alternative interpretation of this concept that is amenable to an extension to manifolds. On the other hand, the log-Cholesky manifold must be characterized through an isometric diffeomorphism with Euclidean space, yielding a perspective that differs from the characterization given by Lin [16]. Furthermore, stochastic analysis concepts such as Brownian motion and Brownian bridges on Riemannian manifolds must be understood and adapted to this specific setting.

The thesis is organized as follows. Chapter 1 is dedicated to introducing $[SF]^2M$ over the Euclidean space and summarizing the main results in Tong et al. [20]. While the first two sections of the chapter are just a summary/rephrasing of [20], in Section 1.3 we define the concept of Markov projection in a different way than what is done in the work, which mimics the projection introduced by Gyöngy [10]; we show how, and under what assumptions, we can actually obtain the Markovization of a mixture of processes. Not only, but we prove that in the case of a mixture of Brownian bridges (relevant for the Schrödinger problem) such assumptions are satisfied. We conclude the chapter by explaining, in Section 1.4, how to practically fit $[SF]^2M$, as explained in Tong et al. [20].

Chapter 2 provides a rigorous overview of the geometry of the SPD manifold under the log-Cholesky metric. This analysis is organized into two primary sections. Section 2.2 studies an intermediate space: the *Cholesky manifold* composed of lower triangular matrices with positive diagonal entries. As shown by Lin [16], an isometric diffeomorphism exists between the SPD manifold and this Cholesky space. We extend this by explicitly constructing an isometric diffeomorphism between the Cholesky manifold and $\mathbb{R}^d$ (or equivalently, the space of lower triangular matrices $\mathcal{L}$ equipped with the standard Frobenius inner product). This intermediate step allows us to treat the log-Cholesky manifold as the image of a flat Euclidean space, which we formalize in Section 2.3. Because this perspective differs slightly from the presentation in Lin [16], we validate our geometric characterization in Section A.2 by providing alternative, isometry-based proofs for several results contained in that work. Additionally, we characterize the tangent spaces of these manifolds, emphasizing their representations in natural bases, which is vital since the coefficients of manifold-valued SDEs are elements of these tangent spaces.

Chapter 3 develops the necessary tools for stochastic analysis on manifolds, which are required to define SDEs, Brownian motion, and Brownian bridges on the log-Cholesky space. Section 3.2 introduces general manifold SDEs and classical path measures, while Section 3.3 specializes these concepts to manifolds admitting an isometric diffeomorphism to Euclidean space. Within this setting, we characterize the Levi-Civita connection, horizontal vector fields, and the stochastic horizontal lift, proving that Brownian motions, Brownian bridges, and Schrödinger bridges on the log-Cholesky manifold are simply the pushed-forward images of their classical Euclidean counterparts.

We synthesize these geometric and stochastic foundations in Chapter 4 to construct the formal extension of $[SF]^2M$ to the log-Cholesky manifold. After defining the model architecture and proving that its Markov projection solves a valid manifold-valued SDE in Section 4.3, we analyze the explicit, entry-wise behavior of these matrix-valued SDEs in Section 4.4.

In Chapter 5, we explore an extension of $[SF]^2M$ to infinite-dimensional Hilbert spaces (e.g., $L^2(\Omega)$ functions over a compact domain $\Omega \subset \mathbb{R}^N$). Since our primary interest lies in parameterized Schrödinger bridges—which manifest as mixtures of Brownian bridges under a Wiener reference measure—we derive representations of infinite-dimensional Brownian bridges as solutions to abstract stochastic evolution equations. This establishes the mathematical framework required to formally extend the Markov projection calculations of Section 1.3 to infinite dimensions.

We conclude in Chapter 6 by giving an overview of the work. We state our contributions as well as their limitations, and suggest some possible future directions.

While each chapter contributes to the core thesis goal realized in Chapter 4, a reader interested primarily in the final model may focus on Chapter 1, the geometric summary in Figure 2.1, and the high-level concepts presented in sections 3.3.3, 3.3.4, and 3.3.5.

The novel contributions of this work are distributed as follows: Chapter 1 contains the formalization and sufficiency proofs for the Markov projection of mixture processes; Chapter 4 provides the complete mathematical framework for $[SF]^2M$ on the log-Cholesky manifold, and Chapter 5 derives the SDE formulations for infinite-dimensional Brownian bridges. Additionally, while Chapter 2 and Chapter 3 cover background material, they feature alternative proofs and geometric derivations not found in the existing literature. All original proofs and derivations throughout this text are explicitly denoted with an asterisk ($*$).

1 | SF^{2}M

In this chapter, we summarize some of the results obtained in Tong et al. [20], and provide a path to prove some of them that is different from the one leveraged by the authors, but more suited to our purposes.

In the paper, the authors introduce a framework to parametrize Schrödinger bridges (but more in general mixtures of SDE solutions) mapping an initial distribution to a target one, as illustrated by an example in Figure 1.1. Since this method extends both score and flow matching models, both of which can be trained in a simulation-free manner, it is referred to $[SF]^2M$ (Simulation-Free Schrödinger bridges via Flow and Score Matching).

In Section 1.2, we define the starting problem that the paper aims to solve, describe the model leveraged to do it and discuss its link to the well-known problem of finding Schrödinger bridges (see [17] for a detailed survey on the argument).

After, in Section 1.3 we define what is meant by Markov projection of a process, and take a different path than [20] to the problem of finding the Markov projection of a Schrödinger bridge. We thus show that under suitable assumptions on the involved distributions we can actually find one.

We conclude in Section 1.4, where the more practical issue of fitting the proposed model is addressed.

1.1. Notations

In the following of the chapter, we will use the following notations. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, with filtration $(\mathcal{F}_t)_{t\in[0,T]}$. All considered processes are adapted to such filtration. A stochastic process X with values in $\mathbb{R}^d$ will be denoted as $X = (\Omega, \mathcal{F}, (\mathcal{F}_t)_t, (X_t)_t, \mathbb{P})$. $\mathbb{R}^d$ is always endowed with the Borel σ-algebra $\mathscr{B}(\mathbb{R}^d)$ (generated by all open sets in $\mathbb{R}^d$). A process X can equivalently be viewed as a random variable from $(\Omega, \mathcal{F}, \mathbb{P})$ to $W(\mathbb{R}^d)$, the space of continuous paths $[0, T] \to \mathbb{R}^d$, endowed with its Borel σ-algebra $\mathscr{B}(W(\mathbb{R}^d))$. The law of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ will be denoted

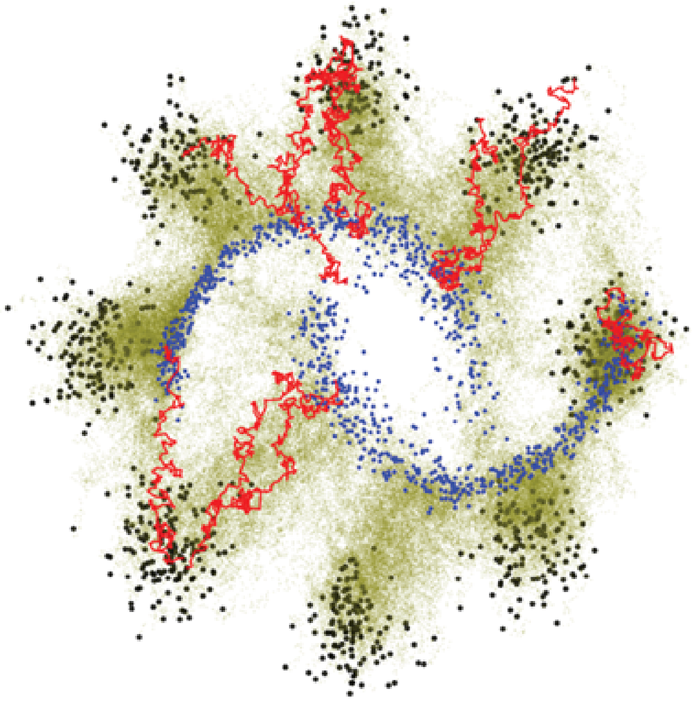

Figure 1.1: Example of trajectories for the learned process. Starting distribution is a two moons, final distribution is a mixture of eight gaussians. Figure taken from Tong et al. [20].

by $\mathbb{P}^X = \mathbb{P} \circ X^{-1}$.

Notice that, when needed, $\mathbb{R}^d$ can be replaced, in the above conventions, with any topological space E (e.g. a smooth manifold M). Though the framework described in Tong et al. [20] puts the focus on $\mathbb{R}^d$, we will introduce some concepts in a more general setting, substituting the Euclidean space with smooth manifolds or general measurable spaces, in order to better exploit them later in our extension effort.

1.2. Problem and Model Definition

The following is the motivating problem

Problem 1.1. Let us consider two probability measures on the probability space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, $\mathbb{Q}_0$ and $\mathbb{Q}_T$. We look for a stochastic process $X = (\Omega, \mathcal{F}, (\mathcal{F}_t)_t, (X_t)_t \subseteq \mathbb{R}^d, \mathbb{P})$ such that $X_0 \sim \mathbb{Q}_0$ and $X_T \sim \mathbb{Q}_T$.

In order to address Problem (1.1), in Tong et al. [20] the search is restricted to a specific type of processes. Loosely, we can define these as a mixture of solutions to stochastic differential equations, and formally write them in the general form:

$$\mathbb{P}((X_t)_{t\in[0,T]} \in \cdot) = \int_{\mathcal{Q}} \mathbb{P}((X_t)_{t\in[0,T]} \in \cdot | Z = z) q(dz) \tag{1.1}$$

where $\mathcal{Q}$ is some (possibly latent) space over which the conditioning variable $Z \sim dq$ takes values.

In the following, we make this definition more precise.

Let $Z : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{Q}, \mathcal{A})$ be a random variable with law $d\mathbb{P}^Z(z) = q(dz)$, where $(\mathcal{Q}, \mathcal{A})$ is a measurable space.

Recall the definition of conditional law:

Definition 1.1. Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ and $Z : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{Q}, \mathcal{A})$ be random variables. We say that, if it exists, $n(z, dx)$ is the **conditional law of X given $Z = z$** if:

- $\forall A \in \mathcal{E}$, the function $z \mapsto n(z, A)$ is $\mathcal{A}$-measurable;
- $\forall A \in \mathcal{E}$ and $\forall B \in \mathcal{A}$:

$$\mathbb{P}(X \in A, Z \in B) = \int_B n(z, A) q(dz) \tag{1.2}$$

  being dq the law of Z.

We assume that the conditional law of X given $Z = z$ exists for all $z \in \mathcal{Q}$, and we denote it with $n(z, dx)$ or equivalently with $\mathbb{P}(dx|Z = z)$. Then by definition:

$$\mathbb{P}(X \in A) = \int_{\mathcal{Q}} \mathbb{P}(X \in A | Z = z) q(dz) \quad \forall A \in \mathscr{B}(W(\mathbb{R}^d))$$

Assume also that for all $z \in \mathcal{Q}$, there exists:

- A filtered probability space $(\Omega^z, \mathcal{F}^z, (\mathcal{F}^z_t)_t, \mathbb{P}^z)$;
- A process Y^z on it, with values in $\mathbb{R}^d$ such that it is the solution to an SDE with properly chosen initial condition (we will come back later to this choice):

$$dY^z_t = b(t, Y^z_t, z) dt + g(t) dW_t \tag{1.3}$$

  where $b : [0, T] \times \mathbb{R}^d \times \mathcal{Q} \to \mathbb{R}^d$, $g : [0, T] \to \mathbb{R}^{d\times m}$ are measurable functions, W is an m-dimensional Brownian motion and the law of Y^z is $n(z, \cdot)$.

This way, we are always justified to write for all $A \in \mathscr{B}(W(\mathbb{R}^d))$:

$$\mathbb{P}(X \in A) = \int_{\mathcal{Q}} \mathbb{P}^z(Y^z \in A) dq(z)$$

recovering what we roughly defined through Equation (1.1).

Once X is defined this way, we need to force it to achieve the right start and end distributions, in order for it to solve problem 1.1.

Recall that a Brownian bridge in $\mathbb{R}^d$ with starting point $a \in \mathbb{R}^d$ and end point $b \in \mathbb{R}^d$ at times, respectively, 0 and $T > 0$, is essentially a Brownian motion that is conditioned to equal a at time 0 and b at time t. A way to achieve what we want is summarized in the following

Proposition 1.1. *Under the above notations, let $\mathcal{Q} = \mathbb{R}^d \times \mathbb{R}^d$, $q \in \Pi(\mathbb{Q}_0, \mathbb{Q}_1)$ (i.e. a transport plan from the initial to the final distribution) and $Y_t^z = Y_t^{(z_0, z_T)}$ be a Brownian bridge from z_0 to z_T at times, resp., 0 and T. Then $X_0 \sim \mathbb{Q}_0$ and $X_T \sim \mathbb{Q}_T$.*

Proof.* Let $A \in \mathscr{B}(\mathbb{R}^d)$: we want to show that $\mathbb{P}(X_0 \in A) = \mathbb{Q}_0(A)$.

$$\begin{aligned}\mathbb{P}(X_0 \in A) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{P}^z(Y_0^{(z_0, z_T)} \in A) dq(z_0, z_T) = \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbf{1}_A(z_0) dq(z_0, z_T) = \\ &= \int_{\mathbb{R}^d} \mathbf{1}_A(z_0) d\mathbb{Q}_0(z_0) = \mathbb{Q}_0(A)\end{aligned}$$

where in the third passage we used the definition of transport plan.

In the exact same way, one proves $\mathbb{P}(X_T \in A) = \mathbb{Q}_T(A)$. □

From now on, unless stated otherwise, we will consider the conditional processes to be Brownian bridges.

The reason for choosing as model a mixture of Brownian bridges weighted by a transport plan is not only that it provides a simple and interpretable way to tackle Problem 1.1, but it allows us to choose the transport plan so that we can obtain the solution to a well studied problem: the Schrödinger bridge.

First, recall the definition of Radon-Nikodym derivative:

Theorem 1.1 (Radon-Nikodym)**.** *Let $(X, \mathscr{M})$ be a measurable space and let μ, ν be two σ-finite measures on it. Then if $\nu << \mu$, there exists a μ-a.e. unique, $\mathscr{M}$-measurable function $f : X \to [0, +\infty)$ such that:*

$$\nu(A) = \int_A f d\mu \tag{1.4}$$

for all $A \in \mathscr{M}$. This is often denoted by $\frac{d\nu}{d\mu}$

Definition 1.2 (Schrödinger Bridge)**.** Let E be a smooth manifold, q_0 and q_T be two probability distributions on $(E, \mathscr{B}(E))$. A **Schrödinger bridge** between q_0 and q_T is defined as a probability measure $\mathbb{P}^*$ on $(W(E), \mathscr{B}(W(E)))$ such that:

$$\mathbb{P}^* = \arg \min_{p_0 = q_0; p_T = q_T} KL(\mathbb{P}|\mathbb{T}) \tag{1.5}$$

where we define:

$$KL(\mathbb{P}|\mathbb{T}) = \int_{W(E)} \log \left(\frac{d\mathbb{P}}{d\mathbb{T}} \right) d\mathbb{P} \tag{1.6}$$

$\mathbb{T}$ is some reference measure on $(W(E), \mathscr{B}(W(E)))$ and p_0 and p_T are the time marginals at times 0 and T, respectively, for $\mathbb{P}$.

Now the following result makes precise what has been explained before, and clarifies the reason for choosing the above form for process X

Proposition 1.2 (Proposition 2.1. in [20])**.** *The unique solution to the Schrödinger bridge problem in $\mathbb{R}^d$ with reference process a Brownian motion with diffusion rate σ is given by:*

$$\mathbb{P}^* = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{P}_{x_0, x_T} d\pi^*_{2\sigma^2}(x_0, x_T) \tag{1.7}$$

where $\mathbb{P}_{x_0, x_T}$ is the law of a Euclidean Brownian bridge from x_0 to x_T with diffusion rate σ and $\pi^*_{2\sigma^2}(x_0, x_T)$ is the entropic OT plan between q_0 and q_T, meaning:

$$\pi^*_\epsilon(q_0, q_T) = \arg \min_{\pi \in \Pi(\hat{q}_0, \hat{q}_T)} \int_{\mathbb{R}^d \times \mathbb{R}^d} d(x_0, x_T)^2 d\pi(x_0, x_T) + \epsilon KL(\pi|q_0 \otimes q_T) \tag{1.8}$$

So far, we constructed a process solving Problem 1.1 by considering a mixture of Brownian bridges weighted by a suitable transport plan, and saw that in the particular case where it happens to be the (entropically regularized) optimal one, the obtained process is actually a Schrödinger bridge. However, we still need a way to practically simulate it. The strategy is as follows: find a process ξ that is Markov and solves an SDE problem whose one-dimensional marginals are the same as X for all $t \in [0, T]$; then, fit the parameters of such an SDE using available data. In the end, we will be able to simulate paths sampled from ξ, with the guarantee that 1D marginals coincide with the ones of X, and the computational advantage of having to simulate a Markov process.

1.3. Markov Projection

In this section we address the problem of starting with a process that is not necessarily Markovian, and ending up with one that is Markov and has the same one-dimensional marginals as the first one. From a computational perspective, simulating a Markov process is easier, since one-dimensional marginals do not depend on the whole past history of the process, meaning that at simulation time we only need to have access to the previous time state, rather than all states from the beginning. In our context, though, there is a second advantage: we will see that the Markov projection of the original process is the solution to an SDE, meaning we can leverage any scheme for numerical solution of stochastic differential equations to simulate it (e.g. Euler-Maruyama, which details can be found in Baldi [1]).

As stated in Léonard [17], when we make the particular choice for the mixture weights being the entropy-regularized optimal transport plan, thus obtaining a Schrödinger bridge, the process is itself Markov. In such a case, we still get the advantage of expressing it as the solution to a stochastic differential equation.

Let us first recall the definition of Markov transition function and Markov process.

Definition 1.3 (Markov transition kernel)**.** Let $(E, \mathcal{E})$ be a measurable space. A **Markov transition function** on $(E, \mathcal{E})$ is a function $p(s, t, x, A)$ for $s, t \in [0, +\infty)$, $x \in E$ and $A \in \mathcal{E}$ such that the following hold true:

1. For all $s, t \in [0, +\infty)$, $A \in \mathcal{E}$ fixed, the function $x \mapsto p(s, t, x, A)$ is $\mathcal{E}$-measurable;
2. For all $s, t \in [0, +\infty)$, $x \in E$ fixed, the map $A \mapsto p(s, t, x, A)$ is a probability measure on $(E, \mathcal{E})$;

3. Chapman-Kolmogorov equation holds true for all $s \leq u \leq t$:

$$p(s,t,x,A) = \int_E p(u,t,y,A)p(s,u,x,dy) \tag{1.9}$$

Definition 1.4 (Markov Process)**.** Let p be a Markov transition function on the measurable space $(E, \mathcal{E})$, and let μ be a probability measure on it. An $(E, \mathcal{E})$-valued stochastic process $X = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq u}, (X_t)_{t \geq u}, \mathbb{P})$ is a **Markov process** associated to p and with initial distribution μ if:

1. $X_u \sim \mu$;
2. Markov property holds true: for any bounded and Borel $f : E \to \mathbb{R}$:

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \int_E f(x)p(s,t,X_s,dx) \tag{1.10}$$

   almost surely.

The definition of Markov projection is as follows

Definition 1.5. Let X be a stochastic process on a smooth manifold M. We say that a stochastic process ξ on M is the **Markovian projection** of X if for all $A \in \mathscr{B}(M)$, $\mathbb{P}(X_t \in A) = \mathbb{P}(\xi \in A)$ for all $t \in [0,T]$. Equivalently, we say $X_t = \xi_t$ in distribution.

We now turn our attention back to process X as defined in Section 1.2, and prove the following result

Theorem 1.2. *Under the above notations, define the following function (measurable by standard results in probability theory):*

$$\bar{b}(t,x) = \mathbb{E}[b(t,X_t,Z)|X_t = x]$$

Suppose moreover that $\bar{b}$ and g satisfy the following regularity conditions on $[0,\tilde{T}]$ for all $\tilde{T} < T$:

- *They are both measurable in (t,x);*
- *$\bar{b}$ has sublinear growth, i.e. there exists $M > 0$ such that for all $x \in \mathbb{R}^d$ and $t \in [0,\tilde{T}]$*

$$|\bar{b}(t,x)| \leq M(1 + |x|)$$

  and g is bounded in $[0,\tilde{T}]$;

- *$\bar{b}$ is locally or globally Lipschitz in space. The (weaker) local condition reads: for all $N > 0$ there exists $L_N > 0$ such that if $|x|, |y| \leq N$ and $t \in [0, \tilde{T}]$, then*

$$|\bar{b}(t, x) - \bar{b}(t, y)| \leq L_N |x - y|$$

and that X_0 is square integrable.

Then process ξ solving the following SDE problem is the Markov projection of X on $[0, \tilde{T}]$ for all $\tilde{T} < T$. Moreover, $\xi_t \xrightarrow[t \to T]{} X_T$ in distribution (equivalently, $\mathbb{P}^{\xi_t} \xrightarrow[t \to T]{} \mathbb{Q}_T$ weakly, being $X_T \sim \mathbb{Q}_T$):

$$\begin{cases} d\xi_t = \bar{b}(t, \xi_t)dt + g(t)dW_t \quad t \in [0, \tilde{T}] \\ \xi_0 \sim X_0 \end{cases} \tag{1.11}$$

Proof.* First, notice that under the assumptions of X_0 being square integrable and the required regularity conditions on the coefficients, there exists a unique strong and Markov solution to (1.11) (see Baldi [1]). This means that if we prove $\xi_t \overset{d}{=} X_t$ for all $t \in [0, \tilde{T}]$, ξ is actually the Markovian projection of X on such an interval.

In order to find a Markov projection for process X, we start by considering $\phi \in C_C^\infty([0, \tilde{T}] \times \mathbb{R}^d)$, and compute $\mathbb{E}[\varphi(t, X_t)]$. This way, we will obtain a weak partial differential equation for the marginals $\mathbb{P}^{X_t}$, which is the Fokker-Planck equation for the solution to the claimed SDE.

We first compute $\mathbb{E}[\phi(t, X_t)]$ by using conditional expectations:

$$\begin{aligned} \mathbb{E}[\varphi(t, X_t)] &= \mathbb{E}[\mathbb{E}[\varphi(t, X_t)|Z]] = \\ &= \int_{\mathcal{Q}} \mathbb{E}[\varphi(t, X_t)|Z = z]dq(z) = \\ &= \int_{\mathcal{Q}} \mathbb{E}[\varphi(t, Y_t^z)]dq(z) \end{aligned} \tag{1.12}$$

Now, since Y_t^z solves an Ito SDE, we can apply Ito formula (we drop the superscript z for ease of reading, since now z is fixed):

$$d\varphi(t, Y_t) = \left(\frac{\partial \varphi}{\partial t}(t, Y_t) + b(t, Y_t, z) \cdot \nabla\varphi(t, Y_t) + \frac{1}{2}\sigma(t) : \nabla^2\varphi(t, Y_t)\right)dt + \\ + \nabla\varphi(t, Y_t)^T g(t) dW_t \tag{1.13}$$

where $\sigma(t) = g(t)g(t)^T$ and $\nabla^2\varphi \in \mathbb{R}^{d\times d}$ such that $\nabla^2\varphi_{ij} = \frac{\partial\varphi}{\partial x^i \partial x^j}$ for $i, j = 1, 2, ..., d$. Being this true for all $t \in [0, \tilde{T}]$, we let $t = \tilde{T}$. Now, if we write Equation (1.13) in the integral form, since φ has compact support and $g(t)$ depends only on time:

$$\begin{aligned} 0 = \mathbb{E}[\varphi(\tilde{T}, Y_{\tilde{T}})] - \mathbb{E}[\varphi(0, Y_0)] = \mathbb{E}\left[\int_0^{\tilde{T}} \frac{\partial\varphi}{\partial t}(t, Y_t)dt\right] + \\ + \mathbb{E}\left[\int_0^{\tilde{T}} b(t, Y_t, z) \cdot \nabla\varphi(t, Y_t)dt\right] + \\ + \mathbb{E}\left[\int_0^{\tilde{T}} \frac{1}{2}\sigma(t) : \nabla^2\varphi(t, Y_t)dt\right] + 0 \end{aligned} \tag{1.14}$$

We now focus on the term involving $b(t, Y_t)$, and plug it into Equation (1.12):

$$\begin{aligned} \int_{\mathcal{Q}} \mathbb{E}\left[\int_0^{\tilde{T}} b(t, Y_t^z, z) \cdot \nabla\varphi(t, Y_t^z)dt\right] dq(z) = \int_0^{\tilde{T}} \int_{\mathcal{Q}} \mathbb{E}[b(t, Y_t^z, z) \cdot \nabla\varphi(t, Y_t^z)]dq(z)dt = \\ = \int_0^{\tilde{T}} \mathbb{E}[b(t, X_t, Z) \cdot \nabla\varphi(t, X_t)]dt \end{aligned} \tag{1.15}$$

where in the first equality we used Fubini's theorem ($|b(t, x, z)|$ is integrable on $\mathcal{Q} \times [0, \tilde{T}]$ being linear in $z \in \mathcal{Q}$ - being the drift of a Brownian bridge - and bounded in $t \in [0, \tilde{T}]$), and in the second one the tower property for conditional expectations.
But notice that:

$$\begin{aligned}\mathbb{E}[b(t,X_t,Z)\cdot\nabla\varphi(t,X_t)] &= \mathbb{E}[\mathbb{E}[b(t,X_t,Z)\cdot\nabla\varphi(t,X_t)|X_t]] = \\ &= \int_{\mathbb{R}^d}\mathbb{E}[b(t,X_t,Z)\cdot\nabla\varphi(t,X_t)|X_t=x]d\mathbb{P}^{X_t}(x) = \\ &= \int_{\mathbb{R}^d}\nabla\varphi(t,x)\cdot\mathbb{E}[b(t,X_t,Z)|X_t=x]d\mathbb{P}^{X_t}(x) = \\ &= \int_{\mathbb{R}^d}\nabla\varphi(t,x)\cdot\bar{b}(t,x)d\mathbb{P}^{X_t}(x)\end{aligned} \tag{1.16}$$

where we defined $\bar{b}(t,x)=\mathbb{E}[b(t,X_t,Z)|X_t=x]$. But then Equation (1.15) equals:

$$\int_0^{\tilde{T}}\int_{\mathbb{R}^d}\bar{b}(t,x)\cdot\nabla\varphi(t,x)d\mathbb{P}^{X_t}(x)dt \tag{1.17}$$

Reasoning the same way for all other terms in (1.14), we get:

$$\int_0^{\tilde{T}}\int_{\mathbb{R}^d}\left\{\frac{\partial}{\partial t}+\bar{b}(t,x)\cdot\nabla+\frac{1}{2}\sigma(t):\nabla^2\right\}\varphi(t,x)d\mathbb{P}^{X_t}(x)dt=0 \tag{1.18}$$

But being φ arbitrary, we have that this is just the weak (distributional) Fokker-Planck equation for $p(t,dx)=\mathbb{P}^{X_t}(dx)$:

$$\frac{\partial p_t}{\partial t}=-\mathrm{div}(\bar{b}(t,x)p_t)+\frac{1}{2}\nabla^2(\sigma(t)p_t) \tag{1.19}$$

for $t\in[0,\tilde{T}]$. Notice that here ∇^2 (applied to a matrix-valued field) has to be interpreted as the double divergence (i.e. $\nabla^2=\mathrm{div}(\mathrm{div})$). Now, under the assumed conditions for $\bar{b}$ and σ, once we choose an initial distribution $p(0,dx)$, Equation (1.19) uniquely determines the 1D marginals. Moreover, its solution $p(t,dx)$ is also the 1D marginals path of the solution to the following SDE:

$$\begin{cases} d\xi_t=\bar{b}(t,\xi_t)dt+g(t)dW_t & ,t\in[0,\tilde{T}] \\ \xi_0\sim\mathbb{P}^{X_0} \end{cases} \tag{1.20}$$

and, being $\tilde{T}<T$ arbitrary:

$$\begin{cases} d\xi_t=\bar{b}(t,\xi_t)dt+g(t)dW_t & ,t\in[0,T) \\ \xi_0\sim\mathbb{P}^{X_0} \end{cases} \tag{1.21}$$

Let us now show the claimed distributional limit. Consider a space test function φ: we want to show that $\int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}^{\xi_t}(x) = \int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}^{X_t}(x) \to \int_{\mathbb{R}^d} \varphi(z_1) d\hat{\mathbb{Q}}_T(x)$.

Notice that:

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}^{X_t}(x) &= \int_{\mathbb{R}^d} \varphi(x) \int_{\mathcal{Q}} d\mathbb{P}^{Y_t^z}(x) dq(z) = \\ &= \int_{\mathcal{Q}} \int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}^{Y_t^z}(x) dq(z)\end{aligned}$$

where we used the definition of X and Fubini-Tonelli theorem for exchanging integrals. Now, recall that Y_t^z is a Brownian bridge, for which it is known that $\mathbb{P}^{Y_t^z} \to \delta_{z_1}$ weakly as $t \to T$. Notice also that we can apply dominated convergence theorem to the previous integral, where the family of integrands is given by $\left(\int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}^{Y_t^z}(x) \right)_{t \in [0,T)}$, and $\int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}^{Y_t^z}(x) \to \varphi(z_1)$ as $t \to T$:

$$\left| \int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}^{Y_t^z}(x) \right| \leq \int_{\mathbb{R}^d} |\varphi(x)| d\mathbb{P}^{Y_t^z}(x) \leq \|\varphi\|_{L^\infty} \cdot 1$$

Then:

$$\int_{\mathcal{Q}} \int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}^{Y_t^z}(x) dq(z) \to \int_{\mathcal{Q}} \varphi(z_T) dq(z_0, z_T) = \int_{\mathbb{R}^d} \varphi(z_T) d\hat{\mathbb{Q}}_T(z_1)$$

which concludes the proof. □

Remark 1.1. *Notice that the first part of the above theorem holds true even when conditioned processes are general solutions to SDE problems, not necessarily Brownian bridges (provided $|b(t, x, z)|$ is integrable on $\mathcal{Q} \times [0, \tilde{T}]$). For our purposes, we are however interested in the latter case.*

Next theorem, instead, gives us a way to guarantee that the conditions on $\bar{b}$ in Theorem 1.2 are satisfied.

Theorem 1.3. *Under the usual notations, if Z is almost surely bounded and admits density f_Z with respect to the d-dimensional Lebesgue measure, then $\bar{b}(t, x)$ satisfies the assumptions in Theorem 1.2.*

Proof.* We will make use of the fact that a Brownian bridge from z_0 to z_1 in $\mathbb{R}^d$ solves:

$$
\begin{cases}
dY_t^z = \frac{z_1 - Y_t^z}{T-t} dt + dW_t \\
Y_0^z = z_0
\end{cases}
$$

so in particular:

$$
b(t, X_t, Z) = \frac{Z_1 - X_t}{T - t}
$$

We restrict to any interval $[0, \tilde{T}]$ with $\tilde{T} < T$. Notice that, if Z has a density with respect to the Lebesgue measure in $\mathbb{R}^d \times \mathbb{R}^d$, and since $X_t | Z = z$ is a Brownian bridge at time t, so admits a density with respect to the Lebsegue measure on $\mathbb{R}^d$, then X_t has a density too, and so do $Z|X_t = x$ (with respect to the Lebesgue measure on $\mathbb{R}^d \times \mathbb{R}^d$), $Z_0|X_t = x$ and $Z_1|X_t = x$ (with respect to the Lebesgue measure on $\mathbb{R}^d$). Indeed:

$$
\begin{aligned}
\mathbb{P}^{X_t}(A) &= \int_{\mathcal{Q}} \mathbb{P}^{Y_t^z}(A) f_Z(z) dz = \\
&= \int_{\mathcal{Q}} \int_A f_{Y_t^z}(x) f_Z(z) dx dz = \\
&= \int_A \left(\int_{\mathcal{Q}} f_{Y_t^z}(x) f_Z(z) dz \right) dx, \quad \forall A \in \mathscr{B}(\mathbb{R}^d)
\end{aligned}
$$

showing that X_t admits density; but also (X_t, Z) admits density, by:

$$
\mathbb{P}(X_t \in A, Z \in B) = \int_A \int_B f_{X_t|Z}(x|z) f_Z(z) dz dx
$$

from which we get $f_{Z|X_t}(z|x)$ by Bayes rule. Finally, if $Z|X_t = x$ has a density, then $Z_0|X_t = x$ and $Z_1|X_t = x$ must have it.

Since $t < \tilde{T}$, the time dependence does not give any trouble, since $T - t \geq T - \tilde{T}$. It is clear that $\mathbb{E}[b(t, X_t, Z)|X_t = x] = -\frac{x}{T-t} + \frac{\mathbb{E}[Z_1|X_t = x]}{T-t}$. Thus, everything amounts to understanding what the dependence of $\mathbb{E}[Z_1|X_t = x]$ is with respect to x.

We can compute such expectation by leveraging Bayes rule:

$$\begin{aligned}
\mathbb{E}[Z_1|X_t = x] &= \int_{\mathbb{R}^d} z_1 f_{Z_1|X_t}(z_1|x) dz_1 = \\
&= \int_{\mathbb{R}^d} z_1 \int_{\mathbb{R}^d} f_{Z|X_t}(z_0, z_1|x) dz_0 dz_1 = \\
&= \int_{\mathbb{R}^d} z_1 \int_{\mathbb{R}^d} \frac{f_{X_t|Z}(x|z) f_Z(z)}{\int_{\mathbb{R}^d \times \mathbb{R}^d} f_{X_t|Z}(x|\eta) f_Z(\eta) d\eta} dz_0 dz_1 = \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} z_1 \frac{f_{X_t|Z}(x|z) f_Z(z)}{\int_{\mathbb{R}^d \times \mathbb{R}^d} f_{X_t|Z}(x|\eta) f_Z(\eta) d\eta} dz = \\
&= \frac{1}{\int_{\mathbb{R}^d \times \mathbb{R}^d} f_{X_t|Z}(x|\eta) f_Z(\eta) d\eta} \int_{\mathbb{R}^d \times \mathbb{R}^d} z_1 f_{X_t|Z}(x|z) f_Z(z) dz
\end{aligned}$$

Now, recall that $X_t|Z = z$ is just a Brownian bridge from z_0 to z_1, so that $X_t|Z = z \sim \mathcal{N}(tz_1 + (1-t)z_0, g(t)^T g(t) t(1-t))$. Let us, for the sake of clarity and without loss of generality, assume that $g(t) = I$. Then we have that the above expression is:

$$\begin{aligned}
&\left(\int_{\mathbb{R}^d \times \mathbb{R}^d} f_Z(\eta) \frac{1}{(2\pi t(1-t))^{\frac{d}{2}}} \exp\left\{ -\frac{1}{2}(t(1-t))^{-2} |x - t\eta_1 + (1-t)\eta_0|^2 \right\} d\eta \right)^{-1} \cdot \\
&\int_{\mathbb{R}^d \times \mathbb{R}^d} z_1 f_Z(z) \frac{1}{(2\pi t(1-t))^{\frac{d}{2}}} \exp\left\{ -\frac{1}{2}(t(1-t))^{-2} |x - tz_1 + (1-t)z_0|^2 \right\} dz = \\
&= \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} f_Z(\eta) \exp\left\{ -\frac{1}{2}(t(1-t))^{-2} |x - t\eta_1 + (1-t)\eta_0|^2 \right\} d\eta \right)^{-1} \cdot \\
&\int_{\mathbb{R}^d \times \mathbb{R}^d} z_1 f_Z(z) \exp\left\{ -\frac{1}{2}(t(1-t))^{-2} |x - tz_1 + (1-t)z_0|^2 \right\} dz
\end{aligned}$$

If we could show that this has a dependence on x fulfilling the requirements of Theorem 1.2, we would be done.

In order to do these computations, let us define the following quantities:

$$\begin{aligned}
s(t) &= \frac{1}{2}(t(1-t))^{-2} \\
k(t, z) &= tz_1 - (1-t)z_0 \\
A(t, x, z) &= \exp\left\{ -s(t)|x - k(t, z)|^2 \right\} \\
D(t, x) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} f_Z(\eta) \exp\left\{ -s(t)|x - k(t, \eta)|^2 \right\} d\eta \\
w(t, x, z) &= \frac{A(t, x, z)}{D(t, x)}
\end{aligned}$$

We are then interested in knowing the dependency on x of $\int_{\mathbb{R}^d \times \mathbb{R}^d} z_1 w(t, x, z) f_Z(z) dz$.

Being Z almost surely bounded, f_Z has compact support. This way, linear growth bound is achieved through boundedness: $|\mathbb{E}[Z_1|X_t = x]| \leq \|Z_1\|_{L^\infty}$.

Then, let us consider the Lipschitz condition. Notice first that $w(t, x, z)$, as a function of x is continuous and infinitely differentiable. Moreover, fixed x, $D(t, x) > 0$, so we can, for a fixed x, compute the gradient of $\mathbb{E}[Z_1|X_t = x]$ as $\int_{\mathbb{R}^d\times\mathbb{R}^d} z_1 \nabla_x w(t, x, z) f_Z(z) dz$.

$$
\begin{aligned}
\nabla_x w(t,x,z) &= \frac{D\nabla A - A\nabla D}{D^2} = \\
&= \frac{-2s(t)(x-k(t,z))A(t,x,z)D(t,x) - A(t,x,z)\int_{\mathbb{R}^d\times\mathbb{R}^d} f_Z(\eta)(-2s(t))(x-k(t,\eta))A(t,x,\eta)d\eta}{D(t,x)^2} = \\
&= \frac{-2s(t)\Big(-k(t,z)A(t,x,z)D(t,x) + A(t,x,z)\int_{\mathbb{R}^d\times\mathbb{R}^d} f_Z(\eta)k(t,\eta)A(t,x,\eta)d\eta\Big)}{D(t,x)^2} = \\
&= 2s(t)w(t,x,z)\left(k(t,z) - \frac{\int_{\mathbb{R}^d\times\mathbb{R}^d} k(t,\eta)f_Z(\eta)A(t,x,\eta)d\eta}{D(t,x)}\right) = \\
&= 2s(t)w(t,x,z)k(t,z) - 2s(t)\frac{\int_{\mathbb{R}^d\times\mathbb{R}^d} k(t,\eta)f_Z(\eta)A(t,x,\eta)d\eta}{D(t,x)}
\end{aligned}
$$

Notice that, using the exact same computations as before, $\frac{\int k(t,\eta)f_Z(\eta)A(t,x,\eta)d\eta}{D(t,x)} = \mathbb{E}[k(t, Z)|X_t = x]$ which is bounded if Z is, being essentially a linear transformation of Z. But then the following quantity is bounded provided Z has bounded first moment (which is satisfied necessarily being Z bounded):

$$
\int_{\mathbb{R}^d\times\mathbb{R}^d} s(t) z_1 \frac{\int_{\mathbb{R}^d\times\mathbb{R}^d} k(t,\eta)f_Z(\eta)A(t,x,\eta)d\eta}{D(t,x)} f_Z(z_0, z_1) dz_1 dz_0 \tag{1.22}
$$

For the first term, once again leveraging boundedness of Z:

$$
\begin{aligned}
&\left| \int_{\mathbb{R}^d \times \mathbb{R}^d} s(t) z_1 k(t,z) w(t,x,z) f_Z(z) dz \right| \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \left| s(t) z_1 k(t,z) w(t,x,z) f_Z(z) \right| dz \leq \\
&\leq \|Z\|_{L^\infty(\mathbb{R}^d)} \|k(t,\cdot)\|_{L^\infty(\mathbb{R})} |s(t)| \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{A(t,x,z) f_Z(z)}{D(t,x)} dz = \\
&= \|Z\|_{L^\infty(\mathbb{R}^d)} \|k(t,\cdot)\|_{L^\infty(\mathbb{R})} |s(t)| \frac{\int_{\mathbb{R}^d \times \mathbb{R}^d} \exp\{-s(t)|x - k(t,z)|^2\} f_Z(z) dz}{\int_{\mathbb{R}^d \times \mathbb{R}^d} \exp\{-s(t)|x - k(t,\eta)|^2\} f_Z(\eta) d\eta} = \\
&= \|Z\|_{L^\infty(\mathbb{R}^d)} \|k(t,\cdot)\|_{L^\infty(\mathbb{R})} |s(t)| \cancel{\frac{\int_{\mathbb{R}^d \times \mathbb{R}^d} \exp\{-s(t)|x - k(t,z)|^2\} f_Z(z) dz}{\int_{\mathbb{R}^d \times \mathbb{R}^d} \exp\{-s(t)|x - k(t,\eta)|^2\} f_Z(\eta) d\eta}} = \\
&= \|Z\|_{L^\infty(\mathbb{R}^d)} \|k(t,\cdot)\|_{L^\infty(\mathbb{R})} |s(t)|
\end{aligned}
$$

We thus get that the gradient of $\bar{b}(t,x)$ is uniformly bounded in x for $t \in [0,\tilde{T}], \tilde{T} < T$, concluding Lipschitz continuity. □

1.4. Fitting SDE coefficients

In this section, we explain how to obtain the SDE coefficients from available data, as explained in Tong et al. [20].

By definition, the probability flow ODE associated to an SDE like (1.3) is:

$$dX_t = b^\circ(t, X_t, z) dt \tag{1.23}$$

where:

$$b^\circ(t,x,z) = b(t,x,z) - \frac{1}{2} g(t)^T g(t) \nabla_x \log f_{X_t|Z}(x|z) \tag{1.24}$$

Provided with a suitable initial distribution, the solution to Equation (1.23) will have the same one-dimensional distributions to the solution of Equation (1.3) (as explained, for example, in [20]).

We can now make the following definitions:

$$b^{\circ}(t,x) = \mathbb{E}\left[\frac{b^{\circ}(t,x,Z) f_{X_t|Z}(x|Z)}{f_{X_t}(x)}\right] \tag{1.25}$$

$$\nabla_x \log f_{X_t}(x) = \mathbb{E}\left[\frac{f_{X_t|Z}(x|Z)}{f_{X_t}(x)} \nabla_x \log f_{X_t|Z}(x|Z)\right] \tag{1.26}$$

$$\bar{b}(t,x) = b^{\circ}(t,x) + \frac{1}{2} g(t)^T g(t) \nabla_x \log f_{X_t}(x) \tag{1.27}$$

and, as shown in Tong et al. [20], $\bar{b}(t,x) = \mathbb{E}[b(t,X_t,Z)|X_t = x]$, ensuring that the solution to the SDE problem with $\bar{b}$ as drift and g as diffusion is the Markov projection of X.

Now, given definitions 1.25 and 1.26, it makes sense to define the following objectives for training two neural networks v_θ and s_θ:

$$\begin{aligned} \mathcal{L}_{[SF]^2M}(\theta) =& \mathbb{E}_{(t,z,x)\sim Q'}\left[\|v_\theta(t,x) - b_t^{\circ}(x|z)\|^2\right] + \\ &+ \mathbb{E}_{(t,z,x)\sim Q'}\left[\|s_\theta(t,x) - \nabla_x \log f_{X_t|Z}(x|z)\|^2\right] \end{aligned} \tag{1.28}$$

where $Q' = \mathcal{U}[0,1] \otimes Q \otimes f_{X_t|Z}(\cdot|z)$.

In Tong et al. [20], they actually prove correctness of this formulation proving its gradient is equivalent to the objective for directly learning $\bar{b}_t(x)$ and $\nabla_x \log f_{X_t}(x)$.

Notice that one could ideally fit a single neural network $v_\theta(t,x)$ by defining the objective:

$$\mathcal{L}_{[SF]^2M}(\theta) = \mathbb{E}_{(t,x,z)\sim Q'}\left[\|v_\theta(t,x) - b(t,x,z)\|^2\right]$$

directly learning the drift $\bar{b}(t,x)$. However, this is not convenient from a practical point of view, since such neural network would be constrained to a unique diffusion $g(t)$. Fitting a separate model for the flow and for the score gives the flexibility of changing at inference time the diffusion rate for the Brownian bridges.

2 | Geometry of the log-Cholesky Manifold

The goal of the present chapter is to study the geometry of two closely related smooth manifolds. The first one, which is mostly of our interest, is the one of SPD matrices, endowed with the *log-Cholesky* Riemaniann metric, introduced in [16]; the second one is a sort of *intermediate* manifold which is useful to describe the former, and is the one of lower triangular matrices with strictly positive diagonal entries (called *Cholesky space*), endowed with a Riemaniann metric $\tilde{g}(\cdot,\cdot)$.

The main result of this chapter is that there exists an isometric diffeomorphism (denoted with θ) between the SPD manifold with the log-Cholesky metric (so called *log-Cholesky manifold*) and $\mathbb{R}^d$ with its usual scalar product (or, equivalently, the space of lower triangular matrices with the Frobenius inner product). In particular, such a map is the composition of two isometries: one between the log-Cholesky manifold and the Cholesky manifold (that is well explored in Lin [16], and is briefly described in Section 2.3), and one between the Cholesky manifold and $\mathbb{R}^d$ (or the lower triangular matrices space), which we will better characterize in Section 2.2, since it is not discussed in these terms in the paper introducing the log-Cholesky metric.

Figure 2.1 provides an exhaustive picture of the relations between the manifolds of interest that we are going to describe. The schema is enough to understand the rest of the work without need to read the detailed derivation of such relationships.

As we will see in the rest of the work, having such a strong map between the log-Cholesky space and the Euclidean one allows us to essentially get stochastic objects that are typically complicated to describe in smooth manifolds (e.g. Brownian motion) as a *composition* of the same objects in the Euclidean setting and the isometry.

2.1. Notations and Preliminaries

We will always deal with n-dimensional matrices, and denote by $\mathcal{S}$ the space of symmetric matrices, $\mathcal{S}_+$ the one of symmetric positive definite matrices, by $\mathcal{L}$ the space of lower triangular matrices and by $\mathcal{L}_+$ the so called *Cholesky space* of lower triangular matrices with positive diagonal elements. If $X \in \mathcal{S}_+$, we can uniquely decompose it using Cholesky decomposition as $X = LL^T$, being $L \in \mathcal{L}_+$. On the other hand, if $L \in \mathcal{L}_+$, then $LL^T \in \mathcal{S}_+$, meaning that the following map is a bijection: $\mathscr{L} : \mathcal{S}_+ \to \mathcal{L}_+$. Actually we can say more: $\mathscr{L}$ is a diffeomorphism between these manifolds [16] (in particular, notice that $\mathcal{S}_+$ is a smooth submanifold of $\mathcal{S}$, which is identifiable with $\mathbb{R}^{\frac{n(n+1)}{2}}$, while $\mathcal{L}_+$ is a smooth submanifold of $\mathcal{L}$, which is also identifiable with $\mathbb{R}^{\frac{n(n+1)}{2}}$). In the following, we will denote with d the value $\frac{n(n+1)}{2}$, so that $\mathbb{R}^{\frac{n(n+1)}{2}}$ is $\mathbb{R}^d$. Also, we will denote Cholesky manifold by M and the one of SPD matrices by N.

In order to make these manifolds Riemaniann, we will endow them with a Riemaniann metric. On N, we put the so called *log-Cholesky metric* $g(\cdot,\cdot)$, and we will refer to the couple $(N, g(\cdot,\cdot))$ as *log-Cholesky space*. $\tilde{g}(\cdot,\cdot)$ is the Riemaniann metric we put on the *Cholesky space* M, whereas $\mathcal{L}$ is endowed with the usual Frobenius inner product $\langle\cdot,\cdot\rangle_F$. Notice that also $\mathbb{R}^d$ is a Riemaniann manifold with the usual euclidean scalar product $\langle\cdot,\cdot\rangle$.

When dealing with matrixes, we will adopt the two following notations, as used in Lin [16]. For a matrix B, we let $\mathbb{D}(B)$ denote its diagonal part (meaning, the matrix with same dimensions of B but equal to B only on the diagonal, null elsewhere), and $\lfloor B \rfloor$ be its strictly lower triangular part.

2.2. Cholesky Manifold

The aim of this section is to characterize the *Cholesky manifold* M, endowed with the Riemaniann metric $\tilde{g}(\cdot,\cdot)$. The main goal of it is to show there exists an isometric diffeomorphism between $(M, \tilde{g}(\cdot,\cdot))$ and $(\mathbb{R}^d, \langle\cdot,\cdot\rangle)$, which will be denoted by φ.

We begin by making explicit the smooth manifold structure of M in Section 2.2.1, and then give a convenient global chart for it that provides a bijection between M and $\mathbb{R}^d$ in Section 2.2.2. We then describe the tangent space to M in Section 2.2.3, and finally show that there exists a such an isometric diffeomorphism in Section 2.2.4.

2.2.1. Smooth Manifold Structure

We claimed M is a smooth submanifold of $\mathcal{L}$, which implies the latter must be a smooth manifold. As explained in Section A.1, in order for it to have the structure of a smooth manifold, we need to find a smooth structure on it. That is, a maximal smooth atlas on $\mathcal{L}$. Being $\mathcal{L}$ a finite-dimensional vector space, we can simply choose the standard smooth structure on it (see Example A.1). In particular, among the charts contained in this smooth structure, we have the most intuitive one. Specifically, let $L = [L_{ij}]_{i\geq j} \in \mathcal{L}$ be a lower triangular matrix: we can define a global chart by $(\mathcal{L}, \psi)$ with $\psi(L) = (L_{11}, L_{22}, ..., L_{nn}, L_{21}, ..., L_{n(n-1)})^T$. Most of the times, the vectorization of L provided by ψ will be written as $(L_1, L_2, ..., L_n, ..., L_d)^T$ with natural notation. Now that we have a smooth structure on $\mathcal{L}$, M is an open subset of it, inheriting its smooth structure. However, the very same global chart will be denoted by (M, η) in order not to make confusion. So $\eta(L)$ is a simple vectorization (with the above convention on indices) for $L \in M$.

2.2.2. Global Chart

One of the starting observations that is key in Lin [16] is that elements of M can be parametrized through a map $\bar{\phi} : M \to \mathcal{L}$ as $\bar{\phi}(L) = [N_{ij}]_{i\geq j}$ such that $N_{ij} = L_{ij}$ if $i \neq j$, else $N_{ii} = \log L_{ii}$.

But now we can define another function on M that is the simple composition of $\bar{\phi}$ and ψ, which we will call $\phi : M \to \mathbb{R}^d$, so that $\phi = \bar{\phi} \circ \psi$. Explicitely, this means $\phi(L) = (\log L_1, ..., \log L_n, L_{n+1}, ..., L_d)^T$.

Such function can be seen as a global chart for M, though it is not derived from any chart in the standard smooth structure on $\mathcal{L}$.

This other chart will be further inspected in Section 2.2.3, and in particular we will see how components of tangent vectors with respect to (the basis of the tangent space induced by) η transform when considering their components with respect to (the basis of the tangent space induced by) ϕ.

Also, we define $\hat{\phi} : \mathbb{R}^d \to \mathbb{R}^d$ by $\hat{\phi} = \psi \circ \bar{\phi} \circ \eta^{-1}$ (i.e. the translation of $\bar{\phi}$ at the level of $\mathbb{R}^d$).

In Figure 2.1 a schema summarizing the geometric relationships is provided.

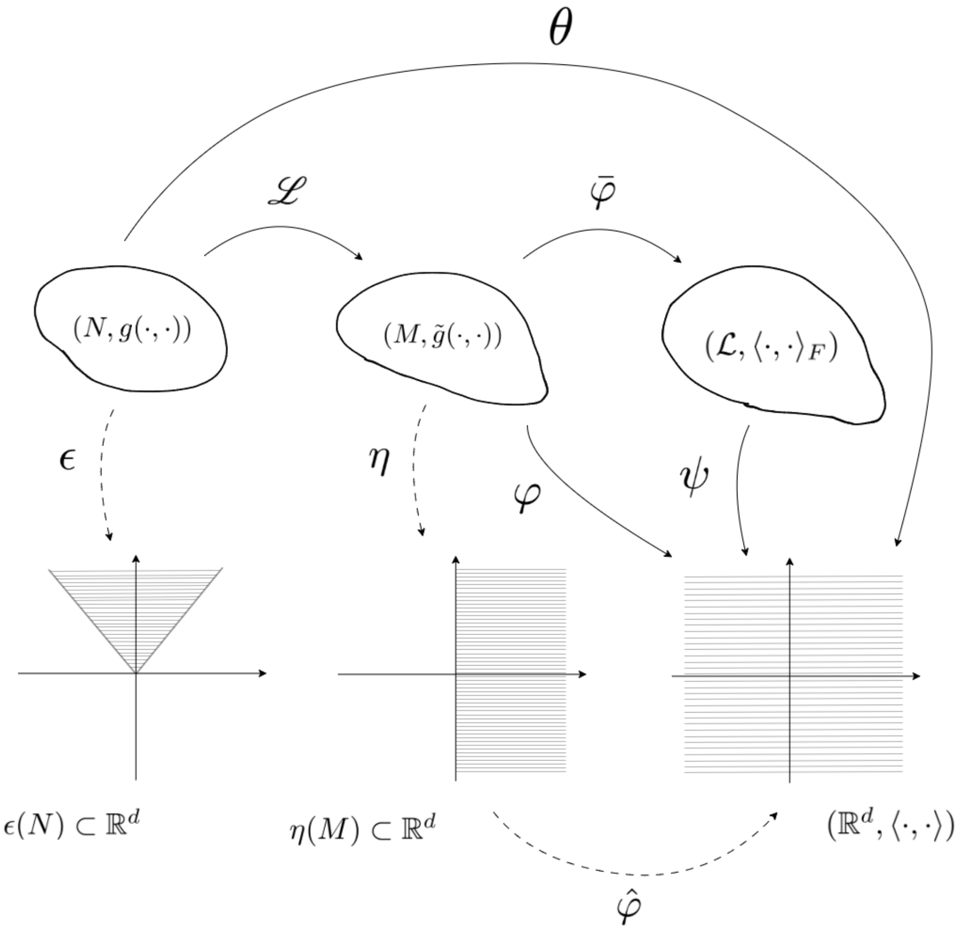

Figure 2.1: **Relations Between Manifolds.** Dashed arrows represent charts or maps between them; continuous arrows represent isometric diffeomorphisms. N is the manifold of SPD matrices, endowed with the log-Cholesky metric $g(\cdot,\cdot)$; M is the one of lower triangular matrices with positive diagonal entries, with the metric $\tilde{g}(\cdot,\cdot)$; $\mathcal{L}$ the one of lower triangular matrices, with the usual Frobenius norm $\langle\cdot,\cdot\rangle$.

2.2.3. Tangent Space

As stated in Proposition 3.13 in Lee [15], the tangent space to a vector space can be identified with the space itself, meaning we can identify the tangent space of $\mathcal{L}$ with $\mathcal{L}$ itself. Also, since by what explained in Section A.1 the tangent space to M, as a smooth submanifold of $\mathcal{L}$, can be identified with the tangent space to $\mathcal{L}$, it makes sense that T_pM is identified with $\mathcal{L}$ for all $p \in M$.

Here notice the following: when seeing M as a submanifold of $\mathcal{L}$, it inherits the standard smooth structure of $\mathcal{L}$, and with that structure we can say M is a smooth manifold and define its tangent spaces. At the same time, when treating M *per se*, we can consider the two charts η (which in the previous scenario is one of the inherited ones from $\mathcal{L}$) and ϕ: the two are smoothly compatible and so we can consider the smooth structure on M containing them. Since, by the first scenario, M with the smooth structure containing η has $\mathcal{L}$ as tangent spaces, we can at least intuitively deduce that M with the smooth structure containing ϕ (thus η too) has $\mathcal{L}$ as tangent spaces. This is not a concern, because the thing is just that ϕ is not a global chart for $\mathcal{L}$, and it is not even extendable to one (we take logarithms on the diagonal).

Now, we proceed to characterize the tangent space to M leveraging results obtained in Section A.1.3. In particular, we see what is the form of the basis induced by η for T_LM for some $L \in M$. We know indeed that a basis for T_LM is provided by elements of the kind $\frac{\partial}{\partial x^i}\Big|_L$, which for $f \in C^\infty(M)$ act as:

$$\begin{aligned}\frac{\partial}{\partial x^i}\Big|_L f &= \frac{\partial}{\partial x^i}\Big|_{\eta(L)} (f \circ \eta^{-1}) = \\ &= \frac{\partial}{\partial x^i}\Big|_{x=(L_1,L_2,...,L_d)^T} \hat{f} = \\ &= \frac{\partial \hat{f}}{\partial x^i}(L_1, L_2, ..., L_d)\end{aligned}$$

where $\hat{f} = f \circ \eta^{-1}$.

Then, if $V \in T_LM$, there exists a set of d components $(V_i)_{i=1}^d \subseteq \mathbb{R}$ such that:

$$Vf = V_i \frac{\partial \hat{f}}{\partial x^i}(L_1, L_2, ..., L_d)$$

for all $f \in C^\infty(M)$. We will denote these components, being them induced by η, V_i^η.

It is clear that one can take the elements V_i and put them in a matrix form adopting the above convention (i.e. the first n elements on the diagonal, the remaining ones filling orderedly the off-diagonal part), to obtain an element of $\mathcal{L}$ in general (indeed, there is no restriction on the sign of the components). This actually makes sense of the above intuition that the tangent space to M is $\mathcal{L}$.

We are now interested in knowing the components induced by ϕ. We do the exact same computations.

$$\begin{aligned}
\left.\frac{\partial}{\partial x^i}\right|_L f &= \left.\frac{\partial}{\partial x^i}\right|_{x=\phi(L)} (f \circ \phi^{-1}) = \\
&= \left.\frac{\partial}{\partial x^i}\right|_{x=\phi(L)} (f \circ \eta^{-1} \circ \hat{\phi}^{-1}) = \\
&= \left.\frac{\partial}{\partial x^i}\right|_{x=\phi(L)} (\hat{f}(\hat{\phi}^{-1})) = \\
&= \left.\frac{\partial \hat{f}}{\partial x^i}(e^{x^1}, e^{x^2}, ..., e^{x^n}, x^{n+1}, ..., x^d)\right|_{x=\phi(L)} = \\
&= \left.\frac{\partial \hat{f}}{\partial y^j}\frac{d\hat{\phi}_j^{-1}(x)}{dx^i}\right|_{x=\phi(L)}
\end{aligned}$$

but

$$\frac{d\hat{\phi}_j^{-1}(x)}{dx^i} = \begin{cases} e^{x^i} \text{ if } i = j \leq n \\ e^{x^i} \text{ if } i = j > n \\ 0 \text{ if } i \neq j \end{cases}$$

Now, since $x = \phi(L)$ means $x = (\log L_1, \log L_2, ..., \log L_n, L_{n+1}, ..., L_d)^T$, we finally get that the new element of the basis are:

$$\left.\frac{\partial}{\partial x^i}\right|_L f = \begin{cases} L_i \frac{\partial \hat{f}}{\partial x^i}(L_1, ..., L_d) \text{ if } i \leq n \\ \frac{\partial \hat{f}}{\partial x^i}(L_1, ..., L_d) \text{ if } i > n \end{cases}$$

Now, calling V_i^ϕ the i^{th} element of the basis induced by ϕ, we have that if $V \in T_L M$:

$$\begin{aligned} Vf &= \sum_{i=1}^{n} V_i^{\phi} L_i \frac{\partial \hat{f}}{\partial x^i}(L_1, ..., L_d) + \sum_{i=n+1}^{d} V_i^{\phi} \frac{\partial \hat{f}}{\partial x^i}(L_1, ..., L_d) = \\ &= \sum_{i=1}^{n} V_i^{\eta} \frac{\partial \hat{f}}{\partial x^i}(L_1, ..., L_d) + \sum_{i=n+1}^{d} V_i^{\eta} \frac{\partial \hat{f}}{\partial x^i}(L_1, ..., L_d) \end{aligned}$$

from which we get the transformation between components:

$$\begin{cases} V_i^{\eta} = V_i^{\phi} L_i \ \text{ if } i \leq n \\ V_i^{\eta} = V_i^{\phi} \ \text{ if } i > n \end{cases} \tag{2.1}$$

or, equivalently, being $L_i > 0$ for $i \leq n$:

$$\begin{cases} V_i^{\phi} = \frac{1}{L_i} V_i^{\eta} \ \text{ if } i \leq n \\ V_i^{\phi} = V_i^{\eta} \ \text{ if } i > n \end{cases} \tag{2.2}$$

If we write elements in $T_L M$ as lower triangular matrices, we can rewrite these transformations more compatly as explained in the following

Proposition 2.1. *Let $B \in \mathcal{L}$ be an element of $T_L M$ with respect to the chart ϕ, and let $\hat{B} \in \mathcal{L}$ the same tangent vector written in terms of the chart η. Then:*

$$\hat{B} = \mathbb{D}(L)\mathbb{D}(B) + \lfloor B \rfloor \tag{2.3}$$

2.2.4. Isometric Diffeomorphism from M to $\mathcal{L}$ (and $\mathbb{R}^d$)

We already saw what the map $\bar{\phi} : M \to \mathcal{L}$ is. It is actually a diffeomorphism bewteen smooth manifolds (as stated in Proposition 9 of Lin [16]). Also, we obtained in Section 2.2.3 the transformation of the components of a tangent vector $V \in T_L M$ from a basis induced by $\phi = \bar{\phi} \circ \psi$ to the one induced by η (which is ψ but on M). We will leverage this to prove that $\bar{\varphi}$, as well as φ are isometric diffeomorphisms.

In particular, we consider M with its global chart η (simply vectorizing the entries of elements in M), and $\mathcal{L}$ with its chart ψ (simply vectorizing the entries of elements in $\mathcal{L}$). Moreover, $\bar{\phi}$ is a diffeomorphism between these two smooth manifolds. Naturally, such diffeomorphism, by what we said in Section A.1.3, has a differential mapping tangent spaces. Leveraging the results obtained in Section 2.2.3, we see that if $V \in T_L M$, $d\bar{\phi}_L(V)$ has components with respect to the basis of $T_{\bar{\phi}(L)}\mathcal{L}$ induced by ψ that are:

$$(d\bar{\phi}_L(V))_i = \begin{cases} \frac{1}{L_i} V_i^\eta \ \text{ if } i \leq n \\ V_i^\eta \ \text{ if } i > n \end{cases}$$

Now, if we endow $\mathcal{L}$ with the standard Riemaniann structure having $\langle \cdot, \cdot \rangle_F$ (Frobenius inner product) as Riemaniann metric, the induced Riemaniann structure on M is precisely the log-Cholesky one. Indeed, by definition, if $X, Y \in T_L M$

$$\begin{aligned} \tilde{g}(X,Y) &= \sum_{i \leq n} \frac{1}{L_i^2} X_i Y_i + \sum_{i > n} X_i Y_i = \\ &= \sum_{i \leq n} (\frac{1}{L_i} X_i)(\frac{1}{L_i} Y_i) + \sum_{i > n} X_i Y_i = \\ &= \langle d\bar{\phi}_L(X), d\bar{\phi}_L(Y) \rangle_F \end{aligned}$$

So we can interpret the action of putting log-Cholesky metric on M as the one of transforming it through a diffeomorphism $\bar{\phi}$ to $\mathcal{L}$, using the usual Frobenius metric on $\mathcal{L}$ and mapping back the Riemaniann structure using $\bar{\phi}^{-1}$. By the very definition of g, we get that the map $\bar{\phi} : (M, \tilde{g}(\cdot, \cdot)) \to (\mathcal{L}, \langle \cdot, \cdot \rangle_F)$ is an isometry between smooth Riemaniann manifolds, thus all geometric properties from one can be easily transported to the other.

In the very same way, one obtains that $\phi : (M, \tilde{g}(\cdot, \cdot)) \to (\mathbb{R}^d, \langle \cdot, \cdot \rangle)$ is an isometry, where $\langle \cdot, \cdot \rangle$ is the usual Euclidean scalar product. Notice indeed that if $V \in T_L M$, then $d\varphi_L(V)$ is an element of $T_{\varphi(L)}\mathbb{R}^d$ consisting of the vectorization of $d\bar{\varphi}_L(V)$, and that the Frobenius scalar product of two elements X, Y tangent to $\mathcal{L}$ is just the Euclidean scalar product of their vectorized forms.

As a by-product of this, and a confirmation that this approach to describe the Cholesky manifold is actually consistent with the one provided by Lin [16], we re-derive some of the results present in that work in a slightly different way leveraging this isometry. These can be found in Section A.2.

2.3. Log-Cholesky Manifold

Now that we have characterized the Cholesky space with $\tilde{g}$, it is easy to describe geometrically the manifold of SPD matrices, endowed with the log-Cholesky metric $g(\cdot, \cdot)$.

Indeed, the map $\theta : N \to \mathbb{R}^d$ defined by $\theta = \phi \circ \mathscr{L}$ can be seen both as a global chart for N and as a diffeomorphism. By what explained in Lin [16], $\mathscr{L} : (N, g(\cdot,\cdot)) \to (M, \tilde{g}(\cdot,\cdot))$ is an isometry, thus $\theta : (N, g(\cdot,\cdot)) \to (\mathbb{R}^d, \langle\cdot,\cdot\rangle)$ is an isometric diffeomorphism (being the composition of isometric diffeomorphisms).

Let $\epsilon : N \to \mathbb{R}^d$ denote the global chart of N obtained by vectorizing its lower triangular part. As done for the Cholesky space, we might be interested in knowing how tangent vectors transform if expressed in terms of one or the other chart.

First, we will need a result from Lin [16]

Proposition 2.2. *Let $S \in N$, $\mathscr{L}(S) = L$ and let $w \in T_S N$ with $W \in \mathcal{L}$ its expression as a lower triangular matrix in terms of the global chart ϵ. Then the differential of $\mathscr{L}$ at S, evaluated at w, is given by:*

$$d\mathscr{L}_S(W) = L(L^{-1}WL^{-T})_{\frac{1}{2}} \tag{2.4}$$

where $X_{\frac{1}{2}} = \lfloor X \rfloor + \mathbb{D}(X)/2$.

Denoting by $\mathscr{S} : M \to N$ the inverse, we have that if $z \in T_L M$ with $Z \in \mathcal{L}$ its expression as a lower triangular matrix in terms of the global chart η, then the differential of $\mathscr{S}$ at L evaluated at z is given by:

$$d\mathscr{S}_L(Z) = LZ^T + ZL^T \tag{2.5}$$

By simply putting together Equation (2.3) and Proposition 2.2, we obtain the following:

Proposition 2.3. *Let $B \in \mathcal{L}$ be an element of $T_S N$ with respect to the chart θ, and let $\hat{B} \in \mathcal{L}$ the same tangent vector written in terms of the chart ϵ. Then:*

$$B = \mathbb{D}(L)^{-1}\mathbb{D}\Big(L(L^{-1}\hat{B}L^{-T})_{\frac{1}{2}}\Big) + \lfloor L(L^{-1}\hat{B}L^{-T})_{\frac{1}{2}} \rfloor \tag{2.6}$$

where $S_{\frac{1}{2}} = \lfloor S \rfloor + \frac{1}{2}\mathbb{D}(S)$.

Whereas:

$$\hat{B} = \mathbb{D}(L)\mathbb{D}(LB^T + BL^T) + \lfloor LB^T + BL^T \rfloor \tag{2.7}$$

3 | Stochastic Analysis on Manifolds

This chapter is dedicated to the study of stochastic analysis concepts defined over certain smooth manifolds. We first fix some notations in Section 3.1.

The rest of the chapter is divided into two parts. Section 3.2 provides an introduction to stochastic analysis on smooth (Riemaniann) manifolds, especially focusing on the notion of stochastic differential equation, Brownian Motion (BM) and Brownian bridge (BB) on them.

In Section 3.3 we focus our attention on those Riemaniann manifolds that satisfy similar properties to the SPD manifold endowed with the *log-Cholesky* Riemaniann metric (see Chapter 2). In particular, we will assume the existence of an isometric diffeomorphism between the manifold and a Euclidean space. This will essentially allow us to study objects defined on the manifold through their Euclidean counterparts. Though all results from this part are stated and obtained in a general setting, unrelated to the log-Cholesky manifold, one should notice that it relies on quite strict assumptions that are typically not fullfilled on general Riemaniann manifolds.

In this chapter, many differential geometry and stochastic differential geometry concepts are given for granted. However, Section A.1 provides a sufficiently extended overview of the main differential geometry concepts, whereas in Chapter B the basics of stochastic differential geometry are explained.

3.1. Notations and Preliminaries

We will adopt the following notations, unless otherwise specified. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, with a filtration $(\mathcal{F}_t)_{t\in[0,+\infty)}$, which will be denoted by $\mathcal{F}_*$. The filtered probability space is sometimes written as $(\Omega, \mathcal{F}_*, \mathbb{P})$.

With M we denote a general smooth manifold (not necessarily Riemaniann), and with $W(M)$ all the continuous paths with values in M with their corresponding explosion time. The space $W(M)$ can be endowed with its Borel σ-algebra $\mathscr{B}(W(M))$, and the Borel filtration $(\mathscr{B}(W(M))_s)_s$: for brevity, we let $\mathscr{B}_s = \mathscr{B}(W(M))_s$. Notice that in some cases M will be substituted with $\mathbb{R}^d$, whenever we mean the Euclidean space as a Riemaniann manifold, endowed with its usual scalar product. For an element $x \in M$, we let T_xM be the tangent space to M at x, TM its tangent bundle and $\Gamma(TM)$ the space of all vector fields over M. The frame bundle for M is $\mathscr{F}(M)$, whereas the orthonormal frame bundle is $\mathscr{O}(M)$ (see Section A.1.4 for an introduction to the concept of frame bundle). We will also denote with π the canonical projection of the frame bundle onto M.

From Section 3.2.3 on, we assume $(M, g(\cdot,\cdot))$ is a Riemaniann manifold, being g its Riemaniann metric, endowed with the Levi-Civita connection ∇ and the Laplace-Beltrami operator Δ_M. For two vector fields V and W over M, $\nabla_V W$ is the derivation (according to the Levi-Civita connection) of W with respect to V.

In Section 3.3, we will assume there exists an isometric diffeomorphism from M to $\mathbb{R}^d$, which we will call θ. In particular, $\theta : M \to \mathbb{R}^d$ is a diffeomorphism such that for all $V, W \in T_SM$ ($S \in M$), it holds $g(V, W) = \langle d\theta_S V, d\theta_S W\rangle$, being $d\theta_S : T_SM \to T_{\theta(x)}\mathbb{R}^d$ the differential of θ at S, and denoting with $\langle\cdot,\cdot\rangle$ the usual Euclidean inner product.

Since θ can be seen as a global chart for M, we will denote for a general $S \in M$ its coordinates with respect to such a global chart by $x(S) = \{x^i(S)\}_{i=1}^d$, often omitting the S dependency. With this convention, we let $X_i = \frac{\partial}{\partial x^i}$ be the fundamental vector field over M, $i = 1, 2, ..., d$.

3.2. Introductive Elements

In this section we give some background material that is needed to study stochastic processes on smooth manifolds. We are in particular interested in defining the concept of stochastic differential equation on a smooth manifold and studying Brownian motion and Brownian bridges on a Riemaniann manifold.

Much of the material for this part has been taken from Hsu [13].

In Section 3.2.1, we recall some basic concepts in stochastic analysis on simple euclidean spaces. We define the concept of semi-martingale, stochastic differential equations and provide an existence and uniqueness result for those. Moreover, we clearify the difference between Ito and Stratonovich form for expressing stochastic differential equations. The latter is indeed more suited to introduce this concept over smooth manifolds.

After this, in Section 3.2.2 we actually introduce SDEs on manifolds, whereas Section 3.2.3 and Section 3.2.4 are about Brownian motion and Brownian bridges on Riemaniann manifolds.

3.2.1. Stochastic Differential Equations on Euclidean Space

Though it is a more general setting than what we need, here we will assume the processes are driven by l-valued semimartingales.

Definition 3.1. Consider the filtered probability space $(\Omega, \mathcal{F}_*, \mathbb{P})$. Z is an $\mathbb{R}^l$-valued, $\mathcal{F}_*$-adapted semimartingale if $Z = M + A$, being M a local $\mathcal{F}_*$-martingale and A an $\mathcal{F}_*$-adapted process with locally bounded first variation such that $A_0 = 0$.

Typically, one deals with $Z_t = (W_t, t)$, where W_t is an $m = (l-1)$-dimensional Brownian motion.

Now, let $\hat{X}_0$ be an $\mathcal{F}_0$-measurable variable ($\mathcal{F}_0$-random variable) with values on $\mathbb{R}^d$, and let τ be an $\mathcal{F}_*$-stopping time. Also, let $\sigma : \mathbb{R}^d \to \mathrm{M}(d, l)$ (matrices with d rows and l columns). We denote the stochastic differential equation on $\mathbb{R}^d$ with $SDE(\sigma, Z, X_0)$ such that:

$$\begin{cases} dX_t = \sigma(X_t) dZ_t \\ X_0 = \hat{X}_0 \end{cases} \tag{3.1}$$

In the more typical case when $Z_t = (t, W_t)$ it becomes the usual:

$$\begin{cases} dX_t = b(X_t)dt + V_\alpha(X_t) dW_t^\alpha \\ X_0 = \hat{X}_0 \end{cases} \tag{3.2}$$

where we used Einstein notation on $\alpha = 1, 2, ..., m$, and b is the last column of σ, while V are the first $m = l - 1$ ones.

The following theorem is a cornerstone for SDEs on Euclidean spaces:

Theorem 3.1. *Suppose σ is globally Lipschitz and X_0 square integrable. Then $SDE(\sigma, Z, X_0)$ has a unique solution X_t.*

On the other hand, if we only assumed σ locally Lipschitz, we would have the same result but up to an explosion time $e(X)$.

When working with SDEs on manifolds, it is typically easier to leverage Stratonovich formulation of SDEs, since formulations have a more familiar form than standard calculus (Ito formula for Stratonovich formulation is just the standard chain rule).

Stratonovich SDE will be denoted as follows:

$$dX_t = \sigma_\alpha(X_t) \circ dZ_t^\alpha \tag{3.3}$$

or, in the integral formulation:

$$X_t = X_0 + \int_0^t \sigma_\alpha(X_s) \circ dZ_s^\alpha \tag{3.4}$$

This is equivalent to Ito SDE:

$$X_t = X_0 + \int_0^t \sigma_\alpha(X_s) dZ_s^\alpha + \frac{1}{2}\int_0^t \nabla_{\sigma_\beta}\sigma_\alpha(X_s) d\langle Z^\alpha, Z^\beta\rangle_s \tag{3.5}$$

Now, if X_t solves Equation (3.4) and $f \in C^2(\mathbb{R}^d)$, we have that (Ito formula):

$$f(X_t) = f(X_0) + \int_0^t \sigma_\alpha f(X_s) \circ dZ_s^\alpha \tag{3.6}$$

where $\sigma_\alpha f$ is the application of σ_α (as a vector field) to f (i.e. $\sigma_\alpha \cdot \nabla_x f$).

3.2.2. Stochastic Differential Equations on Manifolds

We now give the definition of an SDE on a manifold M.

Definition 3.2. Let σ_α, with $\alpha = 1, 2, ..., l$ be l vector fields on M (so $\sigma_\alpha(X_t) \in T_{X_t}M$), X_0 an $\mathcal{F}_0$-random variable on M and Z an $\mathbb{R}^l$-valued $\mathcal{F}_*$-semimartingale. An M-valued semimartingale solves the $SDE(\sigma_1, \sigma_2, .., \sigma_l; Z; X_0)$ denoted by $dX_t = \sigma_\alpha(X_s) \circ dZ_s^\alpha$ up to an $\mathcal{F}_*$-stopping time τ if:

$$f(X_t) = f(X_0) + \int_0^t \sigma_\alpha f(X_s) \circ dZ_s^\alpha, \;\; 0 \leq t < \tau \tag{3.7}$$

for all $\mathbb{R}$-valued functions $f \in C^\infty(M)$.

The following proposition will be useful in our case.

Proposition 3.1. *Let M be a closed submanifold of $\mathbb{R}^d$, and let $f^1, f^2, ..., f^d$ denote the coordinate functions. Then X solves $SDE(\sigma_1, \sigma_2, ..., \sigma_l; Z; X_0)$ up to a stopping time τ if and only if for all $i = 1, 2, ..., d$:*

$$f^i(X_t) = f^i(X_0) + \int_0^t \sigma_\alpha f^i(X_s) \circ dZ_s^\alpha \tag{3.8}$$

Proof. One implication is trivial, being coordinate functions $C^\infty(M)$, using the definition of solution to $SDE(\sigma_1, \sigma_2, ..., \sigma_l; Z; X_0)$. Now, suppose that Equation (3.8) holds. Being M closed, for any $f \in C^\infty(M)$ we can find an extension $\tilde{f} \in C^\infty(\mathbb{R}^d)$ of f. Then $f(X_t) = \tilde{f}(f^1(X_t), f^2(X_t), ..., f^d(X_t))$. Applying Ito formula:

$$\begin{aligned} df(X_t) &= \tilde{f}_{x_i}(f^1(X_t), f^2(X_t), ..., f^d(X_t)) \circ df^i(X_t) = \\ &= \tilde{f}_{x_i}(f^1(X_t), f^2(X_t), ..., f^d(X_t))\sigma_\alpha f^i(X_t) \circ dZ_t^\alpha = \\ &= \sigma_\alpha f(X_t) \circ dZ_t^\alpha \end{aligned}$$

□

Remark 3.1. *Though in Hsu [13] it is required for M to be a differentiable manifold without boundary, so that one can leverage Whitney's embedding theorem to regard M as a closed submanifold of some $\mathbb{R}^d$, one can take a different path under some assumptions. Indeed, suppose there exists a diffeomorphism $\theta : M \to \mathbb{R}^d$, so that M can essentially be regarded as $\mathbb{R}^d$ itself through θ. Then $\theta(M) = \mathbb{R}^d$, and $\mathbb{R}^d$ is a closed submanifold of itself.*

We could also use another argument for the result to work under the diffeomorphism assumption. Considering the proof, we can establish its result as long as we are able to find an extension $\tilde{f}$ to a function f (no matter M to be closed or not). If in particular we have $\theta : M \to \mathbb{R}^d$ as above, given $f \in C^\infty(M)$ we can always extend it as:

$$\tilde{f}(x^1, x^2, ..., x^d) = f(\theta^{-1}(x^1, x^2, ..., x^d))$$

and get the same conclusion.

An important question is of course the one of solvability of $SDE(\sigma_1, \sigma_2, ..., \sigma_d; Z; X_0)$.

Supposing M can be regarded as a closed submanifold of $\mathbb{R}^d$, each vector field σ_α is extendable to a vector field $\tilde{\sigma}_\alpha$ on $\mathbb{R}^d$ (i.e. an element of $\mathbb{R}^d$ itself). One can then consider the extended SDE on the ambient space $\mathbb{R}^d$:

$$X_t = X_0 + \int_0^t \tilde{\sigma}_\alpha(X_s) \circ dZ_s^\alpha \tag{3.9}$$

and this, by standard SDE theory on Euclidean spaces, has a unique solution up to its explosion time $e(X)$.

Now, next result tells us that if we start on M with X_0, we are guaranteed to stay on M until explosion time (provided M is a closed submanifold of $\mathbb{R}^d$). The fact that X_t stays on M makes sense, being the coefficients of $SDE(\sigma_1, \sigma_2, ..., \sigma_d; Z; X_0)$ tangent to the manifold at every point: this is exactly what happens for standard ODEs (if x_t is a continuous curve, $dx_t = \dot{x}_t dt$, being the velocity $\dot{x}_t$ tangent to x_t).

Proposition 3.2. *If X solves Equation (3.9) up to its explosion time $e(X)$ and $X_0 \in M$, having the possibility of regarding M as a closed submanifold of $\mathbb{R}^d$, then $X_t \in M$ for all $0 \leq t < e(X)$.*

The following is an important result which is another justification for using Stratonovich formalism when studying SDEs on manifolds. In a nutshell, it tells us Stratonovich SDEs transform consistently under diffeomorphisms between manifolds.

A diffeomorphism $\Phi : M \to N$ between two manifolds induces a map $\Phi_* : \Gamma(TM) \to \Gamma(TN)$ by:

$$(\Phi_* V)f(y) = V(f \circ \Phi)(x), \;\; y = \Phi(x), \;\; f \in C^\infty(N)$$

Proposition 3.3. *Let $\Phi : M \to N$ be a diffeomorphism and X solution to $SDE(\sigma_1, \sigma_2, ..., \sigma_l; Z; X_0)$. Then $\Phi(X)$ solves $SDE(\Phi_*\sigma_1, \Phi_*\sigma_2, ..., \Phi_*\sigma_l; Z; \Phi(X_0))$ on N.*

3.2.3. Brownian Motion on Manifolds

It is known that given a probability measure μ on M, there exists a unique $\frac{\Delta_M}{2}$-diffusion measure starting at μ on $(W(M), \mathscr{B}_*)$, denoted by $\mathbb{P}_\mu$. Any measure of this kind is called a Wiener measure on $W(M)$. Roughly speaking, a Brownian motion (BM) on M is an M-valued stochastic process X which law is a Wiener measure. The formal definition is provided in the following:

Proposition 3.4. *Let $X : \Omega \to W(M)$ be a measurable map defined on a probabilty space $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mu = \mathbb{P} \circ X_0^{-1}$ be its initial distribution. The following three statements are equivalent:*

1. *X is a $\frac{\Delta_M}{2}$-diffusion measure, i.e. a solution to the martingale problem for $\frac{\Delta_M}{2}$, meaning:*
$$M^f(X) = f(X_t) - f(X_0) - \frac{1}{2}\int_0^t \Delta_M f(X_s)ds$$
is an $\mathscr{F}_*^X$-local martingale for all $f \in C^\infty(M)$ and $0 \leq t < e(X)$;
2. *The law $\mathbb{P}^X = \mathbb{P} \circ X^{-1}$ is a Wiener measure on $(W(M), \mathscr{B}(W(M)))$;*
3. *X is a $\mathscr{F}_*^X$-semimartingale on M whose anti-development is a standard euclidean Brownian motion.*

Any X satisfying the above is a ***Brownian motion on*** *M.*

In the particular case where the Riemaniann manifold $(M, \langle\cdot,\cdot\rangle)$ ($\langle\cdot,\cdot\rangle$ is the Riemaniann metric) is a submanifold of $\mathbb{R}^d$ with the induced Riemaniann metric, then there is a very easy way to obtain a BM on M as solution to an SDE:

Proposition 3.5. *Let $(M, \langle\cdot,\cdot\rangle)$ be a submanifold of $\mathbb{R}^N$ with the induced Riemaniann metric. Let $P_\alpha(x) : \mathbb{R}^N \to T_xM$ be the orthogonal projection of e_α (α^{th} element of the canonical basis) onto T_xM, seen as a subspace of $\mathbb{R}^N$, for $\alpha = 1, 2, ..., N$. Also, let W_t denote the N-dimensional Brownian motion. Then X solving the following SDE is a Brownian motion on M:*

$$dX_t = P_\alpha(X_t) \circ dW_t^\alpha \tag{3.10}$$

Next, we define the concept of transition density function for a Brownian motion.

Let $\mathbb{P}_x$ denote the Wiener measure starting at $x \in M$ on $(W(M), \mathscr{B}(W(M))_*)$. The coordinate process on this path space is a RV $X : (W(M), \mathscr{B}(W(M))_*, \mathbb{P}_x) \to (W(M), \mathscr{B}(W(M))_*)$ defined by $X_t(\omega) = \omega_t$ for $\omega \in W(M)$.

Definition 3.3. Let X be the coordinate process on $W(M)$ with the usual Borel filtration and under $\mathbb{P}_x$. The **transition density function** of BM on M is a function p_M : $(0, +\infty) \times M \times M \to \mathbb{R}$ such that:

$$\mathbb{P}_x(X_t \in C) = \int_C p_M(t, x, y)dM(y) \tag{3.11}$$

for all $C \in \mathscr{B}(M)$ and $t < e$. Here $dM(y)$ is the volume measure on M.

It is known that if $M = \mathbb{R}^d$:

$$p_M(t, x, y) = \left(\frac{1}{2\pi t}\right)^{\frac{d}{2}} \exp\left\{ -\frac{\|x-y\|^2}{2t} \right\} \tag{3.12}$$

3.2.4. Brownian Bridge on Manifolds

Roughly speaking, a Brownian bridge (denoted as BB from now) is just a process obtained by conditioning BM to have some value at the terminal time T. The *bridge space* is:

$$L_{x,y;T}(M) = \{\omega \in W(M) s.t. \omega_0 = x \in M; \omega_T = y \in M\}$$

so that the law of a BB is just a Wiener measure starting at x, restricted to $L_{x,y;T}$. In this section, we consider the measurable space $(W(M), \mathscr{B}(W(M)))$ with the usual filtration $\mathscr{B}_s$ for all $s \in [0, T]$. Moreover, we define $X : (W(M), \mathscr{B}(W), \mathbb{P}_x) \to (W(M), \mathscr{B}(W))$ to be the coordinate process on $W(M)$ under $\mathbb{P}_x$. Denoting by $\mathbb{P}_{x,y;T}$ the BB measure on $W(M)$, we would like it to be something like:

$$\mathbb{P}_{x,y;T}(C) = \mathbb{P}_x(C | X_T = y) \quad \forall C \in \mathscr{B}(W)$$

To proceed more formally, we will need to pass through conditional expectation. Recall that if $K : (W(M), \mathscr{B}(W), \mathbb{P}_x) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is a R.V., $\mathbb{E}[K|X_T] : (W(M), \mathscr{B}(W), \mathbb{P}_x) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is the unique (up to negligible sets) $\sigma(X_T)$-measurable function (call it Z) such that for any $f : (M, \mathscr{B}(M)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ function:

$$\mathbb{E}[f(X_T)Z] = \mathbb{E}[f(X_T)K] \tag{3.13}$$

Moreover, there exists $g : (M, \mathscr{B}(M)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\mathbb{E}[K|X_T] = g(X_T)$. Then, we will let $\mathbb{E}_{x,y;T}[K] = g(y)$. Now, the idea to define $\mathbb{P}_{x,y;T}$ is to treat $\mathbb{E}_{x,y;T}$ as if it was the expectation of the coordinate function (i.e. the identity function on $W(M)$) under $\mathbb{P}_{x,y;T}$, i.e.

$$\mathbb{E}_{x,y;T}[h(Y)] = \int_{W(M)} h(y) d\mathbb{P}_{x,y;T}(y) \quad \forall h : (W(M), \mathscr{B}(W)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R})) \tag{3.14}$$

where $Y : (W(M), \mathscr{B}(W), \mathbb{P}_{x,y;T}) \to (W(M), \mathscr{B}(W))$ is the coordinate function under $\mathbb{P}_{x,y;T}$. So, for any $C \in \mathscr{B}(W)$:

$$\mathbb{P}_{x,y;T}(C) = \mathbb{E}_{x,y;T}[1_C] \tag{3.15}$$

Defined as in Equation (3.15), $\mathbb{P}_{x,y;T}$ does not need to be a probability measure on $(W(M), \mathscr{B}(W))$. Instead of checking it directly, we will do something more constructive: we assume it actually is a probability measure on the right space, and derive its finite dimensional marginals; if one proves that a probability exists on $(W(M), \mathscr{B}(W))$ with the same finite dimensional marginals, then it is necessarily the one we were looking for (because such marginals uniquely determine the measure).

Let $s < T$, and consider $F : (W(M), \mathscr{B}_s) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and $f : (M, W(M)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. By definition of conditional expectation:

$$\mathbb{E}[f(X_T)F(X)] = \mathbb{E}[f(X_T)\mathbb{E}_{x,X_T,T}[F(X)]] \tag{3.16}$$

The left hand side can be rewritten by leveraging the strong Markov property of Wiener measure. In particular, let $\mathbb{E}_{X_s}$ denote the expectation when the probability measure on $(W(M), \mathscr{B}(W))$ is $\mathbb{P}_{X_s}$:

$$\mathbb{E}[f(X_T)F(X)] = \mathbb{E}[F(X)\mathbb{E}_{X_s}[f(X_{T-s})]] \tag{3.17}$$

which equals, being p_M the infinitesimal transition density for BM on M:

$$\mathbb{E}[F(X)\int_M p_M(T-s, X_s, y)f(y)dM(y)] = \int_M \mathbb{E}[F(X)p_M(T-s, X_s, y)]f(y)dM(y) \tag{3.18}$$

The right hand side can be written as:

$$\int_M \mathbb{E}_{x,y;T}[F(X)]p_M(T, x, y)f(y)dM(y) \tag{3.19}$$

(simply see $\mathbb{E}_{x,X_T;T}[\cdot]$ as a function of X_T and use transition density again).

By arbitrarity of f, this means that for all F $\mathscr{B}_s$-measurable, and $s < T$:

$$\mathbb{E}_{x,y;T}[F(X)] = \frac{\mathbb{E}[F(X)p_M(T-s, X_s, y)]}{p_M(T,x,y)} \tag{3.20}$$

i.e. as a measure on $(W(M), \mathscr{B}_s)$, $\mathbb{P}_{x,y;T}$ is absolutely continuous with respect to $\mathbb{P}_x$ and its Radon-Nikodyn derivative is given by:

$$\left.\frac{d\mathbb{P}_{x,y;T}}{d\mathbb{P}_x}\right|_{\mathscr{B}_s} = \frac{p_M(T-s, X_s, y)}{p_M(T,x,y)} = e_s \tag{3.21}$$

Now, if $\mathbb{P}_{x,y;T}$ exists (as a probability measure on $(W(M), \mathscr{B}(W))$), the joint density function of $Y_{s_1}, Y_{s_2}, ..., Y_{s_l}$ with $0 = s_0 < s_1 < ... < s_l < s_{l+1} = T$ is:

$$p_M(T,x,y)^{-1} \prod_{i=0}^{l} p_M(s_{i+1} - s_i, y_i, y_{i+1}) \tag{3.22}$$

with $y_0 = x$ and $y_{l+1} = y$. The following finally establishes the existence of $\mathbb{P}_{x,y;T}$:

Theorem 3.2. *Let M be a compact Riemaniann manifold. Then there is a probability measure $\mathbb{P}_{x,y;T}$ on $(W(M), \mathscr{B}(W))$ (which can be seen also as a probability on the bridge space with Borel sets over it) whose finite-dimensional marginal distributions are given by* (3.22)*.*

Once existence is established, we would like to know whether there is an SDE whose solution provides a BB on M. Recall indeed that in $\mathbb{R}^d$ a BB X from x to y is characterized by:

$$\begin{cases} dX_t = b_t - \frac{X_t - y}{T-t} dt \\ X_0 = x \end{cases} \tag{3.23}$$

where b_t is a Euclidean BM on $(W(\mathbb{R}^d), \mathscr{B}(W(\mathbb{R}^d))_*, \mathbb{P}_{x,y;T})$.

On a Riemaniann manifold we cannot, in general, characterize the process itself, but rather its horizontal lift.

Proposition 3.6. *Let $\mathbb{P}_{x,y;T}$ be the law of a BB from $x \in M$ to $y \in M$ at time T. Then there is a euclidean Brownian motion b_s such that the horizontal lift of the Brownian bridge X, U, to the orthonormal frame $O(M)$ satisfies the following SDE:*

$$dU_s = H_i(U_S) \circ \{db_s^i + H_i \log \tilde{p}_M(s, U_s, y)ds\} \tag{3.24}$$

where $\tilde{p}_M(s, u, y) = p_M(s, \pi u, y)$.

3.3. (Stochastic) Differential Geometry Under Isometric Diffeomorphism

In this section we state some results that hold true an isometric diffeomorphism with $\mathbb{R}^d$ exists. We will consider a general Riemaniann manifold $(M, g(\cdot,\cdot))$, and denote by $\theta : M \to \mathbb{R}^d$ the isometric diffeomorphism we assume to exist. It is quite natural to expect that, under such a strong assumption, all (stochastic) differential geometric properties are simply "transferred" from a manifold to the other. This is very convenient, because on $\mathbb{R}^d$ many objects (such as Brownian motion) are very well understood.

In Section 3.3.1, we see what the local form of the Levi-Civita connection is, how parallel transport on M is achieved and how to obtain the horizontal lift of a curve on it to the (orthonormal) frame bundle. Then, we will study stochastic horizontal lift in Section 3.3.2, to finally discuss Brownian motion and Brownian bridge in Section 3.3.3 and Section 3.3.4 respectively.

3.3.1. Levi-Civita Connection, Parallel Transport and Horizontal Lift

Recall that we can see $\theta : M \to \mathbb{R}^d$ as a global chart for M, and if $S \in M$, then $x(S) = \{x^i(S)\}$ are the coordinates with respect to this global chart. With this convention, we let $X_i = \frac{\partial}{\partial x^i}$.

Proposition 3.7. *The basis for T_SM provided by $\{X_i\}_{i=1}^d$ is orthonormal with respect to the Riemaniann metric $g(\cdot,\cdot)$.*

Proof. Remember that on T_SM, $X_i = \frac{\partial}{\partial x^i} = d\theta_S(e_i)$ where e_i is the i^{th} element of the canonical basis in $\mathbb{R}^d$. Then simply using the definition of isometry:

$$g_S(X_i, X_j) = g(d\theta_S(e_i), d\theta_S(e_j)) = \langle e_i, e_j\rangle = \delta_{ij}$$

□

Since $g(\cdot,\cdot)$ is essentially the metric induced through θ by the usual Riemaniann metric on a flat Euclidean space, it is quite natural to expect that $(M, g(\cdot,\cdot))$ is, in some sense, *flat*. This is made precise in the following:

Proposition 3.8. *Let ∇ be the Levi-Civita connection for the Riemaniann manfiold $(M, g(\cdot,\cdot))$. Then its Christoffel symbols are all null:*

$$\Gamma^k_{ij} = 0 \quad \forall i, j, k \in \{1, 2, ..., d\}$$

Proof. Recall that, from Section A.1.4, we know how to find Christoffel symbols of the Levi-Civita connection as:

$$\Gamma^i_{jk} = \frac{1}{2} g^{im}(g_{jm,k} + g_{mk,j} - g_{jk,m}) \tag{3.25}$$

where $g_{ij} = g(X_i, X_j)$ and denoting by $\{g_{ij}\}$ the corresponding matrix, $\{g^{ij}\}$ is its inverse.

By Proposition 3.7, we know that $g_{ij} = \delta_{ij}$ (being δ the Kroncker delta), so both $\{g_{ij}\}$ and $\{g^{ij}\}$ are the identity matrix, and $g_{jm,k} = 0$ for all $j, m, k \in \{1, 2, ..., d\}$. It is then clear from Equation (3.25) that Christoffel symbols are all zero. □

Direct consequence of this and the local ODE defining parallel transport in Section A.1.4 is that parallel transport of a vector is obtained by simply maintaining the vector unchanged.

Proposition 3.9. *Let $\{x_t\}_{t\in[0,T]} \subseteq M$ be a differentiable curve in M, and let $V_t \in T_{x_t}M$ be a sequence of tangent vectors, and suppose V_t can be written as $V_t = v^i(t)X_i$. Then V_t is the parallel transport of V_0 if and only if:*

$$\dot{v}^k(t) = 0 \quad \forall k = 1, 2, ..., d \tag{3.26}$$

with initial condition v^i_0 (where $V_0 = v^i_0 X_i$). This is equivalent to:

$$V_t = v^i_0 X_i \quad \forall t \in [0, T] \tag{3.27}$$

Recall now that any frame $u \in \mathscr{F}(M)$ can be globally described (globally in our specific case being θ a global chart) by couples $(x, e) = (x^i, e^j_i) \in \mathbb{R}^{d+d^2}$ such that $ue_i = e^j_i X_j$.

Now, take a curve $\{u_t\}_{t\in[0,T]} \subseteq \mathscr{F}(M)$. We know it is horizontal, by definition, if $\{u_t e\}$ is transported parallelly along $\{\pi u_t\}$. From the above, this means that, if we denote

$u_t = (x_t, e_t)$ and if (x_0, e_0) is the initial condition, then $u_t a = u_0 a$ for all $a \in \mathbb{R}^d$. But since $u_t a = e_t a$, we have that $e_0 a = e_t a$ for all $a \in \mathbb{R}^d$, meaning $e_t = e_0$ for all $t \in [0, T]$. This is a proof of the following:

Proposition 3.10. *On $(M, g(\cdot,\cdot))$ the horizontal lift of a differentiable curve x_t must be necessarily of the form $u_t = (x_t, e_0)$ if written with respect to the above defined global chart. In particular, $u_0 = (x_0, e_0)$.*

So, essentially, if we have a curve x_t, to get its horizontal lift we choose the initial basis for $T_{x_0}M$ and we keep it unchanged over time.

From Proposition A.5 we know how to write the fundamental horizontal vector fields in terms of a (global) chart for M. In this specific case, we have that:

Proposition 3.11. *Let $u \in \mathscr{F}(M)$, and let (x^i, e^j_i) be its expression in terms of the above defined global chart. Then the fundamental horizontal vector fields at u, $H_i(u)$, are:*

$$H_i(u) = e^j_i X_j = ue_i \tag{3.28}$$

Proof. This is straightforward. Recall that $ue_i = e^j_i X_j$ and that:

$$H_i(u) = e^j_i X_j - e^j_i e^l_m \Gamma^k_{jl}(x) X_{km}$$

But since Christoffel symbols are zero by Proposition 3.8, we have the required expression. □

Remark. Can we make sense of $H_i(u) = ue_i$ when $H_i(u) \in H_u\mathscr{F}(M)$ and $ue_i \in T_{\pi u}M$? Take $f \in C^\infty(\mathscr{F}(M))$: we can write $f(u) = \hat{f}((x, e)$ where (x, e) is the expression of u in terms of the usual global chart for $\mathscr{F}(M)$. Now, explicitely:

$$H_i f(u) = e^j_i X_j f(u) = e^j_i(u) \frac{\partial \hat{f}}{\partial x^j}(x, e)$$

meaning that $H_i(u)$ acts on f as ue_i, properly extended to act on functions $C^\infty(\mathscr{F}(M))$, would do.

3.3.2. Stochastic Horizontal Lift

First of all, let us prove the stochastic counterpart of Proposition 3.10:

Proposition 3.12. *On $(M, g(\cdot,\cdot))$ the stochastic horizontal lift of a semimartingale X_t must be necessarily of the form $U_t = (X_t, E_0)$ if written with respect to the above defined global chart. In particular, $U_0 = (X_0, E_0)$.*

Proof.* Recall that M can be regarded, through θ, as $\mathbb{R}^d$ itself, so, based on Section B.1, if X_t is an M-valued semimartingale, we obtain its horizontal lift U_t by:

$$dU_t = P^*_\alpha(U_t) \circ dX^\alpha_t \tag{3.29}$$

that is, explicitely, for all $f \in C^\infty(\mathscr{F}(M))$:

$$df(U_t) = P^*_\alpha f(U_t) \circ dX^\alpha_t \tag{3.30}$$

where X_t is seen as an $\mathbb{R}^d$-valued semimartingale. Now, let $\hat{f}(x, e) = f(u)$ where $u = (x, e)$ in the usual global chart for $\mathscr{F}(M)$. Then, since M seen as a closed submanifold of $\mathbb{R}^d$ (so technically speaking $\theta(M)$) is $\mathbb{R}^d$ itself, and since we know from Lee [15] that the tangent space at each point to a euclidean space is identifiable with the space itself, $P_\alpha(x) = e_\alpha$ for each $x \in M$ and $\alpha = 1, 2, ..., d$. Now, $P^*_\alpha(x) = (P_\alpha(x))^*$, so $P^*_\alpha(x) = (e_\alpha)^*$, but by Section 3.3.1 $(e_\alpha)^* = e_\alpha$. This means that:

$$P^*_\alpha f(u) = \frac{\partial \hat{f}}{\partial x^\alpha}(x, e) \tag{3.31}$$

Now, recall that Equation (3.29) is equivalent to testing it on the coordinate functions when seeing $\mathscr{F}(M)$ as $\mathbb{R}^{d+d^2}$ (thanks to the global chart (x, e)).

First, let $f = x^i(U_t)$:

$$\begin{aligned} dx^i(U_t) = & \\ &= df(U_t) = P^*_\alpha f(U_t) \circ dX^\alpha_t = \\ &= \frac{\partial x^i}{\partial x^\alpha}(x(U_t), e(U_t)) \circ dX^\alpha_t = \\ &= dX^i_t \end{aligned}$$

that is quite obvious.

On the other hand, the very same calculations for $f(u) = e^i_j(u)$ shows that it must hold $de^i_j(U_t) = 0$, since $\frac{\partial e^i_j}{\partial x^\alpha} = 0$ for all $i, j, \alpha = 1, 2, .., d$. $\square$

A natural and straightforward consequence of this fact is that when Assumption 4.3 holds true, an SDE for the horizontal lift of a semimartingale X_t is equivalent to a properly defined SDE on X_t. This will be of particular interest when dealing with Brownian bridge. For general Riemaniann manifolds, an SDE for BB can only be given for its horizontal lift. In such case, however, this will be equivalent to an SDE for the process itself.

Proposition 3.13. *Let X_t be a semimartingale on M, and U_t its horizontal lift starting at U_0. Moreover, denote $U_t = (X_t, E_0)$ in terms of the usual global chart. Then an SDE on U_t:*

$$dU_t = \bar{V}_\alpha(t, U_t) \circ dZ_t^\alpha \tag{3.32}$$

is equivalent to an SDE on X_t:

$$dX_t = V_\alpha(t, X_t) \circ dZ_t^\alpha \tag{3.33}$$

where Z_t is an m-dimensional semimartingale, and where we define the relation between $\bar{V}$ and V as follows. Take $\bar{f} \in C^\infty(\mathscr{F}(M))$, and, with an abuse of notation, let $\bar{f}(u) = \bar{f}(x, e)$. Let $f \in C^\infty(M)$ defined by $f(x) = \bar{f}(x, E_0)$. Then:

$$\bar{V}_\alpha \bar{f}(x, E_0) = V_\alpha f(x) \tag{3.34}$$

Proof.* First, start from an SDE for U_t:

$$dU_t = \bar{V}_\alpha(t, U_t) \circ dZ_t^\alpha \tag{3.35}$$

Let $\bar{f}(x, e) = x^i$, so that $\bar{f}(U_t) = f(x(U_t)) = X_t^i$ (i^{th} coordinate function for X_t). Then we can test the SDE with f to get an SDE for the i^{th} coordinate function for X_t:

$$\begin{aligned} dX_t^i &= \bar{V}_\alpha \bar{f}(U_t) \circ dZ_t^\alpha = \\ &= V_\alpha f(U_t) \circ dZ_t^\alpha \end{aligned}$$

that is equivalent to the SDE on X_t:

$$dX_t = V_\alpha(t, X_t) \circ dZ_t^\alpha \tag{3.36}$$

Now, let us start with an SDE on M:

$$dX_t = V_\alpha(t, X_t) \circ dZ_t^\alpha \tag{3.37}$$

Chosen $U_0 = (X_0, E_0)$, we know that its horizontal lift must be of the form (X_t, E_0). Now, let $\bar{f} \in C^\infty(\mathscr{F}(M))$. Since $\bar{f}$ is always evaluated on elements (X_t, E_0), we can restrict it to $f \in C^\infty(M)$ as stated in the proposition. By definition of SDE on M:

$$\begin{aligned} d\bar{f}(U_t) &= \\ &= df(X_t) = V_\alpha f(X_t) \circ dZ_t^\alpha = \\ &= \bar{V}_\alpha \bar{f}(U_t) \circ dZ_t^\alpha \end{aligned}$$

by arbitrarity of $\bar{f} \in C^\infty(\mathscr{F}(M))$, the following holds true:

$$dU_t = \bar{V}_\alpha(t, U_t) \circ dZ_t^\alpha \tag{3.38}$$

□

Example 3.1 (Vectorization of Coefficients)**.** Suppose we start with an SDE on U_t, horizontal lift of L_t, of the type:

$$dU_t = \bar{V}_\alpha(t, U_t) \circ dW_t^\alpha + b(t, U_t)dt \tag{3.39}$$

Recall that $V_\alpha(u) = \bar{V}_\alpha^j H_j(u) = \bar{V}_\alpha^j (E_0)_j^i X_i$. We then look for the components with respect to the basis $\{X_i\}$ of V_α as defined in Proposition 3.13. Take $f \in C^\infty(M)$ and let $\bar{f}(x, E_0) = f(x)$. Then:

$$\begin{aligned} V_\alpha f(L) &= \bar{V}_\alpha \bar{f}(L, E_0) = \\ &= \bar{V}_\alpha^i (E_0)_i^j X_j \end{aligned}$$

so that $V_\alpha^j = \bar{V}_\alpha^i (E_0)_i^j$ (one can write $V_\alpha = \bar{V}_\alpha^T E_0$), and similarly for b.

So the equivalent SDE is:

$$dL_t = V_\alpha(t, L_t) \circ dW_t^\alpha + b(t, L_t)dt \tag{3.40}$$

where we gave above expressions of V_α and b in terms of the global basis. In particular, this can be recasted as an SDE on $\mathbb{R}^d$ leveraging the global chart $x^i(L) = \theta^i(L) = X^i$. This leads to:

$$\begin{aligned} dX_t^i &= V_\alpha^j(t, L_t)\frac{\partial x^i}{\partial x^j}(L_t) \circ dW_t^\alpha + b^j(t, L_t)\frac{\partial x^i}{\partial x^j}(L_t)dt = \\ &= V_\alpha^i(t, \theta^{-1}(X_t)) \circ dW_t^\alpha + b^i(t, \theta^{-1}(X_t))dt \end{aligned}$$

□

3.3.3. Brownian Motion on M

In order to understand what the form of a Brownian motion on $(M, g(\cdot,\cdot))$ is, we will leverage once again the fact that θ provides a way to see M as a closed submanifold of $\mathbb{R}^d$ (i.e. $\mathbb{R}^d$ itself) and the fact that endowing M with $g(\cdot,\cdot)$ is equivalent to simply endowning it with the usual Riemaniann structure of $\mathbb{R}^d$, once it is seen as a submanifold. Now, since $P_\alpha = e_\alpha$, Proposition 3.5 directly gives us the SDE that a BM on M (seen as a submanifold of $\mathbb{R}^d$) solves. If B_t is a Brownian motion on M and W_t the standard d-dimensional one:

$$d\theta(B_t) = e_\alpha \circ dW_t^\alpha = dW_t \tag{3.41}$$

so essentially we can obtain a Brownian motion on M by $B_t = \theta^{-1}(W_t)$.

Thanks to this, we can easily obtain the transition density function $P_M(t, x, y)$ on M. Let indeed $C \in \mathscr{B}(M)$, B_t the BM on M and $\mathbb{P}_x$ the Wiener measure on $W(M)$ starting at $x \in M$. Then:

$$\begin{aligned} \int_C p_M(t, x, y)dM(y) &= \mathbb{P}_x(B_t \in C) = \\ &= \mathbb{P}_x(\theta(W_t) \in C) = \\ &= \mathbb{P}_{\theta(x)}(W_t \in \theta(C)) = \\ &= \int_{\theta(C)} p_{\mathbb{R}^d}(t, \theta(x), z)dz = \\ &= \int_C p_{\mathbb{R}^d}(t, \theta(x), \theta(y))|\text{det}J_\theta(y)|dy = \\ &= \int_C p_{\mathbb{R}^d}(t, \theta(x), \theta(y))dM(y) \end{aligned}$$

Which by arbitarity of C gives:

$$p_M(t,x,y) = p_{\mathbb{R}^d}(t,\theta(x),\theta(y))$$

3.3.4. Brownian Bridge on M

We will address two things: first, we discuss the construction of BB over M. Secondly, we express Brownian bridges as solutions to a proper M-valued SDE.

The construction of Brownian bridge on M follows what has been explained in Section 3.2.4. Though Theorem 3.2 assumes M is a compact manifold to prove existence of $\mathbb{P}_{x,y;T}$, we can avoid requiring compactness under Assumption 4.3, and obtain existence of the law of the BB in a slightly different way, that of course leverages the fact that we can, in some sense, think of M as $\mathbb{R}^d$ itself. What we can do is construct such a measure by specifying its finite-dimensional joint distributions, and directly showing that there actually exists a probability measure having such finite-dimensionals, without passing through Kolmogorov extension theorem (as done in Hsu [13]).

Given $x, y \in M$ and a final time T, we want to show that there exists a probability measure (which we will denote as $\mathbb{P}_{x,y;T}$) such that, if $Y : (W(M), \mathscr{B}(W), \mathbb{P}_{x,y;T}) \to (W(M), \mathscr{B}(W))$ is the coordinate process, the joint density of $Y_{s_1}, Y_{s_2}, ..., Y_{s_l}$ $(0 = s_0 < s_1 < ... < s_l < s_{l+1} = T)$ is:

$$p_M(T,x,y)^{-1} \prod_{i=0}^{l} p_M(s_{i+1}-s_i, y_i, y_{i+1}) \tag{3.42}$$

where $y_0 = x$, $y_{l+1} = y$ and p_M is the transition density for the BM on M, which we derived in Section 3.3.3.

Let $A_{s_i} \in \mathscr{B}(M)$ for $i = 1, 2, ..., l$, and define $\mathbb{P}_{x,y;T}$ by:

$$\begin{aligned}
&\mathbb{P}_{x,y;T}(Y_{s_1} \in A_{s_1}, Y_{s_2} \in A_{s_2}, ..., Y_{s_l} \in A_{s_l}) = \\
&= \int_M \int_M ... \int_M p_M(T,x,y)^{-1} \prod_{i=0}^{l} p_M(s_{i+1}-s_i, y_i, y_{i+1}) dM(y_1)dM(y_2)...dM(y_l)
\end{aligned}$$

Now, by what we already saw, this can be rewritten as follows:

$$\int_M \int_M \dots \int_M p_M(T,x,y)^{-1} \prod_{i=0}^{l} p_M(s_{i+1}-s_i, y_i, y_{i+1}) dM(y_1) dM(y_2) \dots dM(y_l) =$$
$$= \int_{\theta(M)} \int_{\theta(M)} \dots \int_{\theta(M)} p_{\mathbb{R}^d}(T,x,y)^{-1} \prod_{i=0}^{l} p_{\mathbb{R}^d}(s_{i+1}-s_i, y_i, y_{i+1}) dy_1 dy_2 \dots dy_l =$$
$$= \hat{\mathbb{P}}_{\theta(x),\theta(y);T}(\theta(Y_{s_1}) \in \theta(A_{s_1}), \theta(Y_{s_2}) \in \theta(A_{s_2}), \dots, \theta(Y_{s_l}) \in \theta(A_{s_l}))$$

where the last expression is just the finite-dimensional joint probability for the Euclidean d-dimensional BM starting at $\theta(x)$ in 0 and ending at $\theta(y)$ at time T (denoted with $\hat{\mathbb{P}}_{\theta(x),\theta(y);T}$). Since this last probability is well known to exist, we have that $\mathbb{P}_{x,y;T}$ exists as a measure on $(W(M), \mathscr{B}(W))$ and is specified by the above relation.

In shorthand, we can summarize this result by saying that if $\mathbb{P}_{x,y;T}$ is the BB on M from x to y and $\hat{\mathbb{P}}_{\theta(x),\theta(y);T}$ the BB on $\mathbb{R}^d$ from $\theta(x)$ to $\theta(y)$, it holds:

$$\hat{\mathbb{P}}_{\theta(x),\theta(y);T} = \Theta_{\#}\mathbb{P}_{x,y;T} \tag{3.43}$$

or equivalently, by invertibility of θ:

$$\mathbb{P}_{x,y;T} = \Theta_{\#}^{-1}\hat{\mathbb{P}}_{\theta(x),\theta(y);T} \tag{3.44}$$

Now that we have dealt with existence, we look for an SDE on M describing BB. In Proposition 3.6, we saw how one can obtain a Brownian bridge on M by solving an SDE for its horizontal lift. As stated in Section 3.3.2, under Assumption 4.3, an SDE for the horizontal lift U_t of a semimartingale X_t is equivalent to an SDE for X_t and the choice of a fixed frame. Our goal is now to rewrite the BB equation for U_t as an SDE for X_t, i.e. the Brownian bridge on M. Since it is relevant for the practical purpose of simulating a BB, we will be interested in writing an SDE for the coordinate functions of X_t. To do so, we essentially follow what has been done in Example 3.1.

Recall that $H_i(U_s) = (E_0)_i^j \frac{\partial}{\partial x^j}$, so that:

$$H_\alpha \log \tilde{p}_M(T-s, U_s, y) = -(E_0)_\alpha^j \frac{1}{2(T-s)} \frac{\partial}{\partial x^j} \left(\sum_{\beta=1}^{d} (x^\beta(L_s) - x^\beta(y))^2 \right) =$$
$$= -(E_0)_\alpha^j \frac{x^j(L_s) - x^j(y)}{T-s}$$

Now, if we let $\bar{V}_\alpha = H_\alpha$, it is clear that, using the notations in Example 3.1, we obtain $V_\alpha^j = (E_0)_\alpha^j$, so that the set of equations we are looking for is just:

$$dX_t^i = (E_0)_\alpha^i \circ db_t^\alpha - (E_0)_\alpha^i (E_0)_j^\alpha \frac{x^j(L_t) - x^j(y)}{T-t} dt \tag{3.45}$$

Now, a question is what (E_0) to consider. Recall that U_t must be an orthonormal frame. Since $U_0 e_i = (E_0)_i^j X_j$ and $g(X_i, X_j) = \delta_{ij}$, we can actually choose E_0 to be the identity matrix, so that the Brownian bridge on M is:

$$dX_t^i = db_t^i - \frac{1}{T-t}(X_t^i - x^i(y)) \tag{3.46}$$

i.e. the usual euclidean Brownian bridge for the process on $\mathbb{R}^d$ given by $X_t = \theta(S_t)$. This makes sense, since θ, as an isometry between $(M, g(\cdot,\cdot)) \to (\mathbb{R}^d, \langle \cdot, \cdot \rangle_2)$, induces all geometric properties from the euclidean space to M.

3.3.5. Schrödinger Bridge on M

In this section, we study how Schrödinger bridges, which we defined in Definition 1.2, change between manifolds under an isometric diffeomorphism. More specifically, we will see that the Schrödinger bridge problem stated in one manifold is in some sense equivalent to the same problem stated in the second one. The solution to the latter is essentially the push-forward of the solution to the former through the diffeomorphism, and vice-versa. This fact will be made precise in Section 3.3.5.

In Section 3.3.5, instead, we restrict our attention to the case one of the two manifolds is $\mathbb{R}^d$. As already seen in Chapter 1, the SB problem has a closed solution on the Euclidean space, and is basically a mixture of Brownian bridges. We will prove that the solution to the SB problem in a smooth manifold isometrically diffeomorphic to $\mathbb{R}^d$ is once again a mixture of manifold-valued Brownian bridges, weighted by an entropically regularized optimal transport plan.

Mapping Schrödinger Bridge

Let us consider two smooth manifolds E and M, endowed with the corresponding Borel sets, so that we have two measurable spaces $(E, \mathscr{B}(E))$, $(M, \mathscr{B}(M))$. Also, let us consider the path spaces on the two manifolds: $(W(E), \mathscr{B}(W(E)))$ and $(W(M), \mathscr{B}(W(M)))$. We consider a diffeomorphism $\theta : M \to E$, which inverse is denoted by $g : E \to M$, and we define another invertible function $\Theta : W(M) \to W(E)$ by:

$$\Theta(\omega)_t = \theta(\omega_t) \quad \forall t \in [0,T], \forall \omega \in W(M)$$

with inverse $G : W(E) \to W(M)$. Moreover, let $\mathbb{Q}$ be a reference probability measure on $(W(M), \mathscr{B}(W(M)))$, and denote by $\hat{\mathbb{Q}}$ the pushforward trough Θ of $\mathbb{Q}$, i.e. $\hat{\mathbb{Q}} = \Theta_{\#}\mathbb{Q}$ is a probability measure on $(W(E), \mathscr{B}(W(E)))$. Let q_0, q_1 be probability measures on $(M, \mathscr{B}(M))$ and let $\hat{q}_0 = \theta_{\#}q_0$ and $\hat{q}_1 = \theta_{\#}q_1$. We denote by p_0 and p_1 the initial and final marginals of $\mathbb{P}$, general probability measure on $(W(M), \mathscr{B}(W(M)))$ (e.g. $\mathbb{P}(Y_0 \in A) = p_0(A)$ where $Y : (W(M), \mathscr{B}(W(M))) \to (W(M), \mathscr{B}(W(M)))$ are the coordinate processes on $W(M)$). Similarly, the starting and end marginals for the general probability measure on $(W(E), \mathscr{B}(W(E)))$ are $\hat{p}_0$ and $\hat{p}_1$.

First of all, recall a result in measure theory:

Proposition 3.14. *Let X and Y be separable metric spaces, and $\mu \in \mathscr{P}(X)$ (i.e. space of probability measures on X). Let $r : X \to Y$ be a μ-measurable function and denote with $\nu = r_{\#}\mu \in \mathscr{P}(Y)$ the push-forward of μ through r. Then for any $\varphi : Y \to \mathbb{R}$, ν-integrable function, we have:*

$$\int_Y \varphi(y) d\nu(y) = \int_X \varphi(r(x)) d\mu(x) \tag{3.47}$$

For the definition of Schrödinger bridge, see Definition 1.2.

Before showing that the SB problem on $W(E)$ and on $W(M)$ are equivalent, we prove some intermediate lemmas.

First, we notice that probability measures on the two probability spaces can essentially be identified through Θ

Lemma 3.1. *Under the above introduced notation, there exists a bijective correspondence between probability measures on $(W(E), \mathscr{B}(W(E)))$ and probability measures on $(W(M), \mathscr{B}(W(M)))$ provided by Θ.*

Proof.* Let $\mathbb{P}$ be a probability on $W(M)$, then it is clear that $\hat{\mathbb{P}} = \Theta_{\#}\mathbb{P}$ is a probability on $W(E)$. On the other hand, let $\hat{\mathbb{P}}$ be a probability measure on $W(E)$. Then $\hat{\mathbb{P}} = (\Theta \circ \Theta^{-1})_{\#}\hat{\mathbb{P}} = \Theta_{\#}((\Theta^{-1})_{\#}\hat{\mathbb{P}})$. Define $\mathbb{P} = (\Theta^{-1})_{\#}\hat{\mathbb{P}}$ (well defined by invertibility of Θ), and thus $\hat{\mathbb{P}} = \Theta_{\#}\mathbb{P}$, being $\mathbb{P}$ a probability measure on $W(M)$. $\square$

From now on, whenever $\mathbb{P}$ is a probability measure on $W(M)$, we will denote by $\hat{\mathbb{P}}$ the unique probability measure on $W(E)$ such that $\hat{\mathbb{P}} = \Theta_{\#}\mathbb{P}$ and viceversa.

A very straightforward result is the following on change of marginals:

Lemma 3.2. *Under the usual notations, $\hat{\mathbb{P}}$ has initial and final marginals $\hat{p}_0$ and $\hat{p}_1$ if and only if $\mathbb{P}$ has as inital and final marginals p_0 and p_1.*

Proof.* Suppose $\hat{\mathbb{P}}$ has the said marginals. Take $B \in \mathscr{B}(M)$, and let Y be the coordinate process on $(W(M), \mathscr{B}(W(M)), \mathbb{P})$ and X the one on $(W(E), \mathscr{B}(W(E)), \hat{\mathbb{P}})$:

$$\begin{aligned}\mathbb{P}(Y_0 \in B) &= \hat{\mathbb{P}}(G^{-1}((Y_0 \in B))) = \\ &= \hat{\mathbb{P}}(X_0 \in g^{-1}(B)) = \\ &= \hat{p}_0(g^{-1}(B)) = \\ &= p_0(B)\end{aligned}$$

where with a small but widespread abuse of notations we denote by $(Y_0 \in B)$ the subset of $W(M)$ given by $Y_0^{-1}(B)$.

Similarly for end marginals, and if we begin with known marginals to be p_0 and p_1. □

As it is natural to expect, absolute continuity is preserved by push-forward:

Lemma 3.3. *Under the above notations, $\hat{\mathbb{P}} << \hat{\mathbb{Q}}$ if and only if $\mathbb{P} << \mathbb{Q}$.*

Moreover, denoting the Radon-Nikodym derivative of $\hat{\mathbb{P}}$ wrt $\hat{\mathbb{Q}}$ by $\hat{e}(x)$ for $x \in W(E)$ and the one of $\mathbb{P}$ wrt $\mathbb{Q}$ as $e(y)$ for $y \in W(M)$, then:

$$e(y) = \hat{e}(\Theta(y)) \quad \forall y \in W(M) \tag{3.48}$$

or, equivalently:

$$\hat{e}(x) = e(G(x)) \quad \forall x \in W(E) \tag{3.49}$$

Proof.* Suppose $\hat{\mathbb{P}} << \hat{\mathbb{Q}}$, i.e. if $\hat{\mathbb{Q}}(A) = 0$, then $\hat{\mathbb{P}}(A) = 0$. Suppose $\mathbb{Q}(B) = 0$, then $\hat{\mathbb{Q}}(G^{-1}(B)) = 0$. But this means $\hat{\mathbb{P}}(G^{-1}(B)) = \mathbb{P}(B) = 0$. The opposite implication can be proved in the same way.

Now, let $B \in \mathscr{B}(W(M))$:

$$\begin{aligned}
\mathbb{P}(B) = \hat{\mathbb{P}}(G^{-1}(B)) &= \\
&= \int_{G^{-1}(B)} \hat{e}(x) d\hat{\mathbb{Q}}(x) = \\
&= \int_X \mathbf{1}_{G^{-1}(B)}(x)\hat{e}(x) d\hat{\mathbb{Q}}(x) = \\
&= \int_X \mathbf{1}_{G^{-1}(B)}(x) e(G(x)) d\hat{\mathbb{Q}}(x) = \\
&= \int_X \mathbf{1}_B(G(x)) e(G(x)) d\hat{\mathbb{Q}}(x) = \\
&= \int_Y \mathbf{1}_B(y) e(y) d\mathbb{Q}(y) = \\
&= \int_B e(y) d\mathbb{Q}(y)
\end{aligned}$$

where we used Proposition 3.14, the fact that $\mathbf{1}_B e$ is $\mathbb{Q}$-integrable function and the claimed form for e. Notice that since the Radon-Nikodym derivative is $\mathbb{Q}$ (or $\hat{\mathbb{Q}}$)-a.e. unique, $e(y)$ is actually the RN derivative of $\mathbb{P}$ wrt $\mathbb{Q}$. Once again by uniqueness, $\hat{e}(x)$ must have the claimed form. □

The following is instead about the equality of objective functions:

Lemma 3.4. *Under the usual notations:*

$$KL(\hat{\mathbb{P}}|\hat{\mathbb{Q}}) = KL(\mathbb{P}|\mathbb{Q})$$

Proof.* By Lemma 3.3, the Radon-Nikodym derivative of $\hat{\mathbb{P}}$ wrt to $\hat{\mathbb{Q}}$, denoted by $\frac{d\hat{\mathbb{P}}}{d\hat{\mathbb{Q}}}$, exists if and only if $\frac{d\mathbb{P}}{d\mathbb{Q}}$ exists.

Then let us assume they exist, and denote RN derivatives as in Lemma 3.3. Then:

$$\begin{aligned}
KL(\hat{\mathbb{P}}|\hat{\mathbb{Q}}) &= \int_X \log(\hat{e}(x)) d\hat{\mathbb{P}}(x) = \\
&= \int_X \log(e(G(x))) d\hat{\mathbb{P}}(x) = \\
&= \int_Y \log(e(y)) d\mathbb{P}(y) = \\
&= KL(\mathbb{P}|\mathbb{Q})
\end{aligned}$$

□

We can now state the main result of this section.

Proposition 3.15. *Under the above notations, $\hat{\mathbb{P}}^*$ is the Schrödinger bridge between $\hat{q}_0$ and $\hat{q}_1$ on $W(E)$ with reference process $\hat{\mathbb{Q}}$ if and only if $\mathbb{P}^*$ is the one between q_0 and q_1 on $W(M)$ with reference process $\mathbb{Q}$.*

Proof.* Suppose by contraddiction that, e.g., $\hat{\mathbb{P}}^*$ is optimal on $W(E)$ but that there exists $\tilde{\mathbb{P}}$ on $W(M)$ such that $KL(\tilde{\mathbb{P}}|\mathbb{Q}) < KL(\mathbb{P}^*|\mathbb{Q})$. Necessarily there must exist $\hat{\tilde{\mathbb{P}}}$ on $W(E)$ such that $\tilde{\mathbb{P}} = G_\# \hat{\tilde{\mathbb{P}}}$. But we know that in such case $\hat{\tilde{\mathbb{P}}}$ has $\hat{q}_0$ and $\hat{q}_1$ as initial and final marginals. Also, we know that $KL(\hat{\tilde{\mathbb{P}}}|\hat{\mathbb{Q}}) = KL(\tilde{\mathbb{P}}|\mathbb{Q})$. But then:

$$\begin{aligned} KL(\hat{\tilde{\mathbb{P}}}|\hat{\mathbb{Q}}) &= KL(\tilde{\mathbb{P}}|\mathbb{Q}) < \\ &< KL(\mathbb{P}^*|\mathbb{Q}) = \\ &= KL(\hat{\mathbb{P}}^*|\hat{\mathbb{Q}}) \end{aligned}$$

which is a contraddiction, being $\hat{\mathbb{P}}^*$ optimal for SB problem on $W(E)$.

In the very same way one can argue that if $\mathbb{P}^*$ is optimal on $W(M)$, then $\hat{\mathbb{P}}^*$ must be optimal on $W(E)$. $\square$

Mapping Solutions

We now focus on the case $E = \mathbb{R}^d$ with its usual Euclidean scalar product.

We know that the solution to the Schrödinger bridge problem in $\mathbb{R}^d$ with reference process a Brownian motion with diffusion rate σ is given by:

$$\hat{\mathbb{P}}^* = \int_{\mathbb{R}^d \times \mathbb{R}^d} \hat{\mathbb{P}}_{x_0,x_1} d\hat{\pi}^*_{2\sigma^2}(x_0, x_1) \tag{3.50}$$

where $\hat{\mathbb{P}}_{x_0,x_1}$ is the law of a Euclidean Brownian bridge from x_0 to x_1 with diffusion rate σ and $\hat{\pi}^*_{2\sigma^2}(x_0, x_1)$ is the entropic OT plan between $\hat{q}_0$ and $\hat{q}_1$, meaning:

$$\pi^*_\epsilon(q_0, q_1) = \arg \min_{\pi \in \Pi(q_0,q_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} d(x_0, x_1)^2 d\pi(x_0, x_1) + \epsilon KL(\pi|q_0 \otimes q_1) \tag{3.51}$$

Let us now define $\bar{g} : \mathbb{R}^d \times \mathbb{R}^d \to M \times M$ by $\bar{g}(x_0, x_1) = (g(x_0), g(x_1))$. Then take $B \in \mathscr{B}(W(M))$:

$$
\begin{aligned}
\mathbb{P}^*(B) = \hat{\mathbb{P}}^*(G^{-1}(B)) = \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} \hat{\mathbb{P}}_{x_0,x_1}(G^{-1}(B))d\hat{\pi}^*_{2\sigma^2}(x_0,x_1) = \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} G_\#\hat{\mathbb{P}}_{x_0,x_1}(B)d\hat{\pi}^*_{2\sigma^2}(x_0,x_1) = \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} G_\#\hat{\mathbb{P}}_{\theta(g(x_0)),\theta(g(x_1))}(B)d\hat{\pi}^*_{2\sigma^2}(x_0,x_1) = \\
&= \int_{M\times M} G_\#\hat{\mathbb{P}}_{\theta(y_0),\theta(y_1)}(B)d(\bar{g}_\#\hat{\pi}^*_{2\sigma^2})(y_0,y_1) = \\
&= \int_{M\times M} \mathbb{P}_{y_0,y_1}(B)d(\bar{g}_\#\hat{\pi}^*_{2\sigma^2})(y_0,y_1)
\end{aligned}
$$

where in the last passage we used Equation (3.44) and Proposition 3.14.

This way, we show that the Schrödinger bridge on M with reference the M-valued Brownian motion has the form of a mixture of M-valued Brownian bridges weighted by $\bar{g}_\#\hat{\pi}^*_\epsilon$.

The obvious question now is: is $\bar{g}_\#\hat{\pi}^*_\epsilon$ the entropic OT plan between q_0 and q_1? We will proceed in a very similar way as done in Section 3.3.5. Being $\bar{g}$ clearly invertible, we have a bijective correspondence between transport plans between $\hat{q}_0$ and $\hat{q}_1$ and transport plans between q_0 and q_1: $\pi \in \Pi(q_0, q_1)$ if and only if $\hat{\pi} \in \Pi(\hat{q}_0, \hat{q}_1)$, where $\pi = \bar{g}_\#\hat{\pi}$ (equivalently, $\hat{\pi} = \bar{\theta}\pi$). We will then prove the equality of the objective functions for the two entropic OT problems, establishing equivalence of these.

Let us take $B \in \mathscr{B}(M)$, then:

$$
\begin{aligned}
\pi(B\times M) &= \hat{\pi}(\bar{\theta}(B\times M)) = \\
&= \hat{\pi}(\theta(B)\times\theta(M)) = \\
&= \hat{q}_0(\theta(B)) = \\
&= q_0(B)
\end{aligned}
$$

Similarly for q_1. Thus we have that $\hat{\pi} \in \Pi(\hat{q}_0, \hat{q}_1)$ if and only if $\pi \in \Pi(q_0, q_1)$.

Let us prove that the first part of the objectives coincide:

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} d(x_0,x_1)^2 d\hat{\pi}(x_0,x_1) = \int_{M\times M} d(\theta(y_0),\theta(y_1))^2 d\pi(y_0,y_1) =$$
$$= \int_{M\times M} d_M(y_0,y_1)^2 d\pi(y_0,y_1)$$

where the second passage is due to the fact that f is an isometry between Riemaniann manifolds (thus geodesic distances map accordingly).

The second part of the objective is the same for the very same arguments used in Section 3.3.5:

$$KL(\pi|q_0\otimes q_1) = KL(\hat{\pi}|\hat{q}_0\otimes\hat{q}_1)$$

Then the two entropic OT problems are equivalent.

4 | Model on SPD Manifold

This chapter is devoted to the gathering of all relevant results obtained in the previous ones and the definition of a generative model over the SPD manifold endowed with the log-Cholesky metric. Such model is essentially the extension to this specific manifold of the framework introduced in Tong et al. [20].

In Section 4.2, we state three assumptions that hold true for our setting, and prove that they are actually satisfied over the log-Cholesky manifold. Though Assumption 4.3 implies Assumption 4.2, which itself implies Assumption 4.1, we prefer to state them separately, so that we can make precise which passages require stricter assumptions and which ones are still valid under looser hypothesis.

We then define the claimed stochastic generative model in Section 4.3, and show that we can obtain its Markovization as done in Section 1.3. Essentially, we define a mixture of SDE solutions and show (see Proposition 4.1 and Proposition 4.2) that this model corresponds to a model defined on the Euclidean space (i.e. one is a bijective transformation of the other). Since we explained in Chapter 1 how to obtain the Markov projection and how to fit such mixture models on $\mathbb{R}^d$, we show in Proposition 4.3 that the transformed Markov projection of the Euclidean process is actually the Markov projection of the manifold-valued process, and we can thus simulate it by simulating the former and transforming it.

We finally do, in Section 4.4, a small detour about the interpretation of our model at the matrix entries' level (meaning, the component-wise behaviour of the matrix-valued process we simulate).

4.1. Notations and Preliminaries

In this chapter, N denotes the manifold of SPD matrices, and $g(\cdot, \cdot)$ is the log-Cholesky Riemaniann metric we endow it with (see Chapter 2 for a detailed discussion on this Riemaniann manifold). $\mathcal{L}$ is instead the manifold of lower triangular matrices, endowed with the Frobenius norm. $\theta : N \to \mathbb{R}^d$ is the isometric diffeomorphism which maps N to a Euclidean space, endowed with the usual scalar product. We let TN be the tangent bundle, $T_S N$ the tangent space to N at $S \in N$ and $\Gamma(TN)$ the set of vector fields over N. Recall, as explained in Chapter 2, that N can be described by two global charts: one is θ itself, the other is $\epsilon : N \to \mathbb{R}^d$, consisting of the vectorization of the lower triangular part of SPD matrices. If S is a matrix, we write its strict lower triangular part (without the diagonal) as $\lfloor S \rfloor$, whereas its diagonal part as $\mathbb{D}(S)$. Moreover, in Section 4.4 we let $E = \lfloor S \rfloor + \mathbb{D}(S)$ (i.e. lower triangular part of S with its diagonal).

We denote with $W(N)$ the space of continuous maps $[0,T] \to N$ with their explosion time (similarly for $W(\mathbb{R}^d)$), and define $\Theta : W(N) \to W(\mathbb{R}^d)$ by $\Theta(\omega)_t = \theta(\omega_t)$ for all $\omega \in W(N)$. W_t denotes the d-dimensional Brownian motion, and W_t^α is its α^{th} component, $\alpha = 1, 2, ..., d$.

$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ is a filtered probability space, whereas $(\mathcal{Q}, \mathcal{A})$ is a measurable space. If a process S is defined over $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ and takes values in N, then we write $S = (\Omega, \mathcal{F}, (\mathcal{F}_t)_t, (S_t)_t \subseteq N, \mathbb{P})$. Finally, for any topological space E, we denote by $\mathscr{B}(E)$ its Borel σ-algebra.

4.2. Assumptions On SPD Manifold

In this section we define three assumptions (in order of strength) and check that all three are satisfied for the manifold of SPD matrices (with the log-Cholesky Riemaniann metric).

Assumption 4.1. An SDE on N of the type:

$$dS_t = b(t, S_t)dt + v_\alpha(t, S_t)dW_t^\alpha \tag{4.1}$$

is satisfied if and only if an SDE on $\mathbb{R}^d$ is satisfied:

$$dX_t = \bar{b}(t, L_t)dt + \bar{v}_\alpha(t, L_t)dW_t^\alpha \tag{4.2}$$

being $b(t,\cdot), v_\alpha(t,\cdot)$ vector fields on N, $\bar{b}(t,\cdot), \bar{v}_\alpha(t,\cdot)$ elements in $\mathbb{R}^d$ for any t, W_t an m-dimensional Brownian motion, and having used Einstein notation on $\alpha = 1, 2, ..., m$.

Assumption 4.2. There exists a homeomorphism $\theta : N \to \mathbb{R}^d$ (continuous, invertible and with continuous inverse).

Assumption 4.3. The homeomorphism θ in Assumption 4.2 is also an isometric diffeomorphism between the two Riemaniann manifolds $(N, g(\cdot,\cdot))$ and $(\mathbb{R}^d, \langle\cdot,\cdot\rangle)$, where $g(\cdot,\cdot)$ is the Riemaniann metric on N and $\langle\cdot,\cdot\rangle$ the usual scalar product in Euclidean spaces.

First of all, notice that, as explained in Section 2.3, the function $\theta : N \to \mathbb{R}^d$ is a homeomorphism (continuous, invertible and with continuous inverse) mapping N into the whole $\mathbb{R}^d$, establishing that Assumption 4.2 holds true.

This allows us to see $\theta^i : N \to \mathbb{R}$ as coordinate functions when seeing N as a closed submanifold of $\mathbb{R}^d$ (as explained in Remark 3.1). Then, considering Proposition 3.1, we get the following result. Let:

$$\begin{aligned} &b, v_\alpha : [0,+\infty) \times N \to TN \\ &b(t,S_t), v_\alpha(t,S_t) \in T_{S_t}N \\ &b(t,\cdot), v_\alpha(t,\cdot) \in \Gamma(TN) \end{aligned}$$

denoting with TN the tangent bundle to N and with $\Gamma(TN)$ the set of vector fields on N, and $\alpha \in \{1,2,...,m\}$.

An SDE on N:

$$dS_t = b(t,S_t)dt + v_\alpha(t,S_t) \circ dW_t^\alpha$$

is equivalent to a set of d real SDEs (i.e. an SDE on $\mathbb{R}^d$):

$$d\theta^i(S_t) = b\theta^i(t,S_t)dt + v_\alpha\theta^i(t,S_t) \circ dW_t^\alpha$$

for all $i = 1,2,...,d$ (here $b\theta^i$ denotes the application of the vector field b to θ^i). We thus established Assumption 4.1. Notice that it is not necessary for θ to be a homeomorphism in order for this result to hold, but only that it is a chart, meaning Assumption 4.2 is stronger than Assumption 4.1.

Finally, in Section 2.2, we already saw that $\theta : (N, g(\cdot,\cdot)) \to (\mathbb{R}^d, \langle\cdot,\cdot\rangle)$ is an isometry. So also Assumption 4.3 is true in the SPD manifold with log-Cholesky metric.

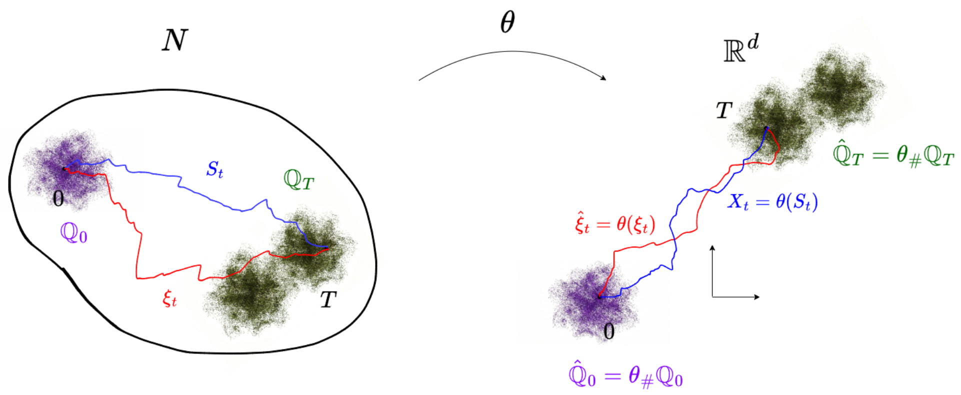

Figure 4.1: Schema of Stochastic processes and their Markov projections.

4.3. Model Definition

Under the same notations used above, we define a model on the smooth manifold N to map the distribution $\mathbb{Q}_0$ to $\mathbb{Q}_T$ through a stochastic process entirely lying on N. The idea is to reconduct the problem to the much more familiar Euclidean setting. In particular, we will see how we can naturally reobtain the same setting as the one tackled by Tong et al. [20] and described in Chapter 1; moreover, we will establish that the properly transformed Markovization of the process on $\mathbb{R}^d$ is actually the Markovization of the process we define on N.

We schematically illustrate this situation in Figure 4.1.

Consider the following stochastic processes and random variables:

$$
\begin{aligned}
&S = (\Omega, \mathcal{F}, (\mathcal{F}_t)_t, (S_t)_t \subseteq M, \mathbb{P})\\
&Z : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{Q}, \mathcal{A})\\
&X = \Theta(S)
\end{aligned}
$$

Assume the conditional law of S given $Z = z$, $n(z, dx)$, exists and that we have $Y^z = (\Omega^z, \mathcal{F}^z, (\mathcal{F}^z_t)_t, (Y^z_t)_t \subseteq \mathcal{Q}, \mathbb{P}^z)$ having $n(z, \cdot)$ as law and solving:

$$dY^z_t = b(t, Y^z_t, z)dt + g(t)dW_t \tag{4.3}$$

Loosely, we can then write the law of process S in a way that mimicks (1.1) as:

$$\mathbb{P}(S \in \cdot) = \int_{\mathcal{Q}} \mathbb{P}^z(Y^z \in \cdot)q(dz)$$

A reason for choosing to parametrize this kind of model is that Schrödinger bridges over N have this form, provided we let $Z = (S_0, S_T)$, the conditional processes be Brownian bridges and the transport plan between initial and final distributions be the optimal for the entropy-regularized optimal transport problem (as explained in Section 3.3.5).

The following result allows us to get back to the Euclidean setting

Proposition 4.1. *$\mathbb{P}(X|Z = z)$ exists, we write it as $\hat{n}(z, dx)$ and is given by:*

$$\hat{n}(z, \cdot) = \Theta_{\#}n(z, \cdot) \quad \forall z \in \mathcal{Q}$$

Proof. Let $A \in \mathscr{B}(W(\mathbb{R}^d))$ and $B \in \mathcal{A}$.

$$\begin{aligned}\mathbb{P}(X \in A, Z \in B) &= \mathbb{P}(S \in \Theta^{-1}(A), Z \in B) = \\ &= \int_B n(z, \Theta^{-1}(A))dq(z) = \\ &= \int_B (\Theta_{\#}n)(z, A)dq(z) = \\ &= \int_B \hat{n}(z, A)dq(z)\end{aligned}$$

Moreover, if we take $A \in \mathscr{B}(W(\mathbb{R}^d))$, then $\hat{n}(y, A) = n(y, \Theta^{-1}(A))$. But Θ is measurable (e.g. by continuity), so $\Theta^{-1}(A) \in \mathscr{B}(W(M))$, thus the map $z \mapsto \hat{n}(z, A) = n(z, \Theta^{-1}(A))$ is $\mathcal{A}$-measurable by definition of conditional law. □

Now, recall that by fullfillment of Assumption 4.1, Equation (4.3) is equivalent to the following SDE for $\hat{Y}^z = \Theta(Y^z)$:

$$d\hat{Y}^z_t = \hat{b}(t, \hat{Y}^z_t, z)dt + g(t)dW_t \tag{4.4}$$

Proposition 4.2. *$\hat{Y}^z$ has law $\hat{n}(z, \cdot)$.*

Proof. Let us consider $A \in \mathscr{B}(W(\mathbb{R}^d))$:

$$\mathbb{P}^z(\hat{Y}^z \in A) = \mathbb{P}^z(Y^z \in \Theta^{-1}(A)) = \\ = n(z, \Theta^{-1}(A)) = \hat{n}(z, A)$$

□

In Section 1.3, we saw that we can find $\hat{\xi}$ being the Markov projection of X fulfilling:

$$d\hat{\xi}_t = \hat{b}(t, \xi_t)dt + \hat{g}(t)dW_t \tag{4.5}$$

so that, once again by Assumption 4.1, $\xi = \Theta^{-1}(\hat{\xi})$ solves:

$$d\xi_t = b(t, \xi_t)dt + g(t)dW_t \tag{4.6}$$

In the following, we prove such ξ is actually the Markov projection of S

Proposition 4.3. *ξ constructed as above is the Markov projection of S.*

Proof. It is easy to see that the one-dimensional marginals of S and ξ actually coincide: let $A \in \mathscr{B}(M)$

$$\mathbb{P}(S_t \in A) = \mathbb{P}(X_t \in \Theta(A)) = \\ = \mathbb{P}(\hat{\xi}_t \in \Theta(A)) = \\ = \mathbb{P}(\xi_t \in A)$$

Now we need to show that actually ξ is Markov. To do it, we need to show that for any bounded borel $f : W(M) \to \mathbb{R}^d$, $\mathbb{E}[f(\xi_t)|\mathcal{F}_s] = \int_{W(M)} f(x)p(s,t,\xi_s,dx)$ for a Markov transition function p.

Let $\hat{p}$ be the Markov transition function associated to $\hat{\xi}$, and let us define, for all $s, t \in [0, T]$ and $x \in M$:

$$p(s,t,x,\cdot) = (\theta^{-1}_{\#}\hat{p})(s,t,\theta(x),\cdot) \tag{4.7}$$

Then:

$$
\begin{aligned}
\mathbb{E}[f(\xi_t)|\mathcal{F}_s] = \mathbb{E}[(f\circ\theta^{-1})(\hat{\xi}_t)|\mathcal{F}_s] = \\
&= \int_{\mathbb{R}^d} (f\circ\theta^{-1})(z)\hat{p}(s,t,\hat{\xi}_s,dz) = \\
&= \int_M f(y)(\theta_{\#}^{-1}\hat{p})(s,t,\hat{\xi}_s,dy) = \\
&= \int_M f(y)p(s,t,\xi_s,dy)
\end{aligned}
$$

We only need to show that actually p, defined as above, is a Markov transition function. It is clear that $x \mapsto p(s,t,x,A)$ is measurable, being composition of $z \mapsto \hat{p}(s,t,z,\theta(A))$ and θ, both of which are measurable. Also, it is clear that $p(s,t,x,\cdot)$ is a valid probability measure, being the pushforward through θ^{-1} of a probability measure. We show that p fulfills the Chapman-Kolmogorov equation:

$$
\begin{aligned}
p(s,t,x,A) = \hat{p}(s,t,\theta(x),\theta(A)) = \\
&= \int_{\mathbb{R}^d} \hat{p}(u,t,y,\theta(A))\hat{p}(s,u,\theta(x),dy) = \\
&= \int_M \hat{p}(u,t,\theta(z),\theta(A))(\theta_{\#}\hat{p})(s,u,\theta(x),dz) = \\
&= \int_M p(u,t,z,A)p(s,u,x,dz)
\end{aligned}
$$

That is precisely the Chapman-Kolmogorov equation. □

To summarize, if we start with process $(S_t)_t$, with "conditional processes" Y^z, we can define $X = \Theta(S)$, obtaining that its "conditional processes" are precisely $\hat{Y}^z = \Theta(Y^z)$, leverage [SF]^{2}M on X to obtain its Markov projection $\hat{\xi}$ in terms of the solution to an SDE problem, and have the guarantee that $\xi = \Theta^{-1}(\hat{\xi})$ is the Markovization of S.

As done in Tong et al. [20], we might consider the important case where $Y^z = Y^{(z_0,z_T)}$ is a Brownian bridge from $z_0 \in N$ and $z_T \in N$. In such case, we already showed in Section 3.2.4 that $\hat{Y}^z = \Theta(Y^z)$ is just the usual Euclidean Brownian bridge from $\theta(z_0)$ to $\theta(z_T)$, and thus $X = \Theta(S)$ is a mixture of Euclidean Brownian bridges.

Not only, but if we require S to be a Schrödinger bridge from $\mathbb{Q}_0$ to $\mathbb{Q}_T$ with N-valued Brownian motion as reference, by what has been explained in Section 3.3.5 this amounts to ask that X is the Schrödinger bridge from $\hat{\mathbb{Q}}_0$ to $\hat{\mathbb{Q}}_T$, which is exactly the setting of Tong et al. [20]. Thus, in order to simulate the Markovization of an N-valued Schrödinger bridge, it is enough to obtain the Markov projection of its mapping through Θ (i.e.

processes X and $\hat{\xi}$, resp.), and simulate $\xi = \Theta^{-1}(\hat{\xi})$.

4.4. Entry-wise Behaviour

We make here a small detour that is useful to interpret SDEs on the manifold of SPD matrices. As we have established that an SDE on N is essentially an SDE on $\mathbb{R}^d$, a convenient way to think of it is as a set of SDEs stacked on a lower triangular matrix. Recall that, if we have an element tangent to N, this can be represented in terms of the manifold charts as lower triangular matrices (see Section 2.2.3 for details). If $B(t, S_t), V_\alpha(t, S_t) \in \mathcal{L}$ are the expression of $b(t, S_t)$ and $v_\alpha(t, S_t)$ in terms of the chart for $T_{S_t}N$ induced by θ, $b\theta^i(t, S_t) = B^i(t, S_t)$ and $v_\alpha\theta^i(t, S_t) = V^i_\alpha(t, S_t)$. So, if $\Theta_t \in \mathcal{L}$ is the stacking of $\theta^i(S_t)$ as lower triangular matrix, we have:

$$d\Theta_t = B(t, S_t)dt + V_\alpha(t, S_t) \circ dW^\alpha_t \tag{4.8}$$

meant as:

$$\begin{pmatrix} d\theta^1_t = B^1(t, S_t)dt & 0 & 0 & \dots & 0 \\ d\theta^{n+1}_t = B^{n+1}(t, S_t)dt & d\theta^2_t = B^2(t, S_t)dt & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & d\theta^n_t = B^n(t, S_t)dt \end{pmatrix} \tag{4.9}$$

where we omitted the diffusion component for space reasons.

Though this form is confortable to work with, it is not really telling us how the single entries of the matrix are evolving, but rather how their transformation is. In order to better interpret Equation (4.8), we only need to use Proposition 2.3. Indeed, if $\hat{B}, \hat{V}_\alpha \in \mathcal{L}$ are lower triangular matrix representations of the vector fields in terms of the (simpler) chart ϵ, Equation (4.8) is just:

$$dE_t = \hat{B}(t, S_t)dt + \hat{V}_\alpha(t, S_t) \circ dW^\alpha_t \tag{4.10}$$

where now E_t is simply the lower triangular part of S_t. More explicitely:

$$
\begin{aligned}
dE_t = &\Bigg\{\mathbb{D}(L_t)\mathbb{D}\Bigg(L_t B^T(t,S_t) + B(t,S_t)L_t^T\Bigg) + \Bigg\lfloor L_t B^T(t,S_t) + B(t,S_t)L_t^T \Bigg\rfloor\Bigg\}dt + \\
&+ \Bigg\{\mathbb{D}(L_t)\mathbb{D}\Bigg(L_t V_\alpha^T(t,S_t) + V_\alpha(t,S_t)L_t^T\Bigg) + \Bigg\lfloor L_t V_\alpha^T(t,S_t) + V_\alpha(t,S_t)L_t^T \Bigg\rfloor\Bigg\} \circ dW_t^\alpha
\end{aligned}
$$

where $L_t L_t^T = S_t$.

5 | Extension to Infinite Dimensional Hilbert Spaces

Once achieved an extension of $[SF]^2M$ to the log-Cholesky manifold, we aim in this chapter at extending the framework to the case of infinite-dimensional Hilbert spaces. Originally, Tong et al. [20] formulated the method in finite-dimensional Euclidean spaces; our previous extension was carried out on a manifold where the choice of the metric made it, in some sense, flat. Another case of interest where curvature is absent is precisely the one of infinite dimensional Hilbert spaces (e.g. $L^2(\Omega)$ functions over some compact set $\Omega \subset \mathbb{R}^N$).

To begin with, we will focus our attention to the restricted problem of finding the Markov projection (as the solution to a stochastic differential problem) of a mixture of Brownian bridges. The importance of this particular processes resides in the fact that, as explained in Föllmer and Gantert [8], also in infinite dimensions solutions to the Schrödinger bridge problem with Brownian motion as reference process are mixtures of Brownian bridges (meant as the very same Brownian motion conditioned on initial and final condition). More specifically, as a first attempt towards the declared goal, we properly define Brownian bridge over an infinite dimensional Hilbert space $\mathcal{H}$, and look for its characterization as the solution to a stochastic differential equation over the same space. This is a fundamental step for reproducing what has been done in Tong et al. [20] and explained in Section 1.3.

The idea to do it is essentially the following: we first state, in Section 5.1.1, the Karhunen-Loève expansion for $\mathcal{H}$-valued Wiener processes, and exploit it to write an infinite-dimensional Brownian bridge as the expansion of an infinite number of independent one dimensional Brownian bridges. Since it is known that 1D Brownian bridges solve a stochastic differential equation, we leverage this to show that formally the $\mathcal{H}$-valued BB solves a stochastic differential problem, which we will later make more precise. Moreover, we show that BB is the unique solution to the obtained equation, and that it satisfies a continuity condition for the final time. All of this is done in Section 5.1.3

All the material that this chapter assumes is explained in Prato and Zabczyk [19] and

Hairer [11].

5.1. Brownian Bridge on Infinite Dimensional Hilbert Space

Let us denote by $\mathcal{H}$ an infinite-dimensional Hilbert space, and denote by Q a positive-definite, trace class operator on it. Let $\tilde{W}$ be a Q-Wiener process on $\mathcal{H}$, and define W to be the $\mathcal{H}$-valued stochastic process defined as $W(t) = \tilde{W}(t) + W_0$, where W_0 is some element of $\mathcal{H}$ (of course $W(0) = W_0$ almost surely). Let us denote by U_0 the Cameron-Martin space of W (indeed, W can be realized as a Gaussian measure on some path space, since $\tilde{W}$ can, and W_0 is essentially a delta distribution, thus Gaussian, so its Cameron-Martin space is well defined).

Our goal is to study the properties of a Brownian bridge, meaning the process having the same law of $W(t)|(W(0) = a, W(T) = b)$ being $a, b \in \mathcal{H}$. Without loss of generality, we will let the final time always be $T = 1$.

5.1.1. Karhunen-Loève Expansion

By standard results in infinite-dimensional Wiener processes, we can write $\tilde{W}$ as:

$$\tilde{W}(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\beta_j(t)e_j \tag{5.1}$$

where $\left\{(\lambda_j, e_j)\right\}_{j=1}^{\infty}$ are eigenvalues and eigenvectors of Q forming an orthonormal basis for $\mathcal{H}$ and $\beta_j(t)$ are one dimensional independent Brownian motions. Now, since this is a basis, we can write $W_0 = \sum_{j=1}^{\infty} W_0^j e_j$ (all these sums converge in the L^2 sense). This means that:

$$W(t) = \sum_{j=1}^{\infty} \left(W_0^j + \sqrt{\lambda_j}\beta_j(t)\right)e_j = \sum_{j=1}^{\infty} c_j(t)e_j \tag{5.2}$$

where now $c_j(t)$ are independent Brownian motions on $\mathbb{R}$ with diffusion rate $\sqrt{\lambda_j}$ and starting at W_0^j.

5.1.2. One Dimensional Brownian Bridge

In one dimension, one can characterize Brownian bridges as solutions to a specific SDE:

$$\begin{cases} dX_t^j = u^j(t, X_t^j)dt + \sqrt{\lambda_j}dB_t^j, & t \in (0,1) \\ X_0^j = a_j \end{cases} \tag{5.3}$$

and $X_t^j \to b_j$ as $t \to 1^-$ in L^2. Here we defined $u^j(t,x)$ as:

$$u^j(t,x) = \frac{b_j - x}{1-t}$$

and B_t^j is a Brownian Motion on $\mathbb{R}$.

5.1.3. Deduction of an $\mathcal{H}$-SDE

Now consider $l \in \mathcal{H}$ and let X_t be the Brownian bridge over $\mathcal{H}$, which is expressible as:

$$X_t = \sum_{j=1}^{\infty} X_t^j e_j \tag{5.4}$$

being X_t^j a one dimensional Brownian bridge from a_j to b_j at times 0 and 1 respectively.

We will let $A_n(t)$ be the partial sum defining $X(t)$ up to the n^{th} component. We test X_t against l:

$$\begin{aligned}
\langle l, X_t \rangle &= \sum_{j=1}^{\infty} \langle l_j, X_t^j \rangle = \\
&= \sum_{j=1}^{\infty} l_j (a^j + \int_0^t u_j(s, X_s^j)ds + \sqrt{\lambda_j} B_t^j) = \\
&= \sum_{j=1}^{\infty} l_j a_j + \sum_{j=1}^{\infty} l_j B_t^j + \sum_{j=1}^{\infty} l_j \int_0^t u_j(s, X_s^j)ds = \\
&= \langle l, a \rangle + \langle l, B_t \rangle + \int_0^t \langle l, u_j(s, X_s^j) \rangle ds
\end{aligned}$$

at least formally, this is the weak solution to an equation on $\mathcal{H}$:

$$\begin{cases} dX_t = dB_t + u(t, X_t)dt, \quad t \in (0,1) \\ X_0 = a \end{cases} \tag{5.5}$$

where we define:

$$u(t,x) = \frac{b-x}{1-t}$$

for $x \in \mathcal{H}$.

The following issues arise:

1. Does the above written equation make sense?
2. Does it have a unique *weak solution*? In Prato and Zabczyk [19], they claim existence and uniqueness of *mild solutions* for this type of equations, but in this particular case we can prove equivalence of the two;
3. Do we still have $X_t \to b$ as $t \to 1^-$ in L^2?

First of all, let us recall the following theory from Prato and Zabczyk [19].

Consider the general nonlinear equation:

$$\begin{cases} dX = (AX + F(t,X))dt + B(t,X)dW(t) \quad ,t \in (0,T] \\ X(0) = \xi \end{cases} \tag{5.6}$$

Assume that we are given a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_t, \mathbb{P})$, call $\mathscr{P}_T$ the predictable σ-algebra on $\Omega_T = [0,T] \times \Omega$. Moreover, for any $T > 0$, define $\mathbb{P}_T = \lambda_T \otimes \mathbb{P}$, being λ_T the Lebesgue measure on $[0,T]$. Also, we denote by U_0 the Cameron-Martin space of the Q-Wiener process W, and L_2^0 the space of Hilbert-Schmidt operators from U_0 to $\mathcal{H}$.

We assume the following assumptions hold true:

Assumption 5.1. 1. A is generator of a C_0-semigroup $S(t)$ in $\mathcal{H}$;

2. The map
$$F : [0,T] \times \Omega \times \mathcal{H} \to \mathcal{H}, \quad (t, \omega, x) \mapsto F(t, \omega, x)$$
is measurable from $(\Omega_T \times \mathcal{H}, \mathscr{P}_T \times \mathscr{B}(\mathcal{H}))$ to $(\mathcal{H}, \mathscr{B}(\mathcal{H}))$;

3. The mapping

$$B : [0,T] \times \Omega \times \mathcal{H} \to L_2^0, \quad (t, \omega, x) \mapsto B(t, \omega, x)$$

   is measurable from $(\Omega_T \times \mathcal{H}, \mathscr{P}_T \times \mathscr{B}(\mathcal{H}))$ to $(L_2^0, \mathscr{B}(L_2^0))$;

4. F and B are Lipschitz and satisfy a linear growth bound condition.

We now define what we mean by solution to this problem.

Definition 5.1 (Weak Solution). We say that the predictable process X is a **weak solution** if $\int_0^T \|X(s)\|_{\mathcal{H}} ds < +\infty$ and for all $l \in \mathcal{H}$ and all $t \in [0,T]$ the following holds true:

$$\langle l, X(t)\rangle = \langle l, \xi\rangle + \int_0^t \langle A^* l, X(s)\rangle ds + \int_0^t \langle l, F(s, X(s))\rangle ds + \int_0^t \langle B^*(s, X(s)) l, dW(s)\rangle$$

$\mathbb{P}$-a.s.

Definition 5.2 (Mild Solution). We say that the predictable process X is a **mild solution** if $\int_0^T \|X(s)\|_{\mathcal{H}}^2 ds < +\infty$ and for all $t \in [0,T]$ we have:

$$X(t) = S(t)\xi + \int_0^t S(t-s) F(s, X(s)) ds + \int_0^t S(t-s) B(s, X(s)) dW(s)$$

$\mathbb{P}$-a.s.

The following is a result about the existence and uniqueness of mild solutions to the problem.

Theorem 5.1 (Theorem 7.2 in [19]). *Assume ξ is an $\mathscr{F}_0$-measurable $\mathcal{H}$-valued random variable and that Assumption 5.1 are satisfied. Then there exists a mild solution X that is unique, up to equivalence, among the processes such that $\int_0^T \|X(s)\|_{\mathcal{H}}^2 ds < +\infty$ $\mathbb{P}$-almost surely. Moreover, X possesses a continuous modification.*

In our specific case of interest, $A = 0$, so that $S(t) = \mathrm{Id}_{\mathcal{H}}$, $F(t,x) = b(t,x) = \frac{b-x}{1-t}$, B is simply the embedding $J : U_0 \to \mathcal{H}$, which is Hilbert-Schmidt (i.e. once we choose Q to be the covariance operator, the Cameron-Martin space U_0 is uniquely determined, and there exists a bigger Hilbert space $\mathcal{H}$ such that the embedding of U_0 on $\mathcal{H}$ is Hilbert-Schmidt and the Q-Wiener process is supported on $\mathcal{H}$). Furthermore, $\xi = a$ $\mathbb{P}$-almost surely. It is then clear that Assumption 5.1 is satisfied, and that Equation (5.5), which we derived only formally, falls under the described general framework.

In this case, a weak solution will satisfy:

$$\langle l, X(t)\rangle = \langle l, a\rangle + \int_0^t \langle l, u(s, X(s))\rangle ds + \int_0^t \langle J^* l, dW(s)\rangle$$

whereas a mild solution must satisfy:

$$X(t) = a + \int_0^t u(s, X(s))ds + \int_0^t JdW(s)$$

In the following, we will prove (in Lemma 5.1 and Proposition 5.1) that if (5.5) admits a weak solution, this is unique, and that $X(t)$ (defined through the expansion) has time-integrable squared norm.

From this, we will conclude that $X(t)$ is the unique weak solution to (5.5). Indeed, notice that, by how we derived Equation (5.5), $X(t)$ must solve it weakly (once again, as $\|X(t)\|^2_{\mathcal{H}}$ is integrable, also $\|X(t)\|_{\mathcal{H}}$ must be because $L^2([0,1]) \subset L^1([0,1])$). But then $X(t)$ is the unique weak solution to Equation (5.5). This means that *on a separable Hilbert space, Brownian bridges are unique weak solutions to certain stochastic differential equations on the space.*

Lemma 5.1. *If a weak solution to* (5.5) *exists, then it is the unique one, up to equivalence, among the processes such that* $\int_0^T \|X(s)\|_{\mathcal{H}} ds < +\infty$ $\mathbb{P}$ *almost surely.*

The proof is very similar to the proof of the uniqueness of mild solutions in Prato and Zabczyk [19].

Proof.* Suppose $X_1(t)$ and $X_2(t)$ are predictable processes on $\mathcal{H}$ solving (5.5). We will prove that $\mathbb{P}(X_1(t) = X_2(t)) = 1$ $\forall t \in [0, T]$.

First, let $R > 0$ be fixed, and define the stopping times:

$$\tau_i = \sup\left\{t \le T : \int_0^T |\langle l, b(s, X(s))\rangle| ds \ge R\right\}, \quad i = 1, 2$$
$$\tau = \min(\tau_1, \tau_2)$$

Now, define:

$$\hat{X}_i(t) = \mathbf{1}_{[0,\tau]}(t) X_i(t), \quad i = 1, 2$$

This way, $\mathbb{P}$-a.s.

$$\langle l, \hat{X}_i(t)\rangle = \mathbf{1}_{[0,\tau]}(t)\langle l, a\rangle + \mathbf{1}_{[0,\tau]}(t)\int_0^t \langle l, b(s, \hat{X}_i(s))\rangle ds + \mathbf{1}_{[0,\tau]}(t)\int_0^t \langle J^* l, dW(s)\rangle$$

We now show that $\mathbb{E}[|\langle l, \hat{X}_1(t)\rangle - \langle l, \hat{X}_2(t)\rangle|]$ for every $t \in [0, T]$ and every $l \in \mathcal{H}$.

Indeed:

$$\begin{aligned}\langle l, \hat{X}_1(t)\rangle - \langle l, \hat{X}_2(t)\rangle = \langle l, \int_0^t \frac{1}{1-s}(\hat{X}_2(s) - \hat{X}_1(s))ds\rangle = \\ = \int_0^t \frac{1}{1-s}\langle l, (\hat{X}_2(s) - \hat{X}_1(s))\rangle ds\end{aligned}$$

From which:

$$\mathbb{E}[|\langle l, \hat{X}_1(t)\rangle - \langle l, \hat{X}_2(t)\rangle|] \leq \frac{1}{1-T}\int_0^T \mathbb{E}[|\langle l, \hat{X}_1(s) - \hat{X}_2(s)\rangle|]ds$$

Since the expectation is bounded by the definition of $\hat{X}_i(t)$ and the one of stopping times, using Gronwall inequality we conclude that $\mathbb{E}[|\langle l, \hat{X}_1(t)\rangle - \langle l, \hat{X}_2(t)\rangle|] = 0$ for all t and all l.

From this, we get that for all $l \in \mathcal{H}$ and all $t \in [0, T]$, $\mathbb{P}(\langle l, \hat{X}_1(t)\rangle = \langle l, \hat{X}_2(t)\rangle) = 1$. Fixed $l \in \mathcal{H}$, this implies that $\langle l, \hat{X}_1\rangle = \langle l, \hat{X}_2\rangle$ $\mathbb{P}_T$-a.s., and since R is arbitrary, $\langle l, X_1\rangle = \langle l, X_2\rangle$ $\mathbb{P}_T$-a.s.. Now, since $\langle l, X_i\rangle$ satisfies the integral equation of the definition of weak solution, we conclude that:

$$\mathbb{P}(\langle l, X_1(t)\rangle = \langle l, X_2(t)\rangle) = 1, \quad \forall t \in [0, T],\ \forall l \in \mathcal{H}$$

Notice that, if we choose any basis $\{e_j\}_{j=1}^{\infty} \subset \mathcal{H}$, this implies:

$$\mathbb{P}(\langle e_j, X_1(t)\rangle = \langle e_j, X_2(t)\rangle) = 1, \quad \forall t \in [0, T],\ \forall j = 1, 2, ...$$

But then, let $N_j = \{\omega \in \Omega : X_1^j(t) \neq X_2^j(t)\}$ where we defined $X_i^j(t)$ to be the j^{th} component of $X_i(t)$ with respect to e_j. Now, $\mathbb{P}(N_j) = 0$, and thus:

$$\mathbb{P}(X_1^j(t) = X_2^j(t), \quad \forall j = 1, 2, ...) = 1, \quad \forall t \in [0, T]$$

which is equivalent to:

$$\mathbb{P}(X_1(t) = X_2(t)) = 1, \quad \forall t \in [0, T]$$

which is the claim. □

Proposition 5.1. *$\mathbb{P}$-almost surely, the following holds:*

$$\int_0^T \|X(s)\|_{\mathcal{H}}^2 ds < +\infty \tag{5.7}$$

Proof.* First of all, let us show that the function $t \mapsto \mathbb{E}[\|X(t)\|_{\mathcal{H}}^2]$ is integrable on $[0, T]$.

Since convergence in L^2 implies convergence of the norms, we know that $\mathbb{E}[\|X(t)\|_{\mathcal{H}}^2] = \lim_n \sum_{j=1}^n \mathbb{E}[\|X(t)^j\|_{\mathcal{H}}^2] = \sum_{j=1}^\infty \mathbb{E}[\|X(t)^j\|_{\mathcal{H}}^2] = \sum_{j=1}^\infty t\lambda_j = tM$, which is integrable.

Then, by Fubini theorem, we get:

$$\mathbb{E}\left[\int_0^T \|X(t)\|_{\mathcal{H}}^2 dt\right] = \int_0^T \mathbb{E}[\|X(t)\|_{\mathcal{H}}^2] dt < +\infty$$

meaning that almost surely $\int_0^T \|X(t)\|_{\mathcal{H}}^2 dt < +\infty$, as claimed. □

For the last of the three questions, let us show that the following limit holds in the L^2 sense:

$$\lim_{t \to 1^-} \lim_{n \to \infty} A_n(t) = b$$

Now:

$$\|A_n(t) - b\|_{L^2} \le \|A_n(t) - b_n\|_{L^2} + \|b_n - b\|_{L^2}$$

where b_n is the n^{th} partial sum for the expansion of b. Being b the L^2 limit of $(b_n)_n$, the second element converges to 0 as we take the limits (being independent of t, in this case the limit reduces to one limit).

For the first one, recall that being X^j a 1D Brownian bridge, $\mathbb{E}[(X_t^j - b_j)^2] = (1-t)^2(a_j - b_j)^2 + t(1-t)\lambda_j$, which converges to 0 as t goes to 1. Thus:

$$\mathbb{E}\left[\sum_{j=0}^{n}(X_t^j - b_j)^2\right] = \sum_{j=1}^{n}\mathbb{E}\left[(X_t^j - b_j)^2\right] =$$
$$= (1-t)\sum_{j=1}^{n}[(1-t)(a_j-b_j)^2 + t\lambda_j]$$

But:

$$\sum_{j=1}^{n}(a_j-b_j)^2 \to \sum_{j=1}^{\infty}(a_j-b_j)^2 = \|a-b\|_{\mathcal{H}}^2 \leq M_1 < +\infty$$

and

$$\sum_{j=1}^{n} t\lambda_j = t\sum_{j=1}^{n}\lambda_j \to t\sum_{j=1}^{\infty}\lambda_j \leq M_2 < +\infty$$

And finally:

$$\mathbb{E}\left[\|A_n(t) - b_n\|_{\mathcal{H}}^2\right] \to_n (1-t)M \to_t 0$$

as required.

We can summarize as follows:

Theorem 5.2. *Under the above notations, the Brownian bridge $X(t)$ over $\mathcal{H}$ is the unique weak solution to:*

$$\begin{cases} dX_t = dB_t + u(t,X_t)dt, & t\in(0,1) \\ X_0 = a \end{cases}$$

where we define:

$$u(t,x) = \frac{b-x}{1-t}$$

Moreover, $X_t \to b$ as $t \to 1^-$ in L^2.

6 | Overview

This thesis developed a mathematical extension of the $[SF]^2M$ framework to the manifold of symmetric positive definite matrices endowed with the log-Cholesky metric. To achieve this goal, we first revisited the notion of Markov projection for mixtures of stochastic processes, providing a formulation that is suitable for manifold-valued settings and establishing sufficient conditions under which the Markovization can be recovered. In particular, these conditions were shown to hold for mixtures of Brownian bridges, which are central to the Schrödinger bridge problem.

Building upon this foundation, we characterized the log-Cholesky manifold through its global isometric diffeomorphism with Euclidean space and exploited this structure to develop the necessary stochastic analysis tools on the manifold. This allowed us to establish manifold analogues of Brownian motion, Brownian bridges, and Schrödinger bridges, and ultimately to construct a complete mathematical framework for $[SF]^2M$ on the log-Cholesky manifold. Finally, we investigated an extension of the theory to infinite-dimensional Hilbert spaces by deriving stochastic evolution equation formulations for infinite-dimensional Brownian bridges, laying the groundwork for a future infinite-dimensional version of the model.

Despite the advantages offered by the log-Cholesky geometry, the present work is constrained by the choice of metric. The log-Cholesky metric provides a particularly convenient setting because it induces a global isometric diffeomorphism with Euclidean space, greatly simplifying both the geometric and stochastic analysis. However, this simplicity may come at the cost of expressive power when compared with alternative geometries on the SPD manifold.

In particular, metrics such as the log-Euclidean metric ([14]) or the Bures–Wasserstein metric (see Bhatia et al. [2]) may provide more meaningful descriptions of the underlying structure of SPD-valued data. The latter is especially appealing because of its close connection to optimal transport and its potential compatibility with infinite-dimensional extensions (see Pigoli et al. [18]). Nevertheless, extending the present framework to such geometries would require a substantially more sophisticated treatment of manifold-valued stochastic analysis, since the convenient Euclidean correspondence available under the log-Cholesky metric would no longer be directly exploitable.

Several natural directions emerge from this work. A first objective is to complete the infinite-dimensional extension outlined in Chapter 5 by fully generalizing the Markov projection framework and the associated Schrödinger bridge constructions to Hilbert spaces. Such an extension would considerably broaden the range of applications of $[SF]^2M$, allowing the model to operate on functional data and other infinite-dimensional objects.

A second direction consists of investigating alternative geometries on the SPD manifold, including the log-Euclidean and Bures–Wasserstein metrics, and studying how the resulting stochastic dynamics differ from those obtained under the log-Cholesky metric.

Finally, from an applied perspective, it would be valuable to implement and evaluate the proposed framework on structural and functional connectivity data. Such experiments would require substantial computational resources, as these connectivities are typically derived from time series associated with large neuroimaging datasets, but they would provide an important assessment of the practical relevance of the theoretical developments presented in this thesis.

Bibliography

[1] P. Baldi. *Stochastic Calculus: An Introduction Through Theory and Exercises*. Universitext. Springer International Publishing, Cham, Switzerland, 2017. ISBN 978-3-319-62225-5, 978-3-319-62226-2. doi: 10.1007/978-3-319-62226-2. URL https://doi.org/10.1007/978-3-319-62226-2.

[2] R. Bhatia, T. Jain, and Y. Lim. On the bures–wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019. ISSN 0723-0869. doi: 10.1016/j.exmath.2018.01.002. URL https://doi.org/10.1016/j.exmath.2018.01.002.

[3] A. Bucci, L. Ippoliti, and P. Valentini. Comparing unconstrained parametrization methods for return covariance matrix prediction. *Statistics and Computing*, 32(90), 2022. doi: 10.1007/s11222-022-10157-4.

[4] R. Caseiro, P. Martins, J. F. Henriques, and J. Batista. A nonparametric riemannian framework on tensor field with application to foreground segmentation. *Pattern Recognition*, 45(11):3997–4017, 2012. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2012.04.011. URL https://www.sciencedirect.com/science/article/pii/S0031320312001689.

[5] R. T. Q. Chen and Y. Lipman. Riemannian flow matching on general geometries. *CoRR*, abs/2302.03660, 2023. doi: 10.48550/arXiv.2302.03660. URL https://arxiv.org/abs/2302.03660.

[6] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

[7] P. Fillard, V. Arsigny, X. Pennec, K. M. Hayashi, P. M. Thompson, and N. Ayache.

Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *NeuroImage*, 34(2):639–650, 2007. doi: 10.1016/j.neuroimage.2006.09.027.

[8] H. Föllmer and N. Gantert. Entropy minimization and Schrödinger processes in infinite dimensions. *The Annals of Probability*, 25(2):901 – 926, 1997. doi: 10.1214/aop/1024404423. URL https://doi.org/10.1214/aop/1024404423.

[9] K. J. Friston. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36, 2011. doi: 10.1089/brain.2011.0008.

[10] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an itô differential. *Probability Theory and Related Fields*, 71(4):501–516, 1986. doi: 10.1007/BF00699039. URL https://doi.org/10.1007/BF00699039.

[11] M. Hairer. An introduction to stochastic pdes, 2023. URL https://arxiv.org/abs/0907.4178.

[12] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

[13] E. P. Hsu. Stochastic analysis on manifolds. Online lecture notes, 2025. URL https://sayanmuk.github.io/StochasticAnalysisManifolds.pdf.

[14] Z. Huang, R. Wang, S. Shan, X. Li, and X. Chen. Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification. In F. Bach and D. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 720–729, Lille, France, July 2015. PMLR. URL https://proceedings.mlr.press/v37/huang15.html.

[15] J. M. Lee. *Smooth Manifolds*, pages 1–29. Springer New York, New York, NY, 2003. ISBN 978-0-387-21752-9. doi: 10.1007/978-0-387-21752-9_1. URL https://doi.org/10.1007/978-0-387-21752-9_1.

[16] Z. Lin. Riemannian geometry of symmetric positive definite matrices via cholesky decomposition. *arXiv e-prints*, Aug. 2019. URL https://arxiv.org/abs/1908.09326. 19 pages.

[17] C. Léonard. A survey of the schrödinger problem and some of its connections with optimal transport, 2014. ISSN 1078-0947. URL https://www.aimsciences.org/article/id/d5bcf817-901d-4104-b7da-eade7847c53e.

[18] D. Pigoli, J. A. D. Aston, I. L. Dryden, and P. Secchi. Distances and inference for covariance operators. *Biometrika*, 101(2):409–422, 2014. doi: 10.1093/biomet/ast074.

[19] G. D. Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 2 edition, 2014.

[20] A. Tong, N. Malkin, K. Fatras, L. Atanackovic, Y. Zhang, G. Huguet, G. Wolf, and Y. Bengio. Simulation-free schrödinger bridges via score and flow matching, 2024. URL `https://arxiv.org/abs/2307.03672`.

[21] Z. Zhou, T. Dan, and G. Wu. Brainflow: A holistic pathway of dynamic neural system on manifold. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL `https://openreview.net/forum?id=fkYZKArFLF`.

A | Differential Geometry Aspects

A.1. Review of Differential Geometry concepts

In this section, we review some basic concepts of differential geometry that will be useful in the following. All the material is taken from Lee [15].

In Section A.1.1, we introduce the notion of smooth manifold, starting from topological manifolds. In Section A.1.2, we define smooth functions/maps, which are central objects in the study of differential geometry. We conclude with Section A.1.3, where tangent vectors are introduced, and the specific example of Cholesky space is considered.

A.1.1. Smooth Manifolds

A smooth manifold is simply a topological manifold with a supplementary structure that lets us do calculus (in particular, take derivatives).

Definition A.1. Let M be a topological space. M is an n-dimensional **topological manifold** if:

- M is Hausdorff ($\forall p, q \in M$, we have U, V open sets such that $p \in U$ and $q \in V$);
- M is second-countable (there is a countable basis for its topology);
- M is locally euclidean of dimension n ($\forall p \in M$, there is $U \subset M$ open such that $p \in U$ and $\phi : U \to \mathbb{R}^n$ is a homomorphism; denote $\phi(U) = \hat{U}$).

Note that even though M, N are only topological manifolds, the notion of *continuous function* exists, because $f : M \to N$ is continuous iff $\forall O \subset N$ open, $f^{-1}(O) \subset M$ is open. However, we cannot say what *smooth function* means (i.e. $C^\infty(M)$). For this, we need some more structure.

Definition A.2. A couple (U, ϕ) as described above is called **local chart**.

The idea of differential geometry is to use local charts to move all computations on (open subsets of) $\mathbb{R}^n$, where everything is well defined. However, we need to ensure charts are compatible one with the other.

Definition A.3. Let (U, ϕ) and (V, ψ) be two local charts.

- If $U \cap V = \emptyset$, the two charts are **smoothly compatible**;
- If $U \cap V \neq \emptyset$, they are **smoothly compatible** if the **transition map** $\psi \circ \phi^{-1} : \phi(U \cap V) \to \psi(U \cap V)$ is a diffeomorphism.

Of course, we need to be able to describe all M. That is why we need the notion of *atlas*.

Definition A.4. $\mathcal{A}$ is an **atlas** of M if it is a collection of local charts covering M. It is a **smooth atlas** if taken any two of its charts, they are smoothly compatible.

In general, a smooth atlas is not enough to make of a topological manifold a differential one. In some sense, we want the atlas to contain all the possible smoothly compatible charts.

Definition A.5. $\mathcal{A}$ is a **smooth structure** if it is a maximal smooth atlas (any chart smoothly compatible with all charts in $\mathcal{A}$ is already in $\mathcal{A}$).

Definition A.6. Let M be a topological manifold, and let $\mathcal{A}$ be a smooth structure on it. $(M, \mathcal{A})$ is a **smooth manifold**. If the smooth structure is understood, then we just write M.

Example A.1 (Finite Dimensional Vector Space)**.** Let V be a finite dimensional vector space. We want to make a smooth manifold out of it.

First of all, we need to make of it a topological manifold. Take any norm on V: it will yield a topology that is independent of its choice. With such topology, V is a topological manifold.

Now we define charts. Take an ordered basis $E_1, ..., E_n$, and define an isomorphism by $E : \mathbb{R}^n \to V$ as:

$$E(x) = x^i E_i \tag{A.1}$$

for $x \in V$. Being a homeomorphism, (V, E^{-1}) is a (global) chart.

What charts are compatible with this? Take any other basis $\tilde{E}_1, ..., \tilde{E}_n$, and let $\tilde{E}$ be the corresponding isomoprhism. There exists an invertible matrix (A_i^j) such that:

$$E_i = A_i^j \tilde{E}_j \tag{A.2}$$

and the transition map is $\tilde{x} = \tilde{E}^{-1} \circ E(x)$, from which we get that $\tilde{x}^j = x^i A_i^j$. This is linear, thus it is a diffeomorphism.

The collection of all such charts is a maximal atlas, and is a smooth strcuture on V. It is called the **standard smooth structure** on V and it makes of V a smooth manifold. □

Example A.2 (Lower Triangular Matrices)**.** Denote with $\mathcal{L}_n$ the space of $n \times n$ matrices that are lower triangular. It is a vector space, and thus is a smooth manifold with the standard smooth structure on it. □

A.1.2. Smooth Functions and Smooth Maps

Notation We denote with the word *function* an object $f : M \to \mathbb{R}^k$ for some $k \in \mathbb{N}$, while we say f is a *map* in all other cases (e.g. $f : M \to N$, being M and N smooth manifolds).

Definition A.7. Let M be a smooth manifold, and $f : M \to \mathbb{R}^k$ be a function. f is a **smooth function** if $\forall p \in M$, there exists a chart containing p, (U, ϕ), such that the function $f \circ \phi^{-1} : \hat{U} \to \mathbb{R}^k$ is smooth in the classical sense (i.e. $C^\infty(\hat{U})$).

Often it is useful to view functions on M as (locally) functions on the open sets of $\mathbb{R}^n$ provided by the local charts.

Definition A.8. If $f : M \to \mathbb{R}^k$ is a smooth function,its **coordinate representations** are $\hat{f} = f \circ \phi^{-1} : \hat{U} \to \mathbb{R}^k$, for all local charts (U, ϕ).

Notice that the definition of smooth function is independent of the choice of the chart on each point, being charts smoothly compatible in the smooth structure that makes of M a smooth manifold.

We now pass to the case of maps between manifolds, and get a definition of diffeomorphism between smooth manifolds.

Definition A.9. Let M, N be smooth manifolds, and $F : M \to N$ a map. F is a **smooth map** if $\forall p \in M$, and chosen any charts (U, ϕ) containing p and (V, ψ) containing $F(p) \in N$, the map $\psi \circ F \circ \phi^{-1} : \hat{U} \to \hat{V}$ is smooth in the usual sense of maps $\mathbb{R}^m \to \mathbb{R}^n$ (recall indeed that $\hat{U} \subseteq \mathbb{R}^m$ and $\hat{V} \subseteq \mathbb{R}^n$ if M is m-dimensional and N is n-dimensional).

Since M and N are topological manifolds, the notion of continuous function exists. A natural question arising is: if F is a smooth map, is it also going to be continuous in the topological sense? The answer is yes.

Proposition A.1. *A smooth map is also continuous in the topological sense.*

We now give the definition of diffeomorphism.

Definition A.10. A smooth map between smooth manifolds $F : M \to N$ is a **diffeomorphism** if it is invertible and its inverse $F^{-1} : N \to M$ is a smooth map.

Note. To check whether $F : M \to N$ is smooth, one has, by definition, to check that coordinate representations $\psi \circ F \circ \phi^{-1} : \hat{U} \to \hat{V}$ are smooth. But for functions $\mathbb{R}^m \to \mathbb{R}^n$, checking smoothness amounts to check that the component functions are C^∞.

Example A.3 (Diffeomorphism between $\mathcal{S}_n^+$ and $\mathcal{L}_n^+$)**.** As explained in Lin [16], the map between the smooth manifold of $n \times n$ SPD matrices $\mathcal{S}_n^+$ and the one of $n \times n$ lower triangular matrices with positive diagonal entries $\mathcal{L}_n^+$ is a diffeomorphism. Such map is denoted by $\mathscr{L}$. In order to prove it, they leverage precisely the above note.

In particular, call $\phi : \mathcal{S}_n^+ \to \mathbb{R}^d$ the global chart for $\mathcal{S}_n^+$ assigning to each SPD matrix the vectorization of its lower triangular part ($d = \frac{n(n+1)}{2}$), and let $\psi : \mathcal{L}_n^+ \to \mathbb{R}^d$ denote the global chart for $\mathcal{L}_n^+$ (referred to as *Cholesky space*) doing the exact same thing. Now, take $P = [P_{ij}]_{i \geq j} \in \mathcal{S}_n^+$, with:

$$\begin{aligned} \phi(P) &= (P_1, P_2, ..., P_n, P_{n+1}, ..., P_d)^T = \\ &= (P_{11}, P_{22}, ..., P_{nn}, P_{21}, ..., P_{n(n-1)})^T \end{aligned}$$

and $\mathscr{L}(P) = L = [L_{ij}]_{i \geq j} \in \mathcal{L}_n^+$, with:

$$\begin{aligned} \psi(L) &= (L_1, L_2, ..., L_n, L_{n+1}, ..., L_d)^T = \\ &= (L_{11}, L_{22}, ..., L_{nn}, L_{21}, ..., L_{n(n-1)})^T \end{aligned}$$

we can prove $\mathscr{L}$ is a diffeomorphism by proving that the map is invertible and (using the more agile and intuitive notation $L_i(P_1, P_2, ..., P_d)$ instead of $\psi(\mathscr{L}(\phi^{-1}(P)))_i$)

that $L_i(P_1, P_2, ..., P_d) \in C^\infty(\phi(\mathcal{S}_n^+))$ for all $i \in 1, 2, ..., d$ and that $P_i(L_1, L_2, ..., L_d) \in C^\infty(\psi(\mathcal{L}_n^+))$ for all $i \in 1, 2, ..., d$. $\square$

A.1.3. Tangent Vectors

The idea behind defining and working with tangent vectors is the one of linear approximation. In some sense, the tangent space to a manifold represents a linear model for it close to a point. Not only, but once we define the tangent space, we can extend the concept of total derivative of a function (defined for functions $\mathbb{R}^m \to \mathbb{R}^n$) to the case of functions between manifolds, introducing the concept of differential.

This part is organized as follows. In Section A.1.3, we consider the more simple case of geometric tangent vectors in $\mathbb{R}^n$, being the most intuitive objects one can always have in mind as a conceptual model for more abstract tangent vectors; in Section A.1.3 we introduce the notion of *derivation* in euclidean spaces: these are objects that can be generalized to get tangent vectors to manifolds. In Section A.1.3, we claim that the set of geometric tangent vectors and the one of derivations are isomorphic and thus deduce a basis for the latter (being a basis of the former very easy to obtain). We then proceed to define tangent vectors on smooth manifolds in Section A.1.3, and discuss the notion of differential of a function in Section A.1.3. Finally, in Section A.1.3 we see how all these objects can be described in local coordinates.

Geometric Tangent Vectors

Geometric tangent vectors are thought of as *rows attached to points* in some euclidean space $\mathbb{R}^n$.

Definition A.11. Let $a \in \mathbb{R}^n$. A **geometric tangent vector to** $\mathbb{R}^n$ **at** a is an element of $\{a\} \times \mathbb{R}^n = \{(a, v), v \in \mathbb{R}^n\}$. Such space is denoted with $\mathbb{R}^n_a$, and its elements will be written as v_a or $v|_a$.

Under the following natural operations, $\mathbb{R}^n_a$ is a vector space:

$$(v + w)_a = v_a + w_a$$

and

$$(cv)_a = c(v_a)$$

for $c \in \mathbb{R}$.

Example A.4 (Spherical Surface in $\mathbb{R}^3$)**.** Take a spherical surface S in $\mathbb{R}^3$ and a point p at "the top" of the surface. Essentially, $\mathbb{R}^3_p$ is just a copy of $\mathbb{R}^3$ whose vectors have tail in p. One can think of the tangent space to S in p simply as a subspace (the (x, y)-plane) of $\mathbb{R}^3_p$. □

Derivations

The question is: how can we generalize the concept of tangent vectors to smooth manifolds. It is quite clear that a construction as the one of Example A.4 cannot be easily carried out in a general smooth manifold. The idea is that on smooth manifolds we only know what smooth maps/functions are, so we want to come up with an action that geometric tangent vectors do on smooth functions in $\mathbb{R}^n$, in order to generalize it to the smooth manifold case. Such action is essentially the directional derivative, and is expressed by the notion of *derivation*.

Take a geometric tangent vector of $\mathbb{R}^n_a$, $v|_a$, and define the functional $D_{v|_a} : C^\infty(\mathbb{R}^n) \to \mathbb{R}$ that takes the directional derivative along v at a, i.e.

$$D_{v|_a}(f) = D_v f(a) = \left.\frac{d}{dt}\right|_{t=0} f(a + tv) \tag{A.3}$$

or, in components:

$$D_{v|_a}(f) = v^i \frac{\partial f}{\partial x^i}(a) \tag{A.4}$$

where of course v^i is the i^{th} component of v with respect to the canonical basis on $\mathbb{R}^n$.

Definition A.12. Let $a \in \mathbb{R}^n$. A map $w : C^\infty(\mathbb{R}^n) \to \mathbb{R}$ is a **derivation at** a if both the following hold true:

- w is $\mathbb{R}$-linear;

- w satisfies the product rule:

$$w(fg) = f(a)wg + g(a)wf \tag{A.5}$$

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

We denote with $T_a\mathbb{R}^n$ the set of all derivations of $C^\infty(\mathbb{R}^n)$ functions at a.

Once again, $T_a\mathbb{R}^n$ is an $\mathbb{R}$-vector space under the operations of sum of maps and multiplication of a map by a scalar.

Isomorphism

The reason why we can intuitively think of elements of $T_a\mathbb{R}^n$ as *rows attached to a point* is the following result.

Proposition A.2. *The map $v_a \to D_{v|_a}$ is an isomorphism from $\mathbb{R}^n_a$ to $T_a\mathbb{R}^n$.*

Let $a \in \mathbb{R}^n$. The following n derivations are a basis for $T_a\mathbb{R}^n$ (thus its dimension is n): $\frac{\partial}{\partial x^i}\big|_a$.

The previous result simply follows from the fact that $e_i|_a$ is a basis for $\mathbb{R}^n_a$, and through the isomorphism we can get the claimed basis for the tangent space.

Tangent Vectors on Smooth Manifolds

We are now ready to give the definition of tangent space for smooth manifolds, by simply generalizing the concept of derivation.

Definition A.13. Let $p \in M$, being M a smooth manifold. A map $v : C^\infty(M) \to \mathbb{R}$ is a **derivation at** p if the two following conditions hold true:

- v is $\mathbb{R}$-linear;
- $v(fg) = f(a)vg + g(a)vf$ for all $f, g \in C^\infty(M)$ (i.e. smooth functions).

The set of derivations at p is the **tangent space to** M **at** p, and is denoted by T_pM. Finally, its elements are called **tangent vectors to** M **at** p.

Differential of a Function

Before introducing the differential of a function on smooth manifolds, let us recall the situation in $\mathbb{R}^n$.

What happens in $\mathbb{R}^n$. Take a function $f : \mathbb{R}^m \to \mathbb{R}^n$. Its *total derivative* at $a \in \mathbb{R}^m$ is a linear function that is the best linear approximation of f close to a (and it is represented by the Jacobian of f at a). In particular, the domain of the total derivative is (a subset of) $\mathbb{R}^m$ itself. In the case of a smooth manifold, however, it does not make sense to talk about linear functions on (a subset of) M itself. The solution is to define a function between two tangent spaces, that are actually vector spaces, so it makes perfectly sense to talk about linear maps. In this sense, the differential is a generalization of the total derivative.

Definition A.14. Let M and N be smooth manifolds with or without boundary, and $F : M \to N$ a smooth map. Let $p \in M$. The **differential of F at p**, denoted by dF_p is a map:

$$dF_p : T_pM \to T_{F(p)}N$$

defined by:

$$dF_p(v)(f) = v(f \circ F)$$

for all $v \in T_pM$, $f \in C^\infty(N)$.

Though its properties are much more, we state two of them in the following

Proposition A.3. *The following two properties hold true:*

- dF_p *is linear;*
- *if F is a diffeomorphism, then dF_p is a homeomorphism between T_pM and $T_{F(p)}N$ (so the two can be identified) and $(dF_p)^{-1} = d(F^{-1})_{F(p)}$.*

Local Coordinates

In this section, we leverage what introduced in Section A.1.3 in order to obtain a local basis of the tangent space of a smooth manifold by using the basis of the tangent space to the euclidean space it is mapped to through local charts. In particular, if M is a smooth manifold, and (U, ϕ) is some local chart containing $p \in M$, we can view ϕ as a diffeomorphism between U and $\hat{U} = \phi(U) \in \mathbb{R}^n$ (denote $\phi(p)$ with $\hat{p}$). The idea is that if we could substitute U with whole M and $\hat{U}$ with whole $\mathbb{R}^n$, we could simply say that

$d\phi_p : T_pM \to T_{\hat{p}}\mathbb{R}^n$ is an isomorphism (by Proposition A.3), and thus obtain a basis of T_pM by leveraging the basis we know of $T_{\hat{p}}\mathbb{R}^n$.

The problem is that tangent vectors are defined globally, while charts are only local. That is why we first need the following

Proposition A.4. *Let M be a smooth manifold, and let $p \in M$, $v \in T_pM$ and $f, g \in C^\infty(M)$. If f and g agree on a neighbourhood of p, then $vf = vg$.*

This in particular tells us that we can identify the tangent space at $p \in M$ to an open submanifold of M $U \subseteq M$ with the one to M itself.

We can now see how to obtain such a claimed basis for T_pM. As explained in Section A.1.3, a basis for $T_{\hat{p}}\mathbb{R}^n$ is given by $\left(\frac{\partial}{\partial x^i}\big|_{\hat{p}}\right)_{i=1}^n$. The isomorphism provided by $d\phi_p$ defines a basis on T_pM which is $\left(\frac{\partial}{\partial x^i}\big|_{p}\right)_{i=1}^n$, which is explicitely:

$$\begin{aligned} \frac{\partial}{\partial x^i}\bigg|_p &= (d\phi_p)^{-1}\left(\frac{\partial}{\partial x^i}\bigg|_{\hat{p}}\right) = \\ &= d(\phi^{-1})_{\hat{p}}\left(\frac{\partial}{\partial x^i}\bigg|_{\hat{p}}\right) \end{aligned}$$

In terms of actions on functions, this is:

$$\frac{\partial}{\partial x^i}\bigg|_p f = \frac{\partial}{\partial x^i}\bigg|_{\hat{p}} (f \circ \phi^{-1}) \tag{A.6}$$

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

A.1.4. Frame Bundle and Connection

In Section A.1.3 we showed how one can obtain a basis of T_pM, which is thus identifiable with $\mathbb{R}^n$. It is natural that such a basis for the tangent space is not unique. Roughly speaking, the set of basis that we can choose for T_pM, for any $p \in M$, builds the frame bundle FM. The reason for introducing this concept is that, when moving along M on a curve, we might want to describe the curve as the solution to an ODE which coefficients are tangent vectors to the curve x_t: we might want to write these vectors in coordinates that are somehow moving with the manifold. In some sense, we want to choose an initial basis on $T_{x_0}M$ and move it along x_t only according to the curvature M prescribes, without introducing artificial rotations. Since the concept of moving the basis only according to what the curvature of M prescribes depends on how we define curvature on M, we will first need to introduce the concept of connection. All the material for this section has been taken from Hsu [13].

Conncetion and Parallel Transport

Let M be a smooth manifold of dimension d. We denote by $\Gamma(TM)$ as the set of vector fields on M. A connection is a rule for differentiating a vector field along another vector field.

Definition A.15 (Connection)**.** A **connection** is a map $\nabla : \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$ such that for each $X, Y, Z \in \Gamma(TM)$ and $f, g \in C^\infty(M)$ the following properties hold true:

1. $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$
2. $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$
3. $\nabla_X (fY) = f\nabla_X Y + X(f)Y$

We call $\nabla_X Y$ the **covariant differentiation of Y along X.**

A connection can be expressed in terms of local coordinates through the so called *Christoffel symbols*. Let $x = \{x^1, x^2, ..., x^d\}$ be a local chart on an open subset $O \subseteq M$. We know that the vector fields $X_i = \frac{\partial}{\partial x^i}$ are a basis for $T_x M$ for all $x \in O$. **Christoffel symbols** are functions $\Gamma^k_{ij} : O \to \mathbb{R}$ defined by:

$$\nabla_{X_i} X_j = \Gamma^k_{ij} X_k \tag{A.7}$$

where Einstein convention on sums has been used.

Now that we have the notion of connection, we can define parallel transport of a vector field along a curve.

Definition A.16 (Parallel Transport of a Vector Field Along a Curve)**.** Let M be a smooth manifold equipped with a connection ∇. A vector field V along a curve $\{x_t\}_{t\in[0,T]} \subset M$ is **parallel along the curve** if:

$$\nabla_{\dot{x}_t} V = 0 \quad \forall t \in [0, T] \tag{A.8}$$

In such a case, we say that V_{x_t} (the vector field evaluated at x_t) is the **parallel transport**

of V_{x_0} along the curve.

Under the same notations of the definition, if we write $x_t = \{x_t^i\}$ in local coordinates, then we can express $V_{x_t} = v^i(t)X_i$. We can thus impose a system of ODEs that $v^i(t)$ have to fulfill in order to have parallel transport:

$$\dot{v}^k(t) + \Gamma^k_{jl}(x_t)\dot{x}^j_t v^l(t) = 0 \tag{A.9}$$

We conclude this paragraph with the definition of geodesic:

Definition A.17 (Geodesic)**.** A curve $\{x_t\}_{t\in[0,T]} \subseteq M$ is a **geodesic** if its tangent vector is transported parallelly along the curve itself, i.e.

$$\nabla_{\dot{x}_t}\dot{x}_t = 0 \tag{A.10}$$

Writing locally $\dot{x}_t = \dot{x}^i_t X_i$, we get the ODE for the geodesics:

$$\ddot{x}^k_t + \Gamma^k_{jl}(x_t)\dot{x}^j_t\dot{x}^l_t = 0 \quad \forall k = 1, 2, ..., d \tag{A.11}$$

Frame Bundle

Definition A.18 (Frame)**.** Let M be a smooth d-dimensional manifold and let $x \in M$ with its tangent space T_xM. A **frame** at x is a linear isomorphism $u : \mathbb{R}^d \to T_xM$.

Equivalently, one can see u as an ordered basis for T_xM: that is why sometimes we will say *frame* to denote the ordered basis itself, though based on the above definition it is the linear isomorphism inducing it. To check this, let $v \in T_xM$ and recall that u is bijective, so there exists $w \in \mathbb{R}^d$ such that $u(w) = v$. If we now consider the canonical basis $(e_i)_{i=1}^d \subset \mathbb{R}^d$, we have that $w = w^i e_i$, so that $v = u(w) = u(w^i e_i)$. By linearity of u, $v = w^i u(e_i)$, meaning that any vector $v \in T_xM$ can be expressed uniquely (because w is uniquely written in terms of the canonical basis) as linear combination of $(ue_1, ue_2, ..., ue_d)$, where we write ue_i for $u(e_i)$. Sometimes we denote by $(u_i)_{i=1}^d \subset T_xM$ such induced basis.

The set of all frames at $x \in M$ is denoted by $\mathscr{F}(M)_x$, whereas the **frame bundle** is $\mathscr{F}(M) = \bigcup_{x\in M}\mathscr{F}(M)_x$. The frame bundle can be endowed with the structure of a smooth manifold of dimension $d + d^2$, and the canonical projection $\pi : \mathscr{F}(M) \to M$ such that for $u \in \mathscr{F}(M)_x$, $\pi(u) = \pi u = x$ is a smooth map. The tangent space to $u \in \mathscr{F}(M)$,

$T_u\mathscr{F}(M)$ is a vector space of dimension $d + d^2$.

Horizontal Vectors and Horizontal Lift

Assume M is equipped with a connection ∇. A curve $\{u_t\} \subseteq \mathscr{F}(M)$ is essentially a smooth choice of frame for the tangent space at each point of the underlying curve on M $\{\pi u_t\} = \{x_t\}$.

Definition A.19 (Horizontal Curve). A curve $\{u_t\} \subseteq \mathscr{F}(M)$ is a **horizontal curve** if for each $e \in \mathbb{R}^d$ the vector field $\{u_t e\}$ is parallel along $\{\pi u_t\}$.

Essentially, a horizontal curve is the choice of basis described at the beginning of this section: we fix a basis for $T_{x_0}M$ and we parallelly transport it along $\{x_t\}$ without introducing artificial rotations.

Definition A.20 (Horizontal Vector). A vector $X \in T_u\mathscr{F}(M)$ is **horizontal** if it is the tangent vector to a horizontal curve starting at u. The space of horizontal vectors at u is denoted by $H_u\mathscr{F}(M)$.

On the opposite side, vertical curves are curves on the frame bundle that "maintain x_0 fixed, and just rotate the basis chosen for $T_{x_0}M$". A vertical vector is just a tangent vector to such curve. Formally:

Definition A.21 (Vertical Vector). $X \in T_u\mathscr{F}(M)$ is **vertical** if it is tangent to the fibre $\mathscr{F}(M)_{\pi u}$. The space of vertical vector fields is denoted by $V_u\mathscr{F}(M)$.

As one would expect by the interpretation we gave, $V_u\mathscr{F}(M)$ is a vector subspace of $T_u\mathscr{F}(M)$ of dimension d^2, whereas $H_u\mathscr{F}(M)$ is a subspace of dimension d. Not only, but actually we have the following decomposition:

$$T_u\mathscr{F}(M) = V_u\mathscr{F}(M) \oplus H_u\mathscr{F}(M)$$

From this, it follows (DIMOSTRA) that the canonical projection $\pi : \mathscr{F}(M) \to M$ induces an isomorphism $\pi_* : H_u\mathscr{F}(M) \to T_{\pi u}M$. This way, the following definition makes sense:

Definition A.22 (Horizontal Lift of Vector). Let $X \in T_xM$ and $u \in \mathscr{F}(M)$ such that $\pi u = x$. The **horizontal lift** of X to u is the unique $X^* \in H_u\mathscr{F}(M)$ such that $\pi_* X^* = X$.

Definition A.23 (Horizontal Lift of Curve). Given a curve $\{x_t\}_{t\in[0,T]} \subseteq M$ and an initial frame u_0 at x_0, there is a unique curve $\{u_t\}_{t\in[0,T]} \subseteq \mathscr{F}(M)$ that is horizontal and such that $\pi u_t = x_t$ for all $t \in [0, T]$. This is the **horizontal lift of x_t from** u_0.

Notice that, even though the horizontal lift of a curve needs to be found by solving an ODE, one can prove that this solution exists and does not blow up.

A question that may arise is whether we can define a basis for $H_u\mathscr{F}(M)$. Let $e \in \mathbb{R}^d$, we can define a vector field H_e on $\mathscr{F}(M)$ by:

$$H_e(u) = (ue)^* \quad \forall u \in \mathscr{F}(M)$$

where X^* is the horizontal lift of X. If $e_1, e_2, ..., e_d$ is the canonical basis in $\mathbb{R}^d$, we call $H_i = H_{e_i}$ the fundamental horizontal vector fields, and $(H_i(u))_i$ is a basis for $H_u\mathscr{F}(M)$.

From a practical perspective, most of the times we are interested in working in local coordinates. The next result tells us how to write H_i in local coordinates.

Let $x = \{x^i\}$ be a local chart on an open subset $O \subseteq M$, and let $\tilde{O} = \pi^{-1}(O) \subseteq \mathscr{F}(M)$. We know tangent vectors $X_i = \frac{\partial}{\partial x^i}$ form a basis for T_xM for each $x \in O$. Now, for each $u \in \tilde{O}$, we have that $ue_i = e_i^j X_j$ for some matrix e_i^j. Then $(x, e) = (x^i, e_j^i) \in \mathbb{R}^{d+d^2}$ is a local chart for $\tilde{O}$. Notice how it makes sense that u can be seen as a couple (x, e) where $x \in M$ and e is some $d \times d$ matrix. Indeed, $u \in \mathscr{F}(M)$ is uniquely characterized by the $x \in M$ such that T_xM is the codomain of u and the effect it has on the basis $e_1, e_2, ..., e_d$, since it is a linear map (such effect is given by the columns of e).

Now, if we define $X_{kj} = \frac{\partial}{\partial e_j^k}$, X_{kj} span the vertical subspace $V_u\mathscr{F}(M)$, whereas $\{X_i, X_{ij}, 1 \leq i, j \leq d\}$ span $T_u\mathscr{F}(M)$ on $\tilde{O}$. The following result gives us, in local coordinates, a basis for $H_u\mathscr{F}(M)$.

Proposition A.5. *With the notation introduced above, at $u = (x, e) = (x^i, e_j^k) \in \mathscr{F}(M)$, we have:*

$$H_i(u) = e_i^j X_j - e_i^j e_m^l \Gamma_{jl}^k(x) X_{km} \tag{A.12}$$

A final definition is the one of *anti-development*. Intuitively, if we have a curve on a smooth manifold and we choose a frame for the tangent space that moves with the curve only according to the curvature of the manifold at each point, the curve that it "prints" on the euclidean space is the anti-development.

Definition A.24 (Anti-development). Let $\{u_t\}$ be horizontal lift of a differentiable curve $\{x_t\}$ on a smooth manifold M. We define the **anti-development** $\{w_t\} \subseteq \mathbb{R}^d$ by:

$$w_t = \int_0^t u_s^{-1} \dot{x}_s ds \quad \forall t \in [0, T] \tag{A.13}$$

or equivalently:

$$\begin{cases} \dot{x}_t = u_t \dot{w}_t \quad \forall t \in [0, T] \\ w_0 = 0 \in \mathbb{R}^d \end{cases} \tag{A.14}$$

It makes sense that the process of passing from $\{w_t\}$ to $\{x_t\}$ is called *rolling-without-slipping*.

Orthonormal Frame Bundle

Of particular interest is the case when we want to restrict ourselves to orthonormal basis of $T_x M$. In order for the concept of orthogonality (and of normality) to be defined we need the notion of inner product on tangent spaces. That is roughly the structure that a Riemaniann manifold has. On this section, we assume M is a Riemaniann manifold with Riemaniann metric $\langle \cdot, \cdot \rangle$. In general, parallel transport according to a connection ∇ does not guarantee that an initially orthonormal frame is transported to an orthonormal one. This happens if and only if for any $X, Y, Z \in \Gamma(TM)$:

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

in which case we say the connection is **compatible** with the Riemaniann metric. Given a Riemaniann metric, there is only one compatible, torsion-free connection: this is called the **Levi-Civita connection**.

If we denote by $\mathscr{O}(M)$ the **orthonormal frame bundle**, everything we said so far for $\mathscr{F}(M)$ can be applied the same way provided the connection is compatible with the Riemaniann metric (so, in particular, if it is the Levi-Civita connection).

If one defines $g_{ij} = \langle X_i, X_j \rangle$, $\{g_i j\}$ the corresponding matrix and $\{g^{im}\}$ its inverse, Christoffel symbols for the Levi-Civita connection are given by:

$$\Gamma^i_{jk} = \frac{1}{2} g^{im} (g_{jm,k} + g_{mk,j} - g_{jk,m}) \tag{A.15}$$

where $g_{jk,m} = \frac{\partial g_{jk}}{\partial x^m}$.

A.2. Alternative Proofs to Lin [16]

We first consider Proposition 3 of Lin [16], claiming the geodesics structure on M.

Proposition A.6. *On $(M, \tilde{g}(\cdot,\cdot))$, the geodesic starting at $L \in M$ with direction $X \in T_L M$ is given by:*

$$\tilde{\gamma}_{L,X}(t) = \lfloor L \rfloor + t\lfloor X \rfloor + \mathbb{D}(L)\exp(t\mathbb{D}(X)\mathbb{D}(L)^{-1}) \tag{A.16}$$

where $\lfloor A \rfloor$ denotes the matrix with same strictly lower diagonal part of A (zero the rest), and $\mathbb{D}(A)$ its diagonal matrix.

Proof. By the above construction, being $\bar{\phi}$ an isometry, we have that $\tilde{\gamma}_{L,X}(t) = \bar{\phi}^{-1}(\gamma_{\bar{\phi}(L),d\bar{\phi}_L(X)}(t))$, being $\gamma_{\bar{\phi}(L),d\bar{\phi}_L(X)}(t)$ the geodesic on $\mathcal{L}$ from $\bar{\phi}(L)$ with direction $d\bar{\phi}_L(X)$.

Being $\mathcal{L}$ linear, it is easy to get such a geodesic: $\gamma_{P,W}(t) = P + tW$. Now, recall that in our case:

$$d\bar{\phi}_L(X) = \begin{pmatrix} \frac{1}{L_{11}}X_{11} & 0 & 0 & \dots & 0 \\ X_{21} & \frac{1}{L_{22}}X_{22} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & X_{n3} & \dots & \frac{1}{L_{nn}}X_{nn} \end{pmatrix}$$

Then:

$$\begin{aligned} \tilde{\gamma}_{L,X}(t) = \bar{\phi}^{-1}(\lfloor L \rfloor + t\lfloor X \rfloor + \log(\mathbb{D}(L)) + t\mathbb{D}(X)\mathbb{D}(L)^{-1}) = \\ = \lfloor L \rfloor + t\lfloor X \rfloor + \mathbb{D}(L)\exp(t\mathbb{D}(X)\mathbb{D}(L)^{-1}) \end{aligned}$$

which concludes the proof. $\square$

This in particular tells us that the geodesic distance between L and K on M can be obtained as:

$$d_M(L,K) = \{\| \lfloor L \rfloor - \lfloor K \rfloor \|_F^2 + \| \log \mathbb{D}(L) - \log \mathbb{D}(K) \|_F^2\}^{\frac{1}{2}} \tag{A.17}$$

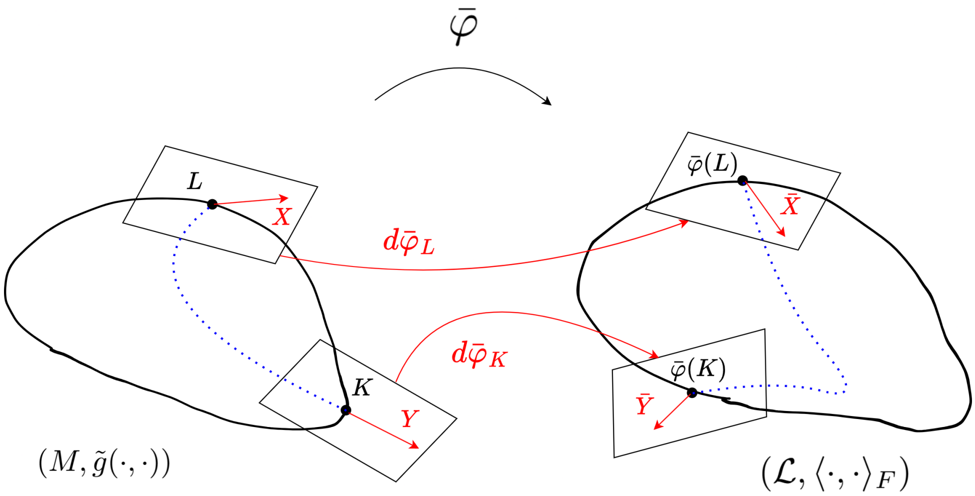

Figure A.1: Schema for Parallel Transport on Cholesky Space

which is precisely the Frobenius distance of the transformations.

Proposition 7 of Lin [16] deals with parallel transport on $(M, \tilde{g}(\cdot,\cdot))$. The way of showing the form of parallel transport leverages the Lie group structure with which one can endow M. We can however obtain it by exploiting once again the fact that $\bar{\phi} : (M, \tilde{g}(\cdot,\cdot)) \to (\mathcal{L}, \langle \cdot,\cdot \rangle_F)$ is an isometry between Riemaniann manifolds.

Proposition A.7. *A tangent vector $X \in T_L M$, being $L \in M$, is parallelly transported to the tangent vector $\lfloor X \rfloor + \mathbb{D}(K)\mathbb{D}(L)^{-1}\mathbb{D}(X)$ at $K \in M$.*

Proof. We will adopt the following notation: $L, K \in M$, $X \in T_L M$ and $Y \in T_K M$ is X parallelly transported to K. $d\bar{\phi}_L : T_L M \to T_{\bar{\phi}(L)}\mathcal{L}$ is the differential at L, while $d\bar{\phi}_K : T_K M \to T_{\bar{\phi}(K)}\mathcal{L}$ is the one at K. Denote $\bar{L} = \bar{\phi}(L)$, $\bar{X} = d\bar{\phi}_L(X)$ and $\bar{Y} = d\bar{\phi}_K(Y)$. For a visual representation of this, see Figure A.1.

As $\mathcal{L}$ is a vector space, with the usual Frobenius norm some $\bar{V} \in T_{\bar{L}}\mathcal{L}$ is transported to $\bar{V} \in T_{\bar{K}}\mathcal{L}$ (recalling the identification between the tangent spaces of $\mathcal{L}$ and itself, $\bar{V}$ is simply identified as the same element in the two tangent spaces). In such case, we can easily obtain Y as:

$$Y = (d\bar{\phi}_K)^{-1}(\bar{X}) \tag{A.18}$$

Now, recall that the differential transforms X by:

$$\bar{X} = \lfloor X \rfloor + \mathbb{D}(L)^{-1}\mathbb{D}(X) \tag{A.19}$$

We simply need a closed form for $(d\bar{\phi}_K)^{-1}$ now. Take $W \in T_{\bar{K}}\mathcal{L}$, and suppose $d\bar{\phi}_K(Z) = W$ for some $Z \in T_K M$ (Z exists and is unique, because being $\bar{\phi}$ a diffeomorphism its differential is an isomorphism, thus in particular an invertible map). We want to find Z. Writing explicitely the differential and separating W into its diagonal and off-diagonal part:

$$\lfloor W \rfloor + \mathbb{D}(W) = \lfloor Z \rfloor + \mathbb{D}(K)^{-1}\mathbb{D}(Z) \tag{A.20}$$

which implies that:

$$\begin{cases} \lfloor Z \rfloor = \lfloor W \rfloor \\ \mathbb{D}(Z) = \mathbb{D}(K)\mathbb{D}(W) \end{cases} \tag{A.21}$$

Now, putting everything together into Equation (A.18), we get that:

$$\begin{aligned} \bar{Y} &= (d\bar{\phi}_K)^{-1}(\lfloor X \rfloor + \mathbb{D}(L)^{-1}\mathbb{D}(X)) = \\ &= \lfloor X \rfloor + \mathbb{D}(K)\mathbb{D}(L)^{-1}\mathbb{D}(X) \end{aligned}$$

that is the claim. □

Another result obtained in Lin [16] is the form of the Frechet mean of a finite number of elements of M.

Proposition A.8. *Let L be a random variable with values in M and suppose for some fixed $A \in M$ it holds that $\mathbb{E}[d_M^2(A, L)] < \infty$. Then the Frechet mean of L, denoted by $\mathbb{E}[L]$, is given by:*

$$\mathbb{E}[L] = \mathbb{E}[\lfloor L \rfloor] + \exp\{\mathbb{E}[\log \mathbb{D}(L)]\} \tag{A.22}$$

where $\mathbb{E}[\lfloor L \rfloor]$ is the Frechet mean of the strictly lower triangular matrix of L and $\mathbb{E}[\log \mathbb{D}(L)]$ is the one of the log-transformed diagonal.

Proof. Since $\bar{\phi}$ is an isometry, we have that:

$$\mathbb{E}[L] = (\bar{\phi})^{-1}(\mathbb{E}[\bar{\phi}(L)]) \tag{A.23}$$

Now, if G is a random variable with values in $\mathcal{L}$ and there is some element $R \in \mathcal{L}$ such that $\mathbb{E}[\|G - R\|_F^2] < \infty$, then the Frechet mean of G exists. We denote it with $\mathbb{E}[G] \in \mathcal{L}$.

The above equation is:

$$\begin{aligned}\mathbb{E}[L] &= (\bar{\phi})^{-1}(\mathbb{E}[\lfloor L \rfloor + \log \mathbb{D}(L)]) = \\ &= (\bar{\phi})^{-1}(\mathbb{E}[\lfloor L \rfloor] + \mathbb{E}[\log \mathbb{D}(L)]) = \\ &= \mathbb{E}[\lfloor L \rfloor] + \exp\{\mathbb{E}[\log \mathbb{D}(L)]\}\end{aligned}$$

where in the first passage we used linearity of expectation, and in the second one the fact that: $\mathbb{E}[\lfloor L \rfloor]$ is a strictly lower triangular matrix, being the space of strictly lower triangular matrixes a vector space, and thus being the Frechet mean well defined in such space,$\mathbb{E}[\log \mathbb{D}(L)]$ is a diagonal matrix for the same reason, and that if A is strictly lower triangular and B is diagonal:

$$(\bar{\phi})^{-1}(A + B) = A + \exp B$$

We thus get the claimed formula for the Frechet mean.

$\square$

B | Stochastic Differential Geometry Aspects

B.1. Stochastic Differential Geometry

Everything we said in Section A.1.4 can be extended to the case of curves on a smooth manifold M that are not differentiable, but are M-valued semimartingales. Specifically, the notions we are interested in generalizing are the ones of horizontal lift, development and anti-development. Intuitively, this is useful because it gives us the possibility to generate semimartingales on M by starting from semimartingales on $\mathbb{R}^d$, developing them as horizontal semimartingales on the frame bundle and finally projecting them onto M, essentially obtaining a curve on M starting from an $\mathbb{R}^d$ curve and *rolling without slipping*.

We will be particularly interested in properly defining Brownian Motion (BM) and Brownian bridge (BB) on M.

In the following, all processes are defined on a fixed filtered probability space $(\Omega, \mathscr{F}_*, \mathbb{P})$.

Definition B.1 (Horizontal Semimartingale)**.** Let U be an $\mathscr{F}(M)$-valued semimartingale. It is a **horizontal semimartingale** if there exists an $\mathbb{R}^d$-valued process W, called the **anti-development** of U (or of πU), such that:

$$dU_t = H_i(U_t) \circ dW_t^i \tag{B.1}$$

where H_i are the fundamental horizontal vector fields.

Definition B.2 (Stochastic Development)**.** Let W be an $\mathbb{R}^d$-valued semimartingale and U_0 an $\mathscr{F}(M)$-valued, $\mathscr{F}_0$-random variable. The solution U of Equation (B.1) with initial condition U_0 is called **stochastic development** of W in $\mathscr{F}(M)$. Its projection πU is the **stochastic development** of W in M.

Definition B.3 (Stochastic Horizontal Lift)**.** Let X be an M-valued semimartingale. Let

U be an $\mathscr{F}(M)$-valued semimartingale such that Equation (B.1) holds true (i.e. U is horizontal) and $X = \pi U$. Then U is called a **stochastic horizontal lift** of X.

Suppose M can be seen as a closed submanifold of $\mathbb{R}^N$, and for any $x \in M$, let $P(x) : \mathbb{R}^N \to T_xM$ be the orthogonal projection onto T_xM, seen as a subspace of $\mathbb{R}^N$. Also, let $P_\alpha(x) = P(x)e^\alpha$, where e^α is the α^{th} element of the canonical basis for $\mathbb{R}^N$, $\alpha = 1, 2, ..., N$. Then one can prove that, given a semimartingale X on M with coordinates X^α, its stochastic horizontal lift U is given by:

$$dU_t = P^*_\alpha(X_t) \circ dX^\alpha_t \tag{B.2}$$

where Einstein notation on sums has been adopted and $P^*_\alpha(X_t)$ is the horizontal lift of $P_\alpha(X_t)$ as defined in Section A.1.4.

Acknowledgements

I would like to express my sincere gratitude to my advisors, Prof. Laura Maria Sangalli, Prof. Victor Panaretos, and Dr. Leonardo Santoro, for their guidance, insightful discussions, and continuous support throughout this work. Their expertise and encouragement have been invaluable to the development of this thesis.

I am also deeply grateful to my family for their constant encouragement, and to my friends for their support, understanding, and for making this journey all the more enjoyable.

Finally, I would like to thank everyone who, in one way or another, has contributed to my academic and personal growth during these years.