Inference on Riemannian Manifolds: Regression and Stochastic Differential Equations

Mai Ngoc Bui

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Statistical Science

University College London

March 25, 2022

I, Mai Ngoc Bui, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work. The ideas in Chapter 4 have been submitted to an academic journal for publication and have also been made available through arXiv, see [1]. An earlier version of this work has been presented at the Greek Stochastics λ workshop. The most recent presentation was at the Oberwolfach workshop titled "Statistics of Stochastic Differential Equations on Manifolds and Stratified Spaces" with a summary available in [2].

Abstract

Statistical inference for manifolds attracts much attention because of its power of working with more general forms of data or geometric objects. We study regression and stochastic differential equations on manifolds from the intrinsic point of view.

Firstly, we are able to provide alternative parametrizations for data that lie on Lie group in the problem of fitting a regression model, by mapping this space intrinsically onto its Lie algebra, while we explore the behaviour of fitted values when this base point is chosen differently. Due to the nature of our data in the application of soft tissue artefacts, we employ two correlation structures, namely Mateŕn and quasi-periodic correlation functions when using the generalized least squares, and show that some patterns of the residuals are removed.

Secondly, we construct a generalization of the Ornstein-Uhlenbeck process on the cone of covariance matrices $\mathcal{SP}(n)$ endowed with two popular Riemannian metrics, namely Log-Euclidean (LE) and Affine-Invariant (AI) metrics. We show that the Riemannian Brownian motion on $\mathcal{SP}(n)$ has infinite explosion time as on the Euclidean space and establish the calculation for the horizontal lifts of smooth curves. Moreover, we provide Bayesian inference for discretely observed diffusion processes of covariance matrices associated with either the LE or the AI metrics, and present a novel diffusion bridge sampling method using guided proposals when equipping $\mathcal{SP}(n)$ with the AI metric. The estimation algorithms are illustrated with an application in finance, together with a goodness-of-fit test comparing models associated with different metrics. Furthermore, we explore the multivariate volatility models via simulation study, in which covariance matrices in the models are assumed to be unobservable.

Impact statement

Statistical inference on manifolds receives much attention in the literature, however, because of non-linearity studying statistical inference on general manifolds is distinct from the Euclidean space. Thus, the Gaussian distribution, while it can be defined, loses its ubiquitous role and many standard statistical procedures behave unexpectedly. While regression analysis and stochastic differential equations have been very successful for several decades, little has been published on the case when the state space on which the system evolves is not simply Euclidean.

In this thesis, we focus on two applications where ignoring the special geometric structures negatively affect model fits. Firstly, we study the human knee joint whose bending at each point in time is described by a rotational matrix, which is an element belonging to the Lie group of rotations of solid three-dimensional bodies. Particularly, we extend the use of generalized least squares to any compact and connected Lie group through a transformation to its tangent space at the identity. Secondly, we generalize the famous Ornstein-Uhlenbeck process on the Euclidean space to the space of covariance matrices, $\mathcal{SP}(n)$. In particular, we develop a novel diffusion bridge sampling strategy from the intrinsic point of view and perform inference for partially observed data when $\mathcal{SP}(n)$ is a complete Riemannian manifold of non-positive sectional curvature. And inference for the multivariate stochastic volatility models is also explored toward the end of the thesis.

In summary, this thesis extends existing methods in statistical science in the case where curvature cannot be ignored and by exploring geometric structure, we can find new statistical methodology that is applicable in improving modelling across a range of applications.

Acknowledgements

First and foremost, I would like to express my deepest and sincere gratitude to my supervisors, Dr. Yvo Pokern and Prof. Petros Dellaportas, for their dedication, patience, enthusiasm and endless support of me throughout this journey. I am very fortunate to working with them, and without their persistent help, this thesis would have never been accomplished. Over the past few years, they have always guided me on different aspects of research and career development, in order for me to become an independent researcher. I also would like to offer my special thanks to Prof. Stephan Huckemann for his insightful comments and suggestions.

Most importantly, none of this could have happened without my family and friends. My parents, who have offered their unconditional love and encouragement through video chats every week. My accomplishments and success are because they have always believed in me. Deepest thanks to my sisters, who keep me grounded, remind me of what is important in life, and are always supportive of my adventures. Special thanks to my uncles, aunts and my cousins for their unfailing support and continuous encouragement. My friends Mariam Adeleke, Nguyen H. Nhu Y, Vu T. Bao Tram, Pham K. Linh, and other friends deserve my special thanks, who directly and indirectly provide me inspirations and valuable suggestion during my study.

Last but not least, I would like to take this opportunity to express gratitude to all member staffs of the Department of Statistical Science for their help and support. And I am thankful for the UCL Research Scholarships, without the funding, this project could not have reached its goal.

Contents

List of Figures

List of Tables

Chapter 1

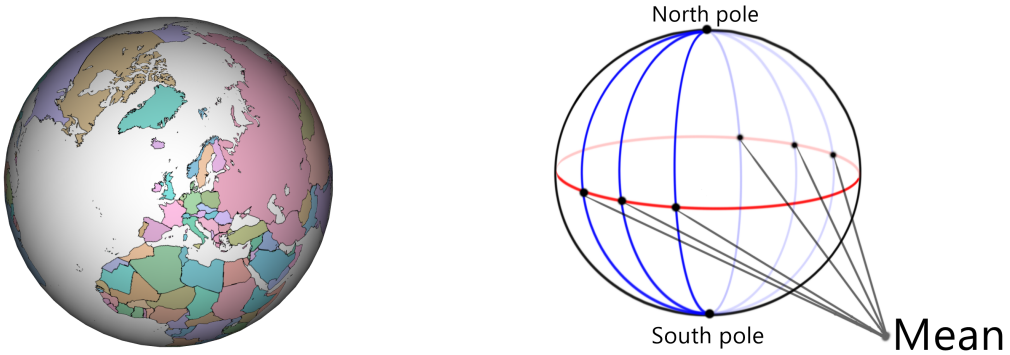
Introduction

1.1 Statistical inference on manifolds

Statistical inference on manifolds has received much attention in the literature because manifolds not only do manifold-valued observations appear in many applications (i.e. directional data [7, 8, 9], shape [10, 11], diffusion tensors [12, 13], etc) but also because manifolds are able to explain geometric concepts (e.g. collinearity in regression analysis [14], etc). However, because of non-linearity, studying statistical inference on general manifolds is distinct from the Euclidean space. For example, the most favoured probability distribution on the Euclidean space, Gaussian distribution, loses many favourable properties, such as that the additive Gaussian error is no longer well-defined on a general manifold. Moreover, many standard statistical procedures behave unexpectedly, such as the existence and uniqueness of the Fréchet mean only being guaranteed for some special conditions [15], such as Riemannian manifold with non-positive sectional curvature, and in general it is not guaranteed, see Example 1.1.

Example 1.1. The Earth (or sphere $\mathcal{S}^2$) in Figure 1.1a is a smooth manifold (Figure 1.1a were created from [16]). It can be easily seen that globally it is a curved surface, whereas locally it is flat. For instance, if we only look at London instead of the whole Earth, we can assume that London is approximately flat.

Furthermore, there is more than one geodesic connecting the North pole and the South pole, see Figure 1.1b. Since the Fréchet mean is defined such that it



(a) The Earth is a smooth manifold. **(b)** Fréchet means on $\mathcal{S}^2$.

Figure 1.1: Fréchet mean on the smooth manifold $\mathcal{S}^2$.

minimizes the sum of squared distances between itself and the observations globally, if our observations are the North pole and the South pole, any points on the equator can become the Fréchet mean for this data set. On the other hand, if we suppose that all observations lie on the equator of $\mathcal{S}^2$, the Fréchet mean can be either the North pole or the South pole, then again the mean is not unique. Now, suppose that we simply pick one of these two points to be the mean, e.g. the South pole, and by moving slightly one observation toward the South pole, the new mean now lies close to the South pole. In other words, the Fréchet mean on $\mathcal{S}^2$ can jump instantaneously by shifting slightly a single data point, and thus this definition of mean is no longer a continuous function with respect to the data.

Firstly, let us discuss some related works of extending a crucial statistical tool on the Euclidean space, namely the Gaussian distribution, onto manifolds. In 1948, M. Fréchet generalized the concept of the Euclidean mean (or expectation) from linear spaces to metric spaces, where distance is well-defined, and the mean is therefore called the Fréchet mean [17]. Example 1.1 illustrates that existence and uniqueness of the Fréchet mean (i.e. intrinsic mean) can not be ensured for a general manifold, and the Karcher mean is sometimes used instead since it minimizes the sum of squared Riemannian distance locally. Besides the extension of the mean, the variance can also be generalized to Riemannian manifolds in a similar manner, which allows us to study statistics on Riemannian manifolds from the intrinsic point of view. In particular, X. Pennec [15] constructs analogues of many usual statistical tools to

geodesically complete Riemannian manifolds. These include notions of mean value, covariance matrix of a random element, as well as Mahalanobis distance. Moreover, the generalization of Gaussian distribution on Riemannian manifold is constructed through the maximization of the entropy given that the mean and covariance of the distribution are available, and this distribution is referred to as normal law. By taking the limit of the variance, one can obtain the uniform distribution on compact manifolds or the point mass distribution. Approximation of this probability density can be efficiently achieved given that variance is small. Furthermore, the $\mathcal{X}^2$-law is studied in [15], which enable hypothesis testing on Riemannian manifolds.

Another successful statistical tools on the Euclidean space are the central limit theorem (CLT) and the law of large numbers (LLN). While the Euclidean version of LLN states that the empirical mean converges in probability and almost surely to the population mean, the corresponding CLT indicates that the limiting probability distribution is simply the Gaussian distribution. In order to extend these theories to Riemannian manifolds from the intrinsic point of view, one needs to overcome two problems: non-linearity of manifolds and the problem of existence, uniqueness and stability of the Fréchet mean. Some generalization works have been developed under some restrictions [18, 19, 20]. For example, on some particular Riemannian manifolds (e.g. open books, etc.) the Fréchet mean does not move when moving one or more data points, i.e. the mean is “sticky", and in this case the classical CLT and LLN are clearly not applicable. S. Huckemann & T. Hotz [19] establish this sticky concept formally and state the corresponding CLT and LLN for three cases of the Fréchet mean being either non-sticky, partial sticky or sticky on these manifolds. On the other hand, a new phenomenon of the Fréchet mean, which is formally defined as “smeariness", have been recently discussed, one up to date example of which is on the circle [20]. The example is a scenario, where there is positive mass near the cut locus of the population mean, and it is shown that the fluctuation of sample means asymptotically converges to a non-Gaussian distribution at a scale n^{α} with $\alpha < 1/2$, whereas in the classical CLT this scale is $\sqrt{n}$ and the limiting distribution is Gaussian.

We have discussed various routes that one can achieve the generalization of Gaussian distribution on manifolds, such as normal law, CLT, etc. However, there is in general no identical version of the Gaussian distribution on a manifold that preserves all properties as on the Euclidean space, and only partial features are retrieved which depend on the aim of use. Overall, studying non-linear spaces brings more complexity, and many Euclidean theories change. However, this study, particularly in statistics, is unavoidable, simply because geometric objects appear in a vast range of applications in many fields, and direct use of the Euclidean approach sometimes raises many issues. For example, Subsection 4.2.3 addresses the necessity for using Rieamannian metrics instead of the Euclidean metric (i.e. the Frobenius inner product) on the space of covariance matrices, and a Brownian bridge simulation exercise is implemented in Subsection 4.6.1 to illustrate the problems that arise when neglecting the Riemannian structures.

1.2 Research activities

In this thesis, we mainly focus on two areas of research, which are regression analysis and stochastic differential equations. More details are discussed below.

1.2.1 Regression on Lie groups

Regression modelling is an ubiquitous statistical tool in understanding the relationship between predictor and response variables, which has been studied for centuries on the Euclidean space and many results are generalized onto manifolds. On the Euclidean space, linear regression is the most popular regression model due to its simplicity and efficiency. If response variables lie on Riemannian manifolds, the natural generalization of linear regression is the geodesic regression, that is to fit the best geodesic curve on these manifolds [21]. Moreover, least squares estimation in this case is defined to minimize the sum of squared Riemannian distance, i.e. from the intrinsic point of view.

In this work, we are mainly interested in the problem of fitting a regression model given that predictor variables lie on Lie groups, i.e. being not only a manifold but also a group (see Definition 2.22). Moreover, we assume this special space is

compact and connected in order to guarantee the existence of a bi-invariant metric. This simply says that we can study the whole space via its tangent space at the identity (i.e. its Lie algebra) by using a fixed base point, and this tangent space is a vector space of same dimension with the space of interest. Thus, usual statistical tools such as generalized least squares can be implemented on the coefficients with respect to some fixed basis on the tangent space.

Next, we look into the space of rotational matrices $\mathcal{SO}(3)$ with the application in human kinematic analysis. Besides providing alternative regression models through the intrinsic view, one of our contributions in this work is to study the effect to fitted values from choosing different base points. Moreover, we explore two different correlation structures, namely Matérn and quasi-periodic correlation functions. These selections are based on the fact that the next movement of the knee is highly correlated to the previous position when observations are collected at high frequency. We show that while using ordinary least squares clearly violates the assumption about randomness of the residuals, using Matérn or quasi-periodic correlation structure removes some patterns among these residuals.

1.2.2 Stochastic differential equations on $\mathcal{SP}(n)$

We now move to a different research activity, in which we have most contributions in this thesis, the construction of time series models for covariance matrices that take the geometric structures into account. In this thesis, we mainly focus on the case that covariance matrices are strictly positive definite, unless indicated otherwise. We denote the space of $n \times n$ covariance matrices $\mathcal{SP}(n)$, and strictly speaking, $\mathcal{SP}(n)$ is the space of symmetric positive definite matrices.

We are interested in a successful class of time series models that not only capture the stochastic effects but also learn about the true trajectories, so-called stochastic differential equations (SDEs). This class of models has been studied intensively on the Euclidean space, but little work of statistical inference has been done on Riemannian manifolds. On $\mathcal{SP}(n)$, some existing time series models like the Wishart process lack geometric structures. O. Pfaffel [22] shows that Wishart process does lie on $\mathcal{SP}(n)$, but the simulation method, which uses the usual Euler-

Maruyama method, depends on some time-adjustment algorithm. Every time the simulated points fail to be symmetric positive definite, the time width is cut half, and this is clearly not an ideal approach since it can produce biased sample paths.

Studying SDEs on manifolds in general can be accomplished through various ways, such as local coordinates, extrinsic approach (i.e. embedding) or intrinsic approach (i.e. exponential map/connection). The first method borrows from the definition of being a manifold, thus using the system of local coordinates will map the space locally to the Euclidean space of same dimension. This approach gains popularity when local charts are not too complicated, and the existing results on Euclidean space can be applied directly [8, 7, 10, 11, 23]. On the other hand, for a more general manifold, where writing down the local charts is not so simple, embedding the manifold in a larger Euclidean space seem to be a natural approach [24]. This is the most common way in proving many important theories of stochastic analysis on manifolds as in theory one always can embed any manifold in some ambient Euclidean space of higher dimension using Nash embedding theorem [25]. However, not many manifolds have a nice embedding, and this approach becomes less appealing due to its practicability.

Alternatively, the intrinsic approach is also a feasible selection as many theoretical works are available, that are mainly based on the concepts of horizontal lift and stochastic development, see for example [26, 25]. Basically, for a smooth manifold equipped with a connection, one can lift a smooth curve horizontally to the frame bundle uniquely, and the resulting curve has a corresponding curve, namely the anti-development curve, on the Euclidean space of the same dimension as the manifold of interest. In other words, we can interpret that there is an one-to-one correspondence between the set of smooth curves of a fixed starting point on the manifolds and their anti-developments on the Euclidean space of same dimension. In practise, the piecewise approximation method using exponential map (sometimes mentioned as injection scheme) is well-received because to some extent it resembles a popular approximation method on the Euclidean space, the Euler-Maruyama approximation [10, 27]. Convergence of this piecewise approximation for some special

manifolds has been shown before the concept of stochastic development of SDEs is established [28, 29]. However, little have been published to connected these two theories in general and people often regard the injection scheme as the stochastic development via smooth approximation, with examples include: the work in [30, 31] that connect the scheme of injection to the tangent plane with the action “rolling without slipping" on $\mathcal{SO}(3)$; in Section 4.3 we demonstrate the approximation of horizontal lifts on $\mathcal{SP}(n)$ endowed with the Affine-Invariant metric.

Back to $\mathcal{SP}(n)$, we concentrate on two popular Riemannian metrics, namely Log-Euclidean (LE) [32, 33] and Affine-Invariant (AI) metrics [33, 34, 35]. Endowing $\mathcal{SP}(n)$ with the LE metric gives us a flat Riemannian manifold, and moreover we can write down the analytic form in local coordinate for the horizontal lift of smooth curves. In fact, we can map $\mathcal{SP}(n)$ globally onto its tangent space, the space of $n \times n$ symmetric matrices $\mathcal{S}(n)$ by the matrix logarithm function, which is a bijective map. The simplicity and efficiency gain from the results in [32],which offer the vector space structure in the logarithm domain of covariance matrices. On the other hand, the AI metric provides a geodesically complete Riemannian manifold of non-positive sectional curvature, i.e. a Cartan-Hadamard manifold. Due to the presence of curvature, many existing theories on the Euclidean space are no longer applicable. Both LE and AI metrics are getting more attentions for statistical inference on $\mathcal{SP}(n)$, such as studying the geometric mean (i.e. Fréchet mean) [32, 36], least squares [37], etc. We discuss the reasons for which Riemannian metrics are more favourable than the Euclidean metric (i.e. the Frobenius inner product) in greater details in Subsection 4.2.3.

Studying SDEs on $\mathcal{SP}(n)$ consists of three main parts, firstly exploring some stochastic properties of Brownian motion on $\mathcal{SP}(n)$ endowed with either the LE or AI metrics, e.g. stochastic completeness; secondly computing horizontal lifts of smooth curves in local coordinate for both metrics; and finally the construction of a parametric family of mean-reverting SDEs, which we refer to as the Ornstein-Uhlenbeck (OU) class of processes on $\mathcal{SP}(n)$. The main challenge in this work is performing statistical inference on $\mathcal{SP}(n)$ endowed with the AI metric, particularly

the capability of sampling diffusion bridges, since existing methods on the Euclidean space are now incompatible. Our contribution includes the development of a diffusion bridge sampler based on guided proposals from the intrinsic point of view. There has been a line of research based on ideas of B. Delyon & Y. Hu [38] on the Euclidean space that uses guided and residual proposal densities, see [39, 40, 41]. There is some extension to landmark manifold [42], but the work involves using local coordinates, in which technically sampling the bridges is still carried out on a flat space. Up to this date, no one has used this method in the intrinsic approach to work with SDEs on Riemannian manifolds, particularly $\mathcal{SP}(n)$. The proposed bridge sampling method can be extended to other Cartan-Hadamard manifolds under the assumption that there exists a suitable diffusion process with an explicit transition density function.

Finally, we illustrate our proposed algorithm in the AI case, and the modification of the strategy by G. O. Roberts & O. Strammer [43] in the LE case to the real application in finance. And for comparison these two metric tensors, we perform a goodness-of-fit test using the transition density-based approach [44]. Furthermore, we study inference for discretely observed data that come from the multivariate stochastic volatility models, in which covariance matrices in these models are indirectly observed and assumed to follow our construction of the OU class.

1.3 Thesis outline

The thesis is organized as follows. Firstly, we give a brief mathematical introduction with definition, mathematical theories and results about Riemannian manifolds, Lie group with Lie algebra and stochastic analysis on manifolds in Chapter 2. Details of two research activities are discussed in-depth in Chapter 3 for regression analysis on Lie groups and Chapter 4 for SDEs on $\mathcal{SP}(n)$.

In Chapter 3, we first review the generalized least squares in the Euclidean setting, and extend them to a compact connected Lie group in Section 3.2. We then apply this methodology to the space of rotational matrices $\mathcal{SO}(3)$, where we aim to fit a regression model of soft tissues artefacts against rotational matrices,

where two correlation structures, namely Matérn and quasi-periodic, are considered in Subsection 3.4.2. Moreover, we compare ordinary least squares estimation and different algorithms of generalized least squares estimation, together with the existing model using Euler angles in Subsection 3.4.1. We also investigate the effect of using different base points on fitted values in Section 3.3.

Next, we start Chapter 4 by revising some results of $\mathcal{SP}(n)$ endowed with either the LE or the AI metrics, such as exponential/logarithm maps, Riemannian distances, etc and following up with a discussion of how important the Riemannian geometric structures when working with covariance matrices. Afterwards, we define the Riemannian Brownian motion class and OU class on $\mathcal{SP}(n)$ in Section 4.3 and 4.4 respectively. Particularly, local expressions of horizontal lifts of smooth curves on $\mathcal{SP}(n)$ equipped with either the LE or the AI metrics are presented in Section 4.3. Furthermore, we construct guided proposals on $\mathcal{SP}(n)$ equipped with the AI metric in Section 4.5 and study the absolute continuity between the target diffusion bridge and the guided proposal. We investigate both simulated (Section 4.6) and real data (Section 4.7) to study the performance of our proposed diffusion processes on $\mathcal{SP}(2)$ with the three metrics: the Euclidean, the Log-Euclidean and the Affine-Invariant metrics. This chapter ends with the multivariate stochastic volatility models in Section 4.8, where we provide a simulation study on $\mathbb{R}^2$.

Finally, we summarize our work and outline future work in Chapter 5. Supplementary figures are presented in Appendix A. In Appendix B we give a brief overview of the guided proposal algorithm on the Euclidean space introduced by M. Schauer & F. V. D. Meulen et al. [40] and the algorithm for low-frequency inference of multi-dimensional diffusion proposed by G. O. Roberts & O. Stramer [43].

Chapter 2

Preliminary Material

We aim to provide a brief overview of the mathematical background on Riemannian geometry and stochastic analysis required for this thesis. Firstly, we introduce the notion of differentiable (or smooth) manifold in Subsection 2.1.1, which generalizes the Euclidean space and differentiation can be applied. We continue to extend these manifolds to metric spaces in Subsection 2.1.2, by equipping some metric tensors. If a Riemannian metric is used, the manifold is called Riemannian manifold.

Thereafter, a special class of manifolds which have algebraic structure, namely Lie groups, are discussed in Section 2.2. Finally, in Section 2.3 we review the theory of stochastic analysis on Euclidean spaces in essence, followed by manifolds in more details, such as the concept of horizontal lift, stochastic development, stochastic completeness etc, which are the key ingredients for Chapter 4.

2.1 Riemannian geometry

2.1.1 Differentiable manifold

A motivating example of a smooth manifold is the Earth (see Example 1.1), where locally resembles $\mathbb{R}^2$. However, being flat locally will not be sufficient to attain the differentiability, more conditions need to be achieved. We state them in the following definition.

Definition 2.1 (Manifold)**.** *[45] An n-dimensional manifold $\mathcal{M}$ is a topological space such that there exists a family $A = \{U_i, \varphi_i \,|\, i \in I\}$ with some indexing set I such that*

(i) U_i *is an open set in* $\mathcal{M}$ *and* $\bigcup_{i\in I} U_i = \mathcal{M}$.

(ii) $\varphi_i : U_i \to \mathbb{R}^n$ *is a continuous bijection onto open set* $\varphi_i(U_i)$ *with continuous inverse (i.e. homeomorphism).*

(iii) Whenever $U_i \cap U_j \neq \emptyset$ *, the map* $\varphi_j \circ \varphi_i^{-1} : \varphi_i(U_i \cap U_j) \to \varphi_j(U_i \cap U_j)$ *is a smooth bijection with smooth inverse (i.e. diffeomorphism)*

If $\{\varphi_i \circ \varphi_j^{-1} \mid \forall i, j\}$ *are differentiable (smooth), we say* $\mathcal{M}$ *is a differentiable (smooth) manifold.* $\varphi_i \circ \varphi_j^{-1}$ *are called transition maps, while the collection of pairs* $\{U_i, \varphi_i\}$ *are called charts (systems of coordinates) of* $\mathcal{M}$*. We call the image* $\varphi_i(U_i)$ *the coordinate neighbourhood at* $p \in U_i \subset \mathcal{M}$.

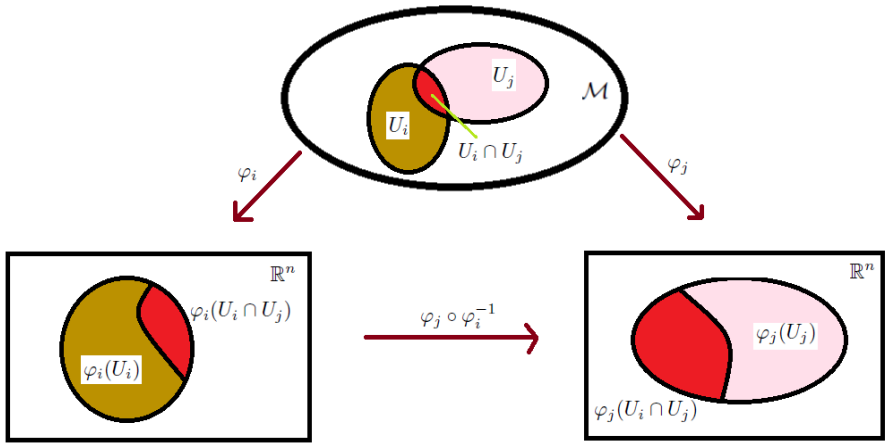


Figure 2.1: Charts of a manifold.

We illustrate transition maps and charts in Figure 2.1. Moreover, there may be more the one choice of charts for a manifold. For example, there are many ways of map projecting the Earth, e.g. cylindrical projection, conformal projection, etc. Furthermore, manifolds can appear as geometry objects like spheres, tori, cones, cylinders, Klein bottles, etc or as a more complicated space in higher dimension like space of symmetric matrices $\mathcal{S}(n)$, space of symmetric positive definite matrices $\mathcal{SP}(n)$, orthogonal group $\mathcal{O}(n)$, special orthogonal group $\mathcal{SO}(n)$, real projective plane and many more.

The special thing about a smooth manifold is the differentiability, so let us review definition of some important terminologies, e.g. tangent vector, differential

map, vector field, etc. Firstly, we fix the notation of $\mathcal{M}$ and $\mathcal{N}$ for differentiable manifolds with dimension m, n respectively.

Definition 2.2 (Differentiable map)**.** *[46] A map $F : \mathcal{M} \to \mathcal{N}$ is differentiable if for every $p \in \mathcal{M}$, there exists a coordinate neighbourhood U covering p with the corresponding transition map φ and $V = F(U) \subset \mathcal{N}$ covering $F(p)$ with the transition map ϕ such that $\phi \circ F \circ \varphi^{-1} : \varphi(U) \to \phi(V)$ is differentiable. This relationship is illustrated in Figure 2.2.*

We call the map $\phi \circ F \circ \varphi^{-1}$ expression of F in local coordinates of ϕ and φ.

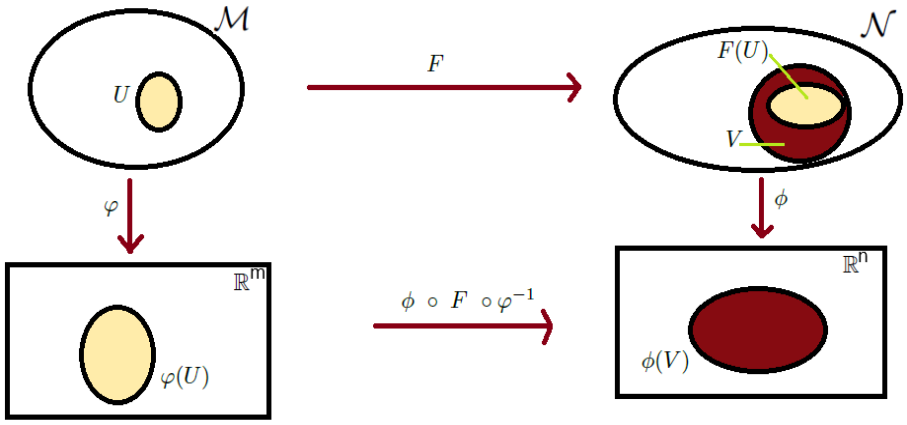


Figure 2.2: Differentiable map between two differentiable manifolds.

Before defining a tangent vector, let us denote the differentiability class for a non-negative integer k:

$$C^k(\mathcal{M}) = \{f : \mathcal{M} \to \mathbb{R} \mid \text{ the derivatives } f', \ldots, f^{(k)} \text{ exist and are continuous}\}$$

$$C^\infty(\mathcal{M}) = \{f : \mathcal{M} \to \mathbb{R} \mid f \text{ has derivatives of all orders}\}$$

Definition 2.3 (Tangent vector)**.** *[47] Let us consider some point $p \in \mathcal{M}$ and the differentiable function $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ such that $\gamma(0) = p$. Then the tangent vector to the curve γ at $t = 0$ is the linear function*

$$\gamma'(0) : C^1(\mathcal{M}) \to \mathbb{R}^m \qquad \gamma'(0)\big(f\big) = \frac{d}{dt}\big(f \circ \gamma\big)\Big|_{t=0} \quad \textit{for } f \in C^1(\mathcal{M}).$$

We define the tangent vector at $p \in \mathcal{M}$ as the tangent vector to some differentiable curve on $\mathcal{M}$ that passes through p at $t = 0$. We denote T_pM, the tangent space of $\mathcal{M}$ at p, as the collection of all tangent vectors at p, that is

$$T_p\mathcal{M} = \{\gamma'(0) \mid \gamma \text{ is a differentiable curve on } \mathcal{M} \text{ such that } \gamma(0) = p\},$$

and $T\mathcal{M} = \bigcup_{p \in M} T_p\mathcal{M}$ is the tangent bundle of $\mathcal{M}$.

In addition, we introduce the definition of differential map, which allows us to map tangent vectors from one tangent space to another tangent space linearly.

Definition 2.4 (Differential map)**.** *[46] Let $F : \mathcal{M} \to \mathcal{N}$ be a differentiable map. Consider a point $p \in \mathcal{M}$ and a tangent vector $v \in T_p\mathcal{M}$ with a differentiable curve γ on $\mathcal{M}$, which satisfies $\gamma(0) = p$, $\gamma'(0) = v$. We denote $\theta = F \circ \gamma$, so θ is a curve on $\mathcal{N}$ with $\theta(0) = F(p)$.*

The differential of F at p is a linear map $dF_p : T_p\mathcal{M} \to T_{F(p)}\mathcal{N}$, given by

$$dF_p(v) = \theta'(0) = \frac{d}{dt}\big(F \circ \gamma\big)\Big|_{t=0}.$$

In the literature, $dF_p(v)$ is sometimes referred to as the push-forward of v by F as it pushes the tangent vector v on tangent space of $\mathcal{M}$ forward onto the tangent space of $\mathcal{N}$, see the illustration in Figure 2.3.

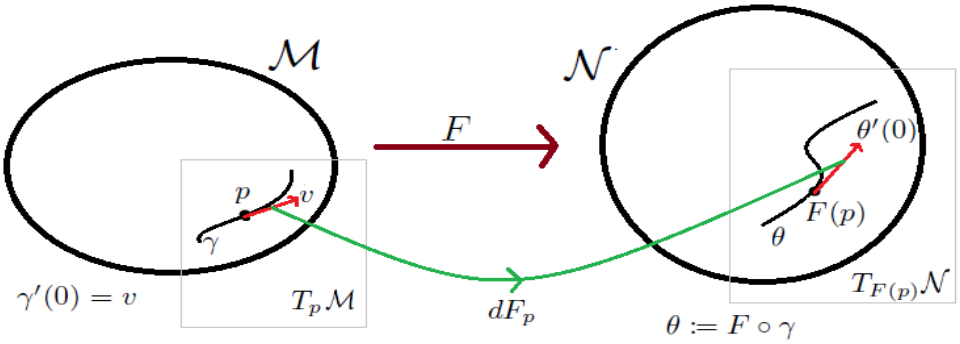


Figure 2.3: Differential map of a differentiable function between two manifolds.

If we want to study the manifold intrinsically, we need to understand carefully what a vector field is. We give two approaches of defining vector fields on a manifold in the following definition together with their expressions in terms of local charts.

Definition 2.5 (Vector field)**.** *[45]*

(i) A vector field X on $\mathcal{M}$ is a function assigning to each point $p \in \mathcal{M}$ an element X_p of the tangent space at p, $T_p\mathcal{M}$. Moreover, we can express the vector field in terms of some local chart $x = \{x^{(1)}, \dots, x^{(m)}\}$ on an open subset $U \subset \mathcal{M}$ covering p as

$$X_p = \sum_{i=1}^{m} x^{(i)}(p) \; \frac{\partial}{\partial x^{(i)}}.$$

(ii) Alternatively, if X satisfies the following for all $f, g \in C^\infty(\mathcal{M})$ and $\alpha, \beta \in \mathbb{R}$:

(a) $[X(\alpha f + \beta g)](p) = \alpha(Xf)(p) + \beta(Xg)(p)$ *(linearity),*

(b) $[X(fg)](p) = [(Xf)(p)]g(p) + f(p)[(Xg)(p)$ *(Leibniz product),*

we say X is a vector field on $\mathcal{M}$. And X is said to be smooth if and only if for all $f \in C^\infty(\mathcal{M})$: $Xf \in C^\infty(\mathcal{M})$. In other words, we define the smooth vector field X as a derivation which maps $C^\infty(\mathcal{M})$ onto itself. Moreover, Xf can also be expressed in terms of the previous local chart x as

$$(Xf)_p = \sum_{i=1}^{m} x^{(i)}(p) \; \frac{\partial f}{\partial x^{(i)}} \qquad\qquad f \in C^\infty(\mathcal{M}).$$

In fact, Xf is proven to be independent of the parametrization x, see [45].

The set of all smooth vector fields on $\mathcal{M}$ is denoted by $\Gamma(T\mathcal{M})$.

On the other hand, if we have a smooth curve $\{\gamma(t) \,|\, t \in I\}$ on $\mathcal{M}$, where I is an interval in $\mathbb{R}$, we call V a *vector field along the curve* γ if V defines a smooth map $I \to T\mathcal{M}$ and $V(t) \in T_{\gamma(t)}\mathcal{M}$, $\forall t \in I$. We denote the collection of all smooth vector fields on $\mathcal{M}$ along γ as $\Gamma_\gamma(T\mathcal{M})$.

Before extending the manifold to a metric space, let us state some important results about vector fields in the following proposition.

Proposition 2.6. *[46, 47] Let X and Y be smooth vector fields on $\mathcal{M}$. There exists a unique smooth vector field Z such that $Zf = (XY - YX)f$, $\forall f \in C^\infty(M)$. We call this operation that acts on vector fields X, Y as the bracket operation, denoted*

by $[X, Y] \coloneqq XY - YX = Z$. *Moreover, this bracket operation satisfies for all* $a, b \in \mathbb{R}$ *and* $f, g \in C^\infty(\mathcal{M})$*:*

(i) $[X, Y] = -[X, Y]$ *(anti-commutativity).*

(ii) $[aX + bY, Z] = a[X, Z] + b[Y, Z]$ *(linearity).*

(iii) $[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$ *(Jacobi identity).*

(iv) $[X, X] = 0$ *(alternativity).*

(v) $[fX, gY] = fg[X, Y] + fX(g)Y - gY(f)X.$

Proposition 2.6 defines a binary operation $[\cdot, \cdot]$ on $\Gamma\big(T\mathcal{M}\big)$, which sometimes is referred to as the Lie bracket. In fact, the space of special vector fields, namely left invariant vector fields, on a smooth manifold forms a Lie algebra equipped with this Lie bracket, see Remark 2.27. We discuss Lie algebra further in Section 2.2.

2.1.2 Riemannian manifold

Since notions of length, angle and shortest distance only make sense on a metric space, ideally we would like to equip the manifold with some metric tensors. First of all, let us look for something similar in a simpler case, the Euclidean space. The inner product of two vectors is commonly used to define these notions, for example:

$$\begin{aligned}
&\text{Length of } x \in \mathbb{R}^m \quad : \quad l(x) = ||x||_2 = \sqrt{\langle x, x\rangle_2} = \sqrt{x^T x},\\
&\text{“Length” of } A \in \mathbb{R}^{m\times m} \quad : \quad l(A) = ||A||_F = \sqrt{\langle A, A\rangle_F} = \sqrt{\text{tr}\,(A^T A)},\\
&\text{Shortest distance between } x,\, y \in \mathbb{R}^m \quad : \quad d(x, y) = ||x - y||_2,\\
&\text{“Shortest distance” between } A,\, B \in \mathbb{R}^{m\times m} \quad : \quad d(A, B) = ||A - B||_F,
\end{aligned}$$

where x^T is the transpose of x and tr is the trace operator. Moreover, the norm $||\cdot||_2$ defined by the inner product $\langle\cdot, \cdot\rangle$ is sometimes referred to as the Euclidean norm or the L^2 norm, while the norm $||\cdot||_F$ is induced from the Frobenius inner product. The fact that we assign the notion of length to something as complicated as a matrix shows that, even in linear algebra, we already use the notion of length generalized beyond everyday meaning of the word “length”.

Accordingly, the required metric tensor for the manifold should preserve properties that the Eucludean inner product has, while act as point-dependent geometry of

inner product spaces. We start this subsection by defining the Riemannian metric on $\mathcal{M}$, and then studying when two Riemannian manifolds have equivalent structures, such as preserving the notion of distance between points (i.e. isometry).

Definition 2.7 (Riemannian metric)**.** *[46] Consider a differentiable manifold $\mathcal{M}$ and a function g which assigns to each point $p \in \mathcal{M}$ a bilinear form (i.e. linear in both arguments) $g_p : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}_{\geq 0}$. We call g a Riemannian metric on $\mathcal{M}$ if for all $p \in \mathcal{M}$, g_p satisfies*

(i)	$g_p(u,v) = g_p(v,u)$	$\forall v, u \in T_p\mathcal{M}$	*(symmetric),*
(ii)	$g_p(u,u) \in \mathbb{R}_{>0}$	$\forall u \in T_p\mathcal{M}\backslash\{0\}$	*(positive definiteness).*

A differentiable manifold equipped with a Riemannian metric is consequently called a Riemannian manifold.

Definition 2.8 (Isometry)**.** *[46] Let $\mathcal{M}$ and $\mathcal{N}$ be Riemannian manifolds equipped with the metric $g^{\mathcal{M}}$ and $g^{\mathcal{N}}$ respectively. A diffeomorphism map $F : \mathcal{M} \to \mathcal{N}$ is called an isometry if for all $p \in \mathcal{M}$ and $u, v \in T_p\mathcal{M}$:*

$$g^{\mathcal{M}}_p(u,v) = g^{\mathcal{N}}_{F(p)}(dF_p(u), dF_p(v)).$$

Here dF_p is the differential of F at p. We therefore said two Riemannian manifolds $\mathcal{M}$ and $\mathcal{N}$ are isometric.

Until now, we review first order differentiability, such as differential map, vector field, etc, the next natural thing is studying higher order. For instance, we want to be able to differentiate a vector field along another vector field, which brings up the notion of a connection on the manifold.

Definition 2.9 (Affine connection)**.** *[47] An affine connection ∇ on a differentiable manifold $\mathcal{M}$ is a map $(X,Y) \overset{\nabla}{\mapsto} \nabla_X Y$, which satisfies for all $X, Y, Z \in \Gamma(T\mathcal{M})$ and $f, g \in C^\infty(\mathcal{M})$:*

(i) $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$,

(ii) $\nabla_X(Y+Z) = \nabla_X Y + \nabla_X Z$,

(iii) $\nabla_X(fY) = f(\nabla_X Y) + X(f)Y$.

$\nabla_X Y$ is called the covariant derivative of Y along X.

In addition, for any $p \in \mathcal{M}$, the tangent vector $\nabla_X Y|_p$ depends only on tangent vector X_p and the values of Y in the neighbourhood of p, see [46]. We call $\nabla_X Y|_p$ the *directional derivative* at p in the direction X_p, and therefore it is often written $\nabla_{X_p} Y$ instead. Whence, let us state the result that the covariant derivative along a differentiable curve on $\mathcal{M}$ can be defined uniquely.

Proposition 2.10. *[47] Suppose that the differentiable manifold $\mathcal{M}$ has an affine connection ∇ and a smooth curve $\{\gamma(t) \,|\, t \in I\}$, where I is an interval in $\mathbb{R}$. There exists a unique operator $\mathscr{D}_t : \Gamma_\gamma(T\mathcal{M}) \to \Gamma_\gamma(T\mathcal{M})$ such that for any $V, W \in \Gamma_\gamma(T\mathcal{M})$ and $f \in C^\infty(\mathcal{M})$:*

(i) $\mathscr{D}_t(V+W) = \mathscr{D}_t V + \mathscr{D}_t W$,

(ii) $\mathscr{D}_t(fV) = f'(t)V + f\mathscr{D}_t V$.

(iii) If V is induced from a vector field $Y \in \Gamma(TM)$, i.e. $\forall t \in I : V(t) = Y_{\gamma(t)}$, then $\mathscr{D}_t V = \nabla_{\gamma'(t)} Y$.

We call $\mathscr{D}_t V$ the covariant derivative of vector field V along the curve γ.

The value of the covariant derivative impacts on the direction of how the corresponding vector fields travel along the curve. Figure 2.5 illustrates examples of vector fields (green color) that have non-zero covariant derivative along the top curve, and vector fields (red color) that has zero covariant derivative at all points on the bottom curve. In addition, the connection ∇ can be expressed as coefficients with respect to some basis in local coordinates, which will become useful in the later contexts, such as parallel transports, geodesics, etc.

Definition 2.11 (Christoffel symbols)**.** *[25] Suppose $\mathcal{M}$ is of dimension m and $x = \{x^{(1)}, \ldots, x^{(m)}\}$ is a local chart on an open subset $U \subset \mathcal{M}$, then the vector*

fields $X_i = \partial/\partial x^{(i)}$ *spans the tangent space* $T_p\mathcal{M}$ *for any* $p \in U$. *The Christoffel symbols* Γ_{ij}^k *are functions on* U *defined uniquely by the following relation*

$$(\nabla_{X_i} X_j)_p = \sum_{k=1}^{m} \Gamma_{ij}^k(p)\, X_k.$$

To put it another way, the Christoffel symbols tell us how the basis field change at different points on the manifold. We know that on $\mathbb{R}^m$ equipped with the Euclidean metric, basis on tangent space does not change its direction, and in this case we have zero Christoffel symbols everywhere. However, Figure 2.4 illustrates an opposite picture on the sphere $\mathcal{S}^2$, where the basis for tangent spaces at different points on $\mathcal{S}^2$ are required to adapt with the bending of the surface, and in this case we say $\mathcal{S}^2$ has non-zero Christoffel symbols. We therefore often distinguish manifolds from the Euclidean space, that have the structure of basis field changing from point to point on the manifold, as curved spaces. An alternative indication is the curvature, in which manifolds having zero curvature everywhere are classified to have Euclidean structure. Since a Riemannian manifold can have dimension higher than 2, there are many ways of expressing the curvature. We name some of them in Definition 2.12.

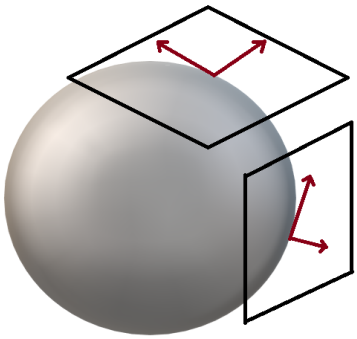


Figure 2.4: Basis for tangent plane changes from point to point on the sphere $\mathcal{S}^2$.

Definition 2.12. *[25] Suppose that the differentiable manifold* $\mathcal{M}$ *of dimension* m *has an affine connection* ∇, *and a Riemannian metric* g.

(i) The Riemannian curvature tensor R *is defined by*

$$R(X,Y)Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z \qquad \forall X,\, Y,\, Z \in \Gamma(T\mathcal{M}),$$

where $[\cdot,\cdot]$ *is the bracket of two vector fields.*

(ii) The sectional curvature tensor K *for any* $p \in \mathcal{M}$ *is defined by* $K(X,Y)_p = g_p\big(R(X,Y)Y,X\big)$ *for any* $X,\ Y \in \Gamma(T\mathcal{M})$.

(iii) The Ricci curvature tensor Ric : $\Gamma(T\mathcal{M}) \times \Gamma(T\mathcal{M}) \to \mathbb{R}_{\geq 0}$ *is defined for any* $p \in \mathcal{M}$ *by*

$$Ric(X,Y)_p = \sum_{i=1}^{m} g_p\big(R(E_i,X)E_i,Y\big) \qquad \text{for } X,Y \in \Gamma(T\mathcal{M}),$$

where $\{E_1,\ldots,E_m\}$ *is an orthonormal basis on* $T_p\mathcal{M}$ *with respect to* g.

(iv) The scalar curvature at any point $p \in \mathcal{M}$, S_p, *is the fully contracted curvature tensor, i.e.* $S_p = \sum_{i=1}^{m} K(E_i,E_i)_p$, *where* $\{E_1,\ldots,E_m\}$ *is an orthonormal basis on* $T_p\mathcal{M}$.

Definition 2.13 (Torsion)**.** *The torsion tensor* $\mathfrak{T}(X,Y)$ *of a smooth manifold* $\mathcal{M}$ *with an affine connection is given by*

$$\mathfrak{T}(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y] \qquad \text{for } X,Y \in \Gamma(T\mathcal{M}).$$

We say $\mathcal{M}$ *is torsion-free if its torsion tensor vanishes identically.*

So far, we define and use the affine connection ∇ in a way which does not relate to the Riemannian metric. However, for each Riemannian metric tensor there is a unique torsion-free affine connection, called the Levi-Civita connection.

Theorem 2.14. *[47] (Levi-Civita) Given a Riemannian manifold* $\mathcal{M}$ *equipped with metric* g, *there exists a unique affine connection* ∇ *on* M *that satisfies:*

(i) ∇ *is symmetric, i.e.* $\mathcal{M}$ *is torsion-free.*

(ii) ∇ *is compatible with the Riemannian metric* g, *that is*

$$Xg(Y,Z) = g(\nabla_X Y,Z) + g(Y,\nabla_X Z) \qquad \forall X,Y,Z \in \Gamma(T\mathcal{M})$$

We refer to this special connection as the Levi-Civita connection.

Hereafter, for any Riemannian manifold we fix the affine connection to be the Levi-Civita connection, which implies the Christoffel symbols are now symmetric, i.e. $\Gamma_{ij}^k = \Gamma_{ji}^k$, $\forall i, j, k \in \{1, \ldots, m\}$.

From now on, we will assume that m-dimensional manifold $\mathcal{M}$ comes equipped with a Riemannian metric g and the Levi-Civita connection ∇. The rest of this subsection is carried out as follows. Firstly, we study how one generalizes the notion of a straight line on the Euclidean space onto a general Riemannian manifold, namely geodesic. Secondly, we present so-called the exponential/logarithm maps, that map from the tangent space onto the manifold itself or vice versa, followed by some important results and definitions related to these maps. And lastly, we introduce the notion of Riemannian gradient (i.e. first order operator) and the Laplace-Beltrami operator (i.e. second order elliptic operator) to a $C^2(\mathcal{M})$ function, which will be required when studying stochastic differential equations on manifolds.

First of all, to achieve the minimum distance between two points on $\mathcal{M}$, intuitively we want the geodesic connecting these two points to have a constant speed from the start, i.e. acceleration equals to zero. In physics, we can interpret the motion on the geodesic as being completely determined by the bending of the surface, which is caused purely by the gravity, and not any other external forces.

Zero acceleration implies the acceleration vector is perpendicular to the tangent space at any point on the curve. Since the tangent space may change when moving along the curve, so we need to be able to move freely between different tangent spaces. Previously, we say that using the differential map one can map linearly a tangent vector from one tangent space onto another tangent space. However, if we want to study how a tangent vectors varies along a smooth curve on $\mathcal{M}$, the concept of differential map is not enough as it does not connect the geometries of nearby points. Thus, we need something which can affix the connection locally, and parallel transport describes exactly this action, see Figure 2.5. Let us consider a smooth curve $\gamma : I \to \mathcal{M}$, where I is a compact interval in $\mathbb{R}$.

Definition 2.15 (Parallel transport)**.** *[46] A vector field $V \in \Gamma_\gamma(T\mathcal{M})$ is said to be parallel along the curve if $\mathscr{D}_t V = 0$ at every point of the curve. In this case, we say*

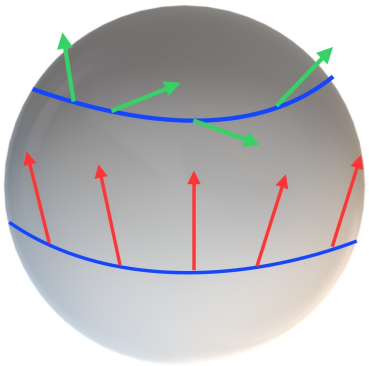


Figure 2.5: Vector fields (red colour) travel parallel along the bottom curve, whereas vector fields (green colour) do not move parallel along the top curve on the sphere.

$V_{\gamma(t)}$ is the parallel transport of $V_{\gamma(0)}$ along the curve γ.

Definition 2.16 (Geodesic)**.** *[47] We call $\gamma(t)$ geodesic if $\forall t \in I : \mathscr{D}_t\gamma'(t) = 0$. In other words, if the vector field $V \in \Gamma_\gamma(T\mathcal{M})$ such that $V(t) = \gamma'(t)$, γ is a geodesic on $\mathcal{M}$ if and only if V is parallel along γ.*

Let us now express conditions of parallel transport and geodesic through Christoffel symbols in local coordinates. Consider a chart $x = \{x^{(1)}, \ldots, x^{(m)}\}$ on an open subset $U \subset \mathcal{M}$, and denote $X_i = \partial/\partial x^{(i)}$.

We write $\gamma(t) = \sum_{i=1}^{n} \gamma^{(i)}(t)\, x^{(i)}$,

$$\Rightarrow \gamma'(t) = \sum_{i=1}^{m} \frac{d\gamma^{(i)}}{dt} X_i \quad \text{and} \quad V_{\gamma(t)} = \sum_{i=1}^{m} v^{(i)}(t)\, X_i.$$

So, V is parallel to γ if and only if for all $k \in \{1, \ldots, m\}$:

$$\frac{dv^{(k)}}{dt} + \sum_{i,j=1}^{m} \Gamma_{ij}^{k}(\gamma(t)) \frac{d\gamma^{(i)}}{dt}\, v^{(j)}(t) = 0. \tag{2.1}$$

Similar to Equation (2.1), γ is a geodesic curve on $\mathcal{M}$ if and only if the following equations hold for all $k \in \{1, \ldots, m\}$:

$$\frac{d^2\gamma^{(k)}}{dt^2} + \sum_{i,j=1}^{m} \Gamma_{ij}^{k}(\gamma(t)) \frac{d\gamma^{(i)}}{dt} \frac{d\gamma^{(j)}}{dt} = 0. \tag{2.2}$$

Here, Equation (2.2) constitutes a system of non-linear second order ordinary dif-

ferential equations (ODEs). Thus, the solution exists and is uniquely determined by the initial condition $\gamma(0)$ and the initial direction $\gamma'(0)$. However, geodesics are only locally distance minimizing paths as these ODEs only hold in U, see [46].

Next, let us introduce a local diffeomorphism map from the tangent space onto the manifold itself, namely the exponential map.

Theorem 2.17 (Exponential map)**.** *[45] Suppose for some point $p \in \mathcal{M}$ with a chart $\{U, \phi\}$ covering p, there exists a neighbourhood $V \subset \mathcal{M}$ covering p and a positive number ϵ such that whenever $q \in V$, $X_q \in T_q\mathcal{M}$ with $||X_q||_2 < \epsilon$. Then there is a unique geodesic $\gamma(t) = \omega(t, q, X_q)$ defined for $t \in (-2, 2)$, and with $\gamma(0) = q$, $\gamma'(t) = X_q$. Moreover, the map ω is a smooth map on the open set $\{t \in (-2, 2),\ q \in V,\ X_q \in T_q\mathcal{M}$ with $||X_q||_2 < \epsilon\}$ and has its value in U.*

Therefore, the exponential map $Exp_q : \mathfrak{V} \to \mathcal{M}$ at q is uniquely defined by

$$Exp_q(X_q) = \gamma(1) = \omega(1, q, X_q),$$

where the open set $\mathfrak{V} = \{X_q \,|\, q \in V, ||X_q||_2 < \epsilon\} \subset T_q\mathcal{M}$.

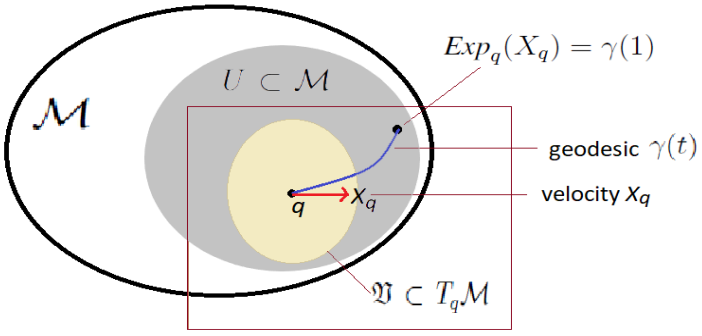


Figure 2.6: The exponential map on a Riemannian manifold.

Intuitively, the exponential map $\text{Exp}_q(X_q) \in \mathcal{M}$ is obtained by starting at q, travelling on the unique geodesic curve $\gamma(t)$ with the velocity X_q after a unit of time, see Figure 2.6. The geodesic $\gamma(t)$ is only locally defined and locally differentiable in Theorem 2.17, because we can't guarantee that the solution of ODEs in Equation (2.2) won't blow up in finite time. This behaviour is similar to the property of locally distance minimizing from geodesics.

Definition 2.18 (Geodesically complete manifold)**.** *We say $\mathcal{M}$ is a geodesically complete manifold if $\mathcal{M}$ is a Riemannian manifold for which, the exponential map at any point $p \in \mathcal{M}$, is well-defined on the entire tangent space $T_p\mathcal{M}$.*

The condition of having geodesical completeness is not a huge issue as it covers a vast variety of Riemanian manifolds, particularly the spaces we are mainly focusing on in this thesis, such as $\mathcal{SO}(3)$, $\mathcal{SP}(n)$, are some typical examples.

Corollary 2.19 (Logarithm map)**.** *Theorem 2.17 implies $Exp_q : \mathfrak{V} \to \mathcal{M}$ is diffeomorphism, it is therefore valid to define the inverse of exponential map on V, the so-called logarithm map $Log_q : V \to \mathfrak{V}$ such that $Exp_q \circ Log_q = Id$, $\forall q \in V$.*

Consequently, logarithm map is also locally differentiable, same as the exponential map. We interpret the logarithm map as a local projection from $\mathcal{M}$ onto its tangent space. In addition, the unique geodesic $\gamma(t)$ in Theorem 2.17 yields smallest distance only when it stays inside U. Thus, there is a threshold for t such that γ will not minimize distance once t passes beyond this threshold.

Definition 2.20 (Cut locus)**.** *[25] Choose an arbitrary $p \in \mathcal{M}$ and some $v \in T_p\mathcal{M}$. Let t_v be the largest t such that the geodesic $\nu(t) = Exp_p(tv)$ is the unique minimizing geodesic between its end points. We define*

$$\tilde{\mathscr{C}_p} = \{t_w w \,|\, w \in T_p\mathcal{M}, ||w||_2 = 1\} \quad \textit{and} \quad \mathscr{C}_p = \{Exp_p(v) \,|\, v \in \tilde{\mathscr{C}_p}\},$$

then $\mathscr{C}_p$ is called the cut locus of p, which is the image of $\tilde{\mathscr{C}_p}$ by the exponential map.

In a similar manner, we can define the set $\mathscr{E}_p$ within the cut locus of p:

$$\tilde{\mathscr{E}_p} = \{tw \,|\, w \in T_p\mathcal{M},\, t \in [0, t_w),\, ||w||_2 = 1\} \quad \textit{and} \quad \mathscr{E}_p = \{Exp_p(v) \,|\, v \in \tilde{\mathscr{E}_p}\}.$$

The following results about the cut locus of a geodesically complete manifold will be requisite for our methodology in Chapter 3.

Theorem 2.21. *[25, 48] Suppose $\mathcal{M}$ is geodesically complete, then*

(i) The exponential map $Exp_p : \tilde{\mathscr{E}_p} \to \mathscr{E}_p$ is a diffeomorphism for any $p \in \mathcal{M}$.

(ii) The cut locus $\mathscr{C}_p$ of p is a closed subset of measure zero for any $p \in \mathcal{M}$.

(iii) For $p, q \in \mathcal{M}$, if $p \in \mathscr{C}_q$ then $q \in \mathscr{C}_p$.

(iv) $\mathscr{E}_p$ and $\mathscr{C}_p$ are disjoint and $\mathcal{M} = \mathscr{E}_p \cup \mathscr{C}_p$ for any $p \in \mathcal{M}$

Lastly, our work in Chapter 4 requires the calculation of the Riemannian gradient of some functions $f \in C^2\big(\mathcal{SP}(n)\big)$, such as squared distance, etc. Thus, we would like to define it formally and give its expression in terms of some local charts.

The Riemannian gradient of f, denoted by ∇f, is a unique vector field (i.e. first order derivative) defined by the following relation

$$g_p(\nabla f, X) = df(X_p) = (Xf)_p \qquad \forall X \in \Gamma(T\mathcal{M}), p \in \mathcal{M}. \tag{2.3}$$

Consider a chart $x = \{x^{(1)}, \ldots, x^{(m)}\}$ on an open subset U covering a point $p \in \mathcal{M}$, and denote $X_i = \partial/\partial x^{(i)}$. Therefore, $\{X_1, \ldots, X_m\}$ forms an basis field defined in U, and the Riemannian metric g can be expressed with respect to this basis as a matrix form, denoted by G. We then can express ∇f in local coordinates:

$$(\nabla f)_q = \sum_{i,j=1}^{m} \frac{\partial f}{\partial x^{(i)}} G^{ij}(q)\, X_j \qquad \forall q \in U,$$

with G^{ij} is the (i, j) entry of G^{-1}. Particularly, let us state a famous result from [49] for some fixed point $q \in \mathcal{M}$:

$$\Big(\nabla d^2(p, q)\Big)_p = -2\,\mathrm{Log}_p(q) \qquad \forall p \in \mathcal{M}. \tag{2.4}$$

In fact, we prove Equation (2.4) for a specific Riemannian manifold, $\mathcal{SP}(n)$ endowed with different Riemannian metrics, see Proposition 4.2 and 4.3.

Moving from first to second order differential operators, among the second order differential operators, the Laplace-Beltrami operator is perhaps of greatest interest. The Lapalace-Beltrami operator on $\mathcal{M}$, denoted by $\Delta_{\mathcal{M}}$, is an intrinsically defined second order elliptic operator, which generalizes the Laplace operator known from Euclidean differential calculus. $\Delta_{\mathcal{M}}$ can be written using the trace of the Hessian of

the function f associated with the connection, that is

$$\Delta_{\mathcal{M}} f = \sum_{i,j=1}^{m} \Delta_{\mathcal{M}} f(X_i, X_j) = \sum_{i,j=1}^{m} G^{ij} \big(\mathrm{Hess} f\big)_{ij}, \tag{2.5}$$

where $\big(\mathrm{Hess} f\big)_{ij} = \mathrm{Hess} f(X_i, X_j) = X_i(X_j f) - (\nabla_{X_i} X_j) f$. Thus, if $\{X_i\}_{i=1}^m$ is orthonormal with respect to g, G equals to an indentity matrix and so $\Delta_{\mathcal{M}} f = \mathrm{tr}\big(\mathrm{Hess} f\big)$. Moreover, we can also express $\Delta_{\mathcal{M}} f$ in local coordinates, see [26]:

$$\Delta_{\mathcal{M}} f = \sum_{i,j=1}^{m} \left\{ G^{ij} \frac{\partial}{\partial x^{(i)}} \frac{\partial f}{\partial x^{(j)}} + \frac{1}{\sqrt{\det G}} \frac{\partial \big(\sqrt{\det G} \cdot G^{ij}\big)}{\partial x^{(i)}} \frac{\partial f}{\partial x^{(j)}} \right\}. \tag{2.6}$$

2.2 Lie group and Lie algebra

Lie groups have been researched extensively in mordern geometry. They appear in many areas, such as physics, biomechanics and engineering, etc. The intuition stems from the idea of replacing the global concept of a group with the infinitesimal group, its local version, which is commonly known as the Lie algebra. In order to achieve this, this group of interest is required to have differentiable structure.

Definition 2.22 (Lie group)**.** *The set $\mathcal{G}$ equipped with a binary operation $* : \mathcal{G} \times \mathcal{G}$ is a Lie group if $(\mathcal{G}, *)$ is a group, that is*

(i) $\forall a, b \in \mathcal{G} \,:\, a * b \in \mathcal{G}$ *(closure),*

(ii) $\forall a, b, c \in \mathcal{G} \,:\, (a * b) * c = a * (b * c)$ *(associativity),*

(iii) $\exists e \in \mathcal{G}$ *satisfies* $\forall a \in \mathcal{G} \,:\, e * a = a = a * e$ *(identity element),*

(iv) $\forall a \in \mathcal{G}$, $\exists a^{-1} \in \mathcal{G}$ *satifies* $a^{-1} * a = e = a * a^{-1}$ *(inverse element),*

and $\mathcal{G}$ is a smooth manifold in which the map $(a, b) \mapsto a * b^{-1}$ is smooth.

Definition 2.23 (Compact connected group)**.** *We say the Lie group $\mathcal{G}$ is connected if it cannot be represented as the union of two or more disjoint non-empty open subsets. Furthermore, if any sequence in $\mathcal{G}$ has a convergent sub-sequence, the connected Lie group $\mathcal{G}$ is now called a compact connected Lie group.*

A typical example of Lie groups, namely the Euclidean group $\mathcal{E}_3$, comes from

isometric action, such as rotations, reflections, translations, etc. Other examples in matrix form, that are Lie groups when equipping the standard matrix multiplication, include the space of invertible matrices $\mathcal{GL}(n)$, space of orthogonal matrices $\mathcal{O}(n)$, space of special orthogonal matrices $\mathcal{SO}(n)$. While $\mathcal{O}(n)$ is not a connected Lie group, $\mathcal{SO}(n)$ is connected.

We fix the notation of $\mathcal{G}$ be a Lie group equipped with the binary operation $*$ and study the invariant Riemannian metrics on $\mathcal{G}$.

Definition 2.24 (Left/right translation)**.** *We define for $a \in \mathcal{G}$ the left translation $L_a : b \mapsto a * b$ and the right translation $R_a : b \mapsto b * a$.*

Definition 2.25 (Invariant Riemannian metric)**.** *[47] Since $\mathcal{G}$ is a Lie group, both left and right translations are diffeomorphisms. Consider a Riemannian metric tensor g on $\mathcal{G}$, then g is said to be left (or right) invariant if L_a (or R_a) are isometric with respect to g for all $a \in \mathcal{G}$. Moreover, a Riemannian metric that is both left and right invariant is said to be bi-invariant.*

In similar manner, we say vector field $X \in \Gamma(T\mathcal{G})$ is left invariant if $\forall a, b \in \mathcal{G} : (dL_a)(X_b) = X_{a*b}$ and right invariant if $\forall a, b \in \mathcal{G} : (dR_a)(X_b) = X_{b*a}$. Furthermore, any tangent vector $X_e \in T_e\mathcal{G}$ at the identity element $e \in \mathcal{G}$ can be extended to a left invariant vector field via the push-forward of X_e by the left translation L_b for any $b \in \mathcal{G}$, i.e. $X_b = (dL_b)X_e$. A similar approach can be applied to construct a right invariant vector field.

Definition 2.26 (Lie algebra)**.** *The tangent space at the identity element $e \in \mathcal{G}$ is called the Lie algebra, denoted by $\mathfrak{g}$. The Lie algebra $\mathfrak{g}$ is a vector space equipped with a binary operation $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, the so-called Lie bracket, which satisfies all five properties in Proposition 2.6.*

Remark 2.27. The bracket of two left invariant vector fields is again a left invariant vector field, that is if $X, Y \in \Gamma(T\mathcal{G})$ are left invariant, $dL_a[X, Y]f = [X, Y]f$, $\forall a \in \mathcal{G}, f \in C^\infty(\mathcal{G})$, see [47]. Therefore, the space of left invariant vector fields on a Lie group is a Lie algebra equipped with the Lie bracket of vector fields.

Proposition 2.28. *[45] Consider an arbitrary inner product $g_e(\cdot,\cdot)$ on the Lie algebra $\mathfrak{g}$ of $\mathcal{G}$, we define the following metric tensor using the left translation:*

$$g_x(u,v) = g_e\big((dL_{x^{-1}})_x(u), (dL_{x^{-1}})_x(v)\big) \qquad \forall u,v \in T_x\mathcal{G},\ x \in \mathcal{G}.$$

This construction induces a left invariant Riemannian metric on $\mathcal{G}$.

Corollary 2.29. *Suppose an inner product g_e on the Lie algebra $\mathfrak{g}$ of $\mathcal{G}$ satisfies*

$$g_e([u,v],w) + g_e(u,[w,v]) = 0 \qquad \forall u,v,w \in \mathfrak{g},$$

then the Riemannian metric constructed from g_e in Proposition 2.28 is bi-invariant.

Proof of Proposition 2.28 and Corollary 2.29 can be easily shown using Definition 2.25 of a Riemannian metric being left invariant or bi-invariant.

Proposition 2.30. *[50] There exists a bi-invariant metric on $\mathcal{G}$ if and only if the adjoint representation $Ad(\mathcal{G})$ of $\mathcal{G}$ is relatively compact, i.e. the group of matrices given by $Ad(\mathcal{G})$ has finite dimension.*

Here, the adjoint representation $Ad(\mathcal{G})$ of the Lie Group $\mathcal{G}$ is defined by the differential of the conjugation action Φ at the identity of $\mathcal{G}$ such that for $x \in \mathcal{G}$, $\Phi_x : (x,y) \mapsto xyx^{-1}$, $\forall y \in \mathcal{G}$.

Consequently, if $\mathcal{G}$ is Abelian (i.e. commutative group) of finite dimension, there exists a bi-invariant metric on $\mathcal{G}$. On the other hand, if $\mathcal{G}$ is non-Abelian group of finite dimension but is compact and connected, the adjoint representation of $\mathcal{G}$ is bounded, Proposition 2.30 therefore can still be applied.

Earlier on, we mention the Lie algebra as the infinitesimal group because elements of the Lie algebra are regarded to lie infinitesimally close to the identity element. We therefore introduce the concept of the one-parameter subgroup, which explains the infinitesimal transformation of the Lie algebra, see Theorem 2.33. This concept is closely related to the exponential map of a Lie group. Let us firstly denote the additive group of real numbers $(\mathbb{R},+)$. The map $F : (\mathbb{R},+) \to (\mathcal{G},*)$ is said to be homomorphism if $\forall x,y \in \mathbb{R} : F(x+y) = F(x) * F(y)$.

Definition 2.31 (One-parameter subgroup)**.** *[45] The sub-group H is a one-parameter subgroup of $\mathcal{G}$ if there exists a homomorphsim map $F : (\mathbb{R}, +) \to (\mathcal{G}, *)$ such thatthe sub-group H is the homomorphic image $F(\mathbb{R})$.*

Definition 2.32 (Infinitesimal generator)**.** *[45] Suppose $\theta : \mathbb{R} \times \mathcal{G} \to \mathcal{G}$ is an smooth action, i.e. it is a smooth map that satisfies $\forall t, s \in \mathbb{R}$, $a \in \mathcal{G} : \theta_0(a) = a$ and $\theta_t \circ \theta_s(a) = \theta_{t+s}(a) = \theta_s \circ \theta_t(a)$.*

For $a \in \mathcal{G}$, we define a map $X_a : C^\infty(\mathcal{G}) \to \mathbb{R}$ by

$$X_a f = \lim_{\delta t \to 0} \frac{f(\theta_{\delta t}(a)) - f(a)}{\delta t} \qquad \forall f \in C^\infty(\mathcal{G}),$$

then clearly X is a vector field on $\mathcal{G}$ and is commonly called the infinitesimal generator of the action θ in the context of algebra.

Theorem 2.33. *[45] Given a homomorphism $F : (\mathbb{R}, +) \to (\mathcal{G}, *)$, let H be the corresponding one-parameter subgroup of $\mathcal{G}$. Consider the left invariant vector field X on $\mathcal{G}$ constructed from tangent vector $X_e = F'(0) \in \mathfrak{g}$. Then $\theta_t(x) = xF(t)$ defines a smooth action of $\mathbb{R}$ on $\mathcal{G}$ having X as its infinitesimal generator.*

Conversely, let X be a left invariant vector field on $\mathcal{G}$, there exists a unique action $\theta : \mathbb{R} \times \mathcal{G} \to \mathcal{G}$ such that the map $F(t) = \theta_t(e)$ defines a one-parameter subgroup of $\mathcal{G}$ and $\theta_t(x) = xF(t)$.

In corollary, there is a one to one correspondence between the elements of $\mathfrak{g}$ and one-parameter subgroup of $\mathcal{G}$. In other words, for each $\nu \in \mathfrak{g}$, let $t \to \tilde{F}(t, \nu)$ be the unique one-parameter subgroup of $\mathcal{G}$, then the homomorphism map $\tilde{F} : \mathbb{R} \times \mathfrak{g} \to \mathcal{G}$ is smooth and satisfies $\tilde{F}(t, s\nu) = \tilde{F}(st, \nu)$, $\forall t, s \in \mathbb{R}$. Thus, the exponential map of a Lie group can be defined using a homomorphism map on $\mathfrak{g}$ via the one-parameter subgroup.

Remark 2.34. The exponential map $\text{Exp}_e : \mathfrak{g} \to G$ is given by $\text{Exp}_e(\nu) = \tilde{F}(1, \nu)$ for $\nu \in \mathfrak{g}$. Moreover, for $p \in \mathcal{G}$ and a tangent vector ω on $T_p\mathcal{G}$, we have $dL_{p^{-1}}(\omega) \in T_e\mathcal{G} = \mathfrak{g}$. Therefore, it is valid to define the exponential map at the point p as $\text{Exp}_p : T_p\mathcal{G} \to \mathcal{G}$ by $\text{Exp}_p(\omega) = \text{Exp}_e\big(dL_{p^{-1}}(\omega)\big)$.

Furthermore, if $\mathcal{G}$ has a bi-invariant Riemannian metric, the exponential map defined in Definition 2.17 using this metric is identical to the exponential map defined in Remark 2.34. In this occasion, the geodesics passing through the identity element form a one-parameter subgroup of $\mathcal{G}$, e.g. $\{\mathrm{Exp}_e(t\nu) \,|\, t \in \mathbb{R}, \nu \in \mathfrak{g}\}$.

For a general Lie group $\mathcal{G}$, existence of a bi-invariant Riemannian metric on $\mathcal{G}$ is not guaranteed. Although there is always an invariant Riemannian metric under either left or right translations (Proposition 2.28), the exponential map defined in Definition 2.17 will not in general agree with the exponential map defined in Remark 2.34. In other words, if $\mathcal{G}$ is not equipped with a bi-invariant Riemannian metric, the geodesics through the identity element, each form a one-parameter subgroups of $\mathcal{G}$.

In addition, the exponential maps of Lie groups consisting of square invertible matrices often coincide with the matrix exponential function.

Definition 2.35 (Matrix exponential/logarithm)**.** *Let $A \in \mathcal{GL}(n)$, then the matrix exponential of A is defined by $\exp A = \sum_{k=0}^{\infty} \left(A^k/k!\right)$ if the series converges.*

Conversely, if $B \in \mathcal{GL}(n)$ such that $\exp A = B$, we say A is the matrix logarithm of B, denoted by $\log B = A$. Moreover, if B is sufficiently close to the identity matrix I_n, i.e. $||B - I_n||_F < 1$, the following power series converges and we can compute $\log B$ as

$$\log B = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(B - I_n)^k}{k}.$$

2.3 Stochastic differential equations

We divide this section into two parts: Euclidean space and Riemannian manifold. On the Euclidean space, literature of SDEs is rich with many well-known results. In particular, we briefly review two popular constructions of stochastic integrals: Itô integral and Stratonovich integral. Moreover, conditions for existence and uniqueness of solutions are also discussed in Subsection 2.3.1. We end this subsection by presenting a popular piecewise approximation method for SDEs in Itô calculus, namely the Euler-Maruyama method.

On the other hand, SDEs on a Riemannian manifold present additional complications over the Euclidean setting. In Subsection 2.3.2, we first introduce the general notion of SDEs on manifolds driven by smooth vector fields, $\mathscr{L}$-diffusion and the concept of stochastic development. Secondly, we discuss how one generalizes Brownian motion on the Euclidean space to a Riemannian manifold. This includes not only definition, local representation, construction via the expoential adapted Euler-Maruyama approximation, but also the explosion time and minimal heat kernel. Lastly, we define a class of SDEs resulting from the convergence of the exponential adapted Euler-Maruyama approximation on some special manifolds, while briefly discuss about different writing forms of SDEs defined by E. P. Hsu [25] and K. D. Elworthy [26]. Furthermore, the extension of Itô's lemma and Girsanov's theorem to a Riemannian manifold is presented towards the end of Subsection 2.3.2.

In this thesis, we work with a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ equipped with the filtration $\mathfrak{F}_* = \{\mathfrak{F}_t, t \geq 0\}$ of σ-fields contained in $\mathfrak{F}$. Here, we assume that $\mathfrak{F} = \lim_{t\uparrow\infty} \mathfrak{F}_t$, while $\mathfrak{F}_*$ is right continuous and every null set (i.e. a subset of a set having measure zero) is contained in $\mathfrak{F}_t$.

2.3.1 Euclidean space

There are different types of driving processes for SDEs, such as Brownian motion (or the Wiener process), jump process, etc. In this thesis, we focus on SDEs driven by Brownian motion, which has remarkable complex mathematical characteristics. One particular feature of Brownian motion is nowhere differentiability almost surely, which makes direct use of standard calculus impossible. Therefore, stochastic integral using the derivative of Brownian motion requires new mathematical establishment. And two most common construction of stochastic integrals are Itô and Stratonovich integrals.

Since both Itô and Stratonovich integrals are defined in a similar manner to the Riemann-Stieltjes integral, which takes a limit in probability of Riemann sums, we demonstrate the difference in the following example. Consider a one-dimensional Brownian motion B_t, and a right continuous, $\mathfrak{F}$-adapted and locally bounded process $f : [0, \infty) \times \Omega \to \mathbb{R}$. For a sequence of partitions π_n of $[0, t]$ with mesh tending to

zero, it is natural to approximate f by

$$f(s,\omega) \approx \sum_{[t_j,t_{j+1})\in\pi_n} f(t_j^*,\omega)\,\mathbb{I}_{\{s\in[t_j,t_{j+1})\}}$$

where $t_j^* \in [t_j, t_{j+1})$ and $\mathbb{I}$ denotes the indicator function. The stochastic integral of f with respect to Brownian motion is defined by $\lim_{n\to\infty} \sum_{[t_j,t_{j+1})\in\pi_n} f(t_j^*,\omega)\left(B_{t_{j+1}} - B_{t_j}\right)(\omega)$. The choice of t_j^* varies, for instance:

- If $t_j^* = t_j$, the integral turns out to be the *Itô integral*, denoted by $\int_0^t f(s,\omega)\,dB_s(\omega)$.
- If $t_j^* = (t_j + t_{j+1})/2$, the integral turns out to be the *Stratonovich integral*, denoted by $\int_0^t f(s,\omega) \circ dB_s(\omega)$.

It is clear that $\int_0^t f(s,\omega)\,dB_s(\omega)$ is a martingale as $f(t_j^*,\omega) \in \mathfrak{F}_{t_j}$ in the Itô case, while $\int_0^t f(s,\omega) \circ dB_s(\omega)$ is not as $f(t_j^*,\omega) \notin \mathfrak{F}_{t_j}$ in the Stratonovich case. This property offers mathematical advantage for the Itô integrals and makes more sense to use in many real applications due to this feature of "not looking into the future", see [51]. Nonetheless, the Stratonovich integrals preserve the ordinary chain rule (i.e. change of variable), whereas Itô integrals do not. The extension of this chain rule in Itô calculus, so-called Itô's lemma, allows conversion between them.

Theorem 2.36 (Itô's lemma)**.** *[52] Suppose that $f \in C^2(\mathbb{R}^m)$ for some positive integer c and $0 \le c \le m$. If $X_t^{(1)}, \ldots, X_t^{(m)}$ are continuous semimartingales (i.e. they can be decomposed as the sum of a local martingale and an adapted finite-variation process) and $X_t^{(c+1)}, \ldots, X_t^{(m)}$ are locally of bounded variation then*

$$\begin{aligned} f(X_t) = f(X_0) &+ \sum_{i=1}^{m} \int_0^t \frac{\partial f}{\partial X^{(i)}}(X_s)\,dX_s^{(i)} \\ &+ \frac{1}{2} \sum_{1\le i,j\le c} \int_0^t \frac{\partial^2 f}{\partial X^{(i)}\partial X^{(j)}}(X_s)\,d[X^{(i)}, X^{(j)}]_s. \end{aligned}$$

Here, $[X^{(i)}, X^{(j)}]_t$ stands for the quadratic covariation of processes $X_t^{(i)}$ and $X_t^{(j)}$.

Similar to the ordinary chain rule, there are no second order terms as in the Itô's lemma when using the Stratonovich integrals, whence they are usually preferred for

SDEs on manifolds. Let us now state conditions under which the solution of an SDE on $\mathbb{R}^m$ exists and is unique.

Theorem 2.37 (Existence and uniqueness theorem for SDEs)**.** *[51] Consider an l-dimensional Brownian motion B_t and some $T > 0$. Let $a : [0,T] \times \mathbb{R}^m \to \mathbb{R}^m$ and $b : [0,T] \times \mathbb{R}^m \to \mathbb{R}^{m\times l}$ be measurable functions with $l \leq m$ for which there exist positive constants C and D such that for all $0 \leq t \leq T$ and $x, y \in \mathbb{R}^m$:*

(i) $||a(t,x)||_2 + ||b(t,x)||_F \leq C(1 + ||x||_2)$ *and*

(ii) $||a(t,x) - a(t,y)||_2 + ||b(t,x - b(t,y)||_F \leq D||x - y||_2$.

Let U be a random variable that is is independent of the σ-algebra generated by $\{B_s,\, s \geq 0\}$ and has finite second moment. Then the stochastic integral equation

$$X_t = U + \int_0^t a(s, X_s)\, ds + \int_0^t b(s, X_s)\, dB_s \quad \textit{for } t \in [0,T] \tag{2.7}$$

has a unique t-continuous solution X_t such that it satisfies two properties

- *X_t is adapted to the filtration generated by U and $\{B_s,\, 0 \leq s \leq t\}$.*
- $\mathbb{E}\left[\int_0^T ||X_t||_2^2\, dt\right]$ *is finite.*

Since explicit solutions of SDEs are not always available, solving a SDE often relies on numerical methods. We end this subsection by introducing a popular numerical approximation method in Itô calculus, the *Euler-Maruyama method* (also called the Euler method). Suppose that we want to approximate the solution of the SDE given in Equation (2.7) with an intial condition $X_0 = u$ using a partition $\pi_n = \{0 = t_0, \ldots, t_n = T\}$. Then the Euler-Maruyama approximation of the solution X_t on $[0,T]$ is the Markov chain $\tilde{X}$ defined by

(i) setting $\tilde{X}_0 = u$,

(ii) defining recursively $\tilde{X}_i$ for $1 \leq i \leq n$ by

$$\tilde{X}_i = \tilde{X}_{i-1} + a\big(t_{i-1}, \tilde{X}_{i-1}\big)\,(t_i - t_{i-1}) + b\big(t_{i-1}, \tilde{X}_{i-1}\big)\,(B_{t_i} - B_{t_{i-1}}).$$

Due to properties of Brownian motion, the random variables $(B_{t_i} - B_{t_{i-1}})$ are independent and identically distributed normal random variables with mean zero, variance $(t_i - t_{i-1})$ for all $1 \leq i \leq n$.

2.3.2 Riemannian manifold

Since the curvature of Riemannian manifold makes direct use of Euclidean stochastic analysis prohibitively hard, a common strategy in stochastic analysis is to adopt the extrinsic view that involves embedding the manifold in a higher dimensional Euclidean space, which is guaranteed by the Nash embedding theorem; see for example [25, 26, 28]. The approach benefits from existing theories on the Euclidean space, but it can be hard to find a suitable embedding and may be even harder to adapt existing SDE algorithms to force the paths to stay on the embedded manifold.

We firstly approach SDEs on manifolds in a similar sense as E. P. Hsu [25], covering SDEs driven by vector fields, diffusion process generated by some differential operators, etc. Afterwards, we discuss the horizontal lift of a smooth curve on manifold to the frame bundle in greater detail and provide intuition in the case of lifting a semimartingale horizontally. Then, we get back to the discussion of the Riemannian Brownian motion, which includes simulation, explosion time, transition probability density, etc. Toward the end of the subsection, we present the extension from K. D. Elworthy [26] to Itô's lemma and Girsanov theorem on the Euclidean space. But, in order to apply these results, we need to review K. D. Elworthy's form of the definition of SDEs on manifolds in [26] and compare this form to the one defined by E. P. Hsu [25]. Although they are equivalent in theory, it is not straightforward to convert explicitly between these two forms. While demonstrating this difficulty on parallelizable manifolds (i.e. existence of a global basis field on the tangent bundle is guaranteed), we define a special class of SDEs on these spaces. This sub-class of SDEs not only provides efficient simulation but also makes many theories in [26] applicable for practical uses. Some examples of these processes are the Riemannian Brownian motion and the Riemannian Ornstein-Uhlenbeck process defined in Chapter 4.

We let $\mathcal{M}$ be a Riemannian manifold of dimension m equipped with a Rie-

mannian metric g. An $\mathcal{M}$-*valued semimartingale* X_t is defined by a continuous $\mathcal{M}$-valued process defined on $[0, \tau)$ for some $\mathfrak{F}_*$-stopping time τ such that the process $f(X_t)$ is a real-valued semimartingale on $[0, \tau)$ for any $f \in C^\infty(\mathcal{M})$.

First of all, we define a SDE on $\mathcal{M}$ driven by $l+1$ smooth vector fields $\{V_i \in \Gamma(T\mathcal{M})\}_{i=0}^{l}$ with a semimartingale Z_t on $\mathbb{R}^{l+1}$ and an $\mathcal{M}$-valued random variable $P \in \mathfrak{F}_0$:

$$dX_t = \sum_{i=0}^{l} V_i(X_t) \circ dZ_t^{(i)} \qquad (X_0 = P) \tag{2.8}$$

An $\mathcal{M}$-valued semimartingale X_t is the solution of Equation (2.8) defined up to a $\mathfrak{F}_*$-stopping time τ if for all $f \in C^\infty(\mathcal{M})$,

$$f(X_t) = f(U) + \sum_{i=1}^{l} \int_0^t V_i f(X_s) \circ dZ_s^{(i)} \qquad (0 \le t < \tau), \tag{2.9}$$

where the integrals above are in the Stratonovich sense, and converting them to Itô sense yields for $0 \le t \le \tau$:

$$\begin{aligned} f(X_t) = f(U) &+ \sum_{i=1}^{l} \int_0^t V_i f(X_s)\, dZ_s^{(i)} \\ &+ \frac{1}{2} \sum_{i,j=1}^{l} \int_0^t (\nabla_{V_j} V_i) f(X_s)\, d[Z^{(i)}, Z^{(j)}]_s. \end{aligned} \tag{2.10}$$

Here, $[Z^{(i)}, Z^{(j)}]_t$ stands for the usual quadratic covariation of $Z_t^{(i)}$ and $Z_t^{(j)}$ defined on the Euclidean space. Comparing to the Itô's lemma (Theorem 2.36), the second order terms (bottom row) of Equation (2.10) are now replaced by the covariant derivatives. Moreover, existence and uniqueness of a SDE's solution up to its explosion time can be shown by reducing this SDE to a corresponding SDE on a Euclidean space of larger dimension and applying the corresponding theorem that was established for Euclidean space [25].

Alternatively, one can also define a $\mathscr{L}$-diffusion process on $\mathcal{M}$ for a smooth second order elliptic differential operator $\mathscr{L}$ on $\mathcal{M}$. An $\mathfrak{F}_*$-adapted stochastic

process X_t on $\mathcal{M}$ is called a diffusion process generated by $\mathscr{L}$ (or *$\mathscr{L}$-diffusion*) if X_t is an $\mathcal{M}$-valued semimartingale up to its explosion time τ and $f(X_t) - f(X_0) - \int_0^t \mathscr{L} f(X_s)\, ds$ is a local $\mathfrak{F}_*$-martingale for all $f \in C^\infty(\mathcal{M})$ and $0 \leq t < \tau$. Moreover, the solution of Equation (2.8) is an $\mathscr{L}$-diffusion generated by a second order parabolic operator

$$\mathscr{L} = \frac{1}{2} \sum_{i=1}^{l} V_i^2 + V_0 \tag{2.11}$$

given that $Z_t^{(i)}$ are $\mathbb{R}$-valued Brownian motion for $1 \leq i \leq l$ and $Z_t^{(0)} = t$, see [25].

On the other hand, inspired by the procedure of "rolling without slipping", K. D. Elworthy [26] has constructed theories about stochastic development of an $\mathcal{M}$-valued semimartingale using the concept of horizontal lift. The fundamental idea here is that one can lift a smooth curve on $\mathcal{M}$ uniquely to a horizontal curve on the frame bundle of $\mathcal{M}$ and there is a unique corresponding smooth curve on $\mathbb{R}^m$ to this horizontal curve. This idea can be generalized to semimartingales on $\mathcal{M}$.

While we discuss horizontal lifts for smooth curves in greater detail later in this subsection, we will only give intuition of the generalization to semimartingales. The stochastic development involves having a semimartingale on $\mathbb{R}^m$, which is the solution of some SDE in terms of Stratonovich integral. The next step is to develop this semimartingale into a horizontal semimartingale on the frame bundle of $\mathcal{M}$. Lastly, we project this horizontal semimartingale down to the manifold $\mathcal{M}$. This procedure induces one-to-one relationships between a semimartingale on $\mathbb{R}^m$, a horizontal semimartingale on the frame bundle of $\mathcal{M}$ and an $\mathcal{M}$-valued semimartingale on $\mathcal{M}$. This approach of stochastic development builds on SDEs, that are written in terms of Stratonovich integrals, however, it is possible to approach stochastic development by smooth approximation [26] with examples: $\mathcal{SO}(3)$ in [31, 30], so-called injection scheme; $\mathcal{SP}(n)$ in Section 4.3. For a moment, let us define some terminologies and notations when dealing with the horizontal lift. We fix a basis $\{e_1, \ldots, e_m\}$ for $\mathbb{R}^m$.

A *frame* at $p \in \mathcal{M}$ is a linear isomorphism $u : \mathbb{R}^m \to T_p\mathcal{M}$, that is $\{u(e_i)\}_{i=1}^m$ is a basis for $T_p\mathcal{M}$. Denoted by $\mathscr{F}(\mathcal{M})_p$ the space containing all frames at p, and

the *frame bundle* of $\mathcal{M}$, $\mathscr{F}(\mathcal{M}) = \bigcup_{p\in\mathcal{M}} \mathscr{F}(\mathcal{M})_p$. Moreover, we can express u in local coordinates $\{E_1, \ldots, E_m, e_1, \ldots, e_m\}$ in some neighbourhood U covering p as $u = (q, \zeta)$ with $\zeta = (\zeta_j^i) \in \mathbb{R}^{m\times m}$ the coefficients with respect to the orthonormal basis $\{E_i(q)\}_{i=1}^m$ on $T_q\mathcal{M}$ for all $q \in U$. That is for any vector $e \in \mathbb{R}^m$, i.e. $e = \sum_{i=1}^m \epsilon_i\, e_i$ for some real numbers ϵ_i, we get $u(e) = \sum_{i,j=1}^m \epsilon_i\, \zeta_j^i\, E_i(p) \in T_p\mathcal{M}$.

Furthermore, $\mathscr{F}(\mathcal{M})$ is a differentiable manifold of dimension $m + m^2$ and there exists a smooth canonical projection map $\pi : \mathscr{F}(\mathcal{M}) \to \mathcal{M}$, see [25]. A smooth curve u_t on $\mathscr{F}(M)$ is defined such that at each time point t, u_t assigns smoothly both a point $\pi(u_t)$ on the manifold as well as a frame for the tangent space $T_{\pi(u_t)}\mathcal{M}$. If for each $e \in \mathbb{R}^d$, $u_t(e)$ is a vector field on $\mathcal{M}$ such that it is parallel along the curve $\pi(u_t)$, then we call u_t a *horizontal curve* on $\mathscr{F}(\mathcal{M})$. Suppose $X \in T_u\mathscr{F}(\mathcal{M})$ for some frame $u \in \mathscr{F}(\mathcal{M})$, then X is called a *horizontal tangent vector* if it is a tangent vector of a horizontal curve u_t with initial frame $u_0 = u$. The subspace of $T_u\mathscr{F}(\mathcal{M})$ contains all horizontal tangent vectors is denoted $\mathcal{H}_u\mathscr{F}(\mathcal{M})$ and has same dimension as the manifold $\mathcal{M}$.

If x_t is a smooth curve on $\mathcal{M}$ with $x_0 = p \in \mathcal{M}$ and u_0 is a frame at p, there exists a unique horizontal curve u_t on $\mathscr{F}(\mathcal{M})$ such that $\pi(u_t) = x_t$, we call u_t the *horizontal lift of the curve* x_t from u_0. Moreover, for $0 < t_0 \le t_1$ the linear map $u_{t_1}u_{t_0}^{-1} : T_{x_{t_0}}\mathcal{M} \to T_{x_{t_1}}\mathcal{M}$ is independent of u_0, while it provides the parallel transport along x_t. Existence and uniqueness of u_t can be shown by expressing the condition for being parallel in local coordinates, and solving a system of ordinary differential equations, see Equation (2.1). Moreover, the projection map π induces an isomorphism for any $u \in \mathscr{F}(\mathcal{M})$, $\pi_* : \mathcal{H}_u\mathscr{F}(\mathcal{M}) \to T_{\pi(u)}\mathcal{M}$. That is, for $p \in \mathcal{M}$, $u \in \mathscr{F}(\mathcal{M})_p$ and $v \in T_p\mathcal{M}$, there exists uniquely $X \in \mathcal{H}_u\mathscr{F}(\mathcal{M})$, we call X the *horizontal lift of the tangent vector* v to $\mathcal{H}_u\mathscr{F}(\mathcal{M})$. Similarly, we can define the *horizontal vector field* H_e on $\mathscr{F}(\mathcal{M})$ for some $e \in \mathbb{R}^m$ such that $H_e(u)$ is the horizontal lift of tangent vector $u(e)$ for any $u \in \mathscr{F}(\mathcal{M})$.

Consider a horizontal lift u_t of a smooth curve x_t on $\mathcal{M}$. Let x_t' denote the tangent vector of x_t for all $t \ge 0$, then $u_t^{-1}(x_t') \in \mathbb{R}^m$. The *anti-development* of the curve x_t on $\mathcal{M}$ is defined by $w_t = \int_0^t u_s^{-1}(x_s')\, ds$, which depends on the initial

frame u_0. In fact, the relationship between w_t and u_t is expressed by the following ordinary differential equation

$$\frac{du}{dt} = \sum_{i=1}^{m} H_{e_i}(u_t) \frac{dw^{(i)}}{dt}. \tag{2.12}$$

The anti-development w_t is guaranteed to exist and is uniquely defined up to initial conditions u_0 and x_0, though its computation is often difficult.

Overall, using stochastic development results in many useful theories, i.e. stochastic line integrals, local martingales on manifolds and sub-manifolds, etc [25]. One of the most important result is that if the anti-development W_t of X_t is a standard Brownian motion on $\mathbb{R}^m$ then X_t is a Riemannian Brownian motion (defined via the Laplace Beltrami operator) on the manifold $\mathcal{M}$ equipped with the affine-connection ∇. In greater detail, a Riemannian Brownian motion, just like a Euclidean Brownian motion, can be defined by the diffusion process generated by $\Delta_{\mathcal{M}}/2$, the Laplace-Beltrami operator. But, on general Riemannian manifolds there is no closed form of expressing Brownian motion via vector fields similar to those in Equation (2.8) because $\Delta_{\mathcal{M}}$ cannot be expressed as similar to $\mathscr{L}$ in Equation (2.11). In theory, there is a lifting of $\Delta_{\mathcal{M}}$ to the orthonormal frame bundle of $\mathcal{M}$, in which it then can be expressed similarly to Equation (2.11). Nevertheless, the practical use of this result is limited as computation of the horizontal lift is usually complex.

On the other hand, since a manifold is made of many flat pieces under transformation of charts, we can write down $\Delta_{\mathcal{M}}$ locally, see Equation (2.6). Consider a chart $x = \{x^{(i)}\}_{i=1}^m$ on an open subset U covering a point $p \in \mathcal{M}$, and the induced basis $\{X_i = \partial/\partial x^{(i)}\}_{i=1}^m$. We write the metric tensor g in the matrix form G with respect to the basis. Consequently, the equation of the Riemannian Brownian motion W_t on $U \subset \mathcal{M}$ in local coordinates is given by

$$dW_t^{(i)} = \sum_{j=1}^{m} \sigma_{ij}(W_t)\, dB_t^{(j)} - \frac{1}{2} \sum_{k,l=1}^{m} G^{lk}(W_t)\Gamma_{kl}^{i}(W_t)\, dt, \tag{2.13}$$

with $1 \leq i \leq m$, $\sigma = \sqrt{G^{-1}}$ and B_t is an $\mathbb{R}^m$-Brownian motion, see also in [25,

Equation (3.3.11)]. Numerical approximation of W_t can be achieved by any existing methods on the Euclidean space, such as the Euler-Maruyama method. However, the SDE (2.13) is well-defined only on U and the matrix function G changes when transitioning to a different chart. Therefore, there is a finite probability of leaving the chart. If one only pays attention in tuning the time width to ensure the next position of W_t stays inside U, this potentially creates biased sample from the true distribution of the SDE of interest, e.g. when simply re-sampling a Brownian increment when the currently proposed increment leads to the next point of the SDE solution falling outside of U. This problem has been addressed in the literature for the Euclidean setting and some suggestions are proposed to overcome it. For example, G. N. Milstein et al. [53] propose a new concept for numerical methods, in which they suggest removing the whole trajectories if any point on these trajectories escapes the restricted region, and clearly this increases the computational cost significantly.

Instead, there is an extension of the Euler-Maruyama method which employs the exponential map [29, 28, 27, 31], we call this method the *exponential adapted Euler-Maruyama method*. At the current time t, we can simulate the next position of the Riemannian Brownian motion W_t as following

$$W_{t+\delta t} = \mathrm{Exp}_{W_t}\left\{\sum_{i=1}^{m}(B^{(i)}_{t+\delta t} - B^{(i)}_t)\, E_i(W_t)\right\} \tag{2.14}$$

where $\{E_1, \ldots, E_m\}$ is an orthonormal basis field with respect to the metric g, defined on a open neighbourhood of W_t and B_t is an $\mathbb{R}^m$-Brownian motion. The exponential adapted Euler-Maruyama method guarantees $W_{t+\delta_t}$ lies on $\mathcal{M}$ if $\mathcal{M}$ is geodesically complete, see Definition 2.18. Moreover, as $\delta t \to 0$, X_t converges to the Riemannian Brownian motion in distribution if there exists a global basis field on the tangent bundle of $\mathcal{M}$ [31, 28, 54], i.e. $\mathcal{M}$ is a *parallelizable manifold* (e.g. Lie groups, torus, $\mathcal{S}^0, \mathcal{S}^1, \mathcal{S}^3$, etc; notice that not all $\mathcal{S}^n$ is parallelizable, e.g. $\mathcal{S}^2$ is not). Additionally, the exponential adapted Euler-Maruyama method only requires solving a system of ordinary differential equations to compute the exponential map, whereas even if we insist to use implicit methods which depend on some clever

choices of integrators during transitioning charts or embedding, this approach often involves solving SDEs which can be much more complicated in the presence of the curvature. Particularly, the space of covariance matrices $\mathcal{SP}(n)$, which is our main focus in Chapter 4, not only is parallelizable, geodesically complete when equipping with either the Log-Eucidean metric or the Affine-Invariant metric, but also offers explicit formulas for the exponential maps. Hence, when working on $\mathcal{SP}(n)$, the exponential adapted Euler-Maruyama method offers greater benefit for practicability.

When performing statistical inference for SDEs, we often require calculation of the transition density function, in particular those of Brownian motions if the process is driven by Brownian motion. We know that the transition density function $f(s, x; t, y)$ of Brownian motion on $\mathbb{R}^m$ is simply a Gaussian distribution $\mathscr{N}\big(x, (t-s)\, I_m\big)$, where I_m is an $m \times m$ identity matrix. However, on a general Riemannian manifold, this density function f exists but usually no explicit expression is available, it is often referred to the *minimal heat kernel* because it solves the heat equation $\mathfrak{L}_{\mathcal{M}} f = 0$, where $\mathfrak{L}_{\mathcal{M}} = \frac{\partial}{\partial t} - \frac{1}{2}\Delta_{\mathcal{M}}$, see more details in [25, 26]. Nonetheless, we can bound the heat kernel below and above for a compact Riemannian manifold.

Theorem 2.38 (Bounds for minimal heat kernel)**.** *[25] Suppose that $\mathcal{M}$ is compact, we denote the Riemannian distance $d(x, y)$ between $x, y \in \mathcal{M}$. There exists constants $C_1, C_2 > 0$ such that for all $t \in (0, 1)$ and $x, y \in \mathcal{M}$, the heat kernel $f_{\mathcal{M}}(0, x; t, y)$ satisfies*

$$\frac{C_1}{\sqrt{t^m}} \exp\left(-\frac{d^2(x,y)}{2t}\right) \leq f_{\mathcal{M}}(0, x; t, y) \leq \frac{C_2}{\sqrt{t^{2m-1}}} \exp\left(-\frac{d^2(x,y)}{2t}\right) \tag{2.15}$$

On the Euclidean space, Brownian motion does not explode in finite time, and if this holds in the Riemannian setting, that is $\int_{\mathcal{M}} f_{\mathcal{M}}(0, p; t, q)\, dq = 1$ for all $p \in \mathcal{M}$ and $0 < t < \infty$, then the manifold $\mathcal{M}$ is said to be *stochastically complete*. A sufficient geometric condition for a complete Riemannian manifold to be stochastically complete depends on the growth of the Ricci curvature.

Theorem 2.39 (Conditions for stochastic completeness)**.** *[25] Consider a fixed point*

$p \in \mathcal{M}$ and denote $d(x,p)$ as the distance between $x \in \mathcal{M}$ and p. Suppose that a negative, non-decreasing, continuous function $\kappa : [0,\infty) \to \mathbb{R}_{<0}$ satisfies

$$\kappa(r) \leq \frac{1}{d-1} \inf_{x \in \mathcal{M}} \{Ric_{\mathcal{M}}(x) : d(x,p) = r\},$$

where $Ric_{\mathcal{M}}(x) = \{Ric(v,v) : v \in T_x\mathcal{M}$ and $||v||_2 = 1\}$. If $\int_c^\infty \big(-\kappa(r)\big)^{-1/2} dr$ is infinite for some constant c then $\mathcal{M}$ is stochastically complete.

Corollary 2.40. *[26] Every geodesically complete Riemannian manifold is stochastically complete if its Ricci curvature is bounded below.*

From this moment, let $\mathcal{M}$ be parallelizable, i.e. there exists a global basis field $\{E_i\}_{i=1}^m$ on the tangent bundle $T\mathcal{M}$. We additionaly introduce $\mathcal{N}$ to denote a general (not necessarily parallelizable) Riemannian manifold. Without loss of generality, we can assume $\{E_i\}_{i=1}^m$ is orthonormal with respect to the metric tensor g on $\mathcal{M}$. Thus, it is possible to define a smooth function $F : \mathcal{M} \times \mathbb{R}^m \to \Gamma(T\mathcal{M})$, in which it assigns smoothly for each $p \in \mathcal{M}$ a frame $F_p : \mathbb{R}^m \to T_p\mathcal{M}$ by

$$F_p(e) = \sum_{i=1}^m \epsilon_i \, E_i(p) \quad \text{with} \quad e = \sum_{i=1}^m \epsilon_i \, e_i \in \mathbb{R}^m. \tag{2.16}$$

Furthermore, suppose that $a : [0,T] \times \mathcal{M} \to \mathbb{R}^m$ and $b : \mathcal{M} \to \mathbb{R}^{m \times m}$ are measurable functions, B_t are $\mathbb{R}^m$-Browninan motion and $P \in \mathfrak{F}_0$ is an $\mathcal{M}$-valued random variable. K. D. Elworthy [26] write SDEs on such manifold $\mathcal{M}$ as

$$dX_t = F_{X_t}\big(a(t,X_t)\,dt\big) + F_{X_t}\big(b(X_t)\,dB_t\big) \qquad (X_0 = P), \tag{2.17}$$

which can be converted to those SDEs in Equation (2.8) (or vice versa) by setting $V_i(X_t) = \sum_{j=1}^m b_{ij}(X_t)\,E_j(X_t)$ for $1 \leq i \leq l = m$ and $V_0(t,X_t) = F_{X_t}\big(a(t,X_t)\big)$, with $Z_t^{(0)} = t$ and $\big(Z_t^{(i)}\big)_{i=1}^m = B_t$. SDEs given in the form of Equation (2.17) presents similar interpretation as the Euclidean SDEs, which are separated into random and deterministic terms.

For a more general manifold $\mathcal{N}$, K. D. Elworthy [26] only requires such function F to be differentiable in the first argument, having locally Lipschitz first derivatives

(in coordinate charts), while it assigns for each point $q \in \mathcal{N}$ an associated frame F_q on $T_q\mathcal{N}$. We distinguish this class of functions when considering $\mathcal{N}$ by denoting $\tilde{F}$. The function F defined in Equation (2.16) is an example of such an $\tilde{F}$. Suppose that $A : [0,T] \times \mathcal{N} \to \Gamma(T\mathcal{N})$ such that A assigns smoothly for each $t \in [0,T]$ a smooth vector field $A(t,\cdot)$ on $\mathcal{N}$. And SDEs on $\mathcal{N}$ are then established in [26] as

$$dX_t = A(t, X_t)\, dt + \tilde{F}_{X_t}\big(b(X_t)\, dB_t\big) \qquad (X_0 = P). \tag{2.18}$$

The assumption of a parallelizable manifold is not only to ensure the definition of F in Equation (2.16) to be well-defined, but also to show examples of possible conversion between SDEs in the forms (2.8) and (2.18). We emphasize two remarks: firstly, Equation (2.17) is another way of writing Equation (2.18) for a parallelizable manifold, given that the formula of $\tilde{F}$ is available explicitly, i.e. $a(t, x_t)$ can be extracted explicitly from $A(t, X_t)$ (or vice versa) and any functions $\tilde{F}$ can be rewritten in term of F with a new function $\tilde{b}$ that plays a similar role to b defined in Equation (2.17). Secondly, although two forms of SDEs in Equation (2.8) and (2.18) are equivalent in theory, it is not obvious how one can move between these forms when dealing with an arbitrary Riemannian manifold, e.g. it is not always possible to extract V_i explicitly from $\tilde{F}$ and vice versa. For instance, if U_t is the horizontal lift of the Riemannian Brownian motion W_t on $\mathcal{N}$, the SDE in the form (2.18) of W_t is simply $dW_t = \tilde{F}_{W_t}(dB_t)$, where $\tilde{F}_{W_t} = U_t$, and we know that SDE of W_t can't be written in the form (2.8) generally. Besides a clear separation between random and deterministic parts, we favour the expression of SDEs in the form (2.17) or (2.18) over the one in Equation (2.8) because many important results, that are established globally can be applied directly, such as the extension of Itô's lemma and the Girsanov theorem to Riemannian manifolds, see Theorem 2.41 and 2.42.

Additionally, let us define a sub-class of SDEs on $\mathcal{M}$ as those processes obtained

as the limit as δ_t tends to zero of the exponential adapted Euler-Maruyama method:

$$X_{t+\delta t} = \mathrm{Exp}_{X_t}\left\{\sum_{j=1}^{m}\left(a^{(j)}(t, X_t)\,\delta t + \sum_{i=1}^{m} b_{ij}(X_t)\,\left(B^{(i)}_{t+\delta t} - B^{(i)}_t\right)\right) E_j(X_t)\right\}, \tag{2.19}$$

where functions $a,\ b$ are those appearing in Equation (2.17). We emphasize that not all functions $a,\ b$ guarantee the convergence, but this special class is non-empty as there are existing works in the literature, which says certain diffusion processes can be achieved by taking the limit $\delta_t \downarrow 0$, see for examples [28, 54, 26]. Besides the example of the Riemannian Brownian motion, the Riemannian Ornstein-Uhlenbeck process defined in Chapter 4 also belongs to this class.

However, the limiting result X_t in Equation (2.19) is not the solution of the SDE (2.17) with a counterexample that having $a = 0,\ b = I_m$. In fact, X_t is the solution of the SDE defined with $\tilde{F}_{X_t} = U_t$, where U_t is the horizontal lift of X_t and in general has no explicit formula. This approach studies stochastic development via smooth approximation. Thus, if we start our SDEs in either forms (2.8) or (2.17) on $\mathcal{M}$, it limits practical uses as one might need to approach embedding or local coordinates for simulation. Instead, we write this special class of stochastic processes on $\mathcal{M}$ as

$$dX_t = A(t, X_t)\,dt + F_{X_t}\big(b(X_t)\,dB_t\big) \qquad (X_0 = P). \tag{2.20}$$

That is to say, when we define X_t as the solution of the SDE (2.20) with explicit formulas for $A(t, X_t)$ and $b(X_t)$, we mean X_t as the limiting results in Equation (2.19) with $a(t, X_t) = F^{-1}_{X_t}\big(A(t, X_t)\big)$. Converting the SDE (2.20) to the form (2.18) is done by setting $\tilde{F}_{X_t} = U_t$, the horizontal lift of X_t. We want to be clear that although the Equation (2.20) and (2.17) look similar, their solutions are different in general.

In favour for practicability, we choose the form given in Equation (2.20) for the construction of the Riemannian Ornstein-Uhlenbeck process in Chapter 4. Moreover, under some restrictions for functions A and b, the map F defined in Equation (2.16) approximates U_t, see the demonstration for $\mathcal{SP}(n)$ in the Subsection 4.3.

When dealing with SDEs in the form (2.20), this approximation procedure allows us to apply existing theories that are directly applicable to SDEs (2.18). While there are many important theories in [26], we end the chapter by presenting only those that will be used later in Chapter 4.

Theorem 2.41 (Global formulations of Itô's lemma)**.** *[26] Consider a stochastic process X_t on $\mathcal{N}$, which is the solution of the SDE* (2.18) *and denote τ the explosion time of X_t. For a function $f \in C^2\big([0,\tau)\times\mathcal{N}\big)$, and any $t<\tau$, we get*

$$f(t,X_t) = f(0,X_0) + \int_0^t \left(\frac{\partial f}{\partial s}\,ds + g_{X_s}\Big(A(s,X_s)\,ds + h(X_s), \nabla f\Big)\right) + \frac{1}{2}\int_0^t \left(\Delta_{\mathcal{N}} f\Big(h(X_s)\,,\,h(X_s)\Big) + g_{X_s}\Big(\nabla_{h(X_s)}h(X_s)\,,\,\nabla f\Big)\right)$$

where $h(X_s) = \tilde{F}_{X_s}(b(X_s)\,dB_s)$, see definition of the Riemannian gradient ∇f in Equation (2.3) *and the Lapalace-Beltrami operator $\Delta_{\mathcal{N}} f$ in Equation* (2.5)*.*

Comparing to the Itô lemma on the Euclidean space (Theorem 2.36), the last additional term in the second line in fact accounts for the curvature of $\mathcal{N}$. Because if $\mathcal{N}$ has zero curvature everywhere, this last term vanishes, and we get a formula identical to the one in Theorem 2.36.

Theorem 2.42 (Girsanov-Cameron-Martin theorem)**.** *[26] Given the stochastic process X_t, Y_t and their explosion time τ_X, τ_Y, their law μ_X, μ_Y respectively, let $\tau = \min\{\tau_X, \tau_Y\}$.*

(i) Suppose X_t, $Y_t \in \mathcal{M}$ are the solutions of the SDEs in the form (2.17)*:*

$$\begin{aligned} dX_t &= F_{X_t}\big(a(t,X_t)\,dt\big) + F_{X_t}\big(b(X_t)\,dB_t\big) && (X_0 = P),\\ dY_t &= F_{Y_t}\big(a(t,Y_t)\,dt\big) + F_{Y_t}\big(b(Y_t)\,dB_t\big) + F_{Y_t}\big(c(t,Y_t)\,dt\big) && (Y_0 = P), \end{aligned}$$

where $c : [0,\tau)\times\mathcal{M} \to \mathbb{R}^m$ is a measurable function. Moreover, assume that function c is locally Lipschitz, bounded above and b is invertible. Then μ_x and

μ_y are absolutely continuous on $[0,\tau)$, and the Radon–Nikodym derivative is

$$\frac{d\mu_x}{d\mu_y}(Y_t) = \exp\left\{ -\int_0^t \left\langle b^{-1}(Y_s)\, c(s,Y_s), dB_s \right\rangle_F - \frac{1}{2}\int_0^t \left|\left| b^{-1}(Y_s)\, c(s,Y_s) \right|\right|_F^2 ds \right\}.$$

(ii) Suppose $X_t,\ Y_t \in \mathcal{N}$ are the solutions of the SDEs in the form (2.18)*:*

$$dX_t = A(t,X_t)\,dt + \tilde{F}_{X_t}\big(\mathfrak{b}\, dB_t\big) \qquad (X_0 = P),$$
$$dY_t = A(t,X_t)\,dt + \tilde{F}_{X_t}\big(\mathfrak{b}\, dB_t\big) + C(t,X_t)\,dt \qquad (Y_0 = P),$$

where $\mathfrak{b}$ is a non-zero constant and $C : [0,T]\times\mathcal{N} \to \Gamma(T\mathcal{N})$ assigns smoothly for each $t \in [0,\tau)$ a smooth vector field $C(t,\cdot)$ on $\mathcal{N}$. Moreover, assume that C is locally Lipschitz, bounded above. Then μ_x and μ_y are absolutely continuous on $[0,\tau)$, and the Radon–Nikodym derivative is given by

$$\frac{d\mu_x}{d\mu_y}(Y_t) = \exp\left\{ -\int_0^t g_{Y_s}\Big(\mathfrak{b}^{-1}\, C(s,Y_s), U_s(dB_s)\Big) - \frac{1}{2}\int_0^t \left|\left| \mathfrak{b}^{-1}\, C(s,Y_s) \right|\right|_{g_{Y_s}}^2 ds \right\},$$

where U_t is the horizontal lift of the process Y_t.

Chapter 3

Regression on Lie groups

3.1 Overview

Regression analysis has been studied for centuries in the literature but most work focuses on the Euclidean space, and more attention to manifolds has recently been received due to geometric features, for examples: parametric circular–circular regression [55], dimension reduction [56], dealing with predictor collinearity [14], etc. Since linearity is not well-defined on a general manifold, the linear model will not be suitable if the variables lie on a manifold. In particular, P. T. Fletcher [21] has constructed an analogue of linear regression to manifold, so-called geodesic regression, to model the relationship between the dependent variables on a Riemannian manifold and the independent variables on the Euclidean space. The theory of least squares for geodesic regression has been investigated in [57, 58], in which the approach basically aims to minimize the sum of squared Riemannian distances between the data and the fitted values from the model. However, as we know that this minimizer in general has no explicit form, one requires an iterative algorithm, e.g. gradient descent method. Besides linear models, more general forms of regression, that have either independent and dependent variables lie on a Riemannian manifold, are proposed in [59] with application in brain imaging via a novel kernel based non-linear regression method.

We consider a simpler scenario in this thesis under two assumptions. Firstly, we assume only the independent variables lie on a Riemannian manifold, whereas

the fitted curve lies on the Euclidean space. Secondly, the Riemannian manifold that the independent variables belong to is assumed to be a compact connected Lie group, so that the existence of a bi-invariant metric is guaranteed. Thus, we can set a principal basis on the tangent plane at the identity equipped with this special metric, and it is sufficient to study the whole space apart from its cut locus.

We map the manifold of interest intrinsically onto its tangent space using the logarithm map. In particular, for some Lie groups of matrix form, such as $\mathcal{GL}(n)$, $\mathcal{SO}(n)$, etc, the exponential/logarithm maps coincide to be the usual matrix exponential/logarithm functions. Many properties of matrix exponential/logarithm can be applied, offering lower computational cost. This parametrization converts locally (i.e. within the cut locus) a more complex space to a Euclidean space, therefore the number of parameters in the model required to estimate does not increase, whereas, using embedding might require more parameters. Although the exponential map is only diffeomorphic within the cut locus, the cut locus C_q is a closed subset with measure zero conditioned on an arbitrary point q on the Lie group of interest (see Theorem 2.21). Therefore, it is valid to restrict our data to stay away from the cut locus. To summarize, the proposed methodology is simply fitting a regression model through the parametrization of Lie groups employing the logarithm map, followed by existing Euclidean estimation methods.

While there are many estimation methods in regression analysis on the Euclidean space, we focus on the generalized least squares (GLS) in this thesis because we would like to make it applicable for our application of soft tissue artefacts, that will come later in this chapter. We briefly review GLS on the Euclidean space, followed by the extension to the compact connected Lie groups in Section 3.2. Furthermore, we explore GLS on the space of rotational matrices $\mathcal{SO}(3)$ in Section 3.3. Besides the parametrization employing the logarithm map, we review another prametrization defined by Euler angles, that is commonly seen in human kinematic analysis; see for examples [3, 4, 60]. Moreover, the impact of different base points of the tangent space on the fitted values are also explored. Lastly, for our application in human kinematic analysis, we implement GLS on $\mathcal{SO}(3)$ for the model that describes the

relationship between soft tissue artefacts and the rotations around the hip/knee joints when using two different parametrizations. In addition, we consider two non-trivial correlation structures, namely Matérn and quasi-periodic correlation functions in the attempt of removing obvious patterns among residuals.

3.2 Generalized least squares

3.2.1 Euclidean space

Linear regression is widely used in practise due to its simplicity, e.g. in agriculture, biology, astronomy, social science, etc [61, 62, 63]. Many applications use linear regression to describe relationships between variables [64] , e.g. prediction, forecasting, explaining the variation in the response variables. The method of least squares is a popular approach in regression analysis [57, 58], which aims to minimize the sum of squared residuals. C. Gauss [65] studies the theory of least squares to a great extend and the Gauss-Markov theorem, which is one fundamental result for general linear models, see Theorem 3.1.

Firstly, let us consider a linear regression model of n observations:

$$y = X\beta + \epsilon \qquad \text{with} \qquad \epsilon \sim \mathcal{N}\left(0_n, \sigma^2\,\Sigma\right) \tag{3.1}$$

where the dependent variables y is an $n \times 1$ vector, the independent variable X is an $n \times p$ matrix with full rank, i.e. $\text{rank}(X) = p$, and β is an $p \times 1$ vector containing p unknown model parameters. Here, we split the covariance matrix of the errors ϵ into σ and Σ. We assume that while σ is an unknown positive constant, Σ is a correlation matrix that is either known or if it is unknown, it has known structure within a parametric family.

We define an estimator $\hat{\beta}$ for β in model (3.1) to be a *linear unbiased estimator* of β if there exists $C \in \mathbb{R}^{p\times n}$ such that $\hat{\beta} = Cy$ and $CX = I_p$. Moreover, the *best linear unbiased estimator* β^* of β is a unique linear unbiased estimator such that for any linear unbiased estimator $\hat{\beta}$ of β, the matrix $\text{Cov}(\hat{\beta}) - \text{Cov}(\beta^*)$ is non-negative definite. In other words, the best linear unbiased estimator has smallest variation

and hence the smallest mean square error among linear unbiased estimators.

Theorem 3.1 (Gauss-Markov theorem)**.** *[66] Consider the linear model (3.1), given that Σ is known. The Gauss-Markov estimator (GME) for β is defined by*

$$\hat{\beta}(\Sigma) = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y, \tag{3.2}$$

then GME is the best linear unbiased estimator of β and Cov$\big(\hat{\beta}(\Sigma)\big) = \sigma^2(X^T\Sigma^{-1}X)^{-1}$.

When Σ equals the $n \times n$ identity matrix I_n, the GME defined in Equation (3.2) reduces to the *ordinary least squares estimator* (OLSE):

$$\hat{\beta}(I_n) = (X^TX)^{-1}X^Ty. \tag{3.3}$$

In real applications, the correlation structure and variation of the data are usually unknown. In practice, an assumption of independent errors with constant variance, i.e. $\Sigma = I_n$, is often made as a first attempt. However, when the errors appear obviously correlated, OLSE does not offer smallest variation and the residuals indicate a clear pattern. In fact, if we decompose Σ using the Cholesky factorization, i.e. $\Sigma = AA^T$ for some lower triangular matrix A with real and positive diagonal entries, the OLSE using transformed data $A^{-1}y$ and $A^{-1}X$ is given by

$$\big\{(A^{-1}X)^T(A^{-1}X)\big\}^{-1}(A^{-1}X)^T(A^{-1}y) = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y$$

which is indeed the GME when using the original data while knowing the correlation matrix Σ. Now suppose that Σ is unavailable but its structure is known, i.e. $\Sigma = \Sigma(\theta)$ for some unknown but estimable parameters θ. It is reasonable to use an estimator for the correlation matrix Σ, say $\hat{\Sigma}$. Thus, the GME in (3.2) reduces to the *generalized least squares estimator* (GLSE),

$$\hat{\beta}(\hat{\Sigma}) = (X^T\hat{\Sigma}^{-1}X)^{-1}X^T\hat{\Sigma}^{-1}y \tag{3.4}$$

From now on, we will denote the OLSE of β as $\hat{\beta}_{\text{OLS}}$, and the GLSE as $\hat{\beta}_{\text{GLS}}$. The OLS residual vector $\mathfrak{e}$ is then given by $\mathfrak{e} = y - X\hat{\beta}_{OLS} = \left(I_n - X(X^TX)^{-1}X^T\right) y$. Moreover, for any linear unbiased estimator $\hat{\beta}$ of β, i.e. $\hat{\beta} = C(\mathfrak{e})y$ for some function C that satisfies $C(\mathfrak{e})X = I_p$, K. Takeaki & H.Kurata [66] show that the estimator $\hat{\beta}$ is in fact a GLSE with a known form of correlation matrix $\hat{\Sigma}$ given by observations,

$$\hat{\Sigma} = \left(C(\mathfrak{e})^T\, C(\mathfrak{e}) + \left(I_n - X(X^TX)^{-1}X^T\right)\right)^{-1}.$$

Lastly, let us introduce two different algorithms of generalized least squares when considering the linear regression model (3.1).

Algorithm 3.2 (Minimized GLS)**.** [66] Suppose the correlation function $\Sigma(\theta)$ is a non-vanishing twice differentiable function with $\theta \in \mathbb{R}^d$ and $d < n$, then $(\beta, \theta) = \left(\hat{\beta}_{\text{GLS}}(\Sigma(\hat{\theta})), \hat{\theta}\right)$ minimizes $\mathcal{F}(\beta, \Sigma) = (y - X\beta)^T\Sigma(\theta)^{-1}(y - X\beta)$, where $\hat{\beta}_{\text{GLS}}(\Sigma(\hat{\theta}))$ is given in Equation (3.4) and

$$\hat{\theta} = \underset{\theta\in\mathbb{R}^d}{\arg\min}\{[y - X\hat{\beta}_{\text{GLS}}(\Sigma(\theta))]^T\Sigma(\theta)^{-1}[y - X\hat{\beta}_{\text{GLS}}(\Sigma(\theta))]\}$$

Algorithm 3.3 (Iterated GLS)**.** [67] Under the model assumption that errors $\epsilon \sim \mathcal{N}(0_n, \sigma^2\Sigma)$, the estimator $\hat{\theta}$ of θ aim to maximize the likelihood, i.e. it minimizes the following function:

$$f(\theta, e) = \frac{n}{2}\,\log\hat{\sigma}^2 + \frac{1}{2}\log\det\Sigma(\theta) + \frac{1}{2\hat{\sigma}^2}(e^T\Sigma^{-1}(\theta)e) \qquad \text{with}$$

$$\Sigma(\theta) = AA^T \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^n(\sum_{j=1}^n A_{ij}^{-1}e_j)^2}{n-1}.$$

1. At $k = 1$,

   (i) Starting with $\Sigma_1 = I_n$, computing $\beta_1 = (X^T\Sigma_1^{-1}X)^{-1}X^T\Sigma_1^{-1}Y$ and the residuals vector $e_1 = y - X\beta_1$.

   (ii) Minimizing function $f(\theta, e_1)$ to get $\theta_1 = \underset{\theta}{\text{argmin}} f(\theta, e_1)$.

2. At $k = 2$,

(i) Computing $\Sigma_2 = \Sigma(\theta_1)$, $\beta_2 = (X^T\Sigma_2^{-1}X)^{-1}X^T\Sigma_2^{-1}Y$ and the residuals vector $e_2 = Y - X\beta_2$.

(ii) Minimizing function $f(e_2)$ to get $\theta_2 = \underset{\theta}{\operatorname{argmin}} f(e_2)$.

3. Repeating for $k \geq 3$, and we will stop the algorithm at $k = N$ where $|\theta_N - \theta_{N-1}| < \delta$ for some small δ.

While Algorithm 3.2 provides an estimation method such that $\{\beta, \theta\}$ minimizes the sum of squared Mahalanobis distance between observations y and its fitted values, Algorithm 3.3 estimates β in a similar manner as linear least squares and at the same time θ maximizes the likelihood of residuals given estimator of β. It is easy to see that Algorithm 3.2 differs from the Algorithm 3.3 through the absence of the calculation $\log\det\Sigma(\theta)$, and we expect these two methods to give similar estimates for (β, θ) when $\det\Sigma(\theta)$ is constant. If $\Sigma = I_n$, i.e. $\hat{\beta}_{\text{GLS}} = \hat{\beta}_{\text{OLS}}$, the Algorithm 3.2 reduces to the usual least squares method.

Particularly, although Algorithm 3.3 may not give the same result as a pure maximum likelihood estimator for $\{\beta, \theta\}$, this algorithm reduces the computational cost significantly as the formula of the estimator for β is available explicitly. However, alongside with the result that the iterated GLSE coming from Algorithm 3.3 offers a GLSE for β, Y. Toyooka [67] also states that "there is no improvement by the use of the iterated residuals up to the second order". For the practical use of iterated GLS, it is recommended to use up to step 2 in Algorithm 3.3 to get an estimate of θ, then compute the GLSE for β given this estimated value of θ. Overall, there is no clear evidence as to which methods is better [66].

3.2.2 Compact connected Lie group

In this subsection, we discuss the extension of the GLS in Subsection 3.2.1 to a compact connected Lie group $\mathcal{G}$ (see Definition 2.23) of dimension p, equipped with the binary operation $*$, the identity element e and its Lie algebra $\mathfrak{g}$. We consider a regression model of y against X, where y is a $n \times 1$ vector and observations $X_i\,(1 \leq i \leq n)$ are elements of $\mathcal{G}$.

The essence of the generalization is based on the following key points. We

construct a bi-invariant Riemannian metric using left translation from an induced Euclidean metric. We then discuss three cases regarding the location where observations lie in order to remove the restriction of exponential/logarithm maps being only locally diffeomorphic. Lastly, we establish the parametrization employing the logarithm map in each case and apply the usual GLS to estimate model parameters.

We intend using Corollary 2.29 to get a bi-invariant Riemannian metric on $\mathcal{G}$, that is to choose an induced Euclidean metric g_e on $\mathfrak{g}$, such as the scalar product or Frobenius inner product, etc, which satisfies

$$g_e([u,v],w) + g_e(u,[w,v]) = 0 \qquad \forall u,v,w \in \mathfrak{g}, \tag{3.5}$$

where $[\cdot,\cdot]$ is the Lie bracket operating on the Lie algebra $\mathfrak{g}$. Using Proposition 2.28, we can define a left invariant Riemannian metric g on $\mathcal{G}$ from this metric g_e through the left translation L, that is

$$g_x(u,v) = g_e((dL_{x^{-1}})_x(u), (dL_{x^{-1}})_x(v) \qquad \forall u,v \in T_x\mathcal{G},\ x \in \mathcal{G},$$

where $L_{x^{-1}}$ stands for the left translation by x^{-1}. Corollary 2.29 ensure g is indeed a bi-invariant Riemannian metric on $\mathcal{G}$. Given that $\mathcal{G}$ is a compact connected Lie group of finite dimension, its adjoint representation is therefore bounded and Proposition 2.30 guarantees the existence of a bi-invariant Riemannian metric on $\mathcal{G}$. In other words, we are always able to obtain a metric g_e which satisfies condition in Equation (3.5), and extend it to a bi-invariant Riemannian metric through the left translation. Accordingly, it is valid to fix a bi-invariant Riemannian metric g on $\mathcal{G}$ for further work.

Since the dimension of the Lie algebra $\mathfrak{g}$ is finite, it is not difficult to find an orthonormal basis for $\mathfrak{g}$ with respect to the metric g, for instance: the Gram-Schmidt algorithm can be applied to convert a basis to an orthonormal basis. Let $\mathscr{B} = \{b_1, \ldots, b_p\}$ be an orthonormal basis for $\mathfrak{g}$, that is for any $x \in \mathfrak{g}$, there exist $x^{(k)} \in \mathbb{R}$ such that $x = \sum_{k=1}^{p} x^{(k)} b_k$ and $g_e(b_j, b_k) = \delta_{jk}$ with the Kronecker delta δ_{jk}. Furthermore, let us denote the set $\mathscr{E}_a$ within the cut locus of $a \in \mathcal{G}$, see for more

details in Definition 2.20. In other words, for some $\epsilon > 0$ with the open ball $B_\epsilon(0)$ at the origin of $\mathfrak{g}$, the following map and its inverse are diffeomorphisms:

$$\mathrm{Exp}_a \circ (dL_a)_e : B_\epsilon(0) \to \mathscr{E}_a \quad \& \quad (dL_{a^{-1}})_a \circ \mathrm{Log}_a : \mathscr{E}_a \to B_\epsilon(0)$$

Noting that not all Lie groups have an exponential map (with inverse given by the logarithm map) at the identity coincides with the matrix exponential (and matrix logarithm, respectively). In order to avoid the confusion, in this subsection we only mean exponential/logarithm map. Finally, let us define our proposed parametrization that employs the logarithm map in three cases:

Case 1. $\{X_1, \ldots, X_n\} \subset \mathscr{E}_e$, then $dL_{e^{-1}} \circ \mathrm{Log}_e = \mathrm{Log}_e$. We compute $x_i \in \mathbb{R}^p$ for $i \in \{1, \ldots, n\}$ such that $\mathrm{Log}_e(X_i) = \sum_{k=1}^p x_i^{(k)} b_k$, i.e. $x_i^{(k)} = \langle \mathrm{Log}_e(X_i), b_k \rangle_F$ for $k \in \{1, \ldots, p\}$.

Case 2. There exists a fixed base point $a \in \mathcal{G}$ such that $\{X_1, \ldots, X_n\} \subset \mathscr{E}_a$, and we compute $x_i \in \mathbb{R}^p$ for $i \in \{1, \ldots, n\}$ such that $\big((dL_{a^{-1}})_a \circ \mathrm{Log}_a\big)(X_i) = \sum_{k=1}^p x_i^{(k)} b_j$, i.e. $x_i^{(k)} = \big\langle \big((dL_{a^{-1}})_a \circ \mathrm{Log}_a\big)(X_i)\,,\, b_k \big\rangle_F$ for $k \in \{1, \ldots, p\}$.

Case 3. There is no such $a \in \mathcal{G}$ as in case 2, which means there is no one-to-one parametrization given our data set. By restricting the domain of the data, we can end up in case 2.

For the first two cases, we get $x_i \in \mathfrak{g}$, uniquely corresponding to our data X_i. In case 2, a can be chosen as any fixed element of $\mathcal{G}$, e.g. a can be one of the data points, the important point here is we need to make sure that $\mathscr{E}_a$ covers the domain of our observations. The Fréchet mean of $\{X_1 \ldots, X_n\}$ could be an alternative choice. Furthermore, since the measure of the cut locus of a point on $\mathcal{G}$ is zero (Theorem 2.21), in probability we will not end up in case 3. However, if restricting the domain is not an option, we need to seek for an alternative parametrization.

Consequently, $\{x_1, \ldots, x_n\}$ lie on $\mathbb{R}^p$, thus standard technique on Euclidean space can be applied from here, such as GLS in Subsection 3.2. Clearly, using the proposed parametrization via logarithm map does not incur extra model parameters.

In fact, any regression models and statistical tools in regression analysis on Euclidean space (e.g. p-values, hypothesis testing) can also be applied at this step. For example, if we assume the data fall into case 1 and the regression model includes an intercept, the OLS aims to minimize $\sum_{i=1}^{n} \left\{ y_i - \beta_0 - \sum_{k=1}^{p} \beta_k \left\langle \text{Log}_e(X_i), b_k \right\rangle_F \right\}^2$.

The GLSE is no longer the best linear estimator as the logarithm map in general is non-linear. But, under the assumption of data points staying away from the cut locus of a fixed base point, the GLSE is unbiased as the logarithm map is a diffeorphism within this cut locus.

In summary, the role of the base point a is simply making sure that the observations stay away from the cut locus. And using different base point a results a different model, because in our methodology we are treating y as a collection of coefficients with respect to the basis on the tangent plane at identity, while transporting X to the tangent plane at a then using the left invariant vector field of a^{-1} to move them to the tangent plane at identity. Depending on the nature of observations (i.e. the spread of the data) and which Lie group we are working with, the impact from the base point varies. We discuss this problem in detail for$\mathcal{SO}(3)$, which is a typical example of a compact connected Lie group in the next section.

3.3 Regression on $\mathcal{SO}(3)$

We briefly review the space of rotational matrices $\mathcal{SO}(3)$, and state the Rodrigues' formula that calculates matrix exponential/logarithm on $\mathcal{SO}(3)$. Followed by studying the effects of using different base points to the fitted values when applying OLS/GLS with our proposed parametrization given in Subsection 3.2.2. Lastly, we explore a convention of Euler angles using the joint coordinate system proposed by E. Grood & W. Suntay [4], which is commonly used in human kinematic study.

The concept of rotation has been studied comprehensively in many areas and there are many ways to parameterize rotational matrices, such as Euler angles, axis angle pair, unit quaternion. For instance, in an application of aircraft rotation occurs when the air-planes lift up from the runway during take-off, Euler angles are regarded as yaw, pitch and roll [68]. In an application of human kinematic

study, movement between femur and hipbone is also a rotation, and Euler angles describe flexion-extension, abduction-adduction and internal-extenal rotation [3, 4]. In addition, in the field of astronomy researchers have shown that stars, planets rotate (or spin) around some axes [69], and in this case expression of rotation can be thought as an axis angle pair; or in quantum mechanics, the spin of the particle (i.e. electron) is indeed undertaken by a rotational transformation [70], which is commonly written in the form of a unit quaternion. Besides these parametrizations, the study in Subsection 3.2.2 suggests another form using coefficients of skew symmetric matrices (via logarithm map) with respect to a fixed basis.

We know that the set of matrices

$$\mathcal{SO}(3) = \{A \in \mathcal{GL}(3) \,|\, A^T A = AA^T = I_3 \text{ and } \det A = 1\}$$

is a group with the binary operation defined by the usual matrix multiplication, and $\mathcal{SO}(3)$ is sometimes referred to as the special orthogonal group. Moreover, every rotational transformation (i.e. rotation of a rigid body which preserves length, angles and orientation) can be represented uniquely by an orthogonal matrix with unit determinant, thus $\mathcal{SO}(3)$ is also often called the space of rotational matrices.

Firstly, it is not difficult to see that $\mathcal{SO}(3)$ is a non-Abelian group (i.e. non-commutative group) since exchanging the order of the rotational transformations implies a different rotation. Furthermore, $\mathcal{SO}(3)$ is a compact connected Lie group, and its Lie algebra, denoted by $\mathfrak{so}(3)$, is the space of skew symmetric matrices [71], i.e. $\mathfrak{so}(3) = \{B \in \mathcal{GL}(3) \,|\, B^T = -B\}$. This can be easily shown by considering a differentiable curve $A : (-\epsilon, \epsilon) \to \mathcal{SO}(3)$ such that $A(0) = I_3$, then differentiating the equation $A^T(t)A(t) = I_3$ with respect to t at $t = 0$, we get

$$\frac{dA^T}{dt}A + A^T \frac{dA}{dt} = \left(\frac{dA}{dt}\right)^T A + A^T \frac{dA}{dt} = 0,$$

$$\Rightarrow \left(\left.\frac{dA}{dt}\right|_{t=0}\right)^T A(0) + A^T(0) \left.\frac{dA}{dt}\right|_{t=0} = 0 \quad \Rightarrow \quad \left(\left.\frac{dA}{dt}\right|_{t=0}\right)^T = -\left.\frac{dA}{dt}\right|_{t=0}.$$

Let us equip the Lie algebra $\mathfrak{so}(3)$ with the Frobennius inner product $\langle \cdot, \cdot \rangle_F$.

It is not difficult to check that Equation (3.5) holds for $g_e = \langle \cdot, \cdot \rangle_F$. Thus, we can define a bi-invariant Riemannian metric g from $\langle \cdot, \cdot \rangle_F$ as discussed in previous section. Particularly, any 3×3 skew symmetric matrix is of the following form

$$\begin{pmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{pmatrix} \qquad\qquad (x, y, z \in \mathbb{R}) \ .$$

It is easy to see that $\mathscr{B} = \{b_i\}_{i=1}^3$ is an orthogonal basis with respect to $\langle \cdot, \cdot \rangle_F$, where

$$b_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \qquad b_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \qquad b_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.6}$$

and the basis $\{b_1/\sqrt{2}\,,\, b_2/\sqrt{2}\,,\, b_3/\sqrt{2}\}$ is orthonormal with respect to $\langle \cdot, \cdot \rangle_F$.

In order to apply GLS in Subsection 3.2.2 to $\mathcal{SO}(3)$, we need to identify the differential of the left translation and the logarithm map. Firstly, given $A \in \mathcal{SO}(3)$ the differential of the left translation L_A (i.e. $B \mapsto AB$) is

$$\forall B \in \mathcal{SO}(3),\ (dL_A)_B : T_B\mathcal{SO}(3) \to T_{L_A(B)}\mathcal{SO}(3) \qquad\qquad v \mapsto Av. \tag{3.7}$$

Secondly, as previously mentioned in Section 2.2, the geodesics passing through the origin of $\mathfrak{so}(3)$ are in fact one-parameter subgroups of $\mathcal{SO}(3)$. Since one-parameter subgroups of $\mathcal{SO}(3)$ are given in the form $\{\exp(tv) \,|\, t \in \mathbb{R},\ v \in \mathfrak{so}(3)\}$, where $\exp$ denotes the matrix exponential function, the curve $\gamma(t) = \exp(tv)$ on $\mathcal{SO}(3)$ is a geodesic that passes through the identity element, i.e. $\gamma(0) = I_3$, in some direction $v \in \mathfrak{so}(3)$. Thus, using the left translation for any point $A \in \mathcal{SO}(3)$, the curve $\gamma(t) = A\exp(tv)$ is a geodesic starting at A with initial velocity $\gamma'(0) = Av$. The exponential/logarithm maps on $\mathcal{SO}(3)$ are therefore defined by

$$\mathrm{Exp}_A(v) = A\exp(A^{-1}v) \quad \text{and} \quad \mathrm{Log}_A(B) = A\log(A^{-1}B), \tag{3.8}$$

for $v \in T_A\mathcal{SO}(3)$, and Log_A is well-defined given that $B \in \mathcal{SO}(3)$ is not in the cut locus of A. Moreover, for any $A_1, A_2 \in \mathcal{SO}(3)$, the minimizing geodesic $\gamma(t)$ joining A_1 and A_2 is $\gamma(t) = A_1 \exp\left\{t \log(A_1^{-1}A_2)\right\}$, which is unique if and only if A_2 is not in the cut locus of A_1, i.e tr $(A_1^{-1}A_2) \neq -1$, see [71].

Therefore, from Equation (3.7) and (3.8), for all $A \in \mathcal{SO}(3)$ we get

$$\left((dL_{A^{-1}})_A \circ \text{Log}_A\right) : \mathcal{SO}(3) \to \mathfrak{so}(3) \qquad B \mapsto \log(A^{-1}B).$$

And since $A \in \mathcal{SO}(3)$, $A^{-1} = A^T$ implies reduction in computational cost. Moreover, given that $A \in \mathcal{SO}(3)$ is not in the cut locus of some $X \in \mathcal{SO}(3)$, we can always express $\log(A^T X)$ in terms of the basis $\mathscr{B}$, i.e. there exists $\{x^{(i)} \in \mathbb{R} \,|\, 1 \leq i \leq 3\}$ such that $\log(A^T X) = \sum_{i=1}^{3} x^{(i)}\, b_i$. In previous chapter, we know that the formulas for the matrix exponential/logarithm in general are complicated, as they involve power series, see Definition 2.35, however, on $\mathcal{SO}(3)$ it is much easier to do so, see the following proposition.

Proposition 3.4 (Rodrigues' formula)**.** *[71] Consider $A \in \mathcal{SO}(3)$ and $a \in \mathfrak{so}(3)$, we define*

$$\xi = \sqrt{\frac{tr\,(a^T a)}{2}} = \frac{||a||_F}{\sqrt{2}} \quad \textit{and} \quad \theta = \cos^{-1}\left(\frac{tr\,A - 1}{2}\right) \quad \textit{with} \quad |\theta| < \pi.$$

Then, the matrix exponential/logarithm are computed by

$$\exp a = \begin{cases} I_3 & \text{if } \xi = 0, \\ I_3 + \frac{\sin\xi}{\xi}\, a + \frac{1-\cos\xi}{\xi^2}\, a^2 & \text{if } \xi \neq 0. \end{cases} \tag{3.9}$$

$$\log A = \begin{cases} 0_3 & \text{if } \theta = 0, \\ \frac{\theta}{2\sin\theta}\,(A - A^T) & \text{if } \theta \neq 0. \end{cases} \tag{3.10}$$

The formula for log has singularities at $\theta = \pm\pi$ (i.e. tr $A = -1$), which are the cut locus of I_3 in $\mathcal{SO}(3)$.

Remark 3.5. [72] For any $A \in \mathcal{SO}(3) \setminus \{I_3\}$ with tr$A \neq -1$, there exists uniquely a skew symmetric matrix ω and a real number ϕ such that $A = \exp(\phi\,\omega)$ and

$||\omega||_F = 2$, i.e. $\omega = \sum_{i=1}^{3} \omega^{(i)} b_i$ with a unit vector $\big(\omega^{(1)}, \omega^{(2)}, \omega^{(3)}\big)$ in $\mathbb{R}^3$. This is a direct consequence from Equation (3.10), that is to set

$$\omega = \frac{2(A - A^T)}{||A - A^T||_F} \qquad \& \qquad \phi = \frac{\theta\, ||A - A^T||_F}{4 \sin\theta}.$$

It is easy to see that ϕ is uniquely determined by A, we call ϕ the *Rodrigues' rotation angle* of A. Moreover, we get $||\log A||_F = |\phi|\, ||\omega||_F = 2\,|\phi|$, and $\phi = 0$ if any only if $A = A^T$, which is equivalent to the case $A = I_3$.

Next, we study the impact of different base points on the fitted values when using GLS, given that the parametrization employs the skew symmetric matrices (i.e. via logarithm map) introduced in Subsection 3.2.2. We intend to bound the difference between fitted values from models associated with different base points when using OLS/GLS, in term of the Rodrigues' rotation angle.

Firstly, we know that $\exp(x)\exp(y) = \exp(x + y)$ holds for real numbers, but on $\mathcal{GL}(n)$ this equation does not hold because the group $\mathcal{GL}(n)$ is non-commutative. In fact, for any Lie group the Baker-Campbell-Hausdorff formula [73] approximate $\mathrm{Log}\big(\mathrm{Exp}(X)\mathrm{Exp}(Y)\big)$, given that X and Y are elements of this Lie group, while Exp and Log are the corresponding exponential/logarithm maps defined in Remark 2.34.

Theorem 3.6 (Baker–Campbell–Hausdorff formula)**.** *[73] Suppose that X, Y, Z are elements of the Lie algebra $\mathfrak{g}$ such that Exp(Z) = Exp(X)Exp(Y), then Z is approximated by the following series expansion in term of the Lie bracket:*

$$Z = X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}([X, [X, Y]] - [[X, Y], Y]) + \cdots \tag{3.11}$$

In particular, for a Lie group of matrices, $[X, Y] = XY - YX$, while Exp/Log coincide with the standard matrix exponential/logarithm. Thus, $Z = \log\big(\exp X \exp Y\big)$ is now approximated by $Z = Z_1 + Z_2 + Z_3 + \cdots$, where

$$Z_1 = X + Y \qquad \& \qquad 2\, Z_2 = XY - YX,$$
$$12\, Z_3 = X(XY - YX) - (XY - YX)X - (XY - YX)Y + Y(XY - YX).$$

It is clear that if the matrix Lie group of interest is commutative, we get back the usual result that applies to the one-dimensional case, i.e. $\log\big(\exp X \exp Y\big) = X + Y$.

Furthermore, we show that the action of Lie bracket on elements of $\mathcal{SO}(3)$ is equivalent to the cross product of coefficient vectors on $\mathfrak{so}(3)$ with respect to the basis $\mathscr{B}$ defined in Equation (3.6), see following lemma.

Lemma 3.7. *For any* $X, Y \in \mathfrak{so}(3)$*, we write*

$$X = \sum_{i=1}^{3} x^{(i)}\, b_i \quad \textit{and} \quad Y = \sum_{i=1}^{3} y^{(i)}\, b_i \;\; \textit{for some } x^{(i)}, y^{(i)} \in \mathbb{R}.$$

If $Z = [X, Y] = XY - YX$*, there exists* $z = \big(z^{(1)}, z^{(2)}, z^{(3)}\big) \in \mathbb{R}^3$ *such that* $z = x \times y$ *(i.e. the cross product of* x *and* y*) and* $Z = \sum_{i=1}^{3} z^{(i)}\, b_i$*.*

Proof. The commutation relations among elements of the basis $\mathscr{B}$ are $[b_1, b_2] = b_3\,;\, [b_2, b_3] = b_1\,;\,, [b_3, b_1] = b_2$ and $[b_i, b_j] = -[b_j, b_i]$ $(\forall 1 \le i, j \le 3)$. Therefore, for $Z = [X, Y]$ we get

$$\begin{aligned} Z &= \left[\sum_{i=1}^{3} x^{(i)}\, b_i\,,\, \sum_{j=1}^{3} y^{(j)}\right] = \sum_{i,j=1} x^{(i)} y^{(j)}\, [b_i, b_j] \\ &= \big(x^{(2)}y^{(3)} - x^{(3)}y^{(2)}\big)\, b_1 + \big(x^{(3)}y^{(1)} - x^{(1)}y^{(3)}\big)\, b_2 + \big(x^{(1)}y^{(2)} - x^{(2)}y^{(1)}\big)\, b_3. \end{aligned}$$

By setting $z^{(1)} = x^{(2)}y^{(3)} - x^{(3)}y^{(2)}\,,\;\; z^{(2)} = x^{(3)}y^{(1)} - x^{(1)}y^{(3)}$ and $z^{(3)} = x^{(1)}y^{(2)} - x^{(2)}y^{(1)}$, it is easy to see that if $z = \big(z^{(1)}, z^{(2)}, z^{(3)}\big)$, then $z = x \times y$ and $Z = \sum_{i=1}^{3} z^{(i)}\, b_i$. □

Let us restate the regression model with intercept using the parametrization that employs the logarithm map introduced in Subsection 3.2.2 for $\mathcal{SO}(3)$, that is for some base point $A \in \mathcal{SO}(3)$:

$$y = x_A \beta + \epsilon \qquad \text{with} \qquad \epsilon \sim \mathcal{N}\big(0_n, \sigma^2\, \Sigma\big). \tag{3.12}$$

Here, we assume that observations for the dependent variables and independent variables are $\{y_i\}_{i=1}^{n}$ and $\{X_i\}_{i=1}^{n}$, such that for all $1 \le i \le n$: $X_i =$

$A \exp\left\{\sum_{j=1}^{3} x_i^{(j)} b_j\right\}$ with

$$x_A = \begin{pmatrix} 1 & x_1^{(1)} & x_1^{(2)} & x_1^{(3)} \\ \dots & \dots & \dots & \dots \\ 1 & x_n^{(1)} & x_n^{(2)} & x_n^{(3)} \end{pmatrix}^T \in \mathbb{R}^{n\times 4}.$$

If $A = I_3$ (i.e. a trivial base point) this parametrization is interpreted by having coefficients in the logarithm domain with respect to $\mathscr{B}$, whereas for a non-trivial base point A, before we perform those steps we need to apply a rotation in the reverse direction of A first, $A^T X_i$ (i.e. left translation with $A^T = A^{-1}$). On the other hand, when choosing a different base point B, we have $A^T X_i = (A^T B) \cdot (B^T X_i)$ for all $1 \leq i \leq n$ in which $A^T B \in \mathcal{SO}(3)$. In other words, changing base point requires secondary rotation $A^T B$, and if $A = B$, it makes sense since this additional rotation equals the identity map.

The essence of proving the main results, Theorem 3.10 for OLS and Theorem 3.11 for GLS, are carried out in two steps. Firstly, we use Theorem 3.6 to approximate the coefficients of $A^T X_i$ in term of those corresponding to $A^T B$ and $B^T X_i$ in the logarithm domain. Secondly, we use results that fitted values are unchanged under mean-centering in Remark 3.8 and a special form of transformation in Lemma 3.9 for independent variables on the Euclidean space, in order to bound the difference of fitted values in L^2-norm.

Remark 3.8. The approach of translation observations by a fixed vector $\mathfrak{c} = \left(\mathfrak{c}^{(1)}, \mathfrak{c}^{(2)}, \mathfrak{c}^{(3)}\right)$, that is for all $i \in \{1, \dots, n\}$

$$\left(1, x_i^{(1)}, x_i^{(2)}, x_i^{(3)}\right) \quad \mapsto \quad \left(1, x_i^{(1)} + \mathfrak{c}^{(1)}, x_i^{(2)} + \mathfrak{c}^{(2)}, x_i^{(3)} + \mathfrak{c}^{(3)}\right)$$

which does not alter fitted values. This is due to the fact that on each direction $j \in \{1, 2, 3\}$ observations were added with same amount $\mathfrak{c}^{(j)}$, and OLS or GLS balance the expected values of response when all predictors are zero with the intercept. Moreover, this transformation is sometimes referred to as mean-centering.

Lemma 3.9. *Suppose* $V \in \mathcal{GL}(p)$ *with* $p < n$ *and some correlation matrix* $\Sigma \in \mathcal{GL}(n)$. *If* $Y = XV$, *the following holds for* $X,\, Y \in \mathbb{R}^{n\times p}$:

$$X(X^T\Sigma^{-1}X)^{-1}X^T = Y(Y^T\Sigma^{-1}Y)^{-1}Y^T,$$

which implies fitted values using OLS/GLS are invariant under given transformation.

Proof. Since $Y = XV$, we have $\left(Y^T\Sigma^{-1}Y\right)^{-1} = V^{-1}\left(X^T\Sigma^{-1}X\right)^{-1}\left(V^T\right)^{-1}$,

$$\begin{aligned}\Rightarrow Y\left(Y^T\Sigma^{-1}Y\right)^{-1}Y^T &= XV\left\{V^{-1}\left(X^T\Sigma^{-1}X\right)^{-1}\left(V^T\right)^{-1}\right\}V^TX^T\\ &= X\left(X^T\Sigma^{-1}X\right)^{-1}X^T.\end{aligned}$$

□

We are now ready to prove our main results by firstly showing the case of OLS, followed by the case of GLS.

Theorem 3.10 (OLS)**.** *Consider two base points* $A, B \in \mathcal{SO}(3)$, *using OLS yields fitted values* $\hat{y}_A,\, \hat{y}_B$ *when using the base point* A *or* B *respectively in the model, i.e.*

$$\hat{y}_A = x_A(x_A^Tx_A)^{-1}x_A^Ty \quad \& \quad \hat{y}_B = x_B(x_B^Tx_B)^{-1}x_B^Ty. \tag{3.13}$$

Suppose that ϕ *is the Rodrigues' rotation angle of* A^TB, *defined in Remark 3.5, then there exists some positive constant* K *such that* $||\hat{y}_A - \hat{y}_B||_2^2 \leq \phi^2K$.

Proof. We express $A^TX_i,\, B^TX_i$ and A^TB in the logarithm domain with respect to the basis $\mathscr{B}$, for all $1 \leq i \leq n$ there exist $\mathfrak{a}_i,\, \mathfrak{b}_i,\, \mathfrak{c} \in \mathbb{R}^3$ such that

$$\log(A^TX_i) = \sum_{j=1}^{3}\mathfrak{a}_i^{(j)}\,b_j \;;\; \log(B^TX_i) = \sum_{j=1}^{3}\mathfrak{b}_i^{(j)}\,b_j \;;\; \log(A^TB) = \sum_{j=1}^{3}\mathfrak{c}^{(j)}\,b_j.$$

In other words,

$$x_A = \begin{pmatrix} 1 & \mathfrak{a}_1^{(1)} & \mathfrak{a}_1^{(2)} & \mathfrak{a}_1^{(3)}\\ \dots & \dots & \dots & \dots\\ 1 & \mathfrak{a}_n^{(1)} & \mathfrak{a}_n^{(2)} & \mathfrak{a}_n^{(3)}\end{pmatrix}^T \quad \& \quad x_B = \begin{pmatrix} 1 & \mathfrak{b}_1^{(1)} & \mathfrak{b}_1^{(2)} & \mathfrak{b}_1^{(3)}\\ \dots & \dots & \dots & \dots\\ 1 & \mathfrak{b}_n^{(1)} & \mathfrak{b}_n^{(2)} & \mathfrak{b}_n^{(3)}\end{pmatrix}^T$$

Applying Theorem 3.6 for $Z = \log(A^T X_i)$, $X = \log(A^T B)$ and $Y = \log(B^T X_i)$, together with Lemma 3.7, we have

$$\mathfrak{a}_i = \mathfrak{c} + \mathfrak{b}_i + \frac{1}{2}\mathfrak{c} \times \mathfrak{b}_i + \frac{1}{12}\{\mathfrak{c} \times (\mathfrak{c} \times \mathfrak{b}_i) - (\mathfrak{c} \times \mathfrak{b}_i) \times \mathfrak{b}_i\} + \cdots$$

Moreover,

$$\begin{aligned}(\mathfrak{c} \times \mathfrak{b}_i) &= \begin{pmatrix} \mathfrak{c}^{(2)}\mathfrak{b}_i^{(3)} - \mathfrak{b}_i^{(2)}\mathfrak{c}^{(3)} \\ \mathfrak{b}_i^{(1)}\mathfrak{c}^{(3)} - \mathfrak{c}^{(1)}\mathfrak{b}_i^{(3)} \\ \mathfrak{c}^{(1)}\mathfrak{b}_i^{(2)} - \mathfrak{c}^{(2)}\mathfrak{b}_i^{(1)} \end{pmatrix} = - \begin{pmatrix} 0 & -\mathfrak{c}^{(3)} & \mathfrak{c}^{(2)} \\ \mathfrak{c}^{(3)} & 0 & -\mathfrak{c}^{(1)} \\ -\mathfrak{c}^{(2)} & \mathfrak{c}^{(1)} & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathfrak{b}_i^{(1)} \\ \mathfrak{b}_i^{(2)} \\ \mathfrak{b}_i^{(3)} \end{pmatrix} \\ &= - \left(\sum_{j=1}^{3} \mathfrak{c}^{(j)}\, b_j \right) \mathfrak{b}_i = -\log(A^T B)\, \mathfrak{b}_i.\end{aligned}$$

$$\begin{aligned}\Rightarrow \mathfrak{c} \times (\mathfrak{c} \times \mathfrak{b}_i) - (\mathfrak{c} \times \mathfrak{b}_i) \times \mathfrak{b}_i &= -(\mathfrak{c} \times \mathfrak{b}_i) \times \mathfrak{c} - (\mathfrak{c} \times \mathfrak{b}_i) \times \mathfrak{b}_i \\ &= \big(\log(A^T B)\, \mathfrak{b}_i\big) \times \mathfrak{c} + \big(\log(A^T B)\, \mathfrak{b}_i\big) \times \mathfrak{b}_i \\ &= \log(H)\, \mathfrak{c} + \log(H)\, \mathfrak{b}_i,\end{aligned}$$

where, $\log(H) = \big(-\mathfrak{c}^{(2)}\mathfrak{b}_i^{(3)} + \mathfrak{b}_i^{(2)}\mathfrak{c}^{(3)}\big)b_1 + \big(-\mathfrak{b}_i^{(1)}\mathfrak{c}^{(3)} + \mathfrak{c}^{(1)}\mathfrak{b}_i^{(3)}\big)b_2 + \big(-\mathfrak{c}^{(1)}\mathfrak{b}_i^{(2)} + \mathfrak{c}^{(2)}\mathfrak{b}_i^{(1)}\big)b_3$, is a skew symmetric matrix. Therefore,

$$\mathfrak{a}_i = \mathfrak{c} + \left\{ I_3 - \frac{1}{2}\log(A^T B) + \frac{1}{12}\log(H) \right\} \mathfrak{b}_i + \boldsymbol{\mathcal{O}}(\mathfrak{c}^2).$$

Here, $\boldsymbol{\mathcal{O}}(\mathfrak{c}^2)$ stands for an infinite sum in which all elements include factors of either $(\mathfrak{c}^{(t)})^2$ or $\mathfrak{c}^{(l)}\mathfrak{c}^{(s)}$ for $l, s \in \{1, 2, 3\}$, and this sum is finite since A, B and X_i are finite real matrices.

Since $A, B \in \mathcal{SO}(3)$ are fixed, the vector $\mathfrak{c}$ is a constant vector. And at each data point, the coefficients of $\log(A^T B)$ are translated by $\mathfrak{c}$, Lemma 3.8 implies $\hat{y}_A$ remains the same if we replace $\mathfrak{a}_i$ by $\mathfrak{a}_i - \mathfrak{c}$. Thus it is sufficient to show that if $\mathfrak{a}_i = \big\{I_3 - \frac{1}{2}\log(A^T B) + \frac{1}{12}\log(H)\big\}\, \mathfrak{b}_i + \boldsymbol{\mathcal{O}}(\mathfrak{c}^2)$ then $||\hat{y}_A - \hat{y}_B||^2 \leq \phi^2 K$ for some positive constant K, where ϕ is the Rodrigues' rotation angle of $A^T B$ defined in Remark 3.5.

On the other hand, Remark 3.5 implies

$$\sum_{j=1}^{3}(\mathfrak{c}^{(j)})^2 = 4\phi^2 \;\Rightarrow\; \max\left\{\left|\mathfrak{c}^{(a)}\right|; 1 \le j \le 3\right\} \le 2\phi$$

$$\Rightarrow \left|\left|\mathfrak{a}_i - \left\{I_3 - \frac{1}{2}\log(A^TB) + \frac{1}{12}\log(H)\right\}\mathfrak{b}_i\right|\right|^2 \le \phi^2 M \text{ for some } M > 0.$$

Let us denote $U = -\frac{1}{2}\log(A^TB) + \frac{1}{12}\log(H)$ then U is also a skew symmetric matrix, i.e. $U = \sum_{j=1}^{3} u^{(j)}b_j$, because both $\log(A^TB)$ and $\log(H)$ are elements of $\mathfrak{so}(3)$. We then set $\tilde{x}_B = x_BU^*$, where

$$U^* = \begin{pmatrix} 1 & 0 \\ 0 & I_3 + U \end{pmatrix} \quad \text{with } \det(U^*) = \det(I_3 + U) = 1 + \sum_{j=1}^{3}\left(u^{(j)}\right)^2 > 0.$$

Thus, U^* is invertible, and using Lemma 3.9, we get

$$\hat{y}_B = x_B(x_B^Tx_B)^{-1}x_B^Ty = \tilde{x}_B(\tilde{x}_B^T\tilde{x}_B)^{-1}\tilde{x}_B^Ty$$

Since each row of $(x_A - \tilde{x}_B)$ has squared length bounded above by ϕ^2M,

$$\Rightarrow ||\hat{y}_A - \hat{y}_B||^2 \le \phi^2K \quad \text{for some positive constant } K.$$

□

Theorem 3.11 (GLS)**.** *Consider two base points $A, B \in \mathcal{SO}(3)$, using GLS associated with a correlation matrix Σ, yields fitted values $\hat{y}_A$, $\hat{y}_B$ when using the base point A or B respectively, i.e.*

$$\hat{y}_A = x_A(x_A^T\Sigma^{-1}x_A)^{-1}x_A^T\Sigma^{-1}y \quad \& \quad \hat{y}_B = x_B(x_B^T\Sigma^{-1}x_B)^{-1}x_B^T\Sigma^{-1}y. \tag{3.14}$$

Suppose that ϕ is the Rodrigues' rotation angle of A^TB, defined in Remark 3.5, then there exists some positive constant K such that $||\hat{y}_A - \hat{y}_B||_2^2 \le \phi^2K$.

Proof. Since the translation invariance and Lemma 3.9 hold for GLS, the same results of OLS will also hold when using GLS instead (i.e. $\Sigma \neq I_3$). □

Both Theorem 3.10 and 3.11 imply that the difference between fitted values changes with rate ϕ, which makes sense as the lower Rodrigues' rotation angle, the closer $A^T B$ to the identity matrix, at which $\phi = 0$ implies $A = B$. Moreover, this could reduce the number of potential base points before proceeding model fitting, by picking only several candidates that have large relative Rodrigues' rotation angles to each other. If observations are close to each other (e.g. the application of human joint movement in Section 3.4), then by picking the base point from one observations or Fréchet mean of all observations, there will be very small difference between fitted values.

Before proceeding to the application of human joint movement in the next section, we give a brief introduction of a convention of Euler angles proposed by E. Grood & W. Suntay [4], because it can happen that even in the same area of research, many convention of Euler angles are used, which might bring some confusion when understanding these angles.

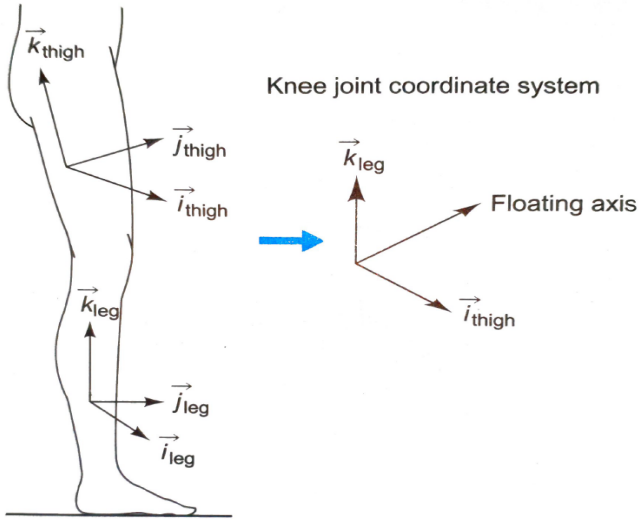


Figure 3.1: The joint coordinate system for the knee proposed by E. Grood & W. Suntay [4] (original source of the image [5]).

Any rotational transformation can be converted to a set of three angles, which describe the orientation of either a rigid object with respect to a fixed coordinate frame (i.e. extrinsic) or a moving frame of reference (i.e. intrinsic). However these sets of angles depend on the definition of the rotational frame, thus whenever using Euler angles, we always need to attach the convention of the frame used, i.e. how it

is defined, the ranges and signs of the angles.

A common way to collect 3D data of joint movement is using multiple cameras, placed around the desired objects, each camera will contribute a 2D coordinate, which is then used to compute 3D coordinates. The global (laboratory) coordinate system is fixed and a local coordinate system is attached to the moving body. E. Grood & W. Suntay [4] proposed a joint coordinate system of the knee consisting of three non-orthogonal axes. It is composed of two body fixed axes and the third axis (floating axis) is the mutually perpendicular of the two (i.e. equals the cross product of the other two axes), see Figure 3.1.

Joint angles (Euler angles) are defined to match clinical terminologies in biomechanics about joint rotation, which are flexion/extension, external/internal rotation and adduction/abduction. Flexion/extension occurs about the femoral fixed axis on the thigh ($\vec{i}_{\text{thigh}}$), external/internal rotation occurs about the tibial fixed axis on the leg ($\vec{k}_{\text{leg}}$) and finally adduction/abduction occurs about the floating axis.

Suppose we have a rotational matrix $R = \left(R_{ij}\right)$ of the right leg, then the computation of the Euler angles (α, β, γ) are

(i) Adduction/abduction $= \cos^{-1} R_{13} - \pi/2 = \beta - \pi/2$ with $\beta \in [0, \pi)$.

(ii) Flexion/extension $= \tan^{-1}(R_{23}/R_{33}) = \alpha$ with $\alpha \in [0, \pi]$.

(iii) External/internal $= \tan^{-1}(R_{12}/R_{11}) = \gamma$ with $\gamma \in [-\pi/2, \pi/2]$.

On the other hand, for the left knee the computation is carried out similarly, except adduction/abduction and external/internal must have their sign reversed in order to maintain the relative orientation between the right and left knee.

When $\beta = 0$ (i.e. 90° abduction), we can only uniquely determine β (i.e $\beta = 0$ or π) and the sum $\alpha + \gamma$, but not the two angles themselves. In this rare case, which is often referred to as gimbal lock, the rotational matrix R reduces to

$$R = \begin{pmatrix} 0 & 0 & 1 \\ -\sin(\alpha+\gamma) & \cos(\alpha+\gamma) & 0 \\ -\cos(\alpha+\gamma) & -\sin(\alpha+\gamma) & 0 \end{pmatrix},$$

and we lose a degree of freedom as changing α or γ does not change the rotation axis, i.e. the rotation is still about the Z-axis, which occurs when $\vec{i}_{\text{thigh}}$ becomes coincident with $\vec{k}_{\text{leg}}$. In general, regardless of the convention of Euler angles used, there will always be some points at which the Euler angles break down, which behaves similar to the cut locus when using exponential/logarithm map. However, this rare scenario is physically impossible with the knee/pelvis joint, Euler angles therefore can always be established uniquely. This indicates that it is possible for us to avoid falling case 3 mentioned in Subsection 3.2.2.

3.4 Application to soft tissue artefacts

In human kinematic study, the interest is often in studying the bone movement, which is a rigid motion. However, measuring the bone movement exactly is very difficult, expensive and time-consuming. Thus people commonly use skin-mounted markers to estimate the actual bone movement. These markers are affixed onto the skin, and get captured by camera during some required activities. One problem arises, that is when the muscle is contracted, which results in skin being stretched or slid under a certain activity, the relative position between the skin and the underlying bone changes. This difference in position is often called soft tissue artefact (STA).

An alternative method uses pin-mounted markers, which can be used to evaluate the validity of using skin-mounted markers, e.g. measuring STAs. Clusters of markers are mounted onto steel pins, which are implanted into the bone (e.g. tibia, femur and hip-bones), and skin displacements encountered (i.e. STAs) during certain activities are computed. These displacements are the differences between the actual skin markers position and the position in which the skin markers would be if only the solid motions (i.e. rotation and translation) that are implied by the change in position of the bone markers had taken place. Even though the use of pin-mounted markers brings more accurate representation of bone movement, it is more dangerous, difficult and expensive than using skin-mounted markers. The impact of STAs is therefore very important when studying the position and orientation of the bone movement.

In this section, we first describe the source of the data, together with reviewing a linear regression proposed by V. Camomilla et al. [3], that uses Euler angles, in Subsection 3.4.1, in which OLS is applied to estimate the regression parameters. Moreover, we see that there is a clear violation of the assumption about residuals, which suggests the use of GLS, which has not been considered elsewhere for this type of data. We implement GLS using two forms of parametrization, defined by either the Euler angles or the coefficients of skew symmetric matrices, to fit a model of STAs in Subsection 3.4.2. Due to the nature of the data, we assume two non-trivial correlation structures: Matérn and quasi-periodic correlations for GLS. We show that there is some reduction of some patterns among residuals.

3.4.1 The regression model proposed by V. Camomilla et al. [3]

The experiment was carried out on four adult cadaver subjects with twelve skin-mounted markers on the thigh skin along three longitudinal lines in antero-lateral, anterior and antero-medial positions. In addition, steel pins equipped with four-marker clusters, are mounted into tibia, femur and hip-bones. These markers are captured by a 9-camera stereophotogrammetric system at frequency 120 frames per second and the number of observations for each subject in each trial is summarized in Table 3.1.

	Subject 1	Subject 2	Subject 3	Subject 4
Trial 1	2423	1932	1671	2133
Trial 2	2498	2004	1860	2316
Trial 3	2042	1838	1968	2471

Table 3.1: Total observations for each subject in each trial from the ex-vivo study in [3].

The subjects are passively performed two movements: a star-movement and an arc-movement over three trials, see Figure 3.2. The reference position at each marker is placed in a femur anatomical frame. STAs at each marker are then computed by the displacement vector relative to the reference position in the global (laboratory) coordinate system. On the other hand, in order to compute rotations of the bone, the estimate of the hip joint centre (HJC) location is required. For our data set, the HJC is determined using a functional approach, see [6] for more details, with two main

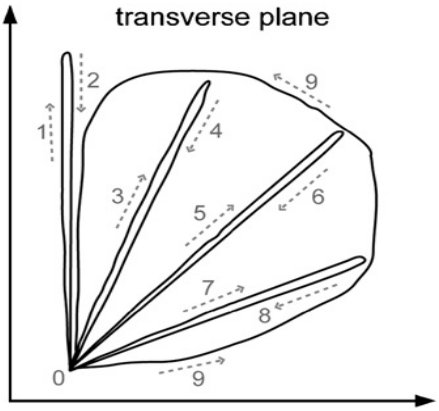


Figure 3.2: The foot trajectory of the lower limb movements relative to the pelvis is projected on the transverse plane during the star-movement followed by the arc movement. The direction of these movement is illustrated by the arrows, numbers 1-9, in which the initial position is indicated with number 0 (original source of the image [6]).

sources of errors for HJC determination coming from the stereophotogrammetric system and the STAs. Therefore, a collection of high-frequency observations of STAs and rotation matrices are recorded over time. Furthermore, trajectories of STAs are filtered with a 4^{th}-order low pass Butterworth filter with a cut-off frequency 10 Hz. The process of smooth filtering reduce the influence of noises, and in the field of biomechanic, practitioners often consider these noises to be undesirable, because it is not only more difficult to see the true pattern of movement, but also computation with differentiation (e.g. velocity, acceleration) can give unreasonably inaccurate values, see [74]. However, using a filter imposes correlation on errors, which in statistics adds more complexity in regression analysis.

V. Camomilla et al. [3] propose a linear regression model of STAs at each marker in each component against three Euler angles (flexion/extension, external/internal and adduction/abduction) computed by the convention from E. Grood & W. Suntay [4] of the thigh and shank, that is

$$\begin{aligned} y_{(m,c)} &= h^{\alpha}_{(m,c,\text{th})}\alpha_{\text{th}} + h^{\beta}_{(m,c,\text{th})}\beta_{\text{th}} + h^{\gamma}_{(m,c,\text{th})}\gamma_{\text{th}} \\ &\quad + h^{\alpha}_{(m,c,\text{sh})}\alpha_{\text{sh}} + h^{\beta}_{(m,c,\text{sh})}\beta_{\text{sh}} + h^{\gamma}_{(m,c,\text{sh})}\gamma_{\text{sh}} + h^{0}_{(m,c)} + \epsilon_{(m,c)} \end{aligned} \tag{3.15}$$

where

- $y_{(m,c)}$: Euclidean component c of the difference between skin-mounted marker m is versus where it would be expected to be under a solid rotation by $\alpha_{\text{th}}, \beta_{\text{th}}, \gamma_{\text{th}}$ in the $(\vec{i}_{\text{thigh}}, \vec{j}_{\text{thigh}}, \vec{k}_{\text{thigh}})$ system.

- $\alpha_{\text{th}}, \alpha_{\text{sh}}$: flexion/extension on the thigh/shank.

- $\beta_{\text{th}}, \beta_{\text{sh}}$: adduction/abduction on the thigh/shank.

- $\gamma_{\text{th}}, \gamma_{\text{sh}}$: external/internal rotation on the thigh/shank.

- $h_{(m,c)} = \left(h^{\alpha}_{(m,c,th)}, h^{\alpha}_{(m,c,sh)}, h^{\beta}_{(m,c,th)}, h^{\beta}_{(m,c,sh)}, h^{\gamma}_{(m,c,th)}, h^{\gamma}_{(m,c,sh)}, h^{0}_{(m,c)}\right)$ are model parameters at marker m in the component c when considering both thigh and shank.

- $\epsilon_{(m,c)}$: residuals.

Similar models are also considered in [75, 76] , all of which assume the errors are independent and identically normally distributed with same variance. Overall, at each marker the model consists of 21 unknown parameters. The intercept terms are added so that the STAs vector is a zero vector at the beginning of the experiment. The model parameters are estimated using OLS. While V. Camomilla et al. [3] use degrees as the unit of Euler angles, we use radian, simply because the coefficients of the skew symmetric matrices computed from the data are very small. This will only change the unit of the model parameters, whereas predictors or fitted values remain the same. While the linear model (3.15) provides a linear correlation of the STA affecting a given marker mounted on the thigh during certain movement with the time history of adjacent joints' rotations, there are high number of independent models corresponding to the numbers of skin markers and their Euclidean components. Alternatively, some works on reducing these numbers include fitting a marker-cluster located on the thigh [77, 78], where the STA of a marker cluster is transformed through a geometrical transformation into a modal series, e.g. projecting the STAs onto an appropriate basis vector representing the components of the soft tissue movement, see for examples [79, 80, 81]. Furthermore, another attempt of data reduction involves using the principal component analysis, see [60]. Here, models

often neglect all geometric nonlinearities, since the STA is relatively small compared to the overall rotation of the bones.

Next, before discussing a downside of the model proposed by V. Camomilla et al. [3], let us explore a problem when processing the data and how one wants to reduce its impact. Due to the method of collecting 3D data from using stereophotogrammetric system and human error when the hip joint is moving, abnormal observations occur as the result of sudden high speed (Figures A.1–A.3, Appendix A). Using camera system leads to apparent "jump", which are covered up by using the smooth filtering. These jumps result from discrete changes in the set of camera images being used to reconstruct the marker position at discrete time points. Without filtering these changes would yield actual jumps. With smooth filtering the trajectory stays continuous but high velocities are observed, e.g. the black dots spread out more during some particular time intervals in Figure 3.3. In other words, the multiple camera system fails to capture all twelve markers for some small time intervals and the filter method aim to guess these missing values.

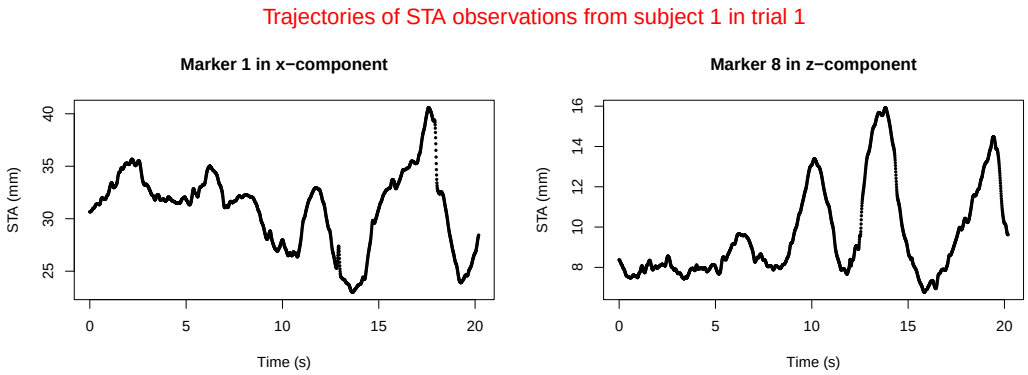


Figure 3.3: Trajectories of 2423 observations of STAs against time from subject 1, trial 1 (before removing abnormal observations).

We notice that the bad time intervals, where these problems happen, not only occurs with STAs observations (i.e. dependent variables y) but also rotation observations (i.e. independent variables X). This is a consequence of using the camera system to capture image of the cluster of four markers on the steel pin, which were implanted into tibia, femur and hip-bones. And observations of rotation and expected displacement of the skin are the results of these images. These bad time

intervals change for different subjects during different trials in both y on $\mathbb{R}^3$ and X on $\mathcal{SO}(3)$. Moreover, they do not always occur at the same time for different locations of markers. When using the model to predict STAs values, we can determine most of these bad time intervals through the independent variables (i.e. rotation matrices), see the comparison between Figure A.3 and A.4.

In order to minimize the impact, we remove partial data at which these problems occur, while we keep a sufficient amount from the dataset to use for parameter estimation. In Algorithm 3.12, we compare estimated speed using Euclidean distance of STAs in some specific markers, trials, subjects and discard observations that have high speed. And for prediction we will use an alternative algorithm, which involves only on the independent variables X, i.e. rotational matrices, see Algorithm 3.13. While Algorithm 3.12 results dependence on trial, subject and location of markers, Algorithm 3.13 implies independence from the location of the markers.

Suppose that we observe rotational matrices $\{X_{t_1}, \ldots, X_{t_n}\} \in \mathcal{SO}(3)$ and the corresponding STAs $\{y_{t_1}, \ldots, y_{t_n}\}$ for observation time $t = \{t_1, \ldots, t_n\}$. Both Algorithm 3.12 and 3.13 involve computation of estimated speed, followed by removal of high speed observations.

Algorithm 3.12 (STAs)**.** We use $\langle \cdot\cdot \rangle_F$ to compute the Euclidean distance.

1. Estimated speed between two adjacent STAs are computed by

$$v_i^{(y)} = \frac{d(y_{t_i}, y_{t_{i+1}})}{t_{i+1} - t_i} = \frac{||y_{t_{i+1}} - y_{t_i}||_F}{t_{i+1} - t_i} \qquad (1 \leq i \leq n-1).$$

2. Remove $\alpha\,\%$ part of data that have higher speed.

Algorithm 3.13 (Rotation)**.** We use a Riemannian metric tensor discussed in [71] to compute the Riemannian distance on $\mathcal{SO}(3)$.

1. Estimated speed between two adjacent rotational matrices are computed by

$$v_i^{(X)} = \frac{d(X_{t_i}, X_{t_{i+1}})}{t_{i+1} - t_i} = \frac{\left|\left|\log\left(X_{t_i}^T X_{t_{i+1}}\right)\right|\right|_F}{t_{i+1} - t_i} \qquad (1 \leq i \leq n-1).$$

2. Remove $\alpha\,\%$ part of data that have higher speed.

The choice of α varies and we recommend to keep in the range $(0, 10]$ as we don't want to remove too many observations. With the given set of data, we set $\alpha = 2$ throughout this study. The impact of performing either Algorithm 3.12 or Algorithm 3.13 on the removal of abnormal observations caused by sudden high speed, together with comparison between these algorithms can be seen in Figure A.3 and A.4 (Appendix A).

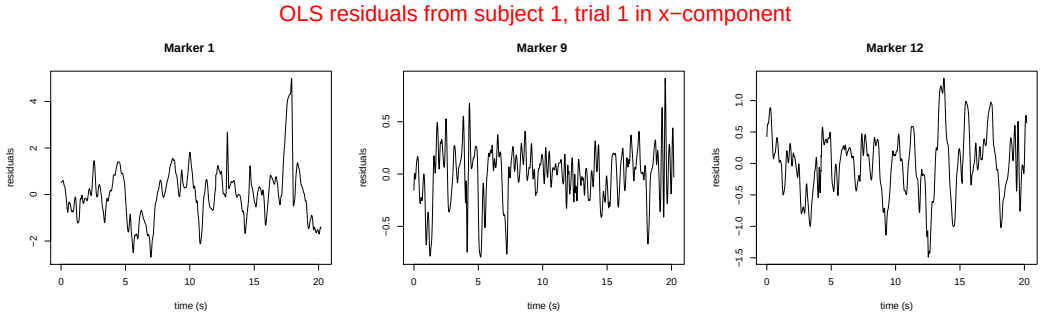


Figure 3.4: OLS residuals plots using regression model (3.15), which consisting 2376 observations (models are fitted after removing abnormal observations).

After prepossessing the data, we fit the regression model in Equation (3.15), in which we use OLS to estimate model parameters. It is clear that there are obvious patterns among residuals, see Figure 3.4. This indicates that the assumption about the independence among errors with same variance is violated. As we mentioned before, observations of STAs have been processed with a smooth filtering, see Figure 3.3. Since we are unable to access the raw data (i.e. before they are processed through the low-pass Butterworth filter), we therefore suggest using GLS to estimate model parameters as it takes into account the correlation between errors.

3.4.2 Model fitting using GLS

We aim to fit a linear regression after parametrization of rotation matrices, either defined by the Euler angles as in model (3.15) or by the coefficients of skew symmetric matrices as in model (3.12), to the experimental data obtained from the ex-vivo study in [3]. In this work, we assume that the model is subject-specific and trial-specific, based on the result from [3], for each subject we therefore construct

36 models for twelve markers in three Euclidean components. We mainly use data of trial 1, and possibly use trial 2 or trial 3 to validate the model. Moreover, we assume a non-trivial correlation structure when applying GLS, i.e. $\Sigma \neq I_n$.

Since the dimension of the $n \times n$ correlation matrix Σ is massive (after removing abnormal observations, $n = 2376$), it is impossible to estimate Σ properly unless we impose an explicit structure on Σ. We know that observations have been processed by a 4^{th}-order low pass Butterworth filter with a cut-off frequency 10 Hz. Since a Butterworth filter shifts the timing of the data, a 2^{nd}-order with a cut-off frequency 10 Hz is applied to the original time series (i.e. forward direction) and then again in the reverse direction (i.e. backward direction), in order to achieve the zero-phase shift, see [74, 82, 83]. This process contains complicated calculation and it is therefore impossible to extract the correlation matrix exactly from the information we have. However, we can still deduce that residuals have time dependence and possibly periodic behaviour. Thus, potential candidates for the correlation functions are the Matérn [84] and the quasi-periodic correlation functions [85], denoted by Σ^M, Σ^Q respectively, that is for $1 \leq i, j \leq n$:

$$\Sigma^M_{ij}(\rho, \nu) = \frac{2^{1-\nu}}{\text{Gam}_1(\nu)} \left(2\sqrt{\nu}|t_i - t_j|10^{\rho}\right)\, \kappa_\nu \left(2\sqrt{\nu}|t_i - t_j|10^{\rho}\right), \tag{3.16}$$

$$\Sigma^Q_{ij}(\rho, \nu, T, \omega) = \Sigma^M_{ij}(\rho, \nu)\, \exp\left(-\frac{\sin^2\left(\frac{\pi|t_i - t_j|}{T}\right)}{2\omega^2}\right). \tag{3.17}$$

with the gamma function for real numbers, Gam_1, and the modified Bessel function of the second kind, κ_ν. The positive parameter ρ plays important role in controlling the correlation decay time, while positive parameter ν has large impact on the roughness of the data, see Figure 3.5. In particular, a Gaussian process with Matérn covariance is $\lceil \nu \rceil - 1$ times differentiable in the mean-squared sense [85], where $\lceil \nu \rceil$ is the ceiling of ν. In other words, we expect that using Matérn with larger ν will produce smoother trajectories. On the other hand, the quasi-periodic correlation function is constructed such that there is an additional periodic term of frequency T and ω, which parametrizes the strength of the periodic effect.

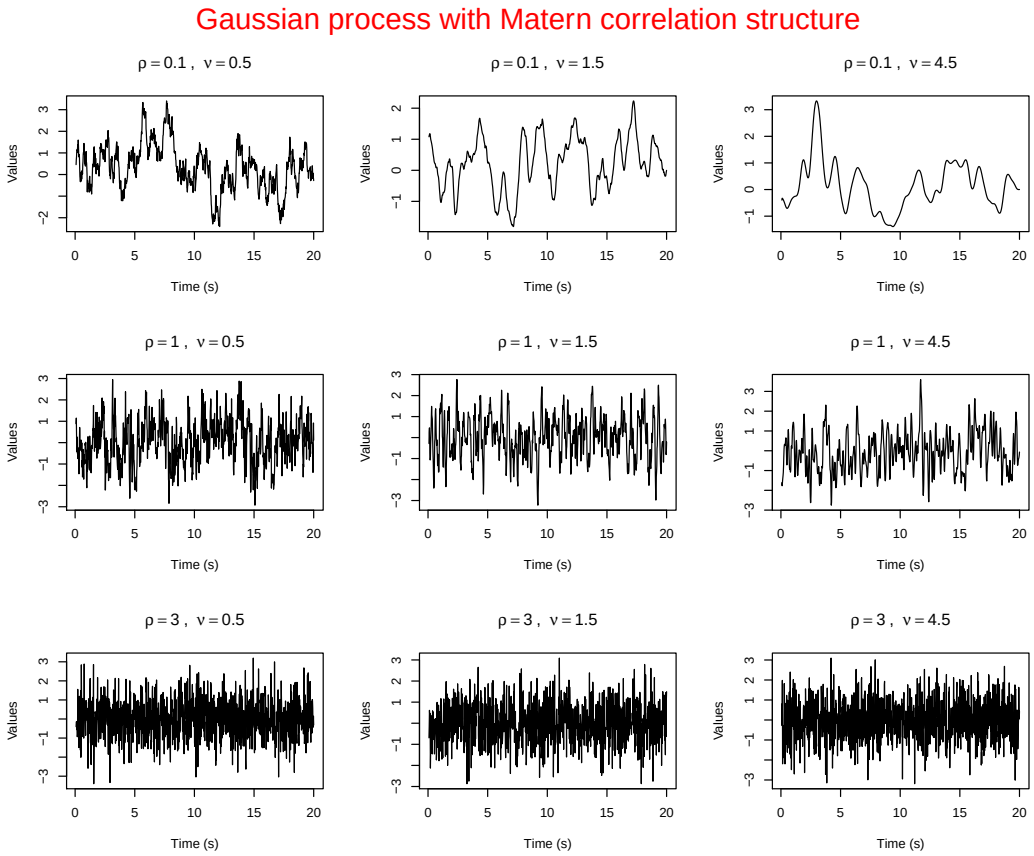


Figure 3.5: Simulation of Gaussian process equipped with the Matérn correlation structure using different pair values of (ρ, ν).

Computational complexity for both Matérn or quasi-periodic correlation function is high because they require solving the Bessel's differential equation. However, we can simplify the calculation for the Matérn correlation matrix in Equation (3.16) and (3.17) when $\nu = q + 1/2$ with $q \in \mathbb{N}$ to a product of an exponential and a polynomial of order q, see [86]. We know that our observations are filtered with a 4^{th}-order Butterworth filter, therefore we expect they are three times differentiable, see [87]. One option is therefore to choose $\nu = 3.5$, but matching only the order of differentiation may not be good thing if the true correlation structure is too different to the proposal form. In general, over-smoothing leads to more rapid loss of rate of convergence than under-smoothing in nonparametric estimation (see for example of the Bayesian context in [88]). Since the true (unknown) correlation structure can be far more different to either the Matérn or the quasi-periodic functions, it is safer to

choose a smaller value for ν. We compare Figure 3.4 versus Figure 3.5 and decide to set $\nu = 1.5$ for all models of twelve markers in three Euclidean components. It is left to estimate ρ for Σ^M and $\{\rho, T, \omega\}$ for Σ^Q. The computational efficiency is improved significantly because not only there are less parameters to estimate, but also the formulas of the correlation matrix are simplified to

$$\Sigma^M_{ij} = \left(1 + \sqrt{3}|t_i - t_j|10^\rho\right) \exp\left(-\sqrt{3}\,|t_i - t_j|\,10^\rho\right), \tag{3.18}$$

$$\Sigma^Q_{ij} = \Sigma^M_{ij} \exp\left(-\frac{\sin^2\left(\frac{\pi|t_i - t_j|}{T}\right)}{2w^2}\right). \tag{3.19}$$

We fit 36 models for each subject in trial 1 independently, using GLS associated with either Σ^M or Σ^Q, defined in Equation (3.18) or (3.19) respectively. We compare the result using either iterated GLS in Algorithm 3.3 and minimized GLS in Algorithm 3.2. We use the Cholesky decomposition on Σ such that $\Sigma = AA^T$, the assumption of GLS is fulfilled if the transformed residuals, i.e. $A^{-1}(y - x\hat{\beta}_{\text{GLS}})$, are independent and identically normally distributed with same variance.

Firstly, we compare results from GLS to the model (3.15) using Euler angles, where correlation matrix Σ equals to I_n (i.e. OLS), Σ^M and Σ^Q. Although using iterated GLS reduces some structure of residuals in both correlation structure (see Figure 3.6), it does not improve the fit (see Figure 3.7). In fact, it performs far worse than the OLS in terms of fitting. On the other hand, compared to the iterated GLS we get a better fit when using minimized GLS but there is less improvement on the independence assumption, see Figures 3.6 and 3.7. We can see from Table 3.2 and 3.3 that estimates for ρ when assuming the Matérn correlation structure tend to be larger when using minimized GLS across four subjects. In other words, the minimized GLS suggests higher correlation decay time than the iterated GLS, which again emphasizes that algorithms used for the minimized GLS and iterated GLS work differently. Moreover, estimates for ρ varies largely across different Euclidean components, different markers and different subjects. This result supports the finding in [3] that we can not assume same model for all markers to be used in a wider population, particularly our data source is only based on four cadaver subjects.

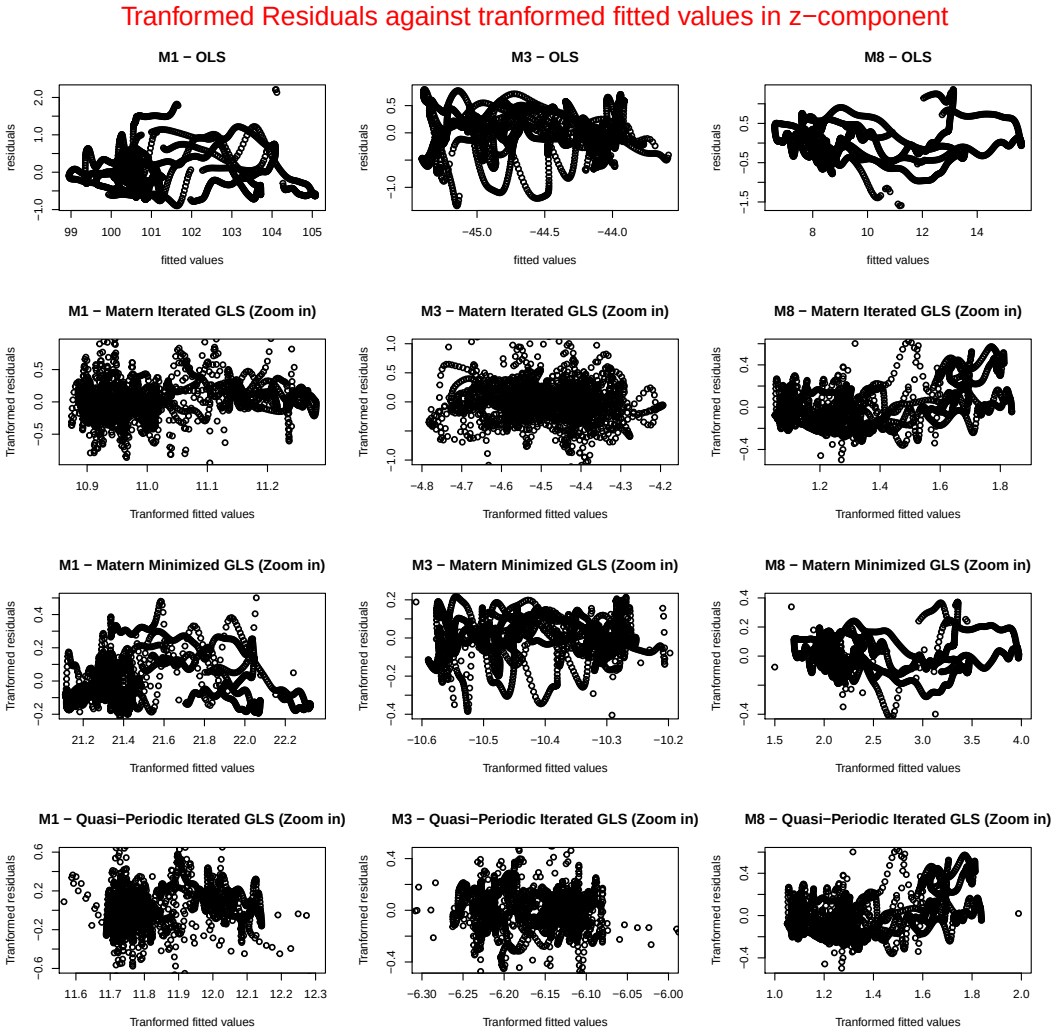


Figure 3.6: The residuals are plotted against fitted values for OLS (top row) and transformed residuals are plotted against transformed fitted values for GLS (bottom three rows), given that the regression model (3.15) fit Euler angles to STAs for subject 1 of 3 markers: 1, 3 and 8.

	Subject 1	Subject 2	Subject 3	Subject 4
x-component	$(0.187, 0.724)$	$(0.848, 1.29)$	$(0.347, 1.07)$	$(0.862, 1.48)$
y-component	$(0.00425, 0.705)$	$(1.27, 1.49)$	$(0.831, 1.24)$	$(0.525, 1.47)$
z-component	$(0.193, 0.757)$	$(0.471, 1.24)$	$(0.145, 1.27)$	$(1.02, 1.57)$

Table 3.2: Range of estimates for ρ accross 12 models (each links to 12 markers) defined in Equation (3.15), that use iterated GLS with Matérn correlation function for each subject in trial 1.

On the other hand, we can see from Figure 3.6 that the transformed residuals using the iterated GLS have similar patterns when using either the Maérn or

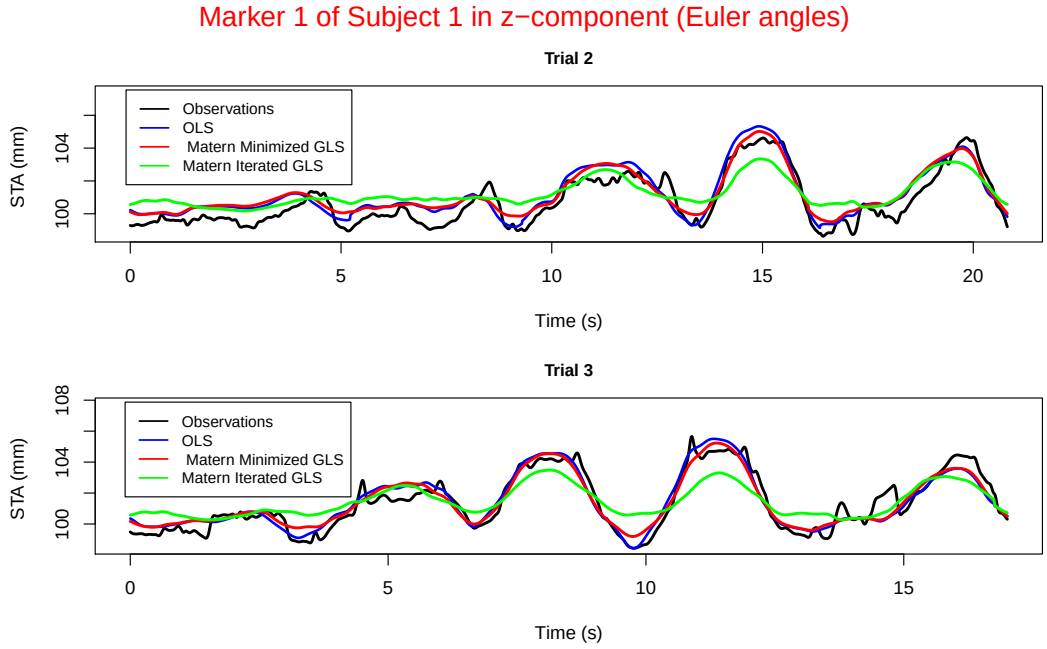


Figure 3.7: STAs predictions from models using OLS or Matérn (iterated/minimized) GLS models (3.15) for data in trial 2 and 3 (parameters are estimated from trial 1).

	Subject 1	Subject 2	Subject 3	Subject 4
x-component	$(1.00, 1.21)$	$(1.03, 1.67)$	$(0.858, 1.37)$	$(0.612, 1.15)$
y-component	$(0.839, 1.15)$	$(1.02, 1.28)$	$(0.605, 1.15)$	$(0.571, 0.818)$
z-component	$(0.972, 1.26)$	$(1.11, 1.55)$	$(0.925, 1.16)$	$(0.658, 1.02)$

Table 3.3: Range of estimates for ρ accross 12 models (each links to 12 markers) defined in Equation (3.15), that use minimized GLS with Matérn correlation function for each subject in trial 1.

the quasi-periodic functions. This suggests the additional periodic term does not help us to get closer to the true correlation structure. Therefore, we expect no significant improvement in removing obvious patterns in transformed residuals, and will not proceed the minimized GLS with the quasi-periodic function because the computational cost is considerably higher.

Secondly, we compare results from GLS to the model (3.12) using coefficients of skew symmetric matrices(via logarithm map) with respect to the basis $\mathscr{B}$ defined in Equation (3.6) using two non-trivial correlation structures Σ^M and Σ^Q. Selection for the base point is simple as the data of rotational matrices lie close to each other on $\mathcal{SO}(3)$. Therefore, we simply choose the first rotation as the base point for our

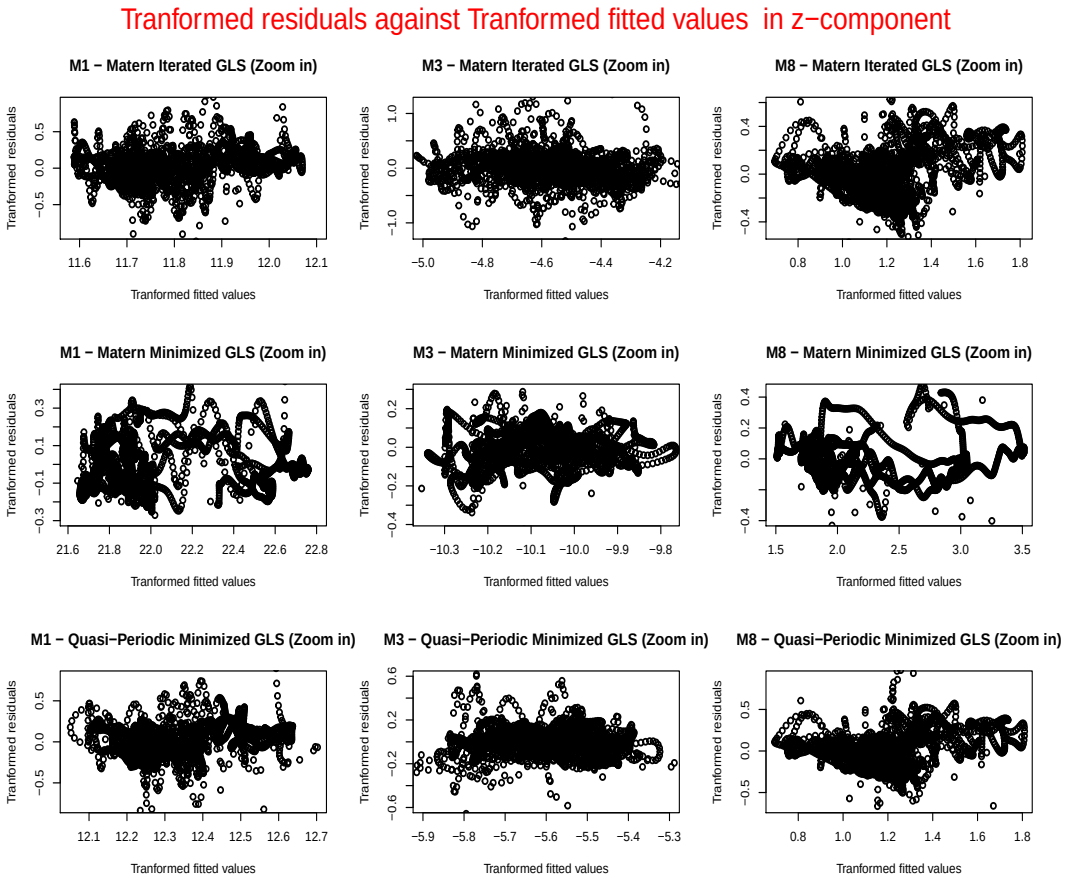


Figure 3.8: The transformed residuals are plotted against transformed fitted values using GLS, given that the regression model (3.12) fit coefficients of skew symmetric matrices (via logarithm map) to STAs for subject 1 of 3 markers: 1, 3 and 8.

model in Equation (3.12), and this is enough to make sure that the matrix logarithm function behaves nicely on $\mathcal{SO}(3)$. Similar to the model (3.15) using Euler angles, using minimized GLS does not worsen the fit, while it remove less structure of the residuals than iterated GLS, and having additional periodic term in the quasi-periodic correlation structure does not improve neither the fit nor the independence assumption on transformed residuals, see Figure 3.8.

For most markers, both models using parametrizations defined by Euler angles or skew symmetric matrices provide very close fit and result in similar patterns of residuals. Figure 3.9 suggests that the model of marker 1 for subject 1 using skew symmetric matrices somehow captures some small movement while the model with Euler angles can not. This could be due to the fact that the logarithm matrix is

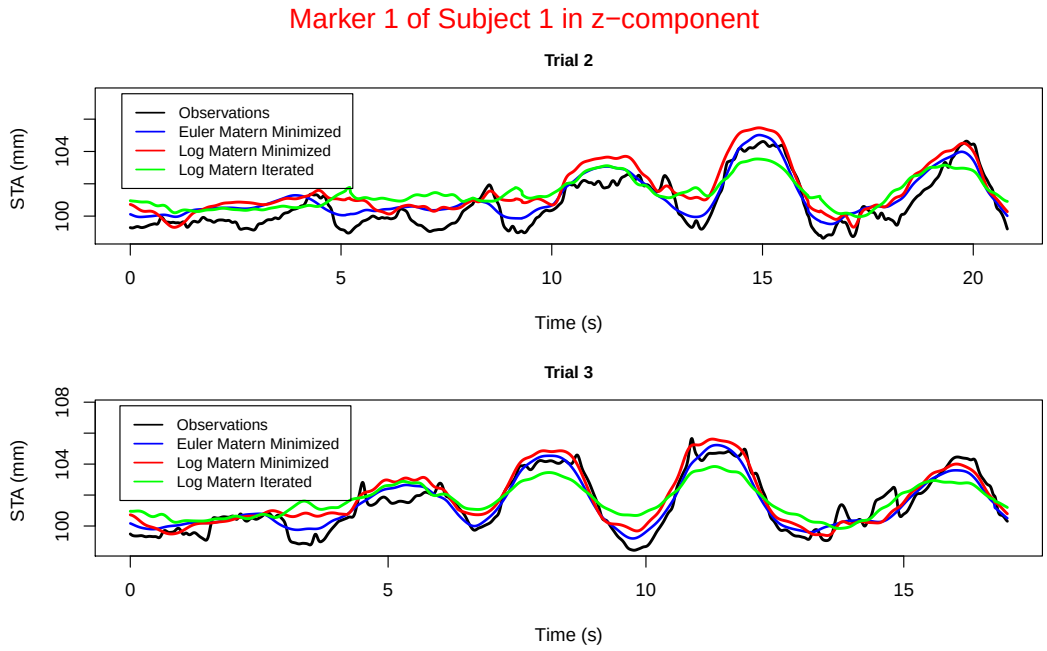


Figure 3.9: Predictions of STAs from models using GLS models equipped with the Matérn correlation structure for data in trial 2 and 3 (parameters are estimated from trial 1), in which parametrizations are defined by either Euler angles in model (3.15) or skew symmetric matrices (via logarithm map) in model (3.12).

non-linear, and thus could capture a smaller oscillation, while models using Euler angles can not.

Overall, models using OLS clearly fail to have approximately scattered residuals, we therefore propose using GLS associated with a non-trivial correlation matrix, instead of the identity matrix. The estimation from using two estimation methods: Algorithm 3.2 and 3.3 differ because they aim to optimize different things. And the difference between two algorithms get more pronounced due to the appearance of sudden high-speed events that we can not remove completely. The Matérn and the quasi-periodic correlation structure only improves to a certain degree, as they do not completely remove the pattern in residuals. This means there are information that both the Matérn and the quasi-periodic don't take into account. Since a Butterworth filter is used, correlation between two observations y_{t_i}, y_{t_j} depends not only themselves, but also observations that are recorded before [82], e.g. $y_{t_{i-2}}, y_{t_{i-1}}, y_{t_{j-2}}, y_{t_{j-1}}$, etc. Thus, one may need to look for alternative correlation structure.

3.5 Summary

In this chapter, by mapping the Lie group to its Lie algebra via the logarithm map with the left-invariant translation from a base point, we are able to apply the usual Euclidean regression tools to the coefficients with respect to some fixed orthonormal basis. Particularly, we revise generalized least squares and ordinary least squares on the Euclidean space, and extend to the compact connected Lie groups. The resulting estimations are unbiased estimators but no longer linear estimations as on the Euclidean space. We then focus on the space of rotational matrices $\mathcal{SO}(3)$, in which the Lie algebra is simply the space of skew symmetric matrices $\mathfrak{so}(3)$. We study the impacts of choosing different base points for our proposed parametrization on $\mathcal{SO}(3)$, and achieve an upper bound for differences of fitted values in L^2-norm in term of the Rodrigues' rotation angle.

We then fit regression models in the application of hip and knee joint movements, which helps learning the relationship between soft tissue artefacts and rotations of the hip/knee bones. The main contribution in this chapter involves the use of GLS in an effort to remove obvious patterns among residuals. Moreover, we compare the existing parametrization defined by Euler angles with our proposed parametrizaton that makes use of the logarithm map, when using two correlation structures, the Matérn and the quasi-periodic functions. The result indicates some reduction of obvious patterns among residuals, however, one may need to find other existing correlation structures or construct an error model to get closer to the true form of correlation among errors. Alternatively, higher order of regression can also be considered, such as the quadratic regression.

Furthermore, the practicability of the models mentioned in this chapter is limited in real life. Because we fit 36 independent regression models for each subject, which is tremendous although all models are quite simple (e.g. linear regression), and these models are subject-specific, therefore cannot be applied in a wider population. Additionally, there is high dependence among markers that are close to each other, reduction of correlated variables are hence recommended. Let us discuss in greater details about future work of this topic in Chapter 5.

Chapter 4

Stochastic differential equations on the cone of covariance matrices

4.1 Overview

Stochastic differential equation (SDE) has been used to model the time evolution of a vast range of applications, such as physics [89], biology [90], finance [91, 92], etc. They express the rate of change of the observed quantity as a function of the quantity's current value. This change can be decomposed into two parts: the drift and an added random component called the diffusivity. SDEs capture the stochastic effects, which offers the possibility of building models that are capable of predicting not only a realistic trajectory, but also the uncertainty of the prediction. While SDEs have been a very successful class of models for time series for several decades, little has been published on the case when the state space on which the system evolves is not simply Euclidean. Even though theory for SDEs on manifolds exists in the literature [25, 26], there is little on statistical inference for these processes, which makes people who have no prior knowledge about geometry find it hard to understand and apply it to statistical situations.

One special SDE on Euclidean space is the Ornstein-Uhlenbeck (OU) process driven by Brownian motion, which appears in many applications, such as biology [93], finance [94],etc. It is a Gaussian process with a bounded variance and there are analytical expressions for its transition probability density as well as the stationary

distribution, which unsurprisingly are Gaussian due to the linearity of the drift, see Equation (4.1). Moreover, it is the unique time-reversible diffusion with the Gaussian equilibrium distribution and constant diffusion coefficient [7], thus it is also ergodic. Another property that characterizes the OU process is mean-reversion such that it is parameterized by a speed and point of attraction and a constant diffusivity. There is another version of OU processes associated with Lévy processes, and OU process can be driven by jump processes, e.g. Poisson process [95]. We will only consider the OU process driven by Brownian motion in this thesis.

The OU process x_t on $\mathbb{R}^m$ that has full-rank diffusivity driven by Brownian motion is defined as the solution to the SDE :

$$dx_t = \theta(\mu - x_t)dt + \sqrt{\Sigma}\, dB_t \qquad (x_0 = u) \tag{4.1}$$

where $x_t : \Omega \to \mathbb{R}^m$ for some sample space Ω and B_t is the standard Euclidean Brownian motion on $\mathbb{R}^m$. The model parameters are $\theta \in \mathbb{R}_{>0}$, the speed of attraction, $\mu \in \mathbb{R}^m$, the point of attraction and $\Sigma \in \mathcal{SP}(m)$, the diffusivity.

Since the Gaussian distribution on a Riemannian manifold, which in general lacks linearity, behaves slightly different from the Euclidean case, thus defining OU process is no longer straightforward on manifolds. Besides non-linearity in the drift, the process might not be tractable, i.e. there might not be an explicit expression of transition probability density. The ergodic property is also very hard to be theoretically proved, thus it might be impossible to write down the expression of equilibrium distribution.

Generalization of OU process to manifolds.

There are some attempts of constructing an analogue of the OU process on several manifolds, see for examples [23, 8, 7, 10, 11]. The first attempt was proposed by J. Kent [23] for the sphere $\mathcal{S}^1$, in which he proposes a circular OU process, also referred to as the von Mises process. This process describes the evolution of angles

through time, and is a solution to the following SDE:

$$d\Theta_t = \theta \sin(\mu - \Theta_t)dt + \sigma dB_t \qquad (\Theta_0 = u),$$

where B_t is the standard Brownian motion on $\mathbb{R}$ with $\theta,\ \sigma \in \mathbb{R}_{>0}$ and $\mu \in [-\pi, \pi)$, see [7]. This process is ergodic and its equilibrium distribution is simply the von Mises distribution $\mathrm{vM}\big(\mu, 2\alpha/\sigma^2\big)$, where the probability density function of $\mathrm{vM}(\mu,\ \kappa)$ is given by

$$f_{\mu,\kappa}(\Theta) = \frac{1}{2\pi\mathfrak{I}_0(\kappa)} \exp\big\{\kappa\cos(\Theta - \mu)\big\}, \qquad \Theta,\ \mu \in [-\pi, \pi);\ \kappa \geq 0.$$

Here, $\mathfrak{I}_0$ is the modified Bessel function of the first kind with order 0. The parameters $\mu,\ \kappa$ depict the location and concentration of the distribution respectively, see [96]. The transition probability density of this von Mises process is not tractable and is therefore usually approximated by the Fokker–Planck or Kolmogorov's equation. A more general theory was introduced afterwards by J. Kent [8], in which a family of densities are established that are the equilibrium distributions of some time-reversible diffusions. Moreover, von Mises processes are also generalized to higher dimension in [8], i.e. sphere $\mathcal{S}^n$ with $n > 1$, which have the von Mises-Fisher distribution as the equilibrium distribution. In a similar manner, M. Wilkinson & P. Alain [9] use angular velocity to characterise the random movement on the circle and the sphere, in which models are then implemented to model objects tumbling in a turbulent environment.

After some time, E. P. García et al. [7] took inspiration from the theory by J. Kent [8], and applied it to the torus. By choosing some well-known probability distribution for directional statistics, the authors constructed Langevin diffusions for which the stationary distribution of those process turn out to be the chosen probability distributions. The reason for choosing Langevin diffusions is that those processes are ergodic and time-reversible with the chosen probability distributions. And by applying the theory proposed by J. Kent [8], we can conclude that the stationary distributions of those processes are in fact the corresponding chosen probability

distributions. Since OU process on the Euclidean space is driven by the Brownian motion whose the stationary distribution is the Gaussian, this construction of Langevin diffusions builds an analogue of the OU process on the torus [7], by choosing probability distributions that have similarities as Gaussian on the Euclidean. In the case of the torus, there are two common choices, which are wrapped normal and von Mises distributions. The same problems occur on $\mathcal{S}^n$, there are no closed forms of transition density, one needs to approximate by solving the Fokker–Planck or Kolmogorov's equation through numerical methods. However, the main challenge in [7] is not caused by the curvature of manifolds but by complex changes in rotations on the torus.

Besides $\mathcal{S}^n$ and the torus, there are some works of constructing OU process on shape manifolds. Some previous work make use of existing Euclidean approaches through local charts, i.e. expressing them in some chosen coordinates systems of shape space. For instance, F. G. Ball & I. L. Dryden et al. [11] initially construct Brownian motion on planar shape space, by expressing them in local coordinate form, i.e. finding the drift and diffusion in term of Kendall and Goodall–Mardia coordinates. Next, some extra drift is added so that there is a tendency of returning to a reference shape and a stationary distribution exists. However, the likelihood cannot be computed explicitly and so some approximation methods are required, e.g. the Euler–Maruyama approximation method (see Subsection 2.3.1) in terms of Kendall and Goodall–Mardia coordinates.

While statistical inference through local charts simply means using usual Euclidean statistical methods locally, an alternative approach is from the intrinsic point of view, e.g. employing the exponential map. For instance, V. Staneva et al. [10] study the OU process intrinsically for shape manifolds, in which they focus on the shapes that do not change their topology, thus can be described by flow of diffeomorphisms. The OU process in [10] is defined to have the mean-reverting property, i.e. adding a drift such that it pushes the process toward the minimizer of the squared distance between the process and some attraction point. This idea is motivated by the fact that the drift of the SDE (4.1) equals to $-0.5\,\nabla_{X_t}(X_t - \mu)^2$ on the

Euclidean space. The likelihood is approximated through the exponential adapted Euler–Maruyama method for high-frequency data. With the corresponding approximated likelihood, the maximum likelihood estimator can be expressed explicitly in some special cases. However, statistical inference for low frequency observations when using this intrinsic approach has not been explored in the literature. Moreover, if we assume the manifold of interest is parallelizable, and consider the special class mentioned in Subsection 2.3.2, where function $A(t, X_t)$ is now replaced by the Riemannian gradient of the squared distance function, we obtain many benefits. For instance, we get an efficient simulation method (i.e. using the exponential adapted Euler-Maruyama method), and many powerful theories are available (e.g. Itô formula, Girsanov-Cameron-Martin theorem, etc) by approximating the horizontal lift. Thus, besides achieving the mean-reverting property, defining the OU process in this approach on a Riemannian manifold simplifies the complexity of working with curved spaces to a certain degree.

Space of covariance matrices $\mathcal{SP}(n)$.

Our main interest is the space of $n \times n$ symmetric positive definite matrices $\mathcal{SP}(n)$, which appears in many contexts, such as in diffusion tensor imaging, volatilities for multiple financial assets, and generally wherever a Gaussian distribution is used in a multivariate context. Here, we only focus on covariance matrices which are strictly positive definite, unless indicated otherwise, i.e. considering only points inside the cone and excluding the boundary in Figure 4.1. We therefore sometimes refer to $\mathcal{SP}(n)$ as the cone of covariance matrices. We will briefly introduce some application fields of symmetric positive definite matrices.

In medical image analysis, covariance matrices arise from the diffusion tensor image (DTI), and these matrices are elements of $\mathcal{SP}(3)$; see for examples [32, 13, 97]. The diffusion of water molecules is used to generate contrast in magnetic resonance images, which reveal the pattern of the diffusion process of water molecules [98, 99]. It is usually assumed that the random motions of water molecules follow Brownian motion. Thus, positions of the water molecules therefore follow a multivariate Gaussian distribution, and the DTI is indeed the covariance matrix of the diffusion

process. Another example is in financial mathematics, where financial time series data of multiple assets (such as stocks, bonds, interest rates, etc.) can be modelled using SDEs, e.g. multivariate geometric Brownian motion is used in the Black-Scholes model to price assets [100, 101]. Covariance matrices in this field sometimes are referred to as volatilities, and they are elements of $\mathcal{SP}(n)$ with n the number of underlying assets.

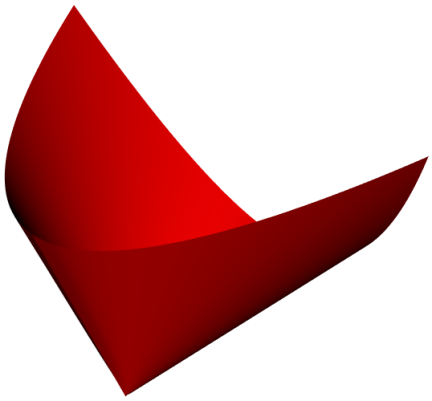


Figure 4.1: Illustration of the cone $\mathcal{SP}(2)$ in $\mathbb{R}^3$, where the constraints are $x, y \geq 0$ and $z = \pm\sqrt{xy}$

It is not difficult to show that the space $\mathcal{SP}(n)$ is a convex cone, for instance Figure 4.1 shows a clear picture for this space when $n = 2$. In fact, $\mathcal{SP}(n)$ is a smooth manifold of dimension $d = n(n+1)/2$, and is an example of a parallelizable manifold, see [102]. While there are many Riemannian metric tensors on $\mathcal{SP}(n)$ [13], we focus on the two most popular ones, namely the Affine-Invariant (AI) metrics [33, 34, 35] and the Log-Euclidean (LE) [32, 33]. In particular, the formulas for the orthonormal global basis field when $\mathcal{SP}(n)$ is equipped either with the AI or LE metrics are available explicitly, see Proposition 4.2 and 4.3. The AI metric can be constructed through various approaches, for instance as Fisher information in the multivariate normal model, see [35, 36], or as group action to an affine group, see [33]. Equipping $\mathcal{SP}(n)$ with the AI metric, we achieve a geodesically complete and simply connected Riemmanian manifold with non-positive sectional curvature, i.e. a Cartan-Hadamard manifold. However, working with a curved surface increases complexity and thus computational cost is high. To overcome this problem, V. Arsigny et al. [32] propose a new matrix operation, i.e. for any $P_1, P_2 \in \mathcal{SP}(n)$,

$P_1 \odot P_2 \coloneqq \exp(\log P_1 + \log P_2)$, and $\mathcal{SP}(n)$ equipped with this new operation is now an Abelian Lie group. Thus, Proposition 2.30 guarantees the existence of a bi-invariant metric on $\mathcal{SP}(n)$. Moreover, Proposition 2.28 defines the LE metric by constructing from an induced Euclidean metric, i.e. the Frobenius inner product on $\mathcal{S}(n)$. It is verified in [32] that the LE metric is a Riemannian metric on $\mathcal{SP}(n)$ and we obtain a vector space of null curvature, i.e. having the Euclidean structure. Thus, many Euclidean statistical tools can be applied, which require less computational resource than when using the AI metric. Moreover, $\mathcal{SP}(n)$ is an example of a symmetric space, so equipping $\mathcal{SP}(n)$ with the AI metric results in a non-compact space [103], and this is also true for the LE metric because we achieve Euclidean structure in this case.

As previously discussed in Section 2.3.2, there are various ways to study SDEs, particularly the OU process, on a Riemannian manifold, such as embedding, local charts, exponential maps, etc. Since $\mathcal{SP}(n)$ endowed with either the LE and the AI metric is geodesically complete, studying the manifold from an intrinsic point of view becomes more beneficial. We therefore define the OU class of processes on $\mathcal{SP}(n)$ in a similar manner to the paper by V. Staneva et al. [10], such that the mean-reverting property is achieved. For high-frequency data, it is not difficult to approximate the likelihood, e.g. via the exponential adapted Euler-Maruyama method, so there is no difficulty for statistical inference in this case.

However, in practice it is not always the case that high-frequency data is available, e.g. because it is too expensive to observe, or missing data, etc. Statistical inference for low-frequency data therefore receives much attention. On the Euclidean space, a common approach is to use an MCMC data-augmentation algorithm, i.e. imputing latent data points between every adjacent observations to allow the Euler-Maruyama method to give better approximation of the likelihood. This would require a diffusion bridge sampler on manifold, in our case on $\mathcal{SP}(n)$. Diffusion bridges in the Euclidean case have been studied for decades, but due to the complexity of manifolds, little has been done in the non-Euclidean case. The most common approach is using embedding or local charts followed by an appropriate Euclidean

method, see [42]. This algorithm will not be suitable in practice when the charts or embedding methods are too complicated. And since exponential maps work nicely on $\mathcal{SP}(n)$, we aim to develop diffusion bridge sampling method which should work well with this extra information. The main challenge in this study is associated with the presence of the curvature, i.e. $\mathcal{SP}(n)$ equipped with AI metric.

Outline.

First of all, we discuss Riemannian metrics on $\mathcal{SP}(n)$ in Section 4.2, where our focus are the Log-Euclidean and Affine-Invariant metrics. We review some calculations including the exponential/logarithm maps, geodesic, Riemannian distance, Riemannian gradient of the squared distance function, etc, because they are essential for the establishment of the Riemannian OU process on $\mathcal{SP}(n)$. Moreover, we also discuss the importance of Riemannian geometry when dealing with covariance matrices toward the end of Section 4.2. It is then followed by Section 4.3 and 4.4, where we discuss the Riemannian Brownian motion class and OU class respectively. Particularly, local expressions of horizontal lifts of smooth curves on $\mathcal{SP}(n)$ equipped with either the LE or the AI metrics are presented in Section 4.3.

The rest of this chapter is structured as follows. We firstly extend the notion of guided proposal process used in MCMC data augmentation sampling strategies to $\mathcal{SP}(n)$ equipped with the AI metric in Section 4.5. We then validate our guided proposal algorithm with simulation study on $\mathcal{SP}(2)$ in Section 4.6. Besides the task of estimating model parameters from simulation data, we also compare Brownian bridges on $\mathcal{SP}(n)$ equipped with either the Euclidean, the LE or the AI metrics; see Subsection 4.6.1. We demonstrate that the LE and the AI metrics should be preferred to the Euclidean metric, the problems (e.g. swelling effect, difficulty of sampling near the boundary of $\mathcal{SP}(n)$, etc) are particularly pronounced when conditioning on matrices with eigenvalues close to zero. Our financial data example in Section 4.7 is chosen to illustrate this exact point: one can use diffusions on $\mathcal{SP}(n)$ with LE and AI metrics for pricing or portfolio construction even when the dynamics on $\mathcal{SP}(n)$ operate near the boundary. Moreover, we perform goodness-of-fit test to compare the fit when considering either the LE or AI metrics.

Lastly, we explore the multivariate stochastic volatility models in Section 4.8, an example of a hidden Markov model, where covariance matrices are not directly observed. We assume the hidden covariance matrices follow the OU class defined in Section 4.4, and the derivative's underlying asset prices follow a linear SDE on the Euclidean space with volatility given by the unobserved covariance matrix process. In this section, we study the Metropolis-adjusted Langevin algorithm (MALA) for parameter estimation and illustrate simulation study on $\mathbb{R}^2$.

4.2 Riemannian metrics on $\mathcal{SP}(n)$

Since the tangent space at any point on $\mathcal{SP}(n)$ is the space of $n \times n$ symmetric matrices $\mathcal{S}(n)$ of dimension $d = n(n+1)/2$, we establish the *standard symmetric basis* $\mathfrak{B}_d = \{S_i\}_{i=1}^d$ for $\mathcal{S}(n)$ by defining an orthonormal basis with respect to $\langle \cdot, \cdot \rangle_F$:

$$\begin{aligned} &S_i = e_{ii}^{(n)} \qquad \text{for } 1 \le i \le n \\ &S_{n+1} = \left(e_{21}^{(n)} + e_{12}^{(n)}\right)/\sqrt{2}\,; \;\; S_{n+2} = \left(e_{31}^{(n)} + e_{13}^{(n)}\right)/\sqrt{2}\,; \ldots \end{aligned} \tag{4.2}$$

Here, $\{S_i\}_{i=1}^n$ has all entries zero except the jth entry on the diagonal being one. The remaining $\{S_i\}_{i=n+1}^d$ are obtained by adding, with $i > j$, the single-entry matrix $e_{ij}^{(n)}$ with one at the (i,j)th entry and zero elsewhere to its transpose and dividing to $\sqrt{2}$ so that it has unit Frobenius norm.

In the first two subsections, Subsections 4.2.1 and 4.2.2, we give a brief summary of existing results about the Log-Euclidean and Affine-Invariant metrics, such as exponential/logarithm maps, geodesics, squared distance function, etc. In preparation for the construction of the OU class on $\mathcal{SP}(n)$, we also compute the Riemannian gradient of squared distance associated with the corresponding Riemannian metrics. Lastly, since one can get different Riemannian structures by endowing with different Riemannian metrics, we compare different geometric structures of $\mathcal{SP}(n)$ when considering the induced Euclidean (or simply the Euclidean), the Log-Euclidean and the Affine-Invariant metrics in the Subsection 4.2.3.

4.2.1 Log-Euclidean metric

It is clear that $\mathcal{SP}(n)$ is not a group with the usual matrix multiplication since the symmetry property is not preserved under this operation. V. Arsigny et al. [32] propose a new matrix operation, referred to as the logarithm product and the logarithm scalar product, that is for any $P_1,\ P_2 \in \mathcal{SP}(n)$ and $\lambda \in \mathbb{R}$:

$$P_1 \odot P_2 = \exp(\log P_1 + \log P_2) \quad \& \quad \lambda \circledast P = \exp(\lambda \log P) = P^{\lambda}.$$

Consequently, we obtain an Abelian Lie group $\big(\mathcal{SP}(n), \odot\big)$, and its associated Lie algebra is $\mathcal{S}(n)$, see [32]. Having a Lie group structure does not on its own guarantee existence of a bi-invariant metric, but having the Abelian property does guarantee it, see Proposition 2.30. Moreover, Proposition 2.28 defines a left invariant Riemannian metric g^{LE} on $\mathcal{S}(n)$ from the Frobenius inner product $\langle \cdot, \cdot \rangle_F$ using the left translation L, that is for any $P \in \mathcal{SP}(n)$ and $X,\ Y \in \mathcal{S}(n)$,

$$g^{\mathrm{LE}}_P(X, Y) = \left\langle (dL_{P^{-1}})_P(X), (dL_{P^{-1}})_P(Y) \right\rangle_F. \tag{4.3}$$

Corollary 2.29 implies g^{LE} is indeed a bi-invariant Riemannian metric, and we call g^{LE} the Log-Euclidean (LE) metric. Moreover, we get a vector space structure with $\circledast$ on top of $\big(\mathcal{SP}(n), \odot\big)$ since $(\mathcal{SP}(n), \odot)$ is isomorphic and diffeomorphic to $\big(\mathcal{S}(n), +\big)$. This simply says even though $\mathcal{SP}(n)$ is not vector space of invertible matrices, we can interpret $\mathcal{SP}(n)$ as a vector space when we identify a symmetric positive definite matrix with its matrix logarithm.

In addition, it is easy to check that $\{\exp(tS) \,|\, t \in \mathbb{R}\}$ for some $S \in \mathcal{S}(n)$ forms an one-parameter subgroup of $\mathcal{SP}(n)$, see Definition 2.31. Furthermore, the geodesics associated with g^{LE} which pass through the identity I_n in the direction S in fact equals to the one-parameter subgroup $\{\exp(tS) \,|\, t \in \mathbb{R}\}$, this is the consequence of equipping $\mathcal{SP}(n)$ with a bi-invariant metric, see Section 2.2. Using the left translation L at some point $P \in \mathcal{SP}(n)$, the curve $\gamma(t) = \exp(\log P + tS)$ is also a geodesic on $\mathcal{SP}(n)$ that satisfies $\gamma(0) = P$ and $\gamma'(0) = D_{\log P} \exp .S$, which represents the derivative of matrix exponential function at $\log P$ in the direction S.

As a result, the exponential/logarithm maps associated with the LE metric at some point $P \in \mathcal{SP}(n)$ are defined by

$$\mathrm{Exp}_P^{\mathrm{LE}}(S) = \exp(\log P + D_P \log .S) \qquad \forall S \in \mathcal{S}(n), \tag{4.4}$$

$$\mathrm{Log}_P^{\mathrm{LE}}(Q) = D_{\log P} \exp .(\log Q - \log P) \qquad \forall Q \in \mathcal{SP}(n). \tag{4.5}$$

Here, we emphasize the derivative of matrix logarithm function $\log$ at P in direction S, i.e. $D_P \log .S$, is identical to the differential of $\log$ at P acts on S, i.e. $d\log_P(S)$. While $\mathcal{SP}(n)$ equipped with the LE metric is geodesically complete since the exponential map is a global diffeomorphism, the LE metric offers Euclidean structure in the domain of logarithm as the name suggests, see the following proposition.

Proposition 4.1. *Two Riemannian manifolds* $\big(\mathcal{SP}(n), g^{LE}\big)$ *and* $(\mathcal{S}(n), \langle\cdot,\cdot\rangle_F)$ *are isometric, where the matrix logarithm function* $\log : \mathcal{SP}(n) \to \mathcal{S}(n)$ *is an isometry.*

Proof. Since $\log : \mathcal{SP}(n) \to \mathcal{S}(n)$ is global diffeomorphism, it is sufficient to show that for all $P \in \mathcal{SP}(n)$ and $X, Y \in \mathcal{S}(n)$, the following equation holds

$$g_P^{\mathrm{LE}}(X, Y) = \big\langle d\log_P(X), d\log_P(Y) \big\rangle_F, \tag{4.6}$$

where $d\log_P$ is the differential of the matrix logarithm function $\log$ at P.

Consider the geodesic curve $\gamma(t) = \exp(\log P + tS)$ passes through $P \in \mathcal{SP}(n)$ and define the curve $\epsilon(t) = \big(L_{P^{-1}} \circ \gamma\big)(t) = \exp(tS)$ using the left translation $L_{P^{-1}}$. In other words, $\epsilon(t)$ is not only the image of $\gamma(t)$ under the left translation $L_{P^{-1}}$, but also a geodesic passing through the identity I_n in the direction S, and the differential map $dL_{P^{-1}}$ at P sends $\gamma'(t) \in T_P\mathcal{SP}(n)$ to $\epsilon'(t) \in T_{I_n}\mathcal{SP}(n) = \mathcal{S}(n)$. So, $d\left(L_{P^{-1}}\right)_P(v) = \epsilon'(0) = S$, where $v = \gamma'(0)$.

On the other hand, applying Definition 2.4 to the differentiable map $\log : \mathcal{SP}(n) \to \mathcal{S}(n)$, we also have $d\log_P(v) = \big(\log \circ \exp(\log P + tS)\big)'(0) = S$. Thus, $d\log = dL_{P^{-1}}$, and Equation (4.3) implies the required result.

□

In consequence, we say the LE metric a *flat Riemannian metric*, and $\mathcal{SP}(n)$

equipped with the LE metric therefore has null curvature everywhere. Moreover, the Riemannian distance associated with the LE metric between any two points $P,\, Q \in \mathcal{SP}(n)$ is simply given by the Euclidean distance in the log-domain, i.e.

$$\mathrm{d}_{\mathrm{LE}}(P,Q) = \sqrt{g_P^{\mathrm{LE}}\Big(\mathrm{Log}_P^{\mathrm{LE}}Q, \mathrm{Log}_P^{\mathrm{LE}}Q)\Big)} = \big|\big|\log P - \log Q\big|\big|_F. \tag{4.7}$$

As a result, any matrices with non-positive eigenvalues are at infinite distance of any covariance matrices. Furthermore, similarity (i.e. $A \mapsto BAB^{-1}$ for any invertible matrix B) and inversion are invariant with respect to the LE metric, see [32].

Lastly, we present a global orthonormal frame on the tangent bundle of $\mathcal{SP}(n)$ equipped with the LE metric and compute the corresponding Riemannian gradient of the squared distance function, which is crucial for our construction of the OU process on $\mathcal{SP}(n)$.

Proposition 4.2. *The set $\mathfrak{B}_d^{LE} = \{E_i^{LE}\}_{i=1}^d$ is an orthonormal frame on the tangent bundle $T\mathcal{SP}(n)$, where for any $P \in \mathcal{SP}(n)$:*

$$E_i^{LE}(P) = (d\log_P)^{-1}(S_i) = D_{\log P}\exp .S_i \qquad\qquad (1 \le i \le d). \tag{4.8}$$

Moreover, the Riemannian gradient of squared distance for any fixed point Q is

$$\Big(\nabla d_{LE}^2(P,Q)\Big)_P = -2\, D_{\log P}\exp .(\log Q - \log P) = -2\, Log_P^{LE}(Q).$$

Proof. It is straightforward to verify that the basis $\mathfrak{B}^{\mathrm{LE}}$ is an orthonormal frame on $T\mathcal{SP}(n)$ using Equation (4.6).

On the other hand, calculation of the Riemannian gradient given in Equation (2.3), together with Theorem 2.14 and Equation (4.5), (4.8) imply

$$\begin{aligned}
&g_P^{\mathrm{LE}}(\nabla f, E_i^{\mathrm{LE}}) = (E_i^{LE})_P f = 2g_P\Big(\nabla_{E_i^{\mathrm{LE}}}\mathrm{Log}_P^{\mathrm{LE}}Q, \mathrm{Log}_P^{\mathrm{LE}}Q)\Big)\\
&= 2\left\langle D_{\log P}(\log Q - \log P).S_i, \log Q - \log P\right\rangle_F = -2\left\langle S_i, \log Q - \log P\right\rangle_F\\
&\Rightarrow (\nabla f)_P = \sum_{i=1}^{d} g_P^{\mathrm{LE}}(\nabla f, E_i^{\mathrm{LE}})\, E_i^{\mathrm{LE}}(P) = -2\, D_{\log P}\exp .(\log Q - \log P).
\end{aligned}$$

□

4.2.2 Affine-Invariant metric

Before the establishment of the LE metric, there is another well-known metric tensor on $\mathcal{SP}(n)$, which also preserves the symmetry with respect to the inversion and offers infinite distance to a non-positive definite matrices, see for examples [32, 33, 34, 104, 36]. There are various approaches to construct this metric. In particular, one way is to define a smooth group action $\psi_R : P \mapsto R \star P = RPR^T$ for any $R \in \mathcal{GL}(n)$ on $\mathcal{SP}(n)$. This group action is linear and its differential map $d\big(\psi_R\big)_P : S \mapsto R \star S$ for $S \in T_P\mathcal{SP}(n)$, see [104]. A Riemannian metric g on $\mathcal{SP}(n)$ is ψ-invariant if for any $R \in \mathcal{GL}(n)$ the map ψ_R preserves the Riemannian distance on $\mathcal{SP}(n)$, that is for any $X, Y \in \mathcal{S}(n)$ and an arbitrary $P \in \mathcal{SP}(n)$:

$$g_P(X, Y) = g_{R\star P}(R \star X, R \star Y).$$

Since the symmetry to the affine transformation is respected, this metric g is therefore referred to as the Affine-Invariant metric [33], denoted by g^{AI}. Choosing $R = P^{-1/2} \in \mathcal{GL}(n)$ and the Frobenius inner product $\langle \cdot, \cdot \rangle_F$ on $T_{I_n}\mathcal{SP}(n)$, we get $R \star P = I_n$ and the equation above yields

$$\begin{aligned} g_P^{\text{AI}}(X, Y) &= \left\langle P^{-1/2} \star X, P^{-1/2} \star Y \right\rangle_F \\ &= \left\langle P^{-1/2} X P^{-1/2}, P^{-1/2} Y P^{-1/2} \right\rangle_F = \text{tr}\left(P^{-1} X P^{-1} Y\right) \end{aligned} \tag{4.9}$$

Consequently, the AI metric at any point $P \in \mathcal{SP}(n)$ can be transported back to the identity I_n through the map $\psi_{P^{-1/2}} : \mathcal{SP}(n) \to \mathcal{S}(n)$, where Equation (4.9) can be rewritten to

$$g_P^{\text{AI}}(X, Y) = \left\langle d\big(\psi_{P^{-1/2}}\big)_P(X), d\big(\psi_{P^{-1/2}}\big)_P(Y) \right\rangle_F. \tag{4.10}$$

Comparison of Equation (4.10) to the LE metric in Equation (4.6), we see that although both metrics are extended from the Frobenius inner product on the tangent space $T_{I_n}\mathcal{SP}(n)$, the function $\log$ offers independence of P, while the function

$\psi_{P^{-1/2}}$ clearly does not. This indicates the presence of non-zero curvature when equipping $\mathcal{SP}(n)$ with the AI metric, which contributes more complexity and therefore results in higher computational burden.

Additionally, the matrix form G_{AI} of g^{AI} with respect to $\mathfrak{B}_d$ is given by:

$$G_{\mathrm{AI}}(P) = D_n^T \cdot \left(P^{-1} \otimes P^{-1}\right) \cdot D_n \;\; \& \;\; G_{\mathrm{AI}}^{-1}(P) = D_n^{\dagger} \cdot (P \otimes P) \cdot (D_n^{\dagger})^T, \quad (4.11)$$

where $D_n \in \mathbb{R}^{n^2 \times d}$ is a constant matrix (referred to as the duplication matrix), that satisfies $\mathrm{vec}(P) = D_n\, \nu(P)$ with $\nu(P)$ containing all independent entries of P and $D_n^{\dagger}$ is the Moore-Penrose inverse of D_n, see details of calculation in [34].

Alternatively, the AI metric also arises from the theory of multivariate normal distribution, see [35, 34]. Here, we regard the collection of multivariate normal distributions of dimension n of zero mean, $\mathcal{M} = \{\mathcal{N}(0, \Sigma) \,|\, \Sigma \in \mathcal{SP}(n)\}$, as a Riemannian manifold of dimension $d = n(n+1)/2$, while endowing to this manifold a Riemannian metric that constructed through the Fisher information [105],

$$1 \leq i, j, k, l \leq n \;\; : \;\; I_{ijkl}(\Sigma) = \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial}{\partial \Sigma_{ij}}\Sigma^{-1}\frac{\partial}{\partial \Sigma_{kl}}\right),$$

where $\partial/\partial\Sigma_{ij}\,, 1 \leq i, j \leq n$ are derivatives of Σ at the $(i,j)^{\text{th}}$ entry.

Since $\mathcal{SP}(n)$ equipped with the AI metric is a symmetric space, the set of all isometries of $\mathcal{SP}(n)$ is a Lie group, see [103, 104, 33]. The geodesics pass through I_n are therefore generated by the action of the one-parameter subgroups of this Lie group, i.e. the curve $\epsilon(t) = \exp(t\tilde{S})$ is a geodesic that satisfies $\epsilon(0) = I_n$, $\epsilon'(0) = \tilde{S}$ for some $\tilde{S} \in \mathcal{S}(n)$. Using the group action $\psi_{P^{-1/2}}$ for an arbitrary $P \in \mathcal{SP}(n)$, the geodesics $\gamma(t)$ with $\gamma(0) = P$ and $\gamma'(0) = S \in \mathcal{S}(n)$ can be constructed such that the curve $\left(\psi_{P^{-1/2}} \circ \gamma\right)(t)$ is a geodesic passing through the identity in the direction $d\left(\psi_{P^{-1/2}}\right)_P(S)$. Since $d\left(\psi_{P^{-1/2}}\right)_P(S) = P^{-1/2} \star \gamma'(0)$, we set $\tilde{S} = P^{-1/2} \star S$, which implies $\gamma(t) = P^{1/2} \star \exp\left(P^{-1/2} \star tS\right)$. As a result, the exponential/logarithm maps associated with the AI metric at some point $P \in \mathcal{SP}(n)$ are defined by

$$\mathrm{Exp}_P^{\mathrm{AI}}(S) = P^{1/2} \star \exp\left(P^{-1/2} \star S\right) = P^{1/2} \exp\left(P^{-1/2} S P^{-1/2}\right) P^{1/2}, \quad (4.12)$$

$$\mathrm{Log}_P^{\mathrm{AI}}(Q) = P^{1/2} \star \log\left(P^{-1/2} \star Q\right) = P^{1/2} \log\left(P^{-1/2} Q P^{-1/2}\right) P^{1/2}, \quad (4.13)$$

for any $S \in \mathcal{S}(n)$ and $Q \in \mathcal{SP}(n)$. Accordingly, we also achieve geodesical completeness for $\mathcal{SP}(n)$ equipped with the AI metric, which is similar in the case of the LE metric, however, $\mathcal{SP}(n)$ has non-positive sectional curvature everywhere in the case of the AI metric, unlike the LE metric which offers null curvature. In fact, $\mathcal{SP}(n)$ equipped with the AI metric offers the geometric structure of a Cartan-Hadamard manifold, and it is parallelizable. The formula for the Riemannian distance $\mathrm{d_{AI}}$ associated with the AI metric between any two points $P, Q \in \mathcal{SP}(n)$ is more complicated than $\mathrm{d_{LE}}$ in Equation (4.7),

$$\begin{aligned} \mathrm{d_{AI}}(P, Q) &= \sqrt{g_P^{\mathrm{AI}}\left(\mathrm{Log}_P^{\mathrm{AI}}(Q), \mathrm{Log}_P^{\mathrm{AI}}(Q)\right)} \\ &= \left|\left| \log\left(P^{-1/2} \star Q\right)\right|\right|_F = \left|\left| \log\left(P^{-1/2} Q P^{-1/2}\right)\right|\right|_F. \end{aligned} \quad (4.14)$$

Even though working with AI metric brings more complications, there are explicit formula for the corresponding connection ∇, that is for any $P \in \mathcal{SP}(n)$,

$$(\nabla_X Y)_P = -\frac{1}{2}(X_P P^{-1} Y_P + Y_P P^{-1} X_P) \qquad \text{for } X, Y \in \Gamma(T\mathcal{SP}(n)), \quad (4.15)$$

which simplifies many calculations, see [34]. For instance, toward the end of this subsection we will show that the Christoffel symbol with respect to a special orthonormal basis is actually a constant in Lemma 4.4.

Similar to previous subsection of the LE metric, we would like to define a global orthonormal frame on the tangent bundle of $\mathcal{SP}(n)$ equipped with the AI metric, and compute the corresponding Riemannian gradient of squared distance, which is crucial for Section 4.4. We will use a special result from [36] when calculating the derivative that involves the matrix logarithm function. For a continuous function $\phi : \mathbb{R} \to \mathcal{SP}(n)$, i.e. all eigenvalues of $\phi(t)$ are strictly positive,

$$\frac{d}{dt} \mathrm{tr}\left\{\log^2\left(\phi(t)\right)\right\} = 2\,\mathrm{tr}\left\{ \log\left[\phi(t)\right]\left[\phi(t)\right]^{-1}\phi'(t)\right\} \quad (4.16)$$

Proposition 4.3. *The set* $\mathfrak{B}_d^{AI} = \{E_i^{AI}\}_{i=1}^d$ *is an orthonormal frame on the tangent bundle* $T\mathcal{SP}(n)$*, where for any* $P \in \mathcal{SP}(n)$*:*

$$E_i^{AI}(P) = d\big(\psi_{P^{1/2}}\big)_P(S_i) = P^{1/2} \star S_i \qquad \textit{for } 1 \leq i \leq d. \tag{4.17}$$

Moreover, the Riemannian gradient of squared distance for any fixed point Q *is*

$$\Big(\nabla d_{AI}^2(P,Q)\Big)_P = -2\sum_{i=1}^{d} \big\langle \log(P^{-1/2} \star Q), S_i \big\rangle_F \, E_i^{AI}(P) = -2\, Log_P^{AI}(Q).$$

Proof. Firstly, we know that $\psi_{P^{1/2}} \circ \psi_{P^{-1/2}}$ is an indentity map, so the differential map $d\psi_{P^{1/2}} = \big(d\psi_{P^{-1/2}}\big)^{-1}$. Using the Equation (4.10) implies $\mathfrak{B}_d^{\mathrm{AI}}$ is an orthonormal frame on $T\mathcal{SP}(n)$.

On the other hand, applying the calculation of the Riemannian gradient given in Equation (2.3) to $f(P) = d_{\mathrm{AI}}^2(P,Q)$, for all $1 \leq i \leq d$: $g_P^{\mathrm{AI}}(\nabla f, E_i^{\mathrm{AI}}) = df_P\big(E_i^{\mathrm{AI}}(P)\big)$. Moreover, using properties of the matrix logarithm function, we get

$$\begin{aligned}
\log(P^{-1/2} \star Q) &= \log\left(P^{-1/2} Q P^{-1/2}\right) = \log\left(P^{1/2} P^{-1} Q P^{-1/2}\right) \\
&= P^{1/2} \log\left(P^{-1} Q\right) P^{-1/2} = -P^{1/2} \log\left(Q^{-1} P\right) P^{-1/2} \\
\Rightarrow f(P) &= \mathrm{tr}\Big(\log^T(P^{-1/2} \star Q) \log(P^{-1/2} \star Q)\Big) = \mathrm{tr}\Big(\log^2\left(Q^{-1} P\right)\Big)
\end{aligned}$$

Let us fix $i \in \{1, \ldots, d\}$ and consider geodesics $\gamma(t) = P^{1/2} \star \exp(tS_i)$, that satisfy $\gamma(0) = P$, $\gamma'(0) = P^{1/2} \star S_i = E_i(P)$. Then $(f \circ \gamma)(t) = \mathrm{tr}\Big\{ \log^2\big[Q^{-1}\big(P^{1/2} \star \exp(tS_i)\big)\big]\Big\}$. We set $\phi(t) = Q^{-1}\big(P^{1/2} \star \exp(tS_i)\big)$ and apply Equation (4.16),

$$\begin{aligned}
df_P(E_i(P)) = \frac{d}{dt}(f \circ \gamma)(t)\Big|_{t=0} &= \frac{d}{dt}\mathrm{tr}\Big\{ \log^2\big(\phi(t)\big)\Big\}\Big|_{t=0} \\
&= 2\,\mathrm{tr}\Big\{ \log\big[\phi(0)\big]\big[\phi(0)\big]^{-1}\phi'(0)\Big\} \\
&= 2\,\mathrm{tr}\Big\{ \log\big[Q^{-1}P\big]\big[Q^{-1}\big(P^{1/2} \star I_n\big)\big]^{-1} Q^{-1}\big(P^{1/2} \star S_i\big)\Big\} \\
&= 2\,\mathrm{tr}\Big\{ \log\big[P^{-1/2}P^{1/2}Q^{-1}P^{1/2}P^{1/2}\big] P^{-1/2} S_i P^{1/2}\Big\} \\
&= -2\,\mathrm{tr}\Big\{ \log\big(P^{-1/2} \star Q\big) S_i\Big\} = -2\big\langle \log\big(P^{-1/2} \star Q\big), S_i\big\rangle_F
\end{aligned}$$

$$\begin{aligned}\Rightarrow (\nabla f)_P &= \sum_{i=1}^{d} g^{\mathrm{AI}}\big(\nabla f, E_i^{\mathrm{AI}}\big)\, E_i^{\mathrm{AI}}(P) \\ &= -2\sum_{i=1}^{d} \big\langle \log(P^{-1/2}\star Q), S_i\big\rangle_F\, E_i^{\mathrm{AI}}(P) \\ &= -2\,\mathrm{Log}_P^{\mathrm{AI}}(Q).\end{aligned}$$

□

We end this subsection by presenting the calculation of the Christoffel symbols, see Definition 2.11, and the Hessian for the squared Riemannian distance with respect to the basis field $\mathfrak{B}_d^{\mathrm{AI}}$ defined in Equation (4.17). Moreover, we show that the Laplace-Beltrami operator $\Delta_{\mathcal{SP}(n)}$ to the squared Riemannian distance with respect to $\mathfrak{B}_d^{\mathrm{AI}}$ is constant when equipping $\mathcal{SP}(n)$ with the AI metric.

Lemma 4.4 (Christoffel symbols)**.** *Consider the orthonormal basis field $\mathfrak{B}_d^{AI}$, then*

$$(\nabla_{E_i^{AI}} E_j^{AI})_P = -\frac{1}{2}\sum_{k=1}^{d} \Big\langle (S_iS_j + S_jS_i)\,,\, S_k\Big\rangle_F E_k^{AI}(P) \quad \forall 1 \le i,j \le d.$$

Thus, Christoffel symbols of the second kind are constant with respect to $\mathfrak{B}_d^{AI}$,

$$\Gamma_{ij}^k(P) = \Gamma_{ij}^k = -\frac{1}{2}\Big\langle (S_iS_j + S_jS_i)\,,\, S_k\Big\rangle_F \qquad \forall 1 \le i,j \le d.$$

Proof. Applying Equation (4.15) with $X = E_i$ and $Y = E_j$,

$$\begin{aligned}(\nabla_{E_i}E_j)_P &= -\frac{1}{2}\Big\{\big(P^{1/2}\star S_i\big)P^{-1}\big(P^{1/2}\star S_j\big) + \big(P^{1/2}\star S_j\big)P^{-1}\big(P^{1/2}\star S_i\big)\Big\} \\ &= -\frac{1}{2}\,P^{1/2}\star\big(S_iS_j + S_jS_i\big) = -\frac{1}{2}\sum_{k=1}^{d}\Big\langle (S_iS_j + S_jS_i)\,,\, S_k\Big\rangle_F E_k^{\mathrm{AI}}(P)\end{aligned}$$

So, $\Gamma_{ij}^k(P) = -\frac{1}{2}\Big\langle (S_iS_j + S_jS_i)\,,\, S_k\Big\rangle_F$. □

Lemma 4.5. *The Laplace-Beltrami operator $\Delta_{\mathcal{SP}(n)}$ and the Hessian to the squared Riemannian distance with respect to $\mathfrak{B}_d^{AI}$ equal to $2\,I_d$, that is*

$$(Hess f)_P\big(E_i^{AI}, E_j^{AI}\big) = 2\,\delta_{ij} \quad \& \quad \Delta_{\mathcal{SP}(n)} f\big(E_i^{AI}, E_j^{AI}\big) = 2\,\delta_{ij} \qquad \forall 1 \le i,j \le d.$$

where $f(P) = d_{AI}^2(P, Q)$ *and* δ_{ij} *is the Kronecker delta.*

Proof. Definition of the Hessian given in Equation (2.5) gives

$$(\text{Hess} f)_P\big(E_i^{\text{AI}}, E_j^{\text{AI}}\big) = E_i^{\text{AI}}(E_j^{\text{AI}} f)_P - \big(\nabla_{E_i^{\text{AI}}} E_j^{\text{AI}} f\big)_P,$$

while from the proof of Proposition 4.2, we know that for any $1 \leq i \leq d$,

$$(E_i^{\text{AI}} f)_P = -2\big\langle \log\big(P^{-1/2} \star Q\big), S_i\big\rangle_F = -2 g_P^{\text{AI}}\big(E_i^{\text{AI}}(P), \text{Log}_P^{\text{AI}}(Q)\big)$$

Moreover, Theorem 2.14 implies $(E_i^{\text{AI}} f)_P = 2 g_P^{\text{AI}}\big(\nabla_{E_i^{\text{AI}}(P)} \text{Log}_P^{\text{AI}}(Q), \text{Log}_P^{\text{AI}}(Q)\big)$. Since $E_i^{\text{AI}}(P)$, $\text{Log}_P^{\text{AI}}(Q)$ and P are invertible, we can conclude $\nabla_{E_i^{\text{AI}}(P)} \text{Log}_P^{\text{AI}}(Q) = -E_i^{\text{AI}}(P)$. Therefore, using Lemma 4.4 we have

$$\begin{aligned}
\big(\nabla_{E_i^{\text{AI}}} E_j^{\text{AI}} f\big)_P &= 2 g_P^{\text{AI}}\Big(\nabla_{(\nabla_{E_i^{\text{AI}}} E_j^{\text{AI}})_P} \text{Log}_P^{\text{AI}}(Q), \text{Log}_P^{\text{AI}}(Q)\Big) \\
&= 2 g_P^{\text{AI}}\Big(\sum_{k=1}^{d} \nabla_{\Gamma_{ij}^k(P) E_k^{\text{AI}}(P)} \text{Log}_P^{\text{AI}}(Q), \text{Log}_P^{\text{AI}}(Q)\Big) \\
&= 2 g_P^{\text{AI}}\Big(\sum_{k=1}^{d} \Gamma_{ij}^k \, \nabla_{E_k^{\text{AI}}(P)} \text{Log}_P^{\text{AI}}(Q), \text{Log}_P^{\text{AI}}(Q)\Big) \\
&= -2 g_P^{\text{AI}}\Big(\big(\nabla_{E_i^{\text{AI}}} E_j^{\text{AI}}\big)_P, \text{Log}_P^{\text{AI}}(Q)\Big). \\
E_i^{\text{AI}}(E_j^{\text{AI}} f)_P &= -2 \, E_i^{\text{AI}} \, g_P^{\text{AI}}\big(E_j^{\text{AI}}(P), \text{Log}_P^{\text{AI}}(Q)\big) \\
&= -2 \, g_P^{\text{AI}}\Big(\big(\nabla_{E_i^{\text{AI}}} E_j^{\text{AI}}\big)_P, \text{Log}_P^{\text{AI}}(Q)\Big) - 2 \, g_P\Big(E_j^{\text{AI}}(P), \nabla_{E_i^{\text{AI}}(P)} \text{Log}_P^{\text{AI}}(Q)\Big) \\
&= -2 \, g_P^{\text{AI}}\Big(\big(\nabla_{E_i^{\text{AI}}} E_j^{\text{AI}}\big)_P, \text{Log}_P^{\text{AI}}(Q)\Big) + 2 \, g_P\Big(E_j^{\text{AI}}(P), E_i^{\text{AI}}(P)\Big) \\
&= -2 \, g_P^{\text{AI}}\Big(\big(\nabla_{E_i^{\text{AI}}} E_j^{\text{AI}}\big)_P, \text{Log}_P^{\text{AI}}(Q)\Big) + 2 \, \delta_{ij}.
\end{aligned}$$

Thus, $\big(\text{Hess} f\big)_{ij} = 2 \, \delta_{ij}$. Since $\mathfrak{B}_{\text{AI}}^d$ is an orthonormal basis with respect to g^{AI}, thus the matrix form G_{AI} of the metric g^{AI} equals to I_d. Therefore, we also achieve $\Delta_{\mathcal{SP}(n)} f\big(E_i^{\text{AI}}, E_j^{\text{AI}}\big) = 2 \, \delta_{ij}$ □

4.2.3 Importance of Riemannian geometry to $\mathcal{SP}(n)$

We discuss two major reasons that necessitate the use of Riemannian geometry: capability of sampling near the boundaries of $\mathcal{SP}(n)$ and no swelling effects. One may

additionally argue that other properties, such as inversion-invariance and similarity-invariance for the LE and AI metrics, and affine-invariance for the AI metric, may be useful in simplifying calculations for complex algorithms. Additionally, we discuss the phenomenon of anisotropy when using either the LE or AI metrics.

Although the Euclidean metric (i.e. Frobenius inner product) on $\mathcal{SP}(n) \subset \mathcal{GL}(n)$ is simple, it is problematic because non-covariance matrices are only a finite distance away from covariance matrices. As Equation (4.7) and (4.14) indicate, LE and AI metrics do not suffer from this problem since they make use of the matrix logarithm function, so that non-covariance matrices are at infinite distance from any points on $\mathcal{SP}(n)$. Therefore, they avoid the undesirable inequality constraints that are required in the Frobenius induced geometry to ensure positive definiteness and whose number grows quadratically with n. As will become evident in our simulation experiments, this property turns out to be a highly desirable because it facilitates sampling, particularly near the boundary of $\mathcal{SP}(n)$.

The determinant of a covariance matrix measures the dispersion of the data points from a multivariate normal distribution. For the Euclidean metric, the geodesic connecting two fixed points often contains points with a larger determinant than the two fixed points, and the difference can get extremely large whenever these fixed points lie close to the boundary of $\mathcal{SP}(n)$. This problem is referred to as the swelling effect [32, 13, 106]. In many contexts, the swelling effect is described as undesirable because the level of dispersion should remain close to the given information obtained by the observations of covariance matrices [32, 107, 108, 109]. The LE and AI metrics avoid this swelling effect and in fact we demonstrate in Corollary 4.6 that points on the geodesics associated with the LE or the AI metrics have their determinants resulted from the linear interpolation in the domain of logarithm.

Corollary 4.6. *Determinant of points on geodesics associated with either the LE and AI metrics are identical.*

Proof. Fix two points $P_0, P_1 \in \mathcal{SP}(n)$ and consider the geodesics associated with the LE and AI metrics $\gamma^{\text{LE}}(t)$, $\gamma^{\text{AI}}(t)$ respectively such that $\gamma^{\text{LE}}(0) = \gamma^{\text{AI}}(0) = P_0$ and $\gamma^{\text{LE}}(1) = \gamma^{\text{AI}}(1) = P_1$. Using the fact that for any $A \in \mathcal{SP}(n)$, $\det \exp A =$

$\exp \mathrm{tr} A$ and $\log \det A = \mathrm{tr} \log A$, we get for any $t \in [0, 1]$:

$$\begin{aligned}
\det \gamma^{\mathrm{LE}}(t) &= \det \Big(\exp(\log P_0 + t(\log P_1 - \log P_0)) \Big) \\
&= \exp \left(\log \det P_0 + t \log \frac{\det P_1}{\det P_0} \right) = \big(\det P_0 \big)^{1-t} \big(\det P_1 \big)^t \\
\det \gamma^{\mathrm{AI}}(t) &= \det \left(P_0^{1/2} \star (P_0^{-1/2} \star P_1)^t \right) \\
&= \det P_0 \det \left(P_0^{-1/2} P_1 P_0^{-1/2} \right)^t = \big(\det P_0 \big)^{1-t} \big(\det P_1 \big)^t
\end{aligned}$$

□

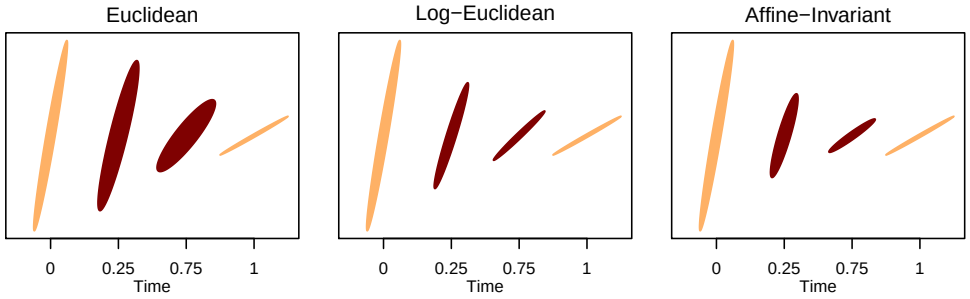


Figure 4.2: Comparison of three metric tensors on $\mathcal{SP}(2)$: Red ellipses represent points at time $t = 0.25, 0.75$ on the geodesic connecting P_0 at $t = 0$ and P_1 at $t = 1$. The length of the axes are the square root of the eigenvalues.

A visual illustration of the swelling effect is provided in Figure 4.2 where two intermediate points on the geodesic connecting

$$P_0 = \begin{pmatrix} 0.4 & 0.3 \\ 0.3 & 0.4 \end{pmatrix} \text{ at } t = 0 \quad \text{and} \quad P_1 = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 0.02 \end{pmatrix} \text{ at } t = 1$$

are shown for each metric. The red ellipses in the case of the Euclidean metric are swelled up noticeably, compared to the other Riemannian metrics: LE and AI metrics. More discussion about this phenomenon can be found in [110, 13].

In addition, the shape (e.g. thin/thick, short/long) of the ellipse tells us about the anisotropy in the covariance matrices, which is understood to occur when the eigenvalues are not identical. If the eigenvalues are equal, the covariance matrix is said to have isotropy. Here, larger anisotropy implies more dependence in some particular directions (i.e. corresponding to the larger eigenvalues), see [110, 13].

In the example, we use two fixed points that have one eigenvalue much larger than the other (i.e. the ellipse is pointy), and Figure 4.2 demonstrates that the LE metric escalates anisotropy. Moreover, the result in [32] says that at a fixed time $t \in [0, 1]$, the trace of points on geodesics associated with the LE metric are always larger or equal to those in the AI case, and since their determinants are identical, the LE metric leads to geodesic with exaggerated anisotropy. In general, whether anisotropy constitutes a problem depends on the application of interest.

Furthermore, as discussed previously in Subsection 4.2.1 and 4.2.2 the LE and the AI metric hold many invariant properties, such as inversion invariance, similarity invariance for the LE metric, and affine invariance for the AI metric, while the Euclidean metric does not achieve any of those. In many complicated computational algorithms, having an invariance property may help to simplify some calculations. Overall, the desirability of the invariance property also depends on the area of the application. For example in diffusion tensor imaging, symmetry with respect to inversion is often required [32, 13, 106], thus the LE and the AI metrics are more suitable. While there are many other Riemannian metrics on either space of symmetry positive semi-definite or symmetry positive definite matrices, such as Cholesky, Procrustes size and shape, etc, explicit formulas for the exponential/logarithm maps when using those metrics are often complicated or unavailable, see [13, 111].

4.3 Brownian motion class

We define the Brownian motion class on $\mathcal{SP}(n)$ by the Riemannian Brownian motion on $\mathcal{SP}(n)$ equipped with different Riemannian metrics, and besides the existing result of the Euclidean metric, in this thesis we consider the two Riemannian metrics: the Log-Euclidean (LE) metric in Subsection 4.2.1 and the Affine-Invariant (AI) metric in Subsection 4.2.2. While on the Euclidean space, the Brownian motion explodes in infinite time, it does not always hold for a general Riemannian manifold. If we achieve this property, the Riemannian manifold obtains stochastic completeness, see Proposition 4.8 for $\mathcal{SP}(n)$ equipped with either the LE or the AI metric.

In order to prove stochastic completeness, we aim to show that Ricci curvatures on $\mathcal{SP}(n)$ equipped with either LE or AI metrics are bounded below, and use Theorem 2.39, i.e. using Corollary 2.40. Firstly, let us state the results of Ricci curvatures in the case of AI metric from [112] and include a concise calculation.

Lemma 4.7. *[112] For any $P \in \mathcal{SP}(n)$, the Ricci curvature is given with respect to the basis $\mathfrak{B}_d^{AI}(P)$, as defined in Equation* (4.17) *:*

$$Ric_P = -\frac{n}{4}\begin{pmatrix} I_n - \frac{1}{n}\mathbb{1}_{n,n} & 0 \\ 0 & I_{n(n-1)/2} \end{pmatrix}.$$

Proof. We apply Equation (4.15) to the Definition 2.12 of curvature tensor and get

$$\begin{aligned} R_P(X,Y)Z &= (\nabla_X\nabla_Y Z)_P - (\nabla_Y\nabla_X Z)_P - (\nabla_{[X,Y]}Z)_P \\ &= \frac{1}{4}\Big(X_P P^{-1}Y_P P^{-1}Z_P + Z_P P^{-1}Y_P P^{-1}X_P \\ &\quad - Y_P P^{-1}X_P P^{-1}Z_P - Z_P P^{-1}X_P P^{-1}Y_P\Big) \end{aligned}$$

Since $1 \le i \le d \ : \ E_i(P) = P^{1/2} \star S_i$, for any $1 \le k \le d$:

$$\Rightarrow R_P(E_k,E_i)E_k = \frac{1}{4}\Big(P^{1/2} \star (2\,S_kS_iS_k - S_iS_kS_k - S_kS_kS_i)\Big).$$

Thus, for any $1 \le i,j \le d \ :$ $\mathrm{Ric}(E_i,E_j)_P = \sum_{k=1}^{d} g_P^{\mathrm{AI}}\big(R_P(E_k,E_i)E_k), E_j(P)\big)$

$$= \frac{1}{4}\sum_{k=1}^{d}\big\langle 2\,S_kS_iS_k - S_iS_kS_k - S_kS_kS_i\,,\,S_j\big\rangle_F.$$

Many terms vanish, and the result is summarised as below :

$$\mathrm{Ric}(E_i,E_j)_P = \begin{cases} -(n-1)/4 & \text{if } 1 \le (i=j) \le n, \\ -n/4 & \text{if } n+1 \le (i=j) \le d, \\ 1/4 & \text{if } 1 \le (i \ne j) \le n, \\ 0 & \text{otherwise.} \end{cases}$$

□

Proposition 4.8 (Stochastic completeness)**.** *The Riemannian manifold $\mathcal{SP}(n)$ is stochastically complete when it is equipped with either*

(i) the LE metric,

(ii) or the AI metric.

Proof. (i) Proposition 4.1 implies $\mathcal{SP}(n)$ equipped with the LE metric has null sectional curvature everywhere, i.e. for all $1 \leq i, j \leq d$ and $P \in \mathcal{SP}(n)$ we have $\nabla_{E_i^{\text{LE}}(P)} E_j^{\text{LE}}(P) = 0$, where $E_i^{\text{LE}}, E_j^{\text{LE}} \in \mathfrak{B}_d^{\text{LE}}$, as defined in Equation (4.8). Thus, the Ricci curvature tensor also vanishes everywhere, and the required result is a direct consequence from Theorem 2.39.

(ii) Let us fix a point $P \in \mathcal{SP}(n)$, and vary some point $Q \in \mathcal{SP}(n)$ such that $d_{\text{AI}}^2(P, Q) = r$ for some $r > 0$. Consider a tangent vector $v \in T_Q\mathcal{SP}(n)$ such that it has unit length, that is $v = \sum_{i=1}^{d} \nu_i E_i^{\text{AI}}(Q) \in T_Q\mathcal{SP}(n)$ and $||\nu||_2 = 1$.

Using Lemma 4.7, and denoting the (i, j)th entry of the Ricci curvature tensor at Q in matrix form as $\text{Ric}_Q^{(i,j)}$, we have

$$\begin{aligned}
\text{Ric}(v, v) &= \sum_{i,j=1}^{d} \nu_i \text{Ric}_Q^{(i,j)} \nu_j \\
&= -\frac{n-1}{4} \sum_{i=j=1}^{n} \nu_i^2 - \frac{n}{4} \sum_{i=j=n+1}^{d} \nu_i^2 + \frac{1}{4} \sum_{i \neq j}^{n} \nu_i \nu_j \\
&= -\frac{n}{4} + \frac{1}{4} \sum_{i,j=1}^{n} \nu_i \nu_j \geq -\frac{n}{4} - \frac{n^2}{4} = -\frac{d}{2}
\end{aligned}$$

The last inequality holds due to the fact that $||\nu||_2 = 1$. Therefore, using Theorem 2.39, we can set $\kappa(r) = -d/\big(2(d-1)\big)$, which is clearly a negative, non-decreasing, continuous function.

$$\Rightarrow \int_c^{\infty} \frac{1}{\sqrt{-\kappa(r)}}\, dr = \sqrt{\frac{2(d-1)}{d}} \int_c^{\infty} 1\, dr = \infty \quad \text{for some constant } c.$$

□

The rest of this section proceeds as follows: firstly establish the local expression of the horizontal lifts of smooth curves on $\mathcal{SP}(n)$ equipped either with the LE or the AI metric; secondly show that any SDEs on $\mathbb{R}^d$ can be extended to $\mathcal{SP}(n)$ equipped with the LE metric in Theorem 4.10; and lastly discuss about the construction of the Riemannian Brownian motion on $\mathcal{SP}(n)$. Particularly, we provide the first order Euler approximation for the horizontal lifts of geodesics when working with the AI metric. But first of all, let us fix the standard basis $\{e_1, \ldots, e_d\}$ for $\mathbb{R}^d$.

Proposition 4.9 (Horizontal lift of smooth curves)**.** *Suppose that $\gamma(t)$ is a smooth curve on $\mathcal{SP}(n)$ with $\gamma(0) = P$ and some smooth curve u_t on the frame bundle $\mathscr{F}\big(\mathcal{SP}(n)\big)$ such that $\pi(u_t) = \gamma(t)$ for all $t > 0$, where $\pi : \mathscr{F}\big(\mathcal{SP}(n)\big) \to \mathcal{SP}(n)$ is the canonical projection map discussed in Subsection 2.3.2.*

(i) (LE metric) If u_t is the unique horizontal lift of $\gamma(t)$ from an initial frame u_0, where $u_0(e_i) = E_i^{LE}(P)$ for all $1 \leq i \leq d$, the expression of u_t in local coordinates with respect to $\big\{E_i^{LE}\big(\gamma(t)\big), e_i\big\}_{i=1}^d$ is $(\gamma(t), \delta)$, where $\delta = \big(\delta_{ij}\big)$, the Kronecker delta.

(ii) (AI metric) Suppose the expression of u_t in local coordinates with respect to $\big\{E_i^{AI}\big(\gamma(t)\big), e_i\big\}_{i=1}^d$ is $u_t = (\gamma(t), \zeta)$ with $\zeta = \big(\zeta_j^i\big)$ and $\zeta_j^i : \mathcal{SP}(n) \to \mathbb{R}$ are differentiable functions, then u_t is the unique horizontal lift of $\gamma(t)$ from an initial frame u_0, where $u_0(e_i) = E_i^{AI}(P)$ for all $1 \leq i \leq d$ if and only if functions ζ_j^i exist uniquely and satisfy for all $1 \leq i, j \leq d$:

$$\sum_{r=1}^{d} \alpha_r\big(\gamma(t)\big) \left\{ \big(E_r^{AI}\zeta_j^i\big)_{\gamma(t)} + \sum_{k=1}^{d} \zeta_k^i\big(\gamma(t)\big)\, \Gamma_{rk}^j\big(\gamma(t)\big) \right\} = 0, \tag{4.18}$$

where the Christoffel symbols $\Gamma_{rk}^j\big(\gamma(t)\big) = \Gamma_{rk}^j$ are given in Lemma 4.4 and smooth functions $\alpha_r : \mathcal{SP}(n) \to \mathbb{R}$ satisfy $\gamma'(t) = \sum_{r=1}^d \alpha_r\big(\gamma(t)\big)\, E_r^{AI}\big(\gamma(t)\big)$. Moreover, if $\alpha_r\big(\gamma(t)\big) \neq 0$ for only one $r \in \{1, \ldots, d\}$, then functions $\zeta_j^i\big(\gamma(t)\big)$ must satisfy

$$\big(E_r^{AI}\zeta_j^i\big)_{\gamma(t)} = -\sum_{k=1}^{d} \zeta_k^i\big(\gamma(t)\big)\, \Gamma_{rk}^j \qquad \text{for all } 1 \leq i, j \leq d.$$

Proof. We are given $\pi(u_t) = \gamma(t)$ with a fixed initial value $\gamma(0) = P$ and an initial frame $u_0 \in \mathscr{F}\big(\mathcal{SP}(n)\big)_P$, thus it is sufficient to show that $u_t(e)$ is parallel along the curve $\gamma(t)$ for any $e \in \mathbb{R}^d$, i.e. $\nabla_{\gamma'(t)} u_t(e) = 0$. Consider an arbitrary $e \in \mathbb{R}^d$, then $e = \sum_{i=1}^d \epsilon_i e_i$ for some $\epsilon_i \in \mathbb{R}$ and we aim to use Definition 2.9 for ∇.

(i) We express $\gamma'(t)$ with respect to the basis $\mathfrak{B}_d^{\text{LE}}$, i.e.

$$\gamma'(t) = \sum_{i=1}^d \alpha_i\big(\gamma(t)\big)\, E_i^{\text{LE}}\big(\gamma(t)\big),$$

where functions $\alpha_i \in C^\infty\big(\mathcal{SP}(n)\big)$. Since $u_t(e) = \sum_{i=1}^d \epsilon_i\, E_i^{\text{LE}}(X_t)$ and ϵ_i does not depend on $\gamma(t)$ for all $1 \le i \le d$,

$$\begin{aligned}\nabla_{\gamma'(t)} u_t(e) &= \nabla_{\left\{\sum_{i=1}^d \alpha_i(\gamma(t))\, E_i^{\text{LE}}\left(\gamma(t)\right)\right\}} \left\{ \sum_{j=1}^d \epsilon_j\, E_j^{\text{LE}}\big(\gamma(t)\big)\right\} \\ &= \sum_{i,j=1}^d \alpha_i\big(\gamma(t)\big)\, \epsilon_j\, \left\{\nabla_{E_i^{\text{LE}}\left(\gamma(t)\right)} E_j^{\text{LE}}\big(\gamma(t)\big)\right\} = 0.\end{aligned}$$

Here we use the result that $\mathcal{SP}(n)$ equipped with the LE metric has null curvature everywhere, i.e. $\nabla_{E_i^{\text{LE}}} E_j^{\text{LE}} = 0$ $(\forall 1 \le i, j \le d)$.

(ii) Firstly, let us suppose that there exist such functions ζ_j^i, which satisfy Equation (4.18). Since $u_t(e) = \sum_{i,j=1}^d \epsilon_i\, \zeta_j^i\big(\gamma(t)\big)\, E_j^{\text{AI}}\big(\gamma(t)\big)$, we use the given condition in Equation (4.18) and get

$$\begin{aligned}\nabla_{\gamma'(t)} u_t(e) &= \sum_{r=1}^d \alpha_r\big(\gamma(t)\big)\, \left(\nabla_{E_r^{\text{AI}}\left(\gamma(t)\right)} \left\{ \sum_{i,j=1}^d e_i\, \zeta_j^i\big(\gamma(t)\big)\, E_j^{\text{AI}}\big(\gamma(t)\big)\right\}\right) \\ &= \sum_{r=1}^d \alpha_r\big(\gamma(t)\big) \Bigg(\sum_{i,j=1}^d e_i \Big\{ \big(E_r^{\text{AI}} \zeta_j^i\big)_{\gamma(t)}\, E_j^{\text{AI}}\big(\gamma(t)\big) \\ &\qquad + \zeta_j^i(X_t)\, \nabla_{E_r^{\text{AI}}\left(\gamma(t)\right)} E_j^{\text{AI}}\big(\gamma(t)\big)\Big\}\Bigg) \\ &= \sum_{r,i,j=1}^d \alpha_r\big(\gamma(t)\big)\, e_i\, \big(E_r^{\text{AI}} \zeta_j^i\big)_{\gamma(t)}\, E_j^{\text{AI}}\big(\gamma(t)\big)\end{aligned}$$

$$+\sum_{r=1}^{d}\alpha_r\big(\gamma(t)\big)\sum_{i,j=1}^{d}e_i\,\zeta_j^i\big(\gamma(t)\big)\left(\sum_{k=1}^{d}\Gamma_{rj}^k\big(\gamma(t)\big)E_k^{\mathrm{AI}}\big(\gamma(t)\big)\right)$$

$$=\sum_{i,j=1}^{d}e_i\,E_j^{\mathrm{AI}}\big(\gamma(t)\big)$$

$$\left(\sum_{r=1}^{d}\alpha_r\big(\gamma(t)\big)\left\{\big(E_r^{\mathrm{AI}}\zeta_j^i\big)_{\gamma(t)}+\sum_{k=1}^{d}\zeta_k^i\big(\gamma(t)\big)\,\Gamma_{rk}^j\right\}\right)=0.$$

On the other hand, if u_t is the horizontal lift of $\gamma(t)$ starting from the initial frame u_0, $\nabla_{\gamma'(t)}u_t(e)=0$. So, for all $1\leq i,j\leq d$:

$$\sum_{r=1}^{d}\alpha_r\big(\gamma(t)\big)\left\{\big(E_r^{\mathrm{AI}}\zeta_j^i\big)_{\gamma(t)}+\sum_{k=1}^{d}\zeta_k^i\big(\gamma(t)\big)\,\Gamma_{rk}^j\big(\gamma(t)\big)\right\}=0.$$

If only one $r\in\{1,\ldots,d\}$ is such that $\alpha_r\big(\gamma(t)\big)\neq 0$, then clearly functions $\zeta_j^i\big(\gamma(t)\big)$ must satisfy

$$\big(E_r^{\mathrm{AI}}\zeta_j^i\big)_{\gamma(t)}=-\sum_{k=1}^{d}\zeta_k^i\big(\gamma(t)\big)\,\Gamma_{rk}^j\qquad\text{for all } 1\leq i,j\leq d.$$

Uniqueness and existence of u_t result from the fact that ζ_j^i are the solution of a system of first order linear ordinary differential equations with initial conditions $\zeta_j^i(P)=\delta_{ij}$.

□

Proposition 4.1 indicates that the LE metric offers the Euclidean structure through the transformation of matrix logarithm. This is consistent with the result in Proposition 4.9, in which the frame u_t does not change its direction while travelling horizontally along the curve $\gamma(t)$. On the other hand, $\mathcal{SP}(n)$ equipped with the AI metric has non-zero curvature and therefore the corresponding frame u_t in this case must adapt, i.e. the functions ζ_j^i are no longer independent of points on $\gamma(t)$. Although in general there are no explicit formulas for ζ_j^i, we will provide instead a first order Euler approximation when assuming $\gamma(t)$ is the geodesic associated with the AI metric toward the end of this section. This is an example where we ap-

proach stochastic development by smooth approximation when using the exponential adapted Euler-Maruyama method.

Before moving to our construction of Riemannian Brownian motion on $\mathcal{SP}(n)$, we present an important result for any SDEs on $\mathcal{SP}(n)$ equipped with the LE metric, where we show an explicit one-to-one transformation between the SDEs on $\mathcal{SP}(n)$ to another SDEs on $\mathbb{R}^d$ using the map $\mathfrak{h}$. Here, we define $\mathfrak{h} = (\mathfrak{h}_j) : \mathcal{SP}(n) \to \mathbb{R}^d$ with an arbitrary $e = \sum_{j=1}^{d} \epsilon_j \, e_j \in \mathbb{R}^d$ and $P \in \mathcal{SP}(n)$ as follows:

$$\mathfrak{h}_j(P) = \langle \log P, S_j \rangle_F \;\; (1 \le j \le d) \quad \text{and} \quad \mathfrak{h}^{-1}(e) = \exp\big(\sum_{j=1}^{d} \epsilon_j \, S_j\big). \tag{4.19}$$

Theorem 4.10. *Suppose the process X_t is the solution of the SDE* (2.17) *on $\mathcal{SP}(n)$ endowed with the LE metric, i.e. for $t \in [0, \tau)$ with a $\mathfrak{F}_*$-stopping time τ:*

$$dX_t = A(t, X_t)\, dt + F_{X_t}\big(b(X_t)\, dB_t\big) \qquad (X_0 = P), \tag{4.20}$$

where A assigns smoothly for each $t \in [0, \tau)$ a smooth vector field $A(t, \cdot)$ on $\mathcal{SP}(n)$ and some smooth function $b : \mathcal{SP}(n) \to \mathbb{R}^{d\times d}$. Moreover, B_t is $\mathbb{R}^d$-valued Brownian motion and the function F is defined in Equation (2.16) *associated with the basis $\mathfrak{B}_d^{LE}$. Then the problem of solving the SDE* (4.20) *on $\mathcal{SP}(n)$ is the same as solving the following SDE on $\mathbb{R}^d$:*

$$dx_t = a(t, x_t)\, dt + \tilde{b}(x_t)\, dB_t \qquad (x_0 = p), \tag{4.21}$$

Here, $p = \mathfrak{h}(P)$, $x_t = \mathfrak{h}(X_t)$ hold for all $t \in [0, \tau)$ and smooth function $\tilde{b}$ is given by $\tilde{b} = b \circ \mathfrak{h}^{-1}$. In addition, smooth function $a = \big(a^{(j)}\big)$ is given by $a^{(j)} : [0, \tau) \times \mathbb{R}^d \to \mathbb{R}$, $(t, x_t) \mapsto \big\langle D_{X_t} \log . A(t, X_t), S_j \big\rangle_F$ for all $1 \le j \le d$.

Proof. As $A(t, \cdot)$ is a smooth vector field on $\mathcal{SP}(n)$ for all $t \in [0, \tau)$, there always exist functions $f^{(j)} \in C^\infty\big([0, \tau) \times \mathcal{SP}(n)\big)$ for all $1 \le j \le d$ such that

$$A(t, X_t) = \sum_{j=1}^{d} f^{(j)}(t, X_t)\, E_j^{\mathrm{LE}}(X_t) = \sum_{j=1}^{d} f^{(j)}(t, X_t) \left(\big(d \log_{X_t}\big)^{-1}(S_j)\right).$$

Then applying the differential map $d\log$ at X_t on both sides, we get

$$d\log_{X_t}\big(A(t,X_t)\big) = \sum_{j=1}^{d} f^{(j)}(t,X_t)\, S_j,$$

which is equivalent to $D_{X_t}\log.A(t,X_t) = \sum_{j=1}^{d} f^{(j)}(X_t)\, S_j$. And since the matrix logarithm function $\log$ is smooth on $\mathcal{SP}(n)$, $D_{X_t}\log.A(t,X_t)$ is smooth with respect to t and X_t. Moreover, we can easily deduce $f^{(j)}(t,X_t) = \big\langle D_{X_t}\log.A(t,X_t), S_j\big\rangle_F$ as $\{S_i\}_{i=1}^d$ are orthonormal. Thus, $a^{(j)}(t,x_t)$ simply equals $f^{(j)}(\mathfrak{h}^{-1}(t,x_t))$, and since $\mathfrak{h}^{-1}$ and f are both smooth, the function a is smooth on $[0,\tau)\times\mathbb{R}^d$.

On the other hand, for any $e = \sum_{j=1}^{d}\epsilon_j\, e_j \in \mathbb{R}^d$, the definition of F in Equation (2.16) implies

$$\begin{aligned} F_{X_t}(e) = \sum_{j=1}^{d}\epsilon_j\, E_j^{\mathrm{LE}}(X_t) &= \big(d\log_{X_t}\big)^{-1}\left(\sum_{j=1}^{d}\epsilon_j\, S_j\right) \\ \iff d\log_{X_t}\big(F_{X_t}(e)\big) &= \sum_{j=1}^{d}\epsilon_j\, S_j. \end{aligned}$$

We know that $\log : \mathcal{SP}(n)\to\mathcal{S}(n)$ is a diffeomorphism, and therefore we can apply the result in [25, Proposition 1.2.4, Page 20], which says if X_t is the solution of the SDE (4.20), the process $\log X_t$ is a solution of the following SDE

$$\begin{aligned} d\log_{X_t}\big(dX_t\big) &= d\log_{X_t}\big(A(t,X_t)\big)\,dt + d\log_{X_t}\Big(F_{X_t}\big\{b(X_t)\,dB_t\big\}\Big) \\ \iff \sum_{j=1}^{d} dx_t^{(j)}\, S_j &= \sum_{j=1}^{d} a^{(j)}(t,x_t)\, S_j\, dt + \sum_{i,j=1}^{d}\tilde{b}_{ij}(x_t)\, S_j\, dB_t^i. \end{aligned}$$

The SDE above is purely in the Itô sense as the second order term in Equation (2.10) of Itô integral vanishes when converting from the corresponding Stratonovich integral. We achieve this because the LE metric offers null curvature everywhere on $\mathcal{SP}(n)$. Lastly, removing the basis $\mathfrak{B}_d = \{S_i\}_{j=1}^d$, we get the desired result. □

The result in Theorem 4.10 brings no surprise because $\mathcal{SP}(n)$ is isomorphic to $\mathcal{S}(n)$, while Proposition 4.9 says the horizontal lift of any semimartingale on $\mathcal{SP}(n)$

equipped with the LE metric is available explicitly and identical to the definition of F in Equation (2.16) associated with the basis $\mathfrak{B}_d^{\mathrm{LE}}$. Consequently, solutions of (4.20) and (4.21) are in one-to-one correspondence. Therefore, when equipping $\mathcal{SP}(n)$ with the LE metric, we only need to study the coefficients of X_t in log-domain with respect to a fixed basis $\mathfrak{B}_d$ instead (i.e. through the function $\mathfrak{h}$), which greatly benefits both theoretical and practical work.

The conditions for existence and uniqueness of the solution for the SDE (4.20) depend directly on the requirements that the drift and diffusivity of the SDE (4.21) satisfy on the Euclidean space, see Theorem 2.37. In particular, the Riemannian Brownian motion associated with the LE metric on $\mathcal{SP}(n)$ has its coefficients in log-domain to equal to an $\mathbb{R}^d$-valued Brownian motion. Similarly, the generalization of OU process can be done in the same way, and we discuss this in more detail later in Section 4.4. For the time being, let us look into the construction of Riemannian Bownian motion associated with the AI metric, which requires an approximation approach and convergence theory.

Firstly, since we have a global basis field $\mathfrak{B}_d^{\mathrm{AI}}$ when dealing with the AI metric on $\mathcal{SP}(n)$, see Proposition 4.3, it is straightforward to use the exponential adapted Euler-Maruyama method when simulating the Riemannian Brownian motion X_t. We substitute $a(t, X_t) = 0$ and $b(X_t) = I_d$ to Equation (2.19) and use the explicit formula of the exponential map associated with the AI metric in Equation (4.12), and get for $\delta_t > 0$:

$$X_{t+\delta_t} = X_t^{1/2} \exp\Big(\sum_{i=1}^{d} \big(B_{t+\delta_t}^{(i)} - B_t^{(i)}\big) S_i \Big) X_t^{1/2},$$

where $B_t = \big(B_t^{(i)}\big)$ is an $\mathbb{R}^d$-valued Brownian motion. As $\delta t \to 0$, X_t converges to the Riemannian Brownian motion in distribution [28]. We achieve this because $\mathcal{SP}(n)$ equipped with the AI metric is parallelizable.

In addition, this approximation method indicates for an infinitesimal time δ_t, $X_{t+\delta_t}$ lies on the geodesic starting from X_t, and we therefore can assume the path connecting X_t and $X_{t+\delta_t}$ is smooth. We highlight that the process X_t is assumed

to be the solution of the SDE (2.20), which can be more complicated than the Riemannian Brownian motion. We firstly approximate function ζ in Proposition 4.9 given that $\gamma(t)$ are geodesics, and aim to use the information that the path between X_t and $X_{t+\delta_t}$ is also a geodesic in providing an approximation of horizontal lifts of the stochastic process X_t. We emphasize this approach of smooth approximation for stochastic development in our work requires the stochastic process X_t to be the limiting result of the exponential adapted Euler-Maruyama method, i.e. X_t belongs to the special class of SDEs discussed at the end of Subsection 2.3.2.

Corollary 4.11 (Approximation of functions ζ)**.** *Consider the geodesic $\gamma(t)$ on $\mathcal{SP}(n)$ equipped with the AI metric with $\gamma(0) = P$. We approximate $\zeta_j^i\big(\gamma(t)\big)$ defined in Proposition 4.9 when $t > 0$ is infinitesimally small under the special scenarios for the initial tangent vector $\gamma'(0) \in T_P\mathcal{SP}(n)$ as follows:*

1. *Choose an integer $l \in \{1, , \ldots, d\}$ and set $\gamma'(0) = P^{1/2} \star S_l$, we therefore get $\gamma(t) = P^{1/2} \star \exp(tS_l)$. Suppose u_t is the unique horizontal lift of $\gamma(t)$ defined in Proposition 4.9. For all $1 \leq r \leq d$ we have*

$$\Rightarrow \big(E_r^{AI}\zeta_j^i\big)_P = \big(d\zeta_j^i\big)_P\big(E_r^{AI}(P)\big) = \frac{d}{dt}(\zeta_j^i \circ \gamma)(t)\bigg|_{t=0}.$$

   Since $\alpha_r(P) = \delta_{rl}$ and $\zeta_i^j(P) = \delta_{ij}$ (i.e. using information of the initial frame), Proposition 4.9 implies

$$\big(E_l^{AI}\zeta_j^i\big)_P = -\sum_{k=1}^{d} \zeta_k^i(P)\,\Gamma_{lk}^j = -\sum_{k=1}^{d} \delta_{ik}\Gamma_{lk}^j = -\Gamma_{li}^j \quad (\forall 1 \leq i, j \leq d)$$
$$\Rightarrow \frac{d}{dt}(\zeta_j^i \circ \gamma)(t)\bigg|_{t=0} = -\Gamma_{li}^j \quad \Rightarrow \lim_{t\to 0}\frac{(\zeta_j^i \circ \gamma)(t) - \delta_{ij}}{t} = -\Gamma_{li}^j,$$

   as $(\zeta_j^i \circ \gamma)(0) = \delta_{ij}$, so $\zeta_j^i\big(\gamma(t)\big) \approx \delta_{ij} - t\Gamma_{li}^j$, given that t is close to zero. Thus, for all $1 \leq i, j \leq d$: $\zeta_j^i\big(P^{1/2} \star \exp(t\,S_l)\big) \approx \delta_{ij} - t\Gamma_{li}^j$ $(0 \leq t \ll 1)$.

2. *We extend case 1 by considering an arbitrary $e = \sum_{l=1}^{d} \alpha_l\, e_l \in \mathbb{R}^d$ such that $\alpha_l \in \mathbb{R} \setminus \{0\}$ for $l \in \mathfrak{I} \subseteq \{1, \ldots, d\}$ and set $\gamma'(0) = P^{1/2} \star \big(u_0(e)\big)$, i.e. $\gamma(t) = Exp_P^{AI}\big(t\,u_0(e)\big)$ with $\alpha_r(X_0)$ in Proposition 4.9 equals to $\alpha_r \neq 0$ for*

$r \in \mathfrak{I}$. *Therefore, Proposition 4.9 implies*

$$\sum_{r\in\mathfrak{I}} \alpha_r \big(E_r^{AI}\zeta_j^i\big)_P = -\sum_{r\in\mathfrak{I}} \alpha_r \sum_{k=1}^{d} \zeta_k^i(P)\,\Gamma_{rk}^j = -\sum_{r\in\mathfrak{I}} \alpha_r\,\Gamma_{ri}^j.$$

On the other hand, we denote $V = \sum_{l=1}^{d} \alpha_l\, E_l^{AI} = \sum_{r\in\mathfrak{I}} \alpha_r\, E_r^{AI}$ *which is a vector field on* $\mathcal{SP}(n)$*, then*

$$\big(V\zeta_j^i\big)_{\gamma(t)} = \sum_{r\in\mathfrak{I}} \alpha_r \left(E_r^{AI}\zeta_j^i\right)_{\gamma(t)} \text{ and } \big(V\zeta_j^i\big)_P = \frac{d}{dt}(\zeta_j^i\circ\gamma)(t)\bigg|_{t=0}$$

$$\Rightarrow \lim_{t\to 0}\frac{(\zeta_j^i\circ\gamma)(t)-\delta_{ij}}{t} = \sum_{r\in\mathfrak{I}} \alpha_r\,(E_r^{AI}\zeta_j^i)_P = -\sum_{r\in\mathfrak{I}} \alpha_r\,\Gamma_{ri}^j.$$

Thus, for all $1 \le i, j \le d$*:*

$$\zeta_j^i\left(P^{1/2}\exp(t\sum_{l=1}^{d}\alpha_l\,S_l)P^{1/2}\right) \approx \delta_{ij} - t\sum_{l=1}^{d}\alpha_l\,\Gamma_{li}^j\,\mathbb{I}_{\{l\in\mathfrak{I}\}},$$

for infinitesimally small $t > 0$*, where* $\mathbb{I}$ *stands for the indicator function.*

From Equation (2.19), we substitute $P = X_t$ and get $\alpha_r = a^{(r)}(t, X_t)\delta_t + \sum_{i=1}^{d} b_{ir}(X_t)\big(B_{t+\delta_t}^{(i)} - B_t^{(i)}\big)$ as given in case 2 of Corollary 4.11, where $a(t, X_t) = F_{X_t}^{-1}\big(A(t, X_t)\big)$, where function F is defined in Equation (2.16) associated with the basis $\mathfrak{B}_d^{\mathrm{AI}}$. Since the difference $B_{t+\delta_t}^{(i)} - B_t^{(i)} \sim \mathscr{N}\big(0, \delta_t\big)$, by assuming that functions $a(t, X_t)$ and $b(X_t)$ are locally bounded (e.g. locally Lipschitz continuous), we can contract the approximation for ζ_j^i in Corollary 4.11 to δ_{ij} given that δ_t is sufficiently small. The conditions for functions a, b clearly hold for the case of Riemannian Brownian motion since $a(t, X_t) = 0$ and $b(X_t) = I_d$. This approximation turns out to be identical to the local expression of the horizontal lift for all smooth curves on $\mathcal{SP}(n)$ equipped with the LE metric, see Proposition 4.9.

To sum up this section, we show that the Riemannian Brownian motion associated with either LE or AI metrics explode in infinite time. In particular, we can preserve many properties of SDEs on $\mathbb{R}^d$ when extending them to $\mathcal{SP}(n)$ equipped with the LE metric via the matrix logarithm function. While the Riemannian Brown-

ian motion associated with the LE metric is straightforward to study, it is not the case with the AI metric with requirement of piecewise approximation. The definition of F in Equation (2.16) associated with the basis $\mathfrak{B}_d^{\mathrm{LE}}$ coincides with the horizontal lifts in the LE case, whereas F associated with the basis $\mathfrak{B}_d^{\mathrm{AI}}$ approximates the horizontal lifts in the AI case under some requirements for functions $a,\ b$ as provided in Equation (2.19). In other words, we are able to approximate the true frame, that is glued to the stochastic process and moving horizontally along with this process, in a very simple formula in the AI case. Thus, on top of efficient simulation, we can make theories of stochastic development more practically accessible. We illustrate this with the task of Bayesian parameter estimation in Section 4.6.2 when assuming the SDE of interest is the generalization of the OU process.

4.4 Ornstein-Uhlenbeck class

Adopting the intrinsic point of view, we present the construction of the Ornstein-Ulenbeck (OU) class of processes on $\mathcal{SP}(n)$ equipped with some Riemannian metric. In analogy with the Euclidean OU process, we start with Brownian motion and add a mean-reverting drift which pushes the process toward the point of attraction. In the Euclidean setting, this drift is simply given as the gradient of the squared distance between the process and the point of attraction. We translate this idea to manifolds and define $A(t, X_t)$ by the Riemannian gradient of the squared distance function in the SDE (2.20). This is similar to the treatment of the drift term by V. Staneva et al. [10] for shape manifolds.

We define the Riemannian OU process X_t on $\mathcal{SP}(n)$ to be the solution of the following SDE with model parameters $\theta \in \mathbb{R}_{>0}$, $M \in \mathcal{SP}(n)$ and $\sigma \in \mathbb{R}_{>0}$:

$$dX_t = -\frac{\theta}{2}\left(\nabla d^2(X_t, M)\right)_{X_t} dt + F_{X_t}(\sigma\, dB_t) \qquad (X_0 = U), \tag{4.22}$$

where F is defined in Equation (2.16) associated with some global basis field, i.e. $F_Q(e) = \sum_{i=1}^d \epsilon_i\, E_i(Q)$ with $e = \sum_{i=1}^d \epsilon_i\, e_i \in \mathbb{R}^d$, $Q \in \mathcal{SP}(n)$, where $\{E_i\}_{i=1}^d$ the orthonormal basis field with respect to the given metric tensor, e.g. $\mathfrak{B}_d^{\mathrm{AI}}$ or $\mathfrak{B}_d^{\mathrm{LE}}$. Simulation for X_t can be easily performed with the ex-

ponential adapted Euler-Maruyama method, see Equation (2.19). In this case, $a^{(j)}(t, X_t) = -g_{X_t}\Big(\theta\big(\nabla d^2(X_t, M)\big)_{X_t}, E_j(X_t)\Big)/2$ for $1 \leq j \leq d$ and $b(X_t) = \sigma$. So, Equation (2.19) can be rewritten as

$$X_{t+\delta_t} = \mathrm{Exp}_{X_t}\left\{ -\frac{\theta}{2}\nabla_{X_t}\{d^2(X_t, M)\}\,\delta_t + \sum_{j=1}^{d}(B^{(j)}_{t+\delta_t} - B^{(j)}_t)\,\sigma\,E_j(X_t)\right\}, \quad (4.23)$$

for $\delta_t > 0$. Here, the exponential map Exp, the squared distance function d^2 and the orthonormal basis frame $\{E_j\}_{j=1}^d$ depend on the chosen metric tensor, i.e. either the LE or AI metrics. Moreover, Proposition 4.2 and 4.3 imply

- LE metric: $a^{(j)}(t, X_t) = \theta\big\langle \log M - \log X_t, S_j\big\rangle_F$.
- AI metric: $a^{(j)}(t, X_t) = \theta\big\langle \log(X_t^{-1/2} \star M), S_j\big\rangle_F$.

Proposition 4.2 and 4.3 demonstrates that the Riemannian gradient chosen for the drift, $\nabla d^2(X_t, M)$, in the SDE (4.22) is explicitly computable, while the definition of E_i^{LE}, E_i^{AI} suggest acceptable computational cost compared to when using the Euclidean metric. In addition, Theorem 4.10 indicates that one can simply study the usual OU process on $\mathbb{R}^d$ and transform the whole process via a global function $\mathfrak{h}$ defined in Equation (4.19), to get the Riemannian OU process associated with the LE metric, so the computational cost is significantly scaled down. Moreover, most favourable properties that the OU process has on the Euclidean space will carry over to $\mathcal{SP}(n)$ equipped with the LE metric, such as existence and uniqueness of the solution and ergodicity. And the transition probability density of the OU process associated with the LE metric is explicitly available up to the Jacobian term involving the derivative of the matrix exponential.

We conclude this section by establishing equivalent results for the AI case in a non-constructive manner. While there is no simple diffeomorphism corresponding to $\mathfrak{h}$, otherwise equivalent results can be obtained for the AI case.

Proposition 4.12. *The existence and uniqueness theorem in [26, Theorem 2E, Page 121] is applicable to the Riemannian OU process associated with the AI metric. And, this diffusion process is also non-explosive, see [26, Corollary 6.1, Page 131].*

4.5 Bayesian parameter estimation

We now focus on the Bayesian estimation for the model parameters in the OU difussion processes on $\mathcal{SP}(n)$ when observations are collected at low frequency. Benefiting from Theorem 4.10, we are able to adopt the MCMC strategy introduced by G. O. Roberts & O. Stramer [43] which requires data imputation through sampling from a diffusion bridge in the case of the LE metric, see Appendix B.2. On the other hand, when working with the AI metric we require a diffusion bridge sampling method that operates on $\mathcal{SP}(n)$ which, unlike the Euclidean case, has not been studied before. So, in the remainder of this section we focus only on the AI metric where no such result exists.

A common approach used in manifolds is to use embedding or local charts followed by an appropriate Euclidean method; see for example, [11, 42, 10], but this strategy is unsuitable when transitioning between charts is required and charts can be cumbersome to work with. We therefore develop a diffusion bridge sampler exploiting the exponential map and adopting an intrinsic viewpoint.

For low frequency data, the transition density, and hence the likelihood, are generally not explicitly available. Hence, one common strategy is to augment the data by filling in a number of data points in between observations and relying on an approximation of the diffusion for short time intervals. This data augmentation problem has appeared in the literature for at least two decades, it is therefore difficult to cover the whole research field of this problem, we instead focus on some key papers. To deal with the data augmentation problem it is either assumed that X_t has constant diffusivity, or that X_t is transformed to a process of constant diffusivity, or existence of a process that is absolutely continuous to X_t and its corresponding transition probability need to be derived. MCMC data augmentation sampling strategies that impute partial trajectories via bridge samplers have been used to numerically approximate the transition densities, see [113, 114, 43]. The use of bridge sampling has a long history in the inference for diffusions starting from A. R. Pedersen [115]. Recent advances include the modified diffusion bridge by G. B. Durham & A. R. Gallant [116] and its modifications, see [117, 118, 119] and ideas

based on sequential Monte Carlo [120, 121]. There has been a line of research based on ideas of B. Delyon & Y. Hu [38] that uses guided and residual proposal densities, see [39, 40, 41]. Finally, a recent promising approach is based on M. Bladt & M. Sorensen [122], which provides a single proposal path (sometimes is referred to as a "confluent diffusion bridge") for the target diffusion bridge by conditioning forward and backward diffusion processes to cross paths, see [123, 124, 125].

In the case of the AI metric, at first glance the SDE seems to have constant diffusivity. However, due to the presence of curvature, the diffusivity does depend on the position of X_t, hence straightforward algorithms from literature are not applicable. For instance, looking at the local coordinate expression of the SDE of the Riemannian Brownian motion W_t on $\mathcal{SP}(n)$ equipped with the AI metric (substituting G_{AI}^{-1} in Equation (4.11) and the Christoffel symbols given in Lemma 4.4 into Equation (2.13)), the dependence of the diffusivity on W_t is evident as is the complexity of the resulting expressions. Furthermore, attempting to sample from this SDE using the standard Euclidean method will lead to symmetric matrices but will fail to preserve positive definiteness.

On the Euclidean space, B. Delyon & Y. Hu [38] and M. Schauer & F. V. D. Meulen et al. [40] suggest adding an extra drift term which guides the SDE solution toward the correct terminal point and leaves its law absolutely continuous with respect to the law of the original conditional diffusion process; the Radon-Nikodym derivative is available explicitly. This results in an easier simulation, much better MCMC mixing rate of convergence and no difficulty of computing acceptance probability when updating the proposal bridge. The additional drift is the gradient of the logarithm of the transition density of an auxiliary process which must have explicitly available transition probability density. We include a short summary about guided proposals on the Euclidean space in the Appendix, see Section B.1. In the manifold setting, we are aware of one attempt to use the above approach to sample diffusion bridges through local coordinates by S. Sommer et al. [42], but using the exponential map in this context is new. Motivated by these ideas, we construct a methodology that allows us to sample a diffusion bridge on $\mathcal{SP}(n)$

equipped with the AI metric using a guided proposal process.

We need to choose a proposal diffusion process, which has both an explicit transition probability density and an analytically tractable gradient of the log transition probability density due to the requirement in obtaining the additional drift for the SDE of the guided proposal. Moreover, the target diffusion bridges that we are unable to sample from and the proposal diffusion process which are much easier to deal with, are required to pull to the terminal point with the same force because it is a mandatory condition for their laws to be equivalent. On the Euclidean space, M. Schauer & F. V. D. Meulen et al. [40] shows that if the proposal diffusion process is a linear process, its diffusion coefficient needs to coincide with the value of the original process at the terminal time, in order to ensure the speed at which the proposal diffusion process tends to the correct endpoints is appropriate with respect to the target (unknown) diffusion bridge.

Unlike the Euclidean space, where many probability distributions have been studied for centuries, there are only few probability distributions on $\mathcal{SP}(n)$. One common distribution on $\mathcal{SP}(n)$ is the Wishart distribution, and the Wishart process on $\mathcal{SP}(n)$ is an extension of the Cox-Ingersoll-Ross process to the space of covariance matrices. The transition probability density of the Wishart process is given in terms of the non-central Wishart distribution $W_n(p, \Sigma, \Theta)$ [22],

$$f(X) = \frac{1}{2^{pn/2}\,\mathrm{Gam}_n\left(\frac{p}{2}\right)\,\sqrt[n]{\det\Sigma}}\,\exp\left\{\mathrm{tr}\left(-\frac{1}{2}\left(\Theta + \Sigma^{-1}X\right)\right)\right\} \\ (\det X)^{(p-n-1)/2}\,{}_0F_1\left(\frac{p}{2}; \frac{1}{4}\,\Theta\,\Sigma^{-1}X\right), \tag{4.24}$$

where $X, \Sigma \in \mathcal{SP}(n)$ and $\Theta \in \mathbb{R}^{n\times n}$, Gam_n is the multivariate Gamma function [126] and ${}_0F_1$ is the hypergeometric function [127]. However, due to the presence of hypergeometric function, there is no closed form for the Riemannian gradient of this transition probability density. If one insists on using this distribution, a numerical approximation method is required to compute this gradient. Alternatively, we choose a diffusion process $\tilde{X}_s$ on $\mathcal{SP}(n)$, whose transition probability is given by the Riemannian Gaussian distribution in [128], that is

$$p(t, \tilde{X}_t; s, \tilde{X}_s) = \frac{1}{K_n(\sigma)} \exp\left(-\frac{\mathrm{d}^2_{\mathrm{AI}}(\tilde{X}_s, \tilde{X}_t)}{2\,(t-s)\,\sigma^2}\right) \qquad \text{for } s > t, \text{ and} \tag{4.25}$$

$$K_n(\sigma) = \frac{\pi^{\frac{n^2}{2}}\, 8^{\frac{n(n-1)}{4}}}{n!\,\mathrm{Gam}_n(\frac{n}{2})} \int_{\mathbb{R}^n} \exp\left(-\frac{\sum_{i=1}^n r_i^2}{2(t-s)\sigma^2}\right) \prod_{i<j} \sinh\frac{|r_i - r_j|}{2} \prod_{i=1}^n dr_i. \tag{4.26}$$

This transition density is symmetric and we denote $\tilde{X}_t|\tilde{X}_s \sim \mathfrak{N}\big(\tilde{X}_s, (s-t)\sigma^2\big)$. If the time difference $s-t$ is fixed, the normalising constant $K_n(\sigma)$ depends only on σ and n. Moreover, this auxiliary process $\tilde{X}_t$ exists but its SDE form is not explicitly available, see [29]. We call $\tilde{X}_t$ the *proposal process* and emphasize that explicit availability of the proposal SDE is not actually a pre-requisite for the guided proposals algorithm. Furthermore, the model parameter σ appears in both the SDE (4.22) and the Equation (4.25) because we want to duplicate the requirement for which the guided proposal process and the target diffusion bridge are absolutely continuous on the Euclidean space, see Appendix B.1. Although the requirement in the Euclidean setting does not directly imply that the absolute continuity will hold on a general manifold, it suggests what the properties of the process $\tilde{X}_t$ should be.

Firstly, the OU process X_t on $\mathcal{SP}(n)$ equipped with the AI metric defined in Section 4.4, is the solution of the following SDE with its law $\mathbb{P}_t$,

$$dX_t = \theta\,\mathrm{Log}^{\mathrm{AI}}_{X_t} M\ dt + F_{X_t}(\sigma\, dB_t) \qquad (X_0 = U). \tag{4.27}$$

and we assume that sampling from the target diffusion bridge $X_t^* = \{X_t, 0 \le t \le T \,|\, X_0 = U, X_T = V\}$ with its corresponding law $\mathbb{P}_t^*$ is required. Subsequently, applying Proposition 4.3 to Equation (4.25), we get the guided drift $\sigma^{-2}\big(\nabla \log p(t, X_t^\diamond; T, V)\big)_{X_t^\diamond} = \big(\mathrm{Log}^{\mathrm{AI}}_{X_t^\diamond} V\big)/(T-t)$. We therefore introduce the *guided proposal* $X_t^\diamond$, which is the solution of the following SDE with its law $\mathbb{P}_t^\diamond$:

$$dX_t^\diamond = \left(\theta\,\mathrm{Log}^{\mathrm{AI}}_{X_t^\diamond} M + \frac{\mathrm{Log}^{\mathrm{AI}}_{X_t^\diamond} V}{T-t}\right) dt + F_{X_t^\diamond}(\sigma\, dB_t) \qquad (X_0^\diamond = U). \tag{4.28}$$

Lastly, we aim to show that the $\mathbb{P}_t^*$ and $\mathbb{P}_t^\diamond$ are equivalent for $t \in [0,T]$ and the Radon-Nikodym derivatives are computable. In essence, we prove that $\mathbb{P}_t^*$ and $\mathbb{P}_t^\diamond$ are equivalent up to time T with the aid of Lemma 4.4 in Theorem 4.13 and compute their corresponding Radon-Nikodym derivatives. We then take limit of $t \to T$, and claim that the measure $\mathbb{P}_t^\diamond$ is absolutely continuous with respect to the measure $\mathbb{P}_t^*$, and obtain the Radon-Nikodym derivative in Theorem 4.15.

Let $p(t, X_t; T, V)$ be the true (unknown) transition density of moving from X_t at time t to V at time T and let $X_{[0:t]}^\diamond$ be the path of $X_t^\diamond$ from time 0 to t.

Theorem 4.13. *For $t \in [0,T)$, the laws $\mathbb{P}_t$, $\mathbb{P}_t^\diamond$ and $\mathbb{P}_t^*$ are absolutely continuous,*

$$\frac{d\mathbb{P}_t}{d\mathbb{P}_t^\diamond}\big(X_{[0:t]}^\diamond\big) = \frac{\exp\big(f(X_0^\diamond)\big)}{\exp\big(f(X_t^\diamond)\big)} \exp\Big\{\Phi(t, X_{[0:t]}^\diamond) + \phi\big(t, X_{[0:t]}^\diamond\big)\Big\}, \tag{4.29}$$

$$\frac{d\mathbb{P}_t^*}{d\mathbb{P}_t^\diamond}\big(X_{[0:t]}^\diamond\big) = \frac{p(t, X_t^\diamond; T, V)}{\exp\big(f(X_t^\diamond)\big)} \frac{\exp\big(f(X_0^\diamond)\big)}{p(0, U; T, V)} \exp\Big\{\Phi\big(t, X_{[0:t]}^\diamond\big) + \phi\big(t, X_{[0:t]}^\diamond\big)\Big\}, \tag{4.30}$$

where the functions f, ϕ and Φ are defined by

$$f(X_t^\diamond) = -\frac{d_{AI}^2(X_t^\diamond, V)}{2\sigma^2(T-t)} = -\frac{\big|\big|\log\big((X_t^\diamond)^{-1/2} V (X_t^\diamond)^{-1/2}\big)\big|\big|_F^2}{2\sigma^2(T-t)}, \tag{4.31}$$

$$\phi\big(t, X_{[0:t]}^\diamond\big) = \sum_{i,j=1}^{d} \int_0^t \frac{\big(\zeta_j^i(X_s^\diamond)\big)^2}{2(T-s)}\, ds, \tag{4.32}$$

$$\begin{aligned}\Phi\big(t, X_{[0:t]}^\diamond\big) = \int_0^t \Bigg(& \frac{\theta\, g^{AI}_{X_s}\big(Log^{AI}_{X_s^\diamond} M, Log^{AI}_{X_s^\diamond} V\big)}{\sigma^2(T-s)} \\ & + \sum_{i,j,r=1}^{d} \frac{g^{AI}_{X_s^\diamond}\Big(\zeta_j^i(X_s^\diamond)\big(\sum_{l=1}^d \zeta_l^i(X_s^\diamond)\Gamma_{jl}^r + (E_j^{AI}\zeta_r^i)_{X_s^\diamond}\big) E_r^{AI}(X_s^\diamond)\,,\, Log^{AI}_{X_s^\diamond} V\Big)}{2(T-s)} \Bigg) ds,\end{aligned} \tag{4.33}$$

with Γ_{jl}^r given in Lemma 4.4. The functions $\zeta = (\zeta_j^i)$ are the coefficients with respect to the basis $\mathfrak{B}_d^{AI}$ in the local expression of the horizontal lift, see Proposition 4.9.

Proof. Using Theorem 2.42 (Girsanov-Cameron-Martin theorem), the measures $\mathbb{P}_t$

and $\mathbb{P}_t^\diamond$ are absolutely continuous, and the Radon-Nikodym derivative is given by

$$\frac{d\mathbb{P}_t}{d\mathbb{P}_t^\diamond}\big(X_{[0:t]}^\diamond\big) = \exp\Bigg\{ -\int_0^t \frac{g_{X_s^\diamond}^{\mathrm{AI}}\Big(\mathrm{Log}_{X_s^\diamond}^{\mathrm{AI}} V\,,\, U_s^\diamond(dB_s)\Big)}{\sigma(T-s)} - \frac{1}{2}\int_0^t \frac{\big|\big|\mathrm{Log}_{X_s^\diamond}^{\mathrm{AI}}\big|\big|^2_{g^{\mathrm{AI}}_{X_s^\diamond}}}{\sigma^2\,(T-s)^2}\,ds\Bigg\} \tag{4.34}$$

where $U_t^\diamond$ is the horizontal lift of the guided proposal process $X_t^\diamond$.

Moreover, Proposition 4.3 and Lemma 4.5 imply the following results:

$$\begin{aligned}
&(\nabla f)_{X_t^\diamond} = \frac{\mathrm{Log}_{X_t^\diamond}^{\mathrm{AI}} V}{\sigma^2(T-t)},\\
&\frac{\partial f}{\partial t} = -\frac{d^2_{\mathrm{AI}}(V, X_t^\diamond)}{2\sigma^2(T-t)^2} = -\frac{\big|\big|\mathrm{Log}_{X_t^\diamond}^{\mathrm{AI}}\big|\big|^2_{g^{\mathrm{AI}}_{X_t^\diamond}}}{2\sigma^2(T-t)^2},\\
&\Delta_{\mathcal{SP}(n)} f\big(\,E_i^{\mathrm{AI}}(X_t^\diamond)\,,\,E_j^{\mathrm{AI}}(X_t^\diamond)\,\big) = \frac{\delta_{ij}}{\sigma^2(T-t)}.
\end{aligned}$$

Since $U_t^\diamond(e_i) = \sum_{j=1}^d \zeta_j^i(X_t^\diamond)\, E_j^{\mathrm{AI}}(X_t^\diamond)$, we get

$$\Rightarrow \nabla_{U_t^\diamond(e_i)} U_t^\diamond(e_k) = \sum_{j,l}^d \zeta_j^i(X_t^\diamond)\Big(\sum_{r=1}^d \zeta_l^k(X_t^\diamond)\Gamma_{jl}^r E_r^{\mathrm{AI}}(X_t^\diamond) + (E_j^{\mathrm{AI}}\zeta_l^k)_{X_t^\diamond} E_l^{\mathrm{AI}}(X_t^\diamond)\Big).$$

We then apply Theorem 2.41 (Global formulations of Itô's lemma) to the smooth function f, and get

$$\begin{aligned}
f(X_t^\diamond) - f(X_0^\diamond) &= \int_0^t \left(\frac{\partial f}{\partial s} + g_{X_s}^{\mathrm{AI}}\left(\theta\,\mathrm{Log}_{X_s^\diamond}^{\mathrm{AI}} M + \frac{\mathrm{Log}_{X_s^\diamond}^{\mathrm{AI}} V}{T-s}, \nabla f\right)\right) ds\\
&+ \int_0^t g_{X_s}^{\mathrm{AI}}\big(U_s^\diamond(\sigma\,dB_s)\,,\,\nabla f\big) + \frac{1}{2}\int_0^t g_{X_s}^{\mathrm{AI}}\left(\nabla_{U_s^\diamond(\sigma\,dB_s)} U_s^\diamond(\sigma\,dB_s)\,,\,\nabla f\right)\\
&+ \frac{1}{2}\int_0^t \Delta_{\mathcal{SP}(n)} f\big(U_s^\diamond(\sigma\,dB_s)\,,\,U_s^\diamond(\sigma\,dB_s)\big)\\
&= -\frac{1}{2}\int_0^t \frac{\big|\big|\mathrm{Log}_{X_s^\diamond}^{\mathrm{AI}}\big|\big|^2_{g^{\mathrm{AI}}_{X_s^\diamond}}}{\sigma^2(T-s)^2}\,ds + \int_0^t \frac{\theta\, g_{X_s}^{\mathrm{AI}}\Big(\mathrm{Log}_{X_s^\diamond}^{\mathrm{AI}} M, \mathrm{Log}_{X_s^\diamond}^{\mathrm{AI}} V\Big)}{\sigma^2(T-s)}\,ds
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^t \frac{\left|\left|\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}\right|\right|^2_{g^{\mathrm{AI}}_{X^\diamond_s}}}{\sigma^2(T-s)^2}\,ds+\int_0^t \frac{g^{\mathrm{AI}}_{X^\diamond_s}\Big(\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}V\,,\,U^\diamond_s(\sigma\,dB_s)\Big)}{\sigma^2\,(T-s)}\\
&+\frac{1}{2\sigma^2\,(T-s)}\int_0^t\sum_{i,k=1}^{d} d\big[B^i,B^k\big]_s\left\{\sum_{j,l=1}^{d}\delta_{jl}\,\zeta^i_j(X^\diamond_s)\,\zeta^k_l(X^\diamond_s)\,\sigma^2\right.\\
&\left.+\sum_{j,r=1}^{d} g^{\mathrm{AI}}_{X^\diamond_s}\left(\sigma^2\,\zeta^i_j(X^\diamond_s)\left(\sum_{l=1}^{d}\zeta^k_l(X^\diamond_s)\Gamma^l_{jr}+(E^{\mathrm{AI}}_j\zeta^k_r)_{X^\diamond_s}\right)E^{\mathrm{AI}}_l(X^\diamond_s)\,,\,\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}V\right)\right\}\\
&=\frac{1}{2}\int_0^t \frac{\left|\left|\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}\right|\right|^2_{g^{\mathrm{AI}}_{X^\diamond_s}}}{\sigma^2(T-s)^2}\,ds+\int_0^t \frac{g^{\mathrm{AI}}_{X^\diamond_s}\Big(\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}V\,,\,U^\diamond_s(dB_s)\Big)}{\sigma(T-s)}\\
&+\int_0^t \theta\,g^{\mathrm{AI}}_{X_s}\left(\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}M,\frac{\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}V}{\sigma^2(T-s)}\right)ds+\frac{1}{2}\sum_{i,j,r=1}^{d}\int_0^t\zeta^i_j(X^\diamond_s)\\
&\qquad g_{X^\diamond_s}\left(\Big(\sum_{l=1}^{d}\zeta^i_l(X^\diamond_s)\Gamma^r_{jl}+(E^{\mathrm{AI}}_j\zeta^i_r)_{X^\diamond_s}\Big)E^{\mathrm{AI}}_r(X^\diamond_s)\,,\,\frac{\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}V}{T-s}\right)ds\\
&+\sum_{i=1}^{d}\int_0^t\frac{\sum_{j=1}^{d}\big(\zeta^i_j(X^\diamond_s)\big)^2}{2(T-s)}\,ds.
\end{aligned}
$$

Substituting into Equation (4.34), we get

$$
\begin{aligned}
&\frac{d\mathbb{P}_t}{d\mathbb{P}^\diamond_t}\big(X^\diamond_{[0:t]}\big)=\exp\left[-\big(f(X^\diamond_t)-f(X^\diamond_0)\big)\right.\\
&+\int_0^t\frac{ds}{2\sigma^2(T-s)}\left\{\theta\,g^{\mathrm{AI}}_{X_s}\Big(\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}M,\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}V\Big)+\sigma^2\sum_{i,j=1}^{d}\big(\zeta^i_j(X^\diamond_s)\big)^2\right.\\
&\left.\left.+\sigma^2\sum_{i,j,r=1}^{d}g^{\mathrm{AI}}_{X^\diamond_s}\Big(\zeta^i_j(X^\diamond_s)\Big(\sum_{l=1}^{d}\zeta^i_l(X^\diamond_s)\Gamma^r_{jl}+(E^{\mathrm{AI}}_j\zeta^i_r)_{X^\diamond_s}\Big)E^{\mathrm{AI}}_r(X^\diamond_s)\,,\,\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s}V\Big)\right\}\right]\\
&=\frac{\exp\big(f(X^\diamond_0)\big)}{\exp\big(f(X^\diamond_t)\big)}\exp\Big\{\Phi\big(t,X^\diamond_{[0:t]}\big)+\phi(t,X^\diamond_{[0:t]})\Big\}.
\end{aligned}
$$

Using the result by M. Schauer & F. Van Der Meulen et al. [40], that is

$$
\frac{d\mathbb{P}^*_t}{d\mathbb{P}_t}(X^\diamond_{[0:t]})=\frac{p(t,X^\diamond_t;T,V)}{p(0,U;T,V)},
$$

we can easily get the desired result in Equation (4.30). □

The Radon-Nikodym derivatives given in Equations (4.29)–(4.30) can not be

computed explicitly due to the presence of ζ. However, we can approximate ϕ and Φ in Theorem 4.13 based on the approximation of ζ, see Corollary 4.14, this is a crucial step to make our proposed algorithm practicable.

Corollary 4.14. *Suppose that we have an $\mathcal{SP}(n)$-valued path $\{X^\diamond_{t_k} = y^\diamond_{t_k}\}_{k=0}^{m+1}$ of the OU process X_t when equipping $\mathcal{SP}(n)$ with the AI metric, in which it is simulated from the exponential adapted Euler-Maruyama method, see Equation* (4.23)*, where $\max\{t_{k+1} - t_k\}_{k=0}^m$ is sufficiently small and $0 = t_0 < \ldots < t_{m+1} = t$. Then the functions ϕ in Equation* (4.32) *and Φ in Equation* (4.33) *can be approximated as follows:*

$$\phi(t, X^\diamond_{[0:t]}) \approx \frac{d}{2} \log \frac{T-t}{T}, \tag{4.35}$$

$$\Phi(t, X^\diamond_{[0:t]}) \approx \sum_{k=0}^{m} \frac{t_{k+1} - t_k}{T - t_k} \left\{ \frac{\theta \left\langle \log\left((y^\diamond_{t_k})^{-1/2} \star M\right), \log\left((y^\diamond_{t_k})^{-1/2} \star V\right) \right\rangle_F}{\sigma^2} + \frac{\left\langle \Gamma, \log\left((y^\diamond_{t_k})^{-1/2} \star V\right) \right\rangle_F}{2} \right\}, \tag{4.36}$$

with $\Gamma = \sum_{i,r=1}^d \Gamma^r_{ii} S_r$ and the Christoffel symbols Γ^r_{ii} are given in Lemma 4.4.

Proof. Using the approximation for ζ^i_j in Corollary 4.11, we get for $0 \le k \le m$ and $1 \le i, j \le d$: $\zeta^i_j(X^\diamond_{t_k}) \approx \delta_{ij}$,

$$\Rightarrow E^{\mathrm{AI}}_l \zeta^i_j = 0, \ \forall 1 \le l \le d \qquad \& \qquad U^\diamond_{t_k}(e_i) \approx E^{\mathrm{AI}}_i(X^\diamond_{t_k}).$$

Thus, we get

$$\phi(t, X^\diamond_{[0:t]}) = \sum_{i,j=1}^d \int_0^t \frac{\left(\zeta^i_j(X^\diamond_s)\right)^2}{2(T-s)}\, ds \approx \sum_{i=1}^d \int_0^t \frac{1}{2(T-s)}\, ds = \frac{d}{2} \log \frac{T-t}{T},$$

$$\Phi\big(t, X^\diamond_{[0:t]}\big) = \int_0^t \Bigg(\frac{\theta\, g^{\mathrm{AI}}_{X_s}\left(\mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s} M, \mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s} V\right)}{\sigma^2 (T-s)} + \sum_{i,j,r=1}^d \frac{g^{\mathrm{AI}}_{X^\diamond_s}\Big(\zeta^i_j(X^\diamond_s)\big(\sum_{l=1}^d \zeta^i_l(X^\diamond_s)\Gamma^r_{jl} + (E^{\mathrm{AI}}_j \zeta^i_r)_{X^\diamond_s}\big) E^{\mathrm{AI}}_r(X^\diamond_s)\,,\, \mathrm{Log}^{\mathrm{AI}}_{X^\diamond_s} V\Big)}{2(T-s)} \Bigg) ds,$$

$$
\begin{aligned}
&\approx \int_0^t \frac{\theta\,\left\langle \log\left((X_s^\diamond)^{-1/2}\star M\right),\,\log\left((X_s^\diamond)^{-1/2}\star V\right)\right\rangle_F}{\sigma^2(T-s)}\,ds \\
&\quad+\sum_{i,j,r=1}^{d}\sum_{k=0}^{m}\frac{\left\langle \delta_{ij}\left(\sum_{l=1}^{d}\delta_{il}\Gamma_{jl}^r\right)S_r\,,\,\log\left(y_{t_k}^\diamond\right)^{-1/2}\star V\right)\right\rangle_F}{2(T-t_k)}\,(t_{k+1}-t_k) \\
&\approx \sum_{k=0}^{m}\frac{t_{k+1}-t_k}{T-t_k}\left\{\frac{\theta\,\left\langle \log\left((y_{t_k}^\diamond)^{-1/2}\star M\right),\,\log\left((y_{t_k}^\diamond)^{-1/2}\star V\right)\right\rangle_F}{\sigma^2}\right. \\
&\quad\left.+\frac{\left\langle \Gamma\,,\,\log\left((y_{t_k}^\diamond)^{-1/2}\star V\right)\right\rangle_F}{2}\right\}.
\end{aligned}
$$

□

Theorem 4.15. *Suppose $h=h(x_{t_1},\ldots,x_{t_N})$ is a bounded $\mathfrak{F}_T$-measurable function for $0<t_1<\cdots<t_N<T$. We write $X=\big(X_{t_1},\ldots,X_{t_N}\big)$ and $X_{[0:T]}=\big(0,X_{t_1},\ldots,X_{t_N},X_T\big)$, then*

$$
\begin{aligned}
\mathbb{E}^*\big[h(X^*)\big] &= \mathbb{E}\big[h(X|X_T=V)\big] \\
&= \mathbb{E}^\diamond\left[h(X^\diamond)\,\frac{\mathscr{H}_{n,T}\,\sigma^{-d}\exp\left(f(X_0^\diamond,\sigma^2)\right)}{p(0,U;T,V)}\,\exp\left(\Phi(T,X_{[0:T]}^\diamond)\right)\right]
\end{aligned}
$$

where $\mathscr{H}_{n,T}$ is a fixed constant that only depends on the dimension n and the terminal time T. Here the notation of expectation $\mathbb{E}$, $\mathbb{E}^\diamond$ and $\mathbb{E}^*$ are corresponding to their laws $\mathbb{P}$, $\mathbb{P}^\diamond$ and $\mathbb{P}^*$ respectively and $f(X_t^\diamond,\sigma^2)$ equals to the function $f(X_t^\diamond)$ defined in Equation (4.31)*. Thus, the measures $\mathbb{P}_t^*$ and $\mathbb{P}_t^\diamond$ are absolutely continuous on $[0,T]$ and the corresponding Radon-Nikodym derivative is given as*

$$
\frac{d\mathbb{P}_T^*}{d\mathbb{P}_T^\diamond}\big(X_{[0:T]}^\diamond\big)=\frac{\mathscr{H}_{n,T}}{p(0,U;T,V)}\,\exp\left\{f(X_0^\diamond,\sigma^2)+\Phi(T,X_{[0:T]}^\diamond)-\frac{d}{2}\log\sigma^2\right\}
$$

Proof plan of Theorem 4.15. We follow similar lines to those proofs in [38, 40].

Firstly, let the function f be defined as in the Proposition 4.13, then $\exp\big(f(X_t^*)\big)/K_n(\sigma)$ is the probability density function of the distribution $\mathfrak{N}\big(V,(T-t)\sigma^2\big)$ on $\mathcal{SP}(n)$, see Equation (4.25). By Proposition 4.13, for any $t\in[0,T)$, and

any bounded $\mathcal{F}_t$-measurable function h, we have

$$\begin{aligned}
&\mathbb{E}^*\left[h(X^*)\,\frac{\exp\big(f(X_t^*)\big)/K_n(\sigma)}{p(t,X_t^*;T,V)}\right]\\
&\quad=\mathbb{E}^\diamond\left[h(X^\diamond)\,\frac{\exp\big(f(X_0^\diamond)\big)}{K_n(\sigma)\,p(0,U;T,V)}\exp\Big\{\Phi\big(t,X_{[0:t]}^\diamond\big)+\phi\big(t,X_{[0:t]}^\diamond\big)\Big\}\right]\\
&\quad=\mathbb{E}^\diamond\left[\frac{h(X^\diamond)}{p(0,U;T,V)}\exp\left\{f(X_0^\diamond)+\Phi\big(t,X_{[0:t]}^\diamond\big)+\log\left(\frac{\exp\phi(t,X_{[0:t]}^\diamond)}{K_n(\sigma)}\right)\right\}\right]
\end{aligned}$$

Then, we need to show the following three equations:

$$\lim_{t\to T}\mathbb{E}^\diamond\left[\frac{\exp\phi\big(t,X_{[0:t]}^\diamond\big)}{K_n(\sigma)}\right]=\log\left(\frac{\mathscr{H}_{n,T}}{\sigma^d}\right). \tag{4.37}$$

Here, the constant $\mathscr{H}_{n,T}$ depends only on the dimension n and the terminal time T.

$$\lim_{t\to T}\mathbb{E}^*\left[h(X^*)\,\frac{\exp\big(f(X_t^*)\big)/K_n(\sigma)}{p(t,X_t^*;T,V)}\right]=\mathbb{E}^*\big[h(X^*)\big]. \tag{4.38}$$

$$\lim_{t\to T}\mathbb{E}^\diamond\Big[\exp\big\{\Phi\big(t,X_{[0:t]}^\diamond\big)\big\}\Big]=\mathbb{E}^\diamond\Big[\exp\big\{\Phi\big(T,X_{[0:T]}^\diamond\big)\big\}\Big]. \tag{4.39}$$

When those statements above are proved (i.e. L^1-convergence), they imply

$$\frac{d\mathbb{P}_T^*}{d\mathbb{P}_T^\diamond}\big(X_{[0:T]}^\diamond\big)=\frac{\mathscr{H}_{n,T}}{p(0,U;T,V)}\,\exp\Big\{f(X_0^\diamond,\sigma^2)+\Phi(T,X_{[0:T]}^\diamond)-\frac{d}{2}\log\sigma^2\Big\}.$$

□

While we are able to prove Equation (4.37) and (4.38), at the moment we are unable to provide a complete proof for Equation (4.39), so we discuss our current proof plan of Equation (4.39) instead. Before proving Equation (4.37), we need to understand the behaviour of the normalising constant $K_n(\sigma)$ of the distribution $\mathfrak{N}\big(V,(T-t)\sigma^2\big)$ when $t\to T$, which is demonstrated in the following lemma.

Lemma 4.16. *Consider the probability distribution* $\mathfrak{N}\big(V,(T-t)\sigma^2\big)$*, see Equation* (4.25)*, for some fixed* $T>0$ *and* $t\in[0,T)$*, then*

$$\lim_{t\to T}\frac{K_n(\sigma)}{(T-t)^{d/2}}=\mathcal{K}_n\,\sigma^d$$

where $\mathcal{K}_n$ is a constant that depends only on $d = n(n+1)/2$.

Proof. Firstly, we note that the Taylor expansion of the $\sinh$ function is

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots$$

$$\Rightarrow \sinh \frac{|r_i - r_j|}{2} = \frac{|r_i - r_j|}{2} + \boldsymbol{\mathcal{O}}\left(\frac{|r_i - r_j|^3}{2^3}\right)$$

We denote $C_1 = \big(\pi^{\frac{n^2}{2}}\, 8^{\frac{n(n-1)}{4}}\big)\big/\big(n!\,\mathrm{Gam}_n(\frac{n}{2})\big)$ and $\rho^2 = (T-t)\sigma^2$, then substitute the variables $u_i = r_i/\rho$ for $1 \le i \le n$ in Equation (4.26), and get

$$\begin{aligned}
K_n(\sigma) &= C_1 \int_{\mathbb{R}^n} \exp\left(-\sum_{i=1}^n u_i^2\right) \prod_{i<j} \sinh \frac{|\rho u_i - \rho u_j|}{2} \prod_{i=1}^n \rho\, du_i \\
&= C_1 \int_{\mathbb{R}^n} \exp\left(-\sum_{i=1}^n u_i^2\right) \prod_{i<j} \left\{\frac{\rho|u_i - u_j|}{2} + \boldsymbol{\mathcal{O}}(\rho^3)\right\} \rho^n \prod_{i=1}^n du_i \\
&= C_1 \int_{\mathbb{R}^n} \exp\left(-\sum_{i=1}^n u_i^2\right) \left\{\rho^{\frac{n(n-1)}{2}} \prod_{i<j} \frac{|u_i - u_j|}{2} + \boldsymbol{\mathcal{O}}\big(\rho^{\frac{n(n-1)}{2}+2}\big)\right\} \rho^n \prod_{i=1}^n du_i \\
&= C_1 \rho^{\frac{n(n+1)}{2}} \int_{\mathbb{R}^n} \exp\left(-\sum_{i=1}^n u_i^2\right) \prod_{i<j} \frac{|u_i - u_j|}{2} \prod_{i=1}^n du_i + \boldsymbol{\mathcal{O}}\big(\rho^{\frac{n(n+1)}{2}+2}\big) \\
&= \mathcal{K}_n\, \rho^d + \boldsymbol{\mathcal{O}}(\rho^{d+2}),
\end{aligned}$$

$$\text{where } \mathcal{K}_n = \frac{\pi^{\frac{n^2}{2}}\, 8^{\frac{n(n-1)}{4}}}{n!\,\mathrm{Gam}_n(\frac{n}{2})} \int_{\mathbb{R}^n} \exp\left(-\sum_{i=1}^n u_i^2\right) \prod_{i<j} \sinh \frac{|u_i - u_j|}{2} \prod_{i=1}^n du_i,$$

which clearly depends only on n.

$$\Rightarrow K_n(\sigma) = \mathcal{K}_n\, \sigma^d (T-t)^{\frac{d}{2}} + \boldsymbol{\mathcal{O}}(\sigma^{d+2}(T-t)^{\frac{d}{2}+1}) \Rightarrow \lim_{t \to T} \frac{K_n(\sigma)}{(T-t)^{d/2}} = \mathcal{K}_n\, \sigma^d.$$

□

Proof of Equation (4.37). Function ϕ is defined in Equation (4.32) by

$$\phi\big(t, X^{\diamond}_{[0:t]}\big) = \sum_{i,j=1}^d \int_0^t \frac{\big(\zeta^i_j(X^{\diamond}_s)\big)^2}{2(T-s)}\, ds,$$

The requirement that functions ζ^i_j need to satisfy in Equation (4.18) (see Proposi-

tion 4.9) and the information that $X_t^\diamond$ is obtained from taking the limit of $\delta_t \to 0$ when using the exponential adapted Euler-Maruyama method, imply $\mathbb{E}^\diamond\big[\,\big|\zeta_j^i(X_t^\diamond)\big|\,\big] < \infty$, and equivalently $\mathbb{E}^\diamond\big[\,\big(\zeta_j^i(X_t^\diamond)\big)^2\,\big] < \infty$ for all $t \in [0,T]$.

Even though ζ_j^i is integrable, the function itself does not have explicit form. We compute this Riemann-Stieltjes integral using a sequence of partitions of $[0,t]$ with mesh tending to zero, that is $t_0 = 0 < \ldots < t_m = t$ and $\Delta_k = (t_{k+1} - t_k) \to 0$ as $m \to \infty$ for all $0 \le k \le m-1$. Moreover, similar to Corollary 4.11, at t_{k+1} we can express $\zeta_j^i(X_{t_{k+1}}^\diamond)$ conditioned on $X_{t_k}^\diamond$ by Taylor series:

$$\begin{aligned}
\zeta_j^i(X_{t_{k+1}}^\diamond) &= \zeta_j^i \circ \gamma(0) + (t_{k+1} - t_k)\big(\zeta_j^i \circ \gamma\big)'(0) + \boldsymbol{\mathcal{O}}\big((t_{k+1} - t_k)^2\big). \\
&\Rightarrow \zeta_j^i(X_{t_{k+1}}^\diamond) = \delta_{ij} - \Delta_k \sum_{i=1}^{d} \alpha_r \Gamma_{ri}^j + \boldsymbol{\mathcal{O}}(\Delta_k^2).
\end{aligned}$$

Here, $\{\gamma(s)\,|\,t_k \le s \le t_{k+1}\}$ is the geodesic curve connecting $X_{t_k}^\diamond$ and $X_{t_{k+1}}^\diamond$ when using the exponential adapted Euler-Maruyama method and

$$\begin{aligned}
\alpha_r &= \theta\Big\langle \log\big((X_{t_k}^\diamond)^{-1/2} \star M\big) + \frac{\log\big((X_{t_k}^\diamond)^{-1/2} \star V\big)}{T - t_k}\,,\, S_r \Big\rangle_F + \sigma\big(B_{t_{k+1}}^{(r)} - B_{t_k}^{(r)}\big). \\
&\Rightarrow \big(\zeta_j^i(X_{t_{k+1}}^\diamond)\big)^2 = \delta_{ij}^2 - 2\,\delta_{ij}\,\Delta_k \sum_{i=1}^{d} \alpha_r \Gamma_{ri}^j + \boldsymbol{\mathcal{O}}(\Delta_k^2). \\
\Rightarrow \phi\big(t, X_{[0,t]}^\diamond\big) &= \sum_{i,j=1}^{d} \lim_{\Delta_k \to 0} \sum_{k=1}^{m-1} \frac{\delta_{ij}^2 - 2\,\delta_{ij}\,\Delta_k \sum_{i=1}^{d} \alpha_r \Gamma_{ri}^j + \boldsymbol{\mathcal{O}}(\Delta_t^2)}{2(T - t_k)}\,\Delta_k \\
&= \sum_{i,j=1}^{d} \lim_{\Delta_k \to 0} \sum_{k=1}^{m-1} \frac{\delta_{ij}^2}{2(T - t_k)}\,\Delta_k \\
&\quad + \sum_{i,j=1}^{d} \lim_{\Delta_k \to 0} \sum_{k=1}^{m-1} \frac{-2\,\delta_{ij}\,\Delta_k^2 \sum_{i=1}^{d} \alpha_r \Gamma_{ri}^j + \boldsymbol{\mathcal{O}}(\Delta_k^3)}{2(T - t_k)}
\end{aligned}$$

The first line simply equals $\sum_{i=1}^{d} \int_0^t \big(2(T-s)\big)^{-1} ds = \log\big(\frac{T-t}{T}\big)^{d/2}$. On the other hand, since our main aim is to take the limit $t \to T$, and this means t_m tends to T. But the term $T - t_k$ stops decreasing when $k = m-1$ so that the denominator is never zero, thus explosion around T will not occur. Hence, the second line converges to zero as the mesh $\Delta_k \to 0$ because the numerator decreases much faster than the

denominator. Now, since $\mathbb{E}^{\diamond}\big[\,\big(\zeta_j^i(X_t^{\diamond})\big)^2\,\big] < \infty$ for all $t \in [0,T]$, Fubini's theorem implies

$$\Rightarrow \mathbb{E}\left[\frac{\exp\phi\big(t, X^{\diamond}_{[0:t]}\big)}{K_n(\sigma)}\right] \to \frac{(T-t)^{d/2}}{T^{d/2}\,K_n(\sigma)} \to \frac{\mathscr{H}_{n,T}}{\sigma^d} \quad \text{as } t \to T,$$

where Lemma 4.16 assists the second limit. □

Proof of Equation (4.38). Using Bayes theorem, Markov property and setting $t = t_N \in (0,T)$, we get

$$\begin{aligned}
&\mathbb{E}^*\left[\frac{h(X^*_{t_1},\ldots,X^*_{t_{N-1}},X^*_t)}{p(t,X^*_t;T,V)}\right]\\
&= \int_{\mathcal{SP}(n)^N} \frac{h(x_1,\ldots,x_{N-1},x_t)\,p(x_1,\ldots,x_{N-1},x_t\,|\,x_T=V)}{p(x_T=V\,|x_t)}\,dx_1\ldots dx_{N-1}\,dx_t\\
&= \int_{\mathcal{SP}(n)^N} \frac{h(x_1,\ldots,x_{N-1},x_t)\,p(x_1,\ldots,x_{N-1},x_t\,|\,x_T=V)}{p(x_T=V\,|x_t,x_{N-1},\ldots,x_1)}\,dx_1\ldots dx_{N-1}\,dx_t\\
&= \int_{\mathcal{SP}(n)^N} \frac{h(x_1,\ldots,x_{N-1},x_t)\,p(x_1,\ldots,x_{N-1},x_t)}{p(x_T=V)}\,dx_1\ldots dx_{N-1}\,dx_t\\
&= \mathbb{E}\left[\frac{h(X^*_{t_1},\ldots,X^*_{t_{N-1}},X^*_t)}{p(0,U;T,V)}\right]
\end{aligned}$$

Now suppose that $t \in (t_N,T)$,

$$\begin{aligned}
&\Rightarrow \mathbb{E}^*\left[h(X^*)\,\frac{\exp\big(f(X^*_t)\big)}{K_n(\sigma)p(t,X^*_t;T,V)}\right] = \mathbb{E}\left[h(X^*)\,\frac{\exp\big(f(X^*_t)\big)/K_n(\sigma)}{p(0,U;T,V)}\right]\\
&= \int_{\mathcal{SP}(n)^{N+1}} \frac{h(x_1,\ldots,x_N)\,\exp f(z)}{K_n(\sigma)\,p(0,U;T,V)}\,p(x_1,\ldots,x_N,z)\,dx_1,\ldots dx_N dz\\
&= \int_{\mathcal{SP}(n)} \Upsilon(t,z)\frac{\exp f(z)}{K_n(\sigma)}dz \qquad (4.40)\\
&\text{where } \Upsilon(t,z) = \int_{\mathcal{SP}(n)^N} \frac{h(x_1,\ldots,x_N)\,p(x_1,\ldots,x_N,z)}{p(0,U;T,V)}dx_1\ldots dx_N.
\end{aligned}$$

In order to compute the limit when t tends to T, we apply change of variables $x = \mathrm{Exp}_V^{\mathrm{AI}}\big\{(T-t)^{-1/2}\,\mathrm{Log}_V^{\mathrm{AI}} z\big\}$, or equivalently $z = \mathrm{Exp}_V\big\{(T-t)^{1/2}\,\mathrm{Log}_V^{\mathrm{AI}} x\big\}$. Then $d^2_{\mathrm{AI}}(z,V) = (T-t)\,d^2_{\mathrm{AI}}(x,V)$ and

$$\lim_{t\to T} \mathrm{Exp}_V\Big((T-t)^{1/2}\,\mathrm{Log}_V x\Big)$$

$$= \lim_{t\to T} \left[V^{1/2} \star \exp\Big((T-t)^{1/2} \log\big(V^{-1/2} \star x \big) \Big) \right] = V.$$

Moreover, since $\exp f(z)/K_n(\sigma)$ is the probability density of $\mathfrak{N}(V, (T-t)\sigma^2)$,

$$\int_{\mathcal{SP}(n)} \frac{\exp f(z)}{K_n(\sigma)}\, dz = 1 \Rightarrow \int_{\mathcal{SP}(n)} \frac{\exp\big(-d^2(x,V)/(2\sigma^2)\big)}{K_n(\sigma)}\, \tilde{h}(x,t)\, dx = 1$$

where $\tilde{h}(x,t) = dz/dx$.

$$\begin{aligned} &\Rightarrow \lim_{t\to T} \int_{\mathcal{SP}(n)} \frac{\exp\big(-d^2(x,V)/(2\sigma^2)\big)}{K_n(\sigma)}\, \tilde{h}(x,t)\, dx = 1 \\ &\Rightarrow \int_{\mathcal{SP}(n)} \exp\big(-d^2(x,V)/(2\sigma^2)\big) \lim_{t\to T} \left(\frac{\tilde{h}(x,t)}{K_n(\sigma)} \right) dx = 1 \end{aligned}$$

Here, we use the dominated convergence theorem to interchange the limiting operation and integral. Moreover, $\lim_{t\to T}\big(\tilde{h}(x,t)/\mathcal{K}_n(t)\big)$ simply is the normalising constant of $\mathfrak{N}(V,\sigma^2)$, thus the limit exists and is finite.

Back to the Equation (4.40), we have

$$\begin{aligned} &\mathbb{E}^* \left[h(X^*)\, \frac{\exp\big(f(X_t^*)\big)}{K_n(\sigma) p(t, X_t^*; T, V)} \right] \\ &= \int_{\mathcal{SP}(n)} \Upsilon\Big(t, \mathrm{Exp}_V\big\{(T-t)^{1/2}\, \mathrm{Log}_V x\big\}\Big) \frac{\exp\big(-d^2(x,V)/(2\sigma^2)\big)}{K_n(\sigma)}\, \tilde{h}(x,t)\, dx \\ &\overset{t\to T}{\longrightarrow} \Upsilon(T,V) \int_{\mathcal{SP}(n)} \exp\left(\frac{-d^2(x,V)}{2\sigma^2} \right) \lim_{t\to T} \left(\frac{\tilde{h}(x,t)}{K_n(\sigma)} \right) dx = \Upsilon(T,V) \end{aligned}$$

Thus, we can deduce Equation (4.38) because

$$\Upsilon(T,V) = \int_{\mathcal{SP}(n)^N} \frac{h(x_1,\ldots,x_N)\, p(x_1,\ldots,x_N,V)}{p(0,U;T,V)} dx_1 \ldots dx_N = \mathbb{E}^*\left[h(X^\star)\right]$$

□

Proof plan of Equation (4.39). Firstly, we denote $\varphi_t = \exp\big\{\Phi\big(t, X^\diamond_{[0:t]}\big)\big\}$. The plan is to show $\varphi_t \to \varphi_T$ as $t \to T$ in L^1-sense, by proving

$$\mathbb{E}^\diamond[\varphi_T] \le \lim_{t\to T} \inf \mathbb{E}^\diamond[\varphi_t] \le \lim_{t\to T} \sup \mathbb{E}^\diamond[\varphi_t] \le \mathbb{E}[\varphi_T]$$

Using Fatou's lemma, we easily get $\mathbb{E}[\varphi_T] \leq \lim_{t\to T} \inf \mathbb{E}^\diamond[\varphi_t]$, whereas the right inequality is not so straightforward to show. We lay out one approach, which we hope can solve the problem and discuss some obstacles of this approach.

For $m \in \mathbb{N}$ and $\epsilon \in (0, 1/6)$, we define

$$\tau_m(Z) = T \wedge \inf_{t\in[0,T]} \left\{ d_{\mathrm{AI}}(Z_t, V) \geq m(T-t)^{1/2-\epsilon} \right\}$$

for some stochastic process Z_t on $\mathcal{SP}(n)$.

Step 1. We want to prove that there exist an $\epsilon \in (0, 1/6)$ and an almost surely finite random variable K such that for all $t \in [0, T)$, it holds almost surely that

$$d_{\mathrm{AI}}(X_t^\diamond, V) \leq K(T-t)^{1/2-\epsilon}.$$

That is to say $X_t^\diamond$ not only converges to V almost surely as t tends to T, but also it needs to converge at an appropriate rate. If we can prove this, it implies $\tau_m(X^\diamond) \uparrow T$ in probability as $m \to \infty$.

The first statement sounds obvious given the way we construct the guided process $X_t^\diamond$, however, proving this formally may be harder than it looks. On the Euclidean space, this is shown by firstly using the Itô's lemma to the function $||X_t^\diamond - V||_F/(T-t)$, followed by the law of the iterated logarithm for the Brownian motion on $\mathbb{R}$, see [38]. The proof of Theorem 4.13 demonstrates the use of Itô lemma on $\mathcal{SP}(n)$ to $d_{\mathrm{AI}}\big(X_t^\diamond, V\big)/(T-t)$. We have seen the extension of the law of the iterated logarithm for the Brownian motion to a compact Riemannian manifold, see [129, 130], we are not sure with the case of non-compact space like $\mathcal{SP}(n)$. And even when there is some kind of formula similar to this iterated logarithm law, making use of this law is a separate problem because of the involvement of the implicit function ζ coming from the horizontal lift of $X_t^\diamond$.

Step 2. Since

$$\frac{\mathbb{E}^\diamond\left[\varphi_t \mathbb{I}_{t\leq\tau_m(X^\diamond)}\right]}{\mathbb{E}^\diamond[\varphi_t]} = 1 - \frac{\mathbb{E}^\diamond\left[\varphi_t \mathbb{I}_{t>\tau_m(X^\diamond)}\right]}{\mathbb{E}^\diamond[\varphi_t]},$$

we are now required to show $\mathbb{E}^{\diamond}\left[\varphi_t \mathbb{I}_{t>\tau_m(X^{\diamond})}\right] / \mathbb{E}^{\diamond}\left[\varphi_t\right] \leq A_m$ for some positive constant A_m such that $\lim_{m\to\infty} A_m = 0$. Because by taking $\liminf_{t\to T}$ of this equation, we get

$$\frac{\mathbb{E}^{\diamond}\left[\varphi_T \mathbb{I}_{T=\tau_m(X^{\diamond})}\right]}{\limsup_{t\to T} \mathbb{E}^{\diamond}\left[\varphi_t\right]} = 1 - \frac{\mathbb{E}^{\diamond}\left[\varphi_t \mathbb{I}_{t>\tau_m(X^{\diamond})}\right]}{\mathbb{E}^{\diamond}\left[\varphi_t\right]} \geq 1 - A_m$$

And since $\tau_m(X^{\diamond})$ converges to T as $m \to \infty$ (result from step 1), then we can deduce

$$\mathbb{E}^{\diamond}\left[\varphi_T\right] \geq \limsup_{t\to T} \mathbb{E}^{\diamond}\left[\varphi_t\right].$$

While we study the proof in [38, 40] for the Euclidean setting, we also seek other options.

□

While we do not present the full argument to prove Theorem 4.15, we do, however, provide a careful numerical validation in Section 4.6. We end this section by presenting the guided proposal algorithm on $\mathcal{SP}(n)$ equipped with the AI metric.

Suppose we have discretely observed data $\mathfrak{D} = \{X_{t_j} = y_j\}_{j=0}^{N}$ at observation times $t_0 = 0 < t_1 < \cdots < t_N = T$, where the diffusion process X_t is the Riemannian OU process on $\mathcal{SP}(n)$ equipped with the AI metric. We wish to sample from the posterior distribution of $\Theta = \{\theta,\, M, \sigma^2\}$. Let us set $\mu = \sum_{i=1}^{d} \mu^{(i)}\, e_i = \mathfrak{h}(M)$, where the function $\mathfrak{h}$ is defined in Equation (4.19). Here, π_0^{θ}, π_0^{μ} and π_0^{σ} denote the prior distributions for $\{\theta, \mu, \sigma^2\}$ respectively. The key step involves first imputing suitable $m_j - 1$ data points between the jth consecutive observations in a way that they are independent of the diffusivity, and then using the exponential adapted Euler-Maruyama method to approximate the likelihood. We choose a random walk proposal, with suitable step size, denoted by $q(\tilde{\theta}|\theta)$, $q(\tilde{\mu}|\mu)$ and $q(\tilde{\sigma}^2|\sigma^2)$ for $\{\theta, \mu, \sigma^2\}$ respectively.

Algorithm 4.17 (Guided proposals on $\mathcal{SP}(n)$)**.**

1. (Iteration $k = 0$). Choose starting values for Θ_0 and sample standard Brownian motions W_j, independently for $1 \leq j \leq N$, each covering the time interval

$t_j - t_{j-1}$, and set $B_j^{(0)} = W_j$.

2. (Iteration $k \geq 1$).

   (a) Update B_j independently $(1 \leq j \leq N)$: sample the proposal $\tilde{W}_j$ and obtain $\tilde{Y}_{[t_{j-1},t_j]}$ from $\{\tilde{W}_j, \Theta_{k-1}, \mathfrak{D}\}$ and $Y_{[t_{j-1},t_j]}$ from $\{B_j^{(k-1)}, \Theta_{k-1}, \mathfrak{D}\}$ using the exponential adapted Euler-Maruyama method to approximately solve the SDE (4.28); then accept $\tilde{W}_j$ with probability

$$\alpha^{(B)} = \exp\Big\{\Phi\big(t_j - t_{j-1}, \tilde{Y}_{[t_{j-1},t_j]}\big) - \Phi\big(t_j - t_{j-1}, Y_{[t_{j-1},t_j]}\big)\Big\}$$

   (b) Sample $\tilde{\sigma}^2$ from $q(\sigma^2|\sigma_{k-1}^2)$ and obtain $\tilde{Y}_{[t_{j-1},t_j]}$ and $Y_{[t_{j-1},t_j]}$ from $\{B_j^{(k)}, \theta_{k-1}, M_{k-1}, \tilde{\sigma}^2, \mathfrak{D}\}$, $\{B_j^{(k)}, \Theta_{k-1}, \mathfrak{D}\}$ respectively using the exponential adapted Euler-Maruyama method to approximately solve the SDE (4.28); then accept $\tilde{\sigma}^2$ with probability

$$\begin{aligned}\alpha^{(\sigma)} = &\frac{\pi_0^\sigma(\tilde{\sigma}^2)\prod_{j=1}^N \exp\big\{f(\tilde{Y}_{t_{j-1}}, \tilde{\sigma}^2) - \frac{d}{2}\tilde{\sigma}^2\big\}}{\pi_0^\sigma(\sigma_{k-1}^2)\prod_{j=1}^N \exp\big\{f(Y_{t_{j-1}}, \sigma_{k-1}^2) - \frac{d}{2}\sigma_{k-1}^2\big\}} \\ &\cdot \prod_{j=1}^N \exp\Big\{\Phi\big(t_j - t_{j-1}, \tilde{Y}_{[t_{j-1},t_j]}\big) - \Phi\big(t_j - t_{j-1}, Y_{[t_{j-1},t_j]}\big)\Big\}.\end{aligned}$$

   (c) Update μ and M : sample $\tilde{\mu}$ from $q(\mu|\mu_{k-1})$, compute the corresponding $\tilde{M} = \mathfrak{h}^{-1}(\tilde{\mu})$ and accept $\tilde{\mu}$, $\tilde{M}$ with probability

$$\begin{aligned}\alpha^{(M)} = &\frac{\pi_0^\mu(\tilde{\mu})}{\pi_0^\mu(\mu_{k-1})} \\ &\cdot \prod_{j=1}^N \exp\Big\{\Phi\big(t_j - t_{j-1}, \tilde{Y}_{[t_{j-1},t_j]}\big) - \Phi\big(t_j - t_{j-1}, Y_{[t_{j-1},t_j]}\big)\Big\},\end{aligned}$$

   where $\tilde{Y}_{[t_{j-1},t_j]}$ and $Y_{[t_{j-1},t_j]}$ are obtained from $\{B_j^{(k)}, \theta_{k-1}, \tilde{M}, \sigma_k^2, \mathfrak{D}\}$ and $\{B_j^{(k)}, \theta_{k-1}, M_{k-1}, \sigma_k^2, \mathfrak{D}\}$ respectively using the exponential adapted Euler-Maruyama method to approximately solve the SDE (4.28).

   (d) Update θ similarly as μ.

Remark 4.18 (Time change)**.** Since Φ in Equation (4.33) explodes as $t \uparrow T$, M.

Schauer & F. Van Der Meulen et al. [40] suggest time change and scaling to reduce the required number of imputed data points. Scaling will not be as effective here as in the univariate setting because it can only fit one of the directions involved, so the effect is less pronounced than in the univariate setting and we expect further lessening as dimension increases. In this thesis, we adopt a time change function τ from [40]: $t \mapsto t\,(2 - t/T)$. And using this time change, our SDE (4.22) becomes

$$dX_{\tau(t)} = -\frac{\theta}{2}\,\tau'(t)\left(\nabla d^2(X_{\tau(t)}, M)\right)_{X_{\tau(t)}}\, dt + F_{X_{\tau(t)}}\left(\sqrt{\tau'(t)}\,\sigma\, dB_t\right) \qquad (X_0 = U),$$

where $\tau'(t) = 2\,(1 - t/T)$.

4.6 Simulation study on $\mathcal{SP}(2)$

4.6.1 Brownian bridges

We perform a simulation exercise to illustrate our proposed bridge sampling method of Algorithm 4.17 and compare among three metric tensors: the Euclidean, the LE and the AI metrics, when sampling the standard Brownian bridges on $\mathcal{SP}(n)$.

The simulation scenario involves sampling a standard Brownian bridge W_t conditioned on $\{W_0 = U, W_T = V\}$ with $T = 0.1$ in two cases:

(i) with U and V lying far away from the boundary,

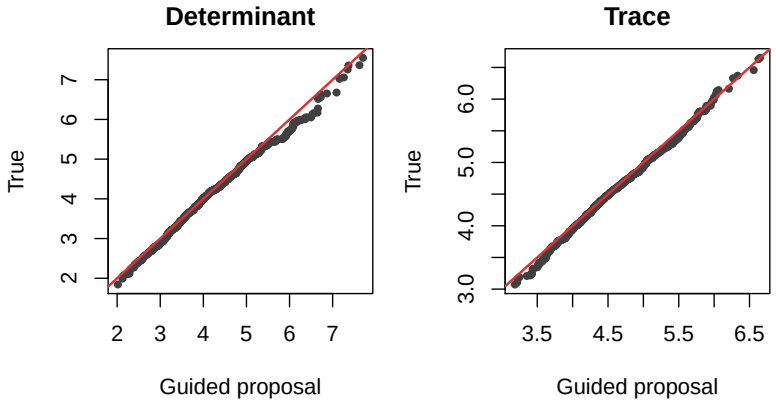
$$U = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad \& \qquad V = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

(ii) with U and V lying close to the boundary,

$$U = \begin{pmatrix} 2 & 1.999 \\ 1.999 & 2 \end{pmatrix} \qquad \& \qquad V = \begin{pmatrix} 3 & 2.435 \\ 2.435 & 2 \end{pmatrix}.$$

For the Euclidean metric, we simply embed $\mathcal{SP}(n)$ in $\mathcal{S}(n)$ endowed with $\langle \cdot, \cdot \rangle_F$, and the Brownian motion on $\mathcal{SP}(n)$ is therefore the solution of the SDE $d\nu(W_t) = dB_t$, with $W_0 = U$, B_t is the standard Brownian motion on $\mathbb{R}^d$ and

$\nu(W_t) \in \mathbb{R}^d$ contains only independent entries of W_t. Working with the LE and AI metrics based on Appendix B.2 (e.g. combining the result from Theorem 4.10) and Algorithm 4.17 respectively.



p-values of K-S tests

		ϵ		
		0.2	0.1	0.05
Determinant	$m = 1000$	0.00412	0.0329	0.288
	$m = 2000$	0.00283	0.0612	0.370
Trace	$m = 1000$	0.0112	0.0293	0.108
	$m = 2000$	0.00123	0.0378	0.288

Figure 4.3: Comparison in distribution at $t = T/2$ of the true bridges and the guided proposal bridges, given that U and V are chosen as in Case 1. Top figures: the Q-Q plots for determinant and trace when $m = 2000$ and $\epsilon = 0.05$. Bottom table: p-values from the K-S tests, where ϵ and m are varied.

Firstly, we evaluate the performance of Algorithm 4.17, compared with the naive simulation approach. For case (i), we obtain samples by forward-simulating the guided proposal in Equation (4.28) and accepting with probability $\alpha^{(B)}$ in Algorithm 4.17. We then compare these bridges with the so-called true bridges, which are generated from the naive simulation approach by forward-solving the Riemannian Brownian motion X_t and picking only those paths X_t that satisfy $\mathrm{d}_{\mathrm{AI}}(X_{0.1}, V) < \epsilon$ for some small positive ϵ. For different values of ϵ and number of imputed points m, we collect 1000 path-valued Brownian bridges and carry out a Kolmogorov–Smirnov (K-S) test to compare the distribution of the true and approximated bridges (i.e. coming from Algorithm 4.17) at $t = T/2 = 0.05$. Note that the number of paths used in the test stays constant as ϵ is decreased. The Q-Q plots and the p-values from K-S

test shown in Figure 4.3 indicate that the values $m = 2000, \epsilon = 0.05$ provide a good approximation and as $\epsilon \downarrow 0$ the two distributions get closer to each other. Note that for case (ii) there is some difficulty in sampling true bridges because V stays very close to the boundary of $\mathcal{SP}(2)$ and points on the boundary are at infinite distance to covariance matrices. In this situation, we observe that the acceptance rate of using naive approach approximately equals zero. This issue escalates as dimension n increases. Overall, apart from high dependence with the diffusivity, the naive approach of simulating bridges has very low acceptance rate when the conditioned observation is close to being non-positive definite, thus it is clearly not a suitable approach to use in the data-augmentation algorithm.

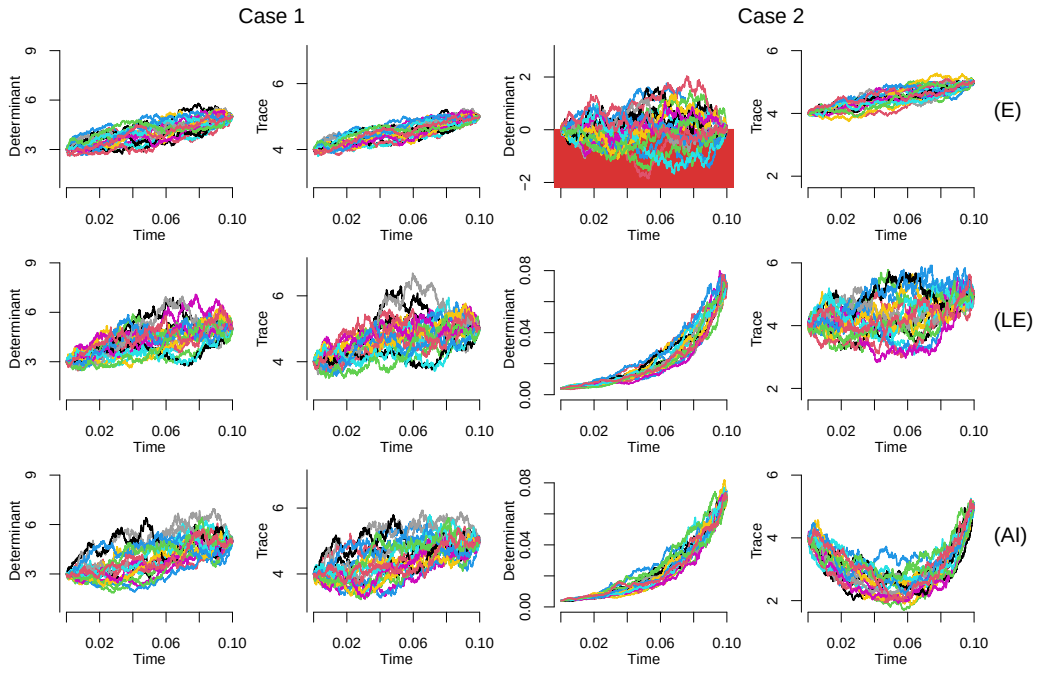


Figure 4.4: Time series for determinant and trace of 20 simulated Brownian bridges on $\mathcal{SP}(2)$ endowed with three different metrics (Euclidean: top row, Log Euclidean: middle row, Affine Invariant: bottom row) in the 2 cases (i) and (ii). Red background area shows failure to be positive definite.

Next, we discuss the problems that arise when neglecting the geometric structure of $\mathcal{SP}(n)$ as mentioned in Subsection 4.2.3. Figure 4.4 depicts traces and determinants of 20 simulated bridges with each metric. Firstly, we observe that there is a clear presence of the swelling effect and it is extremely difficult to get a whole matrix-valued path that preserves the positive definiteness in the case of the

Euclidean metric. Nonetheless, sample points in the case of either the LE or the AI metrics, lie completely in $\mathcal{SP}(n)$ while their determinants behave reasonably with respect to the conditional points. In fact, Corollary 4.6 implies distribution of the determinant is the same for the LE and the AI metrics. The AI metric leads to less anisotropy that becomes more noticeable when the conditional points have eigenvalues close to zero, see plots about the trace in Figure 4.4, indicating that the selection between LE and AI depends on the desirable property that one wishes to achieve.

4.6.2 Parameter estimation for the Affine-Invariant metric

We simulate $10^6 + 1$ equidistant time points from the Riemannian OU process X_t on $\mathcal{SP}(n)$ equipped with the AI metric for $0 \leq t \leq 100$, i.e. using Equation (4.23), with following model parameters

$$\theta = 0.5 \ \& \ M = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} \ \& \ \sigma^2 = 1 \ \& \ X_0 = I_2.$$

We then take sub-samples at time points $\{0, 0.2, \ldots, 100\}$ and apply Algorithm 4.17 with time change, assuming the prior distributions $\log \sigma^2, \log \theta \sim \mathcal{N}(0, 4)$ and $\mu \sim \mathcal{N}(0, 4 \times I_3)$.

Figure 4.5, based on 1,000 burn-in and 4,000 MCMC iterations, indicates that increasing m does not affect the mixing of the chain and improves, in some cases, the approximation to the marginal densities. We then run a longer MCMC chain of 50000 iterations while varying the value of m over $10,\ 50,\ 100$ and 200. These chains are thinned out after a burn-in period of 2000 iterations and samples of 4000 points are collected from the target distributions. We can see from Figure 4.6 that the kernel density estimations of the marginal posterior distributions of $\{\theta,\ \mu,\ \sigma^2\}$ are approximately the same for $m = 100,\ 200$. Thus, $m = 100$ is considered to provide a sufficiently fine discretization for these data. The average proportions of accepting the bridges after the burn-in period are $72.7\%, 70.2\%, 68.5\%$ and 67.3% for $m = 10, 50, 100$ and 200 respectively.

Our final investigation is the prior reproduction test in [131].

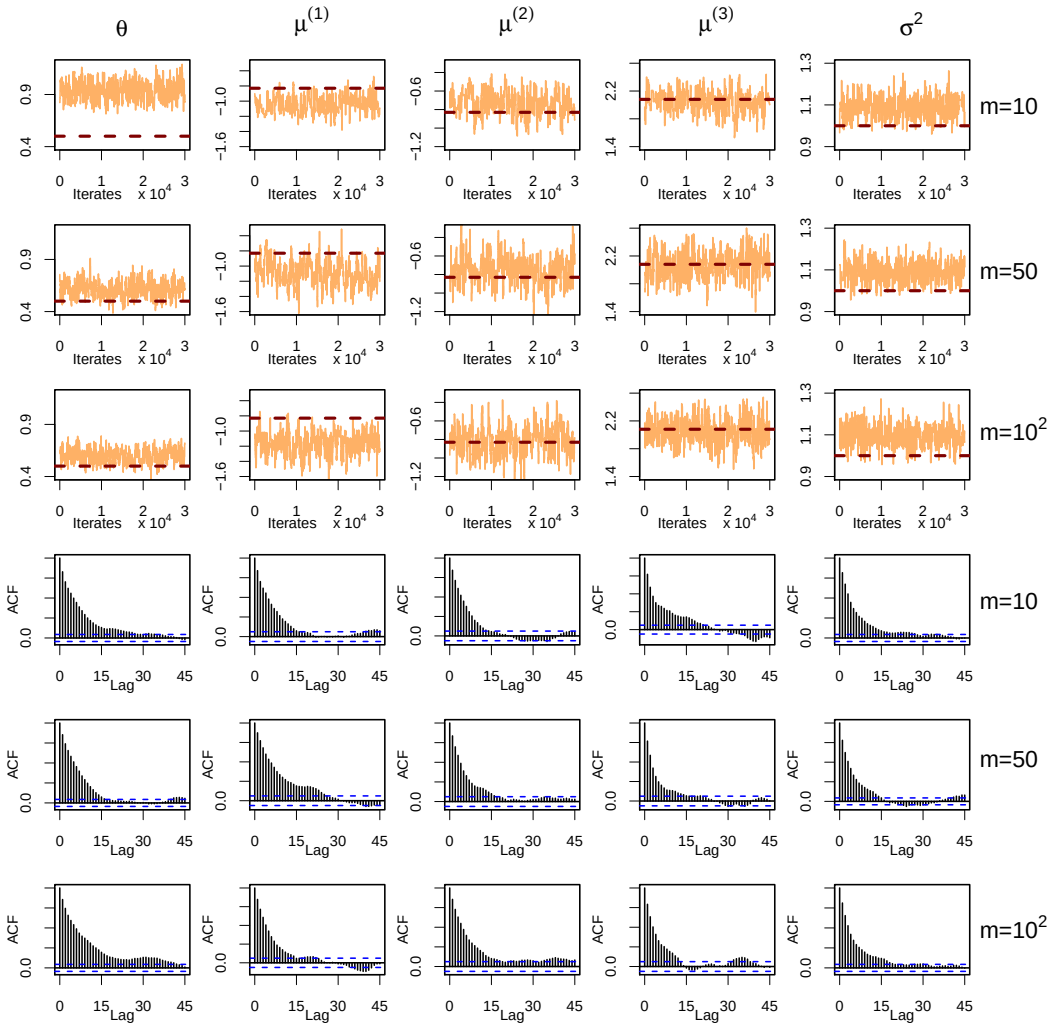


Figure 4.5: (Simulation study). m is varied over $10,\ 50$ and 100. Top three rows: Traceplots from 4×10^3 iterates after discarding 10^3 iterations of burn-in, where true values are indicated with the red dashed lines. Bottom three rows: ACF plots based on these MCMC chains.

Algorithm 4.19 (Prior reproduction test)**.** Assume that prior π_0 for our parameters Θ is proper, and we wish to validate the algorithm $\mathfrak{M}$, that allows sampling from the posterior distribution of Θ, given that observations come from some model $f(X|\Theta)$.

1. Sample $\{\Theta_1,\ldots,\Theta_l\}$ from the prior distribution $\pi_0(\Theta)$.

2. Following steps are carried out independently for each $1\leq i\leq l$:

   - Simulate high-frequency observations and take sub-samples to get $\{x_1^{(\Theta_i)},\ldots,x_N^{(\Theta_i)}\}$, from the model $f(X|\Theta=\Theta_i)$.

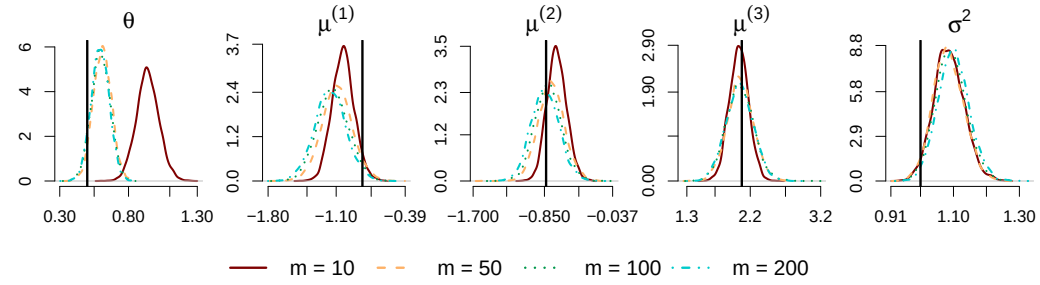


Figure 4.6: (Simulation study). Estimated posterior distribution of $\{\theta, \mu, \sigma^2\}$ using 5×10^4 MCMC iterations (2×10^3 burn-in discarded, thinned by 12), and the result for $M = \mathfrak{h}^{-1}(\mu)$ is given in Figure A.5 (Appendix). True values are indicated by solid vertical black lines.

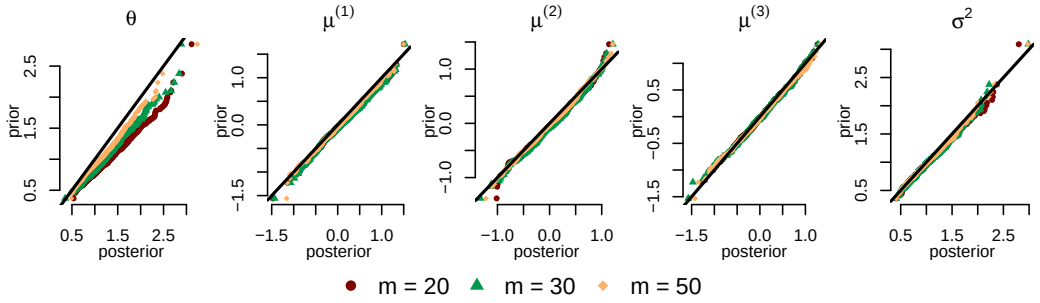


Figure 4.7: Prior reproduction test to validate the Algorithm 4.17 with time change: Q-Q plots of priors against posteriors from three model parameters.

- Use algorithm $\mathfrak{M}$ to draw one sample $\hat{\Theta}_i$ from the posterior distribution $p(\Theta|\{x_1^{(\Theta_i)}, \ldots, x_N^{(\Theta_i)}\})$.

3. Collect a sample $\{\hat{\Theta}_1, \ldots, \hat{\Theta}_l\}$, which comes from some distribution $\pi(\Theta)$.

If algorithm $\mathfrak{M}$ works properly, we expect $\pi = \pi_0$. For sufficiently large l, we can get a good approximate of the distribution π from the sample $\{\hat{\Theta}_1, \ldots, \hat{\Theta}_l\}$.

We assume proper priors for $\{\theta, \mu, \sigma^2\}$: $\log\theta$, $\log\sigma^2 \sim \mathcal{N}(0, 0.3^2)$, $\mu \sim \mathcal{N}(0, 0.2 \times I_3)$ and $N = 500$. And we generate, in turn, 1000 samples from the prior distributions (i.e. $l = 1000$ in Algorithm 4.19) and then, conditional on each sampled parameter vector, high-frequency observations on $[0, 50]$ at $5 \times 10^5 + 1$ equidistant time points and keep sub-samples at time points $\{0, 0.1, \ldots, 50\}$. For the 1000 generated datasets , we estimate the corresponding posterior densities using Algorithm 4.17 and test whether they come from the same distribution as the prior as this validates that our algorithm works properly. Figure 4.7 illustrates that the prior

has been successfully replicated while, as expected, the parameters approximation improves as number of imputed points m increases.

4.7 Application to finance

4.7.1 Data preprocessing

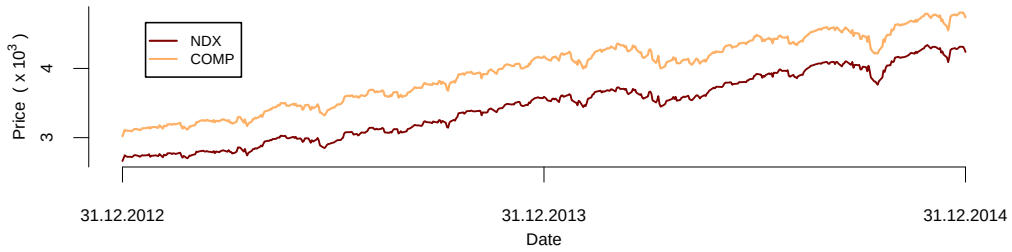


Figure 4.8: Time series of the closing price from two indices: NASDAQ Composite (COMP) and NASDAQ 100 (NDX) at the end of each working day from 31.12.2012 to 31.12.2014.

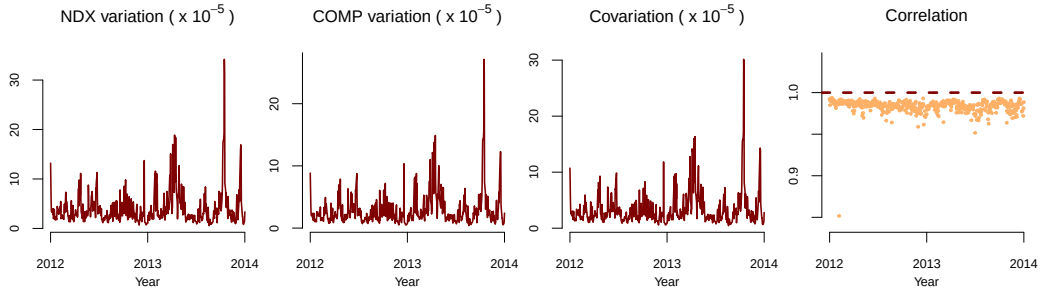


Figure 4.9: Time series of estimated covariance matrices based on the price of two indices NDX and COMP. Estimated correlations between indices are also plotted against time (red dashed line indicates correlation 1).

We estimate 2×2 daily covariance matrices (volatilities) from NASDAQ Composite (COMP) and NASDAQ 100 (NDX) indices, where the data is obtained over 504 working days from 31.12.2012 to 31.12.2014 at $1-$minute intervals, see Figure 4.8; the data is downloaded from [132]. For the estimation we used quadratic variation/covariation in the logarithm domain. For each day i, we estimate a corresponding covariance matrix y_i, based on L high-frequency observations $\{a_j^{(i)}\}_{j=1}^{L}$ (NDX) and $\{b_j^{(i)}\}_{j=1}^{L}$ (COMP) on that day.

$$(y_i)_{11} = \sum_{j=1}^{L-1} \frac{\left(\log a_{j+1}^{(i)} - \log a_j^{(i)}\right)^2}{\mathcal{T}} \quad \& \quad (y_i)_{22} = \sum_{j=1}^{L-1} \frac{\left(\log b_{j+1}^{(i)} - \log b_j^{(i)}\right)^2}{\mathcal{T}},$$

$$(y_i)_{12} = \sum_{j=1}^{L-1} \frac{\left(\log a_{j+1}^{(i)} - \log a_j^{(i)}\right)\left(\log b_{j+1}^{(i)} - \log b_j^{(i)}\right)}{\mathcal{T}}.$$

Here $\mathcal{T}$ is regarded as the time unit, e.g. $\mathcal{T} = 1$ means x_i are in unit of day. We assume that the microstructure noise does not impact our estimates because the indices are very liquid, see [133]. We verify this assumption by observing that estimates based on 5-minute inter-observation intervals are very similar; see Figure 4.10.

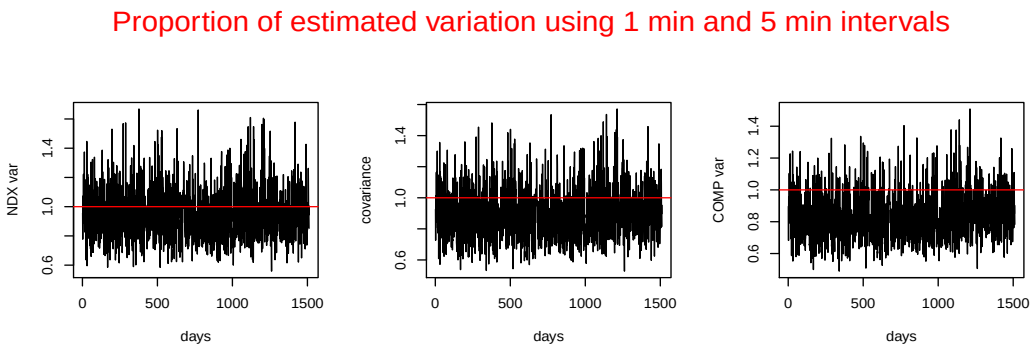


Figure 4.10: Ratio of the volatility estimation when using data observed every 1 minutes or 5 minutes.

Moreover, the pattern of the time series from the entries of covariance matrices in Figure 4.9 indicates a mean-reverting tendency, together with many observations lying close to the boundary of $\mathcal{SP}(2)$, making the Euclidean metric inappropriate and the Riemannian structures suitable.

4.7.2 Model fitting

Our time series have unevenly spaced observations due to weekends and holidays, so the imputed points m_j between the $(j-1)$th and jth consecutive observations are carefully chosen such that $\delta_t = (t_j - t_{j-1})/m_j$ is constant. We choose vague proper priors, e.g. using the Gaussian distribution $\mathcal{N}$ for model parameters $\{\theta, \mu, \sigma^2\}$:

$$\log\theta \sim \mathcal{N}(0,4) \;\;\&\;\; \log\sigma^2 \sim \mathcal{N}(-1,4) \;\;\&\;\; \mu \sim \mathcal{N}\left(\left(-12,-12,3\right)^T, 4\,I_3\right).$$

We combine the approach proposed by G. O. Roberts & O. Stramer [43] in Appendix B.2 and Theorem 4.10 for the case of the LE metric, while use Algo-

rithm 4.17 with time change for the case of the AI metric. MCMC samples based on 10^5 iterations with different values of δ_t were collected after a burn-in period of 4000 iterations and kernel density estimations are depicted in Figure 4.11.

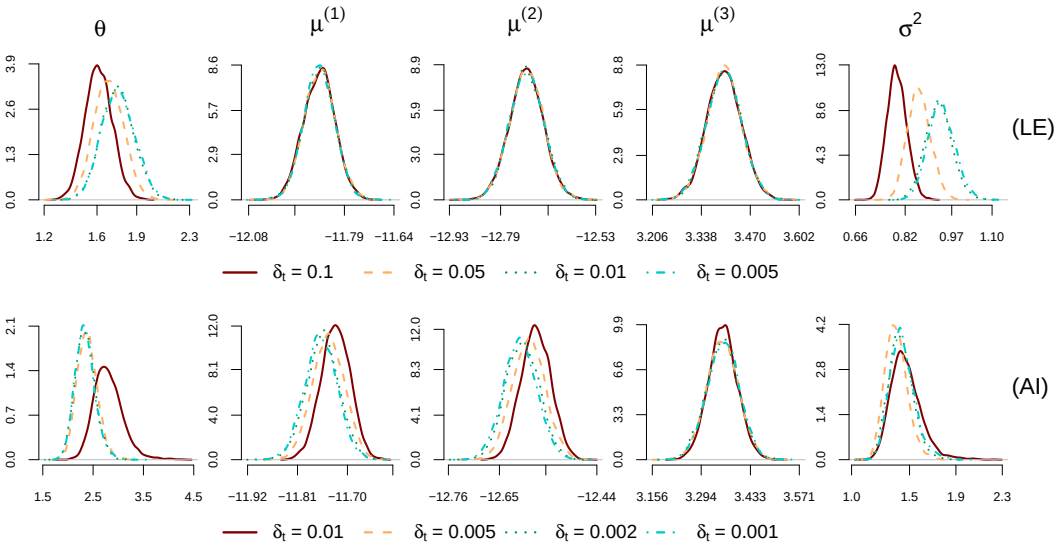


Figure 4.11: Estimated posterior distribution for $\{\theta, \mu, \sigma^2\}$ from the financial data example using either the LE (top row) or the AI metric (bottom row), based on 10^5 MCMC iterations (4×10^3 burn-in discarded, thinned by 19). Moreover, the result for $M = \mathfrak{h}^{-1}(\mu)$ is given in Figure A.7 (Appendix).

Fewer imputed points are required for the LE metric than for the AI metric to adequately approximate the posterior densities which we attribute to different degrees of non-linearity: after transformation through the matrix logarithm, the LE problem is reduced to a linear problem whereas with the AI metric, discretization including approximation of the horizontal lift takes place in the original domain. This domain can be seen to be less linear from the fact that covariance matrices are commutative under the logarithm product which is at the heart of the LE metric whereas exchanging the order of multiplication causes a different result for the AI metric. This difference is more pronounced near the boundary of the cone which is where the majority of our observations lie.

Let us illustrate this with a simple example. By forward-simulating the estimated OU process through the exponential adapted Euler-Maruyama method (using the posterior mean for $\{\theta, M\}$ and $\sigma^2 = 0$), we plot the ratio of corresponding eigenvalues when using either the LE or the AI metric in Figure 4.12. Here, we

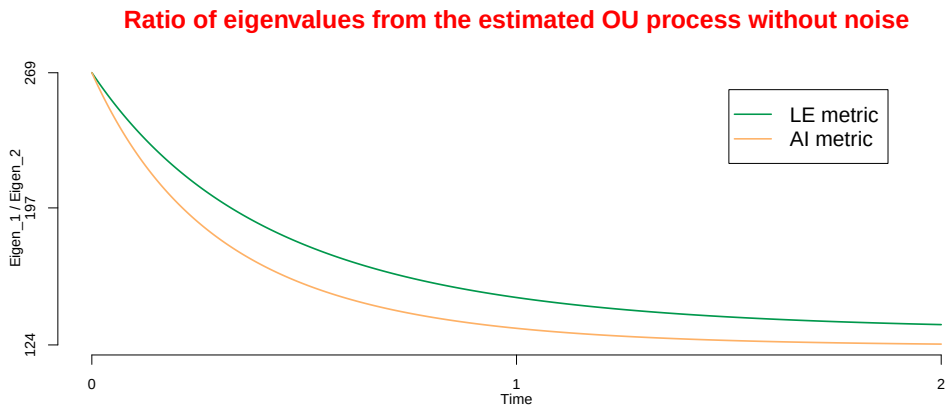


Figure 4.12: Forward-simulating the estimated OU process through the exponential adapted Euler-Maruyama method (using the posterior mean for $\{\theta, M\}$ and $\sigma^2 = 0$), we compare the ratio of corresponding eigenvalues when using either the LE or the AI metric. Here, the ratios of eigenvalues of the estimated posterior mean for M are approximately 132 and 124 in the LE and the AI case respectively.

simply start the OU process from the first observed covariance matrix, and simulate at high frequency for two units of time. This emphasizes again the difference between the LE and the AI metrics that the OU process associated with the AI metric provides less anisotropy, since the larger eigenvalues dominate more when using the LE metric, which has been discussed in Subsection 4.2.3 and 4.6.1. In other words, we get much higher NDX variations compared to the COMP variation in the model with LE metric. In fact, the ratios of eigenvalues of the estimated posterior mean for M associated with the LE metric and the AI metric are approximately 132 and 124 respectively, while the estimated posterior means for θ equal to 1.75 in the model with the LE metric and 2.34 in the model with the AI metric. And we observe the difference in Figure 4.12 due to two reasons: the geometry of geodesics associated with two different Riemannian metrics and the mean reversion speed parameter θ being different. While the LE metric has geodesics that have quite large anisotropy, the AI metric has geodesics that pass through points of low anisotropy. Since our data always has strong correlation (e.g. imputed matrices with descreased anistropy don't fit into the data), in order to compensate for this, the AI model yields a larger θ so that trajectories are quickly pulled away from the geodesics (with its low anisotropy points) and towards the attractive point M.

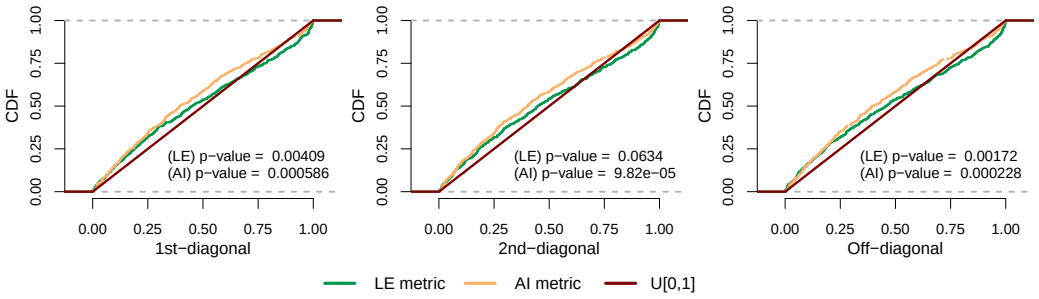


Figure 4.13: Empirical cumulative distribution of the generalized residuals for the entries from volatilities under two models using either the LE and the AI metric, in which model parameter estimates are the posterior mean. Moreover, p-values are obtained from the K-S tests of comparison to $U[0,1]$ (red line).

Next, we test the fit of the two models using the transition density-based approach by Y. Hong & H. Li [44]. For each model, we choose the posterior mean as the point estimates for $\{\theta, M, \sigma^2\}$ and compute the generalized residuals $Z_j^{(i)}$, for $1 \leq j \leq 504$, by

$$Z_j^{(i)} = \int_{-\infty}^{y_j^{(i)}} p^{(i)}(t_j, v | t_{j-1}, y_{j-1}) dv. \tag{4.41}$$

where $\{y_j^{(i)}\}_{j=0,\, i=1}^{504,\, 3}$ denote the observations, $p^{(i)}$ are the marginal transition densities and $i = 1, 2, 3$ denotes the two diagonal and the off-diagonal entries respectively. The integral in equation (4.41) is estimated by simulating $k = 3000$ points at t_j via equation (4.23) starting from y_{j-1} at time t_{j-1}. Under the null hypothesis that the observations come from the model, the realized generalized residuals $\{Z_j^{(i)}\}_{j=0}^{504}$ are i.i.d and follow the standard uniform distribution $U[0,1]$ for all $i \in \{1, 2, 3\}$. The empirical cumulative distribution of these $Z_j^{(i)}$ are shown in Figure 4.13 with p-values obtained from the K-S test when comparing to $U[0,1]$.

Figure 4.13 clearly indicates that the model using the LE metric fits the data better than the one using the AI metric for this particular dataset.

4.8 Multivariate stochastic volatility models.

In many real applications, covariance matrices are not observed from a direct source, and usually are estimated from other quantities. For instance, Section 4.7 illustrates an example, where we estimate covariance matrices from high-frequency observations of the price indices. It is therefore more natural to construct a financial time

series model for the log return of multiple underlying assets $Y_t \in \mathbb{R}^n$ for $n > 1$. We assume that Y_t is the solution of the following system of SDEs:

$$dY_t = a(t, Y_t)\, dt + \sqrt{X_t}\, dW_t \qquad (Y_0 = \mathfrak{y}), \tag{4.42}$$

$$dX_t = \theta\, \mathrm{Log}_{X_t} M\, dt + F_{X_t}(\sigma\, dB_t) \qquad (X_0 = U). \tag{4.43}$$

Here, X_t is assumed to follow the OU class introduced in Section 4.4, while W_t and B_t are independent standard Euclidean Brownian motion on $\mathbb{R}^n$ and $\mathbb{R}^d$ respectively with $d = n(n+1)/2$. Moreover, the drift of the SDE (4.42) is assumed to satisfy conditions in Theorem 2.37.

Observing volatilities only indirectly, i.e. through the price process, has been studied intensively in the univariate setting, see [134], while extension to higher dimensions has been considered less frequently. For example, K. Kalageropoulos et al. [135] propose time change transformation for Bayesian inference and assume the diffusivity in the SDE (4.42), denoted instead by Σ_t, to be a lower triangle matrix, which yields the Cholesky decomposition for $\Sigma_t\, \Sigma_t^T$. Our proposal models expand the choice for the diffusivity of the SDE (4.42), while incorporating the fact that $\Sigma_t\, \Sigma_t^T$ lies on a Riemannian manifold. But the time change transformation introduced in [135] is now no longer applicable, we therefore explore other options.

In this thesis, we put an extra assumption on the drift function $a(t, Y_t)$ such that the SDE (4.42) has explicit solution and the transition probability density is given in an analytic form in terms of X_t. This assumption implies the approximation of the likelihood can be purely computed by high-frequency observations of X_t. We continue the work with a choice of $a(t, Y_t)$ in the following form:

$$a(t, Y_t) = -\kappa Y_t \quad \text{for some } \kappa > 0 \;\Rightarrow\; Y_t = Y_0\, e^{-\kappa\, t} + \int_0^t e^{-\kappa(t-r)} \sqrt{X_r}\, dW_r$$

and the transition probability density $p(s, Y_s; t, Y_t)$ for $0 \le s < t$ is simply the density of the Gaussian distribution $\mathscr{N}\Big(Y_s\, e^{-\kappa(t-s)}\,,\, \int_s^t e^{-2\kappa(t-r)}\, X_r\, dr\Big)$, i.e.

$$p(s, Y_s; t, Y_t) = \mathscr{L}\Big(Y_t\,;\, Y_s\, e^{-\kappa(t-s)}\,;\, \int_s^t e^{-2\kappa(t-r)}\, X_r\, dr\Big). \tag{4.44}$$

4.8.1 Parameter estimation

Let us denote the model parameters by $\Theta = \{\kappa, \theta, M, \sigma^2\}$. From Equation (4.44), it is straightforward to estimate κ through the Metropolis-Hasting algorithm, that is to alternate updating κ and X_t. The tricky part is to find a good proposal for X_t because such a proposal has to be consistent with the model parameters θ, M and especially σ^2, and any proposed path for X_t must also match the observations whose distribution involves a path-integral of the process X_t. Naturally, the algorithms used in Section 4.7 are not applicable because we do not have observations of X_t. There are some works making use of algorithms based on diffusion bridges of the hidden process on the Euclidean space, see for examples [136, 137, 117, 138]. They provide appropriate proposal processes that depend on observations of the primary process Y_t, which are aimed to improve the MCMC convergence and mixing when using a block updating scheme. However, in our case the hidden process lies on a Riemannian manifold and to come up with similar proposal processes to those on the Euclidean space is hard. We made some attempts by proposing independent samplers for the target diffusion bridge on $\mathcal{SP}(2)$, however we encounter MCMC convergence and mixing problems and we expect these problems to be worsen as dimension n increases. While the literature for hidden Markov models (HMMs) is rich, such as particle filtering MCMC and sequential Monte Carlo methods [139, 140, 141, 142], we however will consider these approaches to future works because of their complexity.

Since the naive simulation of X_t is highly correlated to σ and can lead to extremely low acceptance rate. To overcome this problem, a common approach is to transform X_t to a diffusion process with constant diffusivity and direct sampling of this process should be doable, e.g. standard Brownian motion. This approach is widely used in the Euclidean space, see examples [43, 135, 143]. In our work, since we employ the exponential adapted Euler-Maruyama method to approximate X_t, conditional on Θ there is a one-to-one relationship between the sample path X_t and the path of the driving Brownian motion B_t over $[0, T]$ for some positive constant T. Thus, we can obtain B_t corresponding to some path X_t by reversing Equation (4.23)

assuming that that value of X_0 and B_0 are fixed, i.e.

$$\sum_{j=1}^{d} \big(B_{t+\delta_t}^{(j)} - B_t^{(j)}\big) E_j(X_t) = \sigma^{-1}\Big(\mathrm{Log}_{X_t} X_{t+\delta_t} - \theta\, \mathrm{Log}_{X_t} M\, \delta_t\Big) \qquad \text{for } \delta_t > 0,$$

$$\Rightarrow B_{t+\delta_t}^{(j)} = B_t^{(j)} + \sigma^{-1}\, g_{X_s}\Big(\mathrm{Log}_{X_t} X_{t+\delta_t} - \theta\, \mathrm{Log}_{X_t} M\, \delta_t\,,\, E_j(X_t)\Big), \tag{4.45}$$

where $1 \leq j \leq d$, which results from the assumption that $\{E_j\}_{j=1}^d$ is an orthonormal basis field with respect to the metric tensor g. The aim is to propose B_t as similar as those coming from the target probability distribution $p(B_t \,|\, \Theta, Y_t)$. We know that B_t is an $\mathbb{R}^d$-Brownian motion, but proposing an independent Brownian motion, i.e. no information of Y_t included, can result in low acceptance rate, which can escalate as dimension n increases. We therefore suggest using the Metropolis-adjusted Langevin algorithm (MALA) to construct an infinite-dimensional MCMC sampler for the $\mathbb{R}^d-$Brownian motion B_t, which uses knowledge about observations of Y_t. This approach follows similar lines to the work of sampling nonlinear diffusion bridges on the Euclidean space in [144].

Suppose that we have low-frequency observations $\mathfrak{D} = \{Y_{t_0} = y_0, Y_{t_1} = y_1 \ldots, Y_{t_N} = y_N\}$ at time points $\{t_0 = 0, t_1, \ldots, t_N = T\}$, and a fixed starting point for volatility $X_0 = U$. Without loss of generality, we also assume the Brownian motion driving the volatility process X_t starts from zero, i.e. $B_0 = 0$.

Let $B_{[0:T]}$ denote the path-valued of the Brownian motion B_t on $\mathbb{R}^d$ for $t \in [0, T]$ and π denote the target probability distribution $p(B_{[0:T]} \,|\, \Theta, \mathfrak{D})$. We suppose that when discretizing $B_{[0:T]}$ into $\mathscr{M} + 1$ points, the function $\pi(x) \propto \exp\big(x^T \mathscr{A}\, x/2 - \Psi(x)\big)$ for some function $\Psi : \mathbb{R}^{\mathscr{M}+1} \to \mathbb{R}$ and some matrix $\mathscr{A} \in \mathcal{GL}(\mathscr{M} + 1)$ is symmetric negative definite such that $\nabla \log \pi(x)$ is continuously differentiable. The path-valued Langevin stochastic partial differential equation (SPDE) of $Z(u, t)$ with $u \in [0, \infty)$ and $t \in [0, T]$, is given by

$$\partial Z_u = \big(\mathscr{A}\, Z_u - \nabla \Psi(Z)\big)\, \partial u + \sqrt{2}\, \partial R_u, \tag{4.46}$$

where R_u is an $\mathscr{M}$-dimensional Brownian motion on the Euclidean space. Moreover,

realizations of solutions of the SDE (4.46) in a limiting sense will provide samples of paths $B_{[0:T]}$ from the target density π. The conditions on Ψ to guarantee ergodicity is provided in [145, 146]. Here, we denote u the time from the Langevin SPDE, and we distinguish it from the real time t that comes from the driving Brownian motion B_t in Equation (4.43).

The resulting MCMC sampler requires an implicit approximation scheme parametrized by $\beta \in [0, 1]$. In this work, we use the implicit Euler method, i.e.

$$\begin{aligned} Z_{u+\delta_u} &= Z_u + \mathscr{A}\big(\beta Z_{u+\delta_u} + (1-\beta)Z_u\big)\,\delta_u - \nabla\Psi(Z_u)\delta_u + \sqrt{2\delta_u}\,\tilde{\xi}, \qquad (4.47)\\ \text{or } \big(I_{\mathscr{M}+1} - \mathscr{A}\beta\delta_u\big)\,Z_{u+\delta_u} &= \big(I_{\mathscr{M}+1} + \mathscr{A}(1-\beta)\delta_u\big)\,Z_u - \nabla\Psi(Z_u)\delta_u + \sqrt{2\delta_u}\,\tilde{\xi}, \end{aligned}$$

with $\tilde{\xi} = \begin{pmatrix} 0 & \xi \end{pmatrix}$ and $\xi \in \mathscr{N}(0, I_{\mathscr{M}})$. In fact, it has been shown that the algorithm is well defined only if $\beta = 1/2$, such as providing the correct quadratic variation, see [144]. Thus, throughout this work we set $\beta = 1/2$. Furthermore, although the stationary distribution is converged to more quickly for larger δ_u, increasing this time step also increases bias. In this case, we use the Metropolis-Hastings correction, that is if we propose a move from the current location B_t to $\tilde{B}_t$, we accept this proposal with the following probability

$$\begin{aligned} \alpha^{(B)} &= \frac{\pi\big(\tilde{B}_t\big)\,q\big(B_t \,|\, \tilde{B}_t\big)}{\pi(B_t)\,q\big(\tilde{B}_t \,|\, B_t\big)} \qquad \text{with}\\ q\big(\tilde{B}_t \,|\, B_t\big) &= \mathscr{L}\Big(\big(I_{\mathscr{M}+1} - \frac{\delta_u}{2}\mathscr{A}\big)\tilde{B}_t\,;\,\big(I_{\mathscr{M}+1} + \frac{\delta_u}{2}\mathscr{A}\big)\,B_t - \nabla\Psi(B_t)\,\delta_u\,;\,2\,\delta_u\Big) \end{aligned} \qquad (4.48)$$

In practice, the Brownian motion path B_t is discretized in the t-direction. Suppose that we discretize $B_{[0,T]}$ to $\{B_l\}_{l=0}^{\mathscr{M}}$ at equi-distant time points $\tilde{t}_l$ such that $\mathscr{M} = \sum_{j=1}^{N} m_j$. These time points are set out to be the same as Algorithm 4.17, that is to impute suitable $m_j - 1$ data points between the jth consecutive observations. To simplify the notation, we assume that our observation time are at equi-distant in time, i.e. $t_j - t_{j-1}$ equals a constant Δ_t for all $0 \le j \le N$. In this case the same number of imputed points between consecutive observations is required, denoted by

m, and thus $t_j = \tilde{t}_{jm}$ for all $1 \leq j \leq N$ and $\mathscr{M} = Nm$. The time-width in the t-direction of the path-valued for B_t therefore equals $\delta_t = T/\mathscr{M}$.

Next, we compute the target probability distribution π by discretization, that is

$$\pi(B_{[0:T]}) \propto p\big(B_{[0:T]}\big) \prod_{j=1}^{N} p\big(y_j \,|\, y_{j-1}, \Theta, B_{[t_{j-1}:t_j]}\big). \tag{4.49}$$

Firstly, discretizing $B_{[0:T]}$ we get

$$\begin{aligned} p\big(B_{[0:T]}\big) &\approx \prod_{l=1}^{\mathscr{M}+1} \mathscr{Z}\big(B_l \,;\, B_{l-1} \,;\, \delta_t\big) \propto \exp\left\{ \sum_{l=1}^{\mathscr{M}+1} -\frac{(B_l - B_{l-1})^T (B_l - B_{l-1})}{2\,\delta_t} \right\} \\ &= \exp\Big\{ \frac{1}{2} \sum_{i=1}^{d} (B^{(i)})^T \mathscr{A} B^{(i)} \Big\}, \end{aligned}$$

$$\text{with} \quad \mathscr{A} = \frac{1}{\delta_t} \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & -2 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & -1 \end{pmatrix} \in \mathcal{GL}\big(\mathscr{M}+1\big) \tag{4.50}$$

Then, we reverse the exponential adapted Euler-Maruyama method applied to the SDE (4.43), approximate the Riemannian integral in Equation (4.44), and get

$$\begin{aligned} p\big(y_j \,|\, y_{j-1}, \Theta, B_{[t_{j-1},t_j]}\big) &= \mathscr{Z}\Big(y_j \,;\, y_{j-1} e^{-\kappa\Delta_t} \,;\, \int_{t_{j-1}}^{t_j} e^{-2\kappa(t_j - r)} X_r dr\Big) \\ &\approx \mathscr{Z}\Big(y_j \,;\, y_{j-1} e^{-\kappa\Delta_t} \,;\, \sum_{l=(j-1)m}^{jm-1} e^{-2\kappa(t_j - \tilde{t}_l)} X_l\, \delta_t\Big) \\ &\propto \exp\Big\{ -\frac{1}{2} \log\det \mathscr{C}_j \\ &\qquad -\frac{1}{2} (y_j - y_{j-1} e^{-\kappa\Delta_t})^T \mathscr{C}_j^{-1} (y_j - y_{j-1} e^{-\kappa\Delta_t}) \Big\}, \end{aligned}$$

where $\mathscr{C}_j = \sum_{l=(j-1)m}^{jm-1} e^{-2\kappa(t_j - \tilde{t}_l)} X_l\, \delta_t \in \mathcal{SP}(2)$. Additionally, X_l is computed

using Equation (4.23), that is for $X_0 = U$ and $l \geq 1$:

$$X_l = \mathrm{Exp}_{X_{l-1}} \left\{ \theta\, \delta_t\, \mathrm{Log}_{X_{l-1}} M + \sum_{i=1}^{d} (B_l^{(i)} - B_{l-1}^{(i)})\, \sigma\, E_i(X_{l-1}) \right\}$$

where the exponential map Exp, logarithm map Log and the orthonormal basis frame $\{E_i\}_{i=1}^d$ depend on the chosen metric tensor. We then set

$$\Psi\big(B_{[0:T]}\big) = \frac{1}{2} \sum_{j=1}^{N} \Bigg(\log\det \mathscr{C}_j + (y_j - y_{j-1} e^{-\kappa\Delta_t})^T\, \mathscr{C}_j^{-1}\, (y_j - y_{j-1} e^{-\kappa\Delta_t}) \Bigg), \tag{4.51}$$

and differentiate the function Ψ with respect to B_l for $1 \leq l \leq \mathscr{M}$,

$$\Rightarrow 1 \leq i \leq d \;\; : \;\; \frac{\partial \Psi}{\partial B_l^{(i)}} = \frac{1}{2} \sum_{j=1}^{N} \mathbb{I}_{\{(j-1)m \leq l < jm\}} \Bigg(\mathrm{tr}\Big(\mathscr{C}_j^{-1}\, \frac{\partial \mathscr{C}_j}{\partial B_l^{(i)}} \Big) - (y_j - y_{j-1} e^{-\kappa\Delta_t})^T\, \mathscr{C}_j^{-1}\, \frac{\partial \mathscr{C}_j}{\partial B_l^{(i)}}\, \mathscr{C}_j^{-1}\, (y_j - y_{j-1} e^{-\kappa\Delta_t}) \Bigg) \tag{4.52}$$

As we fix $B_0 = 0$, $\partial\Psi / \partial B_0^{(i)} = 0$ for all $1 \leq i \leq d$. And since $\mathscr{C}_j$ contains matrix exponential and possibly matrix logarithm (i.e. in the AI case), there is no explicit formula for the derivative $\partial \mathscr{C}_j / \partial B_l^{(i)}$. Instead, we use a first order approximation for these derivatives, e.g.

$$\frac{\partial \mathscr{C}_j}{\partial B_l^{(i)}} \approx \mathbb{I}_{\{(j-1)m \leq l < jm\}} \Big(e^{-2\kappa(\tilde{t}_{jm} - \tilde{t}_l)} \delta_t \Big)\, \frac{X_l\big(B_l^{(i)} + \mathfrak{s}\big) - X_l\big(B_l^{(i)}\big)}{\mathfrak{s}},$$

with an infinitesimally small $\mathfrak{s} > 0$.

In summary, the essence of the MALA in our work is based on the following steps. Firstly, given the current location of $B_{[0:T]}$, we propose the next move through Equation (4.47) and accept this move with probability $\alpha^{(B)}$ in Equation (4.48). Here, the gradient of Ψ is approximated in Equation (4.52). Next, we alternate updating each model parameter for Θ using random walk (e.g. symmetric proposal) and use the usual Metropolis–Hastings algorithm based on the likelihood function π, which

are given in terms of the matrix $\mathscr{A}$, see Equation (4.50) and the function Ψ, see Equation (4.51).

4.8.2 Simulation study

4.8.2.1 Log-Euclidean metric

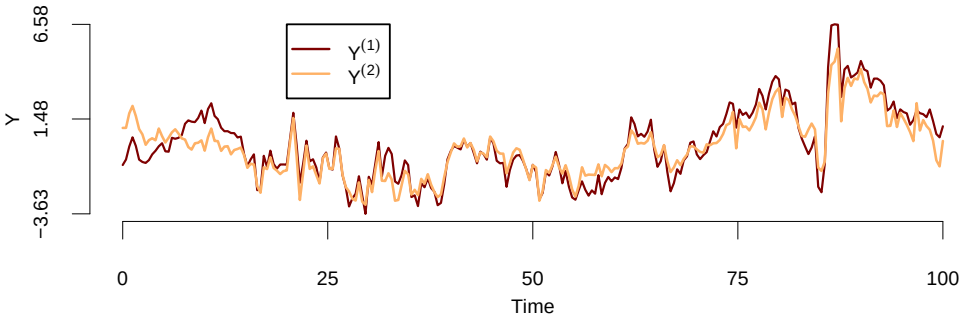


Figure 4.14: Time series of the log return of underlying assets Y_t simulated from Equation (4.42) and (4.43) associated with the LE metric.

We simulate $10^6 + 1$ equidistant time points X_t of the OU process associated with the LE metric using the SDE (4.43) on $[0, 100]$, then use these path-valued of X_t to simulate Y_t via the SDE (4.42). Here, we set the model parameters as follows:

$$\kappa = 0.2 \ \& \ \theta = 0.5 \ \& \ \sigma^2 = 1 \ \& \ X_0 = M = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 0.8 \end{pmatrix} \ \& \ Y_0 = \begin{pmatrix} -1 & 1 \end{pmatrix}.$$

We take sub-samples at time points $\{0, 0.4, \ldots, 100\}$. Due to the choice of the covariance matrix M, it is not difficult to see that $Y^{(1)}$ is highly correlated with $Y^{(2)}$, see Figure 4.14. We apply MALA in Subsection 4.8.1 and assume the prior distributions $\log \kappa,\ \log \theta,\ \log \sigma^2 \sim \mathscr{N}(0, 4)$ and $\mu = \mathfrak{h}(M) \sim \mathscr{N}(0, 4 \times I_3)$, where $\mathfrak{h}$ is defined in Equation (4.19).

Firstly, we vary the number of imputed point m and initialize different values for model parameters $\{\kappa,\ \theta,\ \sigma^2,\ \mu\}$, then we run 25×10^4 MCMC iterations. The left two columns of Figure 4.15 illustrate traceplots of the first 15×10^4 iterates when $m = 2,\ 4$, and they clearly show that the chains associated with different starting points for all parameters will eventually overlap. Moreover, we present the ACF plots based on 15×10^4 burn-in and 10^5 MCMC iterations on the right

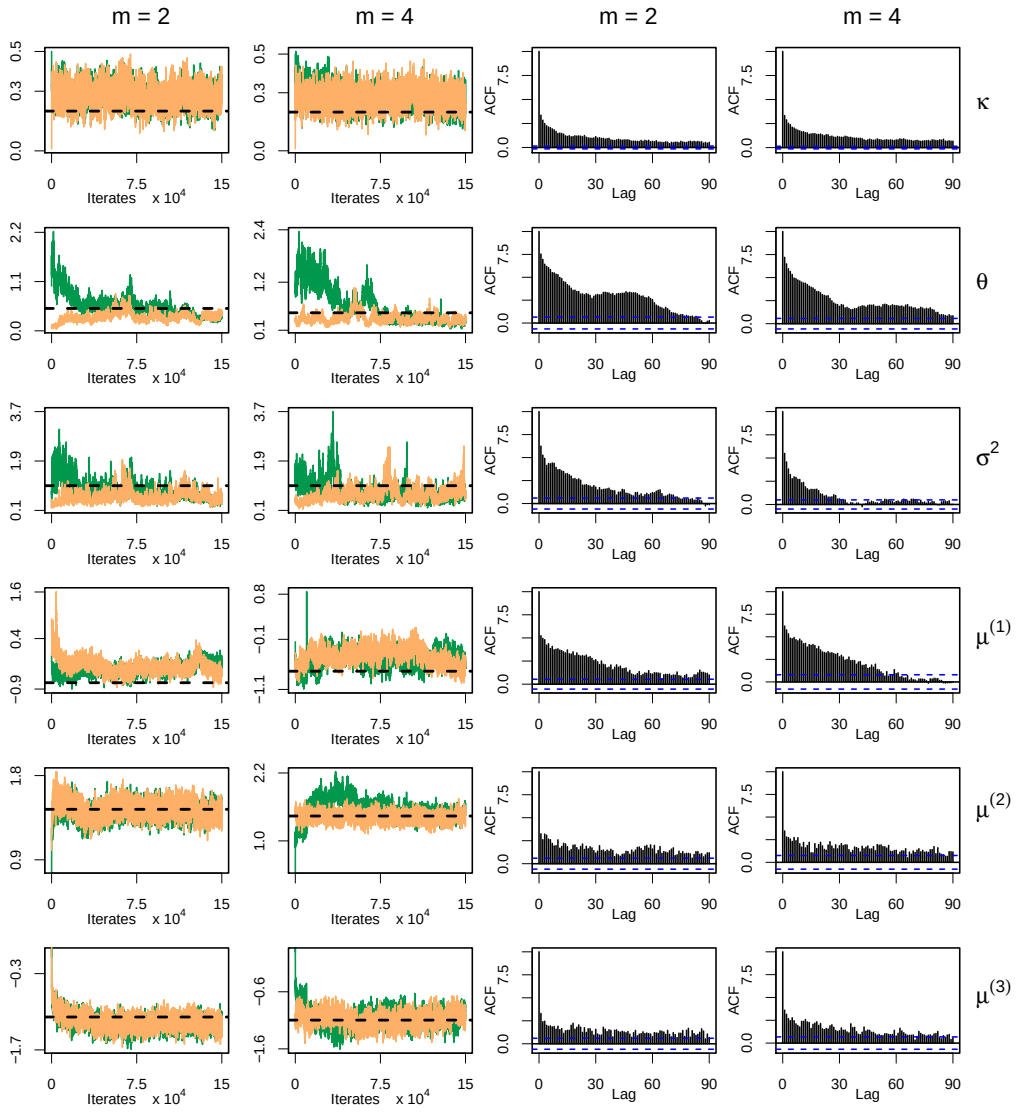


Figure 4.15: Left two columns: Traceplots from 15×10^4 iterations, where true values are indicated with the black dashed lines. Right two columns: ACF plots based on the MCMC iterates 15×10^4 - 25×10^4 (thinned by 5).

two columns of Figure 4.15 when $m = 2, 4$. These indicates that increasing m does not affect the mixing of the chain and improves, in some cases (e.g. μ), the approximation to the marginal densities. Following up, we run a longer MCMC chain of 5×10^5 iterations while varying the value of m over $2,\ 4,\ 10$ and 20. These chains are thinned out after 10^5 iterations and samples of 5000 points are collected from the target distributions. Figure 4.16 depicts the kernel density estimations of the marginal posterior distributions of all parameters are approximately the same for

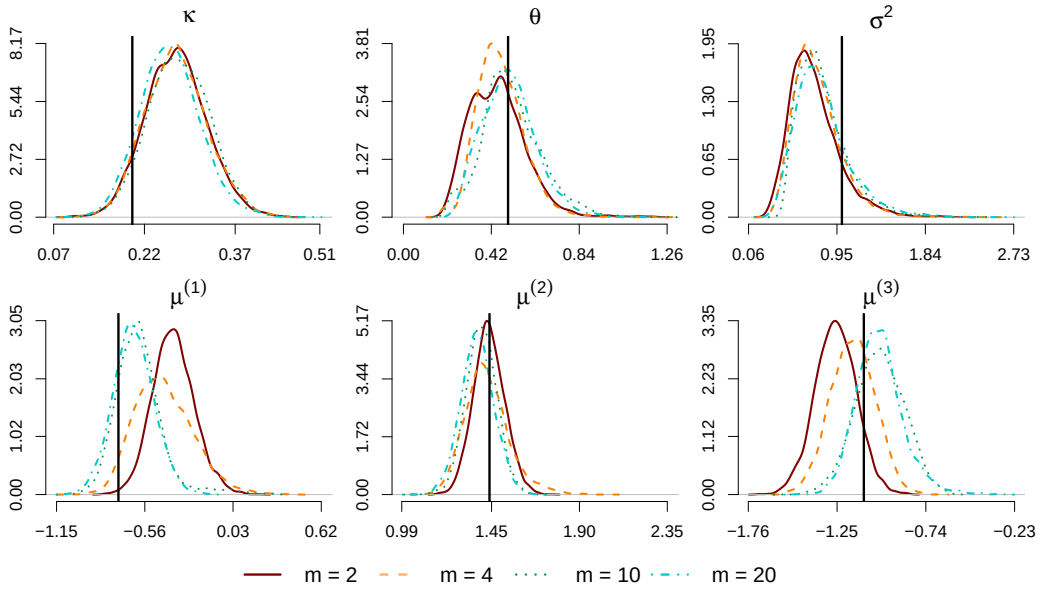


Figure 4.16: Estimated posterior distribution of $\{\kappa, \theta, \sigma^2, \mu\}$ using 5×10^5 MCMC iterations (15×10^4 burn-in discarded, thinned by 60). True values are indicated by solid vertical black lines.

$m = 10$ and $m = 20$. In this case, $m = 10$ is considered to provide a sufficiently fine discretization for these synthetic data.

Similarly to the validation for the guided proposal algorithm in Subsection 4.6.2, we also perform the prior reproduction test on the estimation for $\{\kappa, \theta, \sigma^2, \mu\}$, see Algorithm 4.19. We assume proper priors for the model parameters:

(i) $\log \kappa \sim \mathcal{N}(-1.5, 0.1)$, $\log \theta \sim \mathcal{N}(-1, 0.1)$ and $\log \sigma^2 \sim \mathcal{N}(0, 0.1)$.

(ii) $\mu^{(1)} \sim \mathcal{N}(-0.7, 0.2)$, $\mu^{(2)} \sim \mathcal{N}(1, 0.2)$ and $\mu^{(3)} \sim \mathcal{N}(-0.7, 0.2)$.

We then generate, in turn, $l = 500$ samples from the prior distributions and conditional on each sampled parameter vector, high-frequency observations on $[0, 100]$ at 10^6+1 equidistant time points and keep sub-samples at time points $\{0, 0.4, \ldots, 100\}$. For the 500 generated data sets, we estimate the corresponding marginal posterior distributions and randomly draw one sample. If the estimation algorithm works properly, we expect these collection of samples follows the corresponding prior distributions. Figure 4.17 illustrates that the replication of prior distributions will improve as m increases. We also want to emphasize that even when $m = 4$, it may not offer a sufficiently fine discretization, see for example in Figure 4.16.

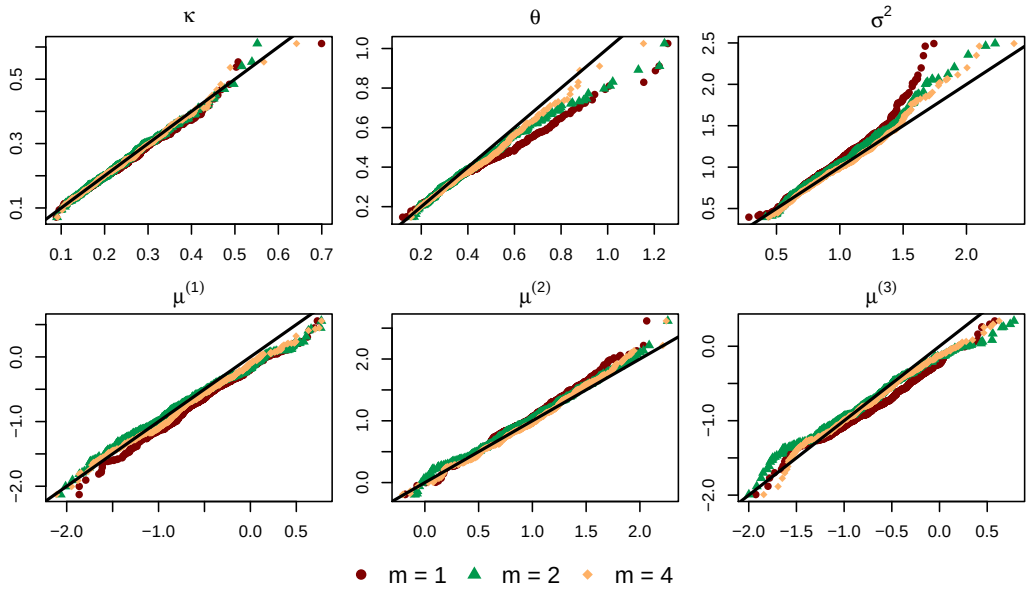


Figure 4.17: Prior reproduction test for the estimation of $\{\kappa, \theta, \sigma^2, \mu\}$ in the case of the LE metric: Q-Q plots of priors against posteriors from these parameters.

Although result from Theorem 4.10 implies when working with the LE metric there is an explicit formula for the transition density in the log-domain, here we illustrate that it is possible to retrieve the true parameters by discretizing the SDE (4.43) finely enough to get the approximation for the likelihood. Clearly knowing the true transition density will improve the bias of likelihood, but in general it is not always the case and it is therefore desirable to test the estimation algorithm with a more complicated scenario, and so we investigate the model with the AI metric next.

4.8.2.2 Affine-Invariant metric

Besides requiring higher computational cost, the impact of large time width on the likelihood approximation when using the exponential adapted Euler-Maruyama method associated with the AI metric is higher compared with the LE metric. We expect the MALA in Subsection 4.8.1 to work in a similar manner for the case of the AI metric but to require significantly increased computational resources. Therefore, we assume that only the parameter σ^2 is unknown, while others are assumed to be available. We wish to estimate the marginal posterior density for σ^2. Moreover, since the number of imputed points m tends to be larger (e.g. Section 4.7) to offer a good approximation for likelihood in the AI case, we retain the same number of

observations, i.e. $N = 250$, but halve the terminal time, i.e. $T = 50$, thus $\Delta_t = 0.2$.

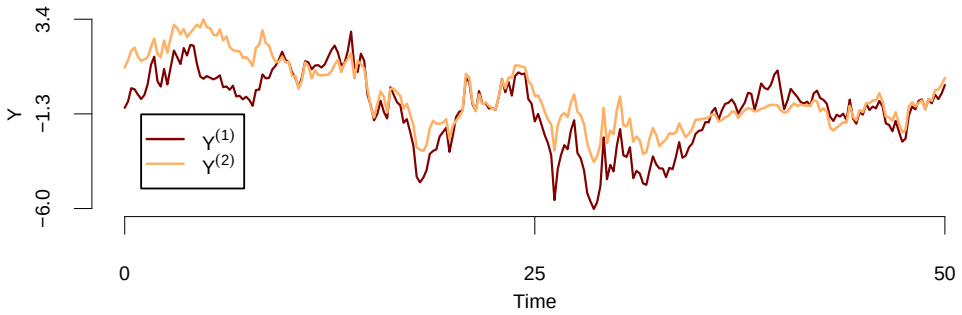


Figure 4.18: Time series of the log return of prices Y_t simulated from Equation (4.42) and (4.43) using the AI metric.

We simulate $10^6 + 1$ equidistant time points X_t of the OU process associated with the AI metric using the SDE (4.43) on $[0, 50]$, then use these paths of X_t to simulate Y_t via the SDE (4.42). Here, we set the model parameters as follows :

$$\kappa = 0.2 \ \& \ \theta = 0.5 \ \& \ \sigma^2 = 0.7 \ \& \ X_0 = M = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 0.8 \end{pmatrix} \ \& \ Y_0 = \begin{pmatrix} -1 & 1 \end{pmatrix}.$$

We take sub-samples at time points $\{0, 0.2, \ldots, 50\}$, and illustrate these data in Figure 4.18. Here, we assume the prior distributions $\log \sigma^2 \sim \mathcal{N}(0, 4)$.

We initialize σ^2 with different values and run 25×10^4 iterations when $m = 2, 4$. The left two columns of Figure 4.19 indicate that increasing m improves the approximation of the likelihood and does not worsen the mixing of the chain. We also plot the estimated kernel density for the posterior distribution for σ^2 using a longer MCMC chain of 5×10^5 iterations (15×10^4 burn-in discarded and thinned by 60), where m is varied over $2, 4, 10, 20$. The right column of Figure 4.19 suggest $m = 10$ is considered to provide a sufficiently fine discretization in this case.

Lastly, we perform the prior reproduction test (Algorithm 4.19) to validate the estimation algorithm for σ^2. We assume proper prior for σ^2: $\log \sigma^2 \sim \mathcal{N}(-0.2, 0.07)$ and generate, in turn, 500 samples from the prior distribution. Conditional on each sample, we simulate high-frequency observations on $[0, 50]$ at $10^6 + 1$ equidistant time points and take sub-samples at time points $\{0, 0.2, \ldots, 50\}$. Figure 4.20 indicates that the distributions of the samples that were collected during

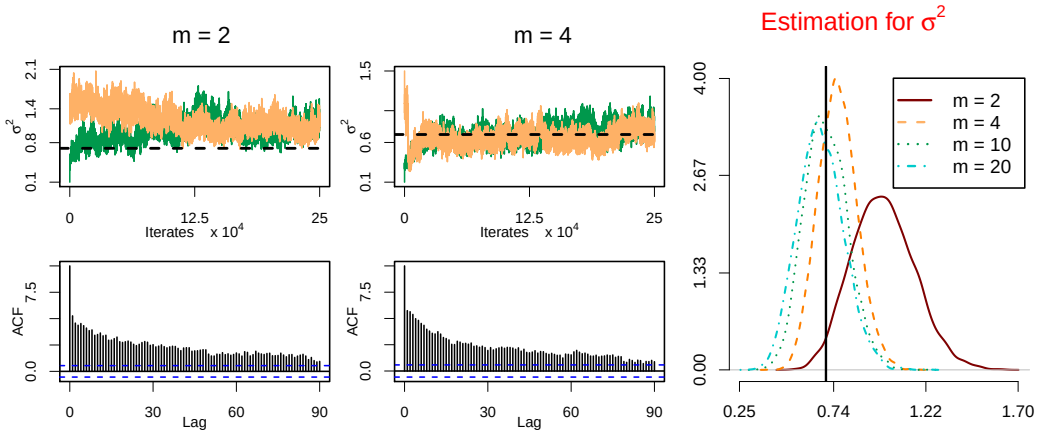


Figure 4.19: Left two columns: Traceplots and ACF plots from 25×10^4 iterations (thinned by 5 for ACF plots), where the true values are indicated with the black dashed lines. Right column: Estimated posterior distribution of σ^2 using 5×10^5 MCMC iterations (15×10^4 burn-in discarded, thinned by 60), where the true values is indicated by a solid vertical black line.

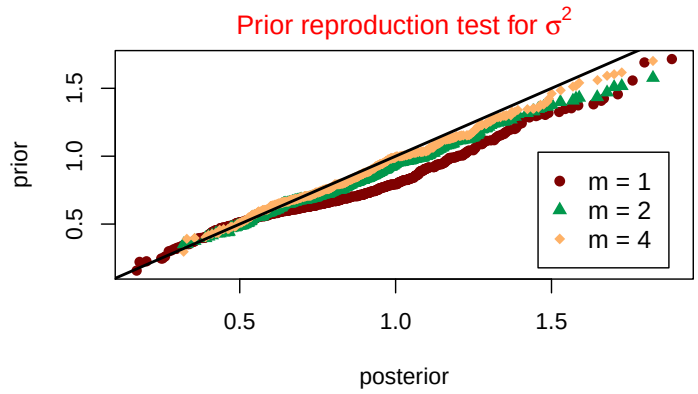


Figure 4.20: Prior reproduction test for the estimation of σ^2 in the case of the AI metric: Q-Q plots of priors against posteriors of σ^2.

the prior reproduction test resemble the prior distribution more as m increases.

4.8.3 Discussion

In brief, we perform MALA in order to estimate the model parameters of the SDE (4.42) and (4.43). Because working with the AI metric requires substantial computational costs, we choose to estimate only the diffusivity parameter σ^2 in the SDE (4.43), and the rest are assumed to be available. For both metrics, we are able to retrieve the true parameters given our choices for these parameters and the time points of observations, and validate the algorithm with the prior reproduction

test. This estimation task is sometimes approximated in the literature as fitting a homogenized single-scale model to multiscale data, see [147, 148]. In this case, model parameters may not be identified and this depends on how often the data is sampled. Particularly, the optimal sampling rate depends heavily on the diffusivity parameter. Nonetheless, if observations are collected too often, it can lead to incorrect parameter inference, see [147]. At this stage of our study, we have not investigated this issue, however, this is one thing that we can look into in the future.

On the other hand, we notice that choosing suitable step sizes when updating proposal parameters in the Metrospolis-Hastings algorithm is hard. It is very laborious to get the optimal MCMC convergence rate in both cases of the LE metric and the AI metric. In this work, we improve the speed of convergence by beginning the algorithm with sampling the path-valued Brownian motion B_t from the distribution $p\big(B_{[0:T]} \,|\, \Theta_0, \mathfrak{D}\big)$. That is to fix the initial values for the model parameters and run the chain with updating only B_t for many iterations, e.g. 5×10^3 - 8×10^4 iterations in our simulation study from the previous subsection. Although the distribution that B_t is initially drawn from may not be the right one, given that the initial values for Θ may be different from their true values, this will ensure that this sampled path of B_t supports the right covariance for the data set Y_t before alternating the model parameters Θ. And after some experiments we are convinced that running these extra iterations can outweigh the additional burn-in iterations for Θ if one uses the usual Metropolis-Hastings set-up. Naturally, further improvements in convergence speed and a consequent reduction in computational cost may be possible but we will leave these as future work.

4.9 Summary

In this chapter, we revise some existing results about two Riemannian metrics on $\mathcal{SP}(n)$: the Log-Euclidean and the Affine-Invariant metrics. We show that $\mathcal{SP}(n)$ endowed with either metric is stochastically complete, that is the Riemannian Brownian motion does not explode at finite time. We then calculate horizontal lifts of smooth curves for both metrics, in which a global form is obtained in the LE

case, and in the AI case, while there is no such global form, one can approximate it in a very simple form given that the piecewise approximation method is the exponential adapted Euler-Maruyama method. Moreover, we construct the OU class on $\mathcal{SP}(n)$, that achieves the mean-reverting property by adding a drift consisting of the Riemannian gradient of the squared distance function.

Due to the presence of curvature, using local coordinates or an embedding bring more complications to the Riemannian manifold $\mathcal{SP}(n)$ endowed with the AI metric. These two approaches work well in theory, but in practise with extreme observations (i.e. near boundary of the cone), the neighbourhood covering these observations needs to be tightly restricted (i.e. Δ_t is sufficient small) so that we are guaranteed to stay on the manifold and these do not fit our problem when we have low-frequency observations. On the other hand, the naive approach of simulating bridges has high correlation with the unknown diffusivity parameters, which will worsen the MCMC convergence rate in the data-augmentation approach. Moreover, it is almost impossible to collect bridges that have end points lying extremely closely to the boundary.

We propose a novel of diffusion bridge sampler on $\mathcal{SP}(n)$ equipped with the AI metric in this thesis, which can be generalized to other Cartan-Hadamard manifolds. In this new approach, the proposal bridge comes from the guided proposals, which can be simulated more efficiently using the exponential adapted Euler-Maruyama method, and acceptance probabilities can be calculated through the approximation of the Radon-Nikodym derivative. In Theorem 4.13, we prove that measures are absolutely continuous up to some terminal time T and write down the Radon-Nikodym derivative. Although, the theorem regarding the limit as $t \to T$ is not fully proved, instead, we validate our estimation algorithm through simulation study on $\mathcal{SP}(2)$. Moreover, we fit the OU class equipped with either the LE metric or the AI metric to data arising from an application in finance, and perform a goodness-of-fit test using a transition density-based approach comparing the models associated with these two Riemannian metrics.

Furthermore, we extend our study with the multivariate stochastic volatility

model. We take advantage that our OU class are limiting processes resulting from the exponential adapted Euler-Maruyama method, so we can discretize the process finely to get the approximation for the likelihood. This allows us to apply MALA, and these estimation steps are now basically Euclidean problems. In this thesis, we assume a very simple model for Y_t, see Equation (4.42), in reality it requires a much more sophisticated model. Therefore, further study is needed in order to provide a more adequate model to use in real life.

In summary, the Euclidean metric brings simplicity and thus uses least computational resources, but it does not guarantee that the simulated path lies completely on manifold, and the swelling effect also appears. While both the LE and the AI metrics remove these two problems, complexities increase, particularly for AI metric as the curvature is no longer zero. Benefiting from the available analytic form of the global map, working with LE metric is much more efficient than with the AI metric. Even though SDEs using the LE and the AI metric are quite similar, in general they are different, in fact covariance matrices tend to be more anisotropic when using LE metric than AI metric. Thus, the choice between LE and AI metrics depends on which applications are considered.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

In this thesis, we explore various statistical inference tools on Riemannian manifolds, and extend their applicability for statistical modelling across a range of applications.

Chapter 3 presents a simple parametrization for observations, which lie on a compact connected Lie group. We map this special space intrinsically onto its tangent space at the identity (i.e. its Lie algebra) using the logarithm map and the differential map of the left translation from a fixed base point. And by using coefficients with respect to a fixed orthonormal basis on this tangent space, there are no extra model parameters incurred by this parametrization, which can occur if one uses embedding. Furthermore, this approach enables existing tools on the Euclidean space to work in the manifold setting, such as GLS in regression analysis. In particular, we investigate $\mathcal{SO}(3)$ and apply our proposed approach for estimating the relationships between soft tissue artefacts and rotations of the bone during certain activities. We compare different correlation structures when using GLS for our proposed parametrization and another transformation that is commonly used in human kinematic analysis, which is defined by the Euler angles. We observe some improvements in terms of residuals, however due to lack of information about the smoothing filter that had been used on the data, while some obvious patterns in the residuals could be removed, model fit is still not quite satisfactory.

Most of our contributions lie in Chapter 4, where we extend the Ornstein-

Uhlenbeck processes to $\mathcal{SP}(n)$ such that they preserve the mean-reverting property. A novel sampling strategy for a diffusion bridge on $\mathcal{SP}(n)$ equipped with the AI metric from the intrinsic point of view is proposed in this thesis. This diffusion bridge sampler is inherited from the guided proposal approach on the Euclidean space. Although only partial mathematical proofs in showing the absolute continuity between the target diffusion bridge and the guided proposal are presented, we validate the proposed estimation algorithm via a simulation study. We further illustrate this algorithm with application in finance, and perform a goodness-of-fit test comparing models based on the LE and AI metrics.

Besides, Chapter 4 addresses the limitation of equipping $\mathcal{SP}(n)$ with the Euclidean metric (i.e. the Frobenius inner product) and we therefore recommend working with Riemannian metrics on $\mathcal{SP}(n)$. We show that the Riemannian Brownian motion on $\mathcal{SP}(n)$ equipped either the LE or the AI metric does not explode in finite time, which is the same behaviour observed on the Euclidean space. Furthermore, we establish the calculation of horizontal lifts of smooth curves and show that the local expressions of horizontal lifts are available explicitly in the case of the LE metric. It is particularly beneficial for practicability since any stochastic process on the Euclidean space can be mapped uniquely onto $\mathcal{SP}(n)$ and therefore the usual tools of statistical inference in the Euclidean setting become compatible. Moreover, we demonstrate the Metropolis-adjusted Langevin algorithm (MALA) to estimate the multivariate stochastic volatility model via simulation study on $\mathbb{R}^2$, where our Ornstein-Uhlenbeck class models the unobserved volatilities.

5.2 Future work

5.2.1 Application to human joint movement

For this topic, our suggestions on future work mainly concern regression problems on the Euclidean space itself, rather than geometric factors.

The total number of model parameters in our setting for one subject (i.e. covering twelve markers in three Euclidean components) is large given the very limited number of data subjects available, while there is while there is high dependence

among markers that stay close to each other. The natural next thing is to reduce these correlated variables. For example, one can firstly perform PCA or evolving factor analysis and then fit the regression model on only some important components. There is an attempt which applies the PCA in the Euclidean sense in [60] using in-vivo data. Alternatively, one can treat the surface containing these model parameters for a single subject as a curved surface, i.e. a Riemannian manifold and apply the principle geodesic analysis as the work in [21] for Lie groups and [149] for Riemannian manifolds.

On the other hand, R. Dumas et al [78] propose the marker-cluster transformations or the skin envelope shape variation as the replacement of coordinates of individual markers. This approach can the reduce number of model parameters as it already includes the correlation among markers that are near by. Moreover, the model for a single subject can be incorporated for a multi-body optimization, see [150], which is hugely beneficial for use in a wider population.

Lastly, based on information about the smooth filter that were applied to the observations, we can not recover the correlation structure explicitly in our work. However, it may be possible to carry out an independent low-budget experiment, that uses solid objects instead of human bodies, to construct an error model. The only requirement here is that the same systems that were used to collect our data, such as the multi-camera systems, the Butterworth filter etc. be used. More generally, a general statistical error model, perplexingly, seems absent from the literature on human joint motion modelling.

5.2.2 SDEs on $\mathcal{SP}(n)$

Firstly, we wish to complete the remaining part of the proof for Theorem 4.15 in Chapter 4, in which the proof plan is discussed comprehensively in Section 4.5. Since Cartan-Hadamard manifolds are diffeomorphic to Euclidean space, and the diffeomorphism maps can be obtained from the exponential map at any point, see [47, 151], we suspect that once the proof for $\mathcal{SP}(n)$ is complete, it is not difficult to generalize this theory to Cartan-Hadamard manifolds. Thus, our approach of sampling diffusion bridges and our proposed class of OU processes can be extended

to any Cartan-Hadamard manifold on which there exists a suitable diffusion process with an explicit transition density function, e.g. hyperbolic spaces [152, 153]. This opens up potential applications to phylogenetic trees [154] and electronic engineering [155] where this or closely related geometries are used.

On the Euclidean space, the OU process is ergodic (i.e. recurrent), and in fact this property continues to hold for the proposed OU process on $\mathcal{SP}(n)$ endowed with the LE metric. Thus the natural question will be about whether one can preserve ergodicity in the AI case, and, if so, what is the equilibrium/invariant distribution. One common and possible approach of proving this is the spectral gap analysis, see [156, 157].

In addition, the family of Gaussian graphical models (GGM) has been widely used in many application areas for learning the conditional independence structure among a collection of variables, which is the case that particular elements of the inverse of the covariance matrix are restricted to be zero, see [158, 159]. The space of covariance matrices under these restrictions turns out to be a sub-manifold of $\mathcal{SP}(n)$. Thus, it would be interesting to explore how one would equip the induced Riemannian metrics: Log-Euclidean and Affine-Invariant, to this sub-manifold, because it then can offer many benefits. For instance, the construction of a reversible jump MCMC that proposes moves between GGM, which will prevents the proposals escaping the sub-manifold (e.g. using the exponential adapted Euler-Maruyama method), may be possible.

Lastly, we perform a very simple goodness-of-fit test to compare models associated with the LE and AI metrics in Section 4.7.2. However, in the literature there are other sophisticated tests, see [160], that offer better performance, and we can experiment with them to our models. Moreover, further work on the multivariate volatility model will include working with real data (e.g. financial data in Section 4.7), or considering more complex drift of the SDEs for the price processes. We suspect that the MALA in Subsection 4.8.1 is still applicable for model parameter estimation even when the drift is non-linear. However, the sensitivity on time width may be exaggerated further because we may require to approximate the transition density

of the price process via discretization. More suitable strategies which can improve the MCMC mixing therefore become necessary. Furthermore, the homogenization problem for multiscale data, when working with our multivariate volatility models, is also an interesting field that can be explored in the future. Since homogenization theory is usually approached via convergence of the infinitesimal generator which is often available for SDEs on manifolds, there are some prospects of extending this theory to the manifold setting.

5.2.3 SDEs on other manifolds

Previously, we discussed the possibility of extending the concept of guided proposals to other Cartan-Hadmard manifolds, which do not have any cut locus. So, the next natural generalization is to explore a space that does have a cut locus. The Lie group $\mathcal{SO}(3)$, discussed in Section 3.3, is a promising curved space to study statistical inference for SDEs after $\mathcal{SP}(n)$ because some extension to a general Lie group might be possible. Moreover, $\mathcal{SO}(3)$ appears in a vast number of applications, thus the demand for time series models, e.g. stochastic processes, is always high, for example Brownian motion has been studied very early when stochastic processes on manifolds were not as popular as they are now, see for examples [24, 31, 30]. Depending on the application, one will decide whether to choose to use local coordinates, embedding or exponential maps.

Notations and Abbreviations

$\mathcal{SP}(n)$ space of $n \times n$ symmetric positive definite matrices. 19

$\mathcal{GL}(n)$ space of $n \times n$ invertible matrices. 40

$\mathcal{S}^n$ the n-sphere. 52, 97

$\mathcal{O}(n)$ space of $n \times n$ orthogonal matrices. 25, 40

$\mathcal{SO}(3)$ space of $n \times n$ rotational matrices. 19, 67

$\mathcal{S}(n)$ space of $n \times n$ symmetric matrices. 21, 102

$\mathscr{C}_p$; $\mathscr{E}_p$ the cut locus of p and the set within the cut locus of p. 37

$\mathscr{F}(\mathcal{M})$ frame bundle of $\mathcal{M}$. 50

$\mathcal{H}_u\mathscr{F}(\mathcal{M})$ space of all horizontal tangent vectors of the frame bundle $\mathscr{F}(\mathcal{M})$. 50

$T_p\mathcal{M}$; $T\mathcal{M}$ tangent space and bundle. 27

$\Gamma_\gamma(T\mathcal{M})$ set of all smooth vector fields on $\mathcal{M}$ along the curve γ. 28

$\Gamma(T\mathcal{M})$ set of all smooth vector fields on $\mathcal{M}$. 28

$\langle\cdot,\cdot\rangle_F$; $||\cdot||_F$ Frobenius inner product and norm. 29

$\langle\cdot,\cdot\rangle_2$; $||\cdot||_2$ L^2 inner product and norm. 29

$\Delta_\mathcal{M}$ Lapalace-Beltrami operator on $\mathcal{M}$. 38, 110

$\mathrm{Ad}(\mathcal{G})$ adjoint representation of the group $\mathcal{G}$. 41

$\mathcal{O}$ Big O notation. 75, 137

Γ_{ij}^{k} Christoffel symbols. 32, 110

∇ affine connection. 30

Σ^{M}; Σ^{Q} Matérn and quasi-periodic correlation functions. 86

$\mathscr{D}_t V$ covariant derivative of a vector filed V along the curve. 31

$R(X,Y)Z$; $K(X,Y)_p$; $\mathrm{Ric}(X,Y)_p$ Riemannian, sectional and Ricci curvature tensors respectively. 32

$D_P f.S$ derivative of function f at P in the direction S. 104

C^k; C^∞ differentiability class. 26

dF_p differential map. 27

$\exp$; $\log$ exponential and logarithm matrix functions. 43

Exp_p; Log_p exponential and logarithm maps. 36, 42

Gam_n multivariate gamma function of real $n \times n$ matrices. 86, 129

∇f Riemannian gradient of function f. 38

$\mathrm{Hess} f$ Hessian of the function f. 38, 111

$\mathbb{I}$ indicator function. 124

d_{LE}; d_{AI} Riemannian distance associated with Log-Eucliudean and Affine-Invariant metric tensor respectively. 105, 108

$\mathrm{Exp}_P^{\mathrm{LE}}$; $\mathrm{Log}_P^{\mathrm{LE}}$; $\mathrm{Exp}_P^{\mathrm{AI}}$; $\mathrm{Log}_P^{\mathrm{AI}}$ exponential/logarithm map at P associated with Log-Eucliudean and Affine-Invariant metric tensor respectively. 104, 107

g^{LE}; g^{AI} Log-Eucliudean and Affine-Invariant metric tensor respectively. 103, 106

L_a; R_a left/right translation. 40

$\otimes$ Kronecker product. 107

$\mathfrak{T}(X, Y)$ torsion tensor. 33

tr trace operator. 29

$\mathscr{B}$ orthogonal basis for $\mathfrak{so}(3)$ with respect to $\langle \cdot, \cdot, \rangle_F$. 69

$\mathfrak{B}_d^{\text{LE}}$; $\mathfrak{B}_d^{\text{AI}}$ orthonormal frame basis associated with Log-Eucliudean and Affine-Invariant metric tensor respectively. 105, 109

$\mathfrak{B}_d$ standard symmetric basis for $\mathcal{S}(n)$ with $d = n(n+1)/2$. 102

I_n $n \times n$ identity matrix. 43

0_n $n \times n$ zero matrix. 70

$\mathscr{Z}(x; 0; I_n)$ density function of the multivariate normal distribution $\mathcal{N}(0, I_n)$, that is evaluated at $x \in \mathbb{R}^n$. 155

$\mathfrak{N}(I_n, \rho^2)$ Riemannian Gaussian distribution on $\mathcal{SP}(n)$. 130

$\mathcal{N}(0, I_n)$ multivariate normal distribution of n-dimension with zero mean and covariance matrix equals to I_n. 53

vM(μ, κ) von Mises distribution. 96

ACF autocorrelation function. 11, 148

AI Affine-Invariant. 21

CLT central limit theorem. 17

DTI diffusion tensor image. 98

GLSE generalized least squares estimator. 62

GME Gauss-Markov estimator. 62

i.i.d independent and identically distributed. 154

K-S Kolmogorov–Smirnov. 145

LE Log-Euclidean. 21, 103

LLN law of large numbers. 17

MALA Metropolis-adjusted Langevin algorithm. 157

ODE ordinary differential equation. 36

OLSE ordinary least squares estimator. 62

OU Ornstein-Uhlenbeck. 21

PCA principal component analysis. 172

SDE stochastic differential equation. 19, 94

SPDE stochastic partial differential equation. 157

STA soft tissue artefact. 79

Index

Appendix A

Supplementary figures

Firstly, we present some figures mentioned in Chapter 3, where we look at the estimated speed for STAs in Figure A.1-A.2 and rotations in Figure A.3-A.4.

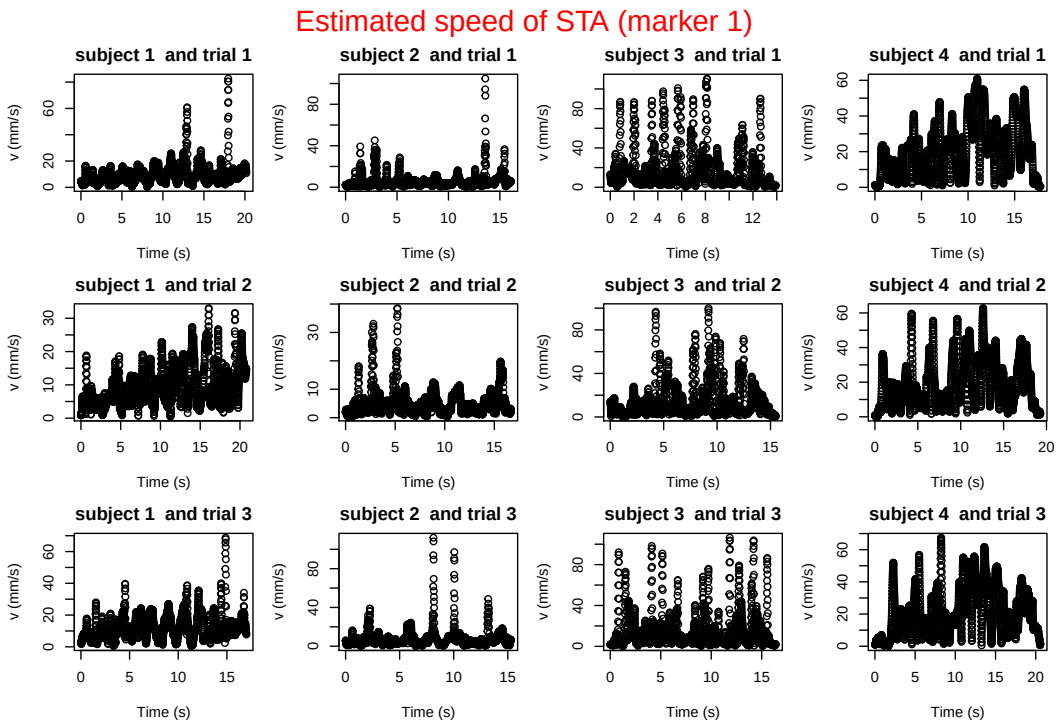


Figure A.1: Euclidean distance between adjacent STAs at marker 1 are plotted against observation time using the data from [3].

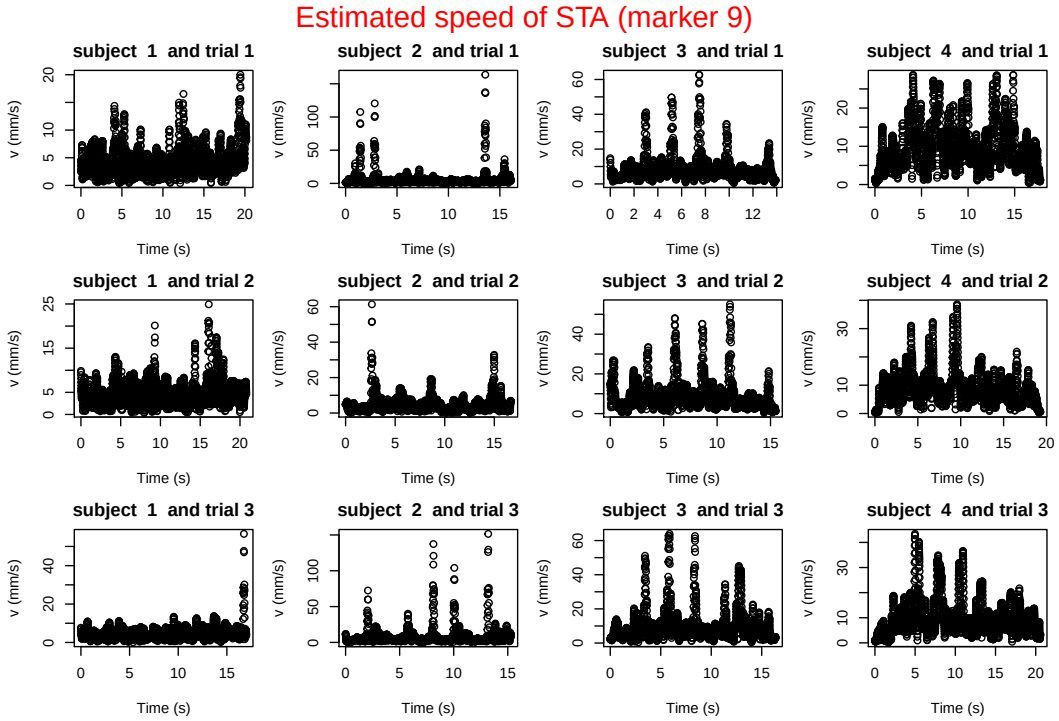


Figure A.2: Euclidean distance between adjacent STAs at marker 9 are plotted against observation time using the data from [3].

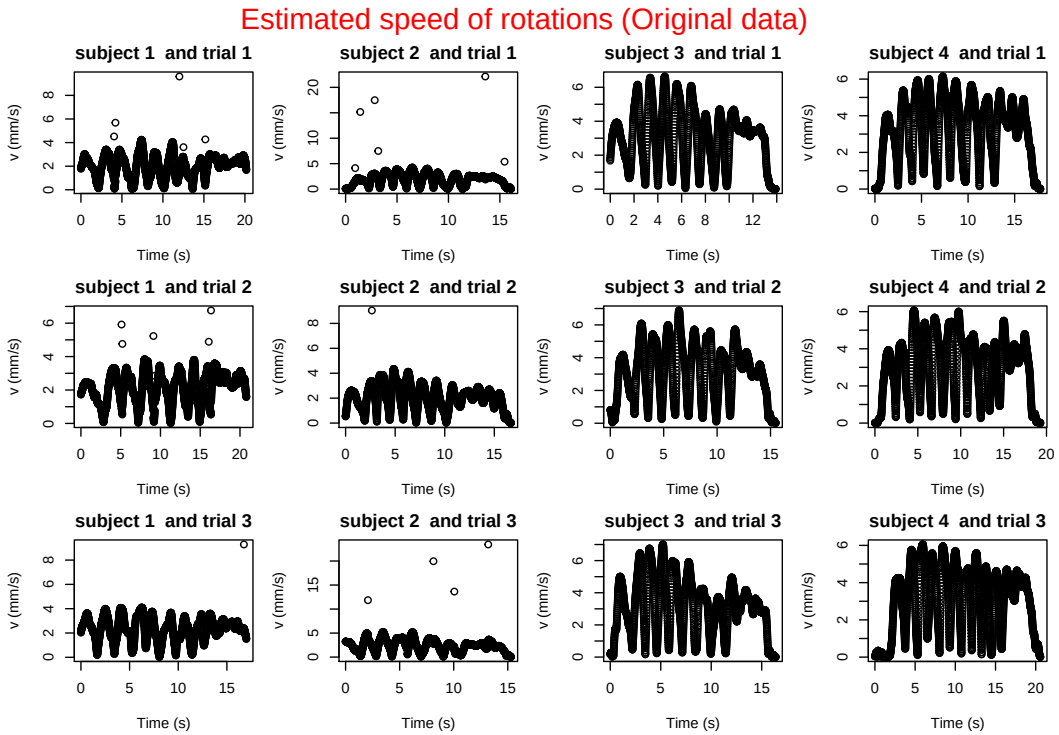


Figure A.3: Riemannian distance between adjacent rotational matrices on the thigh are plotted against observation time using the data from [3].

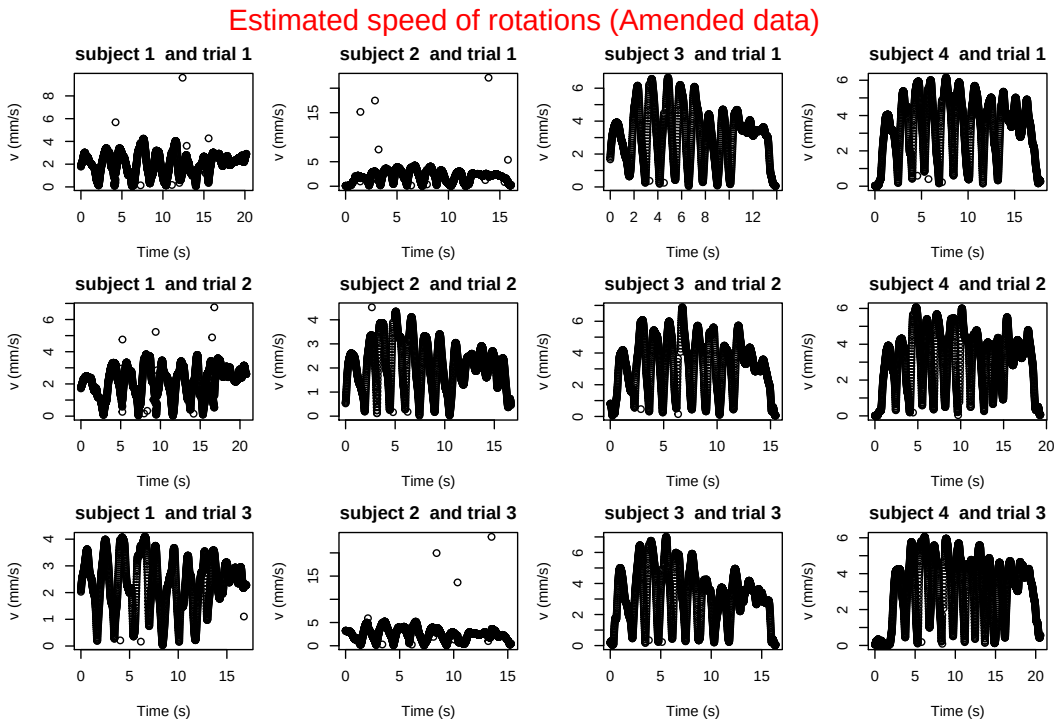


Figure A.4: Riemannian distance between adjacent rotational matrices on the thigh are plotted against observation time after applying Algorithm 3.13 with $\alpha = 2\%$.

Lastly, we presents some figures mentioned in Chapter 4. We include the kernel density estimations for the marginal posterior distribution of the model parameters $\{\theta, M, \sigma^2\}$ in the simulation study (Figure A.5) and the application in finance (Figure A.7). Furthermore, trace plots and ACF plots when using either the LE metric with $\delta_t = 0.01$ or the AI metric with $\delta_t = 0.001$ are shown in Figure A.6.

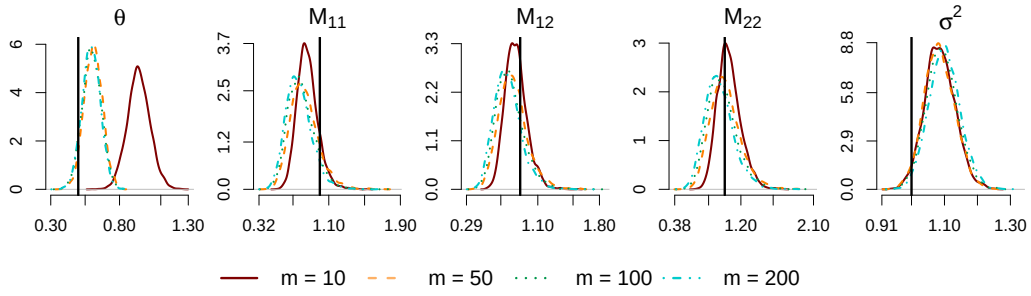


Figure A.5: (Simulation study). Estimated posterior distribution of $\{\theta, M, \sigma^2\}$ using 5×10^4 MCMC iterations (2×10^3 burn-in discarded, thinned by 12). True values are indicated by solid vertical black lines.

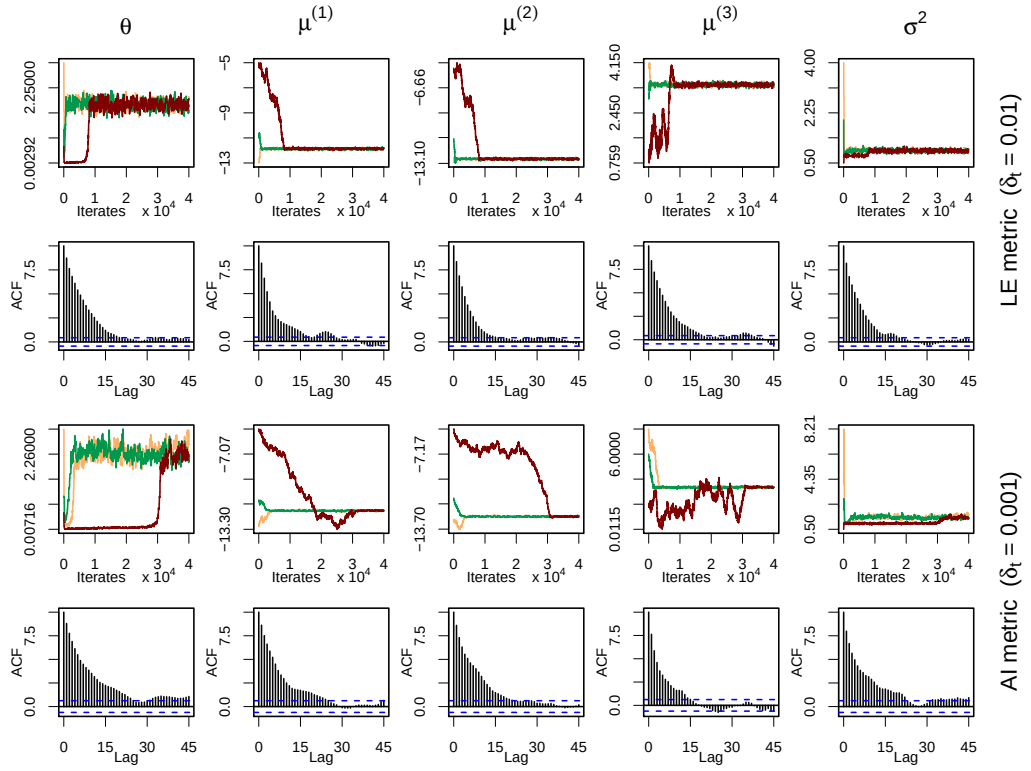


Figure A.6: (Application in finance). MCMC trace plots of 4000 iterations using different starting points (orange, green and red) and ACF plots based on iterates $1000 - 4000$ of the green chain with $\delta_t = 0.01$ in the case of the LE metric and $\delta_t = 0.001$ in the case of the AI metric .

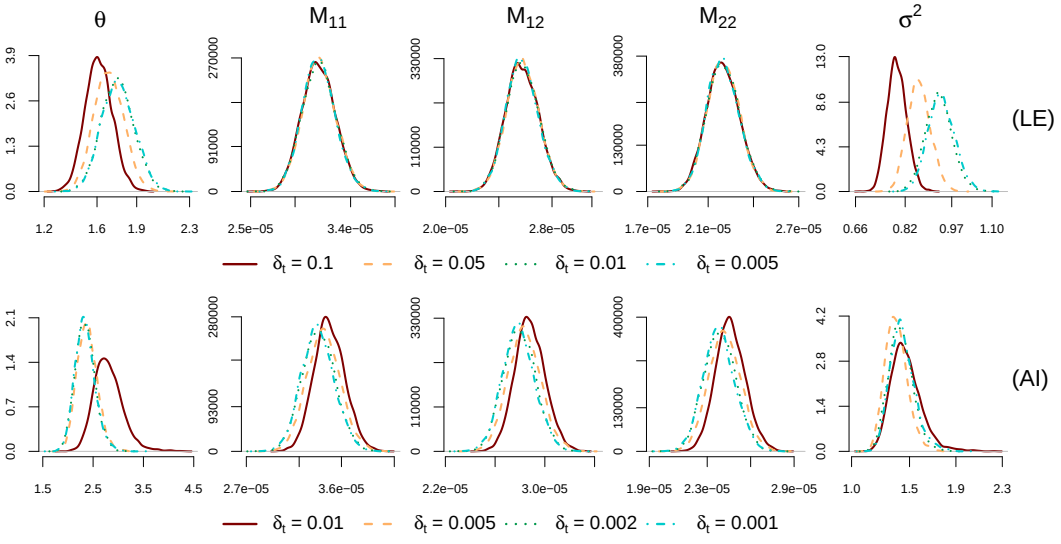


Figure A.7: (Application in finance). Estimated posterior distribution for $\{\theta, M, \sigma^2\}$ using either the LE metric (top row) or the AI metric (bottom row), based on 10^5 MCMC iterations (4×10^3 burn-in discarded, thinned by 19).

Appendix B

Supplementary algorithms

B.1 Guided proposals on the Euclidean space

Suppose that we need to sample the diffusion bridge $y_t^* = \{y_t,\, 0 \leq t \leq T \,|\, y_0 = u,\ y_T = v\}$, where the process $y_t \in \mathbb{R}^d$ is the solution of the following SDE :

$$dy_t = a(t, y_t)\, dt + b(t, y_t)\, dB_t \qquad\qquad (y_0 = u) \tag{B.1}$$

with $a : \mathbb{R}_{\geq 0} \times \mathbb{R}^d \to \mathbb{R}^d$ and $b : \mathbb{R}_{\geq 0} \times \mathbb{R}^d \to \mathcal{SP}(d)$ are measurable functions and locally Lipschitz with respect to y_t. Here, B_t is a standard $\mathbb{R}^d$- Brownian motion and we assume the existence of strong solution for the SDE (B.1).

Under weak assumptions, see [161], characteristic of the diffusion bridge y_t^* is actually the solution of the following SDE for $y_0^* = u$ and $0 \leq t \leq T$:

$$\begin{aligned} dy_t^* &= a(t, y_t^*)\, dt + b(t, y_t^*)\, dB_t \\ &\quad + \big(b(t, y_t^*) b^T(t, y_t^*)\big)\, \nabla_{y_t^*} \big(\log p(t, y_t^*; T, v)\big)\, dt, \end{aligned} \tag{B.2}$$

with $p(t, y_t^*; T, v)$ is the (unknown) transition density of the original stochastic process y_t from time t to the terminal time T. Since sampling y_t^* directly from the SDE (B.2) is impossible, the idea is to propose for y_t^* from a so-called *guided proposal*, that is easier for simulation, via the acceptance/rejection sampling. In this section, we discuss details of the conditions for which the laws of the target diffusion bridge and the guided proposal are equivalent on the Euclidean space in [38, 40].

M. Schauer et al. [40] define the class of guided proposals $y_t^\diamond$, that are the solution of the following SDEs, with $y_0^\diamond = u$ and $0 \leq t \leq T$:

$$\begin{aligned} dy_t^\diamond &= a(t, y_t^*)\, dt + b(t, y_t^\diamond)\, dB_t \\ &\quad + \big(b(t, y_t^\diamond) b^T(t, y_t^\diamond)\big)\, \nabla_{y_t^\diamond} \big(\log \tilde{p}(t, y_t^\diamond; T, v)\big)\, dt, \end{aligned} \tag{B.3}$$

where $\tilde{p}(t, y_t^\diamond; T, v)$ is the transition density of some *proposal (auxiliary) processes* $\tilde{y}_t$, in which $\tilde{p}$ has explicit formula.

B. Delyon & Y. Hu [38] starts this concept earlier by having $\tilde{y}_t$ as a standard $\mathbb{R}^d$-Brownian motion, i.e. $y_t^\diamond$ is given by

$$dy_t^\diamond = \left(a(t, y_t^\diamond) - \frac{y_t^\diamond - v}{T - t}\right) dt + b(t, y_t^\diamond)\, dB_t \qquad (y_0^\diamond = u)$$

The law of the guided proposal is shown to be absolutely continuous with respect to the distribution of the target (unknown) diffusion bridge in [38] under the condition that the function $b(t, y) \in C^2(\mathbb{R}^d)$ are bounded with bounded derivative. Moreover, $b(t, y)$ is assumed to be invertible with respect to y and its inverse is also bounded.

In general, M. Schauer et al. [40] shows that there are three required assumptions B.1-B.3 to ensure the absolute continuity between the guided proposals and the target diffusion bridge.

Assumption B.1. [40] Suppose that the diffusivity of the proposal process $\tilde{y}_t$ is $\tilde{b}(t, \tilde{y})$. The proposal process is assumed to satify $\tilde{b}(T, v)\tilde{b}^T(T, v) = b(T, v)b^T(T, v)$ and the following

(i) The transition density $\tilde{p}$ of $\tilde{y}_t$ satisfy

$$\lim_{t \uparrow T} \int f(t, x)\, \tilde{p}(t, y; T, v)\, dy = f(T, v),$$

for all bounded, continuous functions $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}$. And there exist positive constants $\tilde{A},\ \tilde{B}$ such that for $0 < s < T$,

$$\tilde{p}(s, y; T, v) \leq \tilde{A}\, (T - s)^{-d/2}\, \exp\left(-\tilde{B}\, \frac{||v - y||_2^2}{T - s}\right) \text{ uniformly in } y.$$

(ii) For all $x, y \in \mathbb{R}^d$ and $t \in [0, T]$, there exist positive constants $\tilde{C}_1,\ \tilde{C}_2,\ \tilde{C}_3$ such that the gradient of $\log \tilde{p}$, $\nabla \log \tilde{p}$, and the Hessian of $\log \tilde{p}$, $\Delta \log \tilde{p}$, satisfy

$$\begin{aligned}
&\left|\left|\nabla \log \tilde{p}(t, y; T, v)\right|\right|_2 \leq \tilde{C}_1 \left(1 + \frac{||y - v||_2}{T - t}\right), \\
&\left|\left|\nabla \log \tilde{p}(t, y; T, v) - \nabla \log \tilde{p}(t, x; T, v)\right|\right|_2 \leq \tilde{C}_2 \left(\frac{||y - x||_2}{T - t}\right), \\
&\left|\left|\Delta \log \tilde{p}(t, y; T, v)\right|\right|_F \leq \tilde{C}_3 \left(\frac{1}{T - t} + \frac{||y - v||_2}{T - t}\right).
\end{aligned}$$

Assumption B.2. [40] There exist constants $A,\ B > 0$ and $C > 1$ and a function $\mu_t(s, y) : \{s, t : 0 \leq s \leq t \leq T\} \times \mathbb{R}^d \to \mathbb{R}^d$ with $\left|\left|\mu_t(s, x) - x\right|\right|_2 < C(t - s)||x||_2$ and $||\mu_t(s, x)||_2^2 \geq \max\{1/C,\ 1 - C(t - s)\}\, ||x||_2^2$, so that for all $s < t \leq T$ and $x, y \in \mathbb{R}^d$, the (unknown) transition density of the original process y_t satisfies

$$p(s, x; t, y) \leq B\,(t - s)^{-d/2} \exp\left(-A\left|\left|y - \mu_t(s, x)\right|\right|_2^2 / (t - s)\right).$$

Assumption B.3. [40] There exist an $\epsilon \in (0, 1/6)$ and an almost surely finite random variable K such that for all $0 \leq t \leq T$, $||y_t^\diamond - v||_2 \leq K(T - t)^{1/2 - \epsilon}$ holds almost surely for the guided proposal $y_t^\diamond$.

In summary, while Assumption B.1 select the proposal processes $\tilde{y}_t$ to behave nicely near the terminal time T, Assumption B.3 ensures the guided proposal $y_t^\diamond$ to converge to v at an appropriate speed when t tends to T, which is essential because the two laws of interest are equivalent only if their speed of reaching to the endpoint are the same. On the other hand, Assumption B.2 apply to the original process y_t and particularly if the drift $a(t, y_t)$ is bounded, this assumption is directly achieved by the Aronson 's estimation, see [38].

The Assumptions B.1 and B.3 can be simplified if the proposal process $\tilde{y}_t$ is a linear process, that is the solution of the following SDE:

$$d\tilde{y}_t = \left(\tilde{\mathfrak{a}}_1(t)\, \tilde{y}_t + \tilde{\mathfrak{a}}_2(t)\right) dt + \tilde{\mathfrak{b}}(t)\, dB_t \tag{B.4}$$

where $\tilde{\mathfrak{a}}_1,\ \tilde{\mathfrak{a}}_2,\ \tilde{\mathfrak{b}}$ are time-dependent functions and $\tilde{\mathfrak{b}}(T) = b(T, v)$. We state these

statements in [40] below.

(i) Suppose that $\tilde{\mathfrak{a}}_1$, $\tilde{\mathfrak{a}}_2$ are continuously differentiable on $[0, T]$ and $\tilde{\mathfrak{b}}$ is Lipchitz on $[0, T]$. There exists positive constant η such that for all $t \in [0, T]$ and all $y \in \mathbb{R}^d$:

$$y^T \, \tilde{\mathfrak{b}}(s)\tilde{\mathfrak{b}}^T(s) \, y \geq \eta ||y||_2^2.$$

then the proposal process $\tilde{y}_t$ satisfies Assumption B.1.

(ii) There exists a positive ϵ such that for all $0 \leq s \leq T$, and $x, \, y \in \mathbb{R}^d$:

$$x^T \, b(s, y)b^T(s, y) \, x \geq \epsilon ||x||_2^2,$$

and that the drift of the original process is in the form $a(t, y) = \mathfrak{a}_1(t, y)y + \mathfrak{a}_2(t, y)$, where $\mathfrak{a}_1$ and $\mathfrak{a}_2$ are bounded function. Then there exists an almost surely finite random variable K such that the following holds almost surely

$$\left|\left|y_t^\diamond - v\right|\right|_2 \leq K \sqrt{(T-t) \, \log\log\left(\frac{1}{T-t} + e\right)}$$

for all $0 \leq t \leq T$ and here e is the constant Euler's number. Consequently, Assumption B.3 holds for any $\epsilon > 0$.

B.2 Low-frequency inference on the Euclidean space

In this section, we present an adapted version of the algorithm proposed by G. O. Roberts & O. Stramer [43], which can be applied to a multivariate diffusion process with constant diffusivity. Suppose that $x_t \in \mathbb{R}^d$ is the solution of the following SDE:

$$dx_t = a_\Theta(t, x_t) + \sqrt{C} \, dB_t \qquad\qquad (x_0 = u) \tag{B.5}$$

where Θ is a vector of model parameters in the drift of x_t and a constant matrix $C \in \mathcal{SP}(d)$. We assume the existence of the solution of the SDE (B.5).

Suppose we have observations $\mathfrak{D} = \{\mathfrak{x}_0, \ldots, \mathfrak{x}_N\}$ at observation times $t_0 = 0 < t_1 < \ldots < t_N = T$. Moreover, assuming without loss of generality, observations

are collected at equidistant time points, and time width equals Δ_t. Direct use of the Euler-Maruyama approximation is not suitable when Δ_t is large. To improve the approximation of the transition probability density, we impute $m-1$ data points between two consecutive observations. Since the matrix C is constant, there is an one-to-one transformation between the diffusion bridge $\{x_t,\, t_{i-1} \le t \le t_i \,|,\, x_{t_{i-1}} = \mathfrak{x}_{i-1},\, x_{t_i} = \mathfrak{x}_i\}$ to another diffusion bridges $\{H_s^i,\, 0 \le s \le \Delta_t \,|\, H_0^i = H_{\Delta_t}^i = 0\}$. Thus, besides estimating $\{\Theta, C\}$, we also estimate the diffusion bridges $\{H_s^i,\, 1 \le i \le N\}$ that fill the unknown gaps between consecutive observations.

There are two choices of the proposal candidates for $\{H_s^i,\, 1 \le i \le N\}$ in [43], namely the standard Brownian bridge and the linear bridge. In this thesis, we choose the standard Brownian bridge due to its simplicity. Let us denote $\mathbb{B}_s(0;\Delta_t)$ the standard Brownian bridge conditioned on $B_0 = B_{\Delta_t} = 0$. In fact, $\mathbb{B}_s(0;\Delta_t)$ is the solution of the following SDE :

$$d\mathbb{B}_s = W_s - \frac{s}{\Delta_t} W_{\Delta_t} \quad \text{for } 0 \le s \le \Delta_t,$$

where $W_s$is a standard $\mathbb{R}^d$-dimensional Brownian motion with $W_0 = 0$.

Furthermore, we propose new values for the model parameters Θ by random walk, and the log-domain is used if any parameter is required to be positive. We denote π_0 the prior distributions for Θ.

1. (Iteration $k = 0$) Choosing starting values for $\{\Theta_0\}$ and compute $c_0 = (C_0)$, see function $\mathfrak{h}$ in Equation (4.19). Then sample standard Brownian bridges $\mathbb{B}_s^i(0;\Delta_t)$ independently for $1 \le i \le N$, each covering the time interval Δ_t and set $H_s^i = B_s^i$ for $1 \le i \le N$.

2. (Iteration $k \ge 1$).

   (a) Update H_s^i independently ($1 \le i \le N$): sample the proposal Brownian bridge $\tilde{\mathbb{B}}_s(0,\Delta_t)$, and accept this with the probability

$$\alpha^{(H)} = \exp\Big\{\mathscr{G}\big(\Theta_{k-1}, C_{k-1}; \tilde{\mathbb{B}}_s(0;\Delta_t)\big) - \mathscr{G}\big(\Theta_{k-1}, C_{k-1}; H_s^i(0;\Delta_t)\big)\Big\},$$

where

$$\mathscr{G}\big(\Theta_{k-1}, C_{k-1}; H_s^i(0;\Delta_t)\big) = \sum_{j=0}^{m} \Bigg\{ a_{\Theta_{k-1}}^T(t, y_{s_j})\, C_{k-1}^{-1}\, (y_{s_{j+1}} - y_{s_j}) - \frac{1}{2}\, \Big(a_{\Theta_{k-1}}^T(t, y_{s_j})\, C_{k-1}^{-1}\, a_{\Theta_{k-1}}(t, y_{s_j}) \Big)\, (s_{j+1} - s_j) \Bigg\}, \tag{B.6}$$

and for $0 = s_0 < \ldots < s_{m+1} = \Delta_t$:

$$y_{s_j} = \sqrt{C}_{k-1}\, H_{s_j}^i + \frac{(\Delta_t - s_j)\, \mathfrak{x}_{i-1} + s_j\, \mathfrak{x}_i}{\Delta_t}$$

(b) Update C: propose $\tilde{c}$ and compute the corresponding $\tilde{C} = \mathfrak{h}^{-1}(c)$ and accept $\tilde{c}$, $\tilde{C}$ with probability

$$\alpha^{(C)} = \frac{\pi_0^{(c)}(\tilde{c}) \prod_{i=1}^N \exp\left\{\mathscr{G}\big(\Theta_{k-1}, C_{k-1}; H_s^i(0;\Delta_t)\big)\right\}}{\pi_0^{(c)}(c_{k-1}) \prod_{i=1}^N \exp\left\{\mathscr{G}\big(\Theta_{k-1}, \tilde{C}; H_s^i(0;\Delta_t)\big)\right\}} \cdot \frac{\prod_{i=1}^N \exp\left\{f(t_{i-1}, \mathfrak{x}_{i-1}; t_i, \mathfrak{x}_i; C_{k-1})\right\}}{\prod_{i=1}^N \exp\left\{f(t_{i-1}, \mathfrak{x}_{i-1}; t_i, \mathfrak{x}_i; \tilde{C})\right\}}$$

where $\mathscr{G}$ is given in Equation (B.6) and

$$f(s, x; t, y; C) = -\frac{(y-x)^T\, C^{-1}\, (y-x)}{2(t-s)} - \frac{d}{2} \log\det C.$$

(c) Update Θ: sample proposal $\tilde{\Theta}$ and accept it with the probability

$$\alpha^{(\Theta)} = \frac{\pi_0^{(\Theta)}(\tilde{\Theta}) \prod_{i=1}^N \exp\left\{\mathscr{G}\big(\tilde{\Theta}, C_k; H_s^i(0;\Delta_t)\big)\right\}}{\pi_0^{(\Theta)}(\Theta_{k-1}) \prod_{i=1}^N \exp\left\{\mathscr{G}\big(\Theta_{k-1}, C_k; H_s^i(0;\Delta_t)\big)\right\}}$$

where $\mathscr{G}$ is given in Equation (B.6).

Bibliography

[1] Mai Ngoc Bui, Yvo Pokern, and Petros Dellaportas. Inference for partially observed Riemannian Ornstein–Uhlenbeck diffusions of covariance matrices. *arXiv preprint arXiv:2104.03193*, 2021.

[2] Stephan Huckemann, XueMei Li, Yvo Pokern, and Anja Sturm. Statistics of stochastic differential equations on manifolds and stratified spaces (hybrid meeting). *Oberwolfach Reports*, 2021.

[3] Valentina Camomilla, Andrea Cereatti, Laurence Chèze, and Aurelio Cappozzo. A hip joint kinematics driven model for the generation of realistic thigh soft tissue artefacts. *Journal of biomechanics*, 46(3):625–630, 2013.

[4] Edward S Grood and Wilfredo J Suntay. A joint coordinate system for the clinical description of three-dimensional motions: application to the knee. *Journal of biomechanical engineering*, 105(2):136–144, 1983.

[5] Gordon E Robertson, Graham E Caldwell, Joseph Hamill, Gary Kamen, and Saunders Whittlesey. *Research methods in biomechanics*. Human kinetics, 2013.

[6] Andrea Cereatti, Marco Donati, Valentina Camomilla, Fabrizio Margheritini, and Aurelio Cappozzo. Hip joint centre location: an ex vivo study. *Journal of biomechanics*, 42(7):818–823, 2009.

[7] Eduardo García-Portugués, Michael Sørensen, Kanti V Mardia, and Thomas Hamelryck. Langevin diffusions on the torus: estimation and applications. *Statistics and Computing*, 29(1):1–22, 2019.

[8] John Kent. Time-reversible diffusions. *Advances in Applied Probability*, 10(4):819–835, 1978.

[9] Michael Wilkinson and Alain Pumir. Spherical Ornstein-Uhlenbeck processes. *Journal of Statistical Physics*, 145(1):113, 2011.

[10] Valentina Staneva and Laurent Younes. Learning shape trends: Parameter estimation in diffusions on shape manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 38–46, 2017.

[11] Frank G Ball, Ian L Dryden, and Mousa Golalizadeh. Brownian motion and Ornstein–Uhlenbeck processes in planar shape space. *Methodology and Computing in Applied Probability*, 10(1):1–22, 2008.

[12] Rui Caseiro, Pedro Martins, João F Henriques, and Jorge Batista. A nonparametric Riemannian framework on tensor field with application to foreground segmentation. *Pattern Recognition*, 45(11):3997–4017, 2012.

[13] Ian L Dryden, Alexey Koloydenko, Diwei Zhou, et al. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, 3(3):1102–1123, 2009.

[14] Anil Aswani, Peter Bickel, Claire Tomlin, et al. Regression on manifolds: Estimation of the exterior derivative. *The Annals of Statistics*, 39(1):48–81, 2011.

[15] Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127, 2006.

[16] Daniel Adler, Duncan Murdoch, and others. *rgl: 3D Visualization Using OpenGL*, 2018. R package version 0.99.16.

[17] Stephan Huckemann and Thomas Hotz. On means and their asymptotics: circles and shape spaces. *Journal of mathematical imaging and vision*, 50(1):98–106, 2014.

[18] Rabi Bhattacharya, Vic Patrangenaru, et al. Large sample theory of intrinsic and extrinsic sample means on manifolds—ii. *The Annals of Statistics*, 33(3):1225–1259, 2005.

[19] Thomas Hotz, Stephan Huckemann, Huiling Le, J Stephen Marron, Jonathan C Mattingly, Ezra Miller, James Nolen, Megan Owen, Vic Patrangenaru, Sean Skwerer, et al. Sticky central limit theorems on open books. *The Annals of Applied Probability*, 23(6):2238–2258, 2013.

[20] Benjamin Eltzner and Stephan F Huckemann. A smeary central limit theorem for manifolds with application to high-dimensional spheres. *The Annals of Statistics*, 47(6):3360–3381, 2019.

[21] P Thomas Fletcher. Geodesic regression and the theory of least squares on Riemannian manifolds. *International journal of computer vision*, 105(2):171–185, 2013.

[22] Oliver Pfaffel. Wishart processes. *arXiv preprint arXiv:1201.3256*, 2012.

[23] J Kent. Discussion of paper by K. V. Mardia. *J. Roy. Statist. Soc. Ser. B*, 37(3):377–378, 1975.

[24] Victor Solo. Attitude estimation and Brownian motion on so (3). In *49th IEEE Conference on Decision and Control (CDC)*, pages 4857–4862. IEEE, 2010.

[25] Elton P Hsu. *Stochastic analysis on manifolds*, volume 38. American Mathematical Soc., 2002.

[26] Kenneth David Elworthy. *Stochastic differential equations on manifolds*, volume 70. Cambridge University Press, 1982.

[27] Jonathan H Manton. A primer on stochastic differential geometry for signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 7(4):681–699, 2013.

[28] Ramesh Gangolli. On the construction of certain diffusions on a differentiable manifold. *Probability Theory and Related Fields*, 2(5):406–419, 1964.

[29] Peter Baxendale. Measures and Markov processes on function spaces. *Mémoires de la Société Mathématique de France*, 46(131-141):3, 1976.

[30] Charles David Gorman. Brownian motion of rotation. *Transactions of the American Mathematical Society*, 94(1):103–117, 1960.

[31] HP Jr McKean et al. Brownian motions on the 3-dimensional rotation group. *Memoirs of the College of Science, University of Kyoto. Series A: Mathematics*, 33(1):25–38, 1960.

[32] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications*, 29(1):328–347, 2007.

[33] Xavier Pennec. *Statistical computing on manifolds for computational anatomy*. PhD thesis, Université Nice Sophia Antipolis, 2006.

[34] Maher Moakher and Mourad Zéraï. The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *Journal of Mathematical Imaging and Vision*, 40(2):171–187, 2011.

[35] Lene Theil Skovgaard. A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, pages 211–223, 1984.

[36] Maher Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, 2005.

[37] Chulwoo Han and Frank C Park. A geometric framework for covariance dynamics. *Journal of Banking & Finance*, page 106319, 2021.

[38] Bernard Delyon and Ying Hu. Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Applications*, 116(11):1660–1675, 2006.

[39] Frank van der Meulen, Moritz Schauer, et al. Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electronic Journal of Statistics*, 11(1):2358–2396, 2017.

[40] Moritz Schauer, Frank Van Der Meulen, Harry Van Zanten, et al. Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli*, 23(4A):2917–2950, 2017.

[41] Gavin A Whitaker, Andrew Golightly, Richard J Boys, and Chris Sherlock. Improved bridge constructs for stochastic differential equations. *Statistics and Computing*, 27(4):885–900, 2017.

[42] S Sommer, A Arnaudon, L Kühnel, and S Joshi. Bridge simulation and metric estimation on landmark manifolds. *arXiv preprint arXiv:1705.10943*, May 2017.

[43] Gareth O Roberts and Osnat Stramer. On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika*, 88(3):603–621, 2001.

[44] Yongmiao Hong and Haitao Li. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *The Review of Financial Studies*, 18(1):37–84, 2005.

[45] William M Boothby. *An introduction to differentiable manifolds and Riemannian geometry*, volume 120. Academic press, 1986.

[46] Stephen T Lovett. *Differential geometry of manifolds*. AK Peters/CRC Press, 2010.

[47] Manfredo Perdigao do Carmo. *Riemannian geometry*. Birkhäuser, 1992.

[48] Takashi Sakai. *Riemannian geometry*, volume 149. American Mathematical Soc., 1996.

[49] Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977.

[50] R Hermann et al. S. Sternberg, lectures on differential geometry. *Bulletin of the American Mathematical Society*, 71(2):332–337, 1965.

[51] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[52] Richard Durrett. *Stochastic calculus: a practical introduction*. CRC press, 2018.

[53] GN Milstein and Michael V Tretyakov. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM journal on numerical analysis*, 43(3):1139–1154, 2005.

[54] E Joergensen. Construction of the Brownian motion and the Ornstein-Uhlenbeck process in a Riemannian manifold on basis of the Gangolli-Mc. Kean injection scheme. *Probability Theory and Related Fields*, 44(1):71–87, 1978.

[55] Orathai Polsen and Charles C Taylor. Parametric circular–circular regression and diagnostic analysis. *Geometry driven statistics*, 121:115, 2015.

[56] Ming-yen Cheng and Hau-tieng Wu. Local linear regression on manifolds and its geometric interpretation. *Journal of the American Statistical Association*, 108(504):1421–1434, 2013.

[57] Adrien-Marie Legendre. New methods for the determination of the orbits of comets. *Courcier, Paris*, 1805.

[58] Carl Friedrich Gauss. *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections: A Translation of Gauss's" Theoria Motus." With an Appendix*. Little, Brown and Company, 1857.

[59] Monami Banerjee, Rudrasis Chakraborty, Edward Ofori, David Vaillancourt, and Baba C Vemuri. Nonlinear regression on Riemannian manifolds and its applications to neuro-image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 719–727. Springer, 2015.

[60] Michael S Andersen, Michael Damsgaard, John Rasmussen, Dan K Ramsey, and Daniel L Benoit. A linear soft tissue artefact model for human movement analysis: proof of concept using in vivo data. *Gait & posture*, 35(4):606–611, 2012.

[61] George W Snedecor. Statistical methods applied to experiments in agriculture and biology, 1956.

[62] Takashi Isobe, Eric D Feigelson, Michael G Akritas, and Gutti Jogesh Babu. Linear regression in astronomy. *The astrophysical journal*, 364:104–113, 1990.

[63] Barbara Finlay and A Agresti. *Statistical methods for the social sciences*. Dellen, 1986.

[64] George AF Seber and Alan J Lee. *Linear regression analysis*, volume 329. John Wiley & Sons, 2012.

[65] C Gauss. Theory of the combination of observations which leads to the smallest errors. *Gauss Werke*, 4:1–93, 1821.

[66] Takeaki Kariya and Hiroshi Kurata. *Generalized least squares*. John Wiley & Sons, 2004.

[67] Yasuyuki Toyooka. An iterated version of the Gauss-Markov theorem in generalized least squares estimation. *Journal of the Japan Statistical Society, Japanese Issue*, 17(2):129–136, 1987.

[68] Juri Kalviste. Spherical mapping and analysis of aircraft angles for maneuvering flight. *Journal of Aircraft*, 24(8):523–530, 1987.

[69] Dimitri Mihalas and James Binney. Galactic astronomy: Structure and kinematics. *San Francisco, CA, WH Freeman and Co., 1981. 608 p.*, 1981.

[70] Stefano De Leo and Waldyr A Rodrigues Jr. Quaternionic electron theory: Geometry, algebra and dirac's spinors. *arXiv preprint hep-th/9806058*, 1998.

[71] Maher Moakher. Means and averaging in the group of rotations. *SIAM journal on matrix analysis and applications*, 24(1):1–16, 2002.

[72] Serge Belongie. Wolfram web resource. rodrigues' rotation formula. `http://mathworld.wolfram.com/RodriguesRotationFormula.html`, created by Eric W. Weisstein.

[73] JA Oteo. The Baker–Campbell–Hausdorff formula and nested commutator identities. *Journal of mathematical physics*, 32(2):419–424, 1991.

[74] Georgios Christodoulakis, Krishna Busawon, Nick Caplan, and Su Stewart. On the filtering and smoothing of biomechanical data. In *2010 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010)*, pages 512–516. IEEE, 2010.

[75] Tecla Bonci, Valentina Camomilla, Raphaël Dumas, Laurence Chèze, and Aurelio Cappozzo. A soft tissue artefact model driven by proximal and distal joint kinematics. *Journal of biomechanics*, 47(10):2354–2361, 2014.

[76] Valentina Camomilla and Tecla Bonci. A joint kinematics driven model of the pelvic soft tissue artefact. *Journal of Biomechanics*, 111:109998, 2020.

[77] Valentina Camomilla, Tecla Bonci, Raphaël Dumas, Laurence Chèze, and Aurelio Cappozzo. A model of the soft tissue artefact rigid component. *Journal of biomechanics*, 48(10):1752–1759, 2015.

[78] Raphaël Dumas, Valentina Camomilla, Tecla Bonci, Laurence Chèze, and Aurelio Cappozzo. Generalized mathematical representation of the soft tissue artefact. *Journal of biomechanics*, 47(2):476–481, 2014.

[79] Vincent Bonnet, Vincent Richard, Valentina Camomilla, Gentiane Venture, Aurelio Cappozzo, and Raphaël Dumas. Joint kinematics estimation using a multi-body kinematics optimisation and an extended Kalman filter, and embedding a soft tissue artefact model. *Journal of biomechanics*, 62:148–155, 2017.

[80] Tecla Bonci, Valentina Camomilla, Raphaël Dumas, Laurence Chèze, and Aurelio Cappozzo. Rigid and non-rigid geometrical transformations of a marker-cluster and their impact on bone-pose estimation. *Journal of Biomechanics*, 48(15):4166–4172, 2015.

[81] E Grimpampi, Valentina Camomilla, A Cereatti, P De Leva, and Aurelio Cappozzo. Metrics for describing soft-tissue artefact and its effect on pose, size, and shape of marker clusters. *IEEE Transactions on Biomedical Engineering*, 61(2):362–367, 2013.

[82] David A Winter. *Biomechanics and motor control of human movement*. John Wiley & Sons, 2009.

[83] Nicholas Stergiou. *Biomechanics and gait analysis*. Academic Press, 2020.

[84] Thomas J Santner, Brian J Williams, and William I Notz. *The design and analysis of computer experiments*. Springer Science & Business Media, 2013.

[85] Stephen Roberts, Michael Osborne, Mark Ebden, Steven Reece, Neale Gibson, and Suzanne Aigrain. Gaussian processes for time-series modelling.

Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1984):20110550, 2013.

[86] Milton Abramowitz, Irene A Stegun, and Robert H Romer. Handbook of mathematical functions with formulas, graphs, and mathematical tables, 1988.

[87] Andrey Popoff. *Fundamentals of Signal Processing in Metric Spaces with Lattice Properties: Algebraic Approach*. CRC Press, 2017.

[88] Aad W van der Vaart and J Harry van Zanten. Rates of contraction of posterior distributions based on Gaussian process priors. *The Annals of Statistics*, 36(3):1435–1463, 2008.

[89] Kazimierz Sobczyk. *Stochastic differential equations: with applications to physics and engineering*, volume 40. Springer Science & Business Media, 2001.

[90] Linda JS Allen. *An introduction to stochastic processes with applications to biology*. CRC press, 2010.

[91] Carlos A Braumann. *Introduction to stochastic differential equations with applications to modelling in biology and finance*. John Wiley & Sons, 2019.

[92] Peter Kloeden and Andreas Neuenkirch. Convergence of numerical methods for stochastic differential equations in mathematical finance. In *Recent Developments in Computational Finance: Foundations, Algorithms and Applications*, pages 49–80. World Scientific, 2013.

[93] Krzysztof Bartoszek, Sylvain Glémin, Ingemar Kaj, and Martin Lascoux. Using the Ornstein–Uhlenbeck process to model the evolution of interacting populations. *Journal of theoretical biology*, 429:35–45, 2017.

[94] Rainer Schöbel and Jianwei Zhu. Stochastic volatility with an Ornstein–Uhlenbeck process: an extension. *Review of Finance*, 3(1):23–46, 1999.

[95] Ross A Maller, Gernot Müller, and Alex Szimayer. Ornstein–Uhlenbeck processes and extensions. *Handbook of Financial Time Series*, pages 421–437, 2009.

[96] Eric W. Weisstein. Wolfram web resource. modified bessel function of the first kind. `http://mathworld.wolfram.com/ModifiedBesselFunctionoftheFirstKind.html`.

[97] Xavier Pennec. Statistical computing on manifolds: from Riemannian geometry to computational anatomy. In *LIX Fall Colloquium on Emerging Trends in Visual Computing*, pages 347–386. Springer, 2008.

[98] Denis Le Bihan, Eric Breton, Denis Lallemand, Philippe Grenier, Emmanuel Cabanis, and Maurice Laval-Jeantet. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407, 1986.

[99] Peter J Basser, James Mattiello, and Denis LeBihan. Estimation of the effective self-diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, 1994.

[100] Neil Chriss. *Black Scholes and beyond: option pricing models*. McGraw-Hill, 1996.

[101] Jitka Dupacova, Jan Hurt, and Josef Stepan. *Stochastic modeling in economics and finance*, volume 75. Springer Science & Business Media, 2006.

[102] Barrett O'neill. *Semi-Riemannian geometry with applications to relativity*, volume 103. Academic press, 1983.

[103] Sigurdur Helgason. *Differential geometry, Lie groups, and symmetric spaces*. Academic press, 1979.

[104] P Thomas Fletcher and Sarang Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer vision and math-*

ematical methods in medical and biomedical image analysis, pages 87–98. Springer, 2004.

[105] Luigi Malagò and Giovanni Pistone. Information geometry of the Gaussian distribution in view of stochastic optimization. In *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*, pages 150–162. ACM, 2015.

[106] Sungkyu Jung, Armin Schwartzman, and David Groisser. Scaling-rotation distance and interpolation of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 36(3):1180–1201, 2015.

[107] Christophe Chefd'Hotel, David Tschumperlé, Rachid Deriche, and Olivier Faugeras. Regularizing flows for constrained matrix-valued images. *Journal of Mathematical Imaging and Vision*, 20(1-2):147–162, 2004.

[108] P Thomas Fletcher and Sarang Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, 2007.

[109] David Tschumperle and Rachid Deriche. Diffusion tensor regularization with constraints preservation. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.

[110] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

[111] Valentina Masarotto, Victor M Panaretos, and Yoav Zemel. Procrustes metrics on covariance operators and optimal transportation of Gaussian processes. *Sankhya A*, 81(1):172–213, 2019.

[112] Xavier Pennec. Manifold-valued image processing with SPD matrices. In *Riemannian Geometric Statistics in Medical Image Analysis*, pages 75–134. Elsevier, 2020.

[113] Ola Elerian, Siddhartha Chib, and Neil Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993, 2001.

[114] Bjørn Eraker. MCMC analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics*, 19(2):177–191, 2001.

[115] Asger Roer Pedersen. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, pages 257–279, 1995.

[116] Garland B Durham and A Ronald Gallant. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20(3):297–338, 2002.

[117] Andrew Golightly and Darren J Wilkinson. Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis*, 52(3):1674–1693, 2008.

[118] Osnat Stramer, Matthew Bognar, and Paul Schneider. Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions. *Journal of Financial Econometrics*, 8(4):450–480, 2010.

[119] Erik Lindström. A regularized bridge sampler for sparsely sampled diffusions. *Statistics and Computing*, 22(2):615–623, 2012.

[120] Pierre Del Moral and Lawrence M Murray. Sequential Monte Carlo with highly informative observations. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):969–997, 2015.

[121] Ming Lin, Rong Chen, and Per Mykland. On generating Monte Carlo samples of continuous diffusion bridges. *Journal of the American Statistical Association*, 105(490):820–838, 2010.

[122] Mogens Bladt and Michael Sørensen. Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3):395–410, 2005.

[123] Mogens Bladt and Michael Sørensen. Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli*, 20(2):645–675, 2014.

[124] Mogens Bladt, Samuel Finch, and Michael Sørensen. Simulation of multivariate diffusion bridges. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(2):343–369, 2016.

[125] Marcin Mider, Paul A Jenkins, Murray Pollock, Gareth O Roberts, and Michael Sørensen. Simulating bridges using confluent diffusions. *arXiv preprint arXiv:1903.10184*, 2019.

[126] Robb J Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009.

[127] Frank WJ Olver, Daniel W Lozier, Ronald F Boisvert, and Charles W Clark. *NIST handbook of mathematical functions hardback and CD-ROM*. Cambridge university press, 2010.

[128] Salem Said, Lionel Bombrun, Yannick Berthoumieu, and Jonathan H Manton. Riemannian Gaussian distributions on the space of symmetric positive definite matrices. *IEEE Transactions on Information Theory*, 63(4):2153–2170, 2017.

[129] TE Duncan. A note on some laws of the iterated logarithm. *Journal of Multivariate Analysis*, 5(4):425–433, 1975.

[130] Cheng Ouyang and Jennifer Pajda-De La O. On the law of the iterated logarithm for Brownian motion on compact manifolds. *arXiv preprint arXiv:1608.08680*, 2016.

[131] Samantha R Cook, Andrew Gelman, and Donald B Rubin. Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, 15(3):675–692, 2006.

[132] FirstRate Data. Historical intraday index price data. `http://firstratedata.com/it/index`.

[133] Lan Zhang. Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47, 2011.

[134] Alexandros Beskos, Konstantinos Kalogeropoulos, and Erik Pazos. Advanced MCMC methods for sampling on diffusion pathspace. *Stochastic Processes and their Applications*, 123(4):1415–1453, 2013.

[135] Konstantinos Kalogeropoulos, Gareth O Roberts, and Petros Dellaportas. Inference for stochastic volatility models using time change transformations. *The Annals of Statistics*, pages 784–807, 2010.

[136] Martin Hairer, Andrew M Stuart, Jochen Voss, and Petter Wiberg. Analysis of SPDEs arising in path sampling. part i: The Gaussian case. *Communications in Mathematical Sciences*, 3(4):587–603, 2005.

[137] Martin Hairer, Andrew M Stuart, and Jochen Voss. Analysis of SPDEs arising in path sampling part ii: The nonlinear case. *The Annals of Applied Probability*, 17(5-6):1657–1706, 2007.

[138] Frank van der Meulen and Moritz Schauer. Bayesian estimation of incompletely observed diffusions. *Stochastics*, 90(5):641–662, 2018.

[139] Roger Frigola, Fredrik Lindsten, Thomas B Schön, and Carl Edward Rasmussen. Bayesian inference and learning in Gaussian process state-space

models with particle MCMC. *Advances in neural information processing systems*, 26, 2013.

[140] Alexandros Beskos, Dan Crisan, and Ajay Jasra. On the stability of sequential Monte Carlo methods in high dimensions. *The Annals of Applied Probability*, 24(4):1396–1445, 2014.

[141] Alexandros Beskos, Ajay Jasra, Nikolas Kantas, and Alexandre Thiery. On the convergence of adaptive sequential Monte Carlo methods. *The Annals of Applied Probability*, 26(2):1111–1146, 2016.

[142] Shouto Yonekura and Alexandros Beskos. Online smoothing for diffusion processes observed with noise. *Journal of Computational and Graphical Statistics*, (just-accepted):1–35, 2022.

[143] Siddhartha Chib, Michael K Pitt, and Neil Shephard. Likelihood based inference for diffusion driven models. Technical report, Economics Group, Nuffield College, University of Oxford, 2004.

[144] Alexandros Beskos, Gareth Roberts, Andrew Stuart, and Jochen Voss. MCMC methods for diffusion bridges. *Stochastics and Dynamics*, 8(03):319–350, 2008.

[145] Alexandros Beskos, Omiros Papaspiliopoulos, Gareth O Roberts, and Paul Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382, 2006.

[146] Gareth O Roberts and Richard L Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.

[147] Grigorios A Pavliotis and AM Stuart. Parameter estimation for multiscale diffusions. *Journal of Statistical Physics*, 127(4):741–781, 2007.

[148] Grigoris Pavliotis and Andrew Stuart. *Multiscale methods: averaging and homogenization*. Springer Science & Business Media, 2008.

[149] Stephan Huckemann and Herbert Ziezold. Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. *Advances in Applied Probability*, 38(2):299–319, 2006.

[150] V Richard, V Camomilla, L Chèze, A Cappozzo, and R Dumas. Feasibility of incorporating a soft tissue artefact model in multi-body optimisation. *Computer methods in biomechanics and biomedical engineering*, 15(sup1):194–196, 2012.

[151] Jürgen Jost. *Riemannian geometry and geometric analysis*, volume 42005. Springer, 2008.

[152] Hiroyuki Matsumoto. Closed form formulae for the heat kernels and the Green functions for the Laplacians on the symmetric spaces of rank one. *Bulletin des sciences mathematiques*, 125(6-7):553–581, 2001.

[153] Yoshihiro Nagano, Shoichiro Yamaguchi, Yasuhiro Fujita, and Masanori Koyama. A wrapped normal distribution on hyperbolic space for gradient-based learning. In *International Conference on Machine Learning*, pages 4693–4702, 2019.

[154] Tom MW Nye et al. Principal components analysis in the space of phylogenetic trees. *The Annals of Statistics*, 39(5):2716–2739, 2011.

[155] Stephan F Huckemann, Peter T Kim, Ja-Yong Koo, Axel Munk, et al. Möbius deconvolution on the hyperbolic plane with application to impedance density estimation. *The Annals of Statistics*, 38(4):2465–2498, 2010.

[156] Ioannis Kontoyiannis and Sean P Meyn. Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probability Theory and Related Fields*, 154(1-2):327–339, 2012.

[157] Mu-Fa Chen. *Eigenvalues, inequalities, and ergodic theory*. Springer Science & Business Media, 2006.

[158] Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005.

[159] Reza Mohammadi, Hélène Massam, and Gérard Letac. Accelerating Bayesian structure learning in sparse Gaussian graphical models. *Journal of the American Statistical Association*, pages 1–14, 2021.

[160] Tjun Yee Hoh. *Bayesian inference and model selection for multi-dimensional diffusion process models with non-parametric drift and constant diffusivity*. PhD thesis, UCL (University College London), 2019.

[161] TJ Lyons and WA Zheng. On conditional diffusion processes. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 115(3-4):243–255, 1990.