Adaptive Filtering and Parametric Estimation for Random Processes on Rotation Groups and Stiefel Manifolds

Jérémie Boulanger

UNIVERSITY OF GRENOBLE, FRANCE

Doctoral School EEATS

SIGNAL, IMAGE, SPEECH AND TELECOMMUNICATIONS

UNIVERSITY OF MELBOURNE, AUSTRALIA

Melbourne School of Engineering

DEPARTMENT OF ELECTRICAL & ELECTRONIC ENGINEERING

THESIS

Submitted in total fulfillment of the requirements of the degree of Doctor of Philosophy

Written and presented by

Jérémie BOULANGER

Adaptive filtering and parametric estimation for random processes on rotation groups and Stiefel manifolds

Defense the 5th of December 2013

Committee:

President:	Olivier J.J. MICHEL	-	Grenoble INP, Grenoble
Reviewers:	Yannick BERTOUMIEUX	-	Bordeaux IMS, Bordeaux
	Éric MOULINES	-	Télécom ParisTech (ENST), Paris
Examiners:	Silvère BONNABEL	-	Mines ParisTech, Paris
	Salem SAID	-	Melbourne University, Melbourne
Supervisors:	Nicolas LE BIHAN	-	Gipsa-Lab, Grenoble
	Jonathan H. MANTON	-	Melbourne University, Melbourne

Abstract

In this thesis, we focus on stochastic problems for processes defined on manifolds such as the special orthogonal group and the Stiefel manifolds. Such processes are useful to model different engineering systems such as the observation, complete and partial, of an aircraft. We show that the rotation group $SO(3)$ can also be used to model the state of a linearly transverse polarized wave.

The observation of these processes might be perturbed by noise and in this case, we are looking at an estimate of the non noisy process. That part is treated in Chapter 2. Usual methods cannot be applied due to the non linearity of the considered equations (multiplicative noise). Nowadays solutions to this kind of problems only consider local linearization and therefore only present an approximation of the optimal solution. We propose here an optimal solution without local approximation based on the antideveloppement. The problem is then reduced to a classic filtering problem with additive noise. We also compare the performances of our method for different numerical scheme.

These processes can also be parametrized and one might be looking at estimate these parameters. An example of this problem is presented in Chapter 3. Considering a polarized wave propagating in a scattering medium, we want to estimate this medium. The propagation is modeled by a compound Poisson process on $SO(3)$ and we present an algorithm to estimate these parameters using the geometric phase of the wave, carried by the direction of polarization. The geometric phase is defined by the path of the wave followed during its propagation in the medium. This phase has been observed for elastic waves in the case of propagation along a waveguide. Usual methods only rely on the distribution of the direction of the scattered wave. The one presented in this thesis also use the information carried by the polarization. We believe that this method can be used as a non intrusive way to perform non intrusive imaging.

Declaration

This is to certify that:

- The thesis comprises only my original work towards the PhD except where indicated in the Preface.
- Due acknowledgement has been made in the text to all other material used.
- The thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices as approved by the Research Higher Degrees Committee.

Jérémie Boulanger

Acknowledgments

I would like to address my deepest gratitude to my supervisors Nicolas Le Bihan and Jonathan H. Manton. Nicolas' advice is extremely useful for the clarity and the readability of my thesis and his dedicated help and enthusiasm makes me proud to be one of his students. Thank you to Jonathan H. Manton for offering me the opportunity to make a cotutelle between Grenoble-INP and the University of Melbourne in Australia. I have gained a lot from the year and half I spent in Melbourne, both from a personal and from a scientific point of view.

A large part of the knowledge I acquired at the University of Melbourne is due to Salem Said, whom I am grateful to. He answered to my questions with calm and seriousness despite my repeated failures to understand. Chapter 2 would not exist without him.

I would like to thank Vincent Rossetto, for our collaboration at the beginning of my thesis and for his permission to include his figures in the Chapter 3 of my thesis.

Finally, thank you to all my colleagues, who supported me during these three years, both from the crazy office and from Walter Boas building. I am really happy to have met you guys and sincerely hope to be able to see you again.

To conclude, thank you to my parents, who are always here to support me and help me when I was in need.

Contents

1 Stochastic Differential Equations on $SO(n)$ **3**
- 1.1 Geometry of $SO(n)$ . . . 4
  - 1.1.1 The special orthogonal group $SO(n)$ . . . 4
  - 1.1.2 Geodesics and connections . . . 6
  - 1.1.3 The case of $SO(3)$ . . . 10
- 1.2 Stochastic processes and Itō calculus . . . 11
  - 1.2.1 Stochastic differential equations on $SO(n)$ . . . 12
  - 1.2.2 Simulations of stochastic processes . . . 13
  - 1.2.3 Markov processes . . . 15
- 1.3 Examples of stochastic differential equations . . . 18
  - 1.3.1 Brownian process . . . 18
  - 1.3.2 Simulated annealing . . . 21
- 1.4 Irreducible representation of $SO(n)$ . . . 23
  - 1.4.1 Characteristic functions . . . 24
  - 1.4.2 Application to the Brownian motion . . . 25
- 1.5 Conclusion . . . 27

2 Filtering from complete and partial observation on $SO(n)$ **29**
- 2.1 Filtering on $\mathbb{R}^n$ . . . 30
  - 2.1.1 Description and solution as a KS formula . . . 30
  - 2.1.2 Zakai equation . . . 32
  - 2.1.3 Solution for linear models . . . 33
- 2.2 Filtering from observations on $SO(n)$, solution and implementation . . . 35
  - 2.2.1 Construction of the antidevelopment . . . 36
  - 2.2.2 Implementation of the solution . . . 38
- 2.3 Monte-Carlo method for filtering from observations on the Stiefel manifold . . . 44
  - 2.3.1 Geometry of the Stiefel manifold . . . 44
  - 2.3.2 Solution to the filtering on the Stiefel manifold . . . 48
  - 2.3.3 Implementation of the solution to filtering on the Stiefel manifold . . . 51
  - 2.3.4 Limitations of the solution . . . 54
- 2.4 Conclusion . . . 55

3 Geometric phase **61**
- 3.1 Geometric phase for classical mechanical systems . . . 61
  - 3.1.1 Foucault pendulum . . . 61
  - 3.1.2 Polarized waves and geometric phase . . . 67
  - 3.1.3 Evidence of the geometric phase for elastic waves . . . 70
  - 3.1.4 Conclusion . . . 75
- 3.2 Multiple scattering of polarized wave . . . 75
  - 3.2.1 Description of the random media . . . 75
  - 3.2.2 Scattering of polarized waves . . . 77
  - 3.2.3 Compound Poisson process . . . 78
  - 3.2.4 Estimation via Expectation-Maximization . . . 81
  - 3.2.5 Conclusion . . . 86

Conclusion **89**

Bibliography **91**

Appendices **97**

A Résumé en Français **99**
A.1 Introduction 100
A.2 Équations différentielles stochastiques sur $SO(n)$ 101
A.2.1 Géométrie de $SO(n)$ 102
A.2.2 Processus stochastiques et calcul de Itō 104
A.2.3 Exemples d'équations différentielles stochastiques 107
A.2.4 Représentation irréductible de $SO(n)$ 110
A.3 Filtrage pour observations complètes et partielles sur $SO(n)$ 111
A.3.1 Filtrage dans $\mathbb{R}^n$ 112
A.3.2 Filtrage à partir d'observations sur $SO(n)$, solution et implémentation 115
A.3.3 Filtrage à partir d'observations sur les variétés de Stiefel, solution Monte-Carlo 118
A.4 Phase géométrique 122
A.4.1 Phase géométrique pour les systèmes mécaniques classiques 123
A.4.2 Diffusion multiple d'une onde polarisée 126
A.5 Conclusion 129

List of Figures

1.1 Parallel transport of a vector . . . 9
1.2 Example of trajectory of Brownian motion in $SO(3)$ . . . 19
1.3 Example of trajectory for a stochastic process with drift . . . 22

2.1 Comparison of estimation methods between linear, geodesic interpolation, and classic Kalman filter with $\delta t_{\text{obs}} = 0.01s$ . . . 42
2.2 Comparison of estimation methods between linear, geodesic interpolation, and classic Kalman filter with $\delta t_{\text{obs}} = 0.2s$ . . . 43
2.3 Cumulated error of estimation for $0 \leq t \leq 50$ for the linear interpolation, the geodesic interpolation, and the classic Kalman filter for different values of δt_{obs} . . . 44
2.4 Example of filtering and reconstruction from observation on $SO(3)$ . . . 45
2.5 Illustration of the different notions introduced to describe the Stiefel manifold $V_{n,k}$ as the image from the projection Π of $SO(n)$ . . . 48
2.6 Example of trajectory of P_t (red) on $V_{3,1} = S^2$ at 4 successive times . . . 50
2.7 Trajectories of 10 particles to estimate a Brownian diffusion . . . 53
2.8 Evolution of the estimation (red) of each component of x_t . . . 54
2.9 Particle filter from Algorithm 1 at different successive times . . . 58
2.10 Projection of the particles, the state to estimate and the estimation on the plane Ox, Oy and Ox, Oz. . . . 59
2.11 For an observation on the Stiefel manifold, only the horizontal component can be estimated . . . 59

3.1 Model of Foucault pendulum . . . 62
3.2 Example of parallel transport . . . 64
3.3 Gauss-Bonnet theorem for a triangle . . . 65
3.4 Representation of different states of polarization . . . 68
3.5 Model used for linear polarized wave . . . 68
3.6 Geometric phase, wave guide . . . 69
3.7 Propagation of a linearly polarized wave along a waveguide . . . 71
3.8 Example of vertical and horizontal displacement signals recorded during the experiment . . . 72
3.9 Setup of the experiment . . . 72
3.10 Power spectrum and parametric plot of the recorded signals . . . 73
3.11 Measured geometric phase as a function of arc length along the helix . . . 74
3.12 Description of the model for multiple scattering . . . 76
3.13 Parallel transport of the polarization along a geodesic . . . 78
3.14 Parallel transport of the polarization during multiple scattering . . . 79
3.15 Evolution of the density $p_{\eta t}(\theta = 0, \beta)$ with time in a medium containing Henyey-Greenstein scatterers with anisotropy $g = 0.8$ . . . 82
3.16 Distribution of the geometric phase β for the values of the Poisson parameter $\nu = \eta t$: $\nu = 20$, $\nu = 25$, $\nu = 30$ and $\nu = 35$ corresponding to the vertical lines of Figure 3.15. . . . 82
3.17 Evolution of the density $p_{\mathcal{D}t}(\theta = 0, \beta)$ with time for a Gaussian phase function 83
3.18 Density $p_{\mathcal{D}t}(\theta = 0, \beta)$ for a Gaussian phase function with the values $\mathcal{D}t = 0.03$, $\mathcal{D}t = 0.05$, $\mathcal{D}t = 0.07$ and $\mathcal{D}t = 0.09$ corresponding to the vertical lines of Figure 3.17. . . . 83

3.19 Convergence of the estimate of the Poisson parameter $\hat{\eta}$ using EM algorithm 85
3.20 Convergence of the estimate of the Poisson parameter $\hat{\eta}$ using EM algorithm for noisy observed samples 85
3.21 Convergence of the estimate of the Poisson parameter $\hat{\eta}$ using EM algorithm for different priori of the anisotropy parameter g 86

Introduction

This thesis focuses on the adaptive filtering and parametric estimation for stochastic processes on rotation groups and Stiefel manifolds. The adaptive filtering problem is discussed in Chapter 2 and the parametric estimation issue is treated in Chapter 3 for application in wave physics. Chapter 1 provides the mathematical background useful for the understanding of the other chapters. During this study, we focuse on both the theoretical solutions to the introduced problems and the numerical implementation of the described solutions.

Stochastic processes on rotation groups can be considered as a generalization of rotation processes. Despite that these processes were mainly used to model the orientation of a system [Lo 1973], they are now used for a wide range of applications in signal processing where orthogonality constraints arise, like in sensor array to separate independent components [Arslan 2007] or information theory to maximize the rate of information transiting through a channel [Thomas 1991]. In order to describe these processes and their constraints, Chapter 1 introduces the geometry of the rotation group $SO(n)$ and uses differential geometry tools to construct and simulate these processes.

Amongst the possible problems that can be defined for rotation process, Chapter 2 considers the filtering problem for observation on $SO(n)$ first, then for observation on the Stiefel manifold later on. In this chapter, we are looking for an estimate of the state of our system, here modeled as the angular velocity, from a noisy observation of its orientation.

Despite that this study enters in the field of non-linear filtering, our approach is sensibly different from the one usually found in literature. In [Hanzon 1998] and [Kunita 1971] for example, the problem is to estimate a state with orthogonality constraints from observation with additive noise. Roughly speaking, the problem is to determine an application from $\mathbb{R}^{n\times n}$ to $SO(n)$, the main issue coming from the constraints of $SO(n)$ and the problem of easily computing an empirical mean.

In this thesis, we focus on a different model, considering the opposite way: from observation on $SO(n)$, we want to estimate the state in $\mathbb{R}^n$. In fact, the space in which lays the state to estimate does not modify our results and a different space would lead to similar results. Our major concern comes from applying usual methods of filtering for non linear model of the observation.

We solve this problem by constructing a one-to-one process with respect to the observation with an additive noise model. A non negligible part of Chapter 2 is also dedicated to the numerical implementation of the presented methods and their performances.

Chapter 3 is based on our articles [Rossetto 2012] and [Rossetto 2013] and focuses on the geometric phase for elastic waves. The propagation of linearly transverse polarized waves can be modeled as a process on $SO(3)$ and are subject to acquire a geometric phase during their propagation. From one hand, this phase can be theoretically determined from the propagation path of the wave [Rossetto 2012]. From the other hand, we show that this geometric phase can be experimentally determined with accelerometers. As a consequence, if the propagation model of the model is parametrized, the geometric phase can be used as a parametric estimator.

A large part of imaging methods nowadays consists in using the amplitude of a wave after a propagation in a random medium to construct estimators of some characteristics of this medium [Larose 2005]. However, even if these results are still functional for transverse polarized waves, they do not use the information acquired by the polarization direction from the propagation in the medium. The geometric phase relies on this information to construct

estimators. Thus, the geometric phase can be used as a non intrusive method of imaging, as it contains information on the 3D path that the wave followed when propagating though the medium.

Chapter 1

Stochastic Differential Equations on $SO(n)$

Contents

1.1 Geometry of $SO(n)$ . . . 4
1.1.1 The special orthogonal group $SO(n)$ . . . 4
1.1.2 Geodesics and connections . . . 6
1.1.3 The case of $SO(3)$ . . . 10
1.2 Stochastic processes and Itō calculus . . . 11
1.2.1 Stochastic differential equations on $SO(n)$ . . . 12
1.2.2 Simulations of stochastic processes . . . 13
1.2.3 Markov processes . . . 15
1.3 Examples of stochastic differential equations . . . 18
1.3.1 Brownian process . . . 18
1.3.2 Simulated annealing . . . 21
1.4 Irreducible representation of $SO(n)$ . . . 23
1.4.1 Characteristic functions . . . 24
1.4.2 Application to the Brownian motion . . . 25
1.5 Conclusion . . . 27

Stochastic differential equations are used in a wide range of engineering applications such as filtering [Jazwinski 1970], physical modeling [Durrett 1996], etc. In this thesis, we will focus on stochastic differential equations defined on the rotation group [Altmann 2005]. Therefore, notions of differential geometry need to be introduced to define these equations [Meyer 1981].

This chapter introduces the mathematical tools to be used in Chapters 2 and 3. Section 1.1 describes the group structure of the special orthogonal group $SO(n)$ and the notions of differential geometry used to define stochastic processes for the rest of the chapter. A class of curves known as geodesics is introduced, as they will be of major importance for simulating stochastic processes on $SO(n)$. Section 1.1 ends with a review of the special case of $SO(3)$. This serves as an example and also deserves a separate account, due to its importance in many applications.

Section 1.2 introduces stochastic calculus for stochastic differential equations on $SO(n)$. They are defined as (right) invariant equation driven by Brownian motion. Simulating stochastic process is useful, for example to compare the performances of different algorithms. The question of being able to simulate stochastic processes is thus important. Classically, the Euler scheme [Iacus 2010] is used to have a discrete approximation of a time continuous process. In our case, as the processes are defined on $SO(n)$, the Euler scheme is no more suitable to make simulations. A different scheme [McKean 1960] is presented here, extending the Euler scheme thanks to the geodesics introduced in Section 1.1. Finally, tools dedicated to the study of Markov processes are introduced, such as the

Markov semigroups and their generator, the forward and backward equation. These notions for Markov processes will be used in Section 1.3 to ensure the proper asymptotic properties of the simulations.

Section 1.3 gives two examples of stochastic processes on $SO(n)$. First of all, a description of the Brownian motion on $SO(n)$ and the way to simulate it are presented. In connection to the Brownian motion, an extension of the central limit theorem is presented for random variables on $SO(n)$. Secondly, a Monte-Carlo method makes use of the results from Section 1.2 about the Markov chains to sample a random variable on $SO(n)$ from a given distribution. This method can be extended and used in an optimization algorithm, namely the simulated annealing [Casella 2010] on $SO(n)$.

Section 1.4 highlights the concept of characteristic functions on $SO(n)$. The examples and properties presented can be understood as generalizations of classic results in harmonic analysis for periodic signals. The characteristic function approach provides a new proof of the central limit theorem on $SO(n)$ which is an alternative to the one presented in Section 1.3.

1.1 Geometry of $SO(n)$

1.1.1 The special orthogonal group $SO(n)$

For any positive integer n, the notation $SO(n)$ [Altmann 2005] stands for the set of all the $n \times n$ real matrices R which verify

$$RR^T = I_n, \qquad \det(R) = 1 \tag{1.1}$$

where R^T denotes the transpose of a matrix R, $\det(R)$ the determinant of a matrix R and I_n is the $n \times n$ identity matrix.

The columns of R form an orthogonal positive oriented basis of $\mathbb{R}^n$. In (1.1), the first condition states that the columns are pairwise orthogonal and represent unit vectors. The second condition from (1.1), involving the determinant, states that the order of these columns corresponds to a positive orientation.

The set $SO(n)$ is a group under the matrix multiplication. It means that $SO(n)$ is stable under multiplication (if $R_1, R_2 \in SO(n)$, then $R_1R_2 \in SO(n)$) and inversion (if $R \in SO(n)$, then $R^{-1} = R^T \in SO(n)$). Its identity element is clearly I_n.

The notation $T_RSO(n)$ denotes the space tangent to $SO(n)$ at the point R. It is the set of all the velocity vectors $\dot{R}_0$ of a differentiable trajectory $R_t \in SO(n)$ such that $R_0 = R$. As $R_t \in SO(n)$, Equation (1.1) applies

$$R_tR_t^T = I_n$$

Considering the derivative at $t = 0$ leads to

$$\dot{R}_0R_0^T + R_0\dot{R}_0^T = 0$$
$$\dot{R}_0R_0^T + \left(\dot{R}_0R_0^T\right)^T = 0$$

In other words, $\dot{R}_0R_0^T$ is skew symmetric

$$\dot{R}_0R_0^T \in \mathfrak{so}(n),$$

where $\mathfrak{so}(n)$ is the set of $n \times n$ skew symmetric matrices. Thus $T_RSO(n)$ is a subset of $\{\sigma R | \sigma \in \mathfrak{so}(n)\}$. As both are spaces of dimension $n(n-1)/2$, then $T_RSO(n) = \{\sigma R | \sigma \in \mathfrak{so}(n)\}$. The set of all the tangent spaces is denoted $TSO(n)$ and is called the tangent bundle [do Carmo 1976].

Equation (1.1) can also be written $R^T R = I_n$. From this definition, a second description of the tangent space $T_R SO(n)$ can be deduced. In this case, $T_R SO(n) = \{R\sigma | \sigma \in \mathfrak{so}(n)\}$. This difference of writing will further lead to respectively right and left invariant vector fields [Arnold 1966].

As $I_n \in SO(n)$, the tangent space $T_{I_n} SO(n)$ can be simply described as $T_{I_n} SO(n) = \mathfrak{so}(n)$. The tangent space $T_R SO(n)$ can then be described by a right (or left) multiplication

$$T_R SO(n) = \{\sigma R | \sigma \in T_{I_n} SO(n)\} = T_{I_n} SO(n)\ R.$$

This description leads to the following question: Is it possible to introduce invariant geometric objects? First of all, we will be interested about defining an invariant metric and invariant vector fields. Later on, an invariant connection will be introduced.

Let consider the application $\chi : \mathfrak{so}(n) \times SO(n) \to TSO(n)$ as

$$\chi(\sigma, R) = \sigma R \tag{1.2}$$

For R fixed, the application $\chi(., R)$ is a isomorphism between $\mathfrak{so}(\mathfrak{n})$ and $T_R SO(n)$. For σ fixed, the application $\chi(\sigma, .)$ is a right invariant vector field, namely

$$\chi(\sigma, R_1)R_2 = \chi(\sigma, R_1 R_2) \qquad \text{for all } R_1, R_2 \in SO(n).$$

Let σ denotes the vector field $\sigma(.) = \chi(\sigma, .)$. Despite that the vector field and the skew symmetric matrix share the same notation, they is no ambiguity on the nature of the considered element as one is an application and the other one a real matrix. Similarly, left invariant vector fields can be obtained, using a left product in (1.2) instead of a right product.

A right invariant metric can be obtained as follow. Let consider a basis of $\mathfrak{so}(n)$, $\{\sigma_i\}_{i \leq n(n-1)/2}$. A right invariant metric is obtained by declaring that the set $\{\sigma_i(R)\}$ is pairwise orthonormal. Similarly, a left invariant metric can be obtained, considering left invariant vector fields.

A metric that is both left and right invariant can be obtained, using the metric induced from the ambient space $\mathbb{R}^{n \times n}$

$$\begin{aligned} < \sigma_1, \sigma_2 > &= \frac{1}{2} \operatorname{trace}(\sigma_1^T \sigma_2) \qquad \text{for } \sigma_1, \sigma_2 \in \mathfrak{so}(n) \\ &= -\frac{1}{2} \operatorname{trace}(\sigma_1 \sigma_2). \end{aligned} \tag{1.3}$$

This metric is indeed right invariant as

$$< \sigma_1, \sigma_2 > = < \sigma_1 R, \sigma_2 R > \qquad \text{for any } R \in SO(n).$$

The left invariance demonstration is also immediate thanks to the circular property of the trace application.

From now on, let define $\{\sigma_i\}$ an orthonormal basis of $\mathfrak{so}(n)$ associated to the metric $< ., . >$. Let also define the linear and invertible application $\hat{.}$ as

$$\hat{a} = \sum_i a^i \sigma_i \tag{1.4}$$

where $a = (a^1, a^2, ...) \in \mathbb{R}^{n(n+1)/2}$.

Thanks to the invariance of $< ., . >$, it can be extended for $v_1, v_2 \in T_R SO(n)$ as

$$< v_1, v_2 >_R = < v_1 R^T, v_2 R^T >$$

As $v_1R^T, v_2R^T \in \mathfrak{so}(n)$, it directly comes from (1.3) that the scalar product reads

$$< v_1, v_2 >_R = \frac{1}{2}\,\mathrm{trace}(v_1^T v_2)$$

It is interesting to notice that the last expression does not involve the use of R or any representation of v_1, v_2. Defining $< v_1, v_2 >_R$ as $< R^T v_1, R^T v_2 >$ would lead to the same definition.

It can be shown [Faraut 2008] that if $\sigma \in \mathfrak{so}(n)$, then the matrix R defined as

$$R = \exp(\sigma),$$

with exp denoting the classic matrix exponential, satisfies the conditions in (1.1). Moreover, if $|\sigma| < \pi$, the previous σ is uniquely defined for a given R [Altmann 2005]. In this case, we can define the logarithm map log as the inverse of exp, *i.e* $\sigma = \log R$. More generally, the log map is defined for every $R \in SO(n)$ if R is "close" to I_n.

1.1.2 Geodesics and connections

In section 1.3, we will be concerned with stochastic processes on $SO(n)$. Such processes have trajectories with very complicated properties, like the diffusion processes for example. Fortunately, it is possible to construct these trajectories from much simpler curves known as geodesics. Amongst all the trajectories, geodesics are the simplest ones. They represent in $SO(n)$ the equivalent of straight lines in $\mathbb{R}^d$. They will be used later on to simulate stochastic processes and will appear naturally when trying to discretize a continuous process. Proposition 1, 2 and 3 give equivalent definitions for geodesics, successively described with different differential geometry tools.

Geodesics are always defined with respect to a given connection. In our case, it is explicitly assumed that they are defined with respect to the Levi-Civita connection [Faure 2013], presented in Equation (1.6). This connection is the most natural one to consider, as geodesics will be defined with the Riemannian metric $< ., . >$ induced from the ambient space. In our case, a geodesic is defined as a trajectory R_t with an acceleration equal to zero.

Proposition 1. *Geodesics of $SO(n)$ are given by the set of curves $R_t = \exp(t\sigma)R_0$ with $\sigma \in \mathfrak{so}(n)$.*

Proof. This proof is completed in two distinct parts. From one hand, we show that the curves R_t with $R_t = \exp(t\sigma)R_0$ are geodesics from the previous definition. From the other hand, we show that given a geodesic R_t, it can be expressed as $R_t = \exp(t\sigma)R_0$.

1. The first part consists in proving that a curve R_t such that $R_t = \exp(t\sigma)R_0$ is a geodesic. The acceleration is defined as the orthogonal projection of $\ddot{R}_t$ on $T_{R_t}SO(n)$, with respect to $< ., . >$, the inner product of the ambient space $\mathbb{R}^{n\times n}$. It is then sufficient to show that for any element $v \in T_{R_t}SO(n)$, the term $< \ddot{R}_t, v >$ is equal to zero.

Decomposing v into $v = \omega R_t$, with $\omega \in \mathfrak{so}(n)$ leads to

$$\begin{aligned} < \ddot{R}_t, v > &= < \sigma^2, \omega > \\ &= \frac{1}{2} \operatorname{trace}(\sigma^2 \omega) \\ &= \frac{1}{2} \operatorname{trace}\left((\sigma^2 \omega)^T\right) \\ &= \frac{1}{2} \operatorname{trace}\left(\omega^T (\sigma^2)^T\right) \\ &= -\frac{1}{2} \operatorname{trace}(\omega \sigma^2) \\ &= -\frac{1}{2} \operatorname{trace}(\sigma^2 \omega) \\ &= 0 \end{aligned}$$

Therefore, the orthogonal projection of $\ddot{R}_t$ on $T_R SO(n)$ is null and R_t is a geodesic.

2. The second part consists in, given a geodesic R_t, showing that there exists $\sigma \in \mathfrak{so}(n)$ constant such that $R_t = \exp(t\sigma)R_0$. Directly proving the existence of σ is quite difficult. However, it is sufficient to show $\dot{R}_t R_t^T = \dot{R}_t R_t^{-1}$ is constant.

   Proving the equality $\ddot{R}_t R_t^T + \dot{R}_t \dot{R}_t^T = 0$ is sufficient as it will show that $\dot{R}_t R_t^T$ is constant.

   From one hand, derivating two times the relation $R_t R_t^T = I_n$ shows that $\ddot{R}_t R_t^T + \dot{R}_t \dot{R}_t^T$ is skew-symmetric. From the other hand, the orthogonal projection of the acceleration is zero by definition of R_t. Therefore, $\ddot{R}_t R_t^T$ is symmetric. But $\dot{R}_t \dot{R}_t^T$ is also clearly symmetric. As a conclusion, $\ddot{R}_t R_t^T + \dot{R}_t \dot{R}_t^T$ is symmetric and skew-symmetric, thus null. In other words, $\dot{R}_t R_t^T$ is constant and R_t can be expressed as $R_t = \exp(t\sigma)R_0$ with $\sigma \in \mathfrak{so}(n)$.

□

From Proposition 1, geodesics can be written $R_t = \exp(t\sigma)R_0$ with $\sigma \in \mathfrak{so}(n)$. This expression will be useful in Section 1.2 with geodesics being used to simulate stochastic processes. Proposition 1 gives a way to compute R_t directly.

Just like straight lines in $\mathbb{R}^d$ can be defined as the solution of a differential equation with constant coefficient, geodesics on $SO(n)$ can also be defined as the flows, i.e the solution of a differential equation, of a constant combination of right invariant vector fields.

Proposition 2. *Geodesics are the solutions of*

$$\dot{R}_t = \sum_i a^i \sigma_i(R_t),$$

with the set $\{a^i\}$ *a set of* $n(n-1)/2$ *real constants.*

Proof. 1. Let R_t be a geodesic. From Proposition 1, there exists $\sigma \in \mathfrak{so}(n)$ such that $R_t = \exp(t\sigma)R_0$. Therefore, R_t is the flow of the differential equation

$$\dot{R}_t = \sigma R_t$$

Decomposing $\sigma = \sum_i a^i \sigma_i$, the previous equation reads

$$\dot{R}_t = \sum_i a^i \sigma_i(R_t)$$

and therefore, R_t is solution of a differential equation with constant right invariant vector fields.

2. Let R_t be the flow of
$$\dot{R}_t = \sum_i a^i \sigma_i(R_t)$$
with initial condition R_0. As $\exp(t\sum a^i\sigma_i)R_0$ is also a solution, by unicity of the flow, $R_t = \exp(\hat{a})R_0$. It is therefore a geodesic.

□

Proposition 2 characterizes geodesics by describing the process in its tangent space. This point of view will be used in Section 1.2, Section 1.3 and Chapter 2 to define stochastic processes.

Considering the acceleration of a trajectory rises the question of differentiating velocities and therefore the question of differentiating vector fields. In Proposition 1, the acceleration was defined as the orthogonal projection of the differential of the velocity. It would then be natural to define the differentiation of a vector field by describing the orthogonal projection of its differential in the tangent space. For a given $R \in SO(n)$, we define $\nabla_{\sigma(R)}\tau(R)$ as the differential of the vector field τ in the direction σ at the position R. The operator ∇ is called the covariant derivative. By definition,

$$\nabla_{\sigma(R)}\tau(R) = P_R \frac{d\tau(R)}{d\sigma}$$

with P_R the orthogonal projection on $T_R SO(n)$. The covariant derivative ∇ is bilinear and respects Liebniz rule

$$\nabla_{\sigma(R)}\left[f(R)\tau(R)\right] = f(R)\nabla_{\sigma(R)}\tau(R) + \frac{df(R)}{d\sigma}\tau(R).$$

By linearity, the covariant derivative is completely described by $\nabla_{\sigma^i(R)}\sigma^j(R)$, where the set $\{\sigma^i\}$ is a orthonormal basis of $\mathfrak{so}(n)$.

$$\begin{aligned}
\nabla_{\sigma^i(R)}\sigma^j(R) &= P_R\left[\frac{d\sigma^j(R)}{d\sigma^i}\right] \\
&= P_R \frac{d}{dt}\left[\sigma^j\left(\exp(t\sigma^i)R\right)|_{t=0}\right] \\
&= P_R\left[\sigma^j\sigma^i R\right] \\
&= \left(\sigma^j\sigma^i - \sigma^i\sigma^j\right)R
\end{aligned} \tag{1.5}$$

The coordinates of $\nabla_{\sigma^i(R)}\sigma^j(R)$ in the basis defined by the vector fields $\{\sigma^k(R)\}$ are denoted $C^k_{i,j}(R)$ and are called the structure constants [Hall 2003], *i.e*

$$\nabla_{\sigma^i(R)}\sigma^j(R) = C^k_{i,j}(R)\sigma^k(R). \tag{1.6}$$

Implicitly, choosing to differentiate the vector fields in the way it was presented in (1.5) is equivalent to choose the coefficients $C^k_{i,j}(R)$. In our case, these coefficients were constant. A different choice for these coefficients would lead to a different definition of the differential.

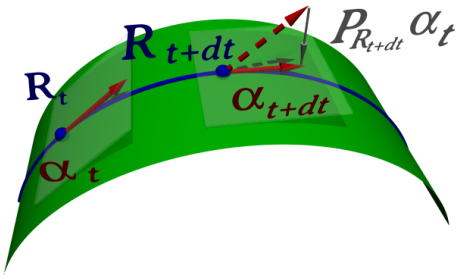


Figure 1.1: Parallel transport of the vector α (red) along R_t (blue). The vector α_{t+dt} is the orthogonal projection (grey) of α_t in the tangent space $T_{R_{t+dt}}SO(n)$.

These differentials are called connections. The connection presented in Equation 1.5 is called the Levi-Civita connection. Therefore, defining geodesics as curves with a vanishing acceleration is implicitly equivalent to considering a given connection. In the sequel, if not specifically said, the used connection will be assumed to be the Levi-Civita connection.

A vector field α is said to be parallel transported along a $\mathcal{C}^1$ trajectory $R_t \in SO(n)$ if

$$\nabla_{\dot{R}_t}\alpha_t = 0 \qquad \text{with } \alpha_t = \alpha(R_t).$$

Figure 1.1 shows an example of vector parallel transported along a curve. Despite that the figure does not represent a trajectory on $SO(n)$, it still gives an intuitive vision of the notion of parallel transport.

Denoting α_t^i the coordinates of α_t in the basis σ^i and $\dot{R}_t^j$ the coordinates of $\dot{R}_t$ in the basis σ^j, the previous equation reads

$$\begin{aligned}
\sum_i \dot{\alpha}_t^i \sigma_i(R_t) + \alpha_t^i \nabla_{\dot{R}_t}\sigma_i(R_t) &= 0 \\
\sum_{i,j} \dot{\alpha}_t^i \sigma_i(R_t) + \alpha_t^i \dot{R}_t^j \nabla_{\sigma_j(R_t)}\sigma_i(R_t) &= 0 \\
\sum_k \left(\sum_{i,j} \dot{\alpha}_t^k + \alpha_t^i \dot{R}_t^j C_{i,j}^k \right) \sigma_k(R_t) &= 0 \\
\dot{\alpha}_t^k + \sum_{i,j} \alpha_t^i \dot{R}_t^j C_{i,j}^k &= 0 \qquad \text{for all } k
\end{aligned} \tag{1.7}$$

The coordinates α_t^i are then solutions of a differential equation with coefficients determined by the velocity $\dot{R}_t$. Thus, the parallel transport and Proposition 2 lead to a third definition for geodesics.

Proposition 3. *For a curve R_t, if the velocity is parallel transported (along the curve), then R_t is a geodesic. The inverse is also true, a geodesic has its velocity parallel transported.*

Proof. 1. Let R_t be a curve such that, $\dot{R}_t$ is parallel transported along $R(t)$. Let show that R_t is a geodesic. The coordinates of $\dot{R}_t$ are denoted $\dot{R}_t^i$.

By assumption, $\dot{R}_t$ is parallel transported. Equation (1.7) holds and therefore

$$\ddot{R}_t^k = -\sum_{i,j} \dot{R}_t^i \dot{R}_t^j C_{i,j}^k$$

By definition of the coefficients $C_{i,j}^k$, and because $\nabla_{\sigma^i(R)}\sigma^j(R) = -\nabla_{\sigma^j(R)}\sigma^i(R)$ from Equation (1.5), the coefficients $C_{i,j}^k$ are skew symmetric, *i.e* $C_{i,j}^k = -C_{j,i}^k$.

Thus, $\ddot{R}^k(t) = 0$ and $R(t)$ is a geodesic (a curve with a vanishing acceleration).

2. Let R_t be a geodesic and let r_t be parallel transported along R_t with the initial condition $r_0 = \dot{R}_0$. To prove that the velocity is parallel transported, one has to show that $r_t = \dot{R}_t$ for all t. From (1.5), transporting r_t along R_t gives

$$\dot{r}_t^k = -\sum_{i,j} r_t^i \dot{R}_t^j C_{i,j}^k$$

As R_t is a geodesic, the coordinates $\dot{R}_t^j$ are constant. The previous equation is then a differential equation with constant coefficients. The initial condition $r(0) = \dot{R}(0)$ leads to $\dot{r}_t^k = 0$. This means that $r_t = \dot{R}_t$ and that $\dot{R}_t$ is parallel transported along R_t.

□

In this section, geodesics have been introduced for the Levi-Civita connection. An explicit description is given by Proposition 1. This will be useful to implement simulations of stochastic processes. These processes are defined in Section 1.2 by describing the process in its tangent space similarly to Proposition 2. Parallel transport introduced in Proposition 3 will be used mainly in Chapter 3. Before describing stochastic processes on $SO(n)$ in Section 1.2, the next paragraph highlights the case of the rotation group $SO(3)$.

1.1.3 The case of $SO(3)$

This subsection is dedicated to the study of $SO(3)$. The group $SO(3)$ has been deeply studied and more specific properties can be found in [Altmann 2005] and [Faraut 2008]. Its main interest is that it models the orientation of a rigid body in space. The orientation would indeed be represented by a basis fixed with respect to the body. The representation of this basis in a fixed referential is then an element of $SO(3)$.

In [Hubbard 1972] for example, the rotational brownian motion is modeled as a process on $SO(3)$. In [Lovera 2002], a law is described to control the orientation of a spacecraft. This orientation is modeled as an element of $SO(3)$.

The usual canonical basis for $\mathfrak{so}(3)$ is usually called J_X, J_Y, J_Z with

$$J_X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, J_Y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, J_Z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1.8}$$

These matrices are mainly used as they represent infinitesimal rotations around the canonical axis of $\mathbb{R}^3$. By infinitesimal rotation around the axis i, it should be understood that a rotation R_i around the axis $i = X, Y$ or Z satisfies, up to a normalization constant

$$\frac{d}{dt} R_t^i|_{t=0} = J_i.$$

This means that a uniform rotation around the axis $i = X$, Y or Z is given by the rotation

$$R_t^i = \exp(tJ_i) \qquad \text{for } i = X, Y, Z \tag{1.9}$$

In a more general way, a uniform rotation around an axis $u = (u^1, u^2, u^3) \in \mathbb{R}^3$ is

$$R_t = \exp\left(t \sum_{i \leq 3} u^i J_i\right) = \exp(t\hat{u}) \qquad \text{with } \hat{u} = \sum_i u^i J_i.$$

In the case of a uniform rotation, the angular velocity described with left (σ_L) and right (σ_R) invariant vector fields remains the same

$$\begin{array}{rcccl} \dot{R}(t) & = & \hat{u}R(t) & = & R(t)\hat{u} \\ & = & \sigma_R R(t) & = & R(t)\sigma_L \end{array}$$

However, in the more general case (non uniform rotations), there is no reason for these decompositions to be equal and $\sigma_L \neq \sigma_R$. In mechanics, σ_L usually stands for the angular velocity of the system in the system referential whereas σ_R represents the angular velocity of the system with respect to a fixed referential [Arnold 1966].

As $SO(3)$ is a manifold of dimension three, its elements need three parameters to be completely described. Amongst all the possible representations, the Euler angles with the convention ZYZ is described here as it will be used in the sequel. This representation will be particularly useful to describe the parallel transport in Chapter 3. A rotation $R \in SO(3)$ is completely described by the three angles $\phi \in [0, 2\pi[$, $\theta \in [0, \pi]$, $\psi \in [0, 2\pi[$ as the composition of successive rotations

$$R(\phi, \theta, \psi) = R^Z(\psi)R^Y(\theta)R^Z(\phi)$$

where R_3, R_2 are defined in (1.9). Despite that every element of $SO(3)$ can be described by a set ϕ, θ, ψ, this set is unique [Altmann 2005] only when $\theta \not\equiv 0[\pi]$. It is noticeable that R^Y and R^Z are sufficient to describe the whole set $SO(3)$ without R^X. This is due to the fact that R^X, R^Y and R^Z are subgroups with dimension 1 that are not independent.

On $SO(3)$, the symbols $C_{i,j}^k$ in (1.6) are given by

$$C_{i,j}^k = \det[e_i, e_j, e_k]$$

where e_1, e_2, e_3 is the canonical basis of $\mathbb{R}^3$.

In this section, we presented the special orthogonal group $SO(n)$ and its geometry. Two particular sets of vector fields, the vector fields respectively invariant by left and right translation were described as they will be used to define stochastic processes in the next section. A first kind of trajectories has also been described, geodesics, as they will be used to simulate processes on $SO(n)$.

1.2 Stochastic processes and Itō calculus

This section introduces stochastic calculus and the description of stochastic processes with right invariant vector fields. Stochastic processes will be used in Chapter 2 and Chapter 3 to model physical phenomenons. After a short introduction describing general tools for stochastic differential equations (denoted as SDE further on), a subsection is dedicated to numerical scheme to simulate these processes. The structure of $SO(n)$ and the multiplicative noise prevent us from using the usual scheme available for spaces like $\mathbb{R}^d$. More information about stochastic calculus on manifolds can be found in [Meyer 1981], [Itō 1950]. Finally, the forward and backward equations [Feller 1971] for Markov processes is given and used to determine the evolution of the distribution of a Markov process.

1.2.1 Stochastic differential equations on $SO(n)$

In the previous section, a particular kind of trajectories on $SO(n)$, called geodesics, were described. They were defined as the solution, up to R_0, of a differential equation such as

$$\dot{R}_t = \sigma R_t$$

for $\sigma \in \mathfrak{so}(n)$. In a more general case, it is possible to consider the same differential equation with $\sigma = \sigma_t$, *i.e* varying with time in the following way

$$\dot{R}_t = \sigma_t R_t = \hat{x}_t R_t \tag{1.10}$$

where $x_t \in \mathbb{R}^{n(n-1)/2}$ such that $\hat{x}_t = \sigma_t$.

In the case of $SO(3)$, if x_t is the instantaneous angular velocity of a rigid body, then R_t in (1.10) is the orientation of this body. If x_t can be measured precisely (in the absence of noise, without drifts, etc), then equation (1.10) can be used to determine the orientation R_t of the considered body.

However, Equation (1.10) is often not sufficient to model a system because of the presence of noise. It is used to model perturbations, interferences [Srivastava 2000] with other systems or imperfections of the sensors for example. Therefore, noise is introduced in (1.10) by using a Stratonovich integral, denoted $\circ$, like

$$dR_t = (\hat{x}_t dt + \circ d\hat{w}_t)\, R_t \tag{1.11}$$

where $w_t \in \mathbb{R}^{n(n-1)/2}$ is a Brownian motion independent from x. In the following, the process x is supposed to be a process with finite variance.

The conversion of the SDE (1.11) in an Itō integral gives

$$dR_t = \left(\hat{x}_t dt + d\hat{w}_t + \frac{1}{2}\mathbb{E}[d\hat{w}_t^2]\right) R_t \tag{1.12}$$

The usual rules of Itō's calculus can be applied to the process R_t defined in (1.12). For any function $f \in \mathcal{C}^2$ on $SO(n)$

$$df(R_t) = \sum_i \partial_i f(R_t) x^i dt + \frac{1}{2}\sum_{i,j} \partial_{i,j} f(R_t)\mathbb{E}[dw^i dw^j] \qquad \text{with } \partial_i f = \frac{df}{d\sigma_i}. \tag{1.13}$$

Working with one form or another (Stratonovich or Itō) depends of the application and the same arguments in favor of one of the forms are the same as in the usual case of SDE defined on $\mathbb{R}^n$.

For example, the Itō lemma (1.13) makes the derivation of composition more complex than the Stratonovich calculus rules, which are simpler. The derivation is similar to the derivation for classic processes

$$df(R_t) = \sum_i \partial_i f \circ dR_t^i \sigma_i(R_t).$$

However, the Itō integral has useful properties like the Itō isometry or the martingale property (the Itō integral remains a martingale if R_t is a martingale [Williams 2000]).

Generally, stochastic processes will be defined with a Stratonovich integral (1.11) but the evolution of functional of these processes will be treated thanks to Equation (1.13) (Section 1.4 for example). Despite the definitions of these processes, one might be looking at numerically manipulate them. In order to do so, the first step is the possibility to simulate these processes.

1.2.2 Simulations of stochastic processes

In order to manipulate stochastic processes as defined in the previous subsection, one might be interested in making a simulation of this process [Iacus 2010]. For example, the simulated annealing presented in Section 1.3 can be used to generate an optimization algorithm based on the simulation of a stochastic process.

Because it will not be numerically possible to generate time continuous trajectories, simulations try to construct a discrete chain which is the sampled version of the considered process. In other words, instead of computing the process on a continuous time period $[0, t]$, the process is sampled at n successive times $0 = t_0 \leq t_1, ..., t_n = t$.

Despite that nothing refrains from using a different time step $\delta t_i = t_{i+1} - t_i$, only the case with a constant time step, denoted δt is considered here.

Before working with SDE on $SO(n)$, we present a reminder for numerical scheme for SDE defined on $\mathbb{R}^d$. Then, an adaptation of the presented method to $SO(n)$ can be realized. Consider the process $y_t \in \mathbb{R}^d$ as

$$dy_t = x_t dt + dw_t$$

with w_t a Brownian motion in $\mathbb{R}^d$.

By definition of dy_t, the term $\delta y_{n\delta t} = y_{(n+1)\delta t} - y_{n\delta t}$ is defined by the integral

$$\delta y_{n\delta t} = \int_{n\delta t}^{(n+1)\delta t} x_s ds + dw_s.$$

Sampling y_t with a time step δt is then the same as computing $y_{n\delta t}$ for $n \geq 0$.

However, x_t is often not known. For example, only discrete values $x_{n\delta t}$ are known (sampling from sensors for example). In this case, the process x_t is approximated as being constant for a time δt. The term $\delta y_{n\delta t}$ is thus approximated by

$$\delta y_{n\delta t} = \int_{n\delta t}^{(n+1)\delta t} x_s ds + dw_s = x_{n\delta t}\delta t + o(\delta t) + \delta w_{n\delta t}$$

Let define the chain $\tilde{y}_n$, indexed by n as

$$\delta \tilde{y}_n = \tilde{y}_{n+1} - \tilde{y}_n = x_{n\delta t}\delta t + \delta w_{n\delta t} \tag{1.14}$$

with $\delta w_{n\delta t} = w_{(n+1)\delta t} - w_{n\delta t}$. By taking $\tilde{y}_0 = y_0$ as initial condition, the whole chain can be generated with Equation (1.14).

If x is constant during δt, the chain $\tilde{y}$ would be a sampling of y instead of an approximation, *i.e* $\tilde{y}_n = y_{n\delta t}$. In a general case, $\tilde{y}$ converges towards y when the time step shrinks to 0. In a intuitive way, this means that the simulation is, over a finite time interval, as accurate as one needs with a sufficient small computation time step.

Proposition 4. *The chain $\tilde{y}$ obtained from (1.14) converges locally uniformly in the mean square sense, towards y, i.e* $\lim_{\delta t} \mathbb{E}\left[\sup_{n\delta t \leq t} |\tilde{y}_n - y_{n\delta t}|^2\right] = 0$

Proof.

$$\begin{aligned}\mathbb{E}\left[\sup_{n\delta t\leq t}|\tilde{y}_n - y_{n\delta t}|^2\right] &= \mathbb{E}\left[\sup_{n\delta t\leq t}\left|\sum_{k\leq n} x_{k\delta t}\delta t - \int_{k\delta t}^{(k+1)\delta t} x_s ds + \delta w_{n\delta t} - \int_{k\delta t}^{(k+1)\delta t} dw_s\right|^2\right] \\ &= \mathbb{E}\left[\sup_{n\delta t\leq t}\left|\sum_{k\leq n}\int_{k\delta t}^{(k+1)\delta t} x_s - x_{k\delta t} ds\right|^2\right] \\ &\leq \sum_{k\leq n}\sup_{k\leq s/\delta t\leq k+1}\mathbb{E}\left[|x_s - x_{k\delta t}|^2\right]\delta t^2\end{aligned} \tag{1.15}$$

As x has a locally finite variance, the term in the sum of the last line is bounded. As the length of the intervals $[k\delta t,(k+1)\delta t]$ is decreasing, it can be majored by a constant, showing the convergence. □

The chain $\tilde{y}_n$ defined from (1.14) can then be used as a simulation of the stochastic process y_t over a finite time interval. The scheme from (1.14) is called Euler scheme [Iacus 2010].

However, even if this method seems intuitive, using a scheme similar to (1.14) to simulate processes R_t with a chain $\tilde{R}_n$ on $SO(n)$ is not possible. This is due to the operator δ in (1.14) for the construction of $\tilde{R}_n$. As $SO(n)$ is not stable under addition, $\tilde{R}_n$ would no more be an element of $SO(n)$.

In fact, (1.12) is considering R_t as a matrix element of the ambient space $\mathbb{R}^{n\times n}$ but does not take into account the structure of $SO(n)$. The solution consists in adapting the Euler scheme to the structure of $SO(n)$. In (1.14), the approximation was to consider x_t constant during δt. The increments are then given by moving in the direction $x_{n\delta t}$ during δt along the geodesic $S_t = \exp(t\hat{x}_{n\delta t})S_0$ with $S_0 = \tilde{R}_{n\delta t}$. This is realized with the operator δ for $\mathbb{R}^d$. However, the same idea extended to $SO(n)$ leads to the following scheme, called Mc Kean scheme [McKean 1960]

$$\tilde{R}_{n+1} = \exp\left(\hat{x}_{n\delta t}\delta t + \delta\hat{w}_{n\delta}\right)\tilde{R}_n \tag{1.16}$$

In this case, it appears clearly that, if $\tilde{R}_0 \in SO(n)$, then the whole chain is in $SO(n)$. Note that the use of the exponential map in (1.16) induces that the considered displacement are made along geodesics of $SO(n)$, described in Proposition 1.

Similarly to the proposition 4 concerning the case of simulation in $\mathbb{R}^d$, there is a uniform convergence for the simulation as δt shrinks to 0.

Proposition 5. *The chain $\tilde{R}$ obtained from (1.16) converges locally uniformly in the mean square sense, towards R, i.e* $\lim_{\delta t}\mathbb{E}\left[\sup_{n\delta t\leq t}|\tilde{R}_n - R_{n\delta t}|^2\right] = 0$

Proof. A development of the exp term in (1.16) gives

$$\tilde{R}_{n+1} = \left(I_n + \hat{x}_{n\delta t} + \delta\hat{w}_{n\delta t} + \frac{1}{2}\mathbb{E}[\delta\hat{w}^2_{n\delta t}] + o(\delta t)\right)\tilde{R}_n$$

The difference between $\tilde{R}_n$ and the chain $\tilde{R}^1_n$ defined by

$$\tilde{R}^1_{n+1} = \left(I_n + \hat{x}_{n\delta t} + \delta\hat{w}_{n\delta t} + \frac{1}{2}\mathbb{E}[\delta\hat{w}^2_{n\delta t}]\right)\tilde{R}^1_n$$

is negligible with respect to δt. Therefore, $\tilde{R}$ converges towards $\tilde{R}^1$. From Proposition 4, $\tilde{R}^1$ converges towards R. The transitivity of the convergence completes the proof. □

The scheme from (1.16) is the numerical scheme that will be used to simulate stochastic processes on $SO(n)$. Because of the error term from (1.15), the smaller δt is for computing $\tilde{R}_n$, the more accurate the simulation is. It is noticeable that the construction of this scheme can be generalized to a different manifold from $SO(n)$, once the geodesics are known. Despite that stochastic processes can now be simulated, other information like the evolution of the distribution or the existence of a stationary distribution might be required. Some of these properties can be determined for a specific class of stochastic processes, namely the Markov processes.

1.2.3 Markov processes

A stochastic process R_t is said to be a Markov process if for any subset $A \subset SO(n)$

$$\mathbb{P}(R_t \in A|\mathcal{F}_s) = \mathbb{P}(R_t \in A|R_s)$$

where $\mathcal{F}_s = \{R_r, r \leq s\}$. This signifies that the evolution of the process can be deduced only with the last observation. Previous observations are not giving more information about the future state of R_t.

For a process R_t defined with an SDE such as

$$dR_t = (\hat{x}_t dt + \circ d\hat{w}_t)\, R_t,$$

it can be a Markov process or not, depending on how x_t is defined.

A Markov process R_t will be called stationary if $\mathbb{P}(R_t \in A|R_s) = \mathbb{P}(R_{t-\tau} \in A|R_{s-\tau})$. This signifies that the evolution of R_t only depends of the time difference with the last observation but not the time of the observation itself.

For a Markovian and stationary process R_t, let T_t be the linear operator

$$[T_t f](r) = \mathbb{E}[f(R_t)|R_0 = r] \tag{1.17}$$

for any function $f \in \mathcal{C}^2(SO(n) \to \mathbb{R})$.

Proposition 6. *The set of operators T_t is a semigroup[1] with $T_s T_t = T_{t+s}$.*

Proof. The proof is here limited to the case when there exists a transition density $p_t^s(r, .)$ to describe the distribution of R_t knowing that $R_s = r$

$$\mathbb{P}(R_t \in A|R_s = r) = \int_{SO(n)} p_t^s(r, R)\mathbf{1}_A(R)dR$$

with $\mathbf{1}_A$ the set function of A.

In this case, T_{t+s} can de decomposed

$$\begin{aligned}
[T_{t+s}f](r) &= \int_{SO(n)} p^0_{t+s}(r,R)f(R)dR \\
&= \iint_{SO(n)^2} p^0_t(r,R')p^t_{t+s}(R',R)f(R)dRdR' \qquad \text{(Markovian property)} \\
&= \iint_{SO(n)^2} p^0_t(r,R')p^0_s(R',R)f(R)dRdR' \qquad \text{(stationarity)} \\
&= \int_{SO(n)} p^0_t(r,R') \int_{SO(n)} p^0_s(R',R)f(R)dRdR' \\
&= \int_{SO(n)} p^0_t(r,R')[T_sf](R')dR' \\
&= [T_sT_tf](r)
\end{aligned}$$

The neutral element is $T_0 = Id$. Thus, the set $\{T_t\}_t$ is a semigroup of operators. □

When the limit exists, let $\mathcal{A}$ be the operator defined as

$$\lim_t \frac{T_tf - T_0f}{t} = \mathcal{A}f.$$

This operator is linear and commutes with any element T_t. Because of these properties, T_t is generally written $T_t = e^{t\mathcal{A}}$. The generator $\mathcal{A}$ is called the Markov generator of R_t [Feller 1971]. The operator $\mathcal{A}$ is the equivalent of the transition matrix for finite state Markov chain. It will be used to describe the infinitesimal evolution of the distribution of R_t.

Proposition 7. *For a stationary Markov process R_t defined via the SDE*

$$dR_t = (\hat{x}_t dt + \circ d\hat{w}_t)R_t,$$

the generator $\mathcal{A}$ can be written

$$\mathcal{A}f(r) = \sum_i x^i(t)\partial_i f(r) + \frac{1}{2}\sum_{i,j} \partial_{i,j}f(r)\mathbb{E}[dw^i_t dw^j_t]. \tag{1.18}$$

Proof.

$$\begin{aligned}
\frac{1}{t}\left([T_tf](r) - [T_0f](r)\right) &= \frac{1}{t}\left([T_tf](r) - f(r)\right) \\
&= \frac{1}{t}\mathbb{E}[f(R_t) - f(R_0)|R_0 = r] \\
&= \frac{1}{t}\int_0^t \sum_i \partial_i f(R_t)x^i + \frac{1}{2}\sum_{i,j}\partial_{i,j}f(R_t)\mathbb{E}[dw^i dw^j]dt
\end{aligned}$$

The last line is obtained thanks to the Itō lemma (1.13). As R_t is continuous and $f \in \mathcal{C}^2$, the limit for $t \to 0$ is given in (1.18). □

The generator $\mathcal{A}$ leads to the following forward and backward equations. By denoting

$u(t,r) = [T_t f](r)$, the evolution of u is given by the backward equation

$$\frac{du}{dt} = \lim_h \frac{T_{t+h}f - T_t f}{h} = \lim_h T_t \frac{T_h f - f}{h}$$
$$\frac{du}{dt} = \mathcal{A}u \tag{1.19}$$

For the inner product $< f, p >= \int f(R)p(R)dR$ on $L^2(SO(n) \to \mathbb{R})$, the operator $\mathcal{A}^\star$ is the operator adjoint to $\mathcal{A}$ with respect to $< ., . >$, *i.e*

$$< \mathcal{A}f, p >=< f, \mathcal{A}^\star p > \qquad \text{for all } f, p \in L^2(SO(n) \to \mathbb{R}).$$

Proposition 8. *The evolution of p_t, the density of R_t, is given by the forward equation*

$$\frac{dp_t(R)}{dt} = \mathcal{A}^\star p_t(R) \tag{1.20}$$

Proof. By definition, $u(t,r) = [T_t f](r)$. Under an integral form

$$u(t,r) = \int_{SO(n)} p_t^0(r,R) f(R) dR$$
$$\frac{du(t,r)}{dt} = \int_{SO(n)} \frac{p_t^0}{dt}(r,R) f(R) dR$$
$$\mathcal{A}u(t,r) = \int_{SO(n)} \frac{p_t^0}{dt}(r,R) f(R) dR \qquad \text{from (1.19)}$$

Replacing u with its definition on the left side of the previous equation gives

$$\int_{SO(n)} p_t^0(r,R) \mathcal{A}f(R) dR = \int_{SO(n)} \frac{p_t^0}{dt}(r,R) f(R) dR$$
$$\int_{SO(n)} \mathcal{A}^\star p_t^0(r,R) f(R) dR = \int_{SO(n)} \frac{p_t^0}{dt}(r,R) f(R) dR$$

This equality is true for all the test functions $f \in \mathcal{C}^2$ so

$$\frac{dp_t^0}{dt}(r,R) = \mathcal{A}^\star p_t^0(r,R)$$

Writing $p_t(R) = p_0(r) p_t^0(r,R)$ with p_0 the distribution of R_0 leads to the result (1.20). □

In the case of a generator $\mathcal{A}$ defined as in (1.18), the adjoint generator (obtained via an integration by part) is given by

$$\mathcal{A}^\star p_t = -\sum_i \partial_i \left[p_t x_t^i\right](x) + \frac{1}{2} \sum_{i,j} \partial_{i,j} \left[p_t \mathbb{E}[dw_t^i dw_t^j]\right] \tag{1.21}$$

The forward equation, also called heat equation, is a very powerful tool for the study of Markov processes. It describes the evolution of the distribution of R_t thanks to linear operators. This allows us to use tools and results from linear algebra to solve stochastic problems.

A distribution $p^\star$ is said stationary if $p_0 = p^\star$ implies $p_t = p^\star$ for every $t \geq 0$. In other words, it means that the distribution of the process does not change with time. In this case, as p_t remains constant, a stationary distribution can be determined, if it exists, as the solution of $\mathcal{A}^\star p^\star = 0$. Aside from the normalization constraints for the distributions, the

stationary distributions are the eigenvectors of $\mathcal{A}^\star$ associated with a null eigenvalue. These distributions will be important as they might be, under certain conditions, the limit of the distribution p_t as $t \to \infty$. In this case, they will give the equilibrium distribution of the process. This will be used in the next section for the example of the simulated annealing. The idea will be to construct a process with a given stationary distribution. After a long enough time, this process can be considered as a random variable sampled from the aimed stationary distribution.

1.3 Examples of stochastic differential equations

This section highlights two different examples of stochastic processes. For each of them, a numerical simulation is realized based on (1.16). These processes will be Markovian and the expression of the generator is given as well as specific properties for each process. First of all, the most basic example is the Brownian motion on $SO(n)$ (subsection 1.3.1). The Markov generator will be used to prove an extension of the central limit theorem for random variables on $SO(n)$. Second, a method for sampling from a density $p^\star$ is presented. It consists in constructing a stochastic process such that its stationary distribution will be $p^\star$. This method is then extended to solve an optimization problem.

1.3.1 Brownian process

The Brownian motion B_t on $SO(n)$ is defined as

$$dB_t = (\circ d\hat{w}_t)\, B_t \qquad \text{where } \hat{w} = \sum_i w_i \sigma^i \tag{1.22}$$

with w a Brownian motion on $\mathbb{R}^{n(n+1)/2}$ with a covariance matrix $\sigma_w I_{n(n+1)/2}$.

To simulate a Brownian motion B_t, one can use the chain $\tilde{B}_n$ defined with the scheme (1.16). For a time step δt, this scheme applied to (1.22) leads to

$$\tilde{B}_{n+1} = \exp\left(\delta \hat{w}_{n\delta t}\right) \tilde{B}_n \tag{1.23}$$

An example of trajectory of B_t is represented in Figure 1.2 for $B_t \in SO(3)$. Among all the possible representations for trajectories in $SO(3)$ (action on S^2, Euler angles, etc), the chosen representation is the display of the trajectories of each column of B_t. Each column representing a unit vector in the canonical basis of $\mathbb{R}^3$, each trajectory is a process on the sphere S^2, here displayed with a distinct color for each column. Another representation would not be able to present the geometry of the process without distortions. For example, the Euler angles of B_t would show discontinuities, even if the Brownian motion is continuous. It is noticeable that even if three processes on S^2 are represented, it does not correspond to a process of dimension 6 as each component is not independent of the other ones.

It is noticeable that the expression of $\tilde{B}_n$ is not direct as it would be on $\mathbb{R}^n$. Because the successive increments $\hat{w}_{n\delta t}$ of (1.23) cannot be gathered under the same exp, every small random rotation $\exp\left(\delta \hat{w}_{n\delta t}\right)$ has to be generated separately:

$$\tilde{B}_n = \exp\left(\delta \hat{w}_{n\delta t}\right) \exp\left(\delta \hat{w}_{(n-1)\delta t}\right) \ldots \exp\left(\delta \hat{w}_{0\delta t}\right) \tilde{B}_0. \tag{1.24}$$

Therefore, $\tilde{B}_n$ cannot be generated before the generation of $\tilde{B}_{n-1}$. To generate one element of the chain, all the previous elements need to be generated before. This is a major difference with the case of a Brownian motion w_t on $\mathbb{R}^d$. However, one should keep in mind that the small random rotation $\exp\left(\delta \hat{w}_{n\delta t}\right)$ is itself, roughly speaking, the approximation of the product of infinitesimal random rotations during a time δt.

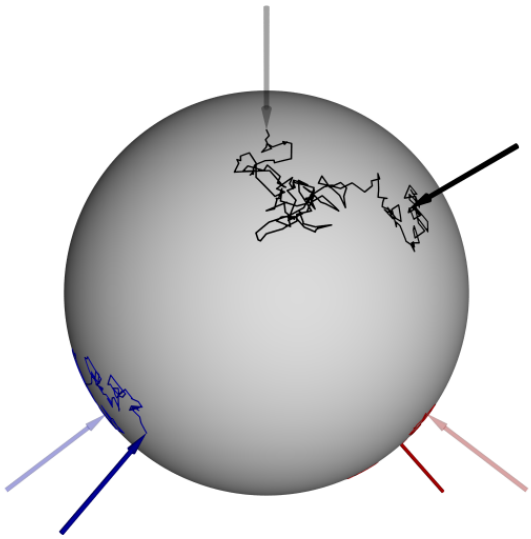


Figure 1.2: Example of trajectory of $B_t \in SO(3)$ for $t \in [0, 0.1]$ and $\sigma_w = 1$. Each color represents the trajectory of a different column of B_t: blue for b_1, red for b_2 and black for b_3 with the convention $B_t = [b_1, b_2, b_3]$. The path on the sphere is the trace of the whole process. Its value at $t = 0$ is marked by transparent arrows pointing at $b_1(0)$, $b_2(0)$ and $b_3(0)$. Its value at $t = 1$ is marked by opaque arrows. Remark: Each individual trajectory is also a Brownian motion on the sphere S^2.

The approximation of Equation (1.24) is constituted of the accumulation of the effect of small Gaussian random variables. It is then legitimate to wonder if this approximation still holds for a different kind of random variables, like it is the case in the usual central limit theorem. First of all, we need to show that B_t is a Markov process. From (1.22), B_t is completely determined by the values of dw_s for $s \leq t$. Therefore, for a subset $A \in SO(n)$,

$$\begin{aligned}\mathbb{P}\left(B_t \in A | B_{s_1}, s_1 \leq s\right) &= \mathbb{P}\left(B_t \in A | dw_{s_1}, s_1 \leq s\right) \\ &= \mathbb{P}\left(B_t \in A | B_s, dw_{s_1}, s_1 \leq s\right) \\ &= \mathbb{P}\left(B_t \in A | B_s\right).\end{aligned}$$

The last line is due to the independence of the increments dw_s.

As B_t is a Markov process, it is possible to define its generator $\mathcal{A}$. From (1.18), this generator is

$$\mathcal{A}f(r) = \frac{\sigma_w^2}{2}\sum_i \partial_{i,i} f(r) = \frac{\sigma_w^2}{2}\Delta f(r). \tag{1.25}$$

We will now look for the stationary distributions. A stationary distribution $p^\star$ is a solution of the equation $\mathcal{A}^\star p^\star = 0$. From the expression (1.25) of $\mathcal{A}$, the equation $\mathcal{A}^\star p^\star = 0$ only has one solution, the uniform distribution over $SO(n)$. Therefore, the stationary distribution for the Brownian motion is the uniform distribution. It is noticeable that the uniform distribution exists on $SO(n)$ because $SO(n)$ is compact. For a Brownian motion in $\mathbb{R}^d$ for example, there is no stationary distribution, the process would diffuse in the whole space.

The expression (1.25) of the generator of B_t can be used to prove an extension of the central limit theorem on $SO(n)$. For a fixed value t and a fixed δ, let R_n^δ be the chain defined as

$$R_{n+1}^\delta = \exp(\sqrt{t\delta}\hat{\zeta}_n) R_n^\delta \tag{1.26}$$

with ζ a centered random variable of $\mathbb{R}^{n(n-1)/2}$ with unit variance. Theorem 1 states that for $\delta = 1/n$, the chain R^δ will tend to have the same distribution as a Brownian motion.

Theorem 1. *The limit $\lim_n R_n^{1/n} = B_t$ for $R^{1/n}$ defined in (1.26) is a Brownian motion, in the distribution sense.*

Proof. Theorem 19.25 of [Kallenberg 2002] states that it is enough to show the convergence of the generators to show the convergence in distribution. In other words, it is sufficient to show that the generator of the chain $\mathcal{A}_n$ defined as

$$\mathcal{A}_n f\left(r\right) = \frac{1}{t/n}\mathbb{E}[f\left(R_{n+1}^{1/n}\right) - f\left(R_n^{1/n}\right) | R_n^{1/n} = r]$$

converges towards the generator of a Brownian motion $\mathcal{A}$ with $\mathcal{A}f = \frac{1}{2}\Delta f$.

A second order development of $f\left(R_{n+1}^{1/n}\right)$ at $R_n^{1/n}$ gives

$$f\left(R_{n+1}^{1/n}\right) = f\left(R_n^{1/n}\right) + \sum_i \partial_i f\left(R_n^{1/n}\right)\zeta_n^i\sqrt{\frac{t}{n}} + \sum_{i,j}\partial_{i,j} f\left(R_n^{1/n}\right)\zeta_n^i\zeta_n^j\frac{t}{n} + o\left(\frac{1}{n}\right)$$

Fixing $R_n^{1/n} = r$, the expected value of the previous equation reads

$$\mathbb{E}[f\left(R_{n+1}^{1/n}\right) - f\left(R_n^{1/n}\right) | R_n^{1/n} = r] = \sum_i \partial_{i,i} f(r)\frac{t}{n} + o\left(\frac{1}{n}\right)$$

as ζ is centered with unit variance.

Therefore, the operator $\mathcal{A}_n f$ converges towards

$$\begin{aligned}\mathcal{A}_n f(r) &= \frac{1}{t/n}\mathbb{E}[f\left(R_{n+1}^{1/n}\right) - f\left(R_n^{1/n}\right)|R_n^{1/n} = r] \\ &= \frac{1}{t/n}\sum_i \partial_{i,i} f(r)\frac{t}{n} + o\left(\frac{1}{n}\right) \\ &= \sum_i \partial_{i,i} f(r) + o(1) \\ &\to \Delta f(r) \\ &= \mathcal{A} f(r)\end{aligned}$$

Thus, the generator of B_t is the Laplacian and B_t is a Brownian motion. □

Therefore, Theorem 1 can be used to replace the increments $\delta\hat{w}_{n\delta t}$ in (1.23) by centered random variables with unit variance. In this case, as δt shrinks to 0, the Markov generator will gets closer to the Laplacian operator Δ and the simulation will be closer to the stochastic process, similarly to the central limit theorem for random variables on $\mathbb{R}^d$.

The next subsection presents a Monte-Carlo method for sampling random variable from a given distribution and the adaptation of this method for the Simulated annealing, an optimization algorithm.

1.3.2 Simulated annealing

The subsection gives another example of a stochastic process. First of all, a Monte-Carlo sampling method is presented, based on the results of Subsection 1.2.3. This method will be used to construct an optimization algorithm called simulated annealing. Deeper results concerning the simulated annealing for the case of processes defined on $\mathbb{R}^d$ can be found in [Johnson 2003].

Several stochastic algorithms [Casella 2010] require to sample a random variable R from a given distribution $p^\star$. A Monte-Carlo method can be used to obtain R by constructing a process R_t whose stationary distribution is $p^\star$ and considering $R = R_T$ for T large enough. Assuming that the distribution of R_t will converge towards $p^\star$ for large values of t [Johnson 2003], Proposition 9 gives a method to construct such a process.

Proposition 9. *Given a function $f \in \mathcal{C}^1(SO(n) \to \mathbb{R})$. Let R_t be defined as*

$$dR_t = \left(\widehat{\text{grad}} f(R_t)dt + \circ d\hat{w}_t\right) R_t \tag{1.27}$$

with w_t a Brownian process with variance $\sigma_w^2 I_{n(n-1)/2}$ and grad *the gradient operator. This process has a stationary distribution $p^\star$, up to a normalization constant, with*

$$p^\star = \exp(2f/\sigma_w^2). \tag{1.28}$$

Proof. Injecting equation (1.27) into (1.21) gives the following expression for $\mathcal{A}^\star$

$$\mathcal{A}^\star p = -\sum_i \partial_i[p\partial_i f] + \frac{\sigma_w^2}{2}\sum_i \partial_{i,i} p.$$

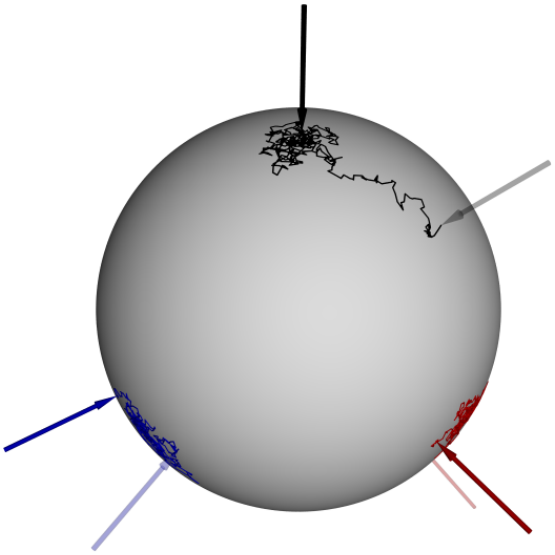


Figure 1.3: Example of trajectory of R_t defined in (1.27) with $f(R) = \text{trace}(R)$. Each color represents the trajectory of a different column of R_t: blue for r_1, red for r_2 and black for r_3 with the convention $R_t = [r_1, r_2, r_3]$. Its value at $t = 0$ is marked by transparent arrows pointing at $r_1(0)$, $r_2(0)$ and $r_3(0)$. Its value at $t = 1$ is marked by opaque arrows.. The process R_t stabilizes itself around I_3

As $\partial_i p^\star = \frac{2}{\sigma_w^2} (\partial_i f) p^\star$, injecting $p^\star$ into the previous equation leads to

$$\begin{aligned}
\mathcal{A}^\star p^\star &= -\sum_i \partial_i [p^\star \partial_i f] + \frac{\sigma_w^2}{2} \sum_i \partial_{i,i} p^\star \\
&= -\sum_i \partial_i [p^\star \partial_i f] + \frac{\sigma_w^2}{2} \sum_i \frac{2}{\sigma_w^2} \partial_i [p^\star \partial_i f] \\
&= 0
\end{aligned}$$

Therefore, $p^\star$ is a stationary distribution. □

To sample from the given distribution $p^\star$, let R_t be the process constructed from Equation (1.27) with $f = \frac{\sigma_w^2}{2} \log(p^\star)$. After simulating R_t for a time T long enough, one can consider that $R_T \sim p^\star$.

An example of trajectory is presented in Figure 1.3. The process tends to be stationary around I_3. This is expected as $f(R) = \text{trace}(R)$ for the simulation. From an intuitive point of view, this is justified as the drift of (1.27) tends to concentrate the process around a local maximum of f. Without noise, this is the classic gradient descent algorithm and the process would just converge to the local maximum. The presence of the noise acts here as a counter-balance to the drift. As the drift tends to stabilize the process, the noise is pushing it away from its equilibrium point. As a result, the process is distributed around the high values of f.

Moreover, it is noticeable that the high values of $p^\star$ coincide with the high values of f.

Therefore, R_t tends to be more concentrated around the high values of f. In this case, the process R_t can then be used to determine the position of the high values of f.

The simulated annealing is a Monte-Carlo method for optimization problems. Its aim is to find the maximum of a given function $f \in \mathcal{C}^1(SO(n) \to \mathbb{R})$. In order to do so, the gradient method, adapted to $SO(n)$, could be used if f is convex [Smith 1998]. It is a method that constructs the flows of the gradient of f. It stabilizes at the maximum, where the gradient is null. In other cases (f non convex), f will exhibit several maxima and the gradient method will not be able to differentiate local and global maxima. To avoid this problem, Monte-Carlo methods using stochastic processes are used.

As for a process R_t defined as in (1.27) the stationary distribution $p^\star$ is given by (1.28), as σ_w shrinks to 0, $p^\star$ will tends to be concentrated around the global maxima of f. The simulated annealing consists in constructing a process R_t as in (1.27) but with the noise parameter σ_w shrinking to 0. Thus, R_t will stabilize around the global maxima of f.

However, in practice, $\sigma_w(t)$ should not decrease too quickly. If that was the case, the process would not be able to explore the whole space $SO(n)$ and the different modes of f. Keeping σ_w with a "high" value during a long enough time is important to ensure that the process visits the different convex areas of f.

From another point of view, it is noticeable that as σ_w is no more a constant, $R(t)$ is no more stationary. If σ_w is varying too quickly, results concerning stationary Markov process like proposition 9 are no more available. It is reasonable to consider that if σ_w is slowly decreasing, the results still hold and the previous stationary distribution can be considered as an initial condition for the next value of σ_w. In practice, a logarithmic decrease should be used.

This method requires the knowledge of ∇f to construct the process R_t. This algorithm is therefore not applicable when f is not $\mathcal{C}^1$ or when the gradient is not available. In this case, a Metropolis-Hasting algorithm can be used [Casella 2010]. Despite that this method does not require a gradient, it might need to generate several candidates before moving to a different position. This is not the case with the method constructed from Proposition 9.

Results from Subsection 1.2.3 have been used for the study of these two examples of stochastic processes, the Brownian motion and the simulated annealing. The simulation of the Brownian motion leads to an extension of the central limit theorem on $SO(n)$. The simulated annealing uses the description of the stationary distribution from the Markov generator to construct a sampling Monte-Carlo method.

The next section describes another tool for random variables on $SO(n)$: the irreducible representation. It will mainly be used in Chapter 3 to determine the distribution of the multiplication of independent and identically distributed random variables on $SO(n)$.

1.4 Irreducible representation of $SO(n)$

This section serves as a quick introduction to the irreducible representation of random variables on $SO(n)$. The presented results are known and only consist in a reminder. More results can be found in [Kyatkin 2001] or [Faraut 2008] amongst the vast literature on the subject. Subsection 1.4.1 gives the definition of an irreducible representation and states the Peter Weyl theorem (Theorem 2). It gives a decomposition formula for distributions on $SO(n)$ thanks to their Fourier coefficients. Subsection 1.4.2 gives an example of use of the Peter-Weyl theorem to determine the Fourier coefficients of the distribution of a Brownian motion. The method of computation of these coefficients leads to another proof of the central limit theorem presented in Section 1.3. The results concerning the composition of independent random variables will mainly be used in Chapter 3.

1.4.1 Characteristic functions

A representation of the group $SO(n)$ is an homomorphism $\lambda : SO(n) \to GL(V)$ where V is a Hilbert space. This representation is said reducible if there exists an invariant subset $W \in V$ (in other words $\lambda_R(W) \subset W$) for all $R \in SO(n)$. A non reducible representation is said irreducible. Thus, the only invariant subspaces are trivial, *i.e* $\{0_V\}$ and V.

Two irreducible representations $\lambda_i : SO(n) \to GL(V_i)$ for $i = 1, 2$ are said equivalent if there exists an invertible map $L : V_1 \to V_2$ such that $L \circ \lambda_1(R) = \lambda_2(R) \circ L$ for $R \in SO(n)$. This relation of equivalence defines classes of equivalence $[\lambda]$ amongst the representations and $\mathrm{Irr}(SO(n))$ will denote the countable set of all the equivalence classes. All the elements of $[\lambda]$ have the same dimension $d_\lambda \times d_\lambda$. They can therefore be represented by a $d_\lambda \times d_\lambda$ matrix. For a better readability, an element of $\mathrm{Irr}(SO(n))$ will be denoted λ instead of $[\lambda]$.

Among all the representations from the class λ, there exists [Lochak 1959] a unitary representation $U^\lambda \in L^2(SO(n))$ (considered as matrix) such that $U^\lambda U^{\lambda T} = I_{d_\lambda}$. Let λ_0 denotes the unit representation $U^{\lambda_0}(R) = I_{d_{\lambda_0}}$. Because U^λ is an homomorphism of $SO(n)$, it means that

$$U^\lambda(R_1 R_2) = U^\lambda(R_1) U^\lambda(R_2) \tag{1.29}$$

For a function $h \in L^2(SO(n))$, let $\hat{h}^\lambda$ be the matrix which elements $\hat{h}^\lambda_{m,n}$ with $-l \leq m, n \leq l$, are defined with the usual inner product for L^2

$$\hat{h}^\lambda_{m,n} = < h, U^\lambda_{m,n} > = \int_{SO(n)} h(R) U^\lambda(R) dR \tag{1.30}$$

The elements of the matrix $\hat{h}^\lambda$ are called the Fourier coefficients of h. Theorem 2 gives a formula to reconstruct h from its Fourier coefficients.

Theorem 2. *Any function $f \in L^2(SO(n), \mathbb{R})$ can be reconstructed via its Fourier transform $\hat{h}^\lambda$ as*

$$h(R) = \sum_{\lambda \in \mathrm{Irr}(SO(n))} d_\lambda \,\mathrm{trace}\left(\hat{h}^\lambda U^\lambda(R)\right). \tag{1.31}$$

The series converges uniformly.

Proof. The proof is omitted here but can be found in [Lochak 1959]. □

This theorem is equivalent to the reconstruction of periodic function from their Fourier transform. It can be used to determine the distribution of a random variable by computing its Fourier coefficients.

For a random variable $R \in SO(n)$, the characteristic function ϕ_R is defined as

$$\phi_R(\lambda) = \mathbb{E}[U^\lambda(R)], \qquad \lambda \in \mathrm{Irr}(SO(n)). \tag{1.32}$$

If the distribution of R can be described with a density p, then $\phi_R(\lambda) = \hat{p}^\lambda$ with $\hat{p}^\lambda$ defined in (1.30).

Proposition 10. *For $R_1, R_2 \in SO(n)$ two random variables, the following properties hold:*

1. *If R_1, R_2 are independent, then*

$$\phi_{R_1 R_2}(\lambda) = \phi_{R_1}(\lambda)\phi_{R_2}(\lambda)$$

2. *If $\phi_{R_1} = \phi_{R_2}$, then R_1 and R_2 are equal in the distribution sense.*

Proof. From (1.29) and Theorem 2, the proof is direct. □

Part 1 of Proposition 10 is useful for example when studying a product of independent identically distributed random variables. Part 2 can be used to demonstrate the equality of two random variables in the sense of the distribution, or the convergence of a sequence of random variables (however, the proper theorem for the convergence is here ommited). They are both adapted for example to demonstrate the central limit theorem on $SO(n)$ (in the next section).

Similarly to the case of real random variables, the characteristic function vanishes for $\lambda \neq \lambda_0$ for uniform distribution.

Proposition 11. *If R is uniformly distributed on $SO(n)$, then for $\lambda \neq \lambda_0$, $\phi_R(\lambda) = 0$.*

Proof. Let us suppose that $\phi_R(\lambda) \neq 0$. It means that there exists $v \in \mathbb{R}^{d_\lambda}$ such that $\phi_R(\lambda)v \neq 0$. Let show that $\phi_R(\lambda)v$ is an eigenvector for any $U^\lambda(S)$ for $S \in SO(n)$.

$$\begin{aligned}
U^\lambda(S)\phi_R(\lambda)v &= U^\lambda(S)\left(\int_{SO(n)} U^\lambda(R)dR\right)v \\
&= \left(\int_{SO(n)} U^\lambda(S)U^\lambda(R)dR\right)v \\
&= \left(\int_{SO(n)} U^\lambda(SR)dR\right)v \\
&= \left(\int_{SO(n)} U^\lambda(R)dR\right)v \qquad \text{as the integral is computed over } SO(n) \\
&= \phi_R(\lambda)v
\end{aligned}$$

And therefore $\phi_R(\lambda)v$ is a eigenvector associated to a non null eigenvalue. As $U^\lambda(S)$ is irreducible, it does not have any invariant subspace. Consequently, $\phi_R(\lambda) = 0$. □

Proposition 11 shows that the uniform distribution is an absorbing element of $L^2(SO(n))$. If R_1 is uniformly distributed, then R_1R_2 is uniformly distributed for any random variable R_2 independent from R_1.

This subsection was an introduction to the irreducible representation of random variables on $SO(n)$ and presented basic properties and examples that will be used in Chapter 3. It can also be used to study the convergence of the product of identically independent random variables. An example is given in the next subsection, demonstrating the result from Theorem 1

1.4.2 Application to the Brownian motion

Let B_t be a Brownian motion on $SO(n)$, defined in (1.22) with a unit variance for w_t. In (1.31), the distribution p_t of B_t can be described thanks to its Fourier coefficients $\hat{p}_t^\lambda$.

Proposition 12. *The coefficients $\hat{p}_t^\lambda$ in (1.31) from the decomposition of the distribution of B_t are given by $\hat{p}_t^\lambda = e^{-\chi_\iota t/2} I_{d_\lambda}$.*

Proof. By definition, $\hat{p}_t^\lambda = \mathbb{E}[U^\lambda(B_t)]$. As the functions U^λ are continuous, B_t can be approximated by the discrete chain $\tilde{B}_N$ defined in (1.23) with $\delta t = t/N$ to determine $\hat{p}_t^\lambda$

$$\mathbb{E}[U^\lambda(B_t)] = \lim_N \mathbb{E}[U^\lambda(\tilde{B}_N)].$$

By definition, $\tilde{B}_N$ can be decomposed as $\tilde{B}_N = \prod_{k\leq N} \exp(\delta d\hat{w}_{k\delta t})$ with $\prod$ a right product (see Equation (1.24)). As w_t is a Brownian motion in $\mathbb{R}^{n(n-1)/2}$, its increments $\delta w_{n\delta t}$ are independent. Thanks to (1.29),

$$\mathbb{E}[U^\lambda(\tilde{B}_N)] = \prod_{k\leq N} \mathbb{E}[U^\lambda(\exp(\delta d\hat{w}_{k\delta t}))].$$

The functions U^λ can then be applied individually to each increment $\exp(\delta d\hat{w}_{k\delta t})$.

Let σ^λ be the $d_\lambda \times d_\lambda$ matrix defined as $\sigma^\lambda = \frac{d}{dt} U^\lambda(\exp(t\sigma))\,|_{t=0}$. As $U^\lambda(R) \in SO(d_\lambda)$, the term $U^\lambda(\exp(\delta d\hat{w}_{k\delta t}))$ can be written

$$U^\lambda(\exp(\delta d\hat{w}_{k\delta t})) = \exp\left(\sum_i \delta dw^i_{k\delta t}\sigma^\lambda_i\right).$$

A second order development of exp gives

$$\prod_{k\leq N} \mathbb{E}[U^\lambda(\exp(\delta d\hat{w}_{k\delta t}))] = \prod_{k\leq N}\left(I_l + \sum_i \mathbb{E}[\delta dw^i_{k\delta t}\sigma^\lambda_i] + \frac{1}{2}\sum_{i,j}\mathbb{E}[\delta dw^i_{k\delta t}\delta dw^j_{k\delta t}]\sigma^\lambda_i\sigma^\lambda_j + o(\delta t)\right). \tag{1.33}$$

By definition of w_t, $\mathbb{E}[\delta dw^i_{k\delta t}\sigma^\lambda_i] = 0$. The unit variance means that $\mathbb{E}[\delta dw^i_{k\delta t}\delta dw^j_{k\delta t}] = \delta t\delta_{i,j}$. The previous equation then becomes

$$\mathbb{E}[U^\lambda(\tilde{B}_N)] = \prod_{k\leq N}\left(I_l + \frac{\delta t}{2}\sum_i \sigma^\lambda_i\sigma^\lambda_i + o(\delta t)\right)$$

We can show [Faraut 2008] that the constant $\sum_i \sigma^\lambda_i\sigma^\lambda_i$ can be written as

$$\sum_i \sigma^\lambda_i\sigma^\lambda_i = -\chi_\lambda I_{d_\lambda} \qquad \text{with } \chi_\lambda > 0.$$

Therefore, as $\mathbb{E}[U^\lambda(B_t)] = \lim_n \mathbb{E}[U^\lambda(\tilde{B}_n)]$ the Fourier coefficients $\hat{p}^\lambda$ can be written $\hat{p}^\lambda = e^{-\chi_\lambda t/2} I_{d_\lambda}$. □

Coupling the proposition 12 with the reconstruction formula (1.31) would give an expression for the distribution of B_t. This can be used for inference on σ_w for example. Amongst other things, it also shows that the coefficients $\hat{p}^\lambda_t$ will converge towards 0, except for $\hat{p}^{\lambda_0}_t = I_{\lambda_0}$ as $\chi_\lambda > 0$ for $\lambda \neq \lambda_0$. This corresponds to the uniform distribution over $SO(n)$ as seen in Proposition 11. Despite that the uniform distribution was known to be the stationary distribution, nothing indicated that it would be the limit of the distribution of the Brownian motion until Proposition 12.

Another interesting aspect from the proposition 12 is the adaptation of the proof to show the central limit theorem on $SO(n)$. The fact that the increments $dw_{k\delta t}$ are Gaussian is not used. The only used properties is their independence, the fact that they are centered and with a variance δt. Therefore, any random variables satisfying these properties can be used. Given a set of random variables ζ_k centered with unit variance, for a fixed value of t and a fixed δ, let the chain R^δ be defined similarly to (1.26) as

$$R^\delta_{n+1} = \exp(\sqrt{t\delta}\hat{\zeta}_k)R^\delta_n.$$

Theorem 3. *The n^{th} element of the chain $R^{1/n}$ will converge in distribution towards B_t, a Brownian motion on $SO(n)$.*

Proof. Using the same method that the proof of proposition 12 by replacing the increments δw_n with $\sqrt{t/n}\zeta_n$ leads to equality of the characteristic function. By proposition 10, the distribution are then equal. □

Theorem 3 gives another proof of the central limit theorem (Theorem 1) presented with Markov generators. Here, a proof based on the characteristic functions gives the same results with a different approach.

1.5 Conclusion

Chapter 1 serves as an introduction of concepts and results used further on. Despite that the presented results are known in literature, this self-contained introduction describes the tools used for the studies in Chapter 2 and Chapter 3. It gives a geometric description of the special orthonormal group $SO(n)$ and defines right invariant processes with value on $SO(n)$.

Section 1.1 is a short introduction to $SO(n)$, introducing a special kind of trajectory, namely the geodesics. These geodesics are used further on to perform simulations of stochastic processes. These simulations rely on the McKean scheme, which can be considered as an extension of the classical Euler scheme from $\mathbb{R}^d$ to $SO(n)$. Two examples of stochastic processes are considered in Section 1.3: the rotation Brownian motion and the simulated annealing. The later is a Monte-Carlo method for optimization problems on $SO(n)$.

Section 1.4 is an introduction to the irreducible representation of $SO(n)$. This section is quite independent from the previous ones as it does not use much previously presented results. A deeper presentation of these results can be found in [Faraut 2008]. It gives a definition to the Fourier coefficients of a $L^2(SO(n))$ function. The Peter Weyl theorem (Theorem 2) gives a reconstruction formula from these coefficients and the irreducible representations. Results from this section will be mainly used in Chapter 3.

Chapter 2

Filtering from complete and partial observation on $SO(n)$

Contents

2.1 Filtering on $\mathbb{R}^n$ 30
2.1.1 Description and solution as a KS formula 30
2.1.2 Zakai equation 32
2.1.3 Solution for linear models 33
2.2 Filtering from observations on $SO(n)$, solution and implementation 35
2.2.1 Construction of the antidevelopment 36
2.2.2 Implementation of the solution 38
2.3 Monte-Carlo method for filtering from observations on the Stiefel manifold 44
2.3.1 Geometry of the Stiefel manifold 44
2.3.2 Solution to the filtering on the Stiefel manifold 48
2.3.3 Implementation of the solution to filtering on the Stiefel manifold . . 51
2.3.4 Limitations of the solution 54
2.4 Conclusion 55

This chapter is dedicated to the filtering problem with observation on $SO(n)$ and the Stiefel manifold $V_{n,k}$. In the filtering problem, a state is estimated from noisy observation. Section 2.1 presents the formalism used in this chapter. More detailed results about filtering can be found in literature [Jazwinski 1970]. A general solution is given, based on the Bayes's rule to describe the distribution of the state, conditioned by the observation, when the observation can be modeled with an additive noise. An adaptive form is derived from the general solution for real-time estimation. This form can be used, in case of linear state model, to express the equations of the famous Kalman Bucy filter [Bucy 1961].

Section 2.2 extends the previous model to observation on $SO(n)$. Contrary to Section 2.1, the noise is no more additive. The model used for the observation is based on the stochastic differential equation from Section 1.2. The main idea presented in this section is the construction of the antidevelopment. The antidevelopment is a stochastic process equivalent, in term of information, to the observed process but modeled with an additive noise. Substituting the antidevelopment to the observation process allows us to use the methods presented in Section 2.1. Moreover, in practice the observation of a process is realized from sampling, leading to a discrete chain instead of a continuous process. Different methods of interpolation are presented and compared in term of complexity of computation and accuracy of approximation. Despite that these interpolations add an error in the estimation, this error becomes negligible as the sampling step becomes small. The question of the choice of the interpolation scheme was already present in [Pontier 1988] for example. However, a comparison in term of computation and accuracy of the estimation has not been presented before.

Section 2.3 describes the Stiefel manifold $V_{n,k}$ as a partial observation from $SO(n)$. The problem of filtering with observation on $V_{n,k}$ can be used to model the problem of Section 2.2 with faulty sensors for example. The lack of observation prevents us from directly using the same method as Section 2.2. However, decomposing the tangent space from $SO(n)$ into horizontal and vertical subspaces, defined from the geometry of $V_{n,k}$ leads to the construction of the horizontal lift, a process on $SO(n)$, and therefore to the construction of the antidevelopment, with an additive noise. Due to the lack of observation, a numerical method must be used to estimate the conditional distribution. A Monte-Carlo method based on a the particle filter is implemented. This technique relies on the use of particles as candidates for the trajectory of the state and computing their likelihood with the observation on $V_{n,k}$. It is shown that this method leads to a partial estimation of the state if this state is modeled as a diffusion. If this state remains constant, a complete estimation can be realized. Despite that particle filters has been heavily used recently, using this technique from observation on the Stiefel manifold has not been proposed before.

2.1 Filtering on $\mathbb{R}^n$

2.1.1 Description and solution as a KS formula

The filtering problem consists in estimating the state of a system from noisy observation. For example, the state to estimate can be the position or the velocity in classical mechanics [Belanger 1992][Canet 1994], the idleness of a telecommunication channel [Arslan 2007], etc. This state is measured via devices or sensors subject to noise caused by the method of measurement or interferences from other systems. Let x_t be the state to estimate at the time t. We will suppose that $x_t \in \mathbb{R}^d$ is a stochastic process following the SDE:

$$dx_t = Fx_t dt + db_t$$

where $b_t \in \mathbb{R}^d$ is a Brownian motion with a variance σ_b^2. In this case, x_t is a Markov process. Its generator will be denoted $\mathcal{A}$ and its priori distribution p_0.

We will suppose that the observations y_t belongs to $\mathbb{R}^n$. The state x_t is measured via the sensor function $H : \mathbb{R}^d \to \mathbb{R}^n$, assumed known. In other words, in the absence of noise:

$$dy_t = H(x_t)dt$$

However, due to the imperfect methods of measurement, those one can be considered as noisy. We will assume that the noise can be modeled as an additive white noise [Jazwinski 1970]

$$dy_t = H(x_t)dt + dw_t \tag{2.1}$$

with w_t a Brownian motion in $\mathbb{R}^n$ independent from x_t with a variance σ_w^2.

Because of the presence of noise in equation (2.1), it will not be possible to retrieve exactly x_t from the observation of y_t. Instead we will be looking for the conditional distribution $\pi_{[0,t]}$ of x given the observation $\mathcal{Y}_t = \{y_s,\ s \leq t\}$. From π_t, all the needed statistics can be extracted. For example, in engineering problems, one is looking for an estimation of x_t. From Proposition 13, the estimator $\mathbb{E}[x_t|\mathcal{Y}_t]$, easily obtained from $\pi_{[0,t]}$, is widely used as it is the optimal estimator [Jazwinski 1970], in the mean square sense.

Proposition 13. *The estimator $\mu = \mathbb{E}[x_t|\mathcal{Y}_t]$ is the optimal estimator of x_t in the mean square sense.*

Proof. Let $\hat{x}_t$ be an estimator of x_t. The mean square error criteria can be written

$$\begin{aligned}
\mathbb{E}\left[(x_t-\hat{x}_t)^2\right] &= \mathbb{E}\left[\mathbb{E}\left[(x_t-\hat{x}_t)^2|\mathcal{Y}_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[(x_t-\mu+\mu-\hat{x}_t)^2|\mathcal{Y}_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[(x_t-\mu)^2|\mathcal{Y}_t\right]\right] - 2\mathbb{E}\left[\mathbb{E}\left[(x_t-\mu)(\hat{x}_t-\mu)|\mathcal{Y}_t\right]\right] + \mathbb{E}\left[\mathbb{E}\left[(\hat{x}_t-\mu)^2|\mathcal{Y}_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[(x_t-\mu)^2|\mathcal{Y}_t\right]\right] - 2\mathbb{E}\left[\mathbb{E}\left[(x_t-\mu)(\hat{x}_t-\mu)|\mathcal{Y}_t\right]\right] + \mathbb{E}\left[\mathbb{E}\left[(\hat{x}_t-\mu)^2|\mathcal{Y}_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[(x_t-\mu)^2|\mathcal{Y}_t\right]\right] - 2\mathbb{E}\left[(\mathbb{E}\left[x_t|\mathcal{Y}_t\right]-\mu)(\hat{x}_t-\mu)\right] + \mathbb{E}\left[(\hat{x}_t-\mu)^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[(x_t-\mu)^2|\mathcal{Y}_t\right]\right] + \mathbb{E}\left[(\hat{x}_t-\mu)^2\right] \qquad \text{by definition of } \mu \\
&= \mathbb{E}[(x_t-\mu)^2] + \mathbb{E}\left[(\hat{x}_t-\mu)^2\right]
\end{aligned}$$

The first term is a constant with respect to $\hat{x}_t$. The last term being minimized by $\hat{x}_t = \mu$, μ is therefore the optimal estimator. □

Determining $\pi_{[0,t]}$ directly is rather complex as in (2.1), it is the distribution of y_t conditioned by the knowledge of x_t that can be easily expressed. The idea is to use the Bayes rule to express $\pi_{[0,t]}$ thanks to the likelihood of y_t given x_t.

First of all, let consider the discrete chain $\tilde{y}_n$ as

$$\delta\tilde{y}_n = H(x_i)\delta t + \delta w_n, \tag{2.2}$$

similarly to (1.14). Let $\pi_n^{\delta t}$ be the conditional distribution of $(x_0, x_{\delta t}, ..., x_{n\delta t})$ under the knowledge of $\tilde{\mathcal{Y}}_n = \{\tilde{y}_k, k \leq n\}$. Using Bayes rule on $\pi_n^{\delta t}$ gives

$$\pi_n^{\delta t}(\zeta_0, ..., \zeta_n) = \frac{p_X(\zeta_0, ..., \zeta_n) L_{\delta t}(\tilde{\mathcal{Y}}_n|\zeta_0, ..., \zeta_n)}{p_Y(\tilde{\mathcal{Y}}_n)} \tag{2.3}$$

where p_X (respectively p_Y) is the distribution of $(x_0, ..., x_{n\delta t})$ (respectively $\tilde{\mathcal{Y}}_n$) and $L_{\delta t}$ the likelihood of $\tilde{\mathcal{Y}}_n$ conditioned by X_n. As the increments δw_n are independent and identically distributed, the likelihood $L_{\delta t}$ can be decomposed as

$$L_{\delta t}(\tilde{\mathcal{Y}}_n|\zeta_0, ..., \zeta_n) = \prod_{k \leq t/\delta t} l^{\delta t}\left(\delta\tilde{y}_k|\zeta_k\right) \tag{2.4}$$

where $l^{\delta t}(\delta\tilde{y}_k|\zeta_k)$ is a Gaussian centered in $H(\zeta_k)\delta t$, with a variance $\sigma_w^2\delta t$.

As $\pi_n^{\delta t}$ in (2.3) is a function of X_n, the denominator $p_Y(\tilde{\mathcal{Y}}_n)$ is a constant. Indeed, this term is a constant and is determined by the observation. Writing p_Y as the marginal of the joint distribution of x and $\tilde{y}$ reads

$$\begin{aligned}
p_Y(\tilde{\mathcal{Y}}_n) &= \int p_X(x'_0, ..., x'_n) L_{\delta t}(\tilde{\mathcal{Y}}_n|x'_0, ..., x'_{n\delta t}) dx'_0 ... dx'_{n\delta t} \\
&= \mathbb{E}[L_{\delta t}(\tilde{\mathcal{Y}}_n|x'_0, ..., x'_{n\delta t})|\tilde{\mathcal{Y}}_n]
\end{aligned}$$

where x' is a copy of x (same distribution) but independent from the observation y. We can therefore express (2.3) only with expectations. Considering a test function $\phi \in \mathcal{C}^2\left(\mathbb{R}^{d\times n} \to \mathbb{R}\right)$, then

$$\pi_n^{\delta t}(\phi) = \frac{\rho_n^{\delta t}(\phi)}{\rho_n^{\delta t}(1)}$$

where $\pi_n^{\delta t}(\phi) = \mathbb{E}\left[\phi(x'_0, ..., x'_{n\delta t})\right]$ and $\rho_n^{\delta t}(\phi) = \mathbb{E}\left[\phi(x'_0, ..., x'_{n\delta t}) L_{\delta t}(\tilde{\mathcal{Y}}_n|x'_0, ..., x'_{n\delta t})|\tilde{\mathcal{Y}}_n\right]$.

In Subsection 1.2.2, it has been shown that the chain $\tilde{y}_n$ approximates y_t for $t = n\delta t$ when δt shrinks to 0. It can be shown (the proof is omitted here but can be found in

[Jazwinski 1970]) that when δt shrinks to 0, the previous equation converges towards an expression of $\pi_{[0,t]}$. For any test function $\phi \in \mathcal{C}^2$,

$$\pi_{[0,t]}(\phi) = \frac{\rho_{[0,t]}(\phi)}{\rho_{[0,t]}(1)}, \tag{2.5}$$

with $\pi_{[0,t]}(\phi) = \mathbb{E}[\phi(x'_{[0,t]})|\mathcal{Y}_t]$, $\rho_{[0,t]}(\phi) = \mathbb{E}\left[\phi(x'_{[0,t]})L_t(y|x')|\mathcal{Y}_t\right]$ and the likelihood L_t is defined as

$$L_t(y|x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < H(x'_s), dy_s > -\frac{1}{2\sigma_w^2}\int_0^t ||H(x'_s)||^2 ds\right) \tag{2.6}$$

Equation (2.5) is called the Kallianpur-Striebel formula. Despite that it gives the solution of the filtering problem, it is not an adaptive solution, *i.e* when a new observation is realised, the determination of the distribution need to be completely computed again, instead of modifying the last estimation. This is precisely the object of the next subsection.

2.1.2 Zakai equation

This paragraph describes the evolution of the conditional distribution of x_t. However, from (2.18), it appears that determining the unnormalized conditional distribution $\rho_{[0,t]}$ is enough. It will be supposed that if one is looking for an adaptive form of $\rho_{[0,t]}$, the computation is done in real-time and the computations only concern the last observation. Therefore, we will only be looking for a recursive expression for ρ_t, the unnormalized conditional distribution of x_t, instead of the whole process.

Similarly to the previous paragraph, the idea to determine a recursive expression for ρ_t consists in working with a time discrete chain and then considering the limit for a small time step.

For the chain $\tilde{y}_n$ defined in (2.2), the Bayes rule gives

$$\rho_n^{\delta t}(\zeta_0, ..., \zeta_n) = p_X(\zeta_0, ..., \zeta_n) L_{\delta t}(\tilde{\mathcal{Y}}_n|\zeta_0, ..., \zeta_n).$$

Replacing the term $L_{\delta t}(\tilde{\mathcal{Y}}_n|\zeta_0, ..., \zeta_n)$ by the decomposition from (2.4) gives

$$\rho_n^{\delta t}(\zeta_0, ..., \zeta_n) = p_X(\zeta_0, ..., \zeta_n) \prod_k l^{\delta t}(\delta\tilde{y}_k, \zeta_k) \tag{2.7}$$

Moreover, it is assumed that x_t is a Markov process with a priori p_0 and a generator $\mathcal{A}$. Consequently, the chain $x_{n\delta t}$ with $n \in \mathbb{N}$ is a Markov chain with a priori distribution and a transition kernel $q_{\delta t}$ defined as

$$\int q_{\delta t}(x', x)p(x)dx = \left[\exp(\mathcal{A}^\star t)p\right](x'). \tag{2.8}$$

Then, the joint distribution p_X can be decomposed into

$$p_X(\zeta_0, ..., \zeta_n) = q_{\delta t}(\zeta_n, \zeta_{n-1}) \ ... \ q_{\delta t}(\zeta_1, \zeta_0)p_0(\zeta_0)$$

Replacing this expression in (2.7), it becomes

$$\begin{aligned}
\rho_n(\zeta_0, ..., \zeta_n) &= q_{\delta t}(\zeta_n, \zeta_{n-1}) \ ... \ q_{\delta}(\zeta_1, \zeta_0)p_0(\zeta_0) \prod_k l^{\delta t}(\delta\tilde{y}_k, \zeta_k) \\
&= q_{\delta t}(\zeta_n, \zeta_{n-1}) \ ... \ q_{\delta t}(\zeta_1, \zeta_0)p_0(\zeta_0) l^{\delta}(\delta\tilde{y}_n, \zeta_n)...l^{\delta t}(\delta\tilde{y}_0, \zeta_0) \\
&= q_{\delta t}(\zeta_n, \zeta_{n-1}) l^{\delta t}(\delta\tilde{y}_n, \zeta_n)... \ q_{\delta t}(\zeta_1, \zeta_0) l^{\delta t}(\delta\tilde{y}_0, \zeta_0)p_0(\zeta_0) \\
&= q_{\delta t}(\zeta_n, \zeta_{n-1}) l^{\delta t}(\delta\tilde{y}_n, \zeta_n)\rho_{n-1}(\zeta_0, ..., \zeta_{n-1}).
\end{aligned}$$

By marginalizing this last expression with respect to the n first elements $\zeta_0, ..., \zeta_{n-1}$, we obtain an integral form for ρ'_n, the unnormalized conditional distribution of ζ_n as

$$\begin{aligned}\rho'_n(\zeta_n) &= \int \rho_n(\zeta_0, ..., \zeta_n) d\zeta_0 ... d\zeta_{n-1} \\ &= \left(\int q_\delta(\zeta_n, \zeta_{n-1}) \rho_{n-1}(\zeta_0, ..., \zeta_{n-1}) d\zeta_0 ... d\zeta_{n-1} \right) l^\delta(\delta y_n, \xi_n) \\ &= \left(\int q_\delta(\zeta_n, \zeta_{n-1}) \rho'_{n-1}(\zeta_{n-1}) ... d\zeta_{n-1} \right) l^\delta(\delta y_n, \xi_n)\end{aligned}$$

The integral term from the previous expression can be written with the Markov adjoint generator $\mathcal{A}^\star$ by replacing $q_{\delta t}$ by its definition in Equation (2.8)

$$\rho'_n(\zeta) = \exp\left(\mathcal{A}^\star \delta t\right) \rho'_{n-1}(\zeta) l^{\delta t}(\delta y_n, \zeta) \tag{2.9}$$

A first order development with respect to δt of the exp operator and of the Gaussian $l^{\delta t}$ gives

$$\rho'_n(\zeta) = \rho_{n-1}(\zeta) + \mathcal{A}^\star \rho'_{n-1}(\zeta) \delta t + \frac{1}{\sigma_w^2} \rho'_{n-1}(\zeta) < H(\zeta), \delta \tilde{y}_n > + o(\delta t)$$

Then, similarly to the previous paragraph, considering the limit when δt shrinks to 0 (and conserving $n\delta t$ constant) gives the following linear stochastic differential equation

$$d\rho_t(\zeta) = \mathcal{L}^\star \rho_t(\zeta) dt + \frac{1}{\sigma_w^2} \rho_t(\zeta) < H(\zeta), dy_t > \tag{2.10}$$

This equation is called the Zakai equation [Zakai 1969]. As new samples are observed, it becomes possible to describe the evolution of ρ_t. Despite that this equation does only give the evolution of the unnormalized conditional distribution, it can be easily normalized by dividing ρ_t by its integral. In practice, using the Zakai equation (2.10) is expensive, in term of computation, due to the integration. An easiest solution might consists in determining the evolution of some of the conditional moments to construct an estimator of x_t, similar to Proposition 13.

2.1.3 Solution for linear models

In this paragraph, we will be using the Zakai equation (2.10) to determine the successive moments of π_t, the conditional distribution of x_t. In order to do so, Equation (2.10) must be modified to have a stochastic differential equation based on π_t instead of ρ_t.

By definition of ρ_t, π_t can be expressed as the quotient

$$\pi_t = \frac{\rho_t}{\int_x \rho_t(x) dx}.$$

Then, using Itō's calculus, its evolution is given by

$$d\pi_t = d\frac{\rho_t}{\int_x \rho_t(x) dx} = \frac{d\rho_t}{\int_x \rho_t(x) dx} + \frac{\rho_t d\left(\int_x \rho_t(x) dx\right)}{\left(\int_x \rho_t(x) dx\right)^2} + \frac{\rho_t \left(d\int_x \rho_t(x) dx\right)^2}{\left(\int_x \rho_t(x) dx\right)^3} - \frac{d\rho_t d\left(\int_x \rho_t(x) dx\right)}{\left(\int_x \rho_t(x) dx\right)^2}.$$

The expression of $d\rho_t$ is given by (2.10). To determine $d\int_x \rho_t(x) dx$, Equation (2.10) is

used again, inside the integral term as

$$\begin{aligned}
d\int_x \rho_t(x)dx &= \int_x d\rho_t(x)dx \\
&= \int_x \mathcal{L}^\star \rho_t(x)dt + \frac{1}{2\sigma_w^2}\rho_t(x) < H(x), dy_t > dx \\
&= \left(\int_x \rho_t(x)dx\right)\int_x \mathcal{L}^\star p_t(x)dt + \frac{1}{2\sigma_w^2}p_t(x) < H(x), dy_t > dx \\
&= \left(\int_x \rho_t(x)dx\right)\mathcal{L}^\star \int_x p_t(x)dxdt + \left(\int_x \rho_t(x)dx\right)\frac{1}{2\sigma_w^2} < \int_x H(x)p_t(x)dx, dy_t > \\
&= 0 + \left(\int_x \rho_t(x)dx\right)\bar{H}dy
\end{aligned}$$

Where $\bar{H} = \int_x H(x)\pi_t(x)dx$.

Substituting these elements gives the following equation, called the Kushner equation [Jazwinski 1970]

$$d\pi_t(\zeta) = \mathcal{A}^\star \pi_t(\zeta)dt + \frac{1}{2\sigma_w^2}\pi_t < H(\zeta) - \bar{H}, dy - \bar{H}dt > \tag{2.11}$$

Now, we want to use (2.11) to extract the first order moment $\mu_t = \int x\pi_t(x)dx$ from π_t. At the beginning of the paragraph, x_t has been defined as solution of

$$dx_t = Fx_t dt + db_t$$

where b_t is a Brownian motion in $\mathbb{R}^d$ with a variance σ_b^2.

Proposition 14. *In the case of linear function H, μ_t is solution of*

$$d\mu_t = F\mu_t dt + P_t H^T(dy - H\mu_t dt)$$

where $P_t = \int (x-\mu)(x-\mu)^T \pi_t(x)dx$ is the second order centered moment.

Proof. Applying (2.11) to $d\mu_t$ gives

$$\begin{aligned}
d\mu_t &= d\int_x x\pi_t(x)dx \\
&= \int_x x d\pi_t(x)dx \\
&= \int_x x\mathcal{L}^\star \pi_t(x)dt + \int_x x\pi_t < H(x) - \bar{H}, dy - \bar{H}dt > dx \\
&= -\sum_i \int_x x\frac{d}{dx_i}\big((Fx)_i\pi_t(x)\big)dxdt + \int_x x\pi_t \left(H(x) - \bar{H}\right)^T \left(dy - \bar{H}dt\right) dx
\end{aligned}$$

Integrating by part the left term, the integral becomes

$$\begin{aligned}
d\mu_t &= \sum_i \int_x \frac{dx}{dx_i}(Fx)_i\pi_t(x)dxdt + \int_x x\pi_t \left(x^T - \mu_t^T\right) H^T \left(dy - H\mu_t dt\right) dx \\
&= \int_x Fx\pi_t(x)dxdt + \int_x (xx^T - \mu_t\mu_t^T)H^T\pi_t \left(dy - H\mu_t dt\right) dx \\
&= F\mu_t dt + P_t H^T(dy - H\mu_t dt)
\end{aligned}$$

□

Therefore, to determine the first order moment, the second order moment is required. It can easily be seen by generalizing the proof of Proposition 14 that when trying to determine the k-th order moment, the Kushner equation will impose the need of the $(k+1)$-th moment. Consequently, trying to solve (2.10) via a moment based method is not appropriate in many cases without considering the other moments as negligible.

However, it is also noticeable that in the special case when H is linear, it is known from (2.18) that π_t will be a Gaussian distribution. Thus, π_t is completely determined by its two first moments (its higher order moments can be described with μ_t and P_t). In this case, it can be show [Bucy 1961] that μ_t and P_t are solutions of

$$\begin{aligned} d\mu_t &= F\mu_t dt + P_t H^T (dy_t - H\mu_t dt) \\ dP_t &= FP_t + P_t F^T - P_t H H^T P_t + \sigma_b^2 \end{aligned} \tag{2.12}$$

where σ_b is the variance of b_t.

This set of equations is known as the Kalman-Bucy filter, the continuous counterpart of the Kalman filter. Just like the Kalman filter, the expression of the covariance P_t does not depend on the observation, only μ_t is affected by y_t.

As an extension, one can notice that these equations are obtained via integrals over the random variables at a fixed given time. In fact, nothing prevents the sensor function H or the drift coefficient F from being constant with respect to time. In the case when H (or F) varies, the same equations would hold with H_t (and F_t) fitting in the equations.

These equations can also be adapted in the case of nonlinear model for x_t and y_t with the stochastic equations

$$\begin{aligned} dx_t &= F(x_t)dt + db_t \\ dy_t &= H(x_t)dt + dw_t \end{aligned}$$

The idea to solve the filtering problem for these models is, roughly speaking, to make a first order approximation of F and H. We will need to assume that the variance σ_b^2 is small with respect to the gradient of the functions F and H. In this case, the variations around x_t are almost linear and F and H can be linearly approximated. In this case, the distributions are almost Gaussian, as the transformation in place are almost linear. By approximating these distributions by Gaussian ones, one would then obtain the equations of the Extended Kalman-Bucy filter

$$\begin{aligned} d\mu_t &= F'\mu_t dt + P_t(dy_t - \mu_t dt) \\ dP_t &= F'P_t + P_t {F'}^T - P_t H' {H'}^T P_t + \sigma_B^2 \end{aligned} \tag{2.13}$$

Where $F' = \nabla F(\mu_t)$ and $H' = \nabla H(\mu_t)$.

The main difference with the classical Kalman-Bucy filter (2.12) is that, in this case, the variance P_t depends on the estimation, and therefore on the observation, leading to coupled stochastic differential equations. For the other cases, numerical solutions can be envisaged, like particle based methods for example.

In this section, the filtering problem has been stated and solved via a Bayesian method in (2.18). However, this solution is not adaptive. A recursive form taking into account new sample has been constructed, finally leading the classic Kalman-Bucy equations. In the next section, these results are extended to filter observation on $SO(n)$.

2.2 Filtering from observations on $SO(n)$, solution and implementation

We saw in the previous section what is the filtering problem for noisy observation on $\mathbb{R}^n$. However, this problem can also be extended to observation with values on $SO(n)$. The

problem consists, in this case, to estimate a state x_t from a noisy observation $R_t \in SO(n)$.

Such problems can arise in mechanics [Lovera 2002] or computer vision for example [Meer 2007], [Chellappa 2008]. As opposed to the previous case of Section 2.1, there are two main differences. First of all, these processes are, by definition, defined on $SO(n)$ and they are submitted to the constraints presented in Chapter 1. Second, the model used for the noise is different from Section 2.1. Despite that R_t can simply be treated as matrix processes with non linear model, using an extended Kalman filter [Jazwinski 1970] to estimate parameters from R_t would lead to equations with non constant coefficients. Contrary to the extended Kalman filter, the method presented in this section gives an optimal solution with constant coefficients.

2.2.1 Construction of the antidevelopment

In this section, the filtering problem is transposed from observation on $\mathbb{R}^n$ to observation on $SO(n)$. As this process is completely defined via its angular velocity, the state to estimate $x_t \in \mathbb{R}^{n(n-1)/2}$ is the angular velocity of $R_t \in SO(n)$. Despite that the major part of the presented method can be generalized to any model for x_t, we restrict ourself to the case when the state x_t can be described by a linear model

$$dx_t = Fx_t dt + db_t \tag{2.14}$$

where $b_t \in \mathbb{R}^{n(n-1)/2}$ is a Brownian motion with a variance σ_b^2.

The noise is modeled similarly to Equation (1.11) from Subsection 1.2.1

$$dR_t = (\hat{x}_t dt + \circ d\hat{w}_t)\, R_t \tag{2.15}$$

where $w_t \in \mathbb{R}^{n(n-1)/2}$ is a Brownian motion with a variance σ_w^2, independent from x_t and the application $x \mapsto \hat{x}$ is defined in (1.4).

The problem of estimating x_t from observation $\mathcal{R}_t = \{R_s, s \leq t\}$ is intrinsically more difficult than the previous case. First, R_t belongs to $SO(n)$ rather than the space $\mathbb{R}^n$. Second, R_t is not modeled with an additive noise as in (2.1). In this case, the noise is multiplicative and the previous solution (2.5) cannot be applied anymore. Indeed, the previous solution relies on the conditional independence of the increments dR_t to express the likelihood L_t. For Equation (2.15), this is no more the case as the increments dR_t depend on R_t.

The solution presented in this section is based on the method described in [Lo 1975]. It consists in the creation of another stochastic process z_t that is constructed from R_t. The process z_t will be uniquely defined for a given R_t and therefore will be identical in term of information to R_t, *i.e* conditioning the distribution of x_t with z_t or R_t will be the same. However, the main point is to construct z_t such that it can be described with a additive noise model, contrary to R_t. Then, classical methods presented in Section 2.1 can be used to realize the estimation of x_t.

Let $z_t \in \mathbb{R}^{n \times n}$ be the process defined as

$$z_t = \int_0^t (\circ dR_s) R_s^T \tag{2.16}$$

Converting (2.16) into an Itō integral stochastic process gives

$$\begin{aligned} dz_t &= (\circ dR_t)\, R_t^T \\ &= (dR_t)\, R_t^T + \frac{1}{2}\mathbb{E}[(dR_t)R_t^T R_t(dR_t^T)] \\ &= (dR_t)\, R_t^T - \frac{1}{2}\mathbb{E}[d\hat{w}_t^2] \\ &= (dR_t)\, R_t^T - \frac{\sigma_w^2}{2}dt \\ &= \hat{x}_t dt + d\hat{w}_t - \frac{\sigma_w^2}{2}dt. \end{aligned} \tag{2.17}$$

Now, contrary to (2.15), the noise in (2.16) is additive.

The process z_t is called the antidevelopment of R_t [Li 2004]. From an intuitive point of view, it can be interpreted as the cumulation of the increments of R_t seen from the tangent space $T_{R_t}SO(n)$. It is noticeable that in the case when R_t is a geodesic in $SO(n)$ (a uniform rotation around a fixed axis), then the antidevelopment is a straight line in $\mathbb{R}^{n\times n}$.

It has been seen earlier that z_t can be constructed from R_t. The contrary is also possible, constructing R_t from the process z_t. Writing

$$\begin{aligned} dR_t &= (\circ dz_t)R_t \\ &= (\hat{x}_t dt + \circ d\hat{w}_t)\, R_t, \end{aligned}$$

gives a way to express R_t from z_t.

This shows the one-to-one character of the antidevelopment z_t with respect to R_t. Therefore, observing R_t or z_t is the same, in term of information. For the filtering problem, the distribution $\pi_{[0,t]}$of the process x_t conditioned by the observations of $\mathcal{R}_t$ is the same as the distribution of x_t conditioned by $\mathcal{Z}_t = \{z_s, s \leq t\}$.

The problem of filtering with observation on $SO(n)$ is then replaced by a problem of filtering on $\mathbb{R}^{n(n-1)/2}$ with a classical model of additive noise presented in Section 2.1 by replacing the observation R_t by the antidevelopment z_t. The problem can then be treated independently from the fact that the observation R_t are initially in $SO(n)$ and the noise is multiplicative.

The solution to the filtering problem is then given by the Kallianpur-Striebel formula (2.18) applied with $\mathcal{Z}_t$ for a test function $\phi \in \mathcal{C}^2$

$$\pi_{[0,t]}(\phi) = \frac{\rho_{[0,t]}(\phi)}{\rho_{[0,t]}(1)}, \tag{2.18}$$

with $\rho_{[0,t]}(\phi) = \mathbb{E}\left[\phi(x')L_t(z,x')|\mathcal{Z}_t\right]$, where x'_t is a copy of x_t independent from R_t. The likelihood L_t is defined as

$$L_t(z,x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < \hat{x}'_s, \circ dz_s > -\frac{1}{2\sigma_w^2}\int_0^t ||\hat{x}'_s||^2 ds\right) \qquad \text{with } \circ\, dz_t = (\circ dR_t)R_t^T.$$

As symmetric and skew-symmetric matrices are orthogonal, the term $< \hat{x}'_s, \circ dz_s >=< \hat{x}'_s, dR_s R_s^T - \sigma_w^2 I_n dt >$ can be replaced by $< x'_s, dR_s R_s^T >$. In this case, the likelihood L_t can be written as

$$L_t(z,x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < x'_s, dR_s R_s^T > -\frac{1}{2\sigma_w^2}\int_0^t ||x'_s||^2 ds\right). \tag{2.19}$$

Equation (2.10) can also be adapted to describe the evolution of the conditional unnormalized distribution ρ_t

$$d\rho_t(\zeta) = \mathcal{A}^\star \rho_t(\zeta)dt + \frac{1}{\sigma_w^2}\rho_t(\zeta) < \hat{\zeta}, dz_t > . \tag{2.20}$$

As the model for x_t is linear (Equation (2.14)), π_t is a Gaussian distribution centered in μ_t and with a variance P_t, solutions of

$$\begin{aligned} d\hat{\mu}_t &= \widehat{F\mu_t}dt + P_t(dz_t - \hat{\mu}_t dt) \\ dP_t &= FP_t + P_t F^T - P_t^2 + \sigma_b^2 \end{aligned} \tag{2.21}$$

These equations deeply differ from the extended Kalman filter as they don't approximate the observation with a linear model. For an extended Kalman filter, this approximation creates a correlation between the coefficients of the equations and the estimation. With the method presented here, this is not the case and the value of the variance P_t can be computed independently from the observation.

Also, in practice, other issues appear, mainly due to the discrete sampling of R_t with a time step δt. Indeed, the integral term in the likelihood L_t (2.19) is then approximated by a Riemannian sum. The differential term dy_s is then replaced by a discrete interpolation $\delta z_{n\delta t}$, itself relying on an interpolation of dR_t. The next part focuses on the conditions required by this interpolation function to make the Riemannian sum converges to the stochastic integral from (2.19) when δt becomes small.

2.2.2 Implementation of the solution

Although the theoretical aspect of the filtering problem with observation on $SO(n)$ has been treated in the previous section, the practical resolution is not direct. Indeed, due to the discretization of the observation, the term dz_t is replaced by a discrete increment $\delta z_{n\delta t}$ with δt the time step of the discretization.

In Subsection 1.2.2, it has been shown that continuous processes with additive noise could be discretized via the Euler scheme. From the expression of the antidevelopment (2.16), applying an Euler scheme with a time step δt gives

$$\begin{aligned} \delta z_{k\delta t} &= (R_{(k+1)\delta t} - R_{k\delta t})R_{k\delta t}^T - \sigma_w^2 I_n \delta t \\ &= R_{(k+1)\delta t}R_{k\delta t}^T - I_n - \sigma_w^2 I_n \delta t \\ &= \Delta R_{k\delta t} - I_n - \sigma_w^2 I_n \delta t \qquad \text{with } \Delta R_{k\delta t} = R_{(k+1)\delta t}R_{k\delta t}^T \end{aligned} \tag{2.22}$$

Because of the continuity of R_t, the term $R_{(k+1)\delta t}R_{k\delta t}^T$ converge towards I_n when δt shrinks to 0. Consequently, $\Delta R_{k\delta t}$ is in the definition domain of the log function, defined in Subsection 1.1.1. It can be shown that (2.22) is the first order development of $\log(\Delta R_{k\delta t})$

$$\delta z_{k\delta t} = \log\left(\Delta R_{k\delta t}\right) + o(\delta t).$$

This expression, in practice, will be simplified into

$$\delta z_{k\delta t} \approx \log(\Delta R_{k\delta t}). \tag{2.23}$$

This interpolation between $R_{k\delta t}$ and $R_{(k+1)\delta t}$ is called geodesic interpolation [Pontier 1988].

In this case, from Proposition 4, the Riemannian sum

$$S^\delta = \sum_{k \leq t/\delta t} < \hat{x}_{k\delta t}, \delta z_{k\delta t} >$$

converges as δt becomes small, in the mean square sense, towards

$$\mathcal{I} = \int_0^t < \hat{x}_s, dz_s >$$

and the likelihood L_t from (2.6) can be approximated by $\tilde{L}_{\delta t}(\tilde{\mathcal{Z}}_{n\delta t})$ from (2.4).

In equation (2.23), the error term was simplified because it was negligible with respect to $\log(\delta R_{k\delta t})$ and it seems natural that any function close enough to $\log(\delta R_{k\delta t})$ could work to approximate $\delta z_{k\delta t}$.

Let $\mathrm{Int} : SO(n) \times SO(n) \to \mathfrak{so}(n)$ be an interpolation function. We are interested in functions Int such that by considering $\delta z_k = \mathrm{Int}(R_{k\delta t}, R_{(k+1)\delta t})$, the solution with a discrete observation will converge to the continuous solution with a small time step. It is noticeable that $\mathrm{Int}(R_{k\delta t}, R_{(k+1)\delta t})$ is an antisymmetric matrix and can therefore be written as an image of the application $x \mapsto \hat{x}$ ($\delta z_{k\delta t}$ in this case).

Theorem 4. *The Riemannian sum*

$$S^\delta = \sum_{k \leq t/\delta t} < \hat{x}_{k\delta t}, \mathrm{Int}(R_{k\delta t}, R_{(k+1)\delta t}) >$$

converges towards

$$\mathcal{I} = \int_0^t < \hat{x}_s, dz_s >$$

in the sense $\mathbb{E}[|S^\delta - \mathcal{I}|^2] \to 0$ *when* $\delta \to 0$ *if the interpolation function* $\mathrm{Int} : SO(n) \times SO(n) \to \mathfrak{so}(n)$ *satisfies the following conditions:*

1. *The function is null on its diagonal* $\mathrm{Int}(R, R) = 0$ *for all* $R \in SO(n)$.
2. *The function* Int *is* $\mathcal{C}^2(SO(n))$.
3. $\nabla \mathrm{Int}(R, R)[v] = vR^T$ *for all* $v \in T_R SO(n)$.
4. $\nabla^2 \mathrm{Int}(R, R)[v] = 0$ *for all* $v \in T_R SO(n)$.

Where the differentials (∇I and $\nabla^2 I$) are computed with respect to the second variable.

Proof. Considering the function $f = \mathrm{Int}(R_{k\delta t}, .)$, condition 2. allows the application of the Itō lemma

$$f(R_{(k+1)\delta t}) - f(R_{k\delta t}) = \int_{k\delta t}^{(k+1)\delta t} \nabla f[dR_s] + \frac{1}{2} \int_{k\delta t}^{(k+1)\delta t} \mathrm{trace}\left([dR_s]^T \nabla^2 f[dR_s]\right)$$

Condition 4. sets the last term to $o(\delta t)$ whereas 3. sets the first term to $\int_{k\delta t}^{(k+1)\delta t} (dR_s) R_s^T + o(\delta t)$. Then, replacing f by $\mathrm{Int}(R_{k\delta t}, .)$

$$\mathrm{Int}(R_{k\delta t}, R_{(k+1)\delta t}) - \mathrm{Int}(R_{k\delta t}, R_{k\delta t}) = \int_{k\delta t}^{(k+1)\delta t} (dR_s) R_s^T$$

Thanks to condition 1, $\mathrm{Int}(R_{k\delta t}, R_{k\delta t}) = 0$.

Separating the integral $\mathcal{I}$ into $t/\delta t$ short integrals gives, up to a remaining integral between t and $kt/\delta t$, the following expression

$$\begin{aligned}
\mathbb{E}\left[|S^\delta - \mathcal{I}|^2\right] &= \mathbb{E}\left[\left|\sum_{k\leq t/\delta t}\int_{k\delta t}^{(k+1)\delta t} < \hat{x}_{k\delta t} - \hat{x}_s, dz_s > \right|^2\right] \\
&= \mathbb{E}\left[\left|\sum_{k\leq t/\delta t}\int_{k\delta t}^{(k+1)\delta t} < \hat{x}_{k\delta t} - \hat{x}_s, (dR_s)R_s^T > \right|^2\right] \qquad \text{as } \sigma_w^2 I_n \text{ is orthogonal to } \mathfrak{so}(n) \\
&\leq \sum_{k\leq t/\delta t}\mathbb{E}\left[\left|\int_{k\delta t}^{(k+1)\delta t} < \hat{x}_{k\delta t} - \hat{x}_s, (dR_s)R_s^T > \right|^2\right]
\end{aligned}$$

The square variations of x_t during a time δt being bounded by $O(\delta t)$, the variation of $\hat{x}_t$ are bounded by $O(\delta t)$ too. The integral is then bounded by $O(\delta t^2)$ thanks to the Itō isometry property of the integral

$$\begin{aligned}
\mathbb{E}\left[|S^\delta - \mathcal{I}|^2\right] &= \mathbb{E}\left[\sum_{k\leq t/\delta t}\int_{k\delta t}^{(k+1)\delta t} O(\delta t)||(dR_s)R_s^T||^2\right] \\
&= \sum_{k\leq t/\delta t} O(\delta t^2) \\
&= O(\delta t)
\end{aligned}$$

This means that S^δ converges towards $\mathcal{I}$ in the mean square error sense when δt shrinks to 0 with a linear convergence rate. □

With Theorem 4, we can verify that the geodesic interpolation from (2.23) ensures a convergence of the solution.

Proposition 15. *The interpolation function* $\mathrm{Int}(R', R) = \log(RR'^T)$ *from (2.23) satisfies the conditions of Theorem 4.*

Proof. The conditions 1 and 2 are direct. For condition 3, an element $v \in T_R SO(n)$ will be described with right invariant vector fields $V = \sigma R$ with $\sigma \in \mathfrak{so}(n)$.

$$\begin{aligned}
\nabla \mathrm{Int}(R, R)[v] &= \frac{d}{dt}\mathrm{Int}(R, \exp(t\sigma)R)|_{t=0} \\
&= \frac{d}{dt}\log\left(\exp(t\sigma)R^T\right)|_{t=0} \\
&= \frac{d}{dt}t\sigma|_{t=0} \\
&= \sigma = vR^T
\end{aligned}$$

The condition 4 is also satisfied as

$$\begin{aligned}\nabla^2 \operatorname{Int}(R,R)[v] &= \frac{d^2}{dtds}\operatorname{Int}(R,\exp(s\sigma)\exp(t\sigma)R)|_{t=0,s=0}\\ &= \frac{d^2}{dtds}\log\left(\exp((t+s)\sigma)R^T\right)|_{t=0}\\ &= \frac{d^2}{dtds}(t+s)\sigma|_{t=0,s=0}\\ &= 0\end{aligned}$$

□

Therefore, the geodesic interpolation used in (2.23) ensures the convergence of the discrete solution to the continuous filtering solution. By considering the geodesic interpolation as a reference, proposition 4 states that any interpolation function whose second order development is the same as the geodesic interpolation ensures a convergence of the discrete solution towards the continuous solution.

In practice, the use of the log function is computationally expensive and the geodesic interpolation is heavy to use. However, other functions, with a smaller computational cost and close enough to the geodesic interpolation can be used.

Proposition 16. *The interpolation function* $\operatorname{Int}(R',R) = \frac{1}{2}(RR'^T - R'R^T)$ *satisfies the conditions of Theorem 4.*

Proof. The conditions 1 and 2 are direct. For conditions 3 and 4, the same notation as proposition 15 is used.

Condition 3 is verified as

$$\begin{aligned}\nabla \operatorname{Int}(R,R)[v] &= \frac{d}{dt}\operatorname{Int}(R,\exp(t\sigma)R)|_{t=0}\\ &= \frac{d}{dt}\frac{1}{2}\left(\exp(t\sigma)RR^T - R^T\exp(t\sigma)^T\right)|_{t=0}\\ &= \frac{d}{dt}\frac{1}{2}\left(\exp(t\sigma)+\exp(t\sigma)\right)|_{t=0} \text{ as } \sigma\in\mathfrak{so}(n)\\ &= \sigma = vR^T\end{aligned}$$

The condition 4 is also satisfied as

$$\begin{aligned}\nabla^2 \operatorname{Int}(R,R)[v] &= \frac{d^2}{dtds}\operatorname{Int}(R,\exp(s\sigma)\exp(t\sigma)R)|_{t=0,s=0}\\ &= \frac{d}{dt}\frac{1}{2}\left(\exp(s\sigma)\exp(t\sigma)RR^T - R^T\exp(t\sigma)^T\exp(s\sigma)^T\right)|_{t=0}\\ &= \frac{d}{dt}\frac{1}{2}\left(\exp((t+s)\sigma) - \exp((t+s)\sigma)\right)|_{t=0}\\ &= 0\end{aligned}$$

□

In practice, the function $\operatorname{Int}(R',R) = \frac{1}{2}(RR'^T - R'R^T)$ from Proposition 16 will be privileged due to its simplicity of computation as it respects the desired properties. This interpolation function is called linear interpolation. The differences for small time steps for both linear (from equation (2.22)) and geodesic (from equation (2.23)) interpolation functions are negligible.

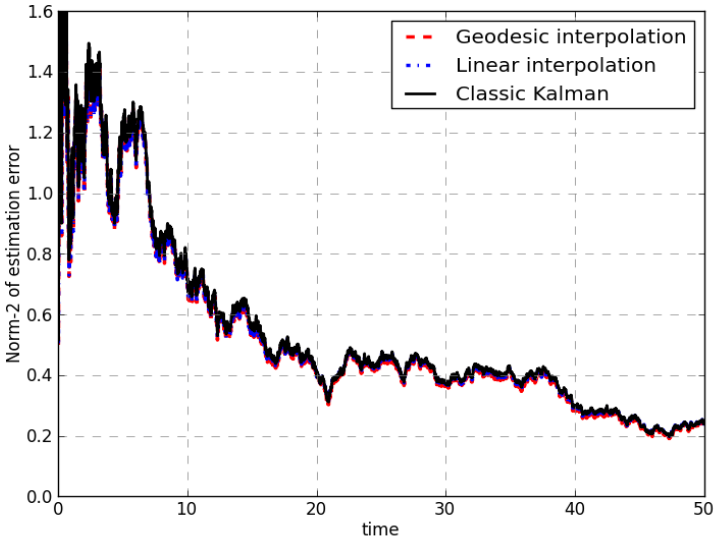


Figure 2.1: Comparison of estimation methods between linear, geodesic interpolation, and classic Kalman filter with $\delta t_{\text{obs}} = 0.01s$. The state to estimate is constant and the variance of the observation is $\sigma_w = 1$

To illustrate this, a simulation which results are presented in Figure 2.1 and Figure 2.2 is realized to show the difference between a classic Kalman filter, for which equations are described at (2.12) and the different methods of interpolation.

Because the equation of the variance P_t does not depend on the observation and is just an isolated differential equation (with respect to the innovation), the variance for each algorithm used in Figure 2.1 is the same (considering that P_0 has always been chosen with the same value) and is consequently not displayed for this kind of comparison.

In the simulation, the process x_t is considered constant, the use of another model might not imply any significant change. The process z_t is first generated with a small time step ($\delta t_{\text{gen}} = 10^{-4}s$) and with a variance $\sigma_w = 0.1$ via an Euler scheme from the stochastic differential equation (2.16).

At the same time, a rotational process R_t is constructed from z_t. The construction of z_t and R_t is then realized with a time step small enough to consider them as time continuous defined.

Then, we will sample the process R_t with a time step δt_{obs} with $\delta t_{\text{obs}} >> \delta t_{\text{gen}}$ and compare the performance of the filter from (2.21) with different methods of interpolation to approximate $\delta z_{k\delta t_{\text{obs}}}$: the linear and the geodesic interpolation. In parallel, as y_t has been continuously generated, these performances are compared to a classical Kalman filter taking directly $\delta z_{k\delta t_{\text{obs}}} = z_{(k+1)\delta t_{\text{obs}}} - z_{k\delta t_{\text{obs}}}$. Knowing that in the case of additive noise, the Kalman filter is optimal [Jazwinski 1970] and because the antidevelopment is one-to-one with the observation, the Kalman filter is used here a reference filter, as we know that it is not possible, in the mean square sense, to outperform it.

The results of this simulation are presented in Figure 2.1 with $\delta t_{\text{obs}} = 10^{-2}s$. This time step is small enough to have similar performances between the different methods of interpolation (geodesic and linear) as they should converge towards the same solution. However, differences start to appear if δt_{obs} has larger values.

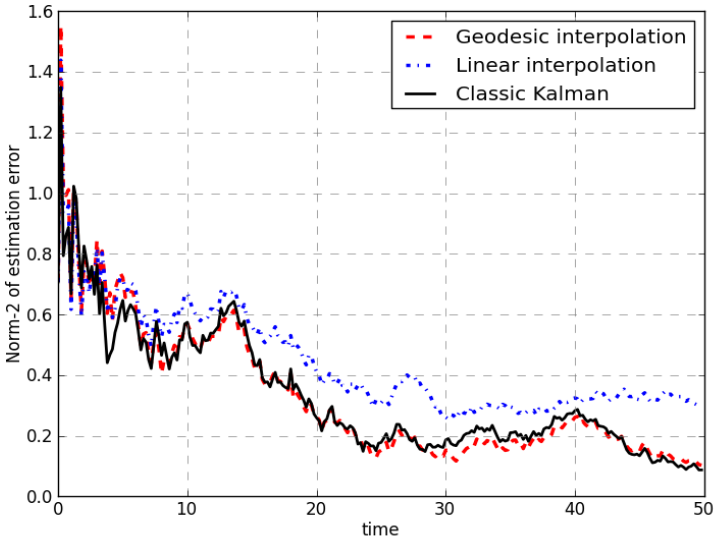


Figure 2.2: Comparison of estimation methods between linear, geodesic interpolation, and classic Kalman filter with $\delta t_{\text{obs}} = 0.2s$. The state to estimate is constant and the variance of the observation is $\sigma_w = 1$.

In Figure 2.2, δt_{obs} is increased to $0.2s$. The difference of observation time step changes the approximation of the process y_t and differences between the two interpolation methods are visible. The linear interpolation is not as accurate as the geodesic interpolation. The linear approximation adds another source of error to the estimation, despite its simpler computational form.

In order to observe the influence of the sampling step δt_{obs} on the performance of each interpolation module, Figure 2.3 illustrates the evolution of the cumulated error for $0 \leq t \leq 50$ at a fixed sample step for each method. Just like previously, the cumulated error should be compared with the cumulated error induced by a proper Kalman filter. It appears that the geodesic interpolation does not create another error term despite that it is an approximation. The cumulated error is the same than for the Kalman filter. The linear interpolation, however, is adding a supplementary error term. As the sampling step is increasing, the approximation is worse and worse. At the end for $\delta t_{\text{obs}} > 0.45$, the sampling step is too high for any filter to perform properly.

In Figure 2.4, an illustration of filtering from observation on $SO(3)$ is presented. The state to estimate is constant: $x = (1, 0, 1)^T$. The sampling step is $\delta t_{\text{obs}} = 0.1$ and the noise variance is $\sigma_w^2 = 1$. The observation is realized only for $0 \leq t \leq 6$, for a better visibility. The Kalman filter from the antidevelopment, the linear and geodesic interpolation based filters roughly estimate the state at $\hat{x} = (1.2, 0.2, 1.1)^T$ and the final variance from 2.21 is $P_6 = 0.15$. The figure represents the reconstruction of the rotation signal from the real time estimation of the angular velocity. Estimating the angular velocity explains the large drift observed in the rotation reconstruction with respect to the non noisy original signal.

In this section, the filtering problem has been extended to observation on $SO(n)$. The presence of multiplicative noise prevents the use of the classical methods of resolution presented in Section 2.1. However, a one-to-one process, the antidevelopment, can be constructed from the observation to satisfy a model with additive noise. The filtering

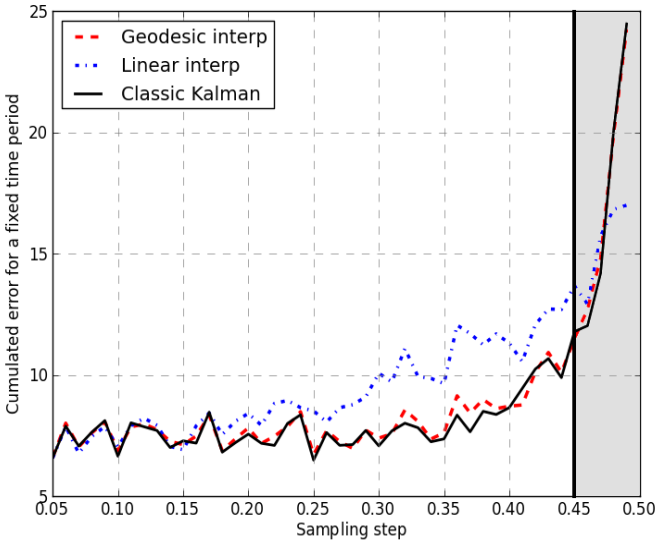


Figure 2.3: Cumulated error of estimation for $0 \leq t \leq 50$ for the linear interpolation, the geodesic interpolation, and the classic Kalman filter for different values of δt_{obs}. The state to estimate is constant and the variance of the observation is $\sigma_w = 1$. Each point represents the average of 20 simulations configured with the same parameters. For a sampling time larger than 0.45, the performances are too poor to consider the filters to converge properly anymore.

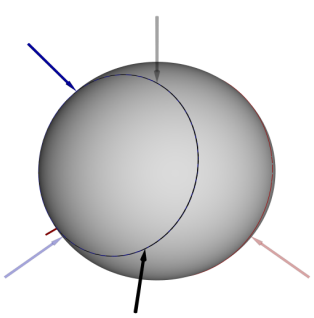
problem with multiplicative noise is then reduced to a classical filtering problem and results from Section 2.1 can be applied. In practice, the computation of the antidevelopment requires an interpolation of the rotation process. Different functions of interpolation have been presented and compared. The geodesic interpolation is accurate but requires heavy computations. The linear interpolation is not so accurate for large sample step but is much simpler in term of computation.

2.3 Monte-Carlo method for filtering from observations on the Stiefel manifold

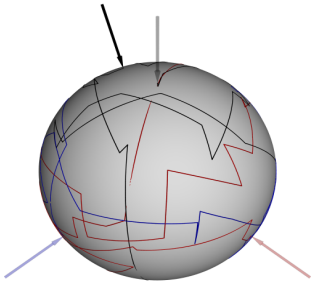
In the previous section, an angular velocity has been estimated from a rotational signal. It is supposed that this signal is observed via sensors or several cameras, depending on the application. However, it might happen that some of the observation devices are faulty or that the complete acquisition is not possible. This section considers this aspect of filtering from partial observation on $SO(n)$. In this case, we will suppose that the observation lays in the Stiefel manifold. After a brief description of the geometry of the Stiefel manifold, a new method of filtering is presented, including its numerical implementation.

2.3.1 Geometry of the Stiefel manifold

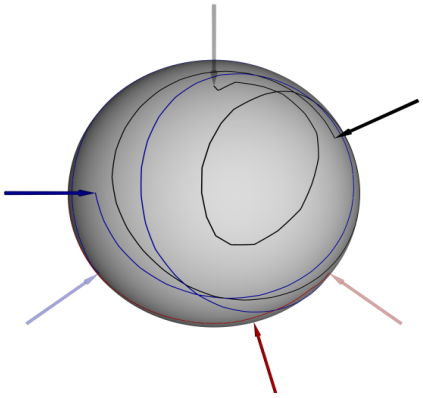
The Stiefel manifold can be described, for a fixed $n, k \in \mathbb{N}$ such that $k \leq n$, roughly speaking, as the set of k orthonormal elements of $\mathbb{R}^n$. It is mainly used in linear algebra to describe principal subspaces [Smith 1998]. Stiefel manifolds can also be used for application in signal processing such as sensor arrays for example [Weiss 1999], in channel estimation



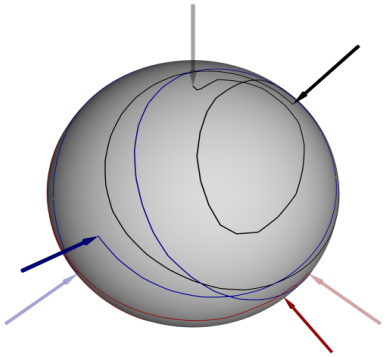
(a) Non noisy signal



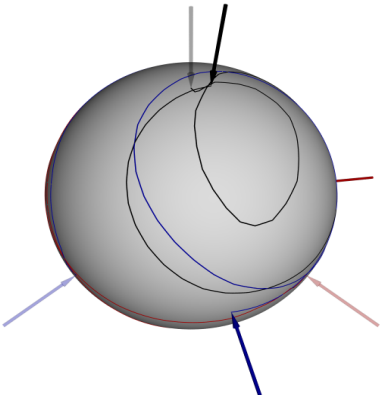
(b) Noisy observation



(c) Kalman filtering from the antidevelopment



(d) Filtering with Geodesic interpolation



(e) Filtering with Linear interpolation

Figure 2.4: Example of filtering and reconstruction from observation on $SO(3)$. Each color represents the trajectory of a different column of the rotation process. The path on the sphere is the trace of the whole process. Its value at $t = 0$ is marked by transparent arrows. Its value at $t = 1$ is marked by opaque arrows. 2.4(a): Original non noisy signal. It corresponds to Equation (2.15) with $\sigma_w^2 = 0$. 2.4(b): Noisy observation with $\sigma_w^2 = 1$. 2.4(c): From the complete knowledge of the antidevelopment, a Kalman filter is applied from the observation of the noisy signal. It constitutes the reference filter, knowing that this is the optimal filter one can achieve. 2.4(d): Reconstruction of the rotation signal from Equation (2.21) using a geodesic interpolation. 2.4(e): Reconstruction of the rotation signal from Equation (2.21) using a geodesic interpolation. The difference between the original and the reconstructed signals is mainly due to the drift (a rotation on $SO(3)$) generated at early times, as we are only looking to estimate the angular velocity.

for wireless communications [Arslan 2007] or in computer vision for representing a scene independently from the lighting characteristics [Kirby 2009].

For any positive integers $n, k \in \mathbb{N}$ with $k \leq n$, the Stiefel manifold $V_{n,k}$ is defined as the set of all matrices $P \in \mathbb{R}^{n\times k}$ such that

$$P^T P = I_k \qquad \text{with } I_k \text{ the } k \times k \text{ identity matrix.} \tag{2.24}$$

In particular, if $k = 1$, then $V_{n,1}$ is the sphere S^{n-1} embedded in $\mathbb{R}^n$. If $k = n$, then $V_{n,n}$ is the orthogonal group $O(n)$.

Let $\Pi : SO(n) \to V_{n,k}$ be the projection truncating the $n-k$ last columns from rotation matrix. In other words,

$$\Pi(R) = P \quad \text{with the concatenation } R = (P,\, P') \text{ and } P \in \mathbb{R}^{n\times k},\ P' \in \mathbb{R}^{n\times(n-k)} \tag{2.25}$$

When $k \leq n-2$, the projection is not injective. In this case, a matrix $P \in V_{n,k}$ can be completed by different sets of orthonormal vectors to form an oriented orthonormal basis of $\mathbb{R}^n$

$$\Pi(R_1) = \Pi(R_2) \Leftrightarrow R_1 = R_2 \begin{pmatrix} I_k & 0 \\ 0 & r \end{pmatrix} \qquad \text{with } R_1, R_2 \in SO(n),\ r \in SO(n-k)$$

However, if $k \leq n$, then Π is clearly surjective, *i.e* $\Pi(SO(n)) \in V_{n,k}$ as the k first columns of a rotation matrix are orthonormal vectors. Therefore, $\Pi(R)^T\Pi(R) = I_k$ for $R \in SO(n)$. The Stiefel manifold $V_{n,k}$ can then be described as

$$V_{n,k} = \{\Pi(R), R \in SO(n)\} \tag{2.26}$$

The case $k = n$ is slightly different. Indeed, $V_{n,n} = O(n)$ is the group of orthonormal matrix and is composed of two connected components: the set of orthonormal matrices with a positive determinant (representing positive oriented basis) $SO(n)$ and the set of orthonormal matrices with a negative determinant (representing negative oriented basis). Further on in this chapter, only continuous processes $P_t \in V_{n,k}$ will be considered. As P_t is continuous, then $\det(P_t) = \det(P_0)$ and the process will stay in the same component as P_0. Therefore, if $\det(P_0) = +1$, then $\Pi = Id$ covers all the reachable points in the Stiefel manifold from $SO(n)$. If $\det(P_0) = -1$, then considering $\Pi(R)$ as the application inverting the sign of the last column of R, then Π also covers all the reachable points in the Stiefel manifold from $SO(n)$. Consequently, the definition (2.26) can be extended to $k = n$, by considering only one connected component.

As $V_{n,k}$ can be constructed from $SO(n)$, it is legitimate to verify if the geometry of $V_{n,k}$ can be described from the geometry of $SO(n)$. The rest of Subsection 2.3.1 is dedicated to describe the geometry of $V_{n,k}$ from the notions introduced in Chapter 1 for $SO(n)$.

From its definition, it can easily be shown that the projection Π is left invariant

$$R_1\Pi(R_2) = \Pi(R_1R_2) \qquad \text{with } R_1, R_2 \in SO(n). \tag{2.27}$$

As Π is surjective, this property also shows the action of $SO(n)$ on $V_{n,k}$, if $R \in SO(n)$ and $P \in V_{n,k}$, then $RP \in V_{n,k}$. This will be useful further on to describe a process on $V_{n,k}$ via the action of the group $SO(n)$.

First of all, we want to determine the tangent space to $V_{n,k}$. This will be useful in Subsection 2.3.2 to describe stochastic processes. Let $\chi : \mathfrak{so}(n) \times V_{n,k} \to TV_{n,k}$ be the application

$$\chi(\sigma, P) = \frac{d}{dt}|_{t=0} \exp(t\sigma)P = \sigma P \tag{2.28}$$

We can show that $\chi(., P)$ is surjective onto $T_PV_{n,k}$, *i.e* $T_PV_{n,k} = \{\sigma P, \sigma \in \mathfrak{so}(n)\}$.

For a given point $P \in V_{n,k}$, let $R \in SO(n)$ be a pre-image of P via Π, *i.e* $P = \Pi(R)$. As Π is surjective, $\Pi^{-1}(P) \neq \varnothing$ and R is well defined. The vertical space [Kirby 2009] at the point P, denoted $\mathcal{V}_R$, is defined as

$$\mathcal{V}_R = \operatorname{Ker} d\Pi_R, \tag{2.29}$$

where $d\Pi_R$ is the differential of Π at the point R. By definition of $d\Pi_R$, the vertical space $\mathcal{V}_R$ is a subspace of the tangent space $T_R SO(n)$.

Thanks to the metric on $SO(n)$ introduced in (1.3), it is possible to construct the orthogonal complement of $\mathcal{V}_R$ called horizontal space and denoted $\mathcal{H}_R$

$$T_R SO(n) = \mathcal{V}_R \oplus \mathcal{H}_R.$$

As $d\Pi_R$ is linear, the restriction of $d\Pi_R$ to $\mathcal{H}_R$ is bijective. In other words, $T_P V_{n,k}$ and $\mathcal{H}_R$ are isomorphic. For a vector $v \in T_P V_{n,k}$, let $v^{\mathcal{H}} \in T_R SO(n)$ be the element defined as

$$d\Pi_R(v^{\mathcal{H}}) = v.$$

For example, let consider $n = 3$ and $k = 1$. The Stiefel manifold $V_{n,k}$ is the sphere S^2. Considering $P = (1, 0, 0)^T \in V_{n,k}$, the matrix $R_1 = (e_1, e_2, e_3)$ and $R_2 = (e_1, e_3, -e_2)$ for $\{e_i\}_{i \leq 3}$ the canonical basis of $\mathbb{R}^3$ are both pre-image of P, $\Pi(R_1) = \Pi(R_2) = P$. The application $\chi(., P)$ describes the tangent space $T_P V_{n,k}$ as $\chi(\sigma, P) = (0, \sigma_{21}, \sigma_{31})^T$ for $\sigma = (\sigma_{i,j})$. At the pre-image R_1, the vertical and horizontal spaces are defined as

$$\mathcal{V}_{R_1} = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \alpha \\ 0 & -\alpha & 0 \end{pmatrix}, \alpha \in \mathbb{R} \right\} \qquad \mathcal{H}_{R_1} = \left\{ \begin{pmatrix} 0 & \beta & \gamma \\ -\beta & 0 & 0 \\ -\gamma & 0 & 0 \end{pmatrix}, \beta, \gamma \in \mathbb{R} \right\}$$

At the pre-image R_2, the vertical and horizontal spaces are defined as

$$\mathcal{V}_{R_2} = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix}, \alpha \in \mathbb{R} \right\} \qquad \mathcal{H}_{R_2} = \left\{ \begin{pmatrix} 0 & \beta & \gamma \\ \gamma & 0 & 0 \\ -\beta & 0 & 0 \end{pmatrix}, \beta, \gamma \in \mathbb{R} \right\}$$

Despite that the horizontal subspaces $\mathcal{H}_{R_1}$ and $\mathcal{H}_{R_2}$ are different, as they are defined by a different pre-image for P, the spaces $\mathcal{H}_{R_1} R_1^T$ and $\mathcal{H}_{R_2} R_2^T$ are identical. An illustration of the notion of horizontal and vertical space is given in figure 2.5

Due to the isomorphism between $T_P V_{n,k}$ and $\mathcal{H}_R$, the application $\chi(., P)$ restricted to $\mathcal{H}_R R^T \in \mathfrak{so}(n)$ with $\Pi(R) = P$ is bijective. Let $\chi^{-1} : T V_{n,k} \to \mathfrak{so}(n)$ denotes its inverse

$$\chi^{-1}(v, P) = v^{\mathcal{H}} R^T. \tag{2.30}$$

The term $v^{\mathcal{H}}$ is the horizontal vector from the tangent space to R. Despite that the definition of χ^{-1} seems to be dependent of R, this is not the case as $v^{\mathcal{H}}$ also depends of R. However, in the end, the term $v^{\mathcal{H}} R^T$ in independent of R.

It is possible to define a metric based on the metric of $SO(n)$. Let $< ., . >_P$ be the metric defined as

$$< v_1, v_2 >_P = < v_1^{\mathcal{H}}, v_2^{\mathcal{H}} >_R \qquad \text{with } v_1, v_2 \in T_P V_{n,k}, \Pi(R) = P \tag{2.31}$$

where $< ., . >_R$ is the metric defined in (1.3).

In the next paragraph, we solve the problem of filtering with observation on the Stiefel manifold $V_{n,k}$. The solution consists in considering a pre-image of the observation and computing a likelihood in the horizontal space.

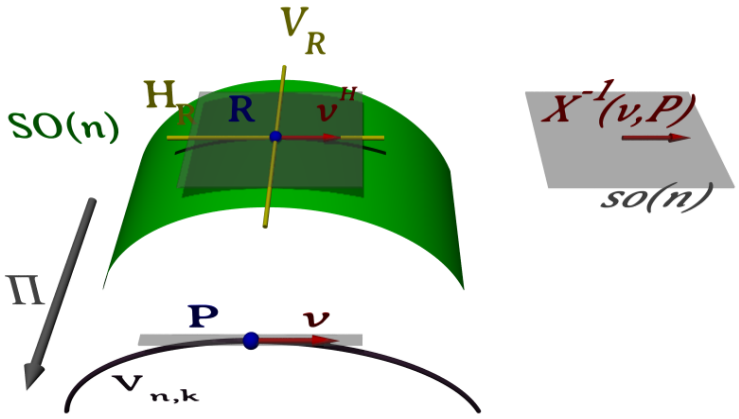


Figure 2.5: Illustration of the different notions introduced to describe the Stiefel manifold $V_{n,k}$ as the image from the projection Π of $SO(n)$. The horizontal $\mathcal{H}_R$ and vertical $\mathcal{V}_R$ spaces are dependent of the chosen pre-image but the translation into $\mathfrak{so}(n)$ via χ^{-1} is invariant with respect to the choice of the preimage.

2.3.2 Solution to the filtering on the Stiefel manifold

In Section 2.2, we were concerned about filtering from observation on $SO(n)$. In practice, this observation comes from sensors or devices that can be flawed and only a limited part of the signal to filter is observed. For example, in satellite control, the existing algorithms require the knowledge of the angular velocity and the orientation of the satellite to control its orientation [Lovera 2002][Bloch 1990]. This angular velocity is determined from different sensors. However, if some of these sensors became faulty, the velocity of the satellite is no more available and the satellite cannot be controlled properly anymore. The proposed algorithm proposes to take advantage of the available data to perform optimal filtering.

In this section, we propose a solution to the filtering problem with partial observation on $SO(n)$. This solution is adapted from the general solution presented in [Manton 2012]. The partial observation from a process R_t on $SO(n)$ is here modeled as a process P_t on the Stiefel manifold $V_{n,k}$, *i.e* only k components amongst the total of n components are known. Using the notions introduced in the previous paragraph, the filtering problem reads: we want to estimate $x_t \in \mathbb{R}^{n(n-1)/2}$ from $y_t \in V_{n,k}$ defined as $P_t = \Pi(R_t)$. Projecting Equation (2.15) becomes

$$dP_t = (\hat{x}_t dt + \circ d\hat{w}_t)\, P_t, \tag{2.32}$$

where $w_t \in \mathbb{R}^{n(n-1)/2}$ is a Brownian motion with variance σ_w^2 independent from x_t and the application $x \mapsto \hat{x}$ is defined in (1.4).

We are looking for $\pi_{[0,t]}$, the distribution of x_t conditioned by the observation $\mathcal{P}_t = \{P_s, s \leq t\}$. Similarly to Section 2.2, the noise is here multiplicative. However, using the solution presented in Section 2.2 is not possible. As $V_{n,k}$ is not a group, it is not possible to construct the antidevelopment from Equation (2.32). The solution presented in this section relies on the same concept: constructing a process z_t similar to P_t in terms of information such that z_t is solution of a stochastic differential equation with additive noise and then, describe $\pi_{[0,t]}$ from Equation (2.18).

Let $z_t \in \mathfrak{so}(n)$ and $R_t \in SO(n)$ be defined as

$$\begin{aligned} dz_t &= \chi^{-1}(\circ dP_t, P_t) \\ dR_t &= (\circ dz_t)R_t \end{aligned} \tag{2.33}$$

where χ^{-1} is defined in (2.30).

The process z_t is called the antidevelopment of P_t and R_t is called the horizontal lift of P_t [Sepulchre 2008]. These processes are equivalent, in terms of information to P_t as P_t can be constructed like

$$P_t = R_t R_0^T P_0 \qquad \text{or} \qquad dP_t = \chi(dz_t, P_t).$$

Conditioning the distribution of x_t by the observation of P_t is the same as conditioning by the observation of z_t. However, the antidevelopment is a solution, as opposed to P_t, of a stochastic differential equation with additive noise.

Replacing dP_t in (2.33) by its expression from (2.32) gives

$$\begin{aligned} dz_t &= \chi^{-1}(dP_t, P_t) \\ &= \chi^{-1}\left((\hat{x}_t dt + \circ d\hat{w}_t)\, P_t, P_t\right) \\ &= \chi^{-1}\left(\left(\chi\left(\hat{x}_t dt + \circ d\hat{w}_t\right), P_t\right), P_t\right) \\ &= \chi^{-1}\left(\chi\left(\hat{x}_t dt, P_t\right), R_t\right) + \circ\chi^{-1}\left(\chi\left(d\hat{w}_t, P_t\right), R_t\right) \\ &= \chi^{-1}(H_t, P_t) + \circ d\hat{\beta}_t, \end{aligned} \tag{2.34}$$

where $H_t = \chi(\hat{x}_t, P_t)$ and $d\hat{\beta}_t = \chi^{-1}\left(\chi\left(\circ d\hat{w}_t, P_t\right), P_t\right)$. By definition of $d\hat{\beta}_t$, the process β_t is constructed from the k first components of $\hat{w}_t$. Therefore, β_t is a Brownian process with a variance that can be diagonalized as $\sigma_w^2 I_{n,k}$ where $I_{n,k} = \text{diag}(1, ..., 1, 0, ..., 0)$ with k non zero elements. An example of antidevelopment is presented in figure 2.6 for $V_{3,1} = S^2$.

Thus, the filtering problem from observation in the Stiefel manifold $V_{n,k}$ with multiplicative noise is now reduced to a filtering problem in $\mathbb{R}^{n(n-1)/2}$ with additive noise. The solution is therefore given by applying (2.18) to the antidevelopment z_t defined in (2.33)

$$\pi_{[0,t]}(\phi) = \frac{\rho_{[0,t]}(\phi)}{\rho_{[0,t]}(1)}, \tag{2.35}$$

with $\rho_{[0,t]}(\phi) = \mathbb{E}\left[\phi(x')L_t(P, x')|\mathcal{P}_t\right]$, where x'_t is a copy of x_t independent of P_t and $\mathcal{P}_t = \{P_s, s \leq t\}$. The likelihood L_t is defined as

$$L_t(P, x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < x'_s, dz_s >_{\mathfrak{so}(n)} - \frac{1}{2\sigma_w^2}\int_0^t ||x'_s||^2 ds\right) \qquad \text{with } dz_t = \chi^{-1}(\circ dP_t, P_t). \tag{2.36}$$

By definition of the inner product in (2.31), the likelihood can be written as

$$L_t(P, x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < H_s, dP_s >_{V_{n,k}} - \frac{1}{2\sigma_w^2}\int_0^t ||H_s||^2 ds\right) \qquad \text{with } H_s = \chi(\hat{x}'_s, P_s). \tag{2.37}$$

The expression (2.37) is preferable to the form from Equation (2.36) as it does not require the computation of R_t. The term under the integral can directly be determined from the observation without constructing an auxiliary process. However, using the expression (2.37), the model for H_t is not linear, even if x_t was solution of a linear model (a model similar to Equation (2.14) for example). Therefore, to use the likelihood from (2.37), the solution presented in Subsection 2.1.3 cannot be used anymore. However, it is still possible to get an approximation of the solution, based on a numerical method, like the particle filter for example.

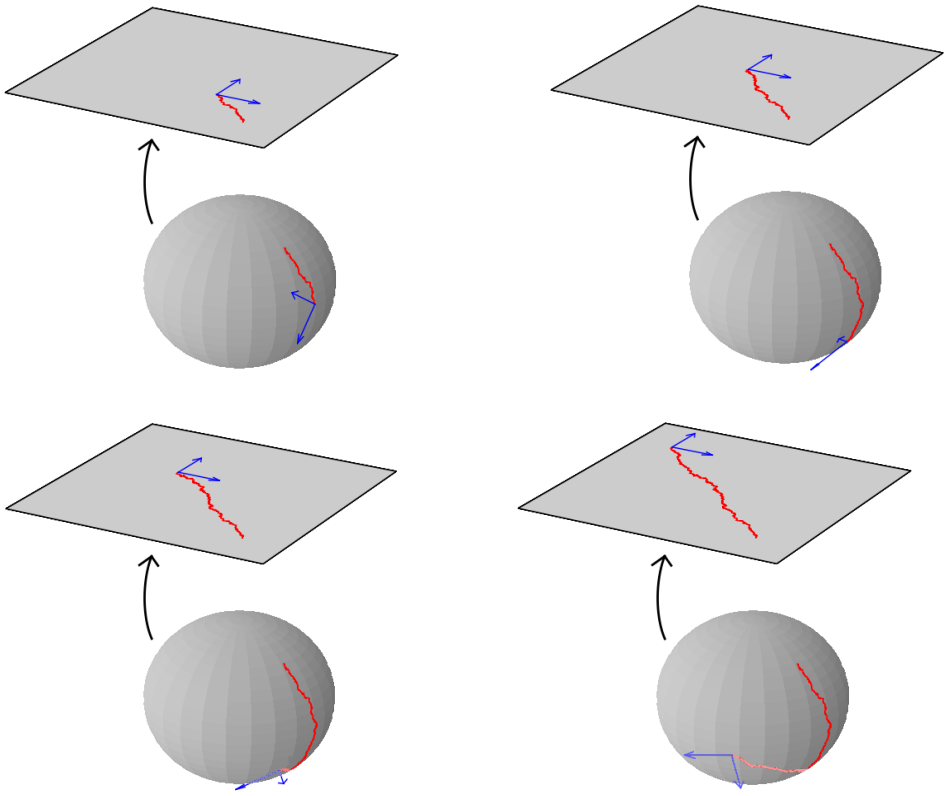


Figure 2.6: Example of trajectory of P_t (red) on $V_{3,1} = S^2$ at 4 successive times (top left, top right, bottom left, bottom right). Over the sphere in a plane, a visualization of the antidevelopment z_t. It can be obtained by considering the trace left by the sphere by rolling without slipping on the plan, with P_t as a contact point. The antidevelopment is solution of a stochastic differential equation with additive noise, as opposed to P_t.

2.3.3 Implementation of the solution to filtering on the Stiefel manifold

The particle filter is a Monte-Carlo method to approximate the solution given by Equation (2.18) for non linear model. Despite that particle filters has been heavily used and studied [Johansen 2011], the application of this method to perform estimation from partial observation on the Stiefel manifold has not been used before.

The main idea of the particle filter is to approximate the expectation in (2.35) using the law of large numbers. Recall

$$\rho_{[0,t]}(\phi) = \mathbb{E}\left[\phi(x')L_t(P,x')|\mathcal{P}_t\right]$$

with L_t the likelihood defined in Equation (2.37). The process x'_t is a copy of x_t (in the sense with the same model of propagation) but, contrary to x_t, independent from P_t.

Let X_t^i be N processes identical to x_t called particles. They represent candidates to estimate the process x_t. The law of large numbers states that $\rho^N_{[0,t]}$ defined as

$$\rho^N_{[0,t]}(\phi) = \frac{1}{N}\sum_{i=1}^{N}\phi(X^i)L_t(P,X^i)$$

will converge with N almost surely to $\rho_{[0,t]}(\phi)$. Therefore, the previous equation gives an approximation of the solution by determining $\rho^N_{[0,t]}$ and normalizing the coefficients $L_t(P,X^i)$. It is noticeable that the particles are not observable and must be simulated. It supposes that the model of propagation for x_t is known.

We have seen in Subsection 1.2.2 that for simulating stochastic processes, they first need to be discretized. Furthermore, it is assumed that the process P_t is not continuously observed. Let δt be the sampling time. These remarks lead to consider a time discretized version of $\rho^N_{[0,t]}$, denoted ρ^N_n, with $n = \lfloor t/\delta t \rfloor$ as

$$\rho^N_n(\phi) = \sum_i \phi(\tilde{X}^i)L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}^i_0, ..., \tilde{X}^i_n)$$

where $\tilde{X}^i_n = X^i_{n\delta t}$, $\tilde{\mathcal{P}}_n = \{P_{k\delta t}, k \leq n\}$ and the likelihood $L_{\delta t}$ is defined as

$$L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}^i_0, ..., \tilde{X}^i_n) = \exp\left(\frac{1}{\sigma_w^2}\sum_{k\leq n} < H^i_k, \delta\tilde{P}_k > -\frac{1}{2\sigma_w^2}\sum_{k\leq n}||H^i_k||^2\right)$$

with $H^i_k = \chi(X^i, \tilde{P}_k)$. It has been proved in Proposition 16 that the discrete likelihood $L_{\delta t}$ converges towards the likelihood L_t from (2.37).

To implement such a solution, there are two independent problems:

- The simulation of the particles $\tilde{X}^i$:
  It will be supposed that x_t is a Markov process with a generator $\mathcal{A}$. Consequently, $\tilde{X}^i$ is a Markov chain with transition kernel $q_{\delta t}$ defined in (2.8) and $\tilde{X}^i_{n+1}$ is directly sampled from $q_{\delta t}(\ .\tilde{X}^i_n)$.

- The computation of the likelihood $L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}^i_0, ..., \tilde{X}^i_n)$:

Considering the last term of the Riemannian sums:

$$\begin{aligned} L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}_0^i, ..., \tilde{X}_n^i) &= \exp\left(\frac{1}{\sigma_w^2}\sum_{k\leq n} < H_k^i, \delta\tilde{P}_k > -\frac{1}{2\sigma_w^2}\sum_{k\leq n}||H_k^i||^2\right) \\ &= \exp\left(\frac{1}{\sigma_w^2}\sum_{k\leq n-1} < H_k^i, \delta\tilde{P}_k > -\frac{1}{2\sigma_w^2}\sum_{k\leq n-1}||H_k^i||^2\right) \\ &\quad \times \exp\left(\frac{1}{\sigma_w^2} < H_n^i, \delta\tilde{P}_n > -\frac{1}{2\sigma_w^2}||H_n^i||^2\right) \\ &= L_{\delta t}(\tilde{\mathcal{P}}_{n-1}, \tilde{X}_0^i, ..., \tilde{X}_{n-1}^i) l_{\delta t}^i \end{aligned}$$

where $l_{\delta t}^i = \exp\left(\frac{1}{\sigma_w^2} < H_n^i, \delta\tilde{P}_n > -\frac{1}{2\sigma_w^2}||H_n^i||^2\right)$. This decomposition shows that the likelihood can be computed adaptively when new samples are available.

This leads to Algorithm 1, here described to estimate the conditional distribution $\pi_n^{N,\delta t}(\phi)$ as

$$\pi_n^{N,\delta t}(\phi) = \frac{\rho_n^{N,\delta t}(\phi)}{\rho_n^{N,\delta t}(1)} = \sum_{i=1}^N \phi(\tilde{X}_n^i) w_n^i$$

where the coefficient

$$w_n^i = \frac{L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}_0^i, ..., \tilde{X}_n^i)}{\sum_{j=1}^N L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}_0^j, ..., \tilde{X}_n^j)}$$

is called the "weight" associated to the particle i.

Algorithm 1 Particle filter algorithm

- For the initialization, generate N particles from a priori p_0: $\tilde{X}_0^i \sim p_0$ and set $w_0^i = 1/N$.
- At a time $n > 0$:
  1. Propagate the particles $\tilde{X}_n^i \sim q_{\delta t}(\ .\ , \tilde{X}_{n-1}^i)$
  2. Update the weight w_n^i of each particle as: $w_n^i = w_{n-1}^i l_{\delta t}^i$ with
  $$l_{\delta t}^i = \exp\left(\frac{1}{\sigma_w^2} < H_n^i, \delta\tilde{P}_n > -\frac{1}{2\sigma_w^2}||H_n^i||^2\right).$$
  3. Normalize the weights: $w_n^i = w_n^i / \sum_j w_n^j$
  4. If $\left(\sum_i (w_n^i)^2\right)^{-1} < N/2$, generates $[m^1\ ...\ m^N] \sim$ multinomial$(w^1\ ...\ w^N)$ such that $\sum_i m^i = N$. Then, clones $\tilde{X}_n^i$ m^i-times and set $w_n^i = \frac{1}{N}$.
  5. Estimate $\pi_t(\phi)$ with $\pi_n^{N,\delta t}(\phi) = \sum_i \phi(\tilde{X}_n^i) w_n^i$

The normalization step (step 3) is not only here to compute $\pi_n^{N,\delta t}$ instead of $\rho_n^{N,\delta t}$ but also to numerically stabilize the computation of the weights.

Step 4 is called resampling. It is here to prevent a degeneracy due to the finite number of particles. Indeed, the particles are propagating without any restriction or drift imposed by the observation. If x_t is a diffusion, then the particles will consequently tends to explore the whole space. For example, in figure 2.7, x_t (black trajectory) is modeled as a Brownian

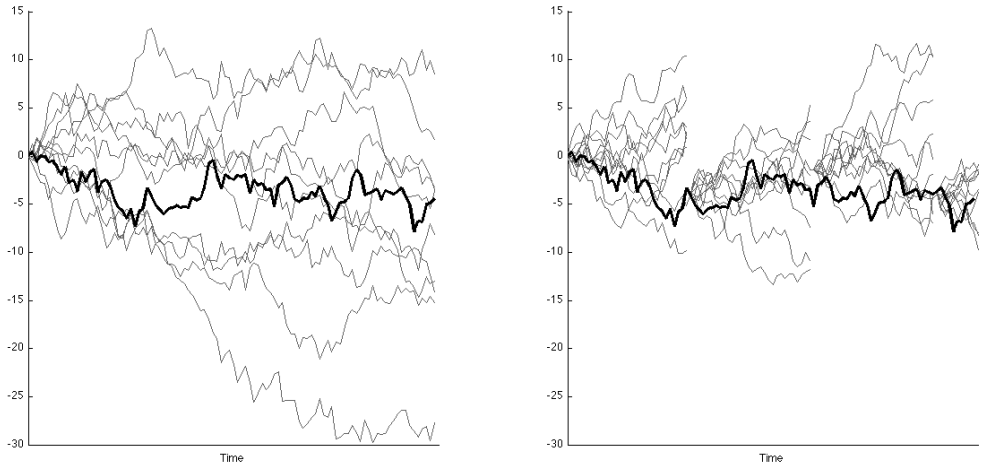


Figure 2.7: Trajectories of 10 particles (grey) to estimate a Brownian diffusion (black). (Left) Without resampling: The trajectory of the particles has no feedback from the likelihood and they propagate without restrictions. Quickly, it degenerates as the weights associated to each particle become very low. (Right) With resampling: After a constant time step, the particles are resampled. Those with a high likelihood are cloned and the particles with a low likelihood are killed. Despite that the independence between the particles is lost, the resampling step prevents the particle filter from degenerating.

motion. The particles (grey) are also modeled as a Brownian motion. Without resampling (left), the particles just explore the space and as they tend to drift away from x_t, they become a bad approximation of x_t, the mean square error $\mathbb{E}[(X_t^i - x_t)^2]$ is linearly growing with respect to the time. The number of particles being fixed, their weights quickly degenerate as they are diffusing away from x_t. Due to the normalization step, this leads to the concentration of all the ponderation into one single particle. Even if this particle is the best candidate amongst the all the particles, the mean square error is still linearly growing.

The resampling step consists in killing the particles far away from x_t (in fact, killing the particles with low weights) and cloning the remaining ones. To measure if the particles are scattered away from x_t, one commonly used criteria is a threshold based on the Effective Sample Size (ESS_w) defined as

$$ESS_w = \left(\sum_i (w_n^i)^2 \right)^{-1}.$$

When ESS_w is lower that say $N/2$, then the particles needs to be resampled (the criteria ESS_w is small when only a few particles have a preponderant weight). To resample the particles, one can for example sample $[m^1 \ ... \ m^N]$ from a multinomial distribution

$$[m^1 \ ... \ m^N] \sim \text{multinomial}(w^1 \ ... \ w^N)$$

such that $\sum_i m^i = N$ (to keep the number of particles constant) and clone the i^{th} particle m^i times. If w^i is high (particle with a good likelihood, thus a good candidate), then m^i should be high too. This effect will tend to keep only the good candidates, based on the likelihood. An example of filtering with resampling is presented in figure 2.7 (right side). However, instead of resampling when the Effective Sample Size becomes too low, resampling is made after a fixed given time. In fact, resampling can be realized at every iteration but it is, computationally speaking, expensive and does not bring noticeable improvements.

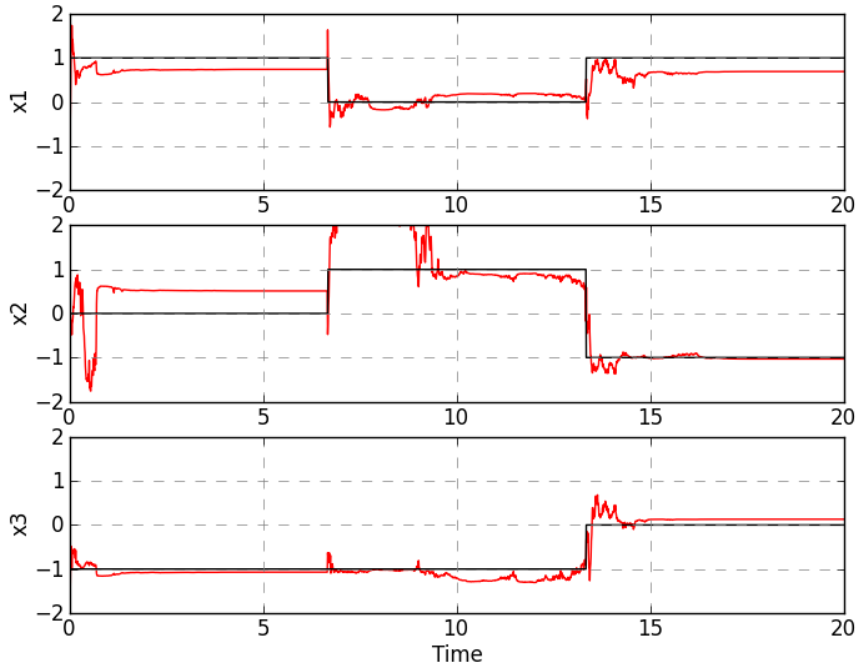


Figure 2.8: Evolution of the estimation (red) of each component of x_t (black), namely x_1, x_2 and x_3. The process x_t is here chosen as a stair function. As the present algorithm can estimate a constant state, it is also able to estimate a stair function, including a proper resampling of the particles when x_t varies abruptly.

2.3.4 Limitations of the solution

This subsection describes the results from a numerical implementation of the particle filter presented in Algorithm 1. For this simulation, the Stiefel manifold was chosen as the sphere $V_{3,1} = S^2$. The process $x_t \in \mathbb{R}^3$ is a Brownian motion with a unit variance. The variance of the noise is fixed to $\sigma_w^2 = 1$.

To approximate x_t, $N = 500$ particles are generated from a normal prior distribution p_0 centered around the origin with a variance 2. Using more particles does not significantly improve the results. Considering a bad prior for generating the particles is not a big issue as the resampling step quickly eliminates the wrong candidates for the estimation. The time step for the observation is $\delta t_{\text{obs}} = 10^{-2}$. However, the time step for the simulation has been fixed to $\delta t_{\text{gen}} = 10^{-3}$. Different time steps are required for the same reasons as Subsection 2.2.2. Figure 2.8 illustrates the results obtained for the estimation when the state x_t is a stair function. As x_t is constant, the algorithm is able to estimate x_t. Even if this has not been implemented for the simulation in Figure 2.8, once could use a classical algorithm to detect abrupt changes in x_t in order to estimate the time where the particles should be sampled [Nikiforov 1993]. When a change is detected, the particles are sampled from the initial priori again to converge towards the new value.

In this case, the particles will not drift away because they are at a constant position (they propagate with the same model than x_t). They only merge when they are resampled, letting less and less possible choice. It has been shown that the vertical component could not be estimated. However, the vertical space is defined with repsect to the observation

point P_t. As P_t will evolve on $V_{n,k}$, the vertical space will change too and the component on the initial vertical space can be estimated. At the end, it will be possible to completely estimate the angular velocity.

However, if the angular velocity is evolving too quickly (with respect to the time the particles need to converge), it will not be possible to estimate the vertical component of x_t, from the observation of P_t. The results presented in figure 2.9 show the evolution of the same simulation at successive times (from top to bottom) with a different model for the state x_t. In this case, x_t is a Brownian motion in $\mathbb{R}^3$ with a variance $\sigma_b^2 = 1$. The variance of the observation has been chosen $\sigma_w^2 = 1$. On the left of each panel of Figure 2.9, the evolution of the particles is illustrated in yellow and the true state x_t in blue. On the right side, the trajectory of the observed process, the spherical process P_t.

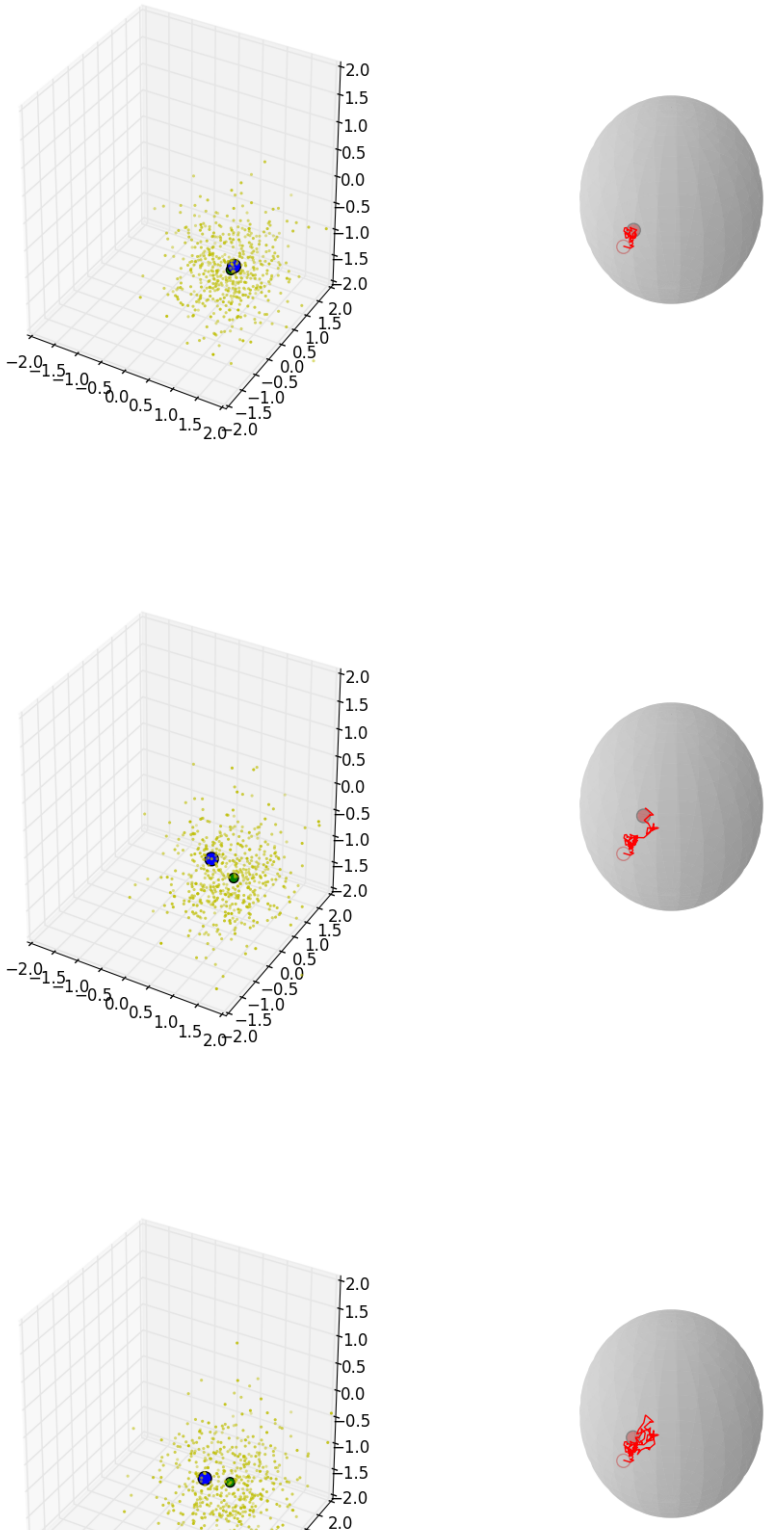
The particles are distributed within an ellipse (see Figure 2.10), whose large axis coincide with the direction of P_t. Along the direction of P_t, the estimation of x_t, the empirical average of the particles (yellow dots), is not satisfactory. However, in the other direction, the estimation is correct. This is due to the fact that the particles can only track the component of x_t that has an impact on P_t. As the innovation term in Equation (2.37) is $< H_s, dP_s >$, only the horizontal component can be observed and therefore estimated. Consequently, particles propagating along the vertical subspace (wich is, for S^2, the line defined by P_t) are not penalized (their weight do not decrease) and are still considered as good candidates. Therefore, the estimate is correct in the horizontal direction, but not in the vertical direction. An illustration of this phenomenon is described in Figure 2.11.

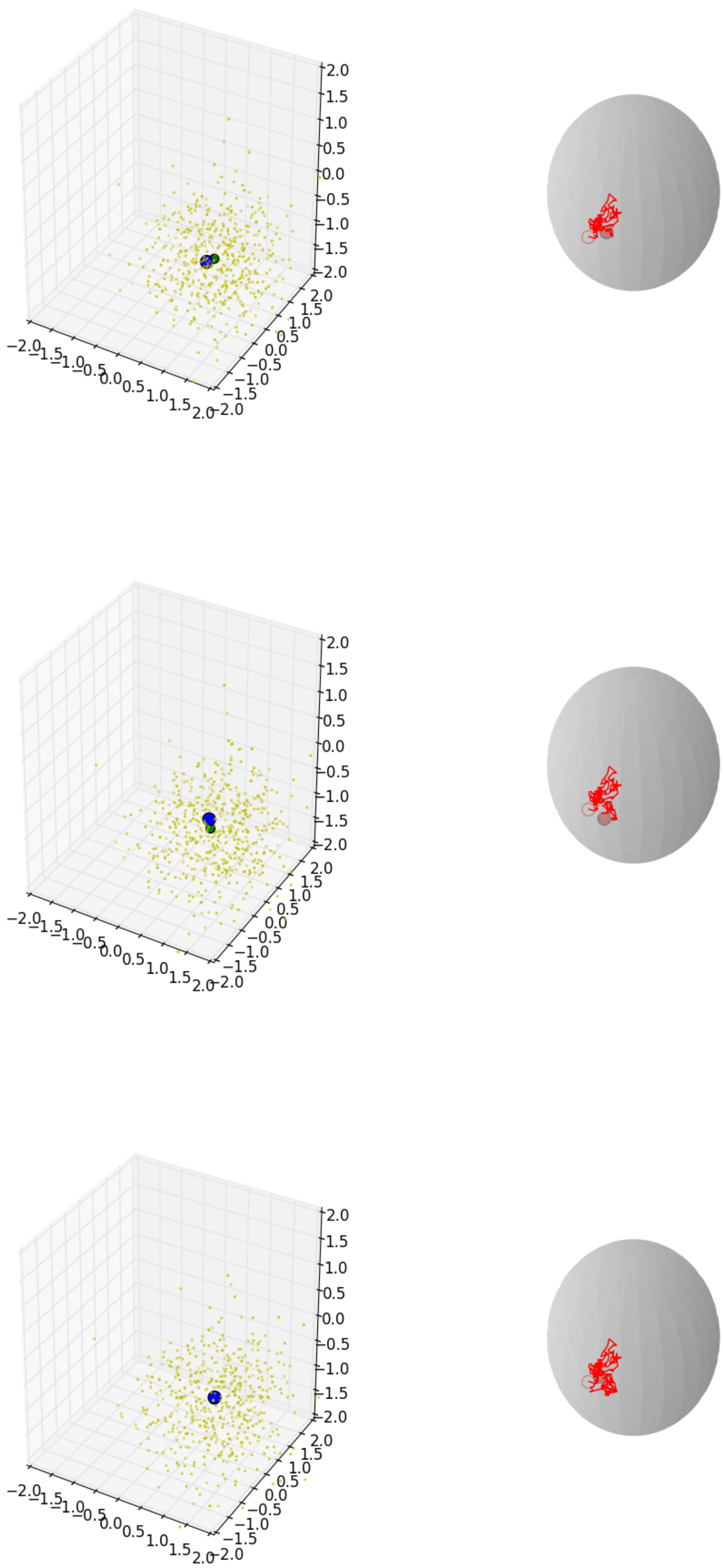
2.4 Conclusion

In this chapter, the filtering problem has been presented and numerically solved for observation in $SO(n)$ and the Stiefel manifold $V_{n,k}$. A rapid introduction presented a general solution for filtering from observation with an additive noise in $\mathbb{R}^n$. The problem of filtering is then described for observation in $SO(n)$, the state to estimate being the angular velocity of the observation process. Contrary to the case from 2.1, observation in $SO(n)$ are modeled with a multiplicative noise, as introduced in Chapter 1.

Methods based on extended Kalman filter create filtering equations with non constant coefficients. These coefficients depend of the current estimation. In Section 2.2, we proposed a solution based on the construction of a one-to-one process, the antidevelopment, constructed from the observation but modeled with an additive noise. Once this process is obtained, the filtering problem is reduced to a classic problem, presented in Section 2.1. Further new results discuss about the case of discrete observation and the way to implement the previous solution.

In the case of observation in the Stiefel manifold $V_{n,k}$, a similar original method is derived. Observation on the Stiefel manifold can be used to model a partial observation from a process on $SO(n)$. However, the lack of information prevents us from using the same method as presented in Section 2.1. The solution requires a numerical resolution. A Monte-Carlo method based on particles is presented. These particles will model candidates for the trajectory of the state to estimate and will be weighted, based on their likelihood with the observation. Yet, this method only gives a partial estimation of the state at a given time. This is due to the fact that some components of the angular velocity leave the observation unchanged.


2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0


2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0

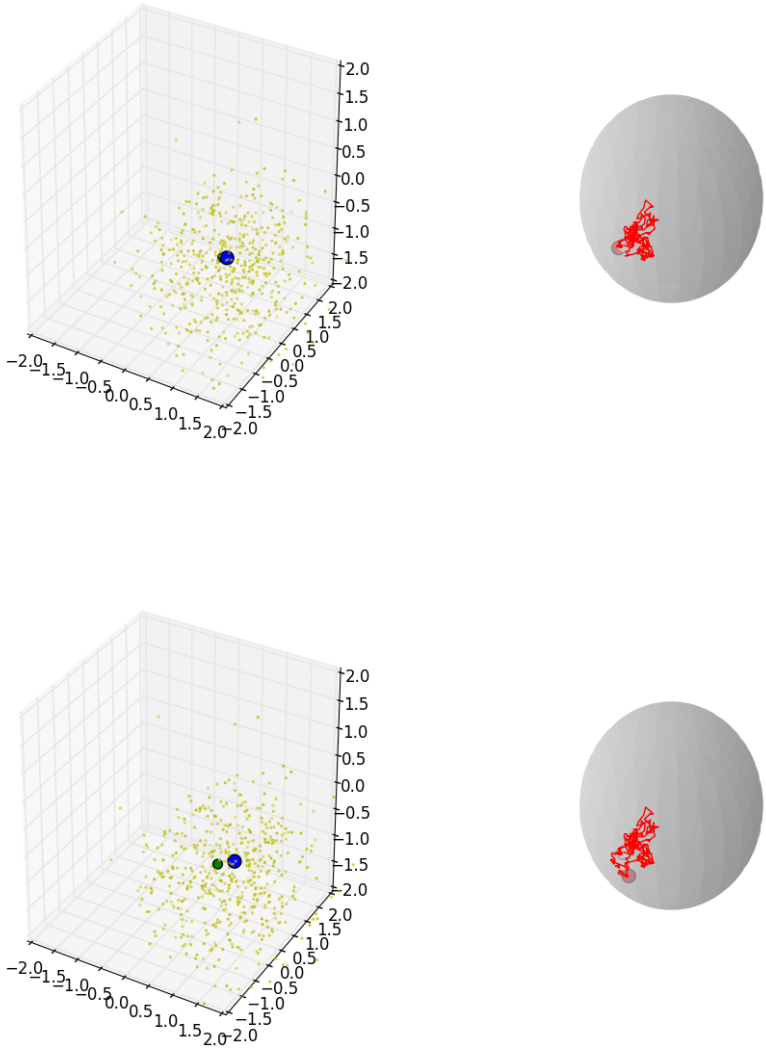


Figure 2.9: Particle filter from Algorithm 1 at different successive times (Top to bottom) for observation on $V_{3,1} = S^2$. On the left side of each image, the particles (yellow), the estimation from the particles (green) and the true state to determine (blue). On the right side of each image, the trajectory of P_t (red, with a red circle for the last observation) observed from the beginning (a red circle for the initial position). In this simulation, the state is modeled as a Brownian motion in $\mathbb{R}^3$ with a unit variance. The variance of the observation has been chosen to be $\sigma_w^2 = 1$.

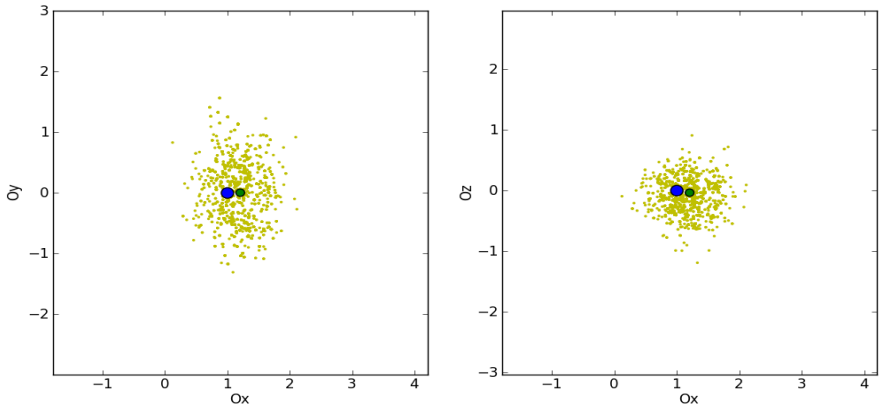


Figure 2.10: Projection of the particles (yellow), the state to estimate (blue) and the estimation (blue) on the plane Ox, Oy and Ox, Oz. At the time of the snapshot, $P_t \approx [0, 1, 0]^T$. Consequently, the component along Oy cannot be estimated. The particles are distributed within an ellipse whose large axis is directed along Oy.

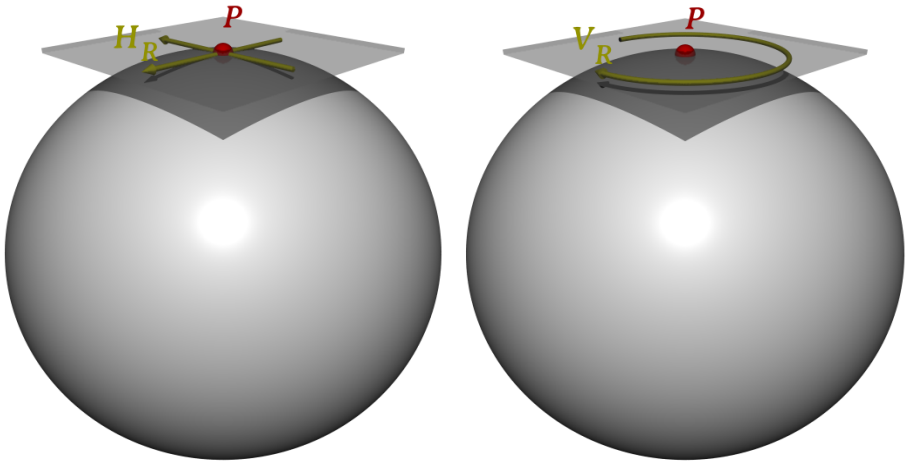


Figure 2.11: For an observation on the Stiefel manifold, only the horizontal component of x_t can be estimated. The action of the vertical component leaves P (red) inchanged. (Left) If the axis of the rotation acting on P is orthogonal to P, then the action is visible, the rotation is in the horizontal space $\mathcal{H}_R$ (yellow). (Right) However, if the axis of the rotation is parallel to P, the action of this rotation is not visible. The rotation is then in the vertical space $\mathcal{V}_R$ (yellow).

Chapter 3

Geometric phase

Contents

3.1 Geometric phase for classical mechanical systems 61
3.1.1 Foucault pendulum 61
3.1.2 Polarized waves and geometric phase 67
3.1.3 Evidence of the geometric phase for elastic waves 70
3.1.4 Conclusion 75
3.2 Multiple scattering of polarized wave 75
3.2.1 Description of the random media 75
3.2.2 Scattering of polarized waves 77
3.2.3 Compound Poisson process 78
3.2.4 Estimation via Expectation-Maximization 81
3.2.5 Conclusion 86

This chapter gathers the results from journal papers [Rossetto 2012] and [Rossetto 2013], respectively in Section 3.1 and Section 3.2. It focuses on the geometric phase acquired by a transverse polarized wave during its propagation.

Section 3.1 defines the notion of geometric phase, initially introduced by Berry in [Berry 1984], using the Foucault pendulum. Geometric phase is a phase acquired by a system laying on S^2 in our case during a closed trajectory. This phase can be expressed only from the geometry of the trajectory of the system. It has been shown [Segert 1987] that transverse polarized waves also acquire a geometric phase on their polarization during an adiabatic propagation. We present in this section an experimental set up showing the existence of the geometric phase and measuring this phase for elastic waves in a waveguide in a non adiabatic regime.

Section 3.2 describes the distribution of the geometric phase when the polarized wave is propagating in a scattering medium. The propagation of the wave can then be modeled as a process on $SO(3)$ with parallel transport. For a multiple scattering regime, the geometric phase can be described with a few parameters like the free mean path (disposition of the scatterers in the medium) or the scattering anisotropy (geometry of one scatterer). This leads to the construction of parametric estimators of these coefficients, constructed from an Expectation-Maximization algorithm, from observation of the geometric phase distribution of the wave.

3.1 Geometric phase for classical mechanical systems

3.1.1 Foucault pendulum

In 1851, Léon Foucault installed a pendulum in the Pantheon in Paris to illustrate the rotation of the Earth without observing the stars or any astronomical object. From the laws of classical mechanics, the direction of oscillation tends to stay constant with respect

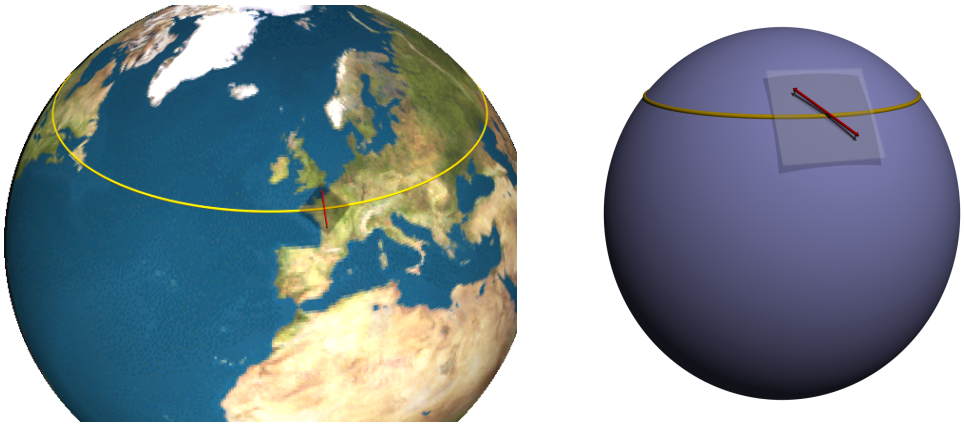


Figure 3.1: Model of the Foucault pendulum. (Left) As Earth is rotating, the position of the pendulum positioned at Paris, France, follows the trajectory γ_t (yellow). The trajectory is a curve with a constant colatitude. The axis of oscillation v_t (red) is, due to gravity, contained in the tangent plan. (Right) Earth is modeled as the sphere S^2. The trajectory γ_t (yellow) is a circle at a constant colatitude. The axis v_t is modeled as an element of $T_{\gamma_t}S^2$.

to the stars. This direction does not remain constant with respect to the room referential as Earth rotates on its axis.

Foucault observed that the pendulum acquired a precession after a complete day of $-271°$ in Paris. After reproducing the same experiment at different places in the world, he deduced an empirical law for the precession acquired by the pendulum after oscillating for a day with respect to the colatitude of place where the pendulum was set up. For a colatitude l, the precession rate is $2\pi(1-\sin l)\,.°/\,\text{day}$. In the extreme case of $l = 0$, *i.e* when the pendulum is set up at the North pole, the precession rate is $360°$ per day. This is due to the axis of oscillation of the pendulum remaining fixed with respect to the stars due to the inertia principle while Earth is making a complete revolution.

Despite that this relation was deduced in an empirical way, the same relation could be obtained by including the Coriolis effect in standard Newton laws of mechanics[1] to describe the precession of the pendulum. However, this phase can also be explained with geometric concepts, *i.e* without using mechanical laws. The model used is the following: Let call v_t the direction of oscillation, where t represents time as this axis is not constant as the surface of the Earth. Due to gravity, v_t is contained in the horizontal plan. Let call γ_t the position of the pendulum at the time t. The previous sentence reads $v_t \in T_{\gamma_t}S^2$, with the sphere S^2 modeling the surface of the Earth and $T_{\gamma_t}S^2$ denoting the plane tangent to S^2 at the point γ_t. In our case, the tangent plane to γ_t is the set of all the vectors orthogonal to γ_t. Figure 3.1 gives a representation of this model.

As $T_{\gamma_t}S^2$ is not constant when γ_t moves on S^2, we want to determine a rule for the evolution of v_t. The inertia principle imposes v_t to remain constant. However, the gravity imposes that $v_t \in T_{\gamma_t}S^2$. Considering the period of oscillation of the pendulum small with respect to the evolution of v_t, the best we can do to respect the inertia principle under the constraint $v_t \in T_{\gamma_t}S^2$ is to consider that for δt small,

$$v_{t+\delta t} = P_{\gamma_{t+\delta t}} v_t$$

with $P_{\gamma_{t+\delta t}}$ the orthogonal projector from $\mathbb{R}^3$ into $T_{\gamma_{t+\delta t}}S^2$. The condition of rapid oscilla-

[1] The works of Coriolis were unknown to Léon Foucault at the time he settled the experiment in Paris

tions with respect to the evolution of v_t is called the adiabatic condition and the previous result can be proven from the adiabatic theorem[2] [Faure 2013].

As the projection is linear and considering the trajectory γ_t continuous, as δt shrinks to 0, one gets

$$P_{\gamma_t}\frac{dv_t}{dt} = \nabla_{\dot{\gamma}_t} v_t = 0,$$

where ∇ is the covariant derivative on S^2. The vector v_t is said to be parallel transported along γ. See Figure 1.1 from Chapter 1 to illustrate the notion of covariant derivative.

Proposition 17. *The parallel transport conserves the inner product. In other words, if v_t and v'_t are parallel transported along γ_t, the inner product $< v_t, v'_t >$ is kept constant.*

Proof. To show that $< v_t, v'_t >$ is constant, it is enough to show that

$$\frac{d}{dt} < v_t, v'_t >= 0$$

As ∇ is the covariant derivative, the previous left term can be decomposed

$$\begin{aligned}\frac{d}{dt} < v_t, v'_t > &=< \nabla_{\dot{\gamma}_t} v_t, v'_t > + < v_t, \nabla_{\dot{\gamma}_t} v'_t > \\ &=< 0, v'_t > + < v_t, 0 > \qquad \text{As } v \text{ and } v' \text{ are parallel transported.} \\ &= 0 + 0.\end{aligned}$$

Therefore, the parallel transport conserves the inner product. □

By extension of Proposition 17, the parallel transport also conserves the norms of a vector. Consequently, a parallel transported vector will not vanish and an orthonormal basis parallel transported along γ_t will remain an orthonormal basis of $T_{\gamma_t}S^2$. Figure 3.2 shows examples of vectors being parallel transported along a portion a great circle on S^2. These vectors have different initial orientations. And the angle between two vectors parallel transported remains constant.

Consider now the case where γ is a closed path on S^2, *i.e*, $\gamma_0 = \gamma_1$ and let v_0 a unitary vector from $T_{\gamma_0}S^2$. Let v_t be the parallel transport version of v_0 along γ_t. As $\gamma_0 = \gamma_1$, the tangent planes $T_{\gamma_0}S^2$ and $T_{\gamma_1}S^2$ are equal and it makes sense to compare v_0 and v_1 as they belong to the same tangent plane.

Let β_γ be the oriented angle between v_0 and v_1

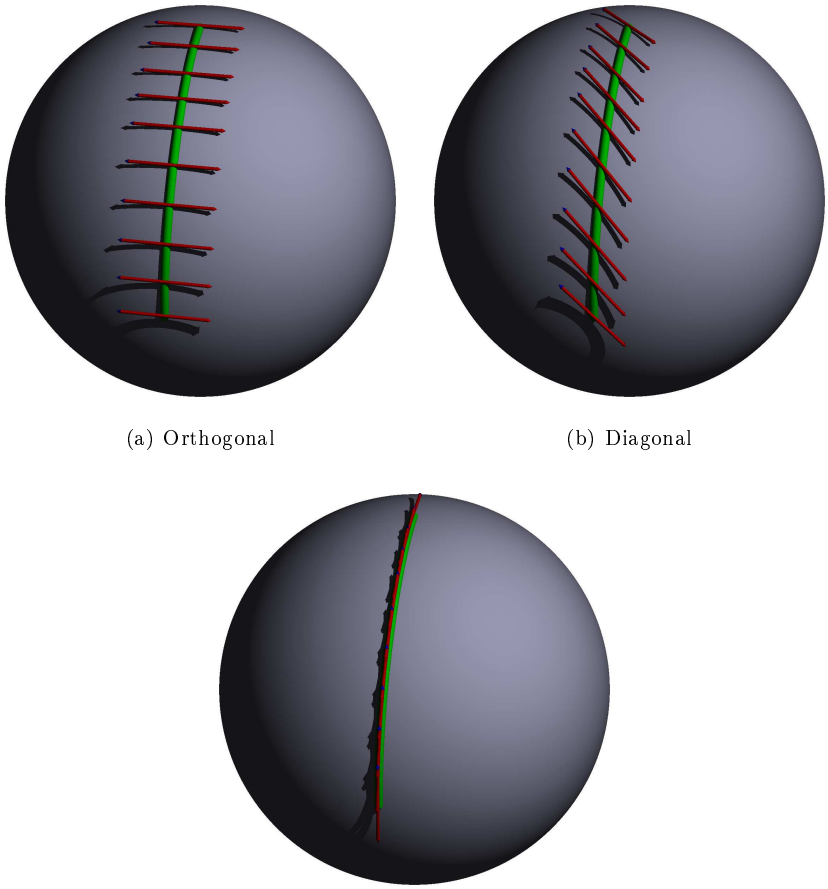
$$\beta_\gamma = \widehat{v_0, v_1}. \tag{3.1}$$

This angle is called the *geometric phase* of γ. First of all, it is noticeable that this angle is well defined as $v_t \neq 0$ from Proposition 17. Also from Proposition 17, the conservation of the inner product makes this definition independent from the choice of the initial vector v_0. For the Foucault pendulum, the independence with respect to the initial vector reads as the independence of the precession with respect to the initial direction of oscillation.

Proposition 18. *The geometric phase β_γ is independent of the parametrization of γ. In other words, for a function $\phi : \mathcal{C}^1([0,1] \to [0,1])$ such that $\phi_0 = 0$ and $\phi_1 = 1$,*

$$\beta_\gamma = \beta_{\gamma\circ\phi}$$

[2] For a system with Hamiltonian slowly evolving with time, this system will remain in a state associated to the same eigenvalues [Berry 1984]



(a) Orthogonal

(b) Diagonal

(c) Parallel

Figure 3.2: Example of vectors (red) parallel transported along a portion of a great circle (green) for different initial orientations. (3.2(a)) The initial orientation is orthogonal with respect to the direction of the transport. The vectors remain parallel during the transport. (3.2(b)) The initial orientation makes a $\pi/4$ rad angle with respect to the direction of the transport. (3.2(c)) The initial orientation is the same that the direction of the transport. In each case, it seems that the initial orientation is conserved during the transport.

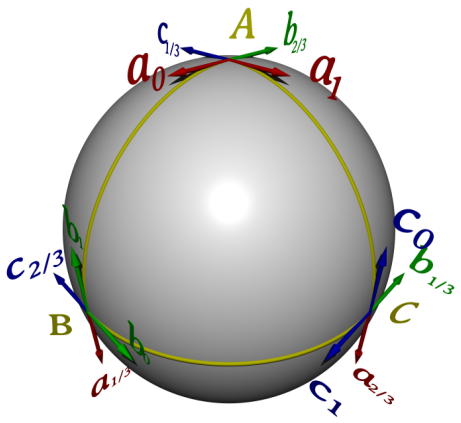


Figure 3.3: The closed path is a triangle ABC (yellow). The vectors a_0 (red), b_0 (green) and c_0 (blue) tangent to each geodesics are anti-clockwise parallel transported along the triangle. The angles $\widehat{a_0, a_1}$, $\widehat{b_0, b_1}$ and $\widehat{c_0, c_1}$ are equal to the enclosed area.

Proof. Let $v_0 \in T_{\gamma_0} S^2$ and let define v_t the parallel transport version of v_0 along γ. To show the independence of the geometric phase with respect to the parametrization, let show that v_{ϕ_t} is parallel transported along $\gamma \circ \phi$.

$$\begin{aligned} P_{\gamma_{\phi_t}} \frac{dv_{\phi_t}}{dt} &= P_{\gamma_{\phi_t}} \frac{dv_{\phi_t}}{d\phi_t} \frac{d\phi_t}{dt} \\ &= P_{\gamma_u} \frac{dv_u}{du} \frac{d\phi_t}{dt} \qquad \text{with } u = \phi_t. \\ &= 0 \frac{d\phi_t}{dt} \qquad \text{As } v_t \text{ is parallel transported along } \gamma. \\ &= 0. \end{aligned}$$

Therefore v_{ϕ_t} is also parallel transported along $\gamma \circ \phi$. □

From a different point of view, Proposition 18 can be interpreted as stating that the geometric phase β_γ only depends of the geometry of the path γ, or the image of the trajectory γ on S^2. However, determining β_γ via its definition requires to determine the parallel transport of a vector along a trajectory. This operation might be rather complex as it requires to solve a differential equation.

Theorem 5. *Local Gauss-Bonnet on S^2: Given a closed path γ on S^2, the geometric phase β_γ is given by the integral*

$$\beta_\gamma = \iint_{\mathcal{C}} ds, \tag{3.2}$$

where $\mathcal{C}$ is the surface enclosed by γ, i.e $\partial\mathcal{C} = \gamma$.

Proof. This proof is inspired by the notes in [Faure 2013]. For a more rigorous proof in the case of a trajectory laying on a generic surface, check [Halper 2008]. Here, the demonstration is realised in two parts. First, the special case when γ is a triangle composed of arc of great circles. Then, Equation (3.2) is proved for a more general trajectory by approximating this trajectory by arc of great circles and applying the first case.

First of all, let γ be a triangle ABC composed of arcs or great circles. From Proposition 18, γ can be parametrized such that $A = \gamma_0$, $B = \gamma_{1/3}$, $C = \gamma_{2/3}$ (see Figure 3.3 for a representation of the different elements). Let $a_0 \in T_{\gamma_0}S^2$ (respectively b_0 and c_0) a unit vector tangent to the first arc of geodesic (respectively second and third arc). Let a_t (respectively b_t and c_t) be the parallel transport version of a_0 (respectively b_0 and c_0) along γ.

By definition of the geometric phase

$$\begin{aligned}\beta_\gamma &= \widehat{a_0, a_1} \\ &= \widehat{a_0, c_{1/3}} + c_{1/3}\widehat{, b_{2/3}} + b_{2/3}\widehat{, a_1}\end{aligned}$$

As parallel transport preserves angles

$$\begin{aligned}\beta_\gamma &= \widehat{a_0, c_{1/3}} + \widehat{c_0, b_{1/3}} + \widehat{b_0, a_{1/3}} \\ &= (\pi - A) + (\pi - B) + (\pi - C) \\ &= \pi - (A + B + C)\end{aligned}$$

For a spherical triangle ABC, the last line is equal to its enclosed area. In other words

$$\beta_\gamma = \iint_{\mathcal{C}} ds.$$

Let now consider a more general case for the trajectory γ. The first part of the proof allows us to consider trajectories that are made of arcs of great circles. In this case, the enclosed area can be decomposed as the sum of triangles and the first part of the proof can be applied to each triangle. The theorem then holds for this special kind of trajectory.

Let γ_n be a sequence of geodesic polygons approximating γ. The theorem holds for each trajectory γ_n. By continuity of the integral,

$$\begin{aligned}\beta_\gamma &= \lim_n \beta_{\gamma_n} \\ &= \lim_n \iint_{\mathcal{C}_n} ds \text{ where } \partial\mathcal{C}_n = \gamma_n \\ &= \iint_{\mathcal{C}} ds.\end{aligned}$$

Therefore, Equation (3.2) holds for any kind of continuous closed curve γ. □

Theorem 5 gives a simple way to determine the geometric phase β_γ for a closed trajectory γ on S^2[3]. It is noticeable that the term "enclosed area" can be ambiguous, depending of the trajectory γ. However, as the complete area of S^2 is 4π and the geometric phase is definite modulo 2π, the enclosed area or the complementary area are just algebraically opposite. As the orientation of the parametrization of γ changes when considering the complementary area, the theorem still holds.

From the considered model for the Foucault pendulum placed at colatitude l, the axis of oscillation is parallel transported during Earth rotation. After a complete revolution, the trajectory γ^l of the position of the pendulum is a closed path and the geometric phase β_{γ^l} represents the angle between the axis of oscillation v_0 and v_1 after a complete revolution. In other words, it represents the precession rate of the Foucault pendulum.

From Theorem 5, this geometric phase can easily be determined as the enclosed area of

[3]For a trajectory on a generic surface, the theorem reads $\beta_\gamma = \iint_{\mathcal{C}} k ds$ with k the Gauss curvature of the support of γ ($k = 1$ for S^2).

the cap delimited by the circle at colatitude l (see Figure 3.1)

$$\beta_{\gamma^l} = 2\pi\,(1 - \sin l)$$

This expression corresponds to the precession rate determined from the classical laws of mechanics with the Coriolis effect. However, this last expression is obtained from geometric considerations.

Despite that the geometric phase was here introduced via an example on S^2 for the Foucault pendulum, the notion of geometric phase still exists for other systems and for trajectories with a support different from S^2. This is a property existing for all processes described in tangent bundle with a parallel transport constraint. For example, the orientation of a falling cat or a free diver can be modeled with a geometric phase. By bending and modifying the shape of its body, a falling cat can modify its general orientation, modeled as a geometric phase. This phase can then be expressed from the squirming of the cat [Montgomery 1993]. Geometric phase also exists for polarized light. In [Segert 1987], Segert shows that transverse polarized waves are subject to acquire a geometric phase. Next subsection describes the model used for polarized waves and the interpretation of the geometric phase in this context.

3.1.2 Polarized waves and geometric phase

The electric field of an electromagnetic wave is transverse to the direction of propagation of the wave. The polarization of the wave is the phase between each orthogonal component of the electric field [Brosseau 1998]. Maxwell laws state that the evolution of the electric field can be represented on an ellipse. Depending on the polarization, this ellipse can be degenerated into a segment. The polarization is then said linear. In the other case, the ellipse can also be degenerated into a circle and the polarization is said circular. Figure 3.4 illustrates these different cases.

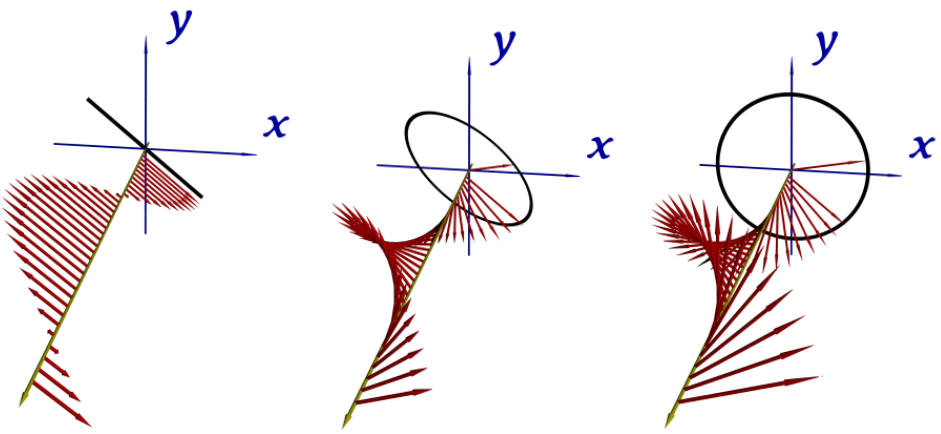
The transverse polarization is however not limited to electromagnetic waves. For example, in elasticity theory, S shear waves induce a displacement in a direction orthogonal to the direction of propagation. The polarization of these waves can then be defined as the displacement induced by the wave (the displacement plays an equivalent of the electric field for electromagnetic waves).

In the next parts of the chapter, a linearly transversed polarized wave is considered. This study can be extended to elliptic polarization[4].

Let γ_t be the unit vector indicating the direction of propagation of a polarized wave at a time t. As γ_t is a unit vector, $\gamma_t \in S^2$. The polarization being linear, let v_t be a unit vector representing the direction of the polarization. As the polarization is transverse, v_t is orthogonal to γ_t. For the sphere S^2, the tangent space $T_{\gamma_t}S^2$ is the set of all the orthogonal vectors to γ_t (see Figure 3.5 for an illustration). This means that $v_t \in T_{\gamma_t}S^2$. In the following, the term polarization will directly refer to the vector v_t.

If the wave is propagating in a constant direction, then γ_t and v_t are constant. However, if the direction of propagation γ_t is changed (waveguide, scattering, etc), the polarization v_t also changes as $v_t \in T_{\gamma_t}S^2$. We suppose here that the state of polarization will remain linear and can still be modeled by a vector v_t. This is the case for isotropic medium for example [Brosseau 1998]. In this section, we are looking at the evolution of v_t when γ_t, and therefore T_{γ_t}, S^2 varies with time.

[4]The geometric phase would also be defined for elliptic polarization. As parallel transport preserves angles and lengths, the great and small axis of the ellipse would remain orthogonal with the same lengths during the propagation. The geometric phase is then defined by working independently on one of these axis.



(a) Linear Polarization (b) Elliptic Polarization (c) Circular Polarization

Figure 3.4: Representation of different states of polarization for an electromagnetic wave or a transverse bulk elastic wave. The direction of propagation is represented in yellow. The transverse electric field is represented in red and the trace of the evolution at different time is depicted in black, in the xy plane. (a) The polarization is linear, the projection of the electric field in the plane orthogonal to the direction of propagation is contained in a segment. For the rest of the study, only linearly polarized waves are considered. (b) General case of a polarization, the electric field describes an ellipse. (c) Circular polarization, the electric field describes a circle.

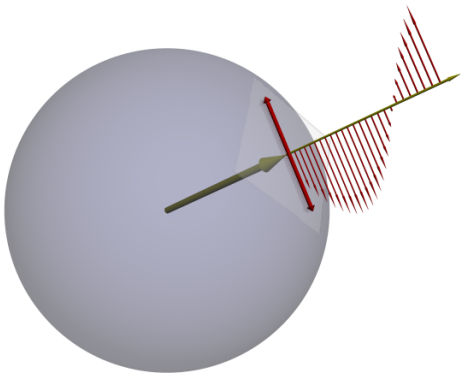


Figure 3.5: Model used for linear polarized wave. The direction of propagation γ_t (big yellow arrow) is a unit vector, *i.e* an element of S^2 (grey sphere) as the polarization v_t (big red double arrows) needs to remain orthogonal to γ_t. The temporal evolution of the polarization would be given by the small red arrows along the thin yellow arrow. Thus, it can be modeled as $v \in T_\gamma S^2$. These vectors are enough to describe the state of the polarized wave. Elliptic polarization can also be modeled this way by considering both axis of the ellipse.

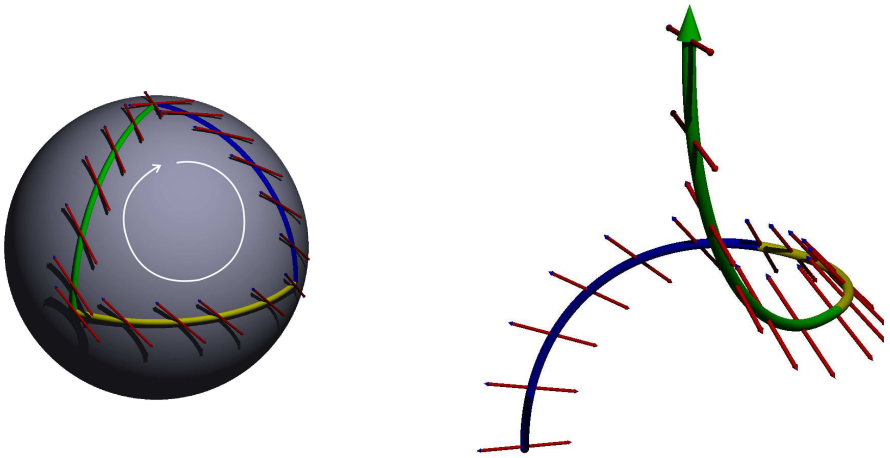


Figure 3.6: (Left) Trajectory of the direction of propagation γ_t defined by 3 portion of great circles defining a triangle. The area enclosed by the triangle $-\pi/2$. The minus sign comes from the clockwise orientation. The polarization v_t (red) is parallel transported along γ. The angle between the initial and final polarization is $-\pi/2$, corresponding to the algebraic enclosed area. (Right) The real trajectory of the wave in 3D space. The fact that γ is a closed curve in S^2 indicates that the initial and final direction of propagation are the same.

In [Segert 1987], Segert shows that if the propagation path of the wave has a small curvature with respect to the wavelength, then the evolution of v_t can be completely described from the geometry of the path. Moreover, v_t is parallel transported on the surface of S^2 following γ_t (see the paragraph 3.1.1 for the definition of the parallel transport on S^2). For electromagnetic light in an optical fiber for example, this condition is always realized as the wavelength is several magnitudes smaller than characteristics of the optical fiber. Under such conditions, the propagation is said to be adiabatic.

If γ_t for $t \in [0,1]$ is a closed path, *i.e* the direction of propagation is the same at $t = 0$ and $t = 1$, then v_t will acquire a geometric phase. This phase, from Proposition 18 only depends of the geometry of the path γ (see Figure 3.6). Thanks to Theorem 5, this phase can be computed from the image of γ_t.

This result still holds for elliptic polarization. As the parallel transport converses the inner product, two orthogonal vectors parallel transported would remain orthogonal. The polarization would remain elliptic, but with a rotation of the major and minus axis of the ellipse by an angle determined by the geometric phase.

In 1986, the existence of the geometric phase for electromagnetic waves (light) was experimentally proved [Chiao 1986]. An optical fiber was used in the experimental set up as a waveguide for the propagation of a linearly polarized wave. By controlling the shape of the optical fiber, the path γ could also be controlled. Therefore, from Theorem 5, the geometric phase β_γ could be controlled too. In the experimentation, the optical fiber followed an helix.

The geometry of the helix is given by the following parameters. Let r be the radius of the helix and P its pitch. The length of a coil is given by $L = \sqrt{(2\pi r)^2 + P^2}$. Let θ be the angle between the direction of propagation γ_t and the main axis of the helix. Thus $\sin\theta = P/L$. As the waveguide is an helix, the angle θ is constant and by considering the main axis vertical, γ_t remains at a constant colatitude on S^2. If $\theta = 0$, the helix degenerates into a straight line. If $\theta = \pi/2$, the helix degenerates into a flat circle (see Figure 3.7).

From Theorem 5, the geometric phase is

$$\beta_\gamma = 2\pi \cos\theta.$$

This is the geometric phase acquired by the polarized wave after propagating along one coil. After a propagation through 3 or 4 coils, the geometric phase will be $3\beta_\gamma$ or $4\beta_\gamma$. However, this result only holds for closed paths and the Gauss-Bonnet theorem needs to be extended to describe the geometric phase for non closed paths, *i.e* to give the orientation of the polarization at each position along the helix.

The Fuller theorem [Fuller 1978] states that if the trajectory γ (in our case, a circle at a constant colatitude) can be described as a smooth deformation from a reference trajectory γ_{ref}, then the geometric phase β_γ can be described from $\beta_{\gamma_{ref}}$ and the area between the path γ and γ_{ref}. More precisely, we will say that after propagating along an arc length s, the geometric phase for γ is

$$\beta_{\gamma,s} = \beta_{\gamma_{ref},s} + \mathcal{A}(\gamma, \gamma_{ref}, s),$$

where $\mathcal{A}(\gamma, \gamma_{ref}, s)$ is the area spanned by $\gamma_{s'}$ for $s' \leq s$ during the smooth deformation of γ_{ref} into γ (see Figure 3.7). This definition compares the geometric phase acquired due to the path deformation with respect to the reference geometric phase. Thus, it makes sense to consider non closed path γ.

The reference trajectory is the flat circle (or an helix with $\theta = \pi/2$) for which it is known that the geometric phase vanishes because the trajectory is two-dimensional. In this case, $\mathcal{A}(\gamma, \gamma_{ref}, s) = (2\pi s/L)(\cos\theta)$. Therefore, the geometric phase acquired after propagating on a distance s is

$$\beta(s) = (2\pi s/L)(\cos\theta) = \tau s, \qquad \text{with } \tau = (2\pi\cos\theta)/L \tag{3.3}$$

compared to the phase after propagating on a distance s on a circle. Equation (3.3) also shows the linear dependance of β with the arc length s for a polarized wave propagating along an helix.

In their experiments from [Chiao 1986], Tomita and Chiao showed that electromagnetic waves acquire a geometric phase given by (3.3). However, due to the short wavelength (compared to the size of the optical fiber), they could not get close to the limits when the propagation cannot be considered adiabatic anymore. In Subsection 3.1.3, an experimental setup similar to the one from [Chiao 1986] is used to measure the geometric phase of elastic waves. The physical parameters of the system make the propagation no more adiabatic. However, the presence of a geometric phase is still visible.

3.1.3 Evidence of the geometric phase for elastic waves

This experiment is the mechanical counterpart of the device from [Chiao 1986]. The aim is to be able to observe the geometric phase for elastic waves and verify the existence of the geometric phase in the non adiabatic limit.

A metallic spring is used as a waveguide for the mechanical waves, taken from a car's rear damper. It has a circular section of 13.5 mm making five coils of radius $r = 75 \pm 1$ mm and with a pitch of $P = 91.5 \pm 1$ mm. The direction of propagation makes a constant angle $\theta = 1.379$ rad with the main axis of the helix. As the cross-section of the helix is circular, the flexural modes are degenerated.

As we will be measuring mechanical vibrations, the system needs to be isolated from the vibrations of the floor or the walls of the building. In order to do so, the helix is suspended by two strings to isolate the system. We used two accelerometers (Brüel & Kjaer, type 4518-003) located at one end of the spring and recorded vibrations in two orthogonal directions

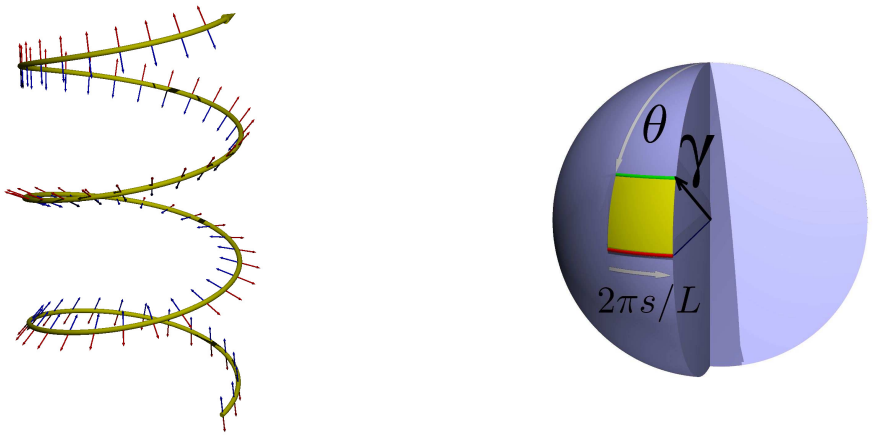


Figure 3.7: (Left)Propagation of a linearly polarized wave along a waveguide (yellow) shaped as an helix. The polarization (red and blue) remains orthogonal to the direction of propagation. It acquires a geometric phase due to the three dimensional path.(Right) Geometry of the transformation from the reference trajectory to the helix trajectory. During the transformation the tangent vector for s fixed follows a meridian (vertical arrow in the figure) between colatitudes $\pi/2$ and θ. A displacement of s along the helix corresponds to an azimuth angle of $2\pi s/L$ (horizontal arrow). The area spanned by the tangent vector during the deformation is equal to the geometric phase difference computed by Fuller's formula.

(see Figure 3.9). The sampling frequency being set to 50 kHz, the information in the signal is thus available up to 25 kHz. Assuming that the accelerometers' power spectrum is stable in the range 1kHz – 25kHz (see Figure 3.10), frequencies above 5 kHz can be ignored. The signal is amplified (Brüel & Kjaer amplifier, type 2694) before being processed.

Measurements are realized at one end of the helix while the source location is moved along the helix with a constant step, see figure 3.9. Because of the temporal reversibility of the waves, displacing the source for fixed sensors does not alter the measure of the geometric phase.

The waves are generated by 32 equally spaced sources, which distance to the accelerometers ranges from $s = 5$ cm to $s = 1.45$ m. Making weak impacts on the metal spring creates bending waves that are linearly polarized. Impacts are made radially with respect to the helix. The source signal is generated manually, by gently hitting the waveguide with a hammer at the different source positions.

The wave propagation modes in helical waveguide constitute a difficult problem. Solutions can only be found numerically [Treyssède 2007] and yield a complex mode structure. The fundamental mode is a flexural mode with linear polarization. Polarization losses appear rapidly after the first waveforms. In order to measure the geometric phase, we use only the first periods of oscillation of the signal, before depolarization has reached a level that does not allow a proper measure of the polarization parameters (orientation of the principal axis).

In figure 3.8, the time windowed signals used for polarization orientation estimation are presented after a hit at a distance $s = 105cm$, together with the complete waveforms and associated polarization parametric plots. In Figure 3.10, we show an example of two

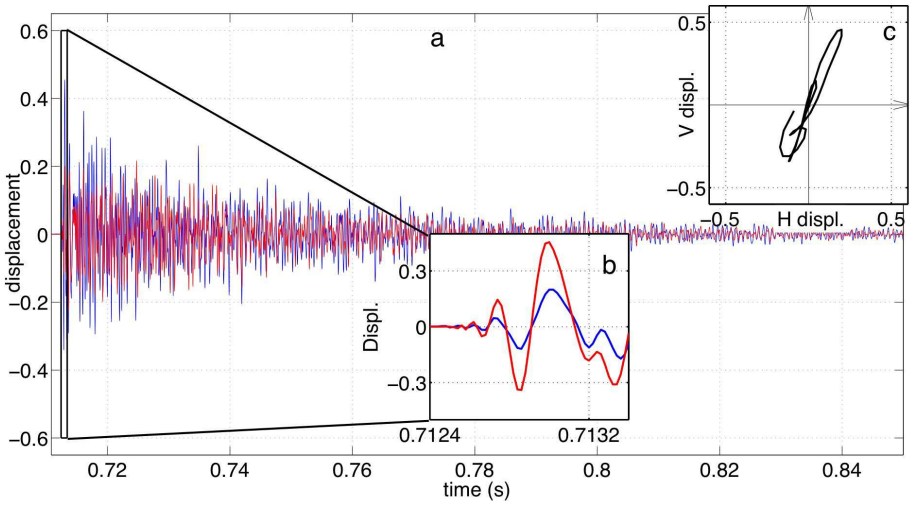


Figure 3.8: Example of vertical (red) and horizontal (blue) displacement signals recorded during the experiment. (a): complete recorded waveforms associated to one source. (b): time windowed signals considered for polarization orientation estimation. (c): associated polarization parametric plot. The truncated (time windowed) signal is chosen in order to isolate direct linearly polarized wave from other modes and reflected waves. Only few first oscillations are considered. Direct waves are the waves that reach the sensors without reflections on the edge of the wave guide

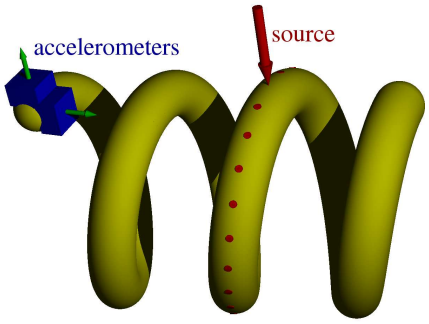


Figure 3.9: Setup of the experiment. Accelerometers are represented as cubes, they record acceleration in two orthogonal directions (small arrows). The large arrow symbolizes the source, a radial impact on the helix. The dots indicate some of the source positions.

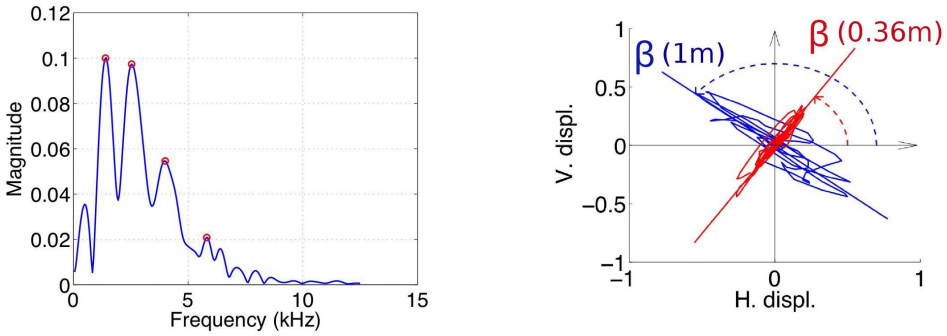


Figure 3.10: (Left) Power spectrum of the first oscillations of the signal corresponding to direct arrival of the flexural mode. Circles indicate the maxima used as central filtering frequencies in the signal analysis.(Right) Parametric plot of the first oscillations (direct arrival) for sources located at $s = 0.36\,\mathrm{m}$ and $s = 1.00\,\mathrm{m}$. The relative angle between polarization orientations is the geometric phase difference.

parametric polarization plots for signals obtained with the source at different distances from the accelerometers. The relative angle between polarization orientations is the geometric phase.

Bending waves in the waveguide are dispersive. We therefore filtered the signal using non-overlapping bandpass filters centered at frequencies: 1.4 kHz, 2.5 kHz, 4.0 kHz and 5.8 kHz. These values correspond to maxima of the power spectrum displayed in figure 3.10 (left). The corresponding velocities are displayed in Table 3.1.

The polarization orientation is obtained from the records using a principal component analysis [Vidale 1986]. This technique consists in obtaining the eigenvectors of the cross-correlation between the two orthogonal signals (recorded on the two accelerometers). The eigenvector with the largest eigenvalue gives the polarization direction. The geometric phase $\beta(s)$ is thus the orientation of polarization measured when the source is located at distance s from the accelerometers (see Figure 3.9).

Several error sources have been taken into account for the measure.

- Accelerometers: The measurement of the direction of polarization relies on the orthogonality of the accelerometers. The inaccuracy of the orthogonality must therefore be considered. Let ε be the discrepancy between the actual angle and $\pi/2$. Namely, the angle between the direction of the measured accelerations is $\pi/2 + \varepsilon$. Furthermore, the coupling between the vibration of the waveguide and the accelerometers cannot be controlled. It is assumed that the coupling of the vertical sensor is $1+\eta$ times the coupling of the horizontal one (mainly due to the method of fixation of the accelerometers on the waveguide, using wax). These errors, once included in the computation of the geometric phase via a principal component analysis give

$$\beta(s) = \beta(0) + \tau s + \sin\beta(s)\big(\eta\cos\beta(s) - \varepsilon\sin\beta(s)\big). \tag{3.4}$$

  The value of ε and η are then determined simultaneously with the rate τ.

- Source: the signal is manually created using a small hammer and hitting radially with respect to the main axis. The position and the direction of hit are therefore submitted to systematic errors. The positional error is estimated around 2mm and the directional error is $\pi/20$rad.

- External sounds: Although the waveguide was isolated as much as possible from the environment by suspending it with strings, ambient sounds can not be completely

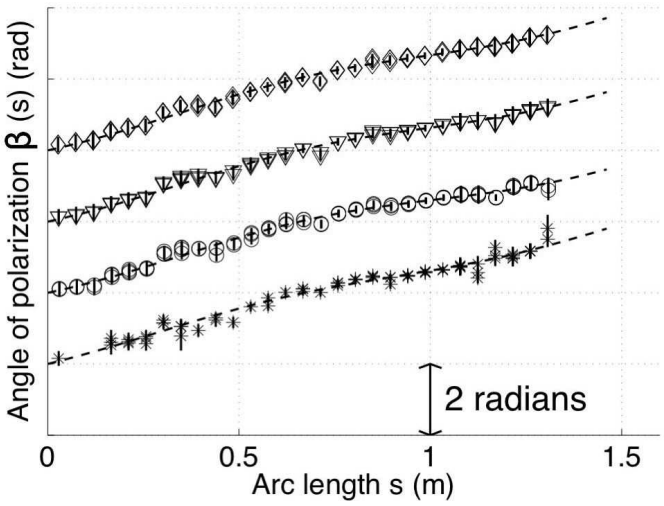


Figure 3.11: Measured geometric phase as a function of arc length along the helix for signals filtered at 1.4 kHz, 2.5 kHz, 4.0 kHz and 5.8 kHz. The dashed line represents the linear regression with correction for the coupling and orthogonality of the accelerometers. The origins of the curves are arbitrary as the initial orientation has no impact on β.

ν(kHz)	c(m.s^{-1})	wavelength (m)	τ (rad.m^{-1})	η	ε (rad)
1.4	$2\,10^3$	1.4	2.53	0.31	-0.03
2.5	$3\,10^3$	1.2	2.35	0.45	0.03
4.0	$4\,10^3$	1.0	2.45	0.39	0.05
5.8	$3.5\,10^3$	0.6	2.50	0.38	0.07

Table 3.1: Frequencies, velocities, wavelengths and fit parameters for the filtered signals. The values of τ have a confidence range of ± 0.1 rad.m^{-1}.

removed from the signal recorded by the accelerometers. Considering these noise as independent, several measures for a same source have been realised and averaged (under the conditions of the central limit theorem, namely considering a white Gaussian noise)

The theoretical value of τ for the helix used in the experiment is $\tau = 2.49 \pm 0.1$ rad.m^{-1}. The fits performed from experimental data give an estimation of $\tau = 2.50 \pm 0.1$ rad.m^{-1}. We do not observe a significant dependency of τ with respect to the frequency or the velocity (see Table 3.1), which justifies the denomination "geometric phase" : the effect we observe is solely due to the geometry of the waveguide.

Apart from bending waves, compressional waves and torsional waves can also propagate in the helix. In a straight rod, these propagation modes are not coupled and bending waves remain polarized after a long time. In helices, the coupling between compression waves and torsional waves increases with curvature and torsion.

Bending waves and torsional waves are therefore partially converted into each other, back and forth, during the propagation. This explains depolarization in our measurements and why we consider only the first oscillations of the signals.

We have filtered the polarized waves at four frequencies of the order of magnitude of

the rate at which the propagation direction evolves $1/T = c/L$. Such evolution rates imply that the regime is not adiabatic. Converted into lengths, this means that the wavelength of the bending wave is of the same order of magnitude as the length of a helix coil.

3.1.4 Conclusion

We have demonstrated the existence of a geometric phase for elastic waves in a waveguide far from the adiabatic regime. This can be considered as an experimental evidence for the universality of the geometric phase in classical one dimensional systems. Because there is no need to invoke the adiabatic approximation to preserve the polarization state, the geometric phase should extend to all frequencies in smoothly shaped waveguides, viewed as one dimensional curved spaces.

In nature, polarized elastic waves, such as bulk S waves (shear waves) are observed under certain conditions, and the concept of geometric phase applies. The degree of polarization and the statistics of the polarization orientation in seismic signals created by a polarized source contain informations concerning the disorder in the propagating medium that can be understood in terms of geometric phases. However, seismic waves trajectories cannot be modeled as a deterministic propagation but rather as the diffusion of waves in a scattering medium [Margerin 2009]. This is the object of the next section.

3.2 Multiple scattering of polarized wave

3.2.1 Description of the random media

If one emits a plane monochromatic wave into a scattering medium, it will on average travel freely along a distance ℓ, called the mean free path, before encountering a scatterer and then change its direction and travel freely to the next scatterer. The multiple scattering events will result in a spreading of the distribution of the direction of propagation with time [de Hulst 1957].

Depending on the strength and density of the scatterers, this spreading will happen at different rates. This rate is characterized by the travel distance along which the memory of the direction of propagation is preserved. We call this distance the *transport mean free path* $\ell^\star$, it is always larger than the mean free path. After traveling the distance $\ell^\star$, the directions of propagation of all the scattered waves are uniformly distributed.

Beyond typically one transport mean free path, the wave energy propagation in the multiple scattering regime is diffuse: it follows the diffusion equation with effective diffusion constant $\mathcal{D}_e = c\ell^\star/3$ [Herbolzheimer 1988] (c is the velocity of the wave). Several multiple scattering regimes exist, depending on the relations between the wavelength λ_w, the diameter a of the scatterers, the mean free path ℓ, the transport mean free path $\ell^\star$ and the depth of the medium L [Ishimaru 1978]. In this section, we will consider the regime where

$$\lambda_w < a \ll \ell \ll \ell^\star \lesssim L. \tag{3.5}$$

The condition $\ell \ll \ell^\star$ states that the wave is preferentially scattered into a direction close to the incoming one. This is the so called "forward scattering" regime. For spherical scatterers, this regime appears under the condition $\lambda_w < a$ [de Hulst 1957, Mishchenko 1991]. An illustration of these length scales is shown in Figure 3.12 (Left).

The transport mean free path $\ell^\star$ is the typical length scale of correlation between waves in the medium, while the mean free path is the length scale of the disorder. Measuring either the mean free path or the transport mean free path is therefore an important issue in experiments because it is a resolution scale for most of the physical quantities measurable

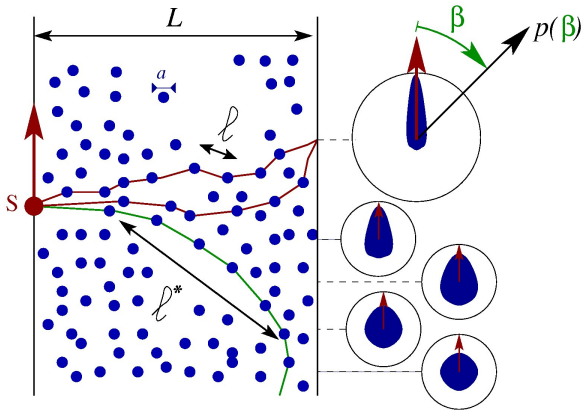


Figure 3.12: (Left) Description of the model: A source (S) emits a linearly polarized plane wave. In the figure, the polarization is represented by the vertical red arrow. The wave is scattered by spherical inclusions of diameter a. The typical distance between these inclusions is ℓ. The memory of the initial (horizontal) direction of propagation is $\ell^\star$. (Right) Probability distribution function $p(\beta)$ of the direction of the polarization for the outgoing wave at different position of the exit surface. The geometric phase β is the angle between the incoming an outgoing polarization. The distribution is more peaked for positions closer to the source, because the average paths are straighter. For a propagation further than $\ell^\star$, the polarization is uniformly distributed (lower right diagram).

using waves. Several techniques estimating the transport mean free path exist, using the outgoing light in various directions. The simplest of them uses the intensity transmission ratio [Lisyansky 1992] to measure $\ell^\star$ in the diffusion approximation. Diffusion models are also useful in anisotropic media [Lagendijk 2008] and to describe intensity leaks in directions orthogonal to the incoming source [López 2011].

In backscattering setup, the opening angle of the weak localization cone, in which an intensity enhancement due to interferences is observed, is related to the transport mean free path in the medium [van Albada & Ad Lagendijk 1985]. All these experiments rely on intensity measurements. An amplitude independent method has been proposed to measure the mean free path ℓ from the phase statistics of a Gaussian field [van Tiggelen & Ludovic Margerin 2009].

It is known that in the forward scattering regime polarization is parallel transported during the propagation, which result in the existence of a geometric phase [Rossetto 2001]. Therefore, each ray possesses a geometric phase depending on the geometry of its path in the medium. At a given point, the observed phase is not uniquely defined, but has a distribution related to the distribution of paths from the source due to the contribution of several rays. In Figure 3.12 (Right), examples of outgoing polarization probability density functions are shown, depending on the position of the exit wave.

The distribution of geometric phase from multiple scattering of a polarized incident wave was adressed ten years ago [Sinha 2000] without any condition on the final direction. The calculation was based on Brownian motion at the surface of a sphere and followed an approach developed earlier in *et al.* [Ouvry 1991] and using Theorem 5. However, multiple scattering is rigorously a Brownian motion on a sphere only in the limit of very weak and dense scatterers. A more general model would be a random walk with finite steps.

In media with a thickness L of the order of $\ell^\star$, the outgoing directions of propagation are not evenly distributed. The joint distribution of direction and polarization state, related to paths statistics, contains informations concerning the scattering events that are not yet

randomized. The number of scattering events is of the order of L/ℓ and in the case where this number is not very large ($L/\ell \simeq 10$) the usual approximations are not suited. For this reason, attention was recently brought to stochastic processes for the description of a scattering system where the number of scattering events is small.

Stochastic models have been introduced fifteen years ago in multiple scattering [Sandison 1995] to solve direct problems. Recently, It has been showed in [Margerin 2009] that a stochastic model for the direction of propagation can be used to solve inverse problems. They computed the mean free path in a medium from the distribution of outgoing directions. We extend this model to take polarization into account and estimate the transport mean free path $\ell^\star$ from the geometric phase distribution.

3.2.2 Scattering of polarized waves

In this paragraph, we consider scattering of polarized waves and we want to determine the evolution of the polarization during scattering.

Adopting a geometrical point of view, polarization can be seen as a particular direction of the electric field. This particular direction evolves along a trajectory. The geometric phase appears in comparing the particular directions of different trajectories. Therefore, without loss of generality, we consider linearly polarized waves for which the particular direction is the polarization itself. This easily extends to elliptical polarized waves, for which the great axis orientation takes the role of the linear polarization. For circular waves, as they have circular symmetry, the measure of the geometric phase must be performed using interferences.

The convention used to describe linearly polarized wave is the same as the one in the previous section: the direction of propagation is denoted $\gamma \in S^2$ and the direction of polarization $v \in T_\gamma S^2$. We consider only media with isotropic uniform absorption, the wave amplitude is therefore completely determined for all paths and does not play any relevant role in our geometric description of the parallel transport.

Let denote F the frame $F = [\gamma, v, \gamma \wedge v]$. By definition of γ and v, $F \in SO(3)$ (orthogonal unitary columns, positively oriented). The wave state along a ray is described by the frame F. Between scattering events, F remains constant. The changes of F occur at scattering events. As F is a frame, any change, of F is realized via a rotation matrix. The scattering angle θ is random, following a probability distribution function called *scattering phase function* $\Phi'(\theta)$ that depends on a/λ_w and on the nature of the scatterer [Greenstein 1941].

The average value of $\langle \cos\theta \rangle = \int \Phi'(\theta) \cos\theta \sin\theta d\theta$ is called the scattering anisotropy and is noted g. Note that we have the relation $\ell^\star = \ell/(1-g)$. In the forward scattering regime, the main scattering angle θ is small and g is close to 1, therefore $\ell^\star \gg \ell$ as mentioned in the Introduction.

The ZYZ convention for the Euler angles describing the rotations of $SO(3)$ is used and angles are denoted $(\psi, \theta, \varphi) \in]-\pi, \pi] \times]0, \pi] \times]-\pi, \pi]$ (see Paragraph 1.1.3).

The rotation matrix R acting on the frame F at a scattering event is fully determined by the incoming direction γ and the outgoing direction γ' and the requirement that the vector v is parallel transported [Rossetto 2001]. By definition of parallel transport on S^2, the angle between γ and the geodesic remains constant. As a consequence the Euler angles ψ and φ of the rotation matrix R are related by (see figure 3.13)

$$\psi = -\varphi. \tag{3.6}$$

Parallel transport of polarization through a scattering event may therefore be described by a rotation matrix $R(\theta, \psi)$, θ is the length of the geodesic on the unit sphere and ψ is the angle between this geodesic and the polarization vector v. The third Euler angle φ does not

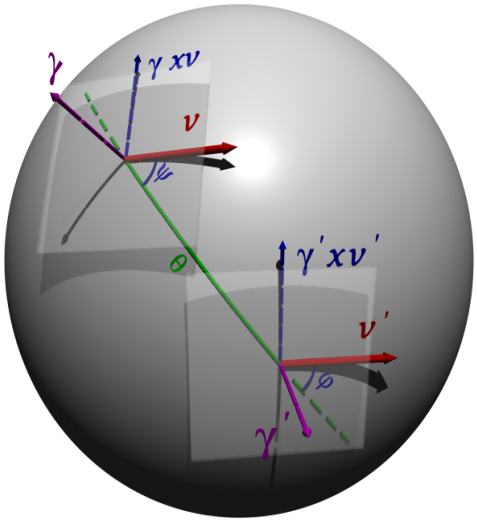


Figure 3.13: Parallel transport of a vector along the geodesic of length θ (solid line). The vector v is transported from the tangent plane $T_\gamma S^2$ to the tangent plane $T_{\gamma'} S^2$. The parallel transport constraint imposes that in these planes, the angle between v and the displacement along a geodesic on the sphere S^2 remains constant. As a result the Euler angles ψ and φ are equal (algebraically opposite).

need to be specified as $\varphi = -\psi$. The relation between the incoming and outgoing frames is

$$F' = F\, R\,(\theta,\, \psi)\,. \tag{3.7}$$

Note that the rotation matrix R acts on the right of F. An expression in terms of "left" action of rotation matrices can be obtained, but would lead to a random process over $SO(3)$ with dependent increments. The independence of increments will be useful for the repeated action (3.7) on F that we present in the next section.

3.2.3 Compound Poisson process

We consider a plane wave with linear polarization, corresponding to the frame F_0 as described in the previous section. The scattering events are described as random rotations acting on the right of F_0. The frame F_0 changes according to (3.7) for each scattering event, so that it becomes the frame F_n after n scattering events (see Figure 3.14):

$$F_n = F_0 R\,(\theta_1,\, \psi_1)\, R\,(\theta_2,\, \psi_2) \cdots R\,(\theta_n,\, \psi_n)\,. \tag{3.8}$$

The scattering events are independent and the scatterers are identical, the random rotations $R\,(\theta_m,\, \psi_m)$ are therefore independent and identically distributed. Let Φ be the distribution of $R\,(\theta_m,\, \psi_m)$. The phase function Φ' is the marginal of Φ with respect to ψ.

From the source to the observer, a wave may encounter a variable number of scatterers. The number of scattering events in a time t can be taken as a Poisson process $N(t)$ [Margerin 2009]. The Poisson parameter η is the average rate at which a wave encounters a scatterer. From the definition of the mean free path ℓ, we have $\eta = c/\ell$: η is the inverse of the mean free time.

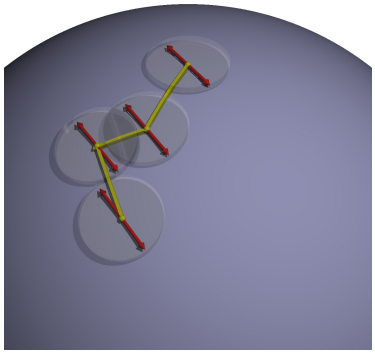


Figure 3.14: Parallel transport of the polarization (red vector) during multiple scattering. Each scattering event modifies the direction of propagation γ and therefore the attached polarization. Between each jump, the polarization converses a constant angle with respect to its trajectory. As multiple jumps will occur, a geometric phase will be accumulated by the polarization.

From these considerations, we obtain the process F_t describing the distribution of polarization after a time t as

$$F_t = F_0 \prod_{k=1}^{N(t)} R\left(\theta_k,\, \psi_k\right), \tag{3.9}$$

where $\prod$ denotes the right-sided product on $SO(3)$. This equation expresses the stochastic process F_t as a compound Poisson process (CPP) on $SO(3)$ with parallel transport. We denote by p_ν the distribution of the process F_t at a given time t, where $\nu = \eta t$ is the average number of scattering events.

A Poisson process is a pure jump process with independent increments and we have

$$p_\nu = \sum_{n=0}^{\infty} e^{-\nu} \frac{\nu^n}{n!}\, p_{F_n}. \tag{3.10}$$

where $e^{-\nu}\frac{\nu^n}{n!}$ is the probability to have exactly n scattering events in the time interval $[0, t]$.

Thanks to (3.8) and the independence of the parallel transport rotations, p_{F_n} is given by [Kyatkin 2001]:

$$p_{F_n} = p_{F_0} * p_{R_1} * \cdots * p_{R_n} = \Phi^{*[n]}. \tag{3.11}$$

Here the symbol $*$ is the convolution product on $SO(3)$ and $\Phi^{*[n]}$ denotes the result of the convolution of n identical functions Φ. We have used the initial condition $p_{F_0}(F) = \delta\left(I_3 - F\right)$. In the case where the source is a distribution of directions of propagation and polarization, the distribution p_{F_n} is the convolution of the initial distribution p_{F_0} with $\Phi^{*[n]}$.

We use harmonic analysis introduced in Section 1.4 on $SO(3)$ to transform the expression (3.11). The irreducible representations λ can be indexed on $\mathbb{N}$ with a dimension $d_\lambda = 2\lambda+1$. The functions U^λ are then $(2\lambda+1)\times(2\lambda+1)$ matrices called Wigner-D matrices. For a rotation matrix $R \in SO(3)$ parametrized with the Euler angles, U^λ can be decomposed as

$$U^\lambda_{m,n}(\psi, \theta, \varphi) = e^{-im\psi} d^\lambda_{m,n}(\theta) e^{-in\varphi} \tag{3.12}$$

with $-l \leq m, n \leq l$ and where $d^\lambda_{m,n}$ is called small Wigner-d functions and can be expressed form Jacobi polynomials.

The next results are reminders from Section 1.4, used here to describe p_ν. For any distribution $p \in L^2(SO(3), \mathbb{R})$, the Fourier coefficients are matrices of the same size and are defined by

$$\hat{p}^\lambda_{m,n} = < p, U^\lambda_{m,n} >,$$

where $<,>$ represents the scalar product for function on $SO(3)$ [Dieudonné 1980].

The Peter-Weyl theorem (Theorem 2) states that any distribution $p \in L^2(SO(3), \mathbb{R})$ can de decomposed as the uniform limit of the series

$$p(R) = \frac{1}{2\pi} \sum_{\lambda \in \mathbb{N}} (2\lambda + 1) \operatorname{trace}\left(\hat{p}^\lambda U^\lambda(R)\right).$$

From Proposition 10, it is known that Fourier coefficients of a convolution are the products of the individual Fourier coefficients

$$\widehat{p_1 * p_2}^\lambda = \hat{p}^\lambda_1 \hat{p}^\lambda_2.$$

We deduce the Fourier coefficients of p_ν from (3.10)

$$\hat{p}^\lambda_\nu = \sum_{n=0}^{\infty} e^{-\nu} \frac{\nu^n}{n!} \left(\widehat{\Phi}^\lambda\right)^n = \exp\left[\nu(\widehat{\Phi}^\lambda - I_{2\lambda+1})\right], \tag{3.13}$$

where exp denotes the matrix exponential.

Using the notation $\Delta(x) = 2\pi \sum_k \delta(x - 2\pi k)$, the parallel transport constraint and the symmetry of the scatterers leads to $\Phi(\phi, \theta, \varphi) = \Delta(\psi + \varphi)\Phi(\theta)$. The Fourier coefficients $\hat{\Phi}^\lambda$ are obtained by the integration formula

$$\hat{\Phi}^\lambda_{m,n} = \frac{1}{8\pi^2} \int_{\theta,\varphi,psi} \Phi(\theta)\Delta(\psi + \varphi) U^\lambda_{m,n}(\psi,\, \theta,\, \varphi) \sin\theta d\theta\, d\varphi\, d\psi. \tag{3.14}$$

Using the definition of the Wigner function (3.12), we obtain the simpler expression

$$\hat{\Phi}^\lambda_{m,n} = \frac{\delta_{mn}}{2} \int_0^\pi \Phi(\theta) d^j_{m,m}(\theta) \sin\theta d\theta, \tag{3.15}$$

showing that the Fourier matrices $\hat{\Phi}^\lambda$ are diagonal for all $\lambda \in \mathbb{N}$. Thanks to (3.13), it is clear that $\hat{p}^\lambda_\nu$ is also diagonal at all orders λ.

As p_ν is a distribution on $SO(3)$, it can be parametrized by the Euler angles ψ, θ, φ and will represent the position of the polarization after a propagation for a time $t = \nu/\eta$ in the scattering medium (see Figure 3.13). As θ represents the latitude of the direction of propagation, the term $\beta = \phi + \varphi$ is the difference of angle between the initial and final polarization (once parallel transported at the origin). Therefore, β is the geometric phase of the polarized wave. By integrating $\Delta(\psi + \varphi)p_\nu$ on $SO(3)$, the distribution of the geometric phase is

$$p_\nu(\theta,\, \beta) = R_0(\theta,\, \nu) + 2 \sum_{m \geq 1} \cos(m\beta) R_m(\theta,\, \nu), \tag{3.16}$$

$$\text{where } R_m(\theta,\, \nu) = \frac{1}{2\pi} \sum_{\lambda \geq m} (2\lambda + 1) e^{\nu(\hat{\Phi}^\lambda_{m,m} - 1)} d^\lambda_{m,m}(\theta). \tag{3.17}$$

Because $\hat{p}^\lambda_\nu$ is diagonal, only the functions U^λ_{mm} appear in the expansion of expression (1.31). Equation (3.16) is a general expression of the distribution of geometric phase in multiple scattering regime for polarized waves. We have only assumed that the scatterers are spherical, identical and that the scattering events were independent. To obtain this

result we have only used the Fourier series on $SO(3)$, summed the Poisson series (3.10) and used the Fourier inversion formula (Theorem 2). As in [Margerin 2009], we get a semi-analytic expression of p_ν that we use for statistical estimation in Section 3.2.4.

The limit of Brownian motion, which is made of isotropic infinitesimal steps of spherical length δ occurring at a strong rate η, corresponds to the physical situation where scatterers are very weak but have a large spatial density. In [Perrin 1936, Eq.(42)], Perrin showed that $\hat{\Phi}^\lambda_{m,m} = \exp[-\frac{1}{2}\left(\lambda(\lambda+1) - m^2\right)\delta^2]$ for $\delta^2 = \mathcal{D}t/\nu$. Thus, formula (3.16) gives the distribution of the geometric phase for a wave propagating in a heterogeneous continuous medium as (3.16) where $\nu(\widehat{\Phi}^\lambda_{m,m} - 1)$ should be replaced by $-\frac{1}{2}[\lambda(\lambda+1) - m^2]\mathcal{D}t$. It is noticeable that the coefficients $\hat{\Phi}^\lambda_{m,m}$ depends on m, contrary to the result of Proposition 12, as the Brownian motion is considered on S^2 with parallel transport, instead of a Brownian motion in $SO(3)$.

Within the Brownian motion limit, taking $\theta = 0$ in (3.16) gives the result demonstrated by Antoine *et al.* [Ouvry 1991], and integrating over θ we find the result obtained by Krishna *et al.* [Sinha 2000]. Equation (3.16) is therefore a generalization of these known formulae that can be applied to an arbitrary scattering phase function Φ, also in the case of a small number of scattering events and for an arbitrary deviation angle θ.

As an illustration we display the distribution obtained from formula (3.16) in the two following cases: discrete Henyey-Greenstein scatterers with anisotropy $g = 0.8$ on a range of Poisson parameter up to $\nu = 20$ on the one hand (Figures 3.15 and 3.16) and diffusive limit with $\mathcal{D} = 1\ \mathrm{rad}^2\,\mathrm{s}^{-1}$ on the other hand (Figures 3.17 and 3.18). Note that any phase function Φ could be used to obtain similar results for Mie scatterers, resonant scatterers, etc. The exit angle is $\theta = 0$ for all of the figures. It means that the polarization is observed for rays propagation in the same direction as the initial (source) direction of propagation.

We can observe that the behavior of the distributions on Figures 3.16 and 3.18 are very different, especially at short times, the distribution p_ν is sharply peaked for the Henyey-Greenstein scatterers but in the case of Brownian motion it is smooth. At long times, depolarization is observed in both cases. The depolarized state is reached with significantly distinct behaviours. This is an illustration that depolarization is significantly dependent on the scattering properties of the medium. It is therefore natural to investigate how these scattering properties can be estimated from observations through signal analysis. In the next section, an EM (Expectation-Maximization) estimation method is used to obtain an estimation of the mean free path ℓ from polarization distribution measurement.

3.2.4 Estimation via Expectation-Maximization

In this section, we take advantage of the Fourier expansion and propose a statistical estimation of the mean free path ℓ obtained from the measured distribution of the geometric phase β. This is achieved by estimating the Poisson parameter $\nu = \eta t$. The estimate of ν is denoted by $\hat{\nu}$.

The inverse problem is solved using an expectation-maximization (EM) approach [Rubin 1977] with the parametric CPP. The EM algorithm is based on the maximization of the log-likelihood of the *a posteriori* distribution given in (3.10). Suppose that we are given a sample of size M of observations (measurements) F_m for $1 \leq m \leq M$. It is assumed that each observation F_m follows independently the probability law (3.16), the joint probability distribution function is simply the product $\prod_m p_\nu(F_m)$, with $p_\nu(F_m) = p_\nu(\theta_m, \psi_m + \varphi_m)$, where φ_m, θ_m and ψ_m are the Euler angles of F_m. The value of ν that maximizes this

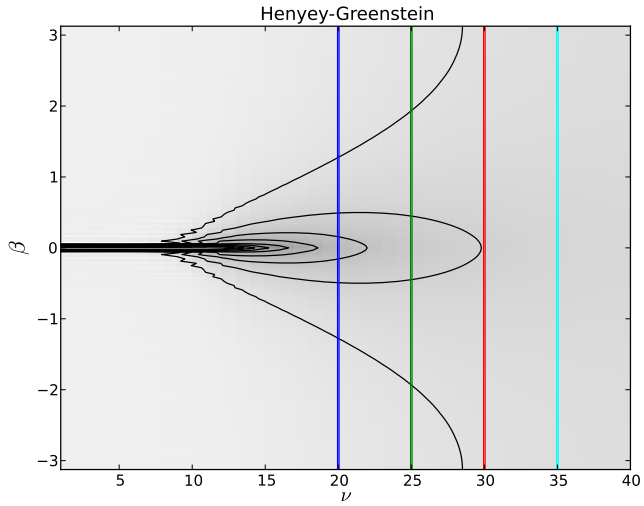


Figure 3.15: Evolution of the density $p_{\eta t}(\theta = 0, \beta)$ with time in a medium containing Henyey-Greenstein scatterers with anisotropy $g = 0.8$. The curves are iso-density levels from 0.05 to 0.45 by increments of 0.05. The vertical lines correspond to the slices displayed on Figure 3.16. The irregularities on the iso-density are due to a truncated Fourier series from the reconstruction formula.

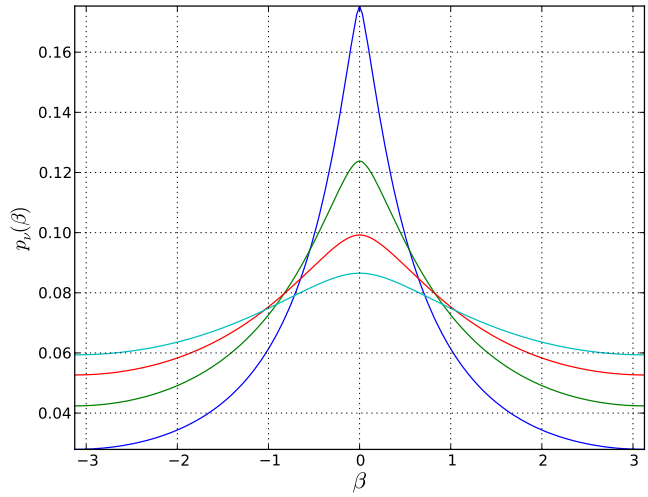


Figure 3.16: Distribution of the geometric phase β for the values of the Poisson parameter $\nu = \eta t$: $\nu = 20$, $\nu = 25$, $\nu = 30$ and $\nu = 35$ corresponding to the vertical lines of Figure 3.15.

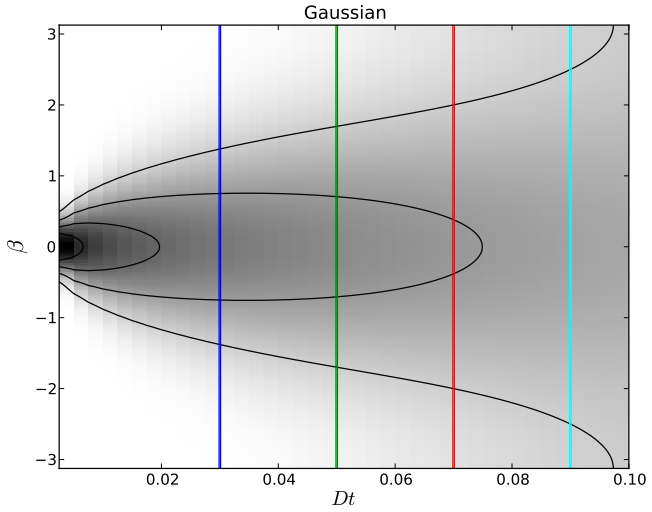


Figure 3.17: Evolution of the density $p_{\mathcal{D}t}(\theta = 0,\, \beta)$ with time for a Gaussian phase function. This corresponds to the rotational Brownian motion on the sphere studied by Perrin in [Perrin 1936]. The curves are iso-density levels from 0.05 to 0.45 by increments of 0.05. The vertical lines correspond to the slices displayed on Figure 3.18.

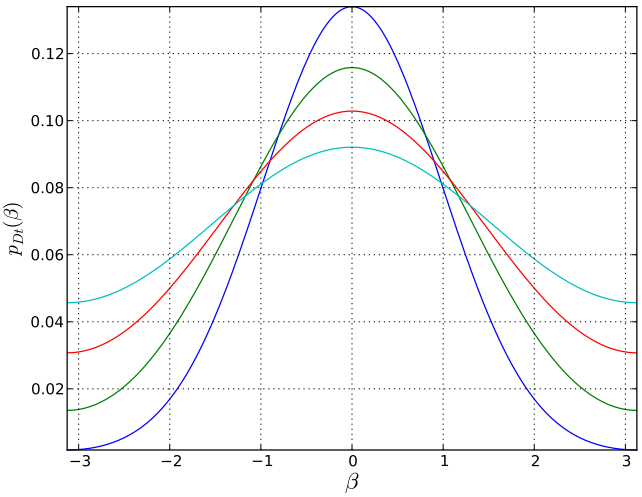


Figure 3.18: Density $p_{\mathcal{D}t}(\theta = 0,\, \beta)$ for a Gaussian phase function with the values $\mathcal{D}t = 0.03$, $\mathcal{D}t = 0.05$, $\mathcal{D}t = 0.07$ and $\mathcal{D}t = 0.09$ corresponding to the vertical lines of Figure 3.17.

log-likelihood is

$$\hat{\nu} = \text{argmax}_\nu \left(\sum_{m=1}^{M} \log p_\nu(F_m) \right). \tag{3.18}$$

The EM algorithm is an iterative procedure that almost surely converges to a local maximum [Rubin 1977]. The main idea consists in introducing a hidden random variable $n = [n_1, ..., n_M]$ that represents the number of scattering events for each observation $F_1, ..., F_M$ and then to construct the sequence of estimator $\hat{\nu}_i$ as

$$\hat{\nu}_{i+1} = \text{argmax}_\nu \sum_{n=0}^{\infty} \mathbb{P}(n \mid F_{1..M}, \hat{\nu}_i) \sum_{m=1}^{M} \log p_\nu(F_m, n_m) \tag{3.19}$$

The computation of the function on the right term is called the expectation step (E-step). Only the terms $\mathbb{P}(n \mid F_{1..M}, \hat{\nu}_i)$ need to be updated during the algorithm as the sum over m does not change. Using Baye's rule, it can be written

$$\mathbb{P}(n \mid F_{1..M}, \hat{\nu}_i) = \frac{\prod_m p_{F_n}(F_m)}{\prod_m p_{\hat{\nu}_i}(F_m)} e^{-\hat{\nu}_i} \frac{\hat{\nu}_i^n}{n!}$$

The elements $p_{F_n}(F_m)$ are independent from $\hat{\nu}_i$ and therefore can be determined once and for all for the complete generation of the sequence $\hat{\nu}_i$. Thus, the E-step is limited to the determination of the coefficients $p_{\hat{\nu}_i}(F_m)$.

The maximization step (M-step) consists in maximizing the function from equation (3.19). Using Bayes' rule to decompose $p_\nu(F_m, n_m)$ leads to the following algorithm

Algorithm 2 Expectation maximization algorithm

- Initiate the algorithm by considering a value $\hat{\nu}_0$.
- Compute off line the set of probabilities $p_{F_n}(F_m)$.
- Iterate on i the recursive operations
  1. Compute the set $p_{\hat{\nu}_i}(F_m)$ from the current estimation.
  2. Determine the next estimation as

$$\hat{\nu}_{i+1} = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{p_{\hat{\nu}_i}(F_m)} \sum_{n=0}^{\infty} n\, p_{F_n}(F_m)\, e^{-\hat{\nu}_i} \frac{\hat{\nu}_i^n}{n!}. \tag{3.20}$$

In practice the sums over n in (3.20) and (3.10) and over j in (3.17) are truncated to an arbitrarily fixed value N. At most N scattering events are described by the model, therefore N should be taken large compared to L/ℓ. This EM-algorithm is able to estimate one parameter, the Poisson parameter ν (or equivalently the mean free path $\ell = ct/\nu$), which means that we have considered that the phase function Φ is known. It is nonetheless possible to estimate simultaneously any finite number of parameters of p_ν, such as the anisotropy g of the scatterers, but the iteration formulas are more sophisticated.

We illustrate the behaviour of the EM algorithm in Figure 3.19 which shows the convergence of the estimator $\hat{\nu}$ for different initial values $\hat{\nu}_0$. Acceleration procedures could be developed in the case of large dataset from experimental setup. Note that no local minima were reached in the simulation, leading to convergence in all the cases. However, as can

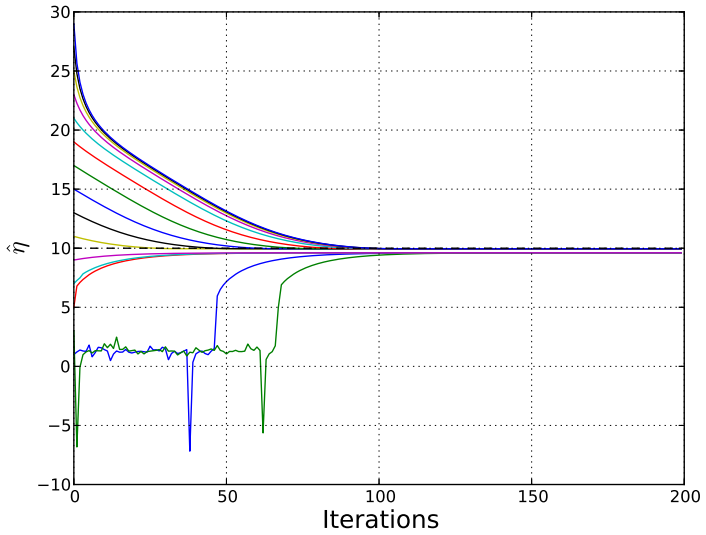


Figure 3.19: Convergence of the estimate of the Poisson parameter $\hat{\eta}$ using EM algorithm 2 with different initialization values ranging from 1 to 29. The actual value of η was set to 10. A total of $M = 100$ samples have been generated to estimate η. Globally, the estimator converges towards the true value of η, independently of the initial condition.

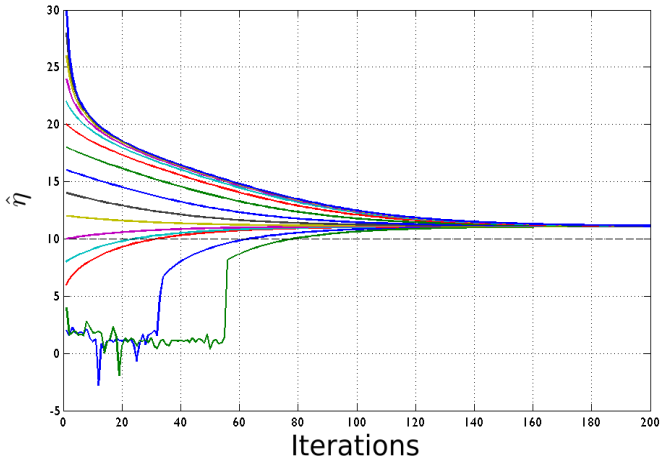


Figure 3.20: Convergence of the estimate of the Poisson parameter $\hat{\eta}$ using EM algorithm 2 with different initialization values ranging from 1 to 29 for noisy observed samples. The variance of the noise was set to 0.1. The actual value of η was set to 10. The estimation still converges but with a positive bias, as the distribution of the geometric phase seems more spread.

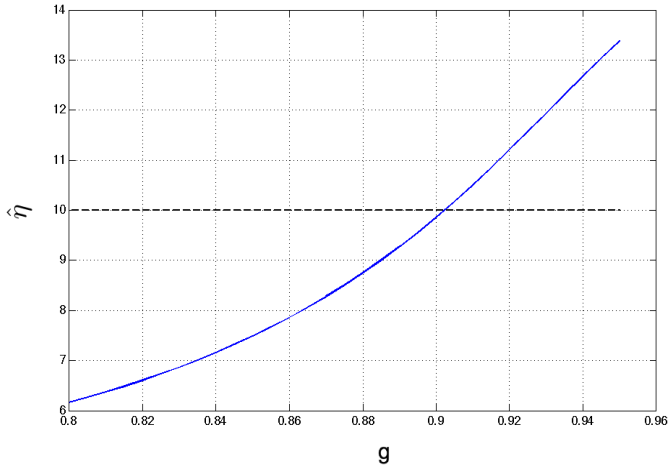


Figure 3.21: Convergence of the estimate of the Poisson parameter $\hat{\eta}$ using EM algorithm 2 for different priori of the anisotropy parameter g. The real value of g for generating the samples was set to 0.9 and η set to 10. The estimation $\hat{\eta}$ appears to be an increasing function of the priori.

be seen for initial values below 10 in Figure 3.19, underestimated initial values lead to a systematic bias in $\hat{\nu}$. This suggests that high values should be privileged when processing dataset, as it does not penalize convergence and leads to a more accurate estimate.

In practice, the samples will be submitted to several measurement errors. Considering a noise with a variance of 0.1 on the observed angles θ and β, the same algorithm creates a bias on the estimator $\hat{\eta}$. In Figure 3.20, it appears that the presence of the noise does not prevent the algorithm from converging with different initialization values. However, the noise is adding a positive bias. Numerically, we can show that the bigger the variance is, the larger the bias is. This effect can be explain by considering that the noise is spreading the values of θ and β. We know that for a propagation further than $\ell^\star$, the geometric phase will be uniform. As the distribution of the angles is spreading with the noise, it gets closer to the uniformization, artificially increasing the number of scattering events. In our case, this reads as a positive bias of the estimator $\hat{\eta}$. This effect is illustrated in Figure 3.20

Another source of error might be due to the use of a wrong model for the scatterers. For example, in our case, a wrong priori on the anisotropy parameter g. Figure 3.21 illustrates the effect or considering a wrong priori model for the scatterers on the estimator $\hat{\eta}$. It appears that $\hat{\eta}$ is a monotonous increasing function of g. A priori on g larger than the actual value signifies that the scatterers does not affect as strongly as they should the trajectory of the wave. For given observations, this is considered by the algorithm as a larger estimation $\hat{\eta}$ to "compensate" this effect and matching properly the likelihood.

3.2.5 Conclusion

The description of the depolarization of multiply scattered waves can be made, in the forward scattering regime, through a compound Poisson process (CPP). This stochastic model predicts the distribution of the geometric phase in all directions, for discrete or continuous scattering media, generalizing existing results (forward outgoing direction or a spatially fluctuating medium). Moreover, the CPP model allows a more detailed description

of the phenomenon as it provides the behaviour of this distribution as a function of the output scattering angle θ.

The present approach allows to design an iterative procedure to estimate properties of the scattering medium through the measurement of the outgoing polarization distribution. We have illustrated this point by presenting an expectation-minimization (EM) algorithm for the estimation of the Poisson parameter, which is directly linked to the transport mean free path ℓ. An interesting feature of this technique is that it relies on polarization rather than on amplitude measurements.

Ongoing work consists in validating the proposed model and estimation algorithm on measurement of polarization distribution from a laboratory experiment.

The stochastic model of section 3.2.3 and the distribution of geometric phase found in Section 1.4 constitute powerful tools for numerical approaches of the polarization in multiple scattering media. It can be used even for small scatterers or in any other situation where the scattering anisotropy vanishes.

The EM procedure can be used on experimental data to estimate the mean free path from polarization statistics of any kind of polarized waves in multiple scattering media, provided that the limit of uniformly distributed polarization is not reached and the source polarization is known. Such waves can be light or elastic waves in solids.

Conclusion

In this thesis, we present and solve two independent problems, respectively in Chapters 2 and 3.

Chapter 2 is dedicated to the problem of filtering from observation on the rotation group $SO(n)$ (complete observation) and on Stiefel manifold (partial observation). The observation processes are solutions of nonlinear stochastic differential equations and we want to estimate the angular velocity of these processes.

The algorithms used nowadays like the extended Kalman filter consider a local linearization of these processes. However, this local linearization needs to be performed at every step and the equations of the filter are dependent of the previous estimations. We propose a different approach, which consist in constructing a new process from the observation called the antidevelopment. The antidevelopment satisfies a linear model and classical solutions can be applied in the case of a complete observation and a numerical solution can be implemented in the case of partial observation.

A particular attention has been given to the practical implementation of the solutions. Because the acquired samples will be discretized, an interpolation scheme needs to be used to determine the likelihood required in the solution. We showed that several interpolation schemes could be chosen, depending on the desired accuracy of the estimation and the computation performances. For a complete observation, our solution is able to properly estimate the angular velocity of a rotation process. For partial observation in the Stiefel manifold, our algorithm only converges if the angular velocity is constant. Despite that the convergence still hold if the state is constant for a sufficient large time (for a stair model for example), the algorithm is only able to estimate partially the state if this velocity is evolving too quickly.

The presented algorithm has potential applications for filtering a partial observation of the trajectory of a robot. The observation could be generated by cameras and we can use the described method to determine its full 3D motion.

In Chapter 3, we focus on the geometric phase acquired during the propagation of a linearly transverse polarized elastic wave. This phase is acquired during a 3D propagation of the wave and corresponds to a rotation of the major axis of the polarization. For an elastic wave propagating in a waveguide, this phase can be determined from one hand with geometrical tools from the path followed by the wave and experimentally from the other hand. By combining both values, one can construct a parametric estimator of the waveguide.

For an elastic wave propagating in a scattering medium, a distribution of the geometric phase based on a compound Poisson process model can be obtained. By sampling from this distribution, an Expectation-Maximation algorithm can be used to realize parametric estimation of the scattering medium or the scatterers. Despite that this distribution has not been observed yet, future works should focus on experimentally sample this distribution. This is conditioned by an accurate measurement of the polarization which is possible for light but less trivial for elastic waves. Once this measurement is made, it should lead to new techniques for non-intrusive imaging and material monitoring that will take advantage of the polarization information in addition to the classically measured amplitude or intensity of wave-fields.

Bibliography

[Altmann 2005] Simon L. Altmann. Rotations, quaternions, and double groups. Dover Books on Mathematics, 2005. 3, 4, 6, 10, 11, 101, 103

[Arnold 1966] Vladimir Arnold. *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits.* Annales de l'institut Fourier, vol. 16, page 319, 1966. 5, 11

[Arslan 2007] Mehmet Kemal Ozdemir & Hueyin Arslan. *Channel estimation for wireless OFDM systems.* Communications Surveys Tutorials, IEEE, vol. 9, no. 2, pages 18–48, 2007. 1, 30, 46, 100

[Belanger 1992] Pierre R. Belanger. *Estimation of angular velocity and acceleration from shaft encoder measurements.* In Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on, pages 585–592 vol.1, 1992. 30

[Berry 1984] Michael V. Berry. *Quantal phase factors accompanying adiabatic changes.* Proc. R. Soc. A, vol. 392, pages 45–57, 1984. 61, 63, 122

[Bloch 1990] Anthony M. Bloch et Jerrold E. Marsden. *Stabilization of rigid body dynamics by the Energy-Casimir method.* Systems & Control Letters, vol. 14, no. 4, pages 341 – 346, 1990. 48

[Brosseau 1998] Christian Brosseau. Fundamentals of polarized light: A statistical optics approach. John Wiley & Sons, 1998. 67

[Bucy 1961] Rudolf E. Kalman & Richard S. Bucy. *New results in linear filtering and prediction theory.* Transactions of the ASME. Series D, Journal of Basic Engineering, vol. 83, pages 95–107, 1961. 29, 35, 112

[Canet 1994] Pierre Canet. *Kalman Filter Estimation of angular velocity and acceleration On-line implementation.* Report, 1994. 30

[Casella 2010] Christian P. Robert & George Casella. Introducing monte carlo methods with R. Springer Series Use R, 2010. 4, 21, 23, 101

[Chellappa 2008] Pavan Turaga & Ashok Veeraraghavan & Rama Chellappa. *Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision.* In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1–8, 2008. 36, 115

[Chiao 1986] Akira Tomita & Raymond Y. Chiao. *Observation of Berry's topological phase by use of an optical fiber.* Phys. Rev. Lett., vol. 57, pages 937–940, 2471, 1986. 69, 70

[de Hulst 1957] Hendrik C Van de Hulst. Light scattering by small particles. Dover, 1957. 75

[Dieudonné 1980] Jean Dieudonné. Special functions and linear representations of lie groups. American Mathematical Society, 1980. 80

[do Carmo 1976] Manfredo P. do Carmo. Differential geometry of curves and surfaces. Prentice-Hall, 1976. 4

[Durrett 1996] Richard Durrett. Stochastic calculus : a practical introduction. Probability and stochastics series. CRC Press, 1996. 3

[Faraut 2008] Jacques Faraut. Analysis on lie groups: An introduction. Cambridge University Press, 2008. 6, 10, 23, 26, 27, 103, 110

[Faure 2013] Frédéric Faure. *Cours de Géométrie et topologie pour la physique.* http://www-fourier.ujf-grenoble.fr/~faure/enseignement/geometrie_topologie_M2/cours_sans_corrections_exercices.pdf, 2013. 6, 63, 65, 103, 123

[Feller 1971] William Feller. An introduction to probability theory and its applications, vol. II. Second edition. John Wiley & Sons Inc., 1971. 11, 16

[Fuller 1978] F. Brock Fuller. *Decomposition of the linking number of a closed ribbon: A problem from molecular biology.* Proc. Natl. Acad. Sci. USA, vol. 75, pages 3557–3561, 1978. 70, 125

[Greenstein 1941] Louis G. Henyey & Jesse L. Greenstein. *Diffuse radiation in the Galaxy*, 1941. 77

[Hall 2003] Brian Hall. Lie groups, lie algebras, and representations: An elementary introduction. Springer, 2003. 8

[Halper 2008] Aaron Halper. *A proof of the Gauss-Bonnet theorem.* http://www.math.uchicago.edu/may/VIGRE/VIGRE2008/REUPapers/Halper.pdf, 2008. 65

[Hanzon 1998] Damiano Brigo& Floyd B. Hanzon et François LeGland. *A differential geometric approach to nonlinear filtering: the projection filter.* Automatic Control, IEEE Transactions on, vol. 43, no. 2, pages 247–252, 1998. 1, 100

[Herbolzheimer 1988] David J. Pine& David A. Weitz & Paul M. Chaikin & E. Herbolzheimer. *Diffusing-wave spectroscopy.* Phys. Rev. Lett., vol. 60, pages 1134–1137, 1988. 75

[Hubbard 1972] Paul S. Hubbard. *Rotational Brownian Motion.* Phys. Rev. A, vol. 6, pages 2421–2433, Dec 1972. 10

[Iacus 2010] Stefano M. Iacus. Simulation and inference for stochastic differential equations. Springer Series in Statistics, 2010. 3, 13, 14, 101

[Ishimaru 1978] Akira Ishimaru. Wave propagation and scattering in random media, vol.1,2. Academic Press, New York, 1978. 75

[Itō 1950] Kiyosi Itō. *Stochastic differential equations in a differentiable manifold.* Nagoya Mathematical Journal, vol. 1, page 35, 1950. 11, 104

[Jazwinski 1970] Andrew H. Jazwinski. Stochastic processes and filtering theory. Academic press, 1970. 3, 29, 30, 32, 34, 36, 42, 101, 112, 113

[Johansen 2011] Arnaud Doucet & Adam M. Johansen. *A tutorial on particle filtering and smoothing: fifteen years later*, 2011. 51

[Johnson 2003] Darall Henderson & Sheldon H. Jacobson & Alan W. Johnson. The theory and practice of simulated annealing. Springer US, 2003. 21

[Kallenberg 2002] Olav Kallenberg. Foundations of modern probability. Probability and its applications. Springer, 2002. 20, 108

[Kirby 2009] Yui Man Lui & J. Ross Beveridge & Michael Kirby. *Canonical Stiefel quotient and its application to generic face recognition in illumination spaces.* In Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems, BTAS'09, pages 431–438. IEEE Press, 2009. 46, 47

[Kunita 1971] Hiroshi Kunita. *Asymptotic behavior of the nonlinear filtering errors of Markov processes.* Journal of Multivariate Analysis, vol. 1, no. 4, pages 365 – 393, 1971. 1, 100

[Kyatkin 2001] Gregory S. Chirikjian & Alexander B. Kyatkin. Engineering applications of noncommutative harmonic analysis. CRC Press, 2001. 23, 79, 110

[Lagendijk 2008] P. M. Johnson& Sanli Faez& Ad Lagendijk. *Full characterization of anisotropic diffuse light.* Optics Express, vol. 16, pages 7435–7446, 2008. 76

[Larose 2005] Eric Larose. *Diffusion multiple des ondes sismiques et expériences analogiques en ultrasons.* PhD thesis, Université Joseph Fourier, 2005. 1, 101

[Li 2004] Xiang Dong Li. *Asymptotic behavior of divergences and Cameron-Martin theorem on loop spaces.* The Annals of Probability, vol. 32, 2004. 37

[Lisyansky 1992] Narciso Garcia & Azriel Z. Genack & Alexander A. Lisyansky. *Measurement of the transport mean free path of diffusing photons.* Phys. Rev. B, vol. 46, pages 14475–14478, 1992. 76

[Lo 1973] James Ting-Ho Lo. *Signal detection of rotational processes and frequency demodulation.* In Decision and Control including the 12th Symposium on Adaptive Processes, 1973 IEEE Conference on, volume 12, pages 139–145, 1973. 1, 100

[Lo 1975] James Ting-Ho Lo. *Signal detection for bilinear systems.* Inf. Sci., pages 249–278, 1975. 36, 116

[Lochak 1959] Mark A. Naimark & Georges Lochak. Les représentations linéaires du groupe de lorentz. Paris Dunot, 1959. 24

[López 2011] Marco Leonetti & Cefe López. *Measurement of transport mean-free path of light in thin systems.* Opt. Letters, vol. 36, pages 2824–2826, 2011. 76

[Lovera 2002] Alessandro Astolfi & Marco Lovera. *Global spacecraft attitude control using magnetic actuators.* In American Control Conference, 2002. Proceedings of the 2002, volume 2, pages 1331–1335 vol.2, 2002. 10, 36, 48, 103, 115

[Manton 2012] S. Said & J.H. Manton. *Particle filtering with observation in a compact manifold.* SIAM Journal on Control and Optimization, 2012. 48

[Margerin 2009] Nicolas Le Bihan & Ludovic Margerin. *Nonparametric estimation of the heterogeneity of a random medium using compound Poisson process modeling of wave multiple scattering.* Phys. Rev. E, vol. 80, page 016601, Jul 2009. 75, 77, 78, 81, 126, 127

[McKean 1960] Henry P. McKean. *Brownian motions on the 3-dimensional rotation group.* Memoirs of the College of Science, University of Kyoto, 1960. 3, 14, 101, 105

[Meer 2007] Oncel Tuzel & Fatih Porikli & Peter Meer. *Human detection via classification on riemannian manifolds.* In In Proc. of the IEEE conference on Computer Vision and Pattern Recognition, pages 1–8, 2007. 36, 115

[Meyer 1981] Paul-André Meyer. *Géométrie stochastique sans larmes.* Séminaire de probabilités de Strasbourg, vol. 15, page 44, 1981. 3, 11, 101, 104

[Mishchenko 1991] Michael Mishchenko. Multiple scattering of light by particles: Radiative transfer and coherent backscattering. Cambridge University Press, 1991. 75

[Montgomery 1993] Richard Montgomery. *Gauge theory of a falling cat.* Fields Institute Communications, 1993. 67

[Nikiforov 1993] Michèle Basseville & Igor Nikiforov. Detection of abrupt changes: Theory and application. Prentice hall information and system sciences series. Prentice Hall, 1993. 54

[Ouvry 1991] Michel Antoine & Alain Comtet & Jean Desbois & Stéphane Ouvry. *Magnetic fields and Brownian motion on the 2-sphere.* J. Phys. A: Math. Gen., vol. 24, pages 2581–2586, 1991. 76, 81

[Perrin 1936] Francis Perrin. *Mouvement browien d'un ellipsoïde (II).* J. Phys. Radium, vol. 1, pages 1–11, 1936. 81, 83

[Pontier 1988] Monique Pontier. *Approximation d'un filtre avec observation sur une variété compacte.* Stochastics, vol. 24, pages 285–304, 1988. 29, 38, 112, 117

[Rossetto 2001] Anthony C. Maggs & Vincent Rossetto. *Writhing Photons and Berry Phases in Polarized Multiple Scattering.* Phys. Rev. Lett., vol. 87, page 253901, Nov 2001. 76, 77, 126

[Rossetto 2012] Jérémie Boulanger & Nicolas Le Bihan & Stefen Catheline & Vincent Rossetto. *Non-adiabatic geometric phase of elastic waves.* Annals of Physics, vol. 327, page 952, 2012. 1, 61, 100, 122

[Rossetto 2013] Jérémie Boulanger & Nicolas Le Bihan & Vincent Rossetto. *Stochastic description of geometric phase for polarized waves in random media.* Journal of Physics A: Mathematical and Theoretical, vol. 46, no. 3, page 035203, 2013. 1, 61, 100, 122

[Rubin 1977] Arthur P. Dempster & Nan M. LairdÊ& Donald B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the royal statistical society, Series B, vol. 39, no. 1, pages 1–38, 1977. 81, 84

[Sandison 1995] Trent Ning & L. Papiez & G. Sandison. *Compound-Poisson-process method for the multiple scattering of charged particles.* Phys. Rev. E, vol. 52, no. 5, pages 5621–5633, Nov 1995. 77

[Segert 1987] Jan Segert. *Photon Berry's phase as a classical topological effect.* Phys. Rev. A, vol. 36, pages 10–15, 1987. 61, 67, 69, 122, 124

[Sepulchre 2008] Pierre-Antoine Absil & Robert Mahony & Rodolphe Sepulchre. Optimization algorithms on matrix manifolds. Princeton University Press, Princeton, NJ, 2008. 49, 120

[Sinha 2000] M.G. Mallela Krishna & Joseph Samuel & Supurna Sinha. *Brownian motion on a sphere: distribution of solid angles.* J. Phys. A: Math. Gen., vol. 33, page 5965, 2000. 76, 81

[Smith 1998] Alan Edelman & Tomás A. Arias & Steven T. Smith. *The geometry of algorithms with orthogonality constraints.* Siam J. Matrix Anal. Appl, vol. 20, no. 2, pages 303–353, 1998. 23, 44

[Srivastava 2000] Anuj Srivastava et Eric Klassen. *Geometric Filtering for Subspace Tracking*, 2000. 12, 104

[Thomas 1991] Thomas M. Cover & Joy Thomas. Elements of information theory. Wiley, 1991. 1, 100

[Treyssède 2007] Fabien Treyssède. *Numerical investigation of elastic modes of propagation in helical waveguides.* J. Acoust. Soc. Am., vol. 121, pages 3398–3408, 2007. 71

[van Albada & Ad Lagendijk 1985] Meint P. van Albada & Ad Lagendijk. *Observation of weak localization of light in a random medium.* Phys. Rev. Lett., vol. 24, pages 2692–2695, 1985. 76

[van Tiggelen & Ludovic Margerin 2009] Domitille Anache-Ménier & Bart van Tiggelen & Ludovic Margerin. *Phase Statistics of Seismic Coda Waves.* Phys. Rev. Lett., vol. 102, page 248501, 2009. 76

[Vidale 1986] John E. Vidale. *Complex polarization analysis of particle motion.* Bull. of the Seismological Soc. of Am., vol. 76, no. 5, pages 1393–1405, 1986. 73

[Weiss 1999] O.J. Micka & Anthony J. Weiss. *Estimating frequencies of exponentials in noise using joint diagonalization.* Signal Processing, IEEE Transactions on, vol. 47, no. 2, pages 341–348, 1999. 44

[Williams 2000] L. Chris G. Rogers & David Williams. Diffusions, markov processes, and martingales, vol. II. Cambridge Mathematical Library. Cambridge University Press, 2000. 12

[Zakai 1969] Moshe Zakai. *On the optimal filtering of diffusion processes.* Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, vol. 11, no. 3, pages 230–243, 1969. 33, 114

Appendices

Appendix A

Résumé en Français

A.1 Introduction

Cette thèse s'intéresse au filtrage adaptative et à l'estimation paramétrique pour des processus stochastiques sur le groupe des rotations et les variétés de Stiefel. Le problème du filtrage adaptatif est traité dans le Chapitre 2 et le problème de l'estimation paramétrique est traité dans le Chapitre 3 avec des applications en physique des ondes. Le Chapitre 1 fournit le bagage mathématique utile pour la compréhension des autres chapitres. Au cours de cette étude seront approfondies à la fois les solutions théoriques aux problèmes proposés ainsi que l'implémentation numérique de ces solutions.

Les processus stochastiques sur le groupe des rotations peuvent être considérés comme une extension des processus rotationnels. Malgré le fait que ces processus étaient à l'origine utilisés pour modéliser l'orientation d'un système [Lo 1973], ils sont maintenant utilisés dans de nombreuses applications en traitement du signal lorsque des contraintes d'orthogonalité surviennent. Cela intervient par exemple dans des domaines tels que le traitement d'antennes pour séparer les composantes indépendantes [Arslan 2007] ou bien en théorie de l'information pour maximiser le taux de transfert dans un canal de communication [Thomas 1991]. Afin de décrire ces processus et leurs contraintes, le Chapitre 1 décrit la géométrie du groupe des rotations $SO(n)$ et utilise des outils de géométrie différentielle pour construire et simuler ces processus.

Parmi les problèmes qu'il est possible d'étudier pour les processus rotationnel, le Chapitre 2 considère le problème du filtrage dans le cas d'observation sur $SO(n)$ dans un premier temps, puis pour des observations sur les variétés de Stiefel $V_{n,k}$ par la suite. Dans ce chapitre, on cherche à estimer l'état de notre système, ici modélisé par la vitesse angulaire du système, à partir d'observations bruitées.

Malgré le fait que cette étude entre dans le champ du filtrage non linéaire, notre approche est sensiblement différente de celles que l'on peut trouver dans la littérature. Dans [Hanzon 1998] et [Kunita 1971] par exemple, le problème revient à estimer un état avec des contraintes d'orthogonalités à partir d'observations avec bruit additif. Grossièrement, le problème revient à déterminer une application de $\mathbb{R}^{n\times n}$ vers $SO(n)$, le principal obstacle provenant des contraintes de $SO(n)$ et la manière de calculer efficacement une moyenne empirique sur $SO(n)$.

Dans cette thèse, l'on se concentre sur un modèle différent, en considérant l'approche inverse: à partir d'observations sur $SO(n)$, l'on cherche à estimer un état dans $\mathbb{R}^{n/timesn}$. Il est à noter que l'espace d'arrivée dans lequel l'état de notre système est défini ne modifie pas nos résultats et un différent espace donnerait des résultats similaires. Notre principal préoccupation étant d'appliquer des méthodes de filtrage classiques pour des modèles d'observations non linéaires.

Ce problème est résolu en construisant un processus satisfaisant un modèle de bruit additif comme image d'une application bijective à partir de l'observation. Une partie non négligeable du Chapitre 2 est consacrée à l'implémentation numérique des méthodes présentées et leurs performances.

Le Chapitre 3 est basé sur nos articles [Rossetto 2012] et [Rossetto 2013] et se concentre sur la phase géométrique des ondes élastiques. La propagation d'ondes avec polarisation transverse et linéaire peut être modélisée par un processus sur $SO(3)$ et est capable d'acquérir une phase géométrique au cours de la propagation. D'une part, cette phase peut être déterminée théoriquement à partir du chemin suivi par l'onde au cours de sa propagation [Rossetto 2012] et d'autre part, nous montrons que cette phase géométrique peut être déterminée expérimentalement avec des accéléromètres. Par conséquent, si la propagation

de l'onde est paramétrée, la phase géométrique peut nous servir à construire des estimateurs de ces paramètres.

Une grande partie des méthodes d'imageries d'aujourd'hui consiste à utiliser l'amplitude de l'onde après la propagation dans un milieu aléatoire pour construire des estimateurs de certaines caractéristiques de ces milieux [Larose 2005]. Cependant, même si ces résultats restent fonctionnels pour des ondes avec une polarisation transverse, ces méthodes n'utilisent pas l'information donnée par la direction de polarisation après propagation dans le milieu. La phase géométrique se base sur cette information pour construire des estimateurs. Autrement dit, la phase géométrique peut être utilisée comme une technique d'imagerie non intrusive car elle contient de l'information sur le chemin en 3D que l'onde a suivi au cours de sa propagation dans le milieu.

A.2 Équations différentielles stochastiques sur $SO(n)$

Les équations stochastiques différentielles sont utilisées dans de nombreuses applications en ingénierie tel que le filtrage [Jazwinski 1970], modélisation de phénomènes physiques [**?**], etc. Cette thèse s'intéresse aux équations stochastiques différentielles définies sur le groupe des rotation [Altmann 2005]. Par conséquent, certaines notions de géométrie différentielles ont besoin d'être introduites pour définir ces équations [Meyer 1981].

Le Chapitre 1 sert à introduire les outils mathématiques utiles pour les Chapitres 2 et 3. La Section 1.1 décrit la structure de groupe du groupe spécial orthogonal $SO(n)$ et les notions de géométrie différentielles utilisées pour définir les processus stochastiques dans le reste du chapitre. Un type de trajectoires, les géodésiques sont introduites car ces dernières seront d'une importance majeure pour simuler des processus stochastiques sur $SO(n)$. La Section 1.1 se termine par une présentation du cas de $SO(3)$. Cette présentation sert d'exemple pour les notions précédentes et mérite un traitement plus complet à cause de son importance dans de nombreuses applications.

La Section 1.2 introduit le calcul stochastique pour les équations différentielles stochastiques sur $SO(n)$. Ces dernières sont définies par des champs vectoriels invariants à droite avec un bruit Brownien. Être capable de simuler de telles processus peut s'avérer utile pour comparer les performances de différentes méthodes ou bien pour valider un modèle d'observation par exemple. Classiquement, le schéma d'Euler [Iacus 2010] est utilisé pour avoir une approximation discrète du processus continu. Dans notre cas, le processus à temps continu est défini sur $SO(n)$ à cause de l'absence d'addition sur $SO(n)$, il n'est pas possible d'utiliser le schéma d'Euler. Un schéma différent est donc présenté ici [McKean 1960], généralisant le schéma d'Euler grâce aux géodésiques introduites à la Section 1.1. Par la suite, des outils dédiés à l'étude des processus Markovien sont introduits, tels que les semi-groupes Markovien et leur générateur, l'équation vers l'avant et vers l'arrière. Ces notions seront utilisées dans la Section 1.3 pour démontrer la convergence des simulations.

La Section 1.3 donne deux exemples de processus stochastiques sur $SO(n)$. En premier, une description du mouvement Brownien sur $SO(n)$ ainsi qu'une méthode de simulation numérique sont présentées. Cela sert à étendre le théorème central limite pour des variables aléatoires de rotations. En second, une méthode Monte-Carlo utilise les résultats de la Section 1.2 sur les chaînes Markoviennes pour échantillonner une distribution sur $SO(n)$. Cette méthode peut être étendue pour construire un algorithme d'optimisation, le refroidissement contrôlé [Casella 2010] sur $SO(n)$.

La Section 1.4 présente le concept de fonctions caractéristiques sur $SO(n)$. Les exemples et propriétés présentés peuvent se voir comme la généralisation de résultats classiques d'analyse harmonique pour les signaux périodiques. Les fonctions caractéristiques fournissent une nouvelle preuve au théorème central limite sur $SO(n)$, en tant qu'alternative à la preuve de la Section 1.3.

A.2.1 Géométrie de $SO(n)$

Pour un entier positif n, la notation $SO(n)$ désigne l'ensemble des matrices de taille $n/timesn$ orthogonales avec un déterminant unitaire. Les colonnes de ces matrices représentent les coordonnées canoniques d'une base orthonormée de $/mathbbR^n$ positivement orientée. $SO(n)$ est stable par multiplication et pour $R \in SO(n)$, $R^{-1} = R^T$. Le groupe $SO(n)$ est aussi appelé groupe des rotations.

On dénote $T_RSO(n)$ l'espace tangent à $SO(n)$ en R. Cela correspond à l'ensemble des vitesses $\dot{R}_0$ de trajectoires R_t différentiables en $R = R_0$. On peut montrer que $T_RSO(n) = \mathfrak{so}(n)R$ où $\mathfrak{so}(n)$ est l'ensemble des matrices antisymétriques $n \times n$. Cet espace est donc de dimension $n(n-1)/2$. On note $TSO(n)$ l'ensemble de tous les espaces tangents à $SO(n)$. Par ailleurs, il est aussi possible de décrire $T_RSO(n)$ à l'aide d'un produit à gauche $T_RSO(n) = R\mathfrak{so}(n)$ plutôt qu'un produit à droite comme précédemment.

En remarquant que $T_{I_n}SO(n) = \mathfrak{so}(n)$, l'on va chercher à définir des objets géométriques invariants par multiplication. En dénotant $\chi(\sigma, R) = \sigma R$ pour $\sigma \in \mathfrak{so}(n)$, on définit un champ vectoriel $\chi(\sigma, .)$, que l'on notera dans la suite $\sigma(.)$. Ce champ vectoriel est invariant à droite, c'est-à-dire que $\sigma(R_1)R_2 = \sigma(R_1R_2)$ pour $R_1, R_2 \in \mathfrak{so}(n)$. De la même manière, il serait possible de définir des champs vectoriels invariants à gauche.

On définit une métrique sur $\mathfrak{so}(n)$ héritée de l'espace ambiant $\mathbb{R}^{n\times n}$:

$$\begin{aligned} < \sigma_1, \sigma_2 > &= \frac{1}{2}\operatorname{trace}(\sigma_1^T\sigma_2) \qquad \text{for } \sigma_1, \sigma_2 \in \mathfrak{so}(n) \\ &= -\frac{1}{2}\operatorname{trace}(\sigma_1\sigma_2). \end{aligned} \tag{A.1}$$

Cette métrique est invariante à droite. On peut alors s'en servir pour définir une métrique sur $T_RSO(n)$. Soit $v_1, v_2 \in T_RSO(n)$, on définit le produit scalaire:

$$< v_1, v_2 >_R = < v_1R^T, v_2R^T > .$$

Comme $v_1R^T, v_2R^T \in \mathfrak{so}(n)$, on déduit de (A.1) que le produit scalaire s'exprime:

$$< v_1, v_2 >_R = \frac{1}{2}\operatorname{trace}(v_1^Tv_2)$$

Il est remarquable que l'expression de ce produit scalaire ne fasse, au final, pas intervenir la connaissance de R.

Etant donné une base orthonormale $\{sigma_i\}$ de $\mathfrak{so}(n)$, on définit pour la suite l'application linéaire et bijective : pour $a \in \mathbb{R}^{n(n-1)/2}$:

$$\hat{a} = \sum_i a^i\sigma_i$$

Par ailleurs, on peut montrer que pour $\sigma \in \mathfrak{so}(n)$, la matrice $R = \exp(\sigma)$ est un élément de $SO(n)$, où exp désigne l'exponentiel matricielle classique. Dans le cas où $|\sigma| < \pi$, l'application exponentiel est inversible. On note log son inverse. On s'intéressera généralement au logarithme de matrices proches de l'identité pour lesquelles cette condition sera vérifiée.

Dans la Section 1.3, on s'intéresse aux processus stochastiques sur $SO(n)$. De tels processus peuvent avoir des trajectoires avec certaines propriétés relativement complexes à étudier, comme c'est le cas pour des processus de diffusions par exemple. Il est heureusement possible de construire ces trajectoires à partir de processus bien plus simples, appelés géodésiques. Les géodésiques peuvent s'interpréter comme l'équivalent dans $SO(n)$ des lignes droites dans $\mathbb{R}^n$. Elles sont utilisées pour simuler des processus stochastiques et apparaissent naturellement lorsque l'on essaie de discrétiser un processus à temps continu.

Les propositions 19 et20 donnent des définitions équivalentes à ces géodésiques, à l'aide d'outils géométriques différents.

Les géodésiques sont toujours définies vis-à-vis d'une connexion. Dans notre cas, on considère la connexion de Levi-Civita [Faure 2013]. Cette connexion est la plus naturelle à utiliser. On définit tout d'abord les géodésiques comme les trajectoires avec une accélération nulle.

Proposition 19. *Les géodésiques de $SO(n)$ sont l'ensemble des trajectoires R_t avec $R_t = \exp(t\sigma)R_0$ pour $\sigma \in \mathfrak{so}(n)$.*

La Proposition 19 est particulièrement utile car elle donne de manière explicite l'expression d'une géodésique. Cela servira à simuler des processus à temps continu.

Tout comme les lignes droites dans $\mathbb{R}^n$ peuvent se définir comme solutions d'une équation différentielle (autrement dit, un flux) avec coefficients constants, on peut définir les géodésiques sur $SO(n)$ comme flux d'équations différentielles avec une combinaison constante de champs vectoriels invariants à droite.

Proposition 20. *Les géodésiques de $SO(n)$ sont les solutions d'équations différentielles du type:*

$$\dot{R}_t = \sum_i a^i \sigma_i(R_t),$$

où les coefficients $\{a^i\}$ sont $n(n-1)/2$ constantes.

La Proposition 20 donne un aperçu de l'utilisation des champs vectoriels définis précédemment. Ces derniers servent à décrire l'évolution du processus dans chaque espace tangent.

Le cas le plus classique des groupes spéciaux orthogonaux reste le groupe $SO(3)$ [Altmann 2005] [Faraut 2008]. Son principal intérêt vient du fait qu'il peut être utilisé pour modéliser l'orientation d'un système dans l'espace vis-à-vis d'un autre référentiel. Dans [Lovera 2002] par exemple, on peut trouver une loi de contrôle pour l'orientation d'un satellite. L'orientation de ce dernier est alors modélisée par un processus sur $SO(3)$.

La base canonique de $\mathfrak{so}(3)$ généralement utilisée est dénotée J_X, J_Y, J_Z avec

$$J_X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, J_Y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, J_Z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Ces matrices peuvent s'interpréter comme des rotations infinitésimal autour des axes canoniques. C'est-à-dire que pour une rotation R_t^i autour de l'axe i peut s'écrire

$$R_t^i = \exp(tJ_i) \qquad \text{pour } i = X, Y, Z.$$

De manière plus générale, une rotation uniforme R_t autour de l'axe $u = (u^1, u^2, u^3)$ s'écrit

$$R_t = \exp\left(t \sum_{i \leq 3} u^i J_i\right) = \exp(t\hat{u}).$$

Comme $SO(3)$ est de dimension trois, ses éléments sont décrits par trois paramètres. Parmi toutes les représentations possibles, les angles d'Euler sont les plus utilisés. C'est d'ailleurs avec les angles d'Euler que sont décrites les rotations du Chapitre 3. Il s'agit de représenter une rotation $R \in SO(3)$ à l'aide de rotations successives autour des axes canoniques. On dira que R est décrite par $\phi \in [0, 2\pi[$, $\theta \in [0, \pi]$, $\psi \in [0, 2\pi[$ avec

$$R(\phi, \theta, \psi) = R^Z(\psi)R^Y(\theta)R^Z(\phi).$$

Cette décomposition est unique dès lors que $\theta \neq 0[\pi]$.

A.2.2 Processus stochastiques et calcul de Itō

Cette section introduit le calcul stochastique et la description de processus sur $SO(n)$ à l'aide de champs vectoriels invariants à droite. Ces processus sont présents dans le Chapitre 2 et Chapitre 3 pour modéliser des phénomènes physiques. Après une brève introduction sur les outils utilisés pour l'étude des équations différentielles stochastiques (abrégées SDE dans la suite), un paragraphe décrit un schéma numérique pour simuler ces processus. A cause de la structure multiplicative de $SO(n)$, il n'est pas possible d'utiliser les schémas classiques de discrétisation. Plus d'information sur le calcul stochastique peut être trouvé dans [Meyer 1981] et [Itō 1950]. Un dernier paragraphe décrit les processus Markoviens de rotation ainsi que les équations vers l'avant et vers l'arrière.

Dans la Section 1.1, un type particulier de trajectoires, les géodésiques sont décrites. Elles sont les solutions d'équations différentielles telles que:

$$\dot{R}_t = \sigma R_t$$

pour $\sigma \in \mathfrak{so}(n)$. Dans un cas plus général, il est possible de considérer la même équation différentielle avec $\sigma = \sigma_t$ qui varie dans le temps

$$\dot{R}_t = \sigma_t R_t = \hat{x}_t R_t$$

où $x_t \in \mathbb{R}^{n(n-1)/2}$ tel que $\hat{x}_t = \sigma_t$. Dans le cas de $SO(3)$, si x_t est la vitesse angulaire instantanée d'un système en rotation, alors R_t décrit dans l'équation précédente est l'orientation de ce système. Si x_t peut être mesuré précisément (en l'absence de bruit, de dérives...), alors l'équation précédente peut servir à connaître l'orientation du système.

Cependant, en pratique, l'équation précédente ne permet pas de modéliser correctement l'orientation du système, à cause de la présence de bruit, de perturbations ou d'interférences [Srivastava 2000]. On introduit le bruit en utilisant une intégrale de Stratonovich, notée $\circ$, de la manière suivante:

$$dR_t = (\hat{x}_t dt + \circ d\hat{w}_t) R_t$$

où $w_t \in \mathbb{R}^{n(n-1)/2}$ est un mouvement Brownien indépendant de x. Dans la suite, on considère que x est un processus à variance finie.

La conversion de l'équation précédente en intégrale de Itō donne

$$dR_t = \left(\hat{x}_t dt + d\hat{w}_t + \frac{1}{2}\mathbb{E}[d\hat{w}_t^2] \right) R_t$$

Les règles de calcul de Itō s'appliquent alors classiquement et pour une fonction $f \in \mathcal{C}^2$ sur $SO(n)$, on a le lemme d'Itō:

$$df(R_t) = \sum_i \partial_i f(R_t) x^i dt + \frac{1}{2} \sum_{i,j} \partial_{i,j} f(R_t) \mathbb{E}[dw^i dw^j] \qquad \text{with } \partial_i f = \frac{df}{d\sigma_i}.$$

Généralement, les processus stochastiques seront définis avec une intégrale de Stratonovich mais l'évolution de sa distribution par exemple, se fera à l'aide du lemme d'Itō.

En plus de la définition des processus stochastiques sur $SO(n)$, on peut chercher à manipuler numériquement ces processus. La première étape pour cela consiste à être capable de simuler ces derniers.

Cela peut aussi être utile pour la réalisation de certains algorithmes, à l'image du refroidissement contrôlé de la Section 1.3.

Parce qu'il est numériquement impossible de générer numériquement des processus à temps continu, une simulation sera une chaine (donc un processus à temps discret) que

l'on cherche à approcher d'un échantillonnage du processus à simuler. On note δt le pas de discrétisation de la simulation. Même si l'on considère ici un pas de discrétisation constant pour plus de simplicité, rien n'empêche de considérer un pas variable.

Pour commencer, considérons un processus $y_t \in \mathbb{R}^d$ défini comme

$$dy_t = x_t dt + dw_t$$

avec w_t un mouvement Brownien sur $\mathbb{R}^d$. Par définition de dy_t, le terme $\delta y_{n\delta t} = y_{(n+1)\delta t} - y_{n\delta t}$ s'exprime

$$\delta y_{n\delta t} = \int_{n\delta t}^{(n+1)\delta t} x_s ds + dw_s.$$

échantillonner y_t avec un pas δt revient alors à calculer $y_{n\delta t}$.

Dans la pratique, x_t n'est généralement pas connu. Par exemple, seul un échantillonnage de x_t est connu. Dans ce cas, le processus x_t est approximé par un processus constant par morceaux durant le temps δt. Le terme $\delta y_{n\delta t}$ est alors approximé par:

$$\delta y_{n\delta t} = \int_{n\delta t}^{(n+1)\delta t} x_s ds + dw_s = x_{n\delta t}\delta t + o(\delta t) + \delta w_{n\delta t}.$$

Définissons par ailleurs la chaîne $\tilde{y}_n$, définie par:

$$\delta \tilde{y}_n = \tilde{y}_{n+1} - \tilde{y}_n = x_{n\delta t}\delta t + \delta w_{n\delta t}.$$

En prenant $\tilde{y}_0 = y_0$, la chaîne est bien définie.

Il est alors possible de montrer la proposition suivante:

Proposition 21. *La chaîne $\tilde{y}_n$ converge localement uniformément au sens de l'erreur moyenne quadratique, vers y_t. Autrement dit:* $\lim_{\delta t} \mathbb{E}\left[\sup_{n\delta t \leq t} |\tilde{y}_n - y_{n\delta t}|^2\right] = 0$

Cela signifie que lorsque le pas de discrétisation δt diminue, la chaîne $\tilde{y}_n$ devient alors une bonne approximation du processus y_t et $\tilde{y}_n$ peut alors être utilisée pour simuler le processus y_t sur un temps fini. Cette méthode de simulation est appelée schéma d'Euler.

Cependant, même si cette méthode semble intuitive, appliquer directement cette méthode dans le cas de processus R_t sur $SO(n)$ avec bruit multiplicatif n'est pas possible. Cela vient de l'opérateur de différence $\delta.$. Comme $SO(n)$ n'est pas stable par addition, la chaîne simulée avec cette méthode $\tilde{R}_n$ ne serait plus constituée d'élément de $SO(n)$.

En regardant de plus près le schéma d'Euler, on comprend que l'opérateur $\delta.$ considère R_t comme un processus de l'espace ambiant $\mathbb{R}^{n\times n}$ et ne tient pas compte de la structure de $SO(n)$. La solution consiste alors à adapter le schéma d'Euler pour prendre en considération la géométrie de $SO(n)$. L'idée du schéma d'Euler est de considérer x_t constant durant un temps δt. L'incrément entre $\tilde{R}^n$ et $\tilde{R}^{n+1}$ (donné dans le cas de processus sur $\mathbb{R}^d$ par l'opérateur $\delta.$) est alors donné par la géodésique partant de $\tilde{R}^n$ dans la direction $x_{n\delta t}$ durant un temps δt. Autrement dit, on souhaite construire la chaîne:

$$\tilde{R}_{n+1} = \exp\left(\hat{x}_{n\delta t}\delta t + \delta \hat{w}_{n\delta}\right)\tilde{R}_n. \tag{A.2}$$

Ce méthode de simulation s'appelle méthode de MacKean [McKean 1960].

Dans ce cas, il apparaît clairement que si $\tilde{R}_0 \in SO(n)$, alors la chaîne au complète est dans $SO(n)$. Il est à remarquer l'utilisation de l'application exp, qui fait que les incréments se font le long de géodésiques de $SO(n)$.

De manière similaire au schéma d'Euler, on peut prouver la convergence de la chaîne pour un pas de calcul qui tend vers 0.

Proposition 22. *La chaîne $\tilde{R}_n$ converge localement uniformément vers R_t au sens de l'erreur moyenne quadratique.*

La méthode de MacKean sera la méthode utilisée pour simuler des processus stochastiques sur $SO(n)$. Plus le pas de simulation δt sera petit, plus précis sera la simulation. Il est remarquable que la méthode de MacKean peut se généraliser à d'autres variétés que $SO(n)$, dès lors que les géodésiques sont connues. Cependant, le fait d'être capable de simuler des processus stochastiques sur $SO(n)$ ne permet pas de connaître l'évolution de leur distribution ou bien l'existence d'une distribution stationnaire. Il est possible de connaître certaines de ces propriétés dans un cas particulier de processus, les processus Markovien.

Un processus stochastique R_t est dit Markovien si pour tout ensemble $A \in SO(n)$:

$$\mathbb{P}(R_t \in A|\mathcal{F}_s) = \mathbb{P}(R_t \in A|R_s)$$

où $\mathcal{F}_s = \{R_r, r \leq s\}$. Cela signifie que l'évolution du processus ne peut être déduit que de la dernière observation. Les observations précédentes n'apportent aucune information supplémentaire.

Un processus Markovien est appelé stationnaire si $\mathbb{P}(R_t \in A|R_s) = \mathbb{P}(R_{t-\tau} \in A|R_{s-\tau})$. Cela signifie que l'évolution de R_t dépend de la différence de temps avec la dernière observation et non pas du temps de la dernière observation lui-même.

Pour un processus Markovien R_t, on définit l'opérateur linéaire T_t:

$$[T_t f](r) = \mathbb{E}[f(R_t)|R_0 = r] \tag{A.3}$$

pour toute fonction $f \in \mathcal{C}^2(SO(n) \to \mathbb{R})$.

Proposition 23. *L'ensemble des opérateurs T_t forme un semi-groupe d'opérateurs avec $T_{t+s} = T_t T_s$. L'élément neutre est $T_0 = Id$.*

Lorsque la limite existe, on définit l'opérateur $\mathcal{A}$ comme:

$$\lim_t \frac{T_t f - T_0 f}{t} = \mathcal{A} f.$$

Cet opérateur est linéaire et commute avec tout élément T_t. A cause des ces propriétés et de sa définition, T_t est généralement écrit $T_t = e^{t\mathcal{A}}$. L'opérateur $\mathcal{A}$ est appelé le générateur Markovien de R_t. Il correspond à la matrice de transition dans le cas de chaînes de Markov à espace d'états finis. Il est utilisé pour décrire l'évolution infinitésimal de la distribution de R_t.

Proposition 24. *Pour un processus de Markov R_t stationnaire défini via la SDE:*

$$dR_t = (\hat{x}_t dt + \circ d\hat{w}_t) R_t,$$

le générateur $\mathcal{A}$ peut s'écrire

$$\mathcal{A}f(r) = \sum_i x^i(t)\partial_i f(r) + \frac{1}{2}\sum_{i,j} \partial_{i,j} f(r)\mathbb{E}[dw_t^i dw_t^j]. \tag{A.4}$$

Le générateur conduit aux équations vers l'avant et vers l'arrière. En notant $u(t,r) = [T_t f](r)$, l'évolution de u est donné par l'équation vers l'arrière

$$\begin{aligned} \frac{du}{dt} &= \lim_h \frac{T_{t+h}f - T_t f}{h} = \lim_h T_t \frac{T_h f - f}{h} \\ \frac{du}{dt} &= \mathcal{A}u \end{aligned} \tag{A.5}$$

Pour le produit scalaire $< f, p >= \int f(R)p(R)dR$ on $L^2(SO(n) \to \mathbb{R})$, l'opérateur $\mathcal{A}^\star$ est l'opérateur adjoint à $\mathcal{A}$ vis-à-vis de $< ., . >$, *i.e*

$$< \mathcal{A}f, p >=< f, \mathcal{A}^\star p > \qquad \text{pour tout } f, p \in L^2(SO(n) \to \mathbb{R}).$$

Proposition 25. *L'évolution de p_t, la densité de R_t, est donnée par l'équation vers l'avant*

$$\frac{dp_t(R)}{dt} = \mathcal{A}^\star p_t(R) \tag{A.6}$$

Dans le cas d'un générateur $\mathcal{A}$ décrit dans l'équation (A.4), le générateur adjoint s'exprime

$$\mathcal{A}^\star p_t = -\sum_i \partial_i \left[p_t x_t^i\right](x) + \frac{1}{2}\sum_{i,j} \partial_{i,j}\left[p_t \mathbb{E}[dw_t^i dw_t^j]\right] \tag{A.7}$$

L'équation vers l'avant, aussi appelée équation de la chaleur, est un outil très puissant pour l'étude des processus Markovien. Elle décrit l'évolution de la distribution de R_t à l'aide d'opérateurs linéaires. Il est donc possible d'utiliser les outils et résultats d'algèbre linéaire pour résoudre certains problèmes liés à ces processus.

Une distribution $p^\star$ est dite stationnaire si $p_0 = p^\star$ implique $p_t = p^\star$ pour tout $t \geq 0$. Autrement dit, la distribution du processus ne change pas au cours du temps si la distribution initiale est une distribution stationnaire. Dans ce cas, comme p_t reste constant, il est possible, grâce à l'équation vers l'avant, de déterminer l'existence de telles distributions en trouvant les solutions de l'équation $\mathcal{A}^\star p^\star = 0$. En dehors des contraintes de normalisation pour les distributions, les distributions stationnaires sont les vecteurs propres de $\mathcal{A}$ associées à une valeur propre nulle. Ces distributions sont importantes car elles peuvent être, sous certaines conditions, la limite de la distribution p_t lorsque $t \to \infty$. Ceci est utilisé dans la section 1.3 pour construire un processus dont la distribution stationnaire sera fixé à l'avance. Après un temps de simulation assez long, on pourra alors considérer que le processus peut être vu comme une variable aléatoire distribué selon la distribution prévue à l'avance. Cette méthode peut par ailleurs être étendue pour résoudre certains problèmes d'optimisation sur $SO(n)$.

A.2.3 Exemples d'équations différentielles stochastiques

La Section 1.3 présente deux exemples différents de processus stochastiques. Pour chacun d'entre eux, une simulation numérique est réalisée, basée sur le schéma de MacKean. Ces processus sont Markoviens et l'expression du générateur est donnée pour chacun d'entre eux. Le premier exemple est le mouvement Brownien. Le second présente une méthode d'échantillonnage pour une distribution donnée.

Le mouvement Brownien B_t sur $SO(n)$ est défini par:

$$dB_t = (\circ d\hat{w}_t)\, B_t \qquad \text{avec } \hat{w} = \sum_i w_i \sigma^i$$

où w est un mouvement Brownien sur $\mathbb{R}^{n(n+1)/2}$ avec une matrice de covariance $\sigma_w I_{n(n+1)/2}$.

Pour simuler B_t, on utilise la chaîne $\tilde{B}_n$ définie avec le schéma de MacKean. Pour un pas de simulation δt, on a

$$\tilde{B}_{n+1} = \exp\left(\delta \hat{w}_{n\delta t}\right) \tilde{B}_n$$

On constate alors que contrairement à ce qu'il se passerait dans le cas d'un processus sur $\mathbb{R}^d$, l'élément $\tilde{B}_n$ ne peut pas se calculer avant de calculer $\tilde{B}_{n-1}$. Pour générer un élément de la chaîne, tous les éléments précédents doivent être générés eux aussi. Il faut par ailleurs

garder à l'esprit que le terme $\exp(\delta\hat{w}_{n\delta t})$ est l'approximation de rotations infinitésimales durant un temps δt. Le terme $\delta\hat{w}_{n\delta t}$ étant une variable aléatoire Gaussienne, on peut alors se demander si d'autres distributions pour les incréments conduisent au même résultat, comme c'est le cas pour des processus sur $\mathbb{R}^d$ grâce au théorème central limite.

Pour cela, il faut d'abord montrer que B_t est un processus de Markov. Cette propriété s'obtient directement à partir de l'indépendance des incréments de w_t.

Son générateur est alors donné par

$$\mathcal{A}f(r) = \frac{\sigma_w^2}{2}\sum_i \partial_{i,i} f(r) = \frac{\sigma_w^2}{2}\Delta f(r).$$

On peut, à partir de $\mathcal{A}$, déterminer l'existence de distributions stationnaires. Une distribution stationnaire $p^\star$ est solution de l'équation $\mathcal{A}^\star p^\star = 0$. De l'expression de $\mathcal{A}$, on ne trouve qu'une seule distribution stationnaire, la distribution uniforme sur $SO(n)$. Il est à noter que la distribution uniforme existe sur $SO(n)$ car ce dernier est compact. Pour un mouvement Brownien sur $\mathbb{R}^d$, il n'existe aucune distribution stationnaire, le processus se diffusant dans tout l'espace.

L'expression du générateur $\mathcal{A}$ peut aussi nous servir à prouver un équivalent du théorème central limite pour les variables aléatoires sur $SO(n)$. Pour une valeur de t et pour une valeur de δ fixées, on définit la chaîne R_n^δ comme

$$R_{n+1}^\delta = \exp(\sqrt{t\delta}\hat{\zeta}_n)R_n^\delta \tag{A.8}$$

avec les variables aléatoires $\zeta_k \in \mathbb{R}^{n(n-1)/2}$ centrées avec une variance unitaire. Le Théorème 6 prouve que pour $\delta = 1/n$, la chaîne R^δ va tendre vers celle du mouvement Brownien sur $SO(n)$.

Theorem 6. *La limite* $\lim_n R_n^{1/n} = B_t$ *for* $R^{1/n}$ *définie à l'équation (A.8) est un mouvement Brownien, au sens des distributions.*

Proof. Le Théorème 19.25 de [Kallenberg 2002] précise qu'il est suffisant pour montrer la convergence en distribution, de montrer la convergence des générateurs. Autrement dit, il est suffisant de montrer que le générateur de la chaîne $\mathcal{A}_n$ défini comme étant

$$\mathcal{A}_n f\,(r) = \frac{1}{t/n}\mathbb{E}[f\left(R_{n+1}^{1/n}\right) - f\left(R_n^{1/n}\right)|R_n^{1/n} = r]$$

converge vers le générateur du mouvement Brownien $\mathcal{A}$ où $\mathcal{A}f = \frac{1}{2}\Delta f$ pour toute fonction $f \in \mathcal{C}^2(SO(n))$

Un développment au second ordre de $f\left(R_{n+1}^{1/n}\right)$ en $R_n^{1/n}$ donne

$$f\left(R_{n+1}^{1/n}\right) = f\left(R_n^{1/n}\right) + \sum_i \partial_i f\left(R_n^{1/n}\right)\zeta_n^i\sqrt{\frac{t}{n}} + \sum_{i,j}\partial_{i,j} f\left(R_n^{1/n}\right)\zeta_n^i\zeta_n^j\frac{t}{n} + o\left(\frac{1}{n}\right)$$

En fixant $R_n^{1/n} = r$, l'espérance de l'équation précédente est alors

$$\mathbb{E}[f\left(R_{n+1}^{1/n}\right) - f\left(R_n^{1/n}\right)|R_n^{1/n} = r] = \sum_i \partial_{i,i} f(r)\frac{t}{n} + o\left(\frac{1}{n}\right)$$

car les variables aléatoires ζ sont centrées avec une variance unitaire.

Par conséquent, l'opérateur $\mathcal{A}_n f$ converge vers

$$\begin{aligned}\mathcal{A}_n f(r) &= \frac{1}{t/n}\mathbb{E}[f\left(R_{n+1}^{1/n}\right) - f\left(R_n^{1/n}\right)|R_n^{1/n} = r] \\ &= \frac{1}{t/n}\sum_i \partial_{i,i} f(r)\frac{t}{n} + o\left(\frac{1}{n}\right) \\ &= \sum_i \partial_{i,i} f(r) + o(1) \\ &\to \Delta f(r) \\ &= \mathcal{A} f(r)\end{aligned}$$

Le générateur de B_t est donc le Laplacien, ce qui signifie que B_t est un mouvement Brownien. □

Ce théorème nous donne donc une version du théorème central limite adapté au cas de variables aléatoires de rotations sur $SO(n)$.

Un autre exemple de processus stochastique est donné dans le paragraphe 1.3.2. De nombreux algorithmes stochastiques ont besoin de variables aléatoires, que l'on note ici R, échantillonnées à partir d'une distribution donnée $p^\star$. Une méthode Monte-Carlo peut être utilisée pour obtenir R en construisant un processus dont la distribution stationnaire est $p^\star$ et en considérant $R = R_T$ pour un temps T suffisamment grand. En supposant que la distribution de R_t va converger vers $p^\star$, la proposition ci-dessous donne une méthode pour construire un tel processus.

Proposition 26. *Étant donné une fonction $f \in \mathcal{C}^1(SO(n) \to \mathbb{R})$, on définit R_t par*

$$dR_t = \left(\widehat{\text{grad}} f(R_t)dt + \circ d\hat{w}_t\right) R_t \tag{A.9}$$

avec w_t un mouvement Brownien avec variance $\sigma_w^2 I_{n(n-1)/2}$ et grad *l'opérateur gradient.*

Ce processus a alors une distribution stationnaire $p^\star$ qui s'exprime (en dehors de toute normalisation)

$$p^\star = \exp(2f/\sigma_w^2). \tag{A.10}$$

Pour échantillonner à partir de la distribution $p^\star$, il suffit alors de définir un processus R_t avec la SDE (A.9) où $f = \frac{\sigma_w^2}{2}\log(p^\star)$. Après avoir simulé R_t pendant un temps T suffisamment long, on peut considérer que $R_T \sim p^\star$.

En prenant $f(R) = \text{trace}(R)$ par exemple, le processus va tendre vers une distribution stationnaire autour de I_n. D'un point de vue intuitif, cela vient du fait que la dérive va tendre le processus à se concentrer vers les maximaux locaux de f. En l'absence de bruit, on obtient l'algorithme classique de descente de gradient. Le bruit présent va tendre à faire se déplacer le processus de manière aléatoire et tend à contre balancer la dérive et à éloigner le processus du point d'équilibre.

Il est de plus remarquable que les hautes valeurs de f coïncident avec les hautes valeurs de $p^\star$. Dans ce cas, le processus R_t peut être utilisé pour déterminer les maximaux de f en regardant à quels endroits le processus tend à se distribuer.

Le refroidissement controlé est une méthode Monte-Carlo d'optimisation utilisant ce principe. Le but est de trouver le maximum d'une fonction $f \in \mathcal{C}^1(SO(n) \to \mathbb{R})$. Pour faire cela, la méthode classique du gradient peut s'adapter au cas de fonction définies sur $SO(n)$. Cette méthode ne fonctionne cependant que si f est convexe. Dans le cas contraire,

la descente de gradient va se stabiliser vers un maximum local. C'est pour éviter ce problème que l'on utilise une méthode Monte-Carlo.

Pour un processus R_t défini par (A.9), la distribution stationnaire $p^\star$ est donnée par l'équation (A.10). D'après l'expression de $p^\star$, lorsque σ_w tend vers 0, $p^\star$ va se concentrer uniquement autour du maximum de f. Le but du refroidissement controlé est alors de construire un tel processus pour lequel la variance du bruit σ_w^2 va diminuer lentement au cours du temps. Il est impératif que la variance diminue lentement afin de laisser le temps au processus à chaque fois d'atteindre la distribution stationnaire et pour laisser le temps au processus de visiter les différentes zones convexes de f. Par ailleurs, si σ_w^2 varie trop rapidement, les résultats concernant les processus Markovien ne s'appliqueront plus. En pratique, une décroissance logarithmique est généralement utilisée.

Il est à noter que cette méthode requiert la connaissance du gradient de f pour construire R_t. Si ce dernier n'est pas connu, il est donc impossible d'utiliser directement la méthode présentée. Un algorithme de type Metropolis-Hasting peut alors être considéré. Cet algorithme n'utilise pas de gradient mais génère plusieurs candidats pour une nouvelle position avant de déplacer le processus.

A.2.4 Représentation irréductible de $SO(n)$

La Section 1.4 sert d'introduction rapide à la représentation irréductible de variables aléatoires sur $SO(n)$. Les résultats présentées dans cette section sont déjà connus et plus de détails peuvent se trouver dans [Kyatkin 2001] et [Faraut 2008].

Une représentation du groupe $SO(n)$ est un homomorphisme $\lambda : SO(n) \to GL(V)$ où V est un espace de Hilbert. La représentation est dite réductible s'il existe un sous-espace invariant $W \in V$ (autrement dit, $\lambda_R(W) \in W$) pour tout élément $R \in SO(n)$. Une représentation irréductible est une représentation qui n'est pas réductible, c'est-à-dire que les seuls sous-espaces invariants sont triviaux, *i.e* $\{0_V\}$ et V.

Deux représentations irréductibles $\lambda_i : SO(n) \to GL(V_i)$ pour $i = 1, 2$ sont dites équivalentes s'il existe une application bijective $L : V_1 \to V_2$ telle que $L \circ \lambda_1(R) = \lambda_2(R) \circ L$ for $R \in SO(n)$. Cette relation d'équivalence définit des classes d'équivalence $[\lambda]$ parmi les représentations. $\mathrm{Irr}(SO(n))$ représente l'ensemble des classes d'équivalence. Tous les éléments de $[\lambda]$ ont la même dimension $d_\lambda \times d\lambda$. Il est donc possible de les représenter par une matrice $d_\lambda \times d_\lambda$. Pour une meilleure lisibilité, la classe $[\lambda]$ sera simplement notée λ.

Parmi toutes les représentations de la classe λ, il existe une représentation unitaire $U^\lambda \in L^2(SO(n))$ (que l'on considère comme une matrice) telle que $U^\lambda U^{\lambda T} = I_{d_\lambda}$. Soit λ_0 la notation pour la représentation identité $U^{\lambda_0}(R) = I_{d_{\lambda_0}}$. Parce que U^λ est un homomorphisme de $SO(n)$, cela signifie que pour tout $R_1, R_2 \in SO(n)$

$$U^\lambda(R_1 R_2) = U^\lambda(R_1) U^\lambda(R_2).$$

Pour une fonction $h \in L^2(SO(n))$, on définit $\hat{h}^\lambda$ la matrice dont les éléments $\hat{h}^\lambda_{m,n}$ pour $-l \leq m, n \leq l$, sont définis par

$$\hat{h}^\lambda_{m,n} = < h, U^\lambda_{m,n} > = \int_{SO(n)} h(R) U^\lambda(R) dR.$$

Les éléments de la matrice $\hat{h}^\lambda$ sont appelés coefficients de Fourier de h. Tout comme dans le cas de fonctions périodiques, il est possible de reconstruire la fonction à partir de ses coefficients de Fourier.

Theorem 7. *Toute fonction $f \in L^2(SO(n), \mathbb{R})$ peut être reconstruite à partir de ses coef-*

ficients de Fourier via la formule de reconstruction

$$h(R) = \sum_{\lambda \in \mathrm{Irr}(SO(n))} d_\lambda \operatorname{trace}\left(\hat{h}^\lambda U^\lambda(R)\right). \tag{A.11}$$

La série converge uniformément.

Pour une variable aléatoire $R \in SO(n)$, on définit la fonction caractéristique ϕ_R de R comme

$$\phi_R(\lambda) = \mathbb{E}[U^\lambda(R)], \qquad \lambda \in \mathrm{Irr}(SO(n)).$$

Si la distribution de R peut s'écrire avec une densité p, alors $\phi_R(\lambda) = \hat{p}^\lambda$ avec $\hat{p}^\lambda$ les coefficients de Fourier de p. On trouve alors les propriétés suivantes

Proposition 27. *Pour $R_1, R_2 \in SO(n)$ deux variables aléatoires, on a*

1. *Si R_1, R_2 sont indépendantes, alors*

$$\phi_{R_1 R_2}(\lambda) = \phi_{R_1}(\lambda)\phi_{R_2}(\lambda)$$

2. *Si $\phi_{R_1} = \phi_{R_2}$, alors R_1 et R_2 sont égales au sens des distributions.*

La première partie de la proposition est utile par exemple pour l'étude d'un produit de variables aléatoires indépendantes identiquement distribuées. La seconde partie peut servir pour démontrer l'égalité de deux variables aléatoires au sens des distribution, ou bien la convergence d'une séquence de variables aléatoires. De manière similaire au cas de variables aléatoires réelles, leur fonction caractéristique s'annule pour $\lambda \neq \lambda_0$ pour une distribution uniforme.

Proposition 28. *Si R est distribué uniformément sur $SO(n)$, alors $\phi_R(\lambda) = 0$ pour $\lambda \neq \lambda_0$.*

Cette proposition montre par ailleurs que la distribution uniforme est un élément absorbant de $L^2(SO(n))$. Si R_1 est uniformément distribué, alors $R_1 R_2$ sera aussi uniformément distribué pour toute variable aléatoire R_2 indépendante de R_1.

Il est possible d'appliquer ces résultats aux processus stochastiques vus auparavant. Soit B_t un mouvement Brownien sur $SO(n)$ avec une variance unitaire. On peut utiliser les représentations irréductibles pour décrire la distribution p_t de B_t à l'aide de ses coefficients de Fourier.

Proposition 29. *Les coefficients de Fourier $\hat{p}_t^\lambda$ de p_t sont donnés par*

$$\hat{p}_t^\lambda = e^{-\chi_l t/2} I_{d_\lambda}.$$

Cette proposition permet aussi de retrouver le fait que la distribution stationnaire de B_t est la distribution uniforme sur $SO(n)$. On trouve cependant un résultat supplémentaire car nous sommes désormais capable de montrer que p_t tend vers la distribution stationnaire. Les résultats sur les fonctions caractéristiques peuvent aussi servir à démontrer avec une approche différente le théorème central limite dans le cas de variables aléatoires sur $SO(n)$.

A.3 Filtrage pour observations complètes et partielles sur $SO(n)$

Le Chapitre 2 traite du problème de filtrage avec observations sur $SO(n)$ et avec observations sur les variétés de Stiefel $V_{n,k}$. Dans le problème du filtrage, on cherche à estimer

l'état d'un système à partir d'observations bruitées. La Section 2.1 présente le formalisme et le filtrage dans le cas d'observation dans $\mathbb{R}^n$. Une solution générale est donnée, basée sur les règles de Bayes pour décrire la distribution de l'état à estimer, conditionnée par les observations. On suppose que les observations sont modélisées avec un bruit additif. Une solution adaptative est alors dérivée de la solution générale pour faire de l'estimation en temps réel. Dans le cas d'un état à estimer qui suit un modèle linéaire, on retrouve le filtre de Kalman Bucy [Bucy 1961].

La Section 2.2 étend le modèle précédent pour des observations sur $SO(n)$. Contrairement à la Section 2.1, le bruit n'est plus additif. Le modèle utilisé pour l'observation est basé sur les SDE de la Section 1.2. L'idée principale présentée dans cette section est la construction de l'antidéveloppement. Il s'agit d'un processus stochastique équivalent, en terme d'informations, au processus observé mais qui se modélise avec un bruit additif. Le problème est alors réduit à un problème de filtrage avec bruit additif. La solution de la Section 2.1 est alors applicable. Par ailleurs, l'observation du processus de rotation se fait de manière discrète. Différentes méthodes d'interpolations sont présentées et comparées en terme de complexité de calcul et en terme de précision de l'approximation. Malgré le fait que ces interpolations ajoutent une erreur dans l'estimation, cette erreur est négligeable pour un faible pas d'observation. La question du choix de la fonction d'interpolation se trouvait déjà dans [Pontier 1988] par exemple. Cependant, aucune comparaison en terme de complexité et précision n'ont été réalisé auparavant.

La Section 2.3 décrit les variétés de Stiefel $V_{n,k}$ comme des observations partielles d'un processus de $SO(n)$. Le problème de filtrage avec observations sur $V_{n,k}$ peut alors être utilisé pour modéliser le filtrage de la Section 2.2 avec des capteurs défectueux par exemple. La décomposition de l'espace tangent de $SO(n)$ en espace horizontal et vertical, définis à partir de la géométrie de $V_{n,k}$, mène à la construction du relèvement horizontal, un processus dans $SO(n)$, et à la construction de l'antidéveloppement, un processus avec bruit additif. A cause du manque d'observations, une méthode numérique doit être employée pour estimer la distribution conditionnelle. Une méthode Monte-Carlo basée sur un filtre à particule est implémentée et présentée. Cette technique utilise les particules comme candidats pour la trajectoire du processus et calcule la vraisemblance avec les observations sur $V_{n,k}$. On montre que cette méthode mène à une estimation partielle si l'état à estimer évolue trop rapidement. Si ce dernier peut se modéliser par des états constants par morceaux, il est possible d'avoir une estimation complète de ce dernier.

A.3.1 Filtrage dans $\mathbb{R}^n$

Le problème du filtrage consiste à estimer l'état d'un système à partir d'observations bruitées. Cet état à estimer peut être une position ou une vitesse dans le cadre de la mécanique classique, etc. Cet état est mesuré à l'aide de procédés sujets à des bruits de mesures ou d'interférences d'autres systèmes. On dénote x_t l'état à estimer au temps t. On suppose que $x_t \in \mathbb{R}^d$ est un processus stochastique défini par la SDE:

$$dx_t = Fx_t dt + db_t$$

où $b_t \in \mathbb{R}^d$ est un mouvement Brownien avec une variance σ_b^2. Dans ce cas, x_t est un processus Markovien. On note $\mathcal{A}$ son générateur et p_0 sa distribution à priori.

On considère dans ce paragraphe que les observations y_t sont des éléments de $\mathbb{R}^n$ pour un n fixé. L'état x_t est mesuré via la fonction $H : \mathbb{R}^d \to \mathbb{R}^n$ que l'on considère connue. On suppose que le bruit d'observation peut se modéliser par un bruit de la façon suivante [Jazwinski 1970]

$$dy_t = H(x_t)dt + dw_t \tag{A.12}$$

avec w_t un mouvement Brownien de variance σ_w^2.

A cause du bruit, il est impossible de retrouver directement x_t à partir de y_t. On peut cependant chercher sa distribution conditionnelle $\pi_[0,t]$ de x étant donné les observations $\mathcal{Y}_t = \{y_s,\ s \leq t\}$. A partir de $\pi_{[0,t]}$, on peut construire aisément un estimateur optimal de x_t, à savoir $\mathbb{E}[x_t|\mathcal{Y}_t]$.

Exprimer directement $\pi_{[0,t]}$ peut être complexe car il (A.12) nous permet d'exprimer facilement la distribution de y_t conditionné par x_t et non l'inverse. L'idée est d'alors d'utiliser la règle de Bayes pour exprimer $\pi_{[0,t]}$ à partir d'une vraisemblance.

Tout d'abord, considérons la chaîne discrète $\tilde{y}_n$

$$\delta\tilde{y}_n = H(x_i)\delta t + \delta w_n, \tag{A.13}$$

Soit $\pi_n^{\delta t}$ la distribution de $(x_0, x_{\delta t}, ..., x_{n\delta t})$ conditionnée par $\tilde{\mathcal{Y}}_n = \{\tilde{y}_k, k \leq n\}$. La règle de Bayes appliquée à $\pi_n^{\delta t}$ donne

$$\pi_n^{\delta t}(\zeta_0, ..., \zeta_n) = \frac{p_X(\zeta_0, ..., \zeta_n) L_{\delta t}(\tilde{\mathcal{Y}}_n|\zeta_0, ..., \zeta_n)}{p_Y(\tilde{\mathcal{Y}}_n)} \tag{A.14}$$

où p_X (respectivement p_Y) est la distribution de $(x_0, ..., x_{n\delta t})$ (respectivement $\tilde{\mathcal{Y}}_n$) et $L_{\delta t}$ la vraisemblance de $\tilde{\mathcal{Y}}_n$ conditionné par X_n. Comme les variables aléatoires δw_n sont indépendantes et identiquement distribuées, $L_{\delta t}$ peut s'écrire

$$L_{\delta t}(\tilde{\mathcal{Y}}_n|\zeta_0, ..., \zeta_n) = \prod_{k \leq t/\delta t} l^{\delta t}(\delta\tilde{y}_k|\zeta_k) \tag{A.15}$$

avec $l^{\delta t}(\delta\tilde{y}_k|\zeta_k)$ une Gaussienne centrée en $H(\zeta_k)\delta t$, avec une variance $\sigma_w^2\delta t$.

Par ailleurs, en écrivant p_Y comme l'espérance de la distribution jointe de x et $\tilde{y}$, on a

$$\begin{aligned} p_Y(\tilde{\mathcal{Y}}_n) &= \int p_X(x'_0, ..., x'_n) L_{\delta t}(\tilde{\mathcal{Y}}_n|x'_0, ..., x'_{n\delta t}) dx'_0 ... dx'_{n\delta t} \\ &= \mathbb{E}[L_{\delta t}(\tilde{\mathcal{Y}}_n|x'_0, ..., x'_{n\delta t})|\tilde{\mathcal{Y}}_n] \end{aligned}$$

où x' est une copie de x (comprendre avec la même distribution) mais indépendant de l'observation y. On peut donc exprimer $\pi_n^{\delta t}$ uniquement à l'aide d'espérances. En considérant une fonction test $\phi \in \mathcal{C}^2\left(\mathbb{R}^{d\times n} \to \mathbb{R}\right)$, on a

$$\pi_n^{\delta t}(\phi) = \frac{\rho_n^{\delta t}(\phi)}{\rho_n^{\delta t}(1)}$$

où $\pi_n^{\delta t}(\phi) = \mathbb{E}\left[\phi(x'_0, ..., x'_{n\delta t})\right]$ et $\rho_n^{\delta t}(\phi) = \mathbb{E}\left[\phi(x'_0, ..., x'_{n\delta t}) L_{\delta t}(\tilde{\mathcal{Y}}_n|x'_0, ..., x'_{n\delta t})|\tilde{\mathcal{Y}}_n\right]$.

On a vu dans le Chapitre 1 que la chaîne $\tilde{y}_n$ approche y_t pour $t = n\delta t$ lorsque δt tend vers 0. On peut montrer [Jazwinski 1970] que cela est encore le cas pour l'équation précédente et que pour δt qui tend vers 0, on trouve, pour toute fonction test $\phi \in \mathcal{C}^2$,

$$\pi_{[0,t]}(\phi) = \frac{\rho_{[0,t]}(\phi)}{\rho_{[0,t]}(1)}, \tag{A.16}$$

avec $\pi_{[0,t]}(\phi) = \mathbb{E}[\phi(x'_{[0,t]})|\mathcal{Y}_t]$, $\rho_{[0,t]}(\phi) = \mathbb{E}\left[\phi(x'_{[0,t]}) L_t(y|x')|\mathcal{Y}_t\right]$ et avec la vraisemblance L_t définie par

$$L_t(y|x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < H(x'_s), dy_s > -\frac{1}{2\sigma_w^2}\int_0^t ||H(x'_s)||^2 ds\right) \tag{A.17}$$

L'équation A.16 est généralement appelée formule de Kallianpur-Striebel. Malgré le fait que cette formule donne la solution générale du problème de filtrage, elle n'est pas adaptative.

Comme le dénominateur dans (A.14) est fixé par l'observation, on ne peut chercher qu'une forme adaptative pour la distribution conditionnelle non normalisée $\rho_[0,t]$.

Pour la chaîne $\tilde{y}_n$ définie précédemment dans ce paragraphe, on a

$$\rho_n^{\delta t}(\zeta_0,...,\zeta_n) = p_X(\zeta_0,...,\zeta_n) L_{\delta t}(\tilde{\mathcal{Y}}_n|\zeta_0,...,\zeta_n).$$

En replaçant le terme $L_{\delta t}(\tilde{\mathcal{Y}}_n|\zeta_0,...,\zeta_n)$ par sa décomposition en produit, on a

$$\rho_n^{\delta t}(\zeta_0,...,\zeta_n) = p_X(\zeta_0,...,\zeta_n) \prod_k l^{\delta t}(\delta\tilde{y}_k,\zeta_k) \tag{A.18}$$

Comme x_t est un processus de Markov, la chaîne $x_{n\delta t}$ est une chaîne de Markov avec un noyau de transition que l'on note $q_{\delta t}$. On peut alors montrer que

$$\rho_n(\zeta_0,...,\zeta_n) = q_{\delta t}(\zeta_n,\zeta_{n-1}) l^{\delta t}(\delta\tilde{y}_n,\zeta_n)\rho_{n-1}(\zeta_0,...,\zeta_{n-1}).$$

En marginalisant cette expression vis-à-vis des $n-1$ premières variables et en exprimant $q_{n\delta t}$ à l'aide du générateur adjoint de x_t, $\mathcal{A}^\star$, on a

$$\rho'_n(\zeta) = \exp(\mathcal{A}^\star \delta t)\, \rho'_{n-1}(\zeta) l^{\delta t}(\delta y_n,\zeta)$$

Un développement au premier ordre de la fonction exp donne

$$\rho'_n(\zeta) = \rho_{n-1}(\zeta) + \mathcal{A}^\star\rho'_{n-1}(\zeta)\delta t + \frac{1}{\sigma_w^2}\rho'_{n-1}(\zeta) < H(\zeta), \delta\tilde{y}_n > + o(\delta t).$$

En considérant la limite pour δt qui tend vers 0, on trouve alors l'équation de Zakai [Zakai 1969]

$$d\rho_t(\zeta) = \mathcal{L}^\star\rho_t(\zeta)dt + \frac{1}{\sigma_w^2}\rho_t(\zeta) < H(\zeta), dy_t > . \tag{A.19}$$

Ainsi, lorsque de nouveaux échantillons sont observables (le terme dy_t), l'on peut déterminer l'évolution de ρ_t. Dans la pratique, cela peut cependant s'avérer coûteux d'utiliser l'équation de Zakai directement. Une solution consiste à l'utiliser pour déterminer l'évolution des moments de π_t.

En écrivant π_t comme un quotient de ρ_t avec son intégrale, on a

$$\pi_t = \frac{\rho_t}{\int_x \rho_t(x)dx}.$$

Le calcul d'Itō nous donne alors l'équation dite de Kushner

$$d\pi_t(\zeta) = \mathcal{A}^\star\pi_t(\zeta)dt + \frac{1}{2\sigma_w^2}\pi_t < H(\zeta) - \bar{H}, dy - \bar{H}dt > \tag{A.20}$$

où $\bar{H} = \int_x H(x)\pi_t(x)dx$.

Proposition 30. *Dans le cas d'une fonction H linéaire, on a μ_t, le moment d'ordre un de π_t solution de*

$$d\mu_t = F\mu_t dt + P_t H^T(dy - H\mu_t dt)$$

où $P_t = \int (x-\mu)(x-\mu)^T\pi_t(x)dx$ est le moment centré de second ordre.

Par conséquent, la détermination du moment d'ordre un requiert le moment d'ordre deux. On peut généraliser et montrer que pour déterminer le moment d'odre k, il nous faut connaître le moment d'ordre $k+1$. Cette méthode ne fonctionne donc qu'en négligeant les moments d'ordres supérieurs.

Il est par ailleurs remarquable que dans le cas où H est linéaire, on peut montrer que π_t est une Gaussienne centrée en μ_t avec une variance P_t dont leur évolution est donnée par le filtre de Kalman-Bucy

$$\begin{aligned} d\mu_t &= F\mu_t dt + P_t H^T (dy_t - H\mu_t dt) \\ dP_t &= FP_t + P_t F^T - P_t H H^T P_t + \sigma_b^2 \end{aligned} \tag{A.21}$$

où σ_b^2 est la variance de x_t. Tout comme le filtre de Kalman, la variance est solution d'une équation indépendante des observations.

Dans le cas où F et H ne sont pas linéaires, c'est-à-dire que x_t et y_t sont solutions de

$$\begin{aligned} dx_t &= F(x_t)dt + db_t \\ dy_t &= H(x_t)dt + dw_t, \end{aligned}$$

on peut étendre la solution précédente en faisant une approximation au premier ordre de F et H. Les distributions sont alors presque Gaussiennes et l'on a une solution approchée

$$\begin{aligned} d\mu_t &= F'\mu_t dt + P_t(dy_t - \mu_t dt) \\ dP_t &= F'P_t + P_t {F'}^T - P_t H' {H'}^T P_t + \sigma_B^2 \end{aligned} \tag{A.22}$$

avec $F' = \nabla F(\mu_t)$ and $H' = \nabla H(\mu_t)$. Ce couple d'équation constitue le filtre de Kalman étendu. Contrairement au cas précédent, la variance P_t dépend ici des observations.

A.3.2 Filtrage à partir d'observations sur $SO(n)$, solution et implémentation

La Section 2.1 est consacrée au filtrage de processus définis dans $\mathbb{R}^n$. Le problème de filtrage peut cependant s'étendre au cas d'observations avec valeurs dans $SO(n)$. De tels problèmes apparaissent en mécanique [Lovera 2002] ou en vision par ordinateur [Meer 2007] [Chellappa 2008]. Les principales différences avec le cas présentée dans la Section 2.1 vient de la contrainte sur l'observation, car le processus est contraint à être défini sur $SO(n)$; la seconde différence vient de la manière dont le bruit perturbe la mesure. Dans le cas de processus sur $SO(n)$, le bruit devient multiplicatif alors que ce dernier est additif dans le cas précédent. Malgré le fait qu'il soit possible d'utiliser des solutions de type filtrage de Kalman étendu, les coefficients des équations du filtre dépendent des observations. La méthode présentée dans la Section 2.2 donne une solution optimale avec des coefficients constants.

Un processus de $SO(n)$ étant complètement défini par sa vitesse angulaire, on considère que l'état à estimer, $x_t \in \mathbb{R}^{n(n-1)/2}$ est la vitesse angulaire du processus observé $R_t \in SO(n)$. Malgré le fait que la majeure partie de cette section peut se traiter dans le cas d'un modèle général pour x_t, on se limite cependant au cas d'un modèle linéaire

$$dx_t = Fx_t dt + db_t$$

avec $b_t \in \mathbb{R}^{n(n-1)/2}$ un mouvement Brownien avec une variance σ_b^2.

On considère que le bruit est modélisé de la même manière que dans la Section 1.2 et que R_t est solution de la SDE

$$dR_t = (\hat{x}_t dt + \circ d\hat{w}_t)\, R_t \tag{A.23}$$

où $w_t \in \mathbb{R}^{n(n-1)/2}$ est un mouvement Brownien de variance σ_w^2 indépendant de x_t.

Rechercher à estimer x_t à partir de l'observation $\mathcal{R}_t = \{R_s, s \leq t\}$ est plus compliqué que dans le cas précédent car les incréments de R_t, conditionnés par x_t, ne sont plus indépendants. Ces derniers dépendent de R_t.

La solution présentée dans cette section est basée sur la méthode décrite dans [Lo 1975]. L'idée est de construire un processus z_t à partir de R_t à l'aide d'une transformation bijective telle que z_t soit solution d'une SDE avec un bruit additif. Le problème se trouve alors réduit au cas de la Section 2.1 et les méthodes classiques de filtrage peuvent s'utiliser.

On définit donc le processus $z_t \in \mathbb{R}^{n\times n}$ par

$$z_t = \int_0^t (\circ dR_s) R_s^T. \tag{A.24}$$

En convertissant cette équation avec une intégrale de Itō, on trouve que z_t est solution de

$$dz_t = \hat{x}_t dt + d\hat{w}_t - \frac{\sigma_w^2}{2} dt.$$

On constate donc que contrairement à R_t, le bruit est additif pour z_t. Ce dernier est appelé antidéveloppement de R_t. D'un point de vue intuitif, il peut s'interpréter comme l'accumulation des incréments de R_t, vus à chaque fois dans le plan tangent $T_{R_t}SO(n)$. Il est remarquable qua dans le cas où R_t est une géodésique de $SO(n)$, son antidéveloppement est une géodésique (une ligne droite) de $\mathbb{R}^{n\times n}$.

On a vu que z_t peut s'écrire à partir de R_t. L'inverse est aussi vrai. En effet, il est possible d'écrire R_t comme solution de

$$\begin{aligned} dR_t &= (\circ dz_t) R_t \\ &= (\hat{x}_t dt + \circ d\hat{w}_t)\, R_t. \end{aligned}$$

Cela montre bien le caractère bijective de la transformation entre le processus et son antidéveloppement. Par conséquent, observer l'un ou l'autre de ces processus revient au même en terme d'information. Pour le problème de filtrage, cela signifie que chercher la distribution de x_t conditionnée par $\mathcal{R}_t$, l'observation de R_t, est identique à chercher la distribution de x_t conditionnée par $\mathcal{Z}_t = \{z_s, s \le t\}$.

Le problème est alors réduit à un problème de filtrage avec bruit additif pour lequel l'observation est l'antidéveloppement. Le problème peut donc se traiter indépendamment du fait que $R_t \in SO(n)$.

La solution est alors donnée par la formule de Kallianpur-Striebel appliquée à $\mathcal{Z}_t$ pour une fonction test $\phi \in \mathcal{C}^2$

$$\pi_{[0,t]}(\phi) = \frac{\rho_{[0,t]}(\phi)}{\rho_{[0,t]}(1)}, \tag{A.25}$$

avec $\rho_{[0,t]}(\phi) = \mathbb{E}\left[\phi(x')L_t(z,x')|\mathcal{Z}_t\right]$ et où x'_t est une copie de x_t indépendante de R_t. La vraisemblance L_t est définie par

$$L_t(z,x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < \hat{x}'_s, \circ dz_s > -\frac{1}{2\sigma_w^2}\int_0^t ||\hat{x}'_s||^2 ds\right) \qquad \text{avec } \circ\, dz_t = (\circ dR_t)R_t^T.$$

Comme les matrices symétriques et anti-symétriques sont orthogonales, le terme $< \hat{x}'_s, \circ dz_s > = < \hat{x}'_s, dR_s R_s^T - \sigma_w^2 I_n dt >$ peut être remplacé par $< x'_s, dR_s R_s^T >$. Dans ce cas, la vraisemblance L_t peut s'écrire comme

$$L_t(z,x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < x'_s, dR_s R_s^T > -\frac{1}{2\sigma_w^2}\int_0^t ||x'_s||^2 ds\right). \tag{A.26}$$

Une solution adaptative est donnée par l'équation de Zakai, qui décrit l'évolution de la ditribution non normalisée ρ_t

$$d\rho_t(\zeta) = \mathcal{A}^\star \rho_t(\zeta) dt + \frac{1}{\sigma_w^2}\rho_t(\zeta) < \hat{\zeta}, dz_t > .$$

Le modèle de x_t étant linéaire, π_t est une Gaussienne centrée en μ_t avec une variance P_t, solutions de

$$\begin{aligned} d\hat{\mu}_t &= \widehat{F\mu}_t dt + P_t(dz_t - \hat{\mu}_t dt) \\ dP_t &= FP_t + P_t F^T - P_t^2 + \sigma_b^2. \end{aligned} \tag{A.27}$$

Ces équations diffèrent du filtre de Kalman étendu car la variance P_t par exemple, est solution d'une équation aux coefficients constants. Il ne s'agit pas non plus d'une approximation à l'aide d'une linéarisation mais d'une solution optimale de filtrage.

Cependant, dans la pratique, d'autres problèmes se posent. Cela vient principalement de la discrétisation des observations. Le terme dz_t doit alors être remplacé par un terme différentiel discret δz_t. Ce terme requiert une interpolation de dR_t. Le paragraphe suivant discute des conditions nécessaires à cette interpolation pour faire converger la solution lorsque le pas d'observation δt tend vers 0.

Dans le paragraphe 1.2.2, on a montré qu'un processus continu avec bruit additif peut être discrétisé via le schéma d'Euler. En appliquant ce même schéma à l'antidéveloppement avec un pas δt, on trouve

$$\begin{aligned} \delta z_{k\delta t} &= (R_{(k+1)\delta t} - R_{k\delta t})R_{k\delta t}^T - \sigma_w^2 I_n \delta t \\ &= R_{(k+1)\delta t}R_{k\delta t}^T - I_n - \sigma_w^2 I_n \delta t \\ &= \Delta R_{k\delta t} - I_n - \sigma_w^2 I_n \delta t \qquad \text{avec } \Delta R_{k\delta t} = R_{(k+1)\delta t}R_{k\delta t}^T \end{aligned}$$

Grâce à la continuité de R_t, le terme $R_{(k+1)\delta t}R_{k\delta t}^T$ converge vers I_n lorsque δt tend vers 0 et $\Delta R_{k\delta t}$ est bien dans le domaine de définition de la fonction log. Un développement au premier ordre donne alors

$$\delta z_{k\delta t} = \log(\Delta R_{k\delta t}) + o(\delta t).$$

Ce terme sera approximé en pratique par

$$\delta z_{k\delta t} \approx \log(\Delta R_{k\delta t}). \tag{A.28}$$

Cette interpolation à l'aide de la fonction log entre $R_{k\delta t}$ et $R_{(k+1)\delta t}$ est appelée interpolation géodésique [Pontier 1988]. Dans ce cas, la somme Riemannienne

$$S^\delta = \sum_{k \le t/\delta t} < \hat{x}_{k\delta t}, \delta z_{k\delta t} >$$

converge, lorsque δt tend vers 0, vers

$$\mathcal{I} = \int_0^t < \hat{x}_s, dz_s > .$$

La vraisemblance L_t (A.17) peut alors être approximée par $\tilde{L}_{\delta t}(\tilde{\mathcal{Z}}_{n\delta t})$ de l'équation (A.15).

Dans l'équation (A.28), le terme d'erreur était négligeable vis-à-vis de $\log(\Delta R_{k\delta t})$. Il semble donc naturel de penser que toutes les fonctions suffisamment proches de cette interpolation fonctionneront pour approximer $\delta z_{k\delta t}$.

On s'intéresse donc aux fonctions $\mathrm{Int} : SO(n) \times SO(n) \to \mathfrak{so}(n)$ telles que en considérant $\delta z_k = \mathrm{Int}(R_{k\delta t}, R_{(k+1)\delta t})$, la solution discrète converge vers la solution continue lorsque δt tend vers 0.

Theorem 8. *La somme Rienmanienne*

$$S^\delta = \sum_{k \le t/\delta t} < \hat{x}_{k\delta t}, \mathrm{Int}(R_{k\delta t}, R_{(k+1)\delta t}) >$$

converge vers

$$\mathcal{I} = \int_0^t < \hat{x}_s, dz_s >$$

dans le sens $\mathbb{E}[|S^\delta - \mathcal{I}|^2] \to 0$ *lorsque* $\delta \to 0$ *si la fonction d'interpolation* $\mathrm{Int} : SO(n) \times SO(n) \to \mathfrak{so}(n)$ *satisfait les conditions suivantes:*

1. *La fonction est nulle sur sa diagonale* $\mathrm{Int}(R, R) = 0$ *pour tout* $R \in SO(n)$.
2. *La fonction* Int *est* $\mathcal{C}^2(SO(n))$.
3. $\nabla \mathrm{Int}(R, R)[v] = vR^T$ *pour tout* $v \in T_R SO(n)$.
4. $\nabla^2 \mathrm{Int}(R, R)[v] = 0$ *pour tout* $v \in T_R SO(n)$.

où les différentielles (∇I *and* $\nabla^2 I$) *sont déterminées par rapport à la seconde variable.*

On peut alors montrer que l'interpolation géodésique donne une solution convergente

Proposition 31. *La fonction d'interpolation* $\mathrm{Int}(R', R) = \log(RR'^T)$ *satisfait les conditions du théorème précédent.*

En pratique, l'utilisation de la fonction log peut s'avérer coûteuse en terme de calcul. Cependant, d'autres fonctions peuvent aussi être utilisées

Proposition 32. *La fonction d'interpolation* $\mathrm{Int}(R', R) = \frac{1}{2}(RR'^T - R'R^T)$ *satisfait les conditions du théorème précédent.*

Cette fonction d'interpolation sera appelée interpolation linéaire. Elle consiste à développer la fonction log au premier ordre et à projeter le développement dans l'espace des matrices anti-symétriques. On peut montrer à l'aide de simulations numériques que pour un pas d'échantillonnage faible, les deux types d'interpolation donnent des résultats similaires. On préférera donc l'interpolation linéaire, à cause de sa simplicité de calcul. Dans le cas d'un pas d'échantillonnage plus grand, l'interpolation géodésique donne un résultat plus précis que l'interpolation linéaire et il conviendra de choisir, suivant le cas d'utilisation, l'une ou l'autre de ces fonctions d'interpolations.

A.3.3 Filtrage à partir d'observations sur les variétés de Stiefel, solution Monte-Carlo

Dans la Section 2.2, on cherche à estimer la vitesse angulaire d'un processus rotationnel bruité. Il est supposé que ce processus est observé grâce à des capteurs tels que des caméras. Il peut cependant arriver que certains capteurs soient défectueux ou bien qu'un nombre restreint de caméras empêche d'acquérir le signal rotationnel dans sa totalité. Le but de la Section 2.3 est de traiter le cas du filtrage pour observation partielle du processus de rotation. On modélise le processus observé par un processus dans les variétés de Stiefel. Une solution numérique est par ailleurs proposée.

La variété de Stiefel $V_{n,k}$ correspond, pour un $k \leq n$ à l'ensemble des matrices P de taille $n \times k$ telles que

$$P^T P = I_k.$$

Si $k = 1$, alors $V_{n,k} = S^{n-1}$, la sphère unité de $\mathbb{R}^n$. Si $k = n$, alors $V_{n,k} = O(n)$, le groupe des matrices orthogonales de $\mathbb{R}^n$. On considère dans notre cas que $k \leq n-1$ pour modéliser le fait que l'observation est partielle. L'idée est de n'observer qu'une partie de la matrice de rotation.

Soit $\Pi : SO(n) \to V_{n,k}$ la projection qui tronque les $n-k$ dernières colonnes d'une matrice de rotation. En d'autres termes,

$$\Pi(R) = P \quad \text{avec la concaténation } R = (P,\, P') \text{ et } P \in \mathbb{R}^{n\times k},\ P' \in \mathbb{R}^{n\times(n-k)}.$$

Lorsque $k \leq n-2$, la projection n'est pas injective, c'est-à-dire qu'un élément $P \in V_{n,k}$ peut se compléter en matrice de rotation de différentes manières

$$\Pi(R_1) = \Pi(R_2) \Leftrightarrow R_1 = R_2 \begin{pmatrix} I_k & 0 \\ 0 & r \end{pmatrix} \qquad \text{avec } R_1, R_2 \in SO(n),\ r \in SO(n-k).$$

L'application Π est par contre clairement surjective, c'est-à-dire que l'image de $SO(n)$ par Π couvre la totalité de $V_{n,k}$. Tout élément de $V_{n,k}$ peut s'écrire comme l'image par Π d'au moins une matrice de rotation. Il est donc légitime de se demander si la géométrie de $V_{n,k}$ peut se déduire de la géométrie de $SO(n)$.

Par sa définition, on peut aisément montrer que Π est invariante à gauche

$$R_1\Pi(R_2) = \Pi(R_1R_2) \qquad \text{with } R_1, R_2 \in SO(n).$$

Cela démontre aussi l'action de $SO(n)$ sur $V_{n,k}$. Si $R \in SO(n)$ et $P \in V_{n,k}$, alors $RP \in V_{n,k}$. Cela servira à décrire des processus sur $V_{n,k}$ à l'aide de l'action du groupe $SO(n)$.

On définit l'application $\chi : \mathfrak{so}(n) \times V_{n,k} \to TV_{n,k}$ de la façon suivante

$$\chi(\sigma, P) = \frac{d}{dt}|_{t=0} \exp(t\sigma)P = \sigma P \tag{A.29}$$

On peut montrer que χ est surjective, *i.e* $T_PV_{n,k} = \{\sigma P, \sigma \in \mathfrak{so}(n)\}$.

Pour un point donné $P \in V_{n,k}$, on définit $R \in SO(n)$ une pré-image de P via Π, *i.e* $P = \Pi(R)$. On définit l'espace vertical $\mathcal{V}_R$ au point P en R par

$$\mathcal{V}_R = \operatorname{Ker} d\Pi_R,$$

où $d\Pi_R$ est la différentielle de Π au point R. Par définition de $d\Pi_R$, l'espace vertical est un sous-espace de l'espace tangent $T_RSO(n)$. Grâce à la métrique introduite sur $SO(n)$, on définit l'espace horizontal $\mathcal{H}_R$ comme le complémentaire orthogonal de $\mathcal{V}_R$ dans $T_RSO(n)$:

$$T_RSO(n) = \mathcal{V}_R \oplus \mathcal{H}_R.$$

Comme $d\Pi_R$ est linéaire, la restriction de $d\Pi_R$ à $\mathcal{H}_R$ est bijective. En d'autres termes, $T_PV_{n,k}$ et $\mathcal{H}_R$ sont isomorphes. Pour un vecteur $v \in T_PV_{n,k}$, on définit $v^{\mathcal{H}}$, son relevé horizontal comme étant

$$d\Pi_R(v^{\mathcal{H}}) = v.$$

Grâce à l'isomorphisme, la restriction de $\chi(.,P)$ à $\mathcal{H}_RR^T \in \mathfrak{so}(n)$ est bijective. On définit χ^{-1} l'inverse de cette restriction:

$$\chi^{-1}(v, P) = v^{\mathcal{H}}R^T.$$

Il est remarquable que malgré le fait que $v^{\mathcal{H}}$ dépende de R, le terme $v^{\mathcal{H}}R^T$ est, lui, indépendant de la matrice R choisie.

On peut par ailleurs définir une métrique basée sur la métrique de $SO(n)$. On définit la métrique $< .,. >_P$ par

$$< v_1, v_2 >_P = < v_1^{\mathcal{H}}, v_2^{\mathcal{H}} >_R \qquad \text{avec } v_1, v_2 \in T_PV_{n,k}, \Pi(R) = P.$$

On considère maintenant le problème de filtrage avec observations partielles. C'est-à-dire, au lieu d'observer le processus de la Section 2.2, on observe sa projection via l'application Π. On note P_t le processus observé. En projetant l'équation (2.15), on obtient

$$dP_t = (\hat{x}_t dt + \circ d\hat{w}_t)\, P_t, \tag{A.30}$$

avec $w_t \in \mathbb{R}^{n(n-1)/2}$ un mouvement Brownien avec variance σ_w^2 indépendant de x_t.

De manière similaire, on cherche la distribution $\pi_[0,t]$ de x_t conditionnée par l'observation $\mathcal{P}_t = \{P_s, s \leq t\}$. De la même façon que dans la Section 2.2, le bruit est multiplicatif. Il n'est cependant pas possible d'utiliser la même solution car $V_{n,k}$ n'est pas un groupe et ses éléments ne sont pas inversibles. Malgré cela, une solution basée sur le même principe est présentée.

Soit $z_t \in \mathfrak{so}(n)$ et $R_t \in SO(n)$ définis par

$$\begin{aligned} dz_t &= \chi^{-1}(\circ dP_t, P_t) \\ dR_t &= (\circ dz_t) R_t \end{aligned} \tag{A.31}$$

Le processus z_t est appelé antidéveloppement de P_t et R_t est appelé relevé horizontal de P_t [Sepulchre 2008]. Ces processus sont équivalents, en terme d'information, à P_t car P_t peut être reconstruit à partir de chacun d'eux:

$$P_t = R_t R_0^T P_0 \qquad \text{ou} \qquad dP_t = \chi(dz_t, P_t).$$

Il est donc possible de chercher la distribution de x_t conditionnée par l'observation de z_t. On peut montrer par ailleurs que l'antidéveloppement est solution d'une SDE avec bruit additif. En effet,

$$\begin{aligned} dz_t &= \chi^{-1}(dP_t, P_t) \\ &= \chi^{-1}\left((\hat{x}_t dt + \circ d\hat{w}_t)\, P_t, P_t\right) \\ &= \chi^{-1}\left(\left(\chi\left(\hat{x}_t dt + \circ d\hat{w}_t\right), P_t\right), P_t\right) \\ &= \chi^{-1}\left(\chi\left(\hat{x}_t dt, P_t\right), R_t\right) + \circ\chi^{-1}\left(\chi\left(d\hat{w}_t, P_t\right), R_t\right) \\ &= \chi^{-1}(H_t, P_t) + \circ d\hat{\beta}_t, \end{aligned} \tag{A.32}$$

où $H_t = \chi(\hat{x}_t, P_t)$ et $d\hat{\beta}_t = \chi^{-1}\left(\chi\left(\circ d\hat{w}_t, P_t\right), P_t\right)$. On peut alors montrer que β_t est un mouvement Brownien avec une variance qui peut se diagonaliser sous la forme $\sigma_w^2 I_{n,k}$ où $I_{n,k} = \mathrm{diag}(1, ..., 1, 0, ..., 0)$ avec k éléments non nuls.

Le problème se trouve ainsi réduit à un problème de filtrage dans $\mathbb{R}^{n(n-1)/2}$ avec du bruit additif. La solution est donc, pour une fonction test ϕ:

$$\pi_{[0,t]}(\phi) = \frac{\rho_{[0,t]}(\phi)}{\rho_{[0,t]}(1)}, \tag{A.33}$$

avec $\rho_{[0,t]}(\phi) = \mathbb{E}\left[\phi(x')L_t(P, x')|\mathcal{P}_t\right]$, où x'_t est une copie de x_t indépendante de P_t et $\mathcal{P}_t = \{P_s, s \leq t\}$. La vraisemblance L_t est définie par

$$L_t(P, x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < x'_s, dz_s >_{\mathfrak{so}(n)} - \frac{1}{2\sigma_w^2}\int_0^t ||x'_s||^2 ds\right) \qquad \text{avec } dz_t = \chi^{-1}(\circ dP_t, P_t).$$

Par définition du produit scalaire dans $V_{n,k}$ hérité de $SO(n)$, la vraisemblance peut s'écrire

$$L_t(P, x') = \exp\left(\frac{1}{\sigma_w^2}\int_0^t < H_s, dP_s >_{V_{n,k}} - \frac{1}{2\sigma_w^2}\int_0^t ||H_s||^2 ds\right) \qquad \text{with } H_s = \chi(\hat{x}'_s, P_s). \tag{A.34}$$

Cependant, le modèle pour H_t n'est pas linéaire, même si x_t suit un modèle linéaire. Utiliser donc la solution (A.33) n'est donc pas faisable de manière adaptative. Il reste tout de même possible d'utiliser une méthode numérique, tel qu'un filtre à particules.

Un filtre à particule est une méthode de filtrage Monte-Carlo. L'idée est d'approximer $\rho_{[0,t]}$ à l'aide de la loi des grands nombres. On rappelle que

$$\rho_{[0,t]}(\phi) = \mathbb{E}\left[\phi(x')L_t(P,x')|\mathcal{P}_t\right]$$

avec L_t la vraisemblance définie à l'équation (A.34). On définit N processus, notés X_t^i et appelés particules avec un modèle similaire à celui de x_t. Ces processus représentent des candidats pour estimer le processus x_t. La loi des grands nombres nous dit que $\rho^N_{[0,t]}$ défini par

$$\rho^N_{[0,t]}(\phi) = \frac{1}{N}\sum_{i=1}^{N}\phi(X^i)L_t(P,X^i)$$

va converger avec N presque sûrement vers $\rho_{[0,t]}(\phi)$. Par conséquent, la précédente équation donne une approximation de la solution, une fois les coefficients $L_t(P,X^i)$ normalisés.

Tout comme dans le cas précédent, les processus ne sont pas observables en temps continus mais en temps discrets, avec un pas δt. On considère donc la version discrétisée de $\rho^N_{[0,t]}$, que l'on note ρ^N_n, avec $n = \lfloor t/\delta t \rfloor$ telle que

$$\rho^N_n(\phi) = \sum_i \phi(\tilde{X}^i)L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}^i_0, ..., \tilde{X}^i_n)$$

où $\tilde{X}^i_n = X^i_{n\delta t}$, $\tilde{\mathcal{P}}_n = \{P_{k\delta t}, k \leq n\}$ et la vraisemblance $L_{\delta t}$ est définie par

$$L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}^i_0, ..., \tilde{X}^i_n) = \exp\left(\frac{1}{\sigma_w^2}\sum_{k\leq n} < H^i_k, \delta\tilde{P}_k > -\frac{1}{2\sigma_w^2}\sum_{k\leq n}||H^i_k||^2\right)$$

avec $H^i_k = \chi(X^i, \tilde{P}_k)$.

Pour implémenter une telle solution de manière adaptative, il faut être capable de traiter ces deux problèmes indépendants

- Simuler les particules $\tilde{X}^i$:
  Comme x_t est un processus de Markov avec un générateur $\mathcal{A}$, on peut définir un noyau de transition $q_{\delta t}$ pour simuler $\tilde{X}^i$.
- Calculer de manière adaptative la vraisemblance $L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}^i_0, ..., \tilde{X}^i_n)$:
  En séparant le dernier terme de la somme et en le factorisant, on obtient que

  $$L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}^i_0, ..., \tilde{X}^i_n) = L_{\delta t}(\tilde{\mathcal{P}}_{n-1}, \tilde{X}^i_0, ..., \tilde{X}^i_{n-1})l^i_{\delta t}$$

  avec $l^i_{\delta t} = \exp\left(\frac{1}{\sigma_w^2} < H^i_n, \delta\tilde{P}_n > -\frac{1}{2\sigma_w^2}||H^i_n||^2\right)$.

Cela mène à l'algorithme suivant, pour lequel on estime la distribution conditionnelle $\pi^{N,\delta t}_n(\phi)$ où

$$\pi^{N,\delta t}_n(\phi) = \frac{\rho^{N,\delta t}_n(\phi)}{\rho^{N,\delta t}_n(1)} = \sum_{i=1}^{N}\phi(\tilde{X}^i_n)w^i_n$$

dont le coefficient

$$w^i_n = \frac{L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}^i_0, ..., \tilde{X}^i_n)}{\sum_{j=1}^{N} L_{\delta t}(\tilde{\mathcal{P}}_n, \tilde{X}^j_0, ..., \tilde{X}^j_n)}$$

Algorithm 3 Filtre à particules

- Pour l'initialisation, génerer N particules à partir de l'apriori p_0: $\tilde{X}_0^i \sim p_0$ et définir $w_0^i = 1/N$.
- Au temps $n > 0$:
  1. Propager les particules $\tilde{X}_n^i \sim q_{\delta t}(\ .\ , \tilde{X}_{n-1}^i)$
  2. Mettre à jour le poids w_n^i de chaque particule: $w_n^i = w_{n-1}^i l_{\delta t}^i$ avec
  $$l_{\delta t}^i = \exp\left(\frac{1}{\sigma_w^2} < H_n^i, \delta\tilde{P}_n > -\frac{1}{2\sigma_w^2}||H_n^i||^2\right).$$
  3. Normaliser les poids: $w_n^i = w_n^i / \sum_j w_n^j$
  4. Si $\left(\sum_i (w_n^i)^2\right)^{-1} < N/2$, génerer $[m^1 \ ... \ m^N] \sim \text{multinomial}(w^1 \ ... \ w^N)$ tels que $\sum_i m^i = N$. Ensuite, cloner la particule $\tilde{X}_n^i$ m^i-fois et redéfinir les poids par $w_n^i = \frac{1}{N}$.
  5. Estimer $\pi_t(\phi)$ avec $\pi_n^{N,\delta t}(\phi) = \sum_i \phi(\tilde{X}_n^i) w_n^i$

est appelé "poids" associé à la particule i.

L'étape 4 s'appelle ré-échantillonnage. Elle est là pour éviter la dégénérescence de l'algorithme due au nombre fini de particules. Dans le cas où x_t est une diffusion par exemple, les particules se propagent dans tout l'espace et deviennent très vite de mauvais candidats. Dans ce cas, seule une particule accumulerait tous les poids. L'idée de cette étape est d'éliminer les particules que l'on sait mauvaises et cloner celles qui sont, jusqu'à maintenant, de bons candidats. Ici, un critère basé sur l'ESS (Effective Sample Size) est donné mais il serait aussi possible de ré-échantillonner à chaque intervalle de temps fixé. Cette étape ralentie cependant grandement l'exécution du filtre car elle empêche tout parallélisation des calculs.

On peut montrer numériquement que cette solution converge lorsque x_t reste constant suffisamment longtemps pour laisser les particules converger. Cela fonctionne donc au cas où x_t est constant par morceaux. Si x_t évolue trop vite, seule la partie horizontale de x_t peut être estimée. En effet, la composante verticale de x_t laissant P_t inchangée, elle reste indétectable tant que le plan tangent à P_t ne change pas. Si ce plan change au cours du temps, il modifie la composante de x_t qui est verticale et permet, au fur et à mesure, de voir l'effet complet (donc d'estimer) de x_t sur P_t.

A.4 Phase géométrique

Le Chapitre 3 traite d'un problème indépendant du Chapitre 2. Il rassemble les résultats des articles [Rossetto 2012] et [Rossetto 2013].

La Section 3.1 définie la notion de phase géométrique, initialement introduite par Berry dans [Berry 1984], à l'aide du pendule de Foucault. La phase géométrique est un phase qu'un système évoluant sur la sphère S^2 acquiert au cours d'un trajectoire fermée. Il a par ailleurs été montré [Segert 1987] que les ondes à polarisation transverse acquièrent aussi une phase géométrique au cours d'une propagation adiabatique. La Section 3.1 présente de plus une expérimentation destinée à mettre en évidence l'existence d'une telle phase et à la mesurer pour des ondes élastiques se propageant dans un guide d'onde.

La Section 3.2 décrit la distribution de la phase géométrique pour une onde polar-

isée qui se propage dans un milieu diffusant. La propagation de l'onde peut se modéliser par un processus sur $SO(3)$ avec transport parallèle. Dans un régime de diffusion multiple, il est possible de décrire la phase géométrique à l'aide de quelques paramètres tels que le libre parcours moyen ou le coefficient de diffusion. Cela mène à la construction d'estimateurs paramétriques de ces coefficients, construits à partir d'un algorithme d'Espérance-Maximisation (EM), à partir de l'observation de la distribution de la phase géométrique.

A.4.1 Phase géométrique pour les systèmes mécaniques classiques

En 1851, Léon Foucault installa dans le panthéon à Paris un pendule afin d'illustrer la rotation de la Terre sans observation d'étoiles ou d'objets astronomiques. A partir des lois classiques de la mécaniques, la direction des oscillations tend à rester constante vis-à-vis des étoiles. Alors que la Terre tourne sur elle-même, cette direction ne reste donc pas constante dans le repère de la salle où le pendule est installé.

Foucault observa que le pendule avait acquis un précession de $-271°$ à Paris après un jour complet. Après avoir reproduit la même manipulation à différents endroits dans le monde, il en déduit une loi empirique pour la précession du pendule après une journée complète d'oscillation. Pour une colatitude l, le taux de précession est de $2\pi(1-\sin l)°$/jour. Dans le cas extrême d'un pendule placé au pôle Nord, *i.e* $l = 0$, on trouve une précession de $360°$ par jour. Cela est due au fait que l'axe d'oscillation reste constant vis-à-vis des étoiles à cause du principe d'inertie.

Malgré le fait que cette relation fut déduite de manière empirique, elle peut s'obtenir en considérant la force de Coriolis dans la mécanique Newtonienne. Il est aussi possible d'expliquer cette phase avec des concepts géométriques, c'est-à-dire sans lois mécaniques. On utilise le modèle suivant: soit v_t la direction des oscillations, où t représente le temps (le vecteur v_t n'étant pas constant sur la sphère). A cause de la gravité, le vecteur v_t est contenu dans le plan horizontal. On appelle γ_t la position du pendule au temps t. On a donc $v_t \in T_{\gamma_t}S^2$, avec S^2 qui modélise la surface de la Terre.

Comme $T_{\gamma_t}S^2$ n'est pas constant lorsque γ_t se déplace sur S^2, on cherche une règle pour déterminer l'évolution de v_t. Le principe d'inertie cherche à conserver v_t constant mais la gravité impose que $v_t \in T_{\gamma_t}S^2$. En considérant de petites périodes d'oscillations pour le pendule, le mieux que l'on puisse faire pour satisfaire le principe d'inertie est de considérer que pour δt petit

$$v_{t+\delta t} = P_{\gamma_{t+\delta t}} v_t$$

avec $P_{\gamma_{t+\delta t}}$ la projection orthogonale de $\mathbb{R}^3$ dans $T_{\gamma_{t+\delta t}}S^2$. La condition de rapides oscillations par rapport à l'évolution de v_t est appelée condition adiabatique et la dernière équation peut se prouver à l'aide du théorème adiabatique [Faure 2013].

Comme la projection est linéaire, en considérant γ_t continue, on a donc, lorsque δt tend vers 0

$$P_{\gamma_t} \frac{dv_t}{dt} = \nabla_{\dot{\gamma}_t} v_t = 0,$$

où ∇ est la dérivée covariante de S^2. Le vecteur v_t est dit transporté parallèlement le long de γ_t.

Proposition 33. *Le transport parallèle conserve le produit scalaire. C'est-à-dire que si v_t et v'_t sont transportés parallèlement le long de γ_t, alors $< v_t, v'_t >$ reste constant.*

Par extension, le transport parallèle conserve aussi la norme. Il est donc possible de transporter une base orthonormée le long d'une courbe. Elle restera une base tout au long de son transport, les vecteurs restant unitaires et orthogonaux entre eux.

On considère maintenant le cas où γ est un chemin fermé de S^2, *i.e* $\gamma_0 = \gamma_1$. Soit v_t un vecteur transporté parallèlement le long de γ. Comme $\gamma_0 = \gamma_1$, les plans tangents $T_{\gamma_0}S^2$ et $T_{\gamma_1}S^2$ sont égaux. Il est donc possible de comparer v_0 et v_1 car ils sont dans le même plan.

Soit β_γ l'angle orienté entre v_0 et v_1

$$\beta_\gamma = \widehat{v_0, v_1}. \tag{A.35}$$

Cet angle est appelé la phase géométrique de γ. Il est à noter que cet angle est bien défini grâce à la proposition précédente. De plus, la conservation du produit scalaire rend cette définition indépendante du choix initiale de v_0.

Proposition 34. *La phase géométrique β_γ est indépendante de la paramétrisation de γ. C'est-à-dire que pour une fonction $\phi : \mathcal{C}^1([0,1] \to [0,1])$ telle que $\phi_0 = 0$ et $\phi_1 = 1$,*

$$\beta_\gamma = \beta_{\gamma \circ \phi}$$

Cette proposition signifie que la phase géométrique ne dépend que de la géométrie du l'image de γ. Cependant, déterminer β_γ directement à partir de sa définition reste difficile car il faudrait transporter un vecteur le long de toute la courbe.

Theorem 9. *Etant donné un chemin fermé γ sur S^2, la phase géométrique β_γ est donnée par*

$$\beta_\gamma = \iint_{\mathcal{C}} ds,$$

où $\mathcal{C}$ est la surface délimitée par γ, i.e $\partial\mathcal{C} = \gamma$.

Ce théorème, appelé Théorème de Gauss-Bonnet local, donne un moyen de calculer facilement la phase géométrique à partir de l'image du chemin γ.

Pour le cas du pendule de Foucault placé à la colatitude l, l'axe d'oscillation est transporté parallèlement durant la rotation de la Terre. Après une révolution complète, la trajectoire du pendule γ est fermée et la phase géométrique β_γ représente la précession du pendule. L'aire enfermé par la calotte à une colatitude l est

$$\beta_\gamma = 2\pi(1 - \sin l).$$

Cette expression correspond au taux de précession déterminé à l'aide des lois de la mécanique classique.

Malgré le fait que la phase géométrique ait été introduite dans la Section 3.1 à l'aide du pendule de Foucault, elle existe aussi pour d'autres systèmes. Dans [Segert 1987] par exemple, il a été montré que les ondes polarisées sont sujettes à acquérir une phase géométrique durant leur propagation.

On considère désormais la propagation d'une onde avec une polarisation transverse linéaire. Il est possible d'étendre les résultats suivant pour des ondes avec une polarisation elliptique mais le problème se trouve plus aisément formulé dans le cas d'une polarisation linéaire.

On note γ_t le vecteur unitaire indiquant la direction de propagation de l'onde. Comme γ_t est unitaire, $\gamma_t \in S^2$. La polarisation étant linéaire, on note v_t le vecteur unitaire représentant la direction de polarisation. La polarisation étant transverse, on a v_t orthogonal à γ_t. Cela signifie aussi que $v_t \in T_{\gamma_t}S^2$. Tant que l'onde se propage dans une direction donnée, γ_t et v_t restent inchangés. Cependant, si la direction de propagation est modifiée (guide d'onde, diffuseur, etc), alors v_t change aussi pour rester dans le plan tangent $T_{\gamma_t}S^2$. On peut montrer que lorsque la trajectoire de l'onde a une faible courbure (par rapport à la

longueur d'onde), alors v_t est transporté parallèlement le long de γ_t. Dans ces conditions, la propagation est dite adiabatique.

Si γ_t pour $t \in [0,1]$ est un chemin fermé, c'est-à-dire si $\gamma_0 = \gamma_1$, alors v_t va acquérir une phase géométrique. En connaissant le chemin suivi par l'onde, on peut prédire la phase géométrique.

Si l'on considère un guide d'onde en forme d'hélice avec un rayon r et un pas P, la longueur d'une spire est $L = \sqrt{(2\pi r)^2 + P^2}$. Soit θ l'angle entre la direction de propagation et l'axe principal de l'hélice. On a $\sin\theta = P/L$. Comme le guide d'onde est une hélice, l'angle θ reste constant au cours de la propagation et quitte à considérer l'axe principal comme étant vertical, γ_t reste à une colatitude constante sur S^2. Grâce au théorème de Gauss-Bonnet, on en déduit que la phase géométrique, après une spire, est de

$$\beta_\gamma = 2\pi\cos\theta.$$

Cependant, ce résultat fonctionne uniquement pour des chemins fermés sur S^2. Dans le cas de la propagation le long de l'hélice, on cherche à déterminer la phase géométrique en tout point le long du guide d'onde.

Pour cela, il faut utiliser le théorème de Fuller [Fuller 1978] qui précise que si la trajectoire γ peut s'écrire comme une déformation d'une trajectoire de référence γ_{ref}, la phase géométrique β_γ peut s'écrire comme la somme de la phase géométrique $\beta_{\gamma_{ref}}$ et l'aire entre le chemin γ et γ_{ref}.Plus précisément, après une propagation le long d'un arc de longueur s, on a

$$\beta_{\gamma,s} = \beta_{\gamma_{ref},s} + \mathcal{A}(\gamma, \gamma_{ref}, s),$$

où $\mathcal{A}(\gamma, \gamma_{ref}, s)$ est l'aire engendrée par $\gamma_{s'}$ pour $s' \leq s$ durant la transformation de γ_{ref} en γ.

Dans le cas du guide d'onde en forme d'hélice, la trajectoire de référence est le cercle (ou bien une hélice avec un angle $\theta = 0$) pour lequel la phase géométrique est nulle. Dans ce cas, $\mathcal{A}(\gamma, \gamma_{ref}, s) = (2\pi s/L)(\cos\theta)$. Par conséquent, la phase géométrique acquise après une propagation sur une distance s est

$$\beta(s) = (2\pi s/L)(\cos\theta) = \tau s, \qquad \text{with } \tau = (2\pi\cos\theta)/L \tag{A.36}$$

Cela montre montre aussi la dépendance linéaire de β avec la longueur s dans le cas d'une hélice.

Une manipulation a été réalisé afin de vérifier l'existence d'une phase géométrique pour les ondes élastiques et afin de valider le modèle précédent.

Pour cela, un guide d'onde en forme d'hélice fut utilisé avec un rayon $r = 75$mm, un pas $P = 91.5$mm et un angle $\theta = 1.38$rad.

Afin de mesurer les vibrations mécaniques, le système fut isolé en le suspendant par deux ficelles. Deux accéléromètres furent utilisés, situés à une extrémité du ressort et placés de manière à enregistrer les vibrations dans deux directions orthogonales, permettant de reconstruire la vibration totale. Les vibrations furent créées en 32 points répartis avec un pas constant le long de l'hélice. En créant de légers impacts radialement (vis-à-vis de l'axe principal de l'hélice), des ondes de flexion avec une polarisation linéaire furent créés. Les impacts furent répétés tout au long de l'hélice.

En mesurant la vibration de l'extrémité du guide d'onde, il est possible avec une méthode de type Analyse en Composantes Principales de retrouver la direction de polarisation de l'onde. L'idée est alors de comparer l'évolution de cette direction (l'angle que fait la direction de polarisation avec une valeur de référence) avec la phase géométrique que l'onde est sensée acquérir le long du guide d'onde.

On trouve alors expérimentalement une phase géométrique avec un taux de $\tau = 2.5 \pm 0.1\text{rad.m}^{-1}$ avec les marges d'erreurs, ce qui correspond à la valeur théorique qui devrait être de $\tau = 2.49 \pm 0.1\text{rad.m}^{-1}$. La fréquence des ondes enregistrées ne modifiaient en rien la valeur du coefficient τ, ce qui montre bien que l'effet observé est uniquement dû à la géométrie du guide d'onde.

A.4.2 Diffusion multiple d'une onde polarisée

Dans la nature, les ondes élastiques polarisées, telles que les ondes S en sismologie, sont observées sous certaines conditions et le concept de phase géométrique s'applique. Le degré de la polarisation et l'orientation de la polarisation contient de l'information sur le désordre du milieu de propagation. Les ondes sismiques ne peuvent cependant se modéliser à l'aide d'une propagation déterministe mais plutôt comme des processus dans des milieux diffusants.

On note ℓ la longueur moyenne que parcourt l'onde entre deux événements de diffusions. Lorsque l'onde rencontre un diffuseur, qui se caractérise par une hétérogénéité dans le milieu, elle change de direction de propagation. On note $\ell^\star$ la longueur à partir de laquelle les directions de propagation des ondes diffusées sont uniformément distribuées.

On considère une onde qui se propage dans un milieu diffusant d'une largeur L telle que $\ell < L < \ell^\star$. Cela signifie que l'onde observée en sortie du milieu a subit un nombre restreint de diffusions lors de sa propagation. Certaines méthodes d'estimation permettent de faire de l'estimation sur le milieu aléatoire à partir de l'observation de la direction de propagation [Margerin 2009]. Ce modèle est ici étendu afin de prendre en compte l'information portée par la polarisation de telles ondes.

On considère que les ondes qui se propagent dans le milieu sont des ondes avec une polarisation linéaire. On néglige tout effet de conversion et de changement de polarisation au cours de la propagation. De même que précédemment, on note $\gamma \in S^2$ la direction de propagation de l'onde et $v \in T_\gamma S^2$ sa polarisation. On note F le repère $F = [\gamma, v, \gamma \wedge v]$. Par définition de γ et v, on a $F \in SO(3)$. Entre deux événements de diffusion, l'état de l'onde, décrit par F, reste constant. Lorsque l'onde rencontre un diffuseur, elle change de direction, toujours sous la contrainte que $v \in T_\gamma S^2$, c'est-à-dire que F change. Comme F est un repère, tout changement peut s'exprimer par une matrice de rotation. L'angle de diffusion, noté θ, est aléatoire. Il est distribué selon une distribution appelée fonction de phase, notée Φ'.

On note g l'anisotropie de la diffusion, qui peut se définir comme la valeur moyenne de $\cos\theta$. Il est a noter que l'on a $\ell^\star = \ell/(1-g)$. Dans le cas d'une diffusion vers l'avant, on a g proche de 1.

La rotation R agissant sur F peut se paramétrer à l'aide des angles d'Euler, dans lesquels on retrouve l'angle de diffusion: $(\psi, \theta, \varphi) \in]-\pi, \pi] \times]0, \pi] \times]-\pi, \pi]$ (voir Paragraphe 1.1.3).

On peut montrer [Rossetto 2001] que durant cet événement de diffusion, la polarisation est transporté parallèlement. Cela se traduit sur les angles de R par la condition que

$$\psi = -\varphi.$$

Il est donc possible de paramétrer R avec seulement 2 angles, le transport parallèle supprimant un degré de liberté. La relation entre le repère avant diffusion F et après diffusion F' s'exprime alors

$$F' = FR(\theta, \psi).$$

Il est remarquable que R agit sur F avec un produit à droite. Il est possible d'écrire une relation similaire avec un produit à gauche mais cela impliquerait une dépendance plus complexe entre les rotations successives qui peuvent s'appliquer.

Pour une onde paramétrée par F_0, son état après n diffusions peut se modéliser par

$$F_n = F_0 R(\theta_1, \psi_1) R(\theta_2, \psi_2) \cdots R(\theta_n, \psi_n). \tag{A.37}$$

Les événements de diffusions sont indépendants et les diffuseurs considérés identiques. Les matrices de rotations sont donc indépendantes et identiquement distribuées. On note Φ la distribution de $R(\theta_m, \psi_m)$. La fonction de phase Φ' est donc la marginale de Φ par rapport à ψ.

Pour l'observateur, l'onde peut rencontrer un nombre variable de diffuseurs. On peut montrer [Margerin 2009] que l'on peut modéliser le nombre d'événements de diffusions après un temps de parcours t par un processus de Poisson $N(t)$ avec un paramètre η.

On peut donc écrire l'état de l'onde après un temps de propagation t par F_t où

$$F_t = F_0 \prod_{k=1}^{N(t)} R(\theta_k, \psi_k), \tag{A.38}$$

pour lequel le produit s'effectue à droite. Cette équation exprime le processus stochastique F_t comme un processus de Poisson composé sur $SO(3)$ avec transport parallèle. On note p_ν la distribution de F_t avec $\nu = \eta t$ le nombre moyen d'événements de diffusions.

Un processus de Poisson est un processus à sauts avec incréments indépendants. On peut écrire que

$$p_\nu = \sum_{n=0}^{\infty} e^{-\nu} \frac{\nu^n}{n!} p_{F_n}. \tag{A.39}$$

où $e^{-\nu}\frac{\nu^n}{n!}$ est la probabilité d'avoir exactement n événements de diffusions dans l'intervalle $[0, t]$.

Grâce à l'indépendance des rotations, on a

$$p_{F_n} = p_{F_0} * p_{R_1} * \cdots * p_{R_n} = \Phi^{*[n]}. \tag{A.40}$$

Ici, le symbole $*$ désigne la convolution sur $SO(3)$ et $\Phi^{*[n]}$ le résultat de la convolution de n fonctions identiques Φ. On considère ici que $p_{F_0}(F) = \delta(I_3 - F)$.

On utilise alors l'analyse harmonique introduite dans la Section 1.4 pour transformer l'équation (A.40). Les représentations irréductibles λ peuvent s'indexer sur $\mathbb{N}$ avec une dimension $d_\lambda = 2\lambda + 1$. Les fonctions U^λ sont alors des matrices $(2\lambda+1) \times (2\lambda+1)$ appelées matrices de Wigner-D. Pour une matrice de rotation paramétrée avec les angles d'Euler, U^λ peut se décomposer en

$$U^\lambda_{m,n}(\psi, \theta, \varphi) = e^{-im\psi} d^\lambda_{m,n}(\theta) e^{-in\varphi}$$

pour $-l \leq m, n \leq l$, où $d^\lambda_{m,n}$ est appelée fonction de Wigner-d et peut s'exprimer à l'aide des polynômes de Jacobi.

Grâce aux résultats de la Section 1.4 et à partir du fait que le transport parallèle contraint la matrice de rotation de diffusion, on déduit les coefficients de Fourier de p_ν

$$\hat{p}^\lambda_\nu = \sum_{n=0}^{\infty} e^{-\nu} \frac{\nu^n}{n!} \left(\widehat{\Phi}^\lambda\right)^n = \exp\left[\nu(\widehat{\Phi}^\lambda - I_{2\lambda+1})\right], \tag{A.41}$$

avec exp, l'exponentielle de matrice.

On paramétre F_t par les angles d'Euler et on note $\beta = \phi + \varphi$ la différence d'angle entre la polarisation initiale et la polarisation en sortie du milieu. Alors que θ représente la latitude

de la direction de propagation, β représente la phase géométrique que l'onde a accumulé le long de son parcours. On trouve alors

$$p_\nu(\theta,\, \beta) = R_0(\theta,\, \nu) + 2\sum_{m\geq 1} \cos(m\beta) R_m(\theta,\, \nu), \tag{A.42}$$

$$\text{where } R_m(\theta,\, \nu) = \frac{1}{2\pi}\sum_{\lambda\geq m} (2\lambda+1) e^{\nu(\hat{\Phi}^\lambda_{m,m}-1)} d^\lambda_{m,m}(\theta). \tag{A.43}$$

Parce que $\hat{p}^\lambda_\nu$ est diagonale, seules les fonctions U^λ_{mm} apparaissent lors de la reconstruction à partir des coefficients de Fourier. On peut constater que la distribution tend vers une distribution uniforme lorsque t devient grand.

Grâce à l'expression de la distribution de la phase géométrique, il est possible de construire un estimateur du libre parcours moyen ℓ. Pour cela, on cherche à estimer le paramètre de Poisson $\nu = \eta t$. On note $\hat{\nu}$ l'estimé de ν. On suppose que l'on a M observations de l'état de l'onde en sortie du milieu F_m pour $1 \leq m \leq M$. On considère que chaque F_m est indépendant des autres observations. On peut donc écrire la distribution jointe par $\prod_m p_\nu(F_m)$, avec $p_\nu(F_m) = p_\nu(\theta_m, \psi_m + \varphi_m)$, où φ_m, θ_m et ψ_m sont les angles d'Euler de F_m. La valeur de ν qui maximise le log-vraisemblance est

$$\hat{\nu} = \text{argmax}_\nu \left(\sum_{m=1}^{M} \log p_\nu(F_m) \right).$$

L'algorithme EM est une procédure itérative qui converge presque sûrement vers un maximum local. L'idée principale est d'introduire une variable aléatoire cachée $n = [n_1, ..., n_M]$ qui représente le nombre d'événements de diffusions pour chaque observation. On construit alors la séquence d'estimateurs

$$\hat{\nu}_{i+1} = \text{argmax}_\nu \sum_{n=0}^{\infty} \mathbb{P}(n \mid F_{1..M}, \hat{\nu}_i) \sum_{m=1}^{M} \log p_\nu(F_m, n_m) \tag{A.44}$$

Seul les termes $\mathbb{P}(n \mid F_{1..M}, \hat{\nu}_i)$ ont besoin d'être mis à jour durant l'algorithme. En utilisant la règle de Bayes, on peut écrire

$$\mathbb{P}(n \mid F_{1..M}, \hat{\nu}_i) = \frac{\prod_m p_{F_n}(F_m)}{\prod_m p_{\hat{\nu}_i}(F_m)} e^{-\hat{\nu}_i} \frac{\hat{\nu}_i^n}{n!}.$$

Le terme $p_{F_n}(F_m)$ est indépendant de $\hat{\nu}_i$ et peut être déterminé une fois pour toutes les itérations de l'algorithme. On a alors l'algorithme 4.

Algorithm 4 Algorithme EM

- Prendre une valeur initiale $\hat{\nu}_0$.
- Calculer hors-ligne l'ensemble de probabilités $p_{F_n}(F_m)$.
- Itérer sur i les opérations:
  1. Calculer l'ensemble $p_{\hat{\nu}_i}(F_m)$ à partir de l'estimation actuelle.
  2. Determiner la prochaine itération

$$\hat{\nu}_{i+1} = \frac{1}{M}\sum_{m=1}^{M} \frac{1}{p_{\hat{\nu}_i}(F_m)} \sum_{n=0}^{\infty} n\, p_{F_n}(F_m)\, e^{-\hat{\nu}_i} \frac{\hat{\nu}_i^n}{n!}.$$

En pratique, la somme sur n pour mettre à jour l'estimateur est tronquée à une valeur de n assez grande. On peut montrer numériquement la convergence de cet algorithme pour une large gamme de priori initial avec des données simulées. Il est à noter cependant que l'algorithme est sensible au bruit et que l'ajout d'un bruit sur les données ajoute un biais positif sur l'estimateur.

A.5 Conclusion

Dans cette thèse, nous présentons deux problèmes indépendants et leurs solutions, dans les Chapitre 2 et Chapitre 3.

Le Chapitre 2 traite du problème de filtrage à partir d'observations sur le groupe des rotations $SO(n)$ dans le cas d'observations complètes et sur les variétés de Stiefel $V_{n,k}$ dans le cas d'observations partielles. Ces processus observés sont solutions d'équations stochastiques différentielles non linéaires L'objectif est d'estimer la vitesse angulaire de ces processus.

Les algorithmes utilisés à ce jour tel que le filtre de Kalman étendu traitent ce problème en utilisant une méthode de linéarisation locale. Cependant, cette linéarisation locale doit être réalisée à chaque nouvel échantillon et les équations du filtre dépendent alors des estimations précédentes. Nous proposons une approche différente, qui consiste à construire un nouveau processus appelé antidéveloppement à partir du processus observé. L'antidéveloppement est solution d'une équation stochastique linéaire pour lequel les méthodes de filtrage classiques peuvent être appliquées dans le cas de l'observation complète. Dans le cas de l'observation partielle, une solution numérique est présentée.

Une attention particulière est apportée à l'implémentation pratique de ces solutions. Les observations étant discrétisées temporellement, une méthode d'interpolation est utilisée pour déterminer la vraisemblance requise dans la solution. Il est par ailleurs démontré que plusieurs fonctions d'interpolations peuvent être choisies, en fonction de la précision désirée de la solution ou des performances de calculs. Dans le cas où le processus rotationnel est complètement observé, notre solution est capable d'estimer proprement la vitesse angulaire. Dans le cas où le processus est partiellement observable, alors modélisé par un processus dans les variétés de Stiefel, notre algorithme converge vers la bonne estimation dans le cas d'une vitesse angulaire constante. Malgré le fait qu'il est toujours possible d'estimer la vitesse angulaire si cette dernière évolue lentement ou bien est constante par morceaux, notre algorithme n'est capable d'estimer que partiellement la vitesse angulaire si cette dernière évolue trop rapidement.

L'algorithme présenté a des applications pour filtrer des observations partielles de trajectoires de robots par exemple. L'observation peut être acquise grâce à des caméras et notre méthode pourrait permettre de retrouver le mouvement 3D complet.

Dans le Chapitre 3, l'on s'intéresse à la phase géométrique acquise durant la propagation d'une onde élastique avec polarisation transverse linéaire. Cette phase résulte de la propagation 3D de l'onde et s'interprète comme une rotation de l'axe principale de la polarisation. Pour une onde élastique qui se propage dans un guide d'onde, cette phase peut être déterminée d'une part avec des outils de géométrie à partir de la trajectoire de l'onde et d'autre part peut être observée expérimentalement. Grâce à ces deux valeurs, il est alors possible de construire un estimateur paramétrique du guide d'onde.

Pour une onde élastique se propageant dans un milieu faiblement diffusant, il est possible d'obtenir une distribution de la phase géométrique due à la contribution des différentes trajectoire de l'onde dans le milieu. Cette distribution est obtenue à partir d'un modèle de processus de Poisson composé sur $SO(3)$. En échantillonnant cette distribution, une méthode d'"Expectation Maximisation" peut être utilisée pour estimer des paramètres du

milieu diffusant ou des diffuseurs. Cette distribution n'a cependant pas encore été observé. De possibles futures travaux se concerneront l'observation expérimentale de cette distribution pour des ondes élastiques. Cette observation n'est réalisable que dans le cas où il est possible de mesurer avec précision la polarisation des ondes. Si cela est possible pour des ondes lumineuses, cela reste bien plus délicat pour des ondes élastiques. Une fois que cette mesure obtenue, il sera possible d'utiliser la méthode présentée pour construire des techniques d'imagerie non intrusive et des techniques de surveillance de matériaux. Ces techniques utiliseront l'information supplémentaire fournie par la polarisation en plus des techniques classiques utilisant seulement l'amplitude ou l'intensité des ondes reçues.