Estimation and filtering of processes in matrix Lie groups

Salem Said

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

$N°$ attribué par la bibliothèque

THESE

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité : signal, image, parole, télécoms

préparée au Département Image et Signal du Laboratoire Gipsa-Lab

dans le cadre de **l'École Doctorale Électronique, Électrotechnique, Automatique, Télécommunications, Signal**

présentée et soutenue publiquement

par

Salem SAID

le 17 Décembre 2009

Titre :

ESTIMATION AND FILTERING OF PROCESSES IN MATRIX LIE GROUPS

Directeurs de thèse : Nicolas Le Bihan & Stephen J. Sangwine

JURY

Monsieur Pierre-Olivier Amblard	, Président
Monsieur Eric Moulines	, Rapporteur
Madame Anne Estrade	, Rapporteur
Monsieur Jonathan H. Manton	, Examinateur
Monsieur Jacques Istas	, Examinateur
Monsieur Nicolas Le Bihan	, Directeur de thèse
Monsieur Stephen J. Sangwine	, Co-directeur de thèse

Acknowledgements

I dedicate this thesis to my father.

I am happy to recognize the efforts and the support of many people without whom this thesis would not have been possible. I am very grateful to my thesis supervisors Nicolas Le Bihan and Steve Sangwine for their solidarity and democratic attitude. Nicolas' support helped me through the worst moments and his experience and advice played a major role in improving the quality and visibility of this thesis. Steve offered me the opportunity to spend four months at the *Department of computing and electronic systems* of the *University of Essex*. This was the happiest and most productive part of my three years as a Ph.D. student.

I would like to thank Daniel Baudois, who by lending me his books introduced me to the challenging field of nonlinear filtering. My thanks also go to Philippe Réfrégier. The whole of chapter 2 of this thesis would not have existed if not for his advice and encouragement. I also thank him for his invitation to the *Institut Fresnel* in Marseille.

I thank Vincent Rossetto-Giaccherino for introducing me to many fascinating topics in wave physics and in DNA dynamics. I look forward to our future collaboration in these fields.

Jonathan Manton offered me the exciting opportunity of visiting the *University of Melbourne*. In addition to getting a first glimpse of the beautiful Australian city of Melbourne, this visit allowed me to appreciate the high scientific and professional standard at the *University of Melbourne*. Jonathan's work concerning signal processing on manifolds was an important motivation for my thesis and my discussions with him were highly enriching.

I would like to thank Professor David Fremlin from the *University of Essex* for sharing with me his long and exciting experience as a pure mathematician. I continue to learn from Professor Fremlin's five book treatise on measure theory!

Many thanks to François Destelle who was a great office partner and will remain a best friend. I also thank Laura Carrea whom I met at the *University of Essex* for her friendship and for sharing her work on spinor models in polarization.

Remerciements

Je dédie cette thèse à mon père.

C'est pour moi un plaisir de reconnaitre les efforts et le soutien de beaucoup de personnes sans lesquelles cette thèse n'aurait pas été possible. Je remercie mes directeurs de thèse Nicolas Le Bihan et Steve Sangwine pour leur solidarité et leur attitude démocratique. Nicolas m'a soutenu pendant les moments les plus difficile et son expérience et ses conseils ont grandement amélioré la qualité et la visibilité de cette thèse. Steve m'a offert l'opportunité de passer quatre mois au *Department of computing and electronic systems* de la *University of Essex*. Cette période a été la plus belle et la plus productive de mes trois ans en tant que thésard.

Je voudrais remercier Daniel Baudois, qui en me prétant ses livres m'a permis d'aborder le domaine du filtrage nonlinéaire. Je remercie également Philippe Réfrégier. La chapitre 2 de cette thèse n'aurait pas existé sans ses conseils et encouragements. Je le remercie de plus pour son invitation à l'*Institut Fresnel* de Marseille.

Je remercie Vincent Rossetto-Giaccherino de m'avoir fait connaitre plusieurs sujets fascinants en physique des ondes et en dynamique de l'ADN. Je suis impatient de poursuivre notre collaboration dans ces domaines après ma thèse.

Grâce à Jonathan Manton j'ai pu visiter la *University of Melbourne*. En plus d'un premier contact avec la très belle ville de Melbourne, cette visite m'a permis d'apprécier le haut niveau scientifique et professionnel de la *University of Melbourne*. Les travaux de Jonathan concernant le traitement du signal sur les variétés étaient une motivation important pour cette thèse et mes discussions avec lui ont été très enrichissantes.

Je voudrais remercier le Professeur David Fremlin de la *University of Essex* pour avoir partagé avec moi sa longue et impressionnante expérience de chercheur en mathématiques pures. Je continue d'apprendre en lisant son traité en cinq tomes sur la théorie de la mesure!

Je remercie François Destelle qui a été un excellent collègue de bureau et restera un meilleur ami. Je remercie également Laura Carrea que j'ai rencontrée à la *University of Essex* de son amitié et d'avoir partagé avec moi ses travaux sur les modèles spinoriels pour la polarisation.

Contents

Chapter 1

Introduction

This thesis addresses estimation and filtering problems stated in terms of stochastic processes with their values in matrix Lie groups. Our interest in these processes is due to direct applications as well as to their general importance in modelling. The thesis is divided into Chapters 2 and 3. Chapter 2 only considers processes with their values in the rotation group, to which we refer shortly as rotation processes. The main problem solved in this chapter is the problem of decompounding, which is a nonparametric estimation problem –see Subsection 2.4.2. Chapter 3 more generally considers processes with their values in any matrix Lie group. It examines a new approach for the filtering of these processes, based on the notion of local linearization. This approach is given its final formulation in Subsection 3.2.5.

Chapter 2 is based on our two papers [63, 66]. It is devoted to those parts of our work which relate to direct applications, namely in the field of wave physics. Section 2.2 presents a new statistical formalism for polarization statistics [63]. Section 2.4 formulates the problem of decompounding as a statistical alternative to the physical problem of inverse scattering. In a more general form, this problem was considered and solved in [66] –to our knowledge, this is the first consideration of this problem.

The necessary mathematical background for Sections 2.2 and 2.4 is given in Sections 2.1 and 2.3. This is essentially based on the theory of characteristic functions of rotation random variables. The use of these characteristic functions is increasingly popular in the literature and they have been especially useful in nonparametric estimation problems –see [67] and discussion in Section 2.1. Our presentation introduces certain new aspects which are important to the applications of Sections 2.2 and 2.4. For example, Paragraph 2.1.2.2 characterizes the symmetry properties of rotation random variables, which are in turn a main ingredient in our solution of the problem of decompounding. Section 2.3 studies an important class of rotation processes, rotation Lévy processes.

Section 2.2 considers a problem of recent relevance in polarization statistics. The classical Stokes formalism of polarization is only based on second order statistics of the optical wavefield and our aim is to generalize this formalism to higher order statistics. The experiments of Ellis and Dogariu which we describe in Subsection 2.2.2 can be seen as our main motivation. In Subsection 2.2.3 we propose a new statistical formalism which generalizes the classical Stokes formalism to higher order statistics. This formalism succeeds in clarifying the contradictions raised by the experiments of Subsection 2.2.2. In Subsection 2.2.3, it is compared to other attempts made in the literature to include higher order statistics in the description of polarization. In Subsection 2.2.4 it is used to study the physical problem of depolarization.

Section 2.4 formulates and solves the problem of decompounding. Subsection 2.4.1 discusses this

problem as a statistical alternative to the physical inverse problem of multiple scattering. Subsection 2.4.2 studies the problem of decompounding in itself. The problem of decompounding is a nonparametric estimation problem stated in terms of rotation compound Poisson processes and interlaced processes, which were introduced in Subsections 2.3.2 and 2.3.4. We propose to solve the problem of decompounding using a characteristic function method. While the use of characteristic function methods for nonparametric estimation with rotation random variables is already well-established, the problem of decompounding presents an essential new difficulty in view of these methods. More precisely, it starts from indirect observations and we will see that its solution requires specific tools from matrix analysis as well as a more detailed statistical treatment. Paragraph 2.4.2.2 provides the mathematical proof for the convergence of our characteristic function method. Paragraph 2.4.2.3 discusses this convergence using numerical simulations.

Chapter 3 is of a more exploratory nature than Chapter 2. It considers processes with their values in any matrix Lie group, rather than only the rotation group. In this general case, there is no useful definition of characteristic functions as in Chapter 2. As such, the processes at hand are considered using a class of defining stochastic differential equations. We are interested in processes with values in matrix Lie groups as a general model for processes subject to nonlinear constraints which express symmetry considerations. Such processes appear in a wide variety of fields: Motion capture, DNA dynamics, medical imaging. We wish to address the applied problem of filtering under nonlinear constraints. The main difficulty to be dealt with is the impossibility of applying linear signal processing operations while at the same time respecting nonlinear constraints.

In practice, this difficulty is often dealt with using so-called global linearization methods. Such methods are quite easy to implement but suffer from many limitations –this has been discussed by Xavier and Manton [75] and Lee and Shin [38]. Motivated by its successful application in motion capture [16, 38], we are interested in the notion of local linearization. Our goal in Chapter 3 is to give a general formulation of the use of local linearization for filtering under nonlinear constraints and to justify the resulting performances. Section 3.1 is based on our paper [64] and takes a practical approach to this goal. Local linearization is considered in the special case of discrete time rotation processes. We study the stability and invariance properties of local linearization as defined in this case. In Subsection 3.1.2 we consider two numerical examples of filtering problems for which we discuss our approach in comparison to other recent works, namely based on optimization.

Section 3.2 corresponds to the main goal of this chapter. We consider processes which can be represented as the solutions of linear stochastic differential equations

$$dY_t = Y_t dX_t \qquad Y_0 = I_d$$

Our unknown process Y and the driving process X have their values in the space of $d \times d$ real matrices for some $d \geq 1$. Moreover, X is a Brownian process with independent increments and the equation is considered as a Stratonovich equation. For the initial condition, I_d denotes the $d \times d$ identity matrix. We refer to processes of the form Y as Brownian matrix processes. Section 3.2 develops the mathematical background needed in order to formulate correctly and in all generality the use of local linearization for Brownian matrix processes.

Subsection 3.2.1 shows that Brownian matrix processes naturally have their values in matrix Lie groups, so that they verify the general type of nonlinear constraints we have in mind. In Subsection 3.2.3, the correspondence $Y \mapsto X$ for a Brownian matrix process Y defined as above is considered. This is a well-defined transformation which we show to transform the process Y subject to nonlinear constraints to the process X subject only to linear constraints. Local linearization of a Brownian matrix process Y refers precisely to the application of the transformation $Y \mapsto X$. This can be

computed in a causal way –based only on current values of $Y-$ and eliminates the nonlinear constraints imposed on Y. Subsection 3.2.5 finally formulates our approach for the filtering of Brownian matrix processes and states the related stability properties. These are considered as the main justification of this approach. Note finally that Chapter 3 has the limited goal of formulating and studying in general the use of local linearization for Brownian matrix processes. While this is realized to a certain extent in Section 3.2, concrete application and detailed study of our approach remain open goals for future work.

The general philosophy of this thesis has been to rely on mathematical tools from Fourier analysis and stochastic calculus to study from an intrinsic point of view processes which have their values in matrix Lie groups. These tools lead to a closer understanding of the dynamics of these processes and a more detailed description of their statistics. This has allowed us to deal correctly with parametric and nonparametric estimation problems and to develop a filtering technique adapted to the specific structure of these processes.

Chapter 2

Rotation processes with applications in wave physics

This chapter is concerned with those parts of our work that were motivated by applications in wave physics, namely in polarization statistics and multiple scattering. It corresponds to our papers [63,66]. Our contribution has been to adapt for these applications a consistent mathematical development based on rotation Lévy processes.

The mathematical content of this chapter is organized into Sections 2.1 and 2.3. Sections 2.2 and 2.4 detail applications based on Sections 2.1 and 2.3, respectively. We have chosen to present our mathematical development independently of applications. Section 2.1 reviews the extension of the notion of characteristic function to rotation random variables. This is made possible by the formalism of Fourier analysis on the rotation group. Section 2.3 introduces rotation Lévy processes. It studies the two essential kinds of such processes, rotation compound Poisson processes and rotation Brownian motion. As such, there is a thematic separation between Sections 2.1 and 2.3. The former studies rotation random variables while the latter studies rotation processes.

Section 2.2 applies Section 2.1 to polarization statistics. Motivated by an experimental setting which shows the deficiency of the classical Stokes formalism of polarization, this section applies the results of Subsection 2.1.3 in order to generalize this formalism to higher order statistics. Within this new generalized formalism, the physical problem of depolarization is addressed using the results of Subsection 2.1.4.

Section 2.4 is based on Section 2.3. It is mainly devoted to the inverse problem of multiple scattering. A stochastic model of multiple scattering is considered within which this inverse problem can be identified with a nonparametric estimation problem. This is the so-called problem of decompounding on the rotation group. A solution of this problem is presented in Subsection 2.4.2.

Each section in the following will contain corresponding references, intended to place it in the context of recent mathematical and applied work. Sufficient background for our mathematical Sections 2.1 and 2.3 can be found in the books of Grenander [21] and Liao [40]. For prerequisite knowledge of fundamental probability we refer to Kallenberg's monograph [27]. Although Section 2.1 is grounded in Fourier analysis on the rotation group, it only requires a working knowledge of this topic. This is provided by the discussion of Subsection 2.1.1. For a detailed rigorous presentation see [74] or [50].

For general reference on Section 2.2, it is possible to use Brosseau's book [7]. For Section 2.4 see Ishimaru [25] or Chandrasekhar's classic [9]. We would like to point out the following two recent contributions which we find complementary to the point of view of this chapter. The paper [6] proposes

a spinor formalism of polarization. This goes beyond the usual Jones formalism, as it introduces a spinor quantity capable of determining the state of polarization of an electromagnetic wave in arbitrary directions. The thesis [58] indicates the importance of rotation Brownian motion to applications in biology and wave physics. Remarkably, the Riemannian geometry of Brownian paths is used to characterize several applications.

2.1 Characteristic functions of rotation random variables

An effective analysis of the probabilistic and algebraic properties of rotation random variables can be achieved using the corresponding characteristic functions. This section gives a self-contained presentation of the characteristic functions of rotation random variables. It is modelled on the recent paper [39]. Basic facts from Fourier analysis on the rotation group are recalled in Subsection 2.1.1. Characteristic functions are introduced in Subsection 2.1.2 where their relation to the symmetry properties of rotation random variables is specified. Subsection 2.1.3 considers the action of rotation random variables on spherical random variables. Subsection 2.1.4 reviews what is perhaps the most famous application of characteristic functions, the asymptotic properties of products of *i.i.d.* rotation random variables.

Currently, the characteristic functions of rotation random variables form an influential new tool in many engineering applications. This is illustrated in the book [67] which catalogues such applications in a wide range of engineering problems. The applications of characteristic functions considered here for rotation random variables are only a special case of similar methods for random variables with values in compact Lie groups. Such methods can be traced back to the work of Grenander [21] and were extensively developed by Heyer [24]. For their recent use in probability see [40].

2.1.1 Fourier analysis on $SO(3)$

Subsection 2.1.2 is based on the formalism of Fourier analysis on the rotation group. In practice, all that will be required is a straightforward application of Theorem 1 which we state in Paragraph 2.1.1.3. This is the Peter-Weyl theorem in the special case of the rotation group. Paragraphs 2.1.1.1 and 2.1.1.2 are intended to place this theorem in an adequate context. For the terminology and expressions introduced in Paragraphs 2.1.1.2 and 2.1.1.3, see [74].

The group of proper rotations of space can be identified with the matrix group $SO(3)$ consisting of all 3×3 real matrices R verifying the following conditions of orthogonality and orientation preservation

$$RR^T = I \qquad \det R = 1 \tag{2.1}$$

where ^{T} stands for the transpose and I is the identity matrix. We have that $SO(3)$ is a compact subset of the vector space of 3×3 real matrices. This is an essential condition for Theorem 1. It follows from the fact that $SO(3)$ is closed and bounded. That $SO(3)$ is closed can be seen from the continuity of the matrix operations in (2.1). From the orthogonality condition in (2.1) we have $|R| = \sqrt{3}$, where $|R|$ is the Euclidean matrix norm of R. This shows that $SO(3)$ is bounded.

The action of $SO(3)$ on $\mathbb{R}^3$ is determined if a canonical basis is chosen. A vector $s \in \mathbb{R}^3$ is then expressed by its coordinates $[s_1, s_2, s_3]$ in the canonical basis. Let $e_1, e_2, e_3 \in \mathbb{R}^3$ have the respective coordinates $[1,0,0]$, $[0,1,0]$ and $[0,0,1]$. For all $R \in SO(3)$ we have a linear transformation $s \mapsto Rs$. This is determined by the following transformation rules for coordinates and vectors

$$[s_1', s_2', s_3'] = R[s_1, s_2, s_3] \qquad [Re_1, Re_2, Re_3] = [e_1, e_2, e_3]^T R \tag{2.2}$$

where square brackets are taken to denote column matrices. Here $[s'_1, s'_2, s'_3]$ are the coordinates of Rs. We define a scalar product on $\mathbb{R}^3$ by $\langle e_i, e_j \rangle = \delta_{ij}$ for $1 \leq i, j \leq 3$, with the Kronecker delta notation. From (2.2), the orthogonality condition in (2.1) is equivalent to $\langle Re_i, Re_j \rangle = \delta_{ij}$.

2.1.1.1 The Euler angles parameterization

We here consider the parameterization of $SO(3)$ by ZYZ Euler angles. An important drawback of this parameterization is the fact that it is not defined at the identity $I \in SO(3)$. However, it will lead in Paragraph 2.1.1.2 to expressions with separated variables. This is desirable for practical computations as in Section 2.4. Note that the choice of ZYZ over ZXZ is the standard one in quantum mechanics [1].

For $1 \leq i \leq 3$ and $-\pi < \theta \leq \pi$ it is straightforward to check in (2.3) below that $R_i(\theta) \in SO(3)$. Geometrically, $R_i(\theta)$ is identified with a rotation of counterclockwise angle θ about the axis e_i. The parameterization of $SO(3)$ by means of ZYZ Euler angles decomposes all $R \in SO(3)$ into a product $R(\alpha, \beta, \gamma) = R_3(\alpha)R_2(\beta)R_3(\gamma)$ where $-\pi < \alpha, \gamma \leq \pi$ and $0 \leq \beta \leq \pi$. These three angles will be referred to as Euler angles of R.

$$R_1(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \quad R_2(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \quad R_3(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.3}$$

The mapping $(\alpha, \beta, \gamma) \mapsto R(\alpha, \beta, \gamma)$ defined for $-\pi < \alpha, \gamma \leq \pi$ and $0 \leq \beta \leq \pi$ is clearly a continuous one. We will state in Proposition 1 below that all $R \in SO(3)$ is of the form $R = R(\alpha, \beta, \gamma)$. In particular, this recovers the result that $SO(3)$ is connected, being the continuous image of a connected set. Proposition 1 can be extended to see the following. The mapping $(\alpha, \beta, \gamma) \mapsto R(\alpha, \beta, \gamma)$ gives a homeomorphism between the set $-\pi < \alpha, \gamma < \pi$ and $0 < \beta < \pi$ and the set of $R \in SO(3)$ verifying the three conditions $|R_{33}| \neq 1$, $R_{13} \neq -1$ and $|R_{31}| \neq 1$.

Proposition 1 *Let $R \in SO(3)$. The equation $R = R(\alpha, \beta, \gamma)$ for unknown $-\pi < \alpha, \gamma \leq \pi$ and $0 \leq \beta \leq \pi$ has at least one solution. This solution is unique iff $|R_{33}| \neq 1$.*

The main step in proving this proposition is to note that $R = R(\alpha, \beta, \gamma)$ for $-\pi < \alpha, \gamma \leq \pi$ and $0 \leq \beta \leq \pi$ only if

$$\begin{array}{lll} R_{33} = \cos\beta & R_{23} = \sin\beta\sin\alpha & R_{13} = \sin\beta\cos\alpha \\ & R_{32} = \sin\beta\sin\gamma & R_{31} = -\sin\beta\cos\gamma \end{array} \tag{2.4}$$

This results by matrix multiplication from (2.3).

Let $-\pi < \alpha, \gamma \leq \pi$ and $0 \leq \beta \leq \pi$ and note $R = R(\alpha, \beta, \gamma)$. The angles α, β and γ give a geometric construction of R. By replacing (2.4) for R and using (2.2) we have

$$\begin{array}{c} Re_3 = \cos\alpha\sin\beta e_1 + \sin\alpha\sin\beta e_2 + \cos\beta e_3 \\ \langle Re_2, R_3(\alpha)e_2 \rangle = \cos\gamma \end{array} \tag{2.5}$$

From the first formula β and α are the spherical angles, respectively the elevation and azimuth, of Re_3 with respect to the canonical basis. From the second formula γ is the counterclockwise angle from $R_3(\alpha)e_2$ to Re_2 about the axis Re_3. Note that the vector $R_3(\alpha)e_2$ is orthogonal to the plane of e_3 and Re_3.

To end this paragraph, we mention an additional aspect of (2.3). This will only be used in Subsection 2.3.3. For $1 \leq i \leq 3$ and $-\pi < \theta \leq \pi$ the matrices $R_i(\theta)$ satisfy the following identities

$$\left[\frac{d}{d\theta}R_i(\theta)\right]_{\theta=0} = J_i \qquad R_i(\theta) = \exp(\theta J_i) \tag{2.6}$$

where exp denotes the matrix exponential and the matrices J_1, J_2 and J_3 are given by

$$J_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad J_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \quad J_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.7}$$

These matrices are said to define infinitesimal rotations around e_1, e_2 and e_3, respectively. For all real numbers a_1, a_2 and a_3 if $J = a_1J_1 + a_2J_2 + a_3J_3$ and $R = \exp(\theta J)$ where $-\pi < \theta \leq \pi$ then $R \in SO(3)$ and R is identified with a rotation of counterclockwise angle θ about the axis $a_1e_1 + a_2e_2 + a_3e_3$. This follows from elementary properties of the matrix exponential, using the fact that the matrices J_1, J_2 and J_3 are antisymmetric.

2.1.1.2 Irreducible representations of $SO(3)$

The most natural statement of Theorem 1 is in terms of the irreducible representations of $SO(3)$. We here review this concept, state Theorem 1 and discuss its proof. We will consider the realization of the irreducible representations of $SO(3)$ in spaces of polynomial functions on the sphere S^2. Another usual realization involves spaces of symmetric tensors [1].

The sphere $S^2 \subset \mathbb{R}^3$ is the set of points $s \in \mathbb{R}^3$ such that $s_1^2 + s_2^2 + s_3^2 = 1$. We note σ the area measure on S^2. This is a rotation invariant measure given in spherical angles $-\pi \leq \varphi \leq \pi$ and $0 \leq \vartheta \leq \pi$ by $d\sigma = \sin\vartheta d\varphi d\vartheta$. Let $E \equiv L^2(S^2, \sigma)$ be the Hilbert space of functions $f : S^2 \to \mathbb{C}$ which are square integrable with respect to σ. This is equipped with the scalar product –since there is no risk of confusion, we use the same notation as for the scalar product on $\mathbb{R}^3$.

$$\langle f, g \rangle = \frac{1}{4\pi} \int_{S^2} f g^* d\sigma \tag{2.8}$$

for $f, g \in E$, where * stands for complex conjugation. The action of $SO(3)$ on E is defined as follows. For all $R \in SO(3)$ we have a unitary operator $U(R)$, where for $f \in E$ the function $U(R)f \in E$ is defined by

$$U(R)f(s) = f(R^T s) \tag{2.9}$$

for $s \in S^2$. The unitarity of $U(R)$ for all $R \in SO(3)$ follows from the rotation invariance of σ. It is straightforward to verify the homomorphism property

$$U(R_1R_2) = U(R_1)U(R_2) \tag{2.10}$$

for all $R_1, R_2 \in SO(3)$.

An invariant subspace $T \subset E$ is a linear subspace T of E such that for all $R \in SO(3)$ and $f \in T$ we have $U(R)f \in T$. If $T \subset E$ is an invariant subspace then the orthogonal complement $T^\perp$ is also invariant. This follows from the fact that $U(R)$ is unitary for all $R \in SO(3)$. Consider the following example. Let $E_e \subset E$ and $E_o \subset E$ be respectively the space of even and odd functions $f \in E$. Recall their definition. For $f \in E$ we have that $f \in E_e$ if $f(-s) = f(s)$ and $f \in E_o$ if $f(-s) = -f(s)$, for $s \in S^2$. We have that $E_e^\perp = E_o$. Each of these subspaces is invariant. For example, for $R \in SO(3)$ and $f \in E_o$ we have

$$U(R)f(-s) = f(-R^T s) = -f(R^T s) = -U(R)f(s)$$

for $s \in S^2$, so that $U(R)f \in E_o$. It is easily verified that E_e is a closed subspace of E. It follows that $E = E_e \oplus E_o$. In other words, E is an orthogonal sum of these two invariant subspaces.

By a representation of $SO(3)$ we mean a finite dimensional invariant subspace $T \subset E$. A representation T is said to be reducible if there exists an invariant subspace $T' \subset T$. If a representation

T is not reducible, it is said to be irreducible. If T is a reducible representation and $T' \subset T$ is an invariant subspace then T' is a representation. Also, the orthogonal complement T'' of T' in T is a representation. A simple recursive reasoning shows that all representation T is a finite orthogonal sum of irreducible representations.

We now introduce a complete family $(T^l)_{l\geq 0}$ of representations of $SO(3)$. This means that all representation T is a finite orthogonal sum

$$T = S^{l_{\min}} \oplus \ldots \oplus S^l \oplus \ldots \oplus S^{l_{\max}} \tag{2.11}$$

where for $0 \leq l_{\min} \leq l \leq l_{\max}$ we have that S^l is a representation in linear isomorphism with T^l. Define the functions $u, z \in E$ by the following expressions for $s \in S^2$

$$u(s) = s_1 + \mathtt{i}s_2 \qquad z(s) = s_3 \tag{2.12}$$

For $l \geq 0$ the representation T^l is of dimension $d_l = 2l + 1$. An orthonormal basis of T^l can be given in terms of spherical harmonics. These are the functions $Y_m^l \in E$ where $-l \leq m \leq l$ and for $s \in S^2$

$$Y_m^l(s) = \mathtt{i}^{m+|m|} \left[\frac{(l-|m|)!}{(l+|m|)!}\right]^{\frac{1}{2}} u^m(s) \left(P_l^{(m)} \circ z\right)(s) \tag{2.13}$$

Here P_l is the Legendre polynomial of degree l and $P_l^{(m)}$ denotes its derivative of order $|m|$. In particular, this is a real-valued polynomial of degree $l - |m|$. It follows from (2.12) and (2.13) that Y_m^l is a polynomial of degree l in s_1, s_2 and s_3. Moreover, Y_m^l is even or odd according to wether l is even or odd. We have the following orthogonality relations. For all $l, s \geq 0$ and corresponding m, p

$$\langle Y_m^l, Y_p^s \rangle = \frac{1}{d_l} \delta_{ls} \delta_{mp} \tag{2.14}$$

The following relation (2.15) results using (2.12). For all $l \geq 0$ and $-l \leq m \leq l$ we have

$$Y_m^{l^*} = (-1)^m Y_{-m}^l \tag{2.15}$$

The usual proof of the affirmation that $(T^l)_{l\geq 0}$ is a complete family of representations proceeds by an analysis of the infinitesimal action of $SO(3)$ in each subspace T^l. This is made tractable by the fact that for all $l \geq 0$ the spherical harmonics Y_m^l for $-l \leq m \leq l$ are eigenfunctions of rotations about e_3. Clearly the action (2.9) of a rotation $R_3(\theta)$ as in (2.3) is to multiply Y_m^l by the factor $e^{-\mathtt{i}m\theta}$. By Proposition 1 it is then enough to consider rotations about e_2.

For $l \geq 0$ and $-l \leq m, n \leq l$ let $U_{mn}^l : SO(3) \to \mathbb{C}$ be given by

$$U_{mn}^l(R) = d_l \langle U(R) Y_n^l, Y_m^l \rangle \tag{2.16}$$

Since T_l is invariant we have by (2.14)

$$U(R) Y_m^l = \sum_{n=-l}^{l} U_{nm}^l(R) Y_n^l \tag{2.17}$$

for all $R \in SO(3)$ and all $l \geq 0$ and $-l \leq m \leq l$.

For $l \geq 0$ the functions U_{mn}^l of (2.16) are called the matrix elements of T^l. For $l \geq 0$ and $R \in SO(3)$ let $U^l(R)$ be the matrix with elements $U_{mn}^l(R)$. The matrix function $R \mapsto U^l(R)$ satisfies the following identities, which can be shown from (2.16) and (2.17) using the fact that $U(R)$ is unitary.

$$U^l(R_1 R_2) = U^l(R_1) U^l(R_2) \qquad [U^l(R)]^{-1} = U^l(R^T) = [U^l(R)]^\dagger \tag{2.18}$$

for all $R, R_1, R_2 \in SO(3)$. Here † stands for the Hermitian transpose. The first identity gives the homomorphism property of U^l. The second one states that for all $R \in SO(3)$ the matrix $U^l(R)$ is unitary.

2.1.1.3 The Peter-Weyl theorem

Theorem 1 below is a version of the Peter-Weyl theorem in the special case of $SO(3)$. In particular, this theorem will state that all continuous function $h : SO(3) \to \mathbb{C}$ is a uniform limit of finite linear combinations of the matrix elements U^l_{mn} defined in (2.16) of the last paragraph. This property, along with identities (2.18), will give rise to the properties of characteristic functions studied in Subsection 2.1.2. For all continuous $h : SO(3) \to \mathbb{C}$ the coefficients of the linear combinations of the U^l_{mn} used to approximate h are given in terms of a scalar product for functions on $SO(3)$. This is defined as in (2.8) by introducing a rotation invariant measure on $SO(3)$. This is the Haar measure μ of $SO(3)$.

We will define μ in terms of the Euler angles. Let $h : SO(3) \to \mathbb{C}$ be continuous. By Paragraph 2.1.1.1, the expression $h(\alpha, \beta, \gamma) = h(R(\alpha, \beta, \gamma))$ defines a continuous function of $-\pi \leq \alpha, \gamma \leq \pi$ and $0 \leq \beta \leq \pi$. The Haar measure μ is defined by the following formula

$$\int_{SO(3)} h d\mu = \frac{1}{8\pi^2} \int_{-\pi}^{\pi} \int_{0}^{\pi} \int_{-\pi}^{\pi} h(\alpha, \beta, \gamma) \sin\beta d\alpha d\beta d\gamma \tag{2.19}$$

for all continuous $h : SO(3) \to \mathbb{C}$. For $P \in SO(3)$ let $\mathcal{L}_P : SO(3) \to SO(3)$ be the mapping given by $\mathcal{L}_P(R) = PR$. The left rotation invariance of μ consists in the following property

$$\int_{SO(3)} (h \circ \mathcal{L}_P) d\mu = \int_{SO(3)} h d\mu \tag{2.20}$$

for all $P \in SO(3)$ and h as above. An intuitive discussion of the validity of (2.19) and (2.20) can be found in [50]. The Haar measure μ has other important properties which we recover in Proposition 4 of the following subsection. Let $H \equiv L^2(SO(3), \mu)$ be the Hilbert space of functions $h : SO(3) \to \mathbb{C}$ which are square integrable with respect to μ. As in (2.8) define a scalar product on H

$$\langle h, k \rangle = \int_{SO(3)} h k^* d\mu \tag{2.21}$$

for $h, k \in H$. We have that the matrix elements (2.16) are in H and satisfy the following orthogonality relations. For all $l, s \geq 0$ and corresponding m, p, n, q

$$\langle U^l_{mn}, U^s_{pq} \rangle = \frac{1}{d_l} \delta_{sl} \delta_{mp} \delta_{nq} \tag{2.22}$$

For $f \in E$ and $l \geq 0$ define the $d_l \times 1$ column matrices $\hat{f}_l$ and Y^l as follows

$$\hat{f}_l = \left[\langle f, Y^l_{-l} \rangle, \ldots, \langle f, Y^l_l \rangle \right] \qquad Y^l = \left[Y^l_{-l}, \ldots, Y^l_l \right] \tag{2.23}$$

Similarly, for $h \in H$ and $l \geq 0$ define $\hat{h}^l$ the $d_l \times d_l$ matrix with elements

$$\hat{h}^l_{mn} = \langle h, U^{l^*}_{mn} \rangle \tag{2.24}$$

In the following statement of Theorem 1 we will use (2.23) and (2.24). We call $\hat{f}_l$ and $\hat{h}^l$ the Fourier coefficients of functions f and h respectively. The series (2.25) and (2.26) are called the Fourier series of f and h. For two different proofs of Theorem 1 see [5, 74] and [50]. The proof in [74] or [5] uses the spectral theory of compact operators. It applies in general when $SO(3)$ is replaced by some other compact group. The proof in [50] uses the Stone-Weierstrass theorem and is specific to the case of $SO(3)$ where it shows that the matrix elements U^l_{mn} are actually polynomial functions in the elements of the matrix $R \in SO(3)$.

Theorem 1 (Peter-Weyl) *The following hold*
(i) *For all $f \in E$ the following series (2.25) converges to f in E. If f is continuous, then (2.25) converges to f uniformly.*

$$f = \sum_{l\geq 0} d_l \hat{f}_l^T Y^l = \sum_{l\geq 0} d_l \sum_{m=-l}^{l} \hat{f}_l^m Y_m^l \tag{2.25}$$

(ii) *For all $h \in H$ the following series (2.26) converges to h in H. If h is continuous, then (2.26) converges to h uniformly.*

$$h = \sum_{l\geq 0} d_l \operatorname{tr}\left(\hat{h}^l U^{l\dagger}\right) = \sum_{l\geq 0} d_l \sum_{m,n=-l}^{l} \hat{h}_{mn}^l U_{mn}^{l^*} \tag{2.26}$$

where tr *denotes the trace of a matrix.*

To conclude this paragraph we express the matrix elements $U_{mn}^l \in H$ in terms of the Euler angles. By Proposition 1 of Paragraph 2.1.1.1 we have that all $R \in SO(3)$ can be written $R = R(\alpha, \beta, \gamma)$ for some Euler angles α, β and γ. Using the homomorphism property (2.18) we have for all $l \geq 0$ that

$$U^l(R) = U^l(R(\alpha,\beta,\gamma)) = U^l(R_3(\alpha))U^l(R_2(\beta))U^l(R_3(\gamma)) \tag{2.27}$$

We have already noted in Paragraph 2.1.1.1 that for all $l \geq 0$ and $-l \leq m \leq l$ the function $Y_m^l \in E$ is an eigenfunction of $U^l(R_3(\theta))$ with corresponding eigenvalues $e^{-\mathtt{i}m\theta}$, where $-\pi < \theta \leq \pi$. Replacing in (2.16) and then in (2.27) we obtain for all $l \geq 0$ and $-1 \leq m, n \leq l$

$$U_{mn}^l(R(\alpha,\beta,\gamma)) = e^{-\mathtt{i}m\alpha} U_{mn}^l(R_2(\beta)) e^{-\mathtt{i}n\gamma}$$

We refer to [74] for the fact that the factor $U_{mn}^l(R_2(\beta))$ above can be written

$$U_{mn}^l(R_2(\beta)) = P_{mn}^l(\cos\beta)$$

where the functions P_{mn}^l can be expressed in terms of Jacobi polynomials. In particular, for all $l \geq 0$ we have $P_{00}^l = P_l$ the Legendre polynomial of degree l. We will use the following expression for (2.27)

$$U_{mn}^l(\alpha,\beta,\gamma) = e^{-\mathtt{i}m\alpha} P_{mn}^l(\cos\beta) e^{-\mathtt{i}n\gamma} \tag{2.28}$$

where we have put $U_{mn}^l(R(\alpha,\beta,\gamma)) = U_{mn}^l(\alpha,\beta,\gamma)$.

As an application of (2.28) it is possible for $l \geq 0$ to calculate the function $\chi^l \in H$ given by (2.29) below. The functions χ^l for $l \geq 0$ are known as the irreducible characters corresponding to the representations T^l of Paragraph 2.1.1.2.

$$\chi^l(R) = \operatorname{tr}(U^l(R)) \tag{2.29}$$

Note that if $R = R_3(\theta)$ for some $-\pi < \theta \leq \pi$ we have by (2.28)

$$\chi^l(R) = \chi^l(R_3(\theta)) = \sum_{m=-l}^{l} e^{-\mathtt{i}m\theta} = \frac{\sin(l+\frac{1}{2})\theta}{\sin\frac{\theta}{2}}$$

where it should be understood that $\chi^l(R) = d_l$ when $\theta = 0$.

For all $R \in SO(3)$ we have that if $0 \leq \theta \leq \pi$ is such that $\operatorname{tr}(R) = 2\cos\theta + 1$ then there exists $P \in SO(3)$ satisfying $PRP^T = R_3(\theta)$. It follows that for $l \geq 0$

$$\chi^l(R) = \operatorname{tr}\left(U^l(P)U^l(R_3(\theta))U^l(P^T)\right) = \operatorname{tr}(U^l(R_3(\theta))) = \frac{\sin(l+\frac{1}{2})\theta}{\sin\frac{\theta}{2}} \tag{2.30}$$

note that θ is uniquely determined by R. It is possible to replace θ in the above expression by the continuous function

$$a(R) = \arccos\left(\frac{\mathrm{tr}(R) - 1}{2}\right) \tag{2.31}$$

For $R \in SO(3)$ we have that $a(R)$ is the angle of the proper rotation R. Formula (2.30) will be useful in Proposition 6 of the following subsection.

2.1.2 Characteristic functions

The characteristic functions of rotation random variables are the main object of the current section. In order to study rotation random variables we need to assume given a complete probability space $(\Omega, \mathcal{A}, \mathbb{P})$. A rotation random variable is then understood as a random matrix X which takes its values in $SO(3)$. Let X be a rotation random variable, the characteristic function of X is the sequence $\phi_X = \{\phi_X(l)\}_{l\geq 0}$ of $d_l \times d_l$ complex matrices given by

$$\phi_X(l) = \mathbb{E}[U^l(X)] \tag{2.32}$$

where $\mathbb{E}$ denotes expectation with respect to $\mathbb{P}$. Paragraph 2.1.2.1 starts by considering the relation of characteristic functions to fundamental probabilistic concepts. Paragraph 2.1.2.2 specifies the symmetry properties of rotation random variables in terms of their characteristic functions. As stated before, only a general understanding of Subsection 2.1.1 is needed. Namely, we will mostly refer to the identities (2.18) of Paragraph 2.1.1.2 and to Theorem 1 of Paragraph 2.1.1.3.

2.1.2.1 General properties

Proposition 2 below ensures that the characteristic function ϕ_X of a rotation random variable X as introduced in formula (2.32) is well-defined. This proposition also generalizes the so-called Glivenko's theorem for vector-valued random variables to rotation random variables. Proposition 3 does the same for Kac's theorem. For these two theorems in the case of vector-valued random variables see for instance [3]. Note that by vector-valued we mean $\mathbb{R}^d$-valued for some $d \geq 1$. Proposition 4 uses characteristic functions to study the Haar measure. For the results of the current paragraph we refer to [21].

Proposition 2 *The following hold*
(i) *Let X and Y be rotation random variables. $X \stackrel{d}{=} Y$ iff $\phi_X = \phi_Y$.*
(ii) *Let $(X_n)_{n\geq 1}$ be rotation random variables. $X_n \stackrel{d}{\to} X$ for some rotation random variable X iff $\phi_{X_n} \to \phi_X$.*

Proof: We start by proving (i). Remember that $X \stackrel{d}{=} Y$ denotes equality in distribution of X and Y. This means that for all continuous $h : SO(3) \to \mathbb{C}$ we have

$$\mathbb{E}[h(X)] = \mathbb{E}[h(Y)] \tag{2.33}$$

Suppose $X \stackrel{d}{=} Y$. For all $l \geq 0$ we have that U^l is a continuous matrix function. In particular, for all $_l \leq m, n \leq l$ the matrix element $U^l_{mn} : SO(3) \to \mathbb{C}$ is continuous. It follows by applying (2.33) that for all $l \geq 0$ we have

$$\phi_X(l) = \mathbb{E}[U^l(X)] = \mathbb{E}[U^l(Y)] = \phi_Y(l)$$

that is, $\phi_X = \phi_Y$.

Suppose now $\phi_X = \phi_Y$. Let $h : SO(3) \to \mathbb{C}$ be a continuous function. We have the uniform limit (2.26) for h. It follows by dominated convergence that

$$\mathbb{E}[h(X)] = \sum_{l \geq 0} d_l \operatorname{tr}\left(\hat{h}^l \phi_X(l)^\dagger\right) = \sum_{l \geq 0} d_l \operatorname{tr}\left(\hat{h}^l \phi_Y(l)^\dagger\right) = \mathbb{E}[h(Y)]$$

This shows that $X \stackrel{d}{=} Y$.

The proof of (ii) can be carried out similarly. Remember that $X_n \stackrel{d}{\to} X$ denotes limit in distribution. This means that for all continuous $h : SO(3) \to \mathbb{C}$ we have

$$\mathbb{E}[h(X)] = \lim_n \mathbb{E}[h(X_n)] \tag{2.34}$$

Compared to the proof of (i), we have the additional difficulty of exchanging limits in distribution and infinite sums. This can again be treated using dominated convergence.▲

Proposition 3 *The following hold*
(i) *Let X be a rotation random variable and $\mathcal{F} \subset \mathcal{A}$ a σ-subalgebra. We have that X is independent of $\mathcal{F}$ iff for all $l \geq 0$*

$$\mathbb{E}[U^l(X)|\mathcal{F}] = \phi_X(l) \tag{2.35}$$

(ii) *Let X and Y be independent rotation random variables and note $Z = XY$. We have for all $l \geq 0$*

$$\phi_Z(l) = \phi_X(l)\phi_Y(l) \tag{2.36}$$

Proof: We start with (i). Remember that X is independent of $\mathcal{F}$ iff for all continuous $h : SO(3) \to \mathbb{C}$ we have

$$\mathbb{E}[h(X)|\mathcal{F}] = \mathbb{E}[h(X)] \tag{2.37}$$

It is clear that the proof can be carried out as for (i) of Proposition 2, using the dominated convergence property of conditional expectation.

We now turn to (ii). For rotation random variables X and Y note $Z = XY$ and $\mathcal{F} = \sigma(Y)$. By the homomorphism property (2.18), we have for $l \geq 0$

$$\phi_Z(l) = \mathbb{E}[U^l(Z)] = \mathbb{E}[U^l(X)U^l(Y)]$$

Noting $\tilde{\phi}_X(l) = \mathbb{E}[U^l(X)|\mathcal{F}]$ and replacing in the last equality we have for $l \geq 0$

$$\phi_Z(l) = \mathbb{E}[\tilde{\phi}_X(l)U^l(Y)]$$

If X and Y are independent we have by (i) that $\tilde{\phi}_X(l) = \phi_X(l)$ for $l \geq 0$. The proof can be completed immediately.▲

Proposition 4 recovers certain properties of the Haar measure using characteristic functions. Indeed replacing X by a constant $R \in SO(3)$ we have in (ii) of this proposition that μ is invariant by left and right rotations. Moreover, it results by comparing (i) to Proposition 5 of the next Paragraph 2.1.2.2 that μ is invariant by inversion. In addition to its invariance properties, μ is especially important due to its role in Subsection 2.1.4.

We will say that a rotation random variable U is uniformly distributed if the probability law of U is the Haar measure μ. In other words, if for all continuous $h : SO(3) \to \mathbb{C}$ we have

$$\mathbb{E}[h(U)] = \int_{SO(3)} h d\mu \tag{2.38}$$

Proposition 4 *The following hold*
(i) *A rotation random variable U is uniformly distributed iff for all $l > 0$ we have $\phi_U(l) = 0$.*
(ii) *Let U be a uniformly distributed rotation random variable. If X is a rotation random variable independent of U then $XU \stackrel{d}{=} U$ and $UX \stackrel{d}{=} U$.*

Proof: We start with (i). The *if* part follows easily using (2.38) and the uniform limit (2.26) for h. For the *only if* part the proof can be carried out by contradiction. Suppose U is uniformly distributed and for some $l > 0$ we have $\phi_U(l) \neq 0$. We have that there exists a $d_l \times 1$ column matrix $f_l \neq 0$ such that

$$\phi_U(l) f_l = \left(\int_{SO(3)} U^l d\mu \right) f_l \neq 0$$

where we have used (2.38). Let $\hat{f}_l = \phi_U(l) f_l \neq 0$, for all $R \in SO(3)$ we have

$$U^l(R)\hat{f}_l = \left(\int_{SO(3)} (U^l \circ \mathcal{L}_R) d\mu \right) f_l = \left(\int_{SO(3)} U^l d\mu \right) f_l = \hat{f}_l \tag{2.39}$$

as follows from the left invariance of μ as expressed by condition (2.20). Consider now the function $f \in T^l$ given by

$$f = \hat{f}_l^T Y^l = \sum_{m=-l}^{l} \hat{f}_l^m Y_m^l \tag{2.40}$$

We have by (2.17) and (2.39) that for all $R \in SO(3)$

$$U(R)f = \sum_{m,n=-l}^{l} U_{mn}^l(R) \hat{f}_l^n Y_m^l = \sum_{m=-l}^{l} \hat{f}_l^m Y_m^l = f$$

Thus $f \in T^l$ is nonzero and invariant by the action of $SO(3)$ as in (2.9). This contradicts the fact that T^l is an irreducible representation. We have achieved our contradiction and it follows that we have $\phi_U(l) = 0$ for all $l > 0$.

We turn to the proof of (ii). This is an immediate application of (i). Suppose X is independent of U and note $Z = XU$. We have by (ii) of Proposition 3 and (i) of the current proposition that for all $l > 0$

$$\phi_Z(l) = \phi_X(l)\phi_U(l) = \phi_U(l)$$

It follows by (i) of Proposition 2 that $Z \stackrel{d}{=} U$. It is shown in the same way that $UX \stackrel{d}{=} U$.▲

2.1.2.2 Symmetries of rotation random variables

The symmetries of a rotation random variable X refer to the invariance properties of its distribution under certain algebraic operations. We have seen in Proposition 2 that the distribution of X is completely determined by its characteristic function ϕ_X. Here we specify certain symmetry properties of X in terms of ϕ_X. The properties of inverse invariance and conjugate invariance are considered in Propositions 5 and 6, respectively. These two properties are considered in [39, 40]. In Proposition 7 we introduce the property of zonal invariance which we will find useful in the multiple scattering problems considered in Section 2.4.

A rotation random variable X is said to be inverse invariant if $X \stackrel{d}{=} X^T$. It is said to be conjugate invariant if $RXR^T \stackrel{d}{=} X$ for all $R \in SO(3)$. We will say that X is zonal invariant if $X \stackrel{d}{=} R_3(\alpha) X R_3(\gamma)$ for all $-\pi < \alpha, \gamma \leq \pi$. When referring to the probability density of X, we mean a probability density

with respect to μ. This is a real-valued function $p \in H$ such that for all continuous $h : SO(3) \to \mathbb{C}$ we have

$$\mathbb{E}[h(X)] = \int_{SO(3)} ph d\mu \tag{2.41}$$

If X has a probability density p then the Fourier series (2.26) of p is given by ϕ_X as follows

$$\phi_X(l) = \int_{SO(3)} pU^l d\mu \qquad p = \sum_{l \geq 0} d_l \operatorname{tr}\left(\phi_X(l) U^{l\dagger}\right) \tag{2.42}$$

Note that (iii) of Proposition 5 leads to a practical way of generating an inverse invariant rotation random variable X from a general rotation random variable Z. The following development is possible using (i) of this proposition. Suppose a rotation random variable X has a probability density p. We have that X is inverse invariant *iff* $p = p \circ J$ in H. Here $J : SO(3) \to SO(3)$ denotes the inverse map $J(R) = R^T$. Indeed, replacing in (2.42) the fact that $\phi_X(l)$ is Hermitian for all $l \geq 0$ we have

$$p = \sum_{l \geq 0} \frac{d_l}{2} \operatorname{tr}\left[\phi_X(l)\left(U^l + U^{l\dagger}\right)\right] \tag{2.43}$$

The required property of p can be checked by replacing the unitarity property (2.18) in this expression.

Proposition 5 (Inverse invariance) *Let X, Z be rotation random variables. The following hold*
(i) *X is inverse invariant iff for all $l \geq 0$ we have that $\phi_X(l)$ is Hermitian.*
(ii) *Suppose X is inverse invariant. If $X_1, \ldots, X_n$ are independent copies of X then their product $X_1 \ldots X_n$ is inverse invariant.*
(iii) *If $Y \stackrel{d}{=} Z$ and Y is independent of Z then ZY^T is inverse invariant.*

Proof: The proof of (i) is as follows. Note $Y = X^T$. By (i) of Proposition 2 we have that X is inverse invariant *iff* $\phi_X = \phi_Y$. We have by the unitarity property (2.18) for $l \geq 0$

$$\phi_Y(l) = \mathbb{E}[U^l(X^T)] = \mathbb{E}[U^l(X)]^\dagger = \phi_X^\dagger(l)$$

It follows that X is inverse invariant *iff* for all $l \geq 0$ we have $\phi_X(l) = \phi_X^\dagger(l)$. This is precisely the condition that $\phi_X(l)$ is Hermitian.

For the proof of (ii) note $Y = X_1 \ldots X_n$. By (ii) of Proposition 3 we have that

$$\phi_Y(l) = \phi_{X_1}(l) \ldots \phi_{X_n}(l)$$

for all $l \geq 0$. It is enough to use (i) after noticing that the powers of a Hermitian matrix are Hermitian. The proof of (iii) uses a similar argument.▲

Proposition 6 (Conjugate invariance) *Let X be a rotation random variable. The following hold*
(i) *X is conjugate invariant iff for all $l \geq 0$ we have $\phi_X(l) = a_l I_l$ where $a_l \in \mathbb{R}$ and I_l is the $d_l \times d_l$ identity matrix.*
(ii) *X is conjugate invariant iff $XZ \stackrel{d}{=} ZX$ for all rotation random variable Z independent of X.*
(iii) *Let Z be a rotation random variable independent of X. If X and Z are conjugate invariant then XZ is conjugate invariant.*
(iv) *Suppose X has probability density p. X is conjugate invariant iff*

$$p = \sum_{l \geq 0} d_l a_l \chi^l \tag{2.44}$$

Proof: (ii) and (iii) result from (i) using (i) of Proposition 2 and (ii) of Proposition 3. Also, (iv) results from (i) and (2.42). We give the proof of (i).

By definition, X is conjugate invariant *iff* for all $R \in SO(3)$ and all $l \geq 0$ we have

$$U^l(R)\phi_X(l) = \phi_X(l)U^l(R) \tag{2.45}$$

this results from (i) of Proposition 2 using the homomorphism property (2.18). For all $R \in SO(3)$ and all $l \geq 0$ (2.45) states that $\phi_X(l)$ commutes with $U^l(R)$. It follows from Schur's lemma that $\phi_X(l) = a_l I_l$ where $a_l \in \mathbb{C}$. It remains to see that for $l \geq 0$ we have $a_l \in \mathbb{R}$. Note that from (2.29) and (2.32)

$$d_l a_l = \mathbb{E}[\chi^l(X)] \tag{2.46}$$

The proof is complete, since by (2.30) the function χ^l is real-valued.▲

For Proposition 7 below, we use the following notation. Let X be a fixed rotation random variable, we write

$$\phi^l_{mn} = \mathbb{E}[U^l_{mn}(X)] \tag{2.47}$$

for $l \geq 0$ and $-l \leq m, n \leq l$. For all $l \geq 0$ we thus have that the complex numbers ϕ^l_{mn} are the elements of the matrix $\phi_X(l)$. Note that (i) of Proposition 7 states that X is zonal invariant *iff* for all $l \geq 0$ the only nonzero element of $\phi_X(l)$ is ϕ^l_{00}. Suppose X has a probability density p. It follows from (2.42) that X is zonal invariant *iff* the expression of p in Euler angles is given by –remember from Paragraph 2.1.1.3 that $P^l_{00} = P_l$.

$$p \equiv p(\cos\beta) = \sum_{l \geq 0} d_l x_l P_l(\cos\beta) \tag{2.48}$$

where for $l \geq 0$ we note $x_l = \phi^l_{00}$. We will return to formula (2.48) in Section 2.4.

Proposition 7 (Zonal invariance) *Let X, Z be rotation random variables. The following hold*
(i) *X is zonal invariant iff for all $l \geq 0$ and $-l \leq m, n \leq l$ we have $\phi^l_{mn} = x_l \delta_{m0}\delta_{n0}$ where $x_l \in \mathbb{R}$.*
(ii) *If X and Z are zonal invariant and independent then XZ is zonal invariant.*

Proof: Note that (ii) follows from (i) and (ii) of Proposition 3. The proof of (i) is as follows. By definition, X is zonal invariant *iff* for all $-\pi < \alpha, \gamma \leq \pi$ and all $l \geq 0$ we have

$$U^l(R_3(\alpha))\phi_X(l)U^l(R_3(\gamma)) = \phi_X(l)$$

this results from (i) of Proposition 2 using the homomorphism property (2.18). Using expression (2.28) we obtain for the matrix elements ϕ^l_{mn} defined in (2.47)

$$e^{-\mathrm{i}m\alpha}\phi^l_{mn}e^{-\mathrm{i}n\gamma} = \phi^l_{mn}$$

for all $l \geq 0$ and $-l \leq m, n \leq l$. Since this relation must be verified for all $-\pi < \alpha, \gamma \leq \pi$ we have $\phi^l_{mn} = x_l \delta_{m0}\delta_{n0}$ where $x_l \in \mathbb{C}$. In order to show that $x_l \in \mathbb{R}$ for $l \geq 0$ it is enough to note by (2.47) that $x_l = \mathbb{E}[U^l_{00}(X)]$ and that U^l_{00} is real-valued.▲

2.1.3 Action on spherical random variables

The current subsection considers the action of rotation random variables on spherical random variables. Its results will be applied in Section 2.2. By a spherical random variable we mean a random vector

S which takes its values in S^2. If S is a spherical random variable and X a rotation random variable then it is natural to consider the spherical random variable S' given by

$$S' = XS \tag{2.49}$$

We are interested in characterizing the relation between the distributions of these three random variables. As in Subsection 2.1.2, the distribution of the rotation random variable X will be studied using its characteristic function ϕ_X. In order to obtain the distributions of the spherical random variables S and S' the current subsection will introduce the characteristic function of a spherical random variable. This will be noted as in Subsection 2.1.2. For example, in relation to (2.49) we will write ϕ_S or $\phi_{S'}$. The following development concerning the characteristic functions of spherical random variables contains many similarities to the development of Subsection 2.1.2.

The characteristic function of a spherical random variable S is the sequence $\phi_S = \{\phi_S(l)\}_{l\geq 0}$ of $d_l \times 1$ column matrices given by

$$\phi_S(l) = \mathbb{E}[Y^{l^*}(S)] \tag{2.50}$$

remember that the column matrices Y^l were defined in (2.23) of Paragraph 2.1.1.3. The probability density of a spherical random variable S is a real-valued function $p \in E$ such that for all continuous $f : SO(3) \to \mathbb{C}$ we have

$$\mathbb{E}[f(S)] = \frac{1}{4\pi}\int_{S^2} pf d\sigma \tag{2.51}$$

If S has a probability density p then the Fourier series (2.25) of p is given by ϕ_S as follows

$$\phi_S(l) = \frac{1}{4\pi}\int_{S^2} pY^{l^*} d\sigma \qquad p = \sum_{l\geq 0} d_l \phi_S^T(l) Y^l \tag{2.52}$$

Our main result in relation to (2.49) is Proposition 9. Proposition 10 is concerned with uniformly distributed spherical random variables, while Proposition 11 introduces the notion of zonal invariant spherical random variable. We start by stating Proposition 8 which ensures that (2.50) leads to a correct definition. The proof of this proposition is identical to that of Proposition 2 of Paragraph 2.1.2.1, using (2.25) instead of (2.26).

Proposition 8 *The following hold*
(i) *Let S and S' be spherical random variables. $S \overset{d}{=} S'$ iff $\phi_S = \phi'_S$.*
(ii) *Let $(S_n)_{n\geq 1}$ be spherical random variables. $S_n \overset{d}{\to} S$ for some spherical random variable S iff we have $\phi_{S_n} \to \phi_S$.*

Proposition 9 *Let S, S' and X be as in (2.49). If S and X are independent then for $l \geq 0$ we have*

$$\phi_{S'}(l) = \phi_X(l)\phi_S(l) \tag{2.53}$$

Proof: We start by rewriting identity (2.17). Replacing R by R^T and conjugating both members of this identity we have for all $l \geq 0$ and $-l \leq m \leq l$

$$U(R^T)Y_m^{l^*}(s) = Y_m^{l^*}(Rs) = \sum_{n=-l}^{l} U_{mn}^l(R) Y_n^{l^*}(s) \tag{2.54}$$

where $R \in SO(3)$ and $s \in S^2$. Using (2.54) in matrix form we have from (2.49)

$$Y^{l^*}(S') = U^l(X) Y^{l^*}(S) \tag{2.55}$$

Note $\mathcal{F} = \sigma(S)$ and $\tilde{\phi}_X(l) = \mathbb{E}[U^l(X)|\mathcal{F}]$. We have from (2.55)

$$\phi_{S'}(l) = \mathbb{E}[\tilde{\phi}_X(l)Y^{l^*}(S)] \tag{2.56}$$

If X and S are independent then by (i) of Proposition 3 we have $\tilde{\phi}_X(l) = \phi_X(l)$. Replacing in (2.56), the proof can be completed.▲

In the current subsection, the following Proposition 10 can be seen as the counterpart of Proposition 4 of Subsection 2.1.2. Uniformly distributed spherical random variables are defined in a similar way to (2.38) using the area measure σ on S^2. We will say that a spherical random variable W is uniformly distributed if the probability law of W is the normalized area measure $(1/4\pi)\sigma$. In other words, if for all continuous $f : S^2 \to \mathbb{C}$ we have

$$\mathbb{E}[f(W)] = \frac{1}{4\pi}\int_{S^2} f d\sigma \tag{2.57}$$

In Proposition 10 we will see that a uniformly distributed spherical random variable can be obtained from a uniformly distributed rotation random variable using (2.49).

Proposition 10 *The following hold*
(i) *A spherical random variable W is uniformly distributed iff for all $l > 0$ we have $\phi_W(l) = 0$.*
(ii) *Let S be a spherical random variable and U be a rotation random variable such that S and U are independent. If U is uniformly distributed then US is uniformly distributed.*

Proof: The proof of (i) will use the same idea as for (i) of Proposition 4. The *if* part follows from (2.57) and the uniform limit (2.25) for f. The *only if* part can be proved by contradiction. Suppose for some $l > 0$ we have $\phi_W(l) \neq 0$ and note ϕ^l_m for $-l \leq m \leq l$ the elements of $\phi_W(l)$ –remember that this is a $d_l \times 1$ column matrix. For all $R \in SO(3)$ and $-l \leq m \leq l$ we have by (2.57) and (2.54)

$$\sum_{n=-l}^{l} U^l_{mn}(R)\phi^l_n = \frac{1}{4\pi}\int_{S^2}\sum_{n=-l}^{l} U^l_{mn}(R)Y^{l^*}_n d\sigma = \frac{1}{4\pi}\int_{S^2} U(R^T)Y^{l^*}_m d\sigma = \frac{1}{4\pi}\int_{S^2} Y^{l^*}_m d\sigma = \phi^l_m \tag{2.58}$$

where for the last step we have used the rotation invariance of σ. Let $f \in T^l$ be given by $f = \phi^T_W(l)Y^l$. We have by (2.17) and (2.58) for all $R \in SO(3)$

$$U(R)f = \sum_{m,n=-l}^{l} U^l_{mn}(R)\phi^l_n Y^l_m = \sum_{m=-l}^{l} \phi^l_n Y^l_m = \phi^T_W(l)Y^l = f$$

That is, $f \in T^l$ is nonzero and invariant by the action of $SO(3)$ as in (2.9). Since T^l is an irreducible representation, we have the desired contradiction.

We now prove (ii). Note $W = US$. Since U and S are independent, it is possible to apply Proposition 9. We have for $l > 0$

$$\phi_W(l) = \phi_U(l)\phi_S(l) = 0$$

It follows by (i) that W is uniformly distributed.▲

Proposition 11 introduces the property of zonal invariance for spherical random variables. A spherical random variable S is said to be be zonal invariant if for all $-\pi < \theta \leq \pi$ we have $R_3(\theta)S \overset{d}{=} S$. This property is closely related to the zonal invariance of rotation random variables. In particular, we will see in Proposition 11 that zonal invariant spherical random variables can be obtained from zonal invariant rotation random variables using (2.49). We will use the following notation. If S is a fixed spherical random variable, we will note for $l \geq 0$ and $-l \leq m \leq l$

$$\phi^l_m = \mathbb{E}[Y^{l^*}_m(S)] \tag{2.59}$$

For all $l \geq 0$, we thus have that the complex numbers ϕ_m^l are the elements of the column matrix $\phi_S(l)$. From (i) of Proposition 11 we have a formula similar to (2.48). Indeed, (i) below states that S is zonal invariant *iff* for all $l \geq 0$ the only nonzero element of $\phi_S(l)$ is ϕ_0^l. Suppose S has a probability density p. It follows from (2.52) that S is zonal invariant *iff* the expression of p in spherical angles $-\pi \leq \varphi \leq \pi$ and $0 \leq \vartheta \leq \pi$ is given by –remember from (2.13) that $Y_0^l = P_l \circ z$.

$$p \equiv p(\cos\vartheta) = \sum_{l \geq 0} d_l x_l P_l(\cos\vartheta) \tag{2.60}$$

where for $l \geq 0$ we note $x_l = \phi_0^l$.

Proposition 11 *Let S, S' and X be as in (2.49) with X and S independent. The following hold*
(i) *S is zonal invariant iff for all $l \geq 0$ and $-l \leq m \leq l$ we have $\phi_m^l = x_l \delta_{m0}$ where $x_l \in \mathbb{R}$.*
(ii) *If X is zonal invariant then S' is zonal invariant.*

Proof: Note that (ii) is an application of (i) and of Proposition 9. In order to prove (i) note that by (2.54) and expression (2.28) we have for all $-\pi < \theta \leq \pi$

$$Y_m^{l^*}(R_3(\theta)S) = \sum_{n=-l}^{l} U_{mn}^l(R_3(\theta)) Y_n^{l^*}(S) = e^{-\mathtt{i}m\theta} Y_m^{l^*}(S)$$

for all $l \geq 0$ and $-l \leq m \leq l$. It follows from (i) of Proposition 8 that S is zonal invariant *iff*

$$\phi_m^l = e^{-\mathtt{i}m\theta} \phi_m^l$$

for all $l \geq 0$ and $-l \leq m \leq l$, where we use the notation (2.59). Since this relation must be verified for all $-\pi < \theta \leq \pi$ it follows that $\phi_m^l = x_l \delta_{m0}$, where by (2.13) we have $x_l = \mathbb{E}[Y_0^l(S)] \in \mathbb{R}$. The proof of (i) is complete.▲

2.1.4 Products of *i.i.d.* rotation random variables

Here we are interested in the asymptotic behavior of products of *i.i.d.* rotation random variables. While this is a vast subject, we only present its most elementary aspect. Theorem 2 gives a necessary and sufficient condition for the convergence in distribution of such products. This will be applied in Section 2.2 in order to address the physical problem of depolarization.

The Haar measure μ plays an important role in the study of products of *i.i.d.* rotation random variables. Based on [21], we consider in the current paragraph a partial result which will occur in Section 2.2. This is stated below in the form of Theorem 2. Let $(Y_n)_{n \geq 1}$ be *i.i.d.* rotation random variables. Note $(X_n)_{n \geq 1}$ their cumulative products

$$X_n = Y_1 \dots Y_n \tag{2.61}$$

For a uniformly distributed rotation random variable U, Theorem 2 states a general algebraic condition necessary and sufficient for $X_n \xrightarrow{d} U$. Remember that U is uniformly distributed if the probability law of U is the Haar measure μ.

Using (ii) of Proposition 2, the convergence $X_n \xrightarrow{d} U$ is equivalent to that of the characteristic functions ϕ_{X_n}. For $n \geq 1$ we will note $\phi_n \equiv \phi_{X_n}$. Since $(Y_n)_{n \geq 1}$ have the same distribution, it follows by (i) of Proposition 2 that they have the same characteristic function. This will be noted $\phi \equiv \phi_{Y_1}$. Applying Proposition 3 to (2.61), it is possible to obtain the following relation in terms of characteristic functions

$$\phi_n(l) = [\phi(l)]^n \tag{2.62}$$

for all $n \geq 1$ and $l \geq 0$. In order to derive (2.62) from (2.61) note that for all $n > 1$ we have that Y_n is independent of X_{n-1}. Thus, it is enough to apply (ii) of Proposition 3 after writing $X_n = X_{n-1}Y_n$. Using (2.62), the study of (2.61) now reduces to that of the matrix powers of $\phi(l)$ for $l > 0$.

The rotation random variable Y_1 is said to be supported by a closed set $B \subset SO(3)$ if $\mathbb{P}(Y_1 \in B) = 1$. Clearly, this depends only on the distribution of Y_1. If Y_1 is supported by a closed set $B \subset SO(3)$ then this also holds for all Y_n. For $l \geq 1$ note $\lambda(l) \subset \mathbb{C}$ the spectrum of $\phi(l)$. Theorem 2 is based on a characterization of the spectra $\lambda(l)$ in terms of the closed sets supporting Y_1. For clarity, this fact will be stated separately in Proposition 12.

Theorem 2 *Let U be a uniformly distributed rotation random variable. $X_n \xrightarrow{d} U$ iff Y_1 is not supported by any closed proper subgroup of $SO(3)$ or by any coset of such a subgroup.*

Proof: We only give a sketch of the proof. A detailed version can be found in [21]. Consider the *only if* part. Suppose Y_1 is supported by a closed proper subgroup $G \subset SO(3)$. For all $n \geq 1$ we have $\mathbb{P}(X_n \in G) = 1$. Indeed,

$$\mathbb{P}(X_n \notin G) \leq \mathbb{P}[\cup_{m=1}^{n}(Y_m \notin G)] \leq \sum_{m=1}^{n} \mathbb{P}(Y_m \notin G) = 0$$

If the X_n converge in distribution to U then U is also supported by G. Note that

$$\mathbb{P}(U \notin G) = \lim_n \mathbb{P}(X_n \notin G) = 0$$

This contradicts the hypothesis that U is uniformly distributed. Suppose now Y_1 is supported by a coset of G. This is the closed set $RG = \{RP | P \in G\}$ where $R \in SO(3)$. The treatment of this case is similar but more complicated. We refer to [21].

We now turn to the *if* part. It follows from Proposition 12 that for all $l > 0$ we have $\lambda(l) \subset D$ where $D = \{z \in \mathbb{C} | |z| < 1\}$. In other words, for all $l > 0$ all the eigenvalues of $\phi(l)$ are < 1 in absolute value. It follows by (2.62) and by a classical application of the Jordan decomposition that $\lim_n \phi_n(l) = 0$. Using (ii) of Proposition 2 and (i) of Proposition 4 we can conclude that $X_n \xrightarrow{d} U$. This completes the proof.▲

Proposition 12 *If Y_1 is not supported by any closed proper subgroup of $SO(3)$ or by any coset of such a subgroup then for all $l > 0$ we have $\lambda(l) \subset D$.*

Proof: We here copy the proof given in [21]. Let $l > 0$ and note $U \equiv U^l(Y_1)$. The random matrix U is unitary. Let $\lambda \in \lambda(l)$ and let f be a nonzero $d_l \times 1$ column matrix such that $\phi(l)f = \lambda f$. Remembering that $\phi(l) = \mathbb{E}[U]$, we have by Jensen's inequality

$$|\lambda|^2 f^\dagger f = f^\dagger \mathbb{E}[U^\dagger]\mathbb{E}[U]f \leq f^\dagger \mathbb{E}[U^\dagger U]f = f^\dagger f \tag{2.63}$$

It follows that $|\lambda| \leq 1$. In order to prove the proposition we will show that $|\lambda| \neq 1$. We proceed by contradiction. If $|\lambda| = 1$, the inequality in (2.63) is replaced by an equality. This is only possible if Y_1 is supported by the set $G \subset SO(3)$ defined as follows. G is the set of $P \in SO(3)$ such that $U^l(P)f = \lambda f$. Suppose $\lambda = 1$. In this case, it can be checked immediately that G is a closed proper subgroup of $SO(3)$. This is a contradiction.

Suppose $\lambda = e^{\mathrm{i}a}$ where $0 < a < 2\pi$. Let $G_0 = \{P \in SO(3) | U^l(P)f = f\}$. We have seen that G_0 is a closed proper subgroup of $SO(3)$. It is clear that for all $R \in G$ we have $G = RG_0$. It follows that G is a coset of the closed proper subgroup G. This is a contradiction. Since our reasoning is independent of l and λ, the proof is complete.▲

Theorem 2 gives a sufficient and necessary condition for the convergence $X_n \xrightarrow{d} U$. This condition excludes the case where Y_1 is supported by a closed proper subgroup of $SO(3)$ or by a coset of such a subgroup. This can be interpreted as follows. In order for the X_n to converge in distribution to U, the values of Y_1 must be sufficient to generate the whole group $SO(3)$ by multiplication. When this is the case, then independently of the precise distribution of Y_1 the products X_n converge in distribution to U. The convergence of the X_n leads to a uniformization of the distribution of Y_1 and has no memory regarding this distribution.

2.2 Application to polarization statistics

The current section is based on our paper [63]. Applying the mathematical development of Subsection 2.1.3, Subsection 2.2.3 introduces a new formalism for the polarization statistics of lightwaves. This formalism is seen to generalize the classical Stokes formalism of polarization to higher order statistics. It leads to a new definition of the notion of degree of polarization. In Subsection 2.2.4, this is used along with the results of Subsection 2.1.4 in order to characterize the physical problem of depolarization.

The state of polarization of a lightwave reflects the joint statistics of the components of the fluctuating transverse electric field. Classically, this is assumed to be contained in the so-called coherence matrix of the lightwave which only contains second order statistics. This assumption relies on a hypothesis of Gaussian fluctuations of the transverse electric field and on the close relation between second order statistics and the physical observables of the wavefield. This approach is equivalent to the classical Stokes formalism which we recall in Subsection 2.2.1. In many situations of interest, the hypothesis of Gaussian fluctuations is broken due to a particular type of interaction between the lightwave and a certain physical medium. While second order statistics contain information as to the observables of the wavefield, they fail to account for the properties of the underlying physical medium. This is illustrated by the experimental setting proposed by Ellis and Dogariu [12,13], which we present in Subsection 2.2.2.

Many approaches have been proposed for including higher order statistics of the transverse electric field into a consistent formalism. These are often driven by an effort to redefine the notion of degree of polarization in a way appropriate to specific applications. In [60, 62], Réfrégier gives a measure of the degree of polarization based on the Kullback relative entropy. Ellis and Dogariu [12] propose to differentiate unpolarized lightwaves using fourth order statistics of the electric field. Mathematically, the definition of degree of polarization which we will obtain in Subsection 2.2.3 is very close to the one derived by Luis in [42]. We will compare our definition to [12] and [42] in Subsection 2.2.3. Let us here note that all these approaches emphasize two aspects. That it is important to include higher order statistics and that this should be done in accordance with the symmetry properties of the wavefield. This is realized in Subsection 2.2.3 using the development of Subsection 2.1.3. After a slight modification, it will be possible to interpret this developmet as giving a decomposition of the higher order statistics of the transverse electric field along the irreducible representations of $SO(3)$.

When considering the interaction of a lightwave with a physical medium, the formalism of Subsection 2.2.3 allows a perspective that is closer to statistical signal processing. The medium can be regarded as a system determined by input/output states of polarization. In general, the response of this system is random and depends on the physical parameters of the medium. Extraction of such parameters from observed reflected or transmitted lightwaves reduces to parametric estimation of the response of the system from sample input/output. This is briefly illustrated in Subsection 2.2.4.

2.2.1 The classical Stokes formalism

We here recall the main features of the classical Stokes formalism of polarization. A thorough presentation of this and other classical polarization descriptions can be found in [7]. We prefer to follow [31] which has a greater emphasis on statistical effects. The Stokes formalism for the state of polarization of a lightwave requires knowledge of the direction of propagation. Let E_1 and E_2 be the complex analytic signals representing the components of the electric field in a plane perpendicular to the direction of propagation. If the electric field is fluctuating, E_1 and E_2 are considered random. The coherence matrix of the lightwave contains the joint second order statistics of the couple (E_1, E_2). Consider the following matrices

$$J = \begin{bmatrix} E_1E_1^* & E_1E_2^* \\ E_2E_1^* & E_2E_2^* \end{bmatrix} \qquad \langle J \rangle = \begin{bmatrix} \langle E_1E_1^* \rangle & \langle E_1E_2^* \rangle \\ \langle E_2E_1^* \rangle & \langle E_2E_2^* \rangle \end{bmatrix} \tag{2.64}$$

The construction of the matrix J can be understood in terms of spinor algebra [6]. We are not interested in this aspect here. For the matrix $\langle J \rangle$, the angular brackets denote averaging over the fluctuations of the electric field. This matrix is known as the coherence matrix of the lightwave. The linear invariants of $\langle J \rangle$ give the observables of the wavefield. In the following formula $\mathcal{I}$ is the average light intensity and P is the degree of polarization.

$$\mathcal{I} = \operatorname{tr} \langle J \rangle \qquad P = \left[1 - 4 \det \langle J \rangle / [\operatorname{tr} \langle J \rangle]^2\right]^{\frac{1}{2}} \tag{2.65}$$

The Stokes formalism can be obtained by decomposing J along a basis of Pauli matrices. These are here chosen to be the following

$$\sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \sigma_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \sigma_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \sigma_3 = \begin{bmatrix} 0 & \mathtt{i} \\ -\mathtt{i} & 0 \end{bmatrix}$$

Since J and $\langle J \rangle$ are Hermitian, there exist unique real J_i where $0 \leq i \leq 3$ such that (2.66) holds. The J_i are given in (2.67).

$$J = J_0\sigma_0 + J_1\sigma_1 + J_2\sigma_2 + J_3\sigma_3 \qquad \langle J \rangle = \langle J_0 \rangle\sigma_0 + \langle J_1 \rangle\sigma_1 + \langle J_2 \rangle\sigma_2 + \langle J_3 \rangle\sigma_3 \tag{2.66}$$

$$J_0 = E_1E_1^* + E_2E_2^* \quad J_1 = E_1E_1^* - E_2E_2^* \qquad J_2 = E_1E_2^* + E_2E_1^* \quad J_3 = \mathtt{i}(E_2E_1^* - E_1E_2^*) \tag{2.67}$$

In particular, it follows from formula (2.65) that $\langle J_0 \rangle = \mathcal{I}$. In describing the state of polarization, it will be convenient to discard the role of $\mathcal{I}$. One considers the so-called Stokes vector S and average Stokes vector $\langle S \rangle$. These are given by (2.68). Note that it is a abuse of notation to write $\langle S \rangle$ since this is actually not the average of S over fluctuations.

$$S = [S_1, S_2, S_3] = (1/J_0)[J_1, J_2, J_3] \qquad \langle S \rangle = [\langle S_1 \rangle, \langle S_2 \rangle, \langle S_3 \rangle] = (1/\langle J_0 \rangle)[\langle J_1 \rangle, \langle J_2 \rangle, \langle J_3 \rangle] \tag{2.68}$$

These "vectors" have simply been defined as triplets of parameters related to E_1 and E_2. Their laws of transformation (2.71) can be seen to give further justification to this terminology. The matrix $\langle J \rangle$ determines the second order statistics of the electric field. In case the fluctuations of this field are Gaussian, they are completely determined by $\langle J \rangle$. The average Stokes vector $\langle S \rangle$ is a normalized version of $\langle J \rangle$. In particular, it determines the eigendirections of $\langle J \rangle$ which are related to the ellipse of polarization. When the fluctuations of the electric field are Gaussian, $\langle S \rangle$ is taken to determine the state of polarization. The degree of polarization P defined in (2.65) is simply given in terms of $\langle S \rangle$ as

$$P = \left[\langle S_1 \rangle^2 + \langle S_2 \rangle^2 + \langle S_3 \rangle^2\right]^{\frac{1}{2}} \tag{2.69}$$

Clearly, in the case of non Gaussian fluctuations, $\langle J \rangle$ and $\langle S \rangle$ are not sufficient. Subsection 2.2.3 is precisely concerned with this problem. It introduces the higher order average Stokes vectors $\langle S^l \rangle$ as in (2.75). These contain higher order statistics and transform according to (2.76) which immediately generalizes (2.71).

In terms of its state of polarization, the interaction of a lightwave with a physical medium can be characterized by the Mueller matrix of the medium. This formulation does not consider the propagation of the wavefield in the medium. Rather, it is assumed that incident and reflected/transmitted states of polarization are related by a linear transformation characterizing the medium. If we refer to these as input and output states of polarization given then the effect of the medium is given by a complex 2×2 matrix T as follows

$$J_{out} = T J_{in} T^{\dagger} \tag{2.70}$$

here the subscripts *in* and *out* refer to the input and output states of polarization and T is known as the Jones matrix. In order to write (2.70) as a transformation of the Stokes vector, we need to impose a restriction on the matrix T. Namely, that T must be unitary. In this case, T preserves the trace of J_{in}. Physically, this means that the average intensity is preserved, $\mathcal{I}_{in} = \mathcal{I}_{out}$. Thus a unitary T represents a medium with no polarization related losses. In optics, such a medium is referred to as a birefringent medium. Under this restriction, relation (2.70) is equivalent to

$$S_{out} = M S_{in} \qquad \langle S_{out} \rangle = \langle M \rangle \langle S_{in} \rangle \tag{2.71}$$

The equivalence between (2.70) and (2.71) is proved in [31]. Here, M is a 3×3 matrix known as the Mueller matrix of the medium. In fact, M arises from T by a tensor operation and it follows from the fact that T is unitary that $M \in SO(3)$. In particular, the terminology Stokes "vector" seems to be justified by the fact that in (2.71) the input S_{in} transforms into the output S_{out} by a proper rotation. In general, the matrix M is also fluctuating. This is indeed the case when lightwaves are transmitted by the atmosphere or by certain particle suspensions. The second relation in (2.71) is based on the hypothesis that the fluctuations of the medium are uncoupled from those of the wavefield. This justifies the following replacement –see discussion of relation (2.73) in Subsection 2.2.3.

$$\langle S_{out} \rangle = \langle M S_{in} \rangle = \langle M \rangle \langle S_{in} \rangle$$

In Subsection 2.2.3 the formalism of the current subsection will be generalized to higher order statistics.

2.2.2 Experimental characterization of unpolarized light

Fluctuations of the transverse electric field of a lightwave arise in many situations. Of special importance to imaging applications are speckle phenomena. These arise from the coherent illumination of a scattering medium or rough surface. Speckles appear as random intensity patterns due to interference. Each speckle is considered as a realization of the fluctuating electric field. To each such realization corresponds a realization of a Stokes vector defined as in (2.68). A statistical analysis of the state of polarization of the lightwave requires observation of individual speckles. In principle, a sufficient number of such observations gives access to the distribution of an underlying random Stokes vector S. In fact, Ellis and Dogariu have shown this to hold using the experimental setting which we describe in the current subsection. We refer to their paper [13].

In order to visualize the realizations of the random Stokes vector S, we consider the Poincaré sphere [7]. Returning to equations (2.67) and (2.68), it is easy to check that individual realizations

of S are restricted to the surface of a unit sphere. On the other hand, the average Stokes vector $\langle S\rangle$ takes its values inside this unit sphere,

$$S_1^2 + S_2^2 + S_3^2 = 1 \qquad 0 \leq P = \left[\langle S_1\rangle^2 + \langle S_2\rangle^2 + \langle S_3\rangle^2\right]^{\frac{1}{2}} \leq 1 \tag{2.72}$$

As explained in Subsection 2.2.1, S is a "vector" in a parameter space. In the same way, one speaks of the unit sphere in this parameter space as the Poincaré sphere. Note in (2.72) that the degree of polarization satisfies $0 \leq P \leq 1$. When $P = 1$, we have that S is constant, in the sense that $S = \langle S\rangle$. This is referred to as a pure state of polarization. When $P < 1$, one speaks of a partially polarized state. In particular, $P = 0$ is said to correspond to an unpolarized state. Thus, the surface of the Poincaré sphere consists of pure states of polarization, while partially polarized states lie inside the Poincaré sphere.

The average Stokes vector $\langle S\rangle$ corresponds to second order statistics of the fluctuating electric field. As mentioned before, $\langle S\rangle$ is closely related to physical observables of the wavefield. These are sufficient in the case of Gaussian fluctuations but provide no information as to the properties of the physical system which has led to the formation of our individual observations. In order to obtain such information, it is necessary to start from observations of individual speckles and obtain higher order statistics or even the whole distribution of the random Stokes vector S. Clearly, these correspond to higher order statistics of the fluctuating electric field. Non Gaussian fluctuations have been observed in several applications, such as surface roughness measurements [19, 52, 59], detection of particles on a surface [53], scattering from small particles and particle shape determination [26]. In all these cases, pertinent information is to be found in the higher order statistics of S.

The experimental setting proposed in [13] gives access to the higher order statistics and the distribution of S and illustrates their pertinence to the properties of an underlying physical medium. The aim is to characterize experimentally the statistics of unpolarized lightwaves. The situation is considered which consists in coherent illumination of a thin layer of small –with respect to the incident wavelength– independent scattering particles. Light is incident on this layer within a solid angle of 2π and transmitted speckle patterns are observed along the optical axis of the layer. Various constitutions for the layer are used, defined in terms of particle shape and orientation. Thus, the layer is either constituted of spherical particles or of cylindrical particles distributed uniformly in a plane or a certain family of planes. In order to prepare observations of individual speckles a Stokes polarimeter is used. This is made up of a rotating quarter-wave plate, a fixed polarizer and a high-resolution CCD camera. A realization of the random Stokes vector S can be obtained for each pixel within a speckle. These realizations belong to the surface of the Poincaré sphere. They can be used to construct empirically the distribution of S or to estimate its higher order statistics.

Resulting states of polarization are close to unpolarized and it is possible to consider that we always have $P = 0$. However, depending on the constitution of the scattering layer, three types of states of polarization were observed. These are differentiated by their higher order statistics. In particular, they have different values of the correlations of the elements of S and for each type the distribution of S has different symmetry properties. These were labelled type I, II and III unpolarized light [12, 13].

- The distribution of S on the Poincaré sphere is invariant by any rotation as well as symmetric about any plane. Thus S is uniformly distributed on the Poincaré sphere. This is called type I unpolarized light.

- The distribution of S is invariant by any rotation preserving S_3 and symmetric about the plane of S_1 and S_2. This is called type II unpolarized light.

- The state of polarization is not of type II but the distribution of S is symmetric about the plane of S_1 and S_2. This is called type III unpolarized light.

This classification cannot be given a physical justification within our limited presentation of the Stokes formalism in Subsection 2.2.1. In fact, the symmetries required for each type correspond to a state of polarization which is invariant under the effect of certain optical devices or physical transformation. For example, symmetry about the plane of S_1 and S_2 implies that the state of polarization is invariant by inversion of the direction of propagation.

A theoretical study of these three types is given in [12]. We will consider them within the generalized Stokes formalism of Subsection 2.2.3. From the experimental results, it is possible to make the following conclusions. It is clear that the average Stokes vector $\langle S \rangle$ fails to give any information on the scattering layer. Such information appears to be contained in the correlations of the elements of the random Stokes vector S. Moreover, the definition of unpolarized state seems to be incorrect. Indeed, unpolarized states are obtained which in fact correspond to different distributions of S. In Subsection 2.2.3 we remove this ambiguity by proposing a stronger definition of the degree of polarization.

2.2.3 A generalized Stokes Formalism

Motivated by the discussion and experimental results presented in the last Subsection 2.2.2, the current subsection uses the mathematical development of Subsection 2.1.3 to generalize the classical Stokes formalism given in Subsection 2.2.1 to higher order statistics. The aim is to preserve the general form of the classical Stokes formalism while avoiding its restriction to second order statistics. By decomposing the higher order statistics of the random Stokes vector along the irreducible representations of $SO(3)$ a new formalism will be obtained which is based on definition (2.75) and relation (2.76) below. These are similar to (2.68) and (2.71) of Subsection 2.2.1. However, they fully include higher order statistics which provide information as to the properties of the underlying physical medium. As an application of this new formalism, a definition of the degree of polarization will be proposed. This definition will be used in characterizing the three types of unpolarized light encountered in the experimental setting of Subsection 2.2.2. It will also be compared to other definitions of the degree of polarization proposed in the literature with the aim of including higher order statistics.

Consider a standard situation consisting in the interaction of a lightwave with a physical medium. The incident and reflected/transmitted states of polarization correspond to random Stokes vectors S_{in} and S_{out} related by the Mueller matrix M of the medium as in (2.71). In general, the matrix M is fluctuating and takes its values in $SO(3)$. For convenience, relation (2.71) is copied here

$$S_{out} = MS_{in} \tag{2.73}$$

The analogy between relation (2.73) and (2.49) of Subsection 2.1.3 is clear. In order to apply the results of Subsection 2.1.3 to (2.73) the following understanding is needed[1]. This was until now kept implicit in the current section. We understand that fluctuating quantities such as S_{in}, S_{out} and M are equivalent to random variables on some fixed probability space. The operation of averaging over fluctuations, denoted for example by $\langle M \rangle$ in (2.71), is considered to be equivalent to mathematical expectation on this probability space. Moreover, S_{in} and M are treated as independent random variables. With this understanding, Proposition 9 of Subsection 2.1.3 will be applied to (2.73).

In order to apply Proposition 9 note that by identity (2.55) it follows from (2.73) that

$$Y^{l^*}(S_{out}) = U^l(M)Y^{l^*}(S_{in})$$

[1]While rigorously justifying such a convention is no simple matter, note that it corresponds to a usual mode of operation in statistical physics. See for instance the textbook [37].

for all $l \geq 0$. Using the independence of S_{in} and M, it is possible to write as in Proposition 9

$$\langle Y^{l^*}(S_{out})\rangle = \langle U^l(M)\rangle \langle Y^{l^*}(S_{in})\rangle \tag{2.74}$$

Our formalism will result from a modification of the averaged relation (2.74) leading to an equivalent real-valued relation. For $l \geq 0$, the $d_l \times 1$ column matrices Y^l are defined by (2.13) and (2.23). In the above relation (2.74), the elements of $Y^l(S_{in})$ and $Y^l(S_{out})$ are polynomials of degree l in the elements of S_{in} and S_{out}. While these polynomials are complex-valued, they can be replaced by real-valued polynomials using (2.15) of Paragraph 2.1.1.2. Let S denote either one of the random Stokes vectors S_{in} and S_{out}. For $l \geq 0$ define S^l as the $d_l \times 1$ column matrices

$$S^l = [S^l_{-l}, \ldots, S^l_l]$$

where for $-l \leq m \leq l$ the elements S^l_m of S^l are given by

$$S^l_m = \begin{cases} \frac{-1}{\sqrt{2}}\left((-1)^m Y^l_{-m}(S) + Y^l_m(S)\right) & \text{for } m > 0 \\ Y^l_0(S) & \text{for } m = 0 \\ \frac{\mathrm{i}}{\sqrt{2}}\left((-1)^{-m} Y^l_m(S) - Y^l_{-m}(S)\right) & \text{for } m < 0 \end{cases} \tag{2.75}$$

For all $l \geq 0$, it follows from (2.15) that S^l is real-valued. Moreover, it can be checked from (2.75) that S^l and $Y^l(S)$ are related by a unitary linear transformation. It follows from (2.74) that for all $l \geq 0$ there exists a real orthogonal $d_l \times d_l$ matrix $\langle M^l\rangle$ such that

$$\langle S^l_{out}\rangle = \langle M^l\rangle \langle S^l_{in}\rangle \tag{2.76}$$

where $\langle M^l\rangle$ is related to $\langle U^l(M)\rangle$ by a similarity transformation given by (2.75).

Definition (2.75) and relation (2.76) are the main equations of our formalism. For $l \geq 0$, the vector $\langle S^l\rangle$ will be called the average Stokes vector of order l. The elements of S^l are real-valued polynomials of order l in the elements of S. Thus, $\langle S^l\rangle$ gives the statistics of order l of S. These correspond to statistics of order $2l$ of the fluctuating electric field. The transformation given by (2.76) is of the same form as (2.71). However, it gives the laws of transformation of the statistics of order l of S for all $l \geq 0$.

Let us now apply (2.75) and (2.76) to the definition of the degree of polarization and to the experimental results of Subsection 2.2.2. By analogy with (2.69) consider for $l \geq 0$ the following quantity P^l

$$P^l = \left[\sum_{m=-l}^{l} \langle S^l_m\rangle^2\right]^{\frac{1}{2}} \tag{2.77}$$

This generalizes the degree of polarization to statistics of order l of the random Stokes vector S. As in (2.72) we have for all $l \geq 0$ that $0 \leq P^l \leq 1$ and $P^l = 1$ implies $S^l = \langle S^l\rangle$. This results from the following identity which holds for $l \geq 0$

$$\sum_{m=-l}^{l} (S^l_m)^2 = \sum_{m=-l}^{l} |Y^l_m(S)|^2 = 1 \tag{2.78}$$

Here, the first equality follows from (2.15) and (2.75). The second equality can be shown using (2.54) and the definition of spherical harmonics (2.13). For all $l \geq 0$, the quantity P^l will be called the degree of polarization of order l.

Let us give an example of average Stokes vectors and degree of polarization of order $l = 1$ and $l = 2$. This is based on a direct calculation from (2.13) and (2.75). For $l = 1$ it follows that

$\langle S^1\rangle = [-\langle S_2\rangle, \langle S_3\rangle, \langle S_1\rangle]$ and $P^1 = P$. In other words, the classical Stokes formalism of Subsection 2.2.1 is recovered using $\langle S^1\rangle$. The average Stokes vector of order $l = 2$ is given by

$$\langle S^2\rangle = [-\sqrt{\frac{3}{4}}\langle 2S_1S_2\rangle, -\sqrt{3}\langle S_2S_3\rangle, \langle\frac{3}{2}S_3^2 - \frac{1}{2}\rangle, \sqrt{3}\langle S_1S_3\rangle, \sqrt{\frac{3}{4}}\langle S_1^2 - S_2^2\rangle] \tag{2.79}$$

Clearly, $\langle S^2\rangle$ completely determines the second order statistics of S. In [12] it was suggested that types I, II and III of unpolarized light could be differentiated by the correlations of the elements of S. Consider the three following distributions for S. (**i**) S is uniformly distributed on the Poincaré sphere. This corresponds to type I unpolarized light. (**ii**) S takes the values $[0, 0, \pm 1]$ with equal probability $1/2$. This corresponds to type II unpolarized light. (**ii**) S is uniformly distributed on the great circle of the Poincaré sphere belonging to the plane $S_1 = 0$. This corresponds to type III unpolarized light. In each of these three cases, $\langle S^1\rangle = [0, 0, 0]$ and $P^1 = 0$. They are however physically different, as can be seen from the average Stokes vector of order $l = 2$. In case (**i**) the vector $\langle S^2\rangle$ is zero and $P^2 = 0$. In case (**ii**), $\langle S^2\rangle = [0, 0, 1, 0, 0]$ and $P^2 = 1$. In case (**iii**), $\langle S^2\rangle = [0, 0, 1/4, 0, -\sqrt{3}/4]$ and $P^2 = 1/2$.

From the above example, it is clear that the degree of polarization P used in the classical Stokes formalism fails to correctly characterize unpolarized light. Indeed, cases (**ii**) and (**iii**) verify $P = 0$. Thus, in terms of the classical Stokes formalism, they are to be considered identical and correspond to an unpolarized state. However, these cases display selection against certain states of polarization. Moreover, case (**ii**) only allows two values of S. These two values in fact correspond to so-called left and right circular polarization. In this sense, this case corresponds to a specific kind of pure state of polarization and cannot be termed unpolarized. In order to avoid such ambiguities, we propose the following stronger definition of the degree of polarization. For $l \geq 0$, we have introduced in (2.77) the state of polarization of order l, noted P^l. We will say that S determines a totally unpolarized state of polarization if for all $l > 0$ we have $P^l = 0$. If for some $l > 0$ we have $P^l = 0$, we will say that S determines a state of polarization which is unpolarized at order l. With this terminology, types I, II and III of unpolarized light can be characterized as follows.

- The distribution of S corresponds to type I unpolarized light *iff* S determines a totally unpolarized state of polarization. That is, *iff* $P^l = 0$ for all $l > 0$.
- The distribution of S corresponds to type II unpolarized light *iff* for all $l > 0$ and $m \neq 0$ we have $\langle S_m^l\rangle = 0$ and for all $l > 0$ such that l is odd we have $P^l = 0$.
- The distribution of S corresponds to type III unpolarized light *iff* $P^1 = P = 0$ and for all $l > 0$ and $-l \leq m \leq l$ we have $\langle S_m^l\rangle = 0$ whenever $l + m$ is odd.

Our characterization of type I unpolarized light follows from Proposition 10. For type II and III unpolarized light, we have used Proposition 11 and the fact that for all $l > 0$ and $-l \leq m \leq l$

$$Y_m^l(S') = (-1)^{l+m}Y_m^l(S)$$

where $S' = [S_1, S_2, -S_3]$ is the image of S by symmetry about the plane of S_1 and S_2.

Other definitions of the degree of polarization have been proposed in the literature with the aim of including higher order statistics. In [42], the degree of polarization is considered in the context of quantum optics. Mathematically, the definition thus obtained is quite similar to ours. Under the assumption that

$$L = \sum_{l>0}(P^l)^2 < \infty$$

the following definition of the degree of polarization P is introduced in [42]

$$0 \leq P = \frac{L}{1+L} \leq 1 \tag{2.80}$$

This is consistent with our terminology for a totally unpolarized state of polarization. Indeed, $P = 0$ in (2.80) *iff* we have in (2.77) that $P^l = 0$ for all $l > 0$. In [60, 62], a different approach is taken to the definition of the degree of polarization. A definition is proposed which is based on the Kullback relative entropy between the distribution of the wavefield and the distribution of a completely isotropic wavefield with the same intensity distribution. Unlike our definition in (2.77) and that of [42] in (2.80), this definition takes into account the intensity distribution of the wavefield and not only the Stokes vector S.

2.2.4 Depolarization

This subsection uses the results of Subsection 2.1.4 to address the physical problem of depolarization. Depolarization will appear as a physical obstacle to the application of relations (2.76) of the last Subsection 2.2.3.

Consider the situation described by relation (2.73). From the point of view of statistical signal processing, this relation describes a random linear system. The response of this system is given by the fluctuating Mueller matrix M. Suppose we seek to infer the physical parameters determining M from observed input/output states of polarization. Information as to such physical parameters is contained in the higher order statistics of S_{out}. Relations (2.76) give the higher order statistics of S_{out} in terms of the matrices $\langle M^l \rangle$. This indicates that our goal should be parametric estimation of the matrices $\langle M^l \rangle$ from sample S_{in} and S_{out}.

This goal seems practically realisable. In many situations, S_{in} is known. The experimental setting of Subsection 2.2.2 gives access to the higher order statistics of S_{out} when the output state of polarization is observed. In the current subsection we will see how depolarization can lead to a breakdown of relations (2.76) so that estimation of the matrices $\langle M^l \rangle$ is not exploitable.

Depolarization appears as a loss of information due to interaction between S_{in} and the physical medium. In particular, the distribution of S_{out} on the Poincaré sphere is made closer to a uniform distribution. This can be quantified in terms of the definition of degree of polarization proposed in the last Subsection 2.2.3. Rather than attempt to do this in general, we will be interested in a specific kind of situation. Namely, we assume that the Mueller matrix M can be written as a product

$$M = M_1 \dots M_n \tag{2.81}$$

where $n \geq 1$ and $M_1, \dots, M_n$ are Mueller matrices with values in $SO(3)$. Such a decomposition holds for many important applications. We will shortly consider a precise example from fiber optics. The decomposition (2.81) corresponds to the physical medium being modelled as a succession of uncoupled identical layers. The matrices $M_1, \dots, M_n$ are then treated as *i.i.d.* random variables. This recovers the context of Subsection 2.1.4. For $l > 0$, note P^l_{in} and P^l_{out} the degree of polarization of order l obtained as in (2.77) for S_{in} and S_{out}. By a reasoning similar to the one of inequality (2.63) in the proof of Proposition 12, we have the following general estimation for P^l_{out}

$$P^l_{out} \leq (\lambda_l)^n P^l_{in} \leq (\lambda_l)^n \tag{2.82}$$

where $\lambda_l \leq 1$ for $l > 0$. If the common distribution of the Mueller matrices $M_1, \dots, M_n$ satisfies the condition of Theorem 2 then $\lambda_l < 1$ for $l > 0$. In the limit of large n we have $P^l_{out} = 0$ for $l > 0$.

The output state of polarization approaches a totally unpolarized state independently of the input state. Thus, the output state contains no information as to the input state. It is important to observe here that depolarization only takes place when the Mueller matrix M is fluctuating. Indeed, if M is constant the condition of Theorem 2 is not satisfied.

We now consider the example of optical fibers with polarization mode dispersion. We refer to [20] for a detailed introduction to this topic and to [73] for this particular example. Polarization mode dispersion is a linear propagation effect in optical fibres which leads to a random mixture between different polarizations. Consider a length Z of optical fiber and note the corresponding Mueller matrix M_Z. In our paper [63], it was shown that the corresponding matrices $\langle M_Z^l \rangle$ as in (2.76) are given by –compare to Subsection 2.3.3.

$$\langle M_Z^l \rangle = e^{-\frac{\mu^2}{2} l(l+1)Z} I_l \tag{2.83}$$

for $l \geq 0$, where I_l is the $d_l \times d_l$ identity matrix. Here $\mu \geq 0$ is a physical parameter characterizing the fiber. For this example, the decomposition (2.81) corresponds to a partition of the length Z into n successive intervals of equal length. This decomposition is discussed in [20]. From relations (2.76) and definition (2.77) we now have

$$P_{out}^l = e^{-\frac{\mu^2}{2} l(l+1)Z} P_{in}^l \tag{2.84}$$

In particular, if the input state of polarization is pure, we have for $P_{out} = P_{out}^1$

$$P_{out} = e^{-\mu^2 Z} \tag{2.85}$$

From formula (2.84) it is clear that the output state of polarization approaches a totally unpolarized state in the limit of large Z. As for (2.82), information as to the input state of polarization is lost. This is clearly an undesirable effect in optical fiber telecommunications. The importance of the physical parameter μ stems from the fact that it determines the rate of convergence $P_{out}^l \downarrow 0$ for $l > 0$. Equations (2.84) give the relation between this physical parameter and the higher order statistics of S_{out}. Let us use (2.85) to estimate μ from observations of S_{out}. Remember that P is given by (2.69). If samples $S_1, \ldots, S_N$ for some $N \geq 1$ are available of S_{out} then empirical estimates $\hat{P}$ and $\hat{\mu}$ of P and μ can be formed

$$\hat{P}^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} S_i^T S_j \qquad \hat{\mu}^2 = -\frac{1}{Z} \log(\hat{P}) \tag{2.86}$$

As an alternative to the empirical estimate $\hat{P}$, note that P can be estimated from intensity measurements [61]. Depolarization poses two problems in using the estimates (2.86). First, when the distribution of S_{out} is closer to a uniform distribution, a higher number N of samples becomes necessary to obtain a good estimated value of P. Second, depolarization implies that $P \downarrow 0$. In this limit, the logarithm in (2.86) is divergent and can lead to numerical problems.

2.3 Rotation Lévy processes

Rotation Lévy processes will be defined as rotation processes with independent stationary increments and which are stochastically continuous. This section studies the two essential kinds of such processes, rotation compound Poisson processes and rotation Brownian motion.

In Subsection 2.3.1, the definition of rotation Lévy processes is given and the resulting expression of their characteristic functions is recalled. Subsection 2.3.2 is devoted to the symmetry properties and asymptotic behavior in distribution of rotation compound Poisson processes. Subsection 2.3.3 carries out a similar objective in the case of rotation Brownian motion. Finally, Subsection 2.3.4 considers

the interlacing construction used to obtain general Lévy processes from rotation compound Poisson processes and Brownian motion. In particular, these two kinds of processes appear as the building blocks of rotation Lévy processes. The main application of the current section will be to the solution of the problem of decompounding in Subsection 2.4.2. In Section 2.4, this problem will be formulated in terms of rotation compound Poisson processes.

Rotation Lévy processes are a special case of Lévy processes in compact Lie groups. A general reference on Lévy processes with values in Lie groups is [40]. The paper [39] details the application of characteristic functions to the study of Lévy processes in compact Lie groups. Compound Poisson processes in Lie groups were introduced by Applebaum in [2]. In Subsection 2.3.3, our approach to rotation Brownian motion will be different from that of [39, 40]. Indeed, we propose to start from the definition of rotation Brownian motion as a multiplicative integral. This choice is discussed in Subsection 2.3.3.

2.3.1 Definition and characteristic functions

This subsection states Proposition 13. This proposition is a direct result of the definition of rotation Lévy processes. It gives a general expression of their characteristic functions. Note that Proposition 13 holds for a general rotation Lévy process, while propositions obtained in the rest of this section concern specific rotation Lévy processes.

As in Section 2.1, we assume a complete probability space $(\Omega, \mathcal{A}, \mathbb{P})$. By a rotation process Y we mean a family of rotation random variables Y_t indexed by $t \geq 0$. In short, we write $Y = (Y_t)_{t\geq 0}$. The following notation will be used. For $0 \leq s \leq t$, note $Y_{(s|t)} \equiv Y_s^T Y_t$. The rotation random variables $Y_{(s|t)}$ are called the increments of Y. This is in the sense that $Y_t = Y_s Y_{(s|t)}$. Lévy processes will be defined in terms of their increments. More precisely, Y will be called a rotation Lévy process if it verifies the following conditions

(L1) $Y_0 = I$ almost surely.

(L2) For $n \geq 1$ and $0 \leq t_1 \leq \ldots \leq t_n$ the rotation random variables $Y_{(0|t_1)}, \ldots, Y_{(t_{n-1}|t_n)}$ are independent.

(L3) For $0 \leq s \leq t$ we have $Y_{(s|t)} \stackrel{d}{=} Y_{t-s}$.

(L4) $Y_t \stackrel{d}{\to} I$ when $t \downarrow 0$.

Conditions (L2) and (L3) respectively state that the increments of Y are independent and stationary. It follows from conditions (L3) and (L4) that Y is stochastically continuous. Strictly speaking, conditions (L1) to (L4) define a *left* Lévy process. In particular, for $t \geq 0$ and $n \geq 1$ the following product is ordered from left to right and is a product of *i.i.d.* rotation random variables

$$Y_t = Y_{(0|t_1)} \ldots Y_{(t_{n-1}|t_n)}$$

where for $1 \leq m \leq n$ we have $t_m = (m/n)t$. Right Lévy processes are defined by putting $Y_{(s|t)} \equiv Y_t Y_s^T$ so that for $0 \leq s \leq t$ we have $Y_t = Y_{(s|t)} Y_s$. It can be seen easily that if Y is a left Lévy process then $Y^T = (Y_t^T)_{t\geq 0}$ is a right Lévy process. In the following we only consider left Lévy processes and the adjective *left* will be dropped.

We now state Proposition 13 and give its proof. Let Y be a rotation Lévy process. For $t \geq 0$ we will note $\phi_t \equiv \phi_{Y_t}$. Formula (2.87) gives the general expression of ϕ_t in terms of the matrices $\{A_l\}_{l\geq 0}$. The precise form of these matrices in the case of rotation compound Poisson processes and Brownian motion is obtained in Propositions 14 and 19 below. In Subsection 2.3.4 we will discuss the form of $\{A_l\}_{l\geq 0}$ for a general rotation Lévy process Y.

Proposition 13 *For all $l \geq 0$ there exists a $d_l \times d_l$ matrix A_l such that for $t \geq 0$*

$$\phi_t(l) = \exp(tA_l) \tag{2.87}$$

Proof: For all $0 \leq s \leq t$ we have $Y_t = Y_s Y_{(s|t)}$. It follows from condition (L2) that Y_s and $Y_{(s|t)}$ are independent. Moreover, by condition (L2) we have $Y_{(s|t)} \stackrel{d}{=} Y_{t-s}$. Applying (i) of Proposition 2 and (ii) of Proposition 3 it follows for $l \geq 0$

$$\phi_t(l) = \phi_s(l)\phi_{t-s}(l) \tag{2.88}$$

For $l \geq 0$, the matrix function $t \mapsto \phi_t(l)$ is continuous. Indeed, by (2.88) we have for $0 \leq s \leq t$

$$|\phi_t(l) - \phi_s(l)| \leq |\phi_s(l)||\phi_{t-s}(l) - I_l| \leq [2l+1]^{\frac{1}{2}} |\phi_{t-s}(l) - I_l|$$

where I_l is the $d_l \times d_l$ identity matrix. By condition (L4) and (ii) of Proposition 2 we have $\phi_{t-s}(l) \to I_l$ when $t - s \downarrow 0$. It follows for $t \geq 0$ that $\phi_s(l) \to \phi_t(l)$ when $s \uparrow t$. The case where $s \downarrow t$ can be treated similarly.

We now have that for $l \geq 0$ the matrix function $t \mapsto \phi_t(l)$ is continuous and satisfies (2.88). It follows that there exists a $d_l \times d_l$ matrix A_l such that for $t \geq 0$ the expression (2.87) holds for $\phi(l)_t$. This completes the proof.▲

2.3.2 Compound Poisson processes in $SO(3)$

After recalling the definition of rotation compound Poisson processes, the current subsection will consider their symmetry properties and asymptotic behavior in distribution. These are obtained in Propositions 15 and 16, respectively. Both can be seen as results of Proposition 14 which gives the characteristic function of a rotation compound Poisson process. Propositions 15 and 16 will be important to the solution of the problem of decompounding in Subsection 2.4.2.

Rotation compound Poisson processes naturally generalize real-valued compound Poisson processes. Let $N = (N_t)_{t \geq 0}$ be a Poisson process with parameter $\lambda > 0$. Suppose $(x_n)_{n \geq 1}$ are *i.i.d.* real-valued random variables and the family $(x_n)_{n \geq 1}$ is itself independent of N. Noting $x_0 = 0$, the following process $y = (y_t)_{t \geq 0}$ is said to be a real-valued compound Poisson process

$$y_t = \sum_{n=0}^{N_t} x_n \tag{2.89}$$

Rotation compound Poisson processes are defined by analogy to this formula. We continue with the process N. Let $(X_n)_{n \geq 1}$ be *i.i.d.* rotation random variables and suppose as before that the family $(X_n)_{n \geq 1}$ is independent of N. Noting $X_0 = I$, the following process $Y = (Y_t)_{t \geq 0}$ is said to be a rotation compound Poisson process

$$Y_t = \prod_{n=0}^{N_t} X_n \tag{2.90}$$

We understand that products are ordered from left to right, in accord with Subsection 2.3.1 which used the definition of left rotation Lévy processes. As in the classical case of the process y, the rotation process Y is a pure jump process. This means that its paths consist of sequences of jumps. Note that, by the definition of N, the paths of Y are *càdlàg* –right continuous with left limits.

Proposition 14 establishes that our definition of rotation compound Poisson processes is correct. Formula (2.91) gives the expression (2.87) of the characteristic function of a rotation compound Poisson process. The notation $\phi_t \equiv \phi_{Y_t}$ is used as for Proposition 13.

Proposition 14 *Let the process $Y = (Y_t)_{t\geq 0}$ be defined as above. Y is a rotation Lévy process. Moreover, for $t \geq 0$ we have*

$$\phi_t(l) = \exp[\lambda t(\phi(l) - I_l)] \tag{2.91}$$

for $l \geq 0$, where $\phi \equiv \phi_{X_1}$.

Proof: In order to see that Y is a Lévy process, we check conditions (L1) to (L4). Condition (L1) follows immediately from the fact that $N_0 = 0$. For condition (L2), we will prove that for all $0 \leq s \leq t$ we have that Y_s and $Y_{(s|t)}$ are independent. A complete proof of (L2) can be carried out similarly.

Let $0 \leq s \leq t$ and note that

$$Y_{(s|t)} = \prod_{r=1}^{N_t - N_s} X_{N_s + r} \tag{2.92}$$

with the understanding that the product is equal to I if $N_t - N_s = 0$. Suppose $h_1, h_2 : SO(3) \to \mathbb{C}$ are continuous functions. Note $Y_1 \equiv Y_s$ and $Y_2 \equiv Y_{(s|t)}$. It is clear that

$$\mathbb{E}[h_1(Y_1)h_2(Y_2)] = \sum_{n\geq 0}\sum_{m\geq 0} \mathbb{E}\left[h_1(Y_1)h_2(Y_2)\mathbf{1}(N_s = n)\mathbf{1}(N_t - N_s = m)\right] \tag{2.93}$$

where the notation $\mathbf{1}(A)$ stands for the indicator of an event $A \in \mathcal{A}$. For all $n, m \geq 1$, it follows from the independence of $(X_n)_{n\geq 1}$ and N and from the fact that N has independent increments that

$$\mathbb{E}\left[h_1(Y_1)h_2(Y_2)\mathbf{1}(N_s = n)\mathbf{1}(N_t - N_s = m)\right] = \\ \mathbb{E}[h_1(X_1 \dots X_n)]\mathbb{E}[h_2(X_{n+1} \dots X_{n+m})]\mathbb{P}(N_s = n)\mathbb{P}(N_t - N_s = m)$$

where we have used (2.92). Since the $(X_n)_{n\geq 1}$ are *i.i.d.*, the last equality can be written

$$\mathbb{E}\left[h_1(Y_1)h_2(Y_2)\mathbf{1}(N_s = n)\mathbf{1}(N_t - N_s = m)\right] = \\ \mathbb{E}[h_1(X_1 \dots X_n)]\mathbb{E}[h_2(X_1 \dots X_m)]\mathbb{P}(N_s = n)\mathbb{P}(N_t - N_s = m)$$

The case where $n = 0$ or $m = 0$ can be treated similarly. It is now possible to evaluate the sum (2.93). This decomposes into a product of two sums as follows

$$\mathbb{E}[h_1(Y_1)h_2(Y_2)] = \sum_{n\geq 0} \mathbb{P}(N_s = n)\mathbb{E}[h_1(X_0 \dots X_n)] \sum_{m\geq 0} \mathbb{P}(N_t - N_s = m)\mathbb{E}[h_1(X_0 \dots X_m)]$$

The first sum is clearly equal to $\mathbb{E}[h_1(Y_1)]$ and the second sum can be identified with $\mathbb{E}[h_2(Y_2)]$ using (2.92). It follows that

$$\mathbb{E}[h_1(Y_1)h_2(Y_2)] = \mathbb{E}[h_1(Y_1)]\mathbb{E}[h_2(Y_2)]$$

This shows that $Y_1 \equiv Y_s$ and $Y_2 \equiv Y_{(s|t)}$ are independent for all $0 \leq s \leq t$. Condition (L3) can also be proved using (2.92).

We now prove (2.91). For $t \geq 0$ and $l \geq 0$

$$\phi_t(l) = \mathbb{E}[U^l(Y_t)] = \sum_{n\geq 0} \mathbb{E}[U^l(X_0 \dots X_n)]\mathbb{P}(N_t = n)$$

where we have used the independence of $(X_n)_{n\geq 1}$ and N. Using the fact that the $(X_n)_{n\geq 1}$ are *i.i.d.* we have

$$\phi_t(l) = \sum_{n\geq 0} [\phi(l)]^n \mathbb{P}(N_t = n) = e^{-\lambda t} \sum_{n\geq 0} \frac{(\lambda t)^n}{n!}[\phi(l)]^n$$

and (2.91) follows from the Taylor series formula for the matrix exponential. To complete the proof, we must obtain condition (L3) for Y. From (2.91) we have for $l \geq 0$

$$\lim_{t\downarrow 0} \phi_t(l) = I_l$$

using (ii) of Proposition 2 it follows that $Y_t \xrightarrow{d} I$ when $t \downarrow 0$. This completes the proof.▲

We now summarize the symmetry properties of the random variables Y_t for $t \geq 0$. Note first that for all $t \geq 0$, Y_t does not have a probability density. Indeed, for all $t \geq 0$ we have $\mathbb{P}(Y_t = I) \geq \mathbb{P}(N_t = 0) = e^{-\lambda t}$. It follows that Y_t has an atom at I. We study Y_t for $t \geq 0$ using its characteristic function which was given in Proposition 14. Combining formula (2.91) and the results of Paragraph 2.1.2.2 we have the following proposition. It states that for all $t \geq 0$ the symmetry properties of Y_t are the same as those of the $(X_n)_{n \geq 1}$.

Proposition 15 *Let $t \geq 0$. If X_1 is inverse (respectively, conjugate, zonal) invariant then Y_t is inverse (respectively, conjugate, zonal) invariant.*

Proposition 16 gives the asymptotic behavior in distribution of Y. This is similar to the behavior of products of *i.i.d.* rotation random variables described in Theorem 2. Proposition 20 of Subsection 2.3.3 gives the same result for rotation Brownian motion. A general version of this proposition which holds for Lévy processes in compact Lie groups is given in [40]. We prove Proposition 16 using Propositions 12 of Subsection 2.1.4 and Proposition 14. Another proof using only Proposition 12 is possible.

Proposition 16 *Let U be a uniformly distributed rotation random variable. If X_1 is not supported by any closed proper subgroup of $SO(3)$ or by any coset of such a subgroup then $Y_t \xrightarrow{d} U$ when $t \uparrow \infty$.*

Proof: Using (ii) of Proposition 2, it is enough to prove that $\phi_t(l) \to 0$ for all $l > 0$ when $t \uparrow \infty$. This can be done using formula (2.91). By Proposition 12 we have for $l > 0$ that the eigenvalues of $\phi(l)$ are all less than unity in absolute value. It follows that the eigenvalues of $\phi(l) - I_l$ all have negative real parts. Replacing in (2.91) we have for $l > 0$ that $\phi_t(l) \to 0$ when $t \uparrow \infty$. This completes the proof.▲

2.3.3 Rotation Brownian motion

In the last subsection, rotation compound Poisson processes were introduced as pure jump rotation Lévy processes. This subsection studies the essential continuous path rotation Lévy process, rotation Brownian motion. The results obtained here are similar to those of the last subsection. Proposition 18 gives the symmetry properties of rotation Brownian motion. Proposition 20 obtains its asymptotic behavior in distribution. We will study rotation Brownian motion using its definition as a multiplicative integral. In Paragraph 2.3.3.1 we recall this definition and discuss that it does indeed lead to a continuous path rotation Lévy process. In Paragraph 2.3.3.2 we will apply the definition of Paragraph 2.3.3.1 to Propositions 18 and 20. Proposition 19 of Paragraph 2.3.3.2 gives the characteristic function of conjugate invariant rotation Brownian motion.

The first consideration of rotation Brownian motion is due to Perrin in 1928 [56]. Rotation Brownian motion was then approached indirectly, using the heat equation on the rotation group. Stochastic calculus makes possible a direct study of rotation Brownian motion, based on its stochastic differential equation. This is the so-called Euler-Langevin equation. A simplified version of the Euler Langevin equation was considered informally by Debye in 1913 [11]. Both [11, 56] consider rotation Brownian motion as a model for the rotational motion of molecules.

The Euler-Langevin equation is a linear stochastic differential equation driven by a $\mathbb{R}^3$-valued Brownian motion process B with coordinates $[B^1, B^2, B^3]$. In matrix form, it is expressed as the following Stratonovich stochastic differential equation [67]

$$dY_t = Y_t dJ_t \qquad Y_0 = I \tag{2.94}$$

Here $Y = (Y_t)_{t\geq 0}$ is the unknown process. The process J is an embedding of B in the space of antisymmetric 3×3 matrices. More precisely, for $t \geq 0$ we have $J_t = B_t^1 J_1 + B_t^2 J_2 + B_t^3 J_3$ where the matrices J_1, J_2, J_3 were given in (2.7) of Paragraph 2.1.1.1. Equation (2.94) is a matrix equation. It is equivalent to the following system of linear stochastic differential equations, verified for $1 \leq i, j \leq 3$

$$dY_t^{ij} = \sum_{l=1}^{3} \left[\sum_{k=1}^{3} Y^{ik} J_l^{kj} \right] dB_t^l \tag{2.95}$$

where superscripts have been used to denote matrix elements. We will shortly see that (2.94) defines a *left* rotation Brownian motion, in the sense explained in Subsection 2.3.1.

From (2.95), it is straightforward to derive the Itô stochastic differential equation equivalent to (2.94). This is given in terms of the shifted process $\tilde{J} = (\tilde{J}_t)_{t\geq 0}$ as follows

$$dY_t = Y_t \tilde{J}_t \qquad \tilde{J}_t = J_t + \frac{1}{2}\mathbb{E}[J_t^2] \tag{2.96}$$

2.3.3.1 Rotation Brownian motion as a multiplicative integral

The definition of rotation Brownian motion as a multiplicative integral was proposed by McKean in 1960 [47]. The use of multiplicative integrals has since evolved into a powerful and very general tool for studying processes with values in Lie groups and even Riemannian manifolds [4, 15]. It will offer us two advantages. First, multiplicative integrals support a simple and intuitive interpretation of rotation Brownian motion. Second, it allows a straightforward probabilistic treatment avoiding more complicated analytical methods.

Equation (2.94) is a linear stochastic differential equation driven by a Brownian motion B. We know from elementary stochastic analysis that it has a solution Y with almost surely continuous paths. This solution is moreover unique. Multiplicative integrals refer to the following technique for approximating Y.

For $n, m \geq 0$ let $t_m^n = m/2^n$. For $n \geq 0$ define the processes J^n and Y^n on each interval $t_m^n \leq t < t_{m+1}^n$ by

$$J_t^n = J_{t_m^n} + (J_{t_{m+1}^n} - J_{t_m^n}) \frac{t - t_m^n}{2^{-n}} \qquad Y_t^n = Y_{t_m^n} \exp\left(J_t^n - J_{t_m^n}^n\right) \tag{2.97}$$

with the initial condition $Y_0^n = I$.

The processes $(J^n)_{n\geq 0}$ perform a linear interpolation of J. For all $n, m \geq 0$ we have $J_{t_m^n}^n = J_{t_m^n}$ and J^n is affine in t on each interval $t_m^n \leq t < t_{m+1}^n$. Since J has almost surely continuous paths, the processes $(J^n)_{n\geq 0}$ converge to J almost surely uniformly. For $n \geq 0$ the process Y^n solves the ordinary differential equation

$$dY_t^n = Y_t^n dJ_t^n \qquad Y_0^n = I \tag{2.98}$$

which is equation (2.94) with J^n instead of J.

In order to approximate Y the sequence (2.98) of ordinary differential equations was introduced. Formally, these equations approximate (2.94). The following Proposition 17 states that the solutions Y^n of the equations (2.98) do in fact converge to the solution Y of (2.94). This leads to the desired definition of rotation Brownian motion. Our aim is here to interpret this proposition and to discuss that the resulting process Y is indeed a continuous path rotation Lévy process.

Proposition 17 *The processes $(Y^n)_{n\geq 0}$ converge locally uniformly in the square mean to the solution Y of (2.94). In other words, for all $T \geq 0$ we have*

$$\lim_n \mathbb{E}\left[\sup_{t\leq T} |Y_t^n - Y_t|^2\right] = 0 \tag{2.99}$$

Proposition 17 is here stated without proof. In Chapter 3, it will be considered in a stronger form which applies to processes in matrix groups in general.

The approximation of Proposition 17 defines Y as a multiplicative integral. This terminology can be justified as follows. Let $n \geq 0$ and $t \geq 0$ and define $N = \lceil 2^n t \rceil$ –the upper integer part. For $0 \leq m < N$ let $t_m = t_m^n$ and note $t_N = t$. Expression (2.97) can be rewritten

$$Y_t^n = \prod_{m=0}^{N-1} \exp\left(J_{t_{m+1}}^n - J_{t_m}^n\right) \tag{2.100}$$

where the product is ordered from left to right. From Proposition 17 we have the limit in the square mean $Y_t^n \to Y_t$. Thus, Proposition 17 gives Y_t as a limit of products which correspond to regular partitions of the interval $[0, t]$. The number of factors in the product (2.100) increases to infinity with n while each individual factor converges to I. This is in clear analogy with the usual definition of an integral in terms of Riemann sums. In the same way, the products (2.100) are referred to as Riemann products.

Returning to expression (2.97), we have that on each interval $t_m^n \leq t < t_{m+1}^n$ the evolution of the process Y^n is given by the exponential factor $\exp\left(J_t^n - J_{t_m^n}^n\right)$. The argument of the exponential is an antisymmetric matrix which is a linear function of t. It follows that these exponential factors describe uniform rotations connecting $Y_{t_m^n}^n$ to $Y_{t_{m+1}^n}^n$. While the processes J^n are piecewise affine, the processes Y^n are piecewise uniform rotations. Thus Proposition 17 approximates Y by a sequence of piecewise uniform rotation processes Y^n. As seen before, the number of such uniform rotations up to a time t increases to infinity while the time during which each uniform rotation is considered decreases to zero.

To end this paragraph, we would like to discuss that Proposition 17 does indeed define Y as a continuous path rotation Lévy process. The fact that Y has continuous paths is already a result of equation (2.94), since this equation is a stochastic differential equation driven by a Brownian motion. The same result can be obtained from Proposition 17. Indeed, for $n \geq 0$ the paths of the process Y^n are clearly continuous. According to Proposition 17, the paths of Y are uniform limits of continuous paths. It follows that they are themselves continuous.

Two things remain to be seen. First that Y is indeed a rotation process. Second, that Y is a Lévy process. That Y is a rotation process can be seen from equation (2.94). Indeed, this equation can be used to check that almost surely Y_t verifies conditions (2.1) for all $t \geq 0$. Consider the process $Y^T = (Y_t^T)_{t \geq 0}$. By transposing equation (2.94) we have

$$dY_t^T = -dJ_t Y_t^T \tag{2.101}$$

where we have used the fact that the values of the process J are antisymmetric matrices. Clearly, $Y_0 Y_0^T = I$, using integration by parts we have –remembering that equation (2.94) is a Stratonovich equation

$$d(YY^T)_t = Y_t dJ_t Y^T - Y_t dJ_t Y_t^T = 0$$

so that $Y_t Y_t^T = I$ for all $t \geq 0$. Note that integration by parts was here carried out for matrix processes. This can be justified by writing down this operation in terms of matrix elements, as in equation (2.95). Since $\det(Y_0) = 1$ and Y has almost surely continuous paths, it now follows that $\det(Y_t) = 1$ almost surely for all $t \geq 0$. Thus, conditions (2.1) are verified.

Proposition 17 leads quite directly to the fact that Y is a rotation Lévy process. For $t \geq 0$ consider the following sequences of rotation random variables

$$E_t^n = \prod_{t_m^n \leq t} \exp\left(J_{t_m^n}^n - J_{t_{m-1}^n}^n\right) \tag{2.102}$$

on each interval $t_m^n \leq t < t_{m+1}^n$ we have from (2.97) that

$$Y_t^n = E_t^n \exp\left(J_t^n - J_{t_m^n}^n\right)$$

Since $Y_t^n \to Y_t$ in the square mean, we also have $E_t^n \to Y_t$ in the square mean. Indeed, the exponential factor on the right hand side converges to I when $n \uparrow \infty$. Let $0 \leq s \leq t$ and define for $n \geq 0$ the rotation random variables E_1^n and E_2^n as follows

$$E_1^n = \prod_{t_m^n \leq s} \exp\left(J_{t_m^n}^n - J_{t_{m-1}^n}^n\right) \qquad E_2^n = \prod_{s < t_m^n \leq t} \exp\left(J_{t_m^n}^n - J_{t_{m-1}^n}^n\right)$$

By the approximation (2.102) we have the limits in the square mean $E_1^n \to Y_s$ and $E_2^n \to Y_{(s|t)}$. Note that for $n \geq 0$ the random variables E_1^n and E_2^n are independent. This follows from the fact that the factors in each of these products depend on the increments of the Brownian process J which are independent. Since independence is preserved by limits in the square mean, we have that Y_s and $Y_{(s|t)}$ are independent. In this way, condition (L2) can be checked for Y. A similar reasoning leads to condition (L3). Condition (L1) is already clear in (2.94) and condition (L4) is verified since Y has almost surely continuous paths.

We now have that Y is a left rotation Lévy process. In order to obtain a right rotation Lévy process from Y, it is enough to consider the process Y^T. We have seen that this is the solution of equation (2.101). It is also possible to define Y^T as a multiplicative integral. Indeed, such an approximation simply follows from Proposition 17 by transposing the processes Y^n.

2.3.3.2 The characteristic function of rotation Brownian motion

The current paragraph obtains the symmetry properties and the asymptotic behavior in distribution of rotation Brownian motion. These are given in Propositions 18 and 20, respectively. Proposition 18 will be obtained directly from the definition of rotation Brownian motion as a multiplicative integral, given in Proposition 17. Proposition 20 will result from Proposition 19 which gives the characteristic function of conjugate invariant rotation Brownian motion. In order to obtain Proposition 19, we will apply Proposition 18 and equation (2.96). Note that we have already met with expression (2.103) of Proposition 19 in formula (2.83) of Subsection 2.2.4.

The symmetry properties of the solution Y of the Euler-Langevin equation are determined by the covariance matrix C of the driving Brownian motion process B. We have the following proposition.

Proposition 18 *Let Y be the solution of (2.94). For $t \geq 0$ we have that Y_t is inverse invariant. Moreover, if the covariance matrix C of B is of the form $a^2 I$ where $a \in \mathbb{R}$ then for $t \geq 0$ we have that Y_t is conjugate invariant.*

Proof: We start with the proof of inverse invariance. Given $t \geq 0$, we will note $E_t^n \equiv E_n$ where for $n \geq 0$ the rotation random variable E_t^n is given in (2.102). For $t \geq 0$, we have the limits in the square mean $\lim_n E_n = Y_t$ and $\lim_n E_n^T = Y_t^T$. For $n \geq 0$ we will prove that $E_n \stackrel{d}{=} E_n^T$.

By (2.102) we have for $n \geq 0$

$$E_n = \prod_{t_m^n \leq t} \exp\left(J_{t_m^n}^n - J_{t_{m-1}^n}^n\right)$$

From the definition of J and (2.97), it can be seen that this is a product of *i.i.d.* rotation random variables. Note that for $n, m \geq 1$

$$\left[\exp\left(J_{t_m^n}^n - J_{t_{m-1}^n}^n\right)\right]^T = \exp\left[-\left(J_{t_m^n}^n - J_{t_{m-1}^n}^n\right)\right]$$

Since $J_t \stackrel{d}{=} -J_t$ for $t \geq 0$, it follows that each factor in the product for E_n is inverse invariant. By (ii) of Proposition 5, we now have that E_n is inverse invariant. In other words $E_n \stackrel{d}{=} E_n^T$. Since equality in distribution is preserved by limits in the square mean, we have $Y_t \stackrel{d}{=} Y_t^T$, so that Y_t is inverse invariant.

The proof of conjugate invariance follows the same steps. If $C = a^2 I$ with $a \in \mathbb{R}$ then for all $R \in SO(3)$ and $t \geq 0$ we have $RJ_tR^T \stackrel{d}{=} J_t$. Using the following identity for $n, m \geq 1$ and $R \in SO(3)$

$$R\left[\exp\left(J^n_{t^n_m} - J^n_{t^n_{m-1}}\right)\right]R^T = \exp\left[R\left(J^n_{t^n_m} - J^n_{t^n_{m-1}}\right)R^T\right]$$

it follows that each factor in the product for E_n is conjugate invariant. Applying (iii) of Proposition 6, we now have that E_n is conjugate invariant. That is, $RE_nR^T \stackrel{d}{=} E_n$ for $R \in SO(3)$. The proof can be completed as before by taking limits in the square mean.▲

Proposition 19 *Let Y be the solution of (2.94) and suppose the covariance matrix C of B is of the form $a^2 I$ where $a \in \mathbb{R}$. For $t \geq 0$ we have*

$$\phi_t(l) = e^{-\frac{a^2}{2}l(l+1)t} I_l \tag{2.103}$$

for $l \geq 0$, where $\phi_t \equiv \phi_{Y_t}$.

Proof: By Proposition 18 we have that for $t \geq 0$ the rotation random variable Y_t is conjugate invariant. It follows by (i) of Proposition 6 that for $t \geq 0$ and $l \geq 0$ we have $\phi_t(l) = a_l(t)I_l$ where $a_l(t) \in \mathbb{R}$. From expression (2.28) of Paragraph 2.1.1.3 and definition (2.32) we have for $t \geq 0$ and $l \geq 0$ that –again superscripts are used to denote matrix elements.

$$a_l(t) = \mathbb{E}[U^l_{00}(Y_t)] = \mathbb{E}[P_l(Y_t^{33})]$$

we will compute this by replacing equation (2.96) in Itô's formula. Note first that we now have in (2.96) $\tilde{J}_t = J_t - a^2 tI$ for $t \geq 0$. It follows that

$$dY_t^{33} = Y_t^{31}dB_t^2 - Y^{32}dB_t^1 - a^2 Y^{33}dt \qquad Y_0^{33} = 1$$

and Itô's formula can now be written down

$$dP_l(Y_t^{33}) = \frac{a^2}{2}[(1-(Y_t^{33})^2)P_l''(Y_t^{33}) - 2Y_t^{33}P_l'(Y_t^{33})]dt + P_l'(Y_t^{33})[Y_t^{31}dB_t^2 - Y^{32}dB_t^1] \tag{2.104}$$

where the prime stands for derivation. The second term only contains Brownian differentials and has no influence on $\mathbb{E}[P_l(Y_t^{33})]$. By discarding this term we obtain

$$\frac{d}{dt}a_l(t) = \frac{a^2}{2}\mathbb{E}[(1-(Y_t^{33})^2)P_l''(Y_t^{33}) - 2Y_t^{33}P_l'(Y_t^{33})] \tag{2.105}$$

Note the differential equation for the Legendre polynomial P_l –see for instance [67].

$$(1-x^2)P_l''(x) - 2xP_l'(x) = -l(l+1)P_l(x)$$

Replacing this in (2.105) we have

$$\frac{d}{dt}a_l(t) = -\frac{a^2}{2}l(l+1)a_l(t) \tag{2.106}$$

which immediately leads to (2.103) given the initial conditions.▲

Based on Proposition 19, we can now give the asymptotic behavior in distribution of conjugate invariant rotation Brownian motion. This is done in Proposition 20 below which gives for conjugate invariant rotation Brownian motion the same result obtained in Proposition 16 for rotation compound Poisson processes. The proof of Proposition 20 is an immediate application of (2.103) and (ii) of Proposition 2. Indeed, it is clear from (2.103) that $\phi_t(l) \to 0$ for all $l > 0$ if $a \neq 0$.

Proposition 20 *Let U be a uniformly distributed rotation random variable. If Y is the solution of (2.94) where $C = a^2 I$ with real $a \neq 0$ then $Y_t \xrightarrow{d} U$ when $t \uparrow \infty$.*

2.3.4 Interlacing

Rotation compound Poisson processes and rotation Brownian motion are the essential rotation Lévy processes. General rotation Lévy processes can be obtained from these two kinds of processes using the interlacing construction. Given a rotation compound Poisson process $(Y_t)_{t\geq 0}$ and a rotation Brownian motion $(W_t)_{t\geq 0}$ we will see in the current subsection how the interlacing construction can be used to obtain a new rotation Lévy process $Z = (Z_t)_{t\geq 0}$ from Y and W. We also consider the problem of calculating the characteristic function of Z when W is a conjugate invariant rotation Brownian motion. This is given in Proposition 21. This can be used to determine the symmetry properties and asymptotic behavior in distribution of Z in the case where W is conjugate invariant. The interlacing construction is well-known in the theory of Lévy processes with values in Lie groups [40]. It plays the same role we have described in the case of rotation processes, leading to the construction of general Lévy processes.

The current subsection has been included for completeness of our presentation of rotation Lévy processes. However, very little use will be made of its content in the following. All that is required for Section 2.4 is a general understanding of Proposition 21. For reasons of space, we do not include a complete proof of this proposition. It is discussed below that such a proof can be obtained by generalizing that of Proposition 19.

We start by describing the interlacing construction. Suppose we have processes Y and W as above. Note $N = (N_t)_{t\geq 0}$ and $(X_n)_{n\geq 1}$ the Poisson process and *i.i.d.* rotation random variables used to define Y as in Subsection 2.3.2. Let $T_0 = 0$ and suppose $(T_n)_{n\geq 1}$ are the jump times of N. The interlacing construction gives the process Z defined as follows. We have the initial condition $Z_0 = I$. For $t > 0$ and $n \geq 1$

$$Z_t = Z_{T_{n-1}} W_{T_{n-1}}^{-1} W_t \quad \text{for } T_{n-1} \leq t < T_n \tag{2.107}$$

where the following formula holds at each time T_n –note that Z_{T_n-} stands for the left limit at T_n.

$$Z_{T_n} = Z_{T_n-} X_n \tag{2.108}$$

This definition is sufficient since $T_n \uparrow \infty$ almost surely. The term interlacing comes from the fact that the paths of Z are obtained by introducing the jumps of Y into the paths of W as these jumps occur. The paths of W are thus interlaced with the jumps of Y. We refer to processes of the form Z as interlaced processes.

Under the assumption that the processes W and N and the family $(X_n)_{n\geq 1}$ are independent, it can be shown that the process Z is a rotation Lévy process with *càdlàg* paths. With this same assumption, Proposition 21 obtains the characteristic function of Z in the case where W is conjugate invariant. The main difficulty in the definition of Z is that it is defined by pieces, on each stochastic interval $T_{n-1} \leq t < T_n$.

Proposition 21 *Let λ be the parameter of the Poisson process N and suppose W is conjugate invariant with its characteristic function given as in formula (2.103) of Proposition 19. For $t \geq 0$ the characteristic function $\phi_t \equiv \phi_{Z_t}$ is given by*

$$\phi_t(l) = \exp\left[t\lambda\phi(l) - tI_l\left(\lambda + \frac{l(l+1)a^2}{2}\right)\right] \tag{2.109}$$

for $l \geq 0$, where $\phi \equiv \phi_{X_1}$.

In order to prove this proposition, it is possible to follow the same steps as for Proposition 19. Given a stochastic differential equation satisfied by Z, Itô's formula can be applied to obtain (2.109). Note that Z is a discontinuous process and one must accordingly apply the so-called discontinuous Itô formula.

The process Z verifies equation (2.111) below. In order to state this equation consider the following processes $J = (J_t)_{t\geq 0}$ and $X = (X_t)_{t\geq 0}$.

$$J_t = B_t^1 J_1 + B_t^2 J_2 + B_t^3 J_3 \qquad X_t = \sum_{n\geq 0} \mathbf{1}_{[0,t]}(T_n)[X_n - I] \tag{2.110}$$

Here $[B^1, B^2, B^3]$ are the coordinates of a $\mathbb{R}^3$-valued Brownian motion B with covariance matrix $C = a^2 I$, for some $a \in \mathbb{R}$. Moreover, we continue to use the notation $X_0 = I$. The process J corresponds to the rotation Brownian motion process W as in equation (2.94). The process X appears in (2.111) in order to account for the jumps of Z as in (2.108). Equation (2.111) is stated in terms of these two processes. In the absence of the second term on the right hand side, this is exactly equation (2.94).

$$dZ_t = Z_{t-}[dJ_t + dX_t] \qquad Z_0 = I \tag{2.111}$$

The differential dJ_t of the Brownian process J is to be understood as a Stratonovich differential. On the other hand, the differential dX_t is an ordinary differential since the process X is of finite variation. The proof that Z does indeed verify (2.111) is straightforward but contains several lengthy technical steps. We do not consider it here.

For $l \geq 0$ we are interested in the process Z^l where for $t \geq 0$ we have $Z_t^l = U^l(Z_t)$. Admitting that Z is a rotation Lévy process it can be shown immediately that Z^l has similar properties. For example, we have for all $0 \leq s \leq t$ that

$$[Z_s^l]^\dagger Z_t^l \stackrel{d}{=} Z_{t-s}^l$$

which results immediately from the property (L3) of Z and the homomorphism property (2.18) of U^l. Generally speaking, we have for $l \geq 0$ that the process Z^l has independent and stationary increments and that $Z_0^l = I_l$ almost surely. Moreover, since Z is stochastically continuous, Z^l is also stochastically continuous.

For $l \geq 0$, Itô's discontinuous formula can be used to give a stochastic differential equation verified by the process Z^l. This is a linear stochastic differential equation similar to (2.111). Without giving this equation explicitly, it is possible to infer (2.109). Indeed, it results from the discontinuous Itô formula that the equation for Z^l is of the form

$$dZ_t^l = Z_{t-}^l[dJ_t^l + dX_t^l] \qquad Z_0^l = I_l \tag{2.112}$$

Where $J^l = (J_t^l)_{t\geq 0}$ is a Brownian process given by $[B^1, B^2, B^3]$ and $X^l = (X_t^l)_{t\geq 0}$ is the process

$$X_t^l = \sum_{n\geq 0} \mathbf{1}_{[0,t]}(T_n)[U^l(X_n) - I]$$

Formula (2.109) can be obtained directly from (2.112). It is enough to average both sides of this equation and to remember that the Brownian part of this equation yields formula (2.106) as in Proposition 19.

It is important that we end with the following clarification. As we have stated, the interlacing construction can be used to describe any rotation Lévy process. It should not however be thought that any rotation Lévy process is an interlaced process. In all generality, a rotation Lévy process can be obtained as a uniform limit of interlaced processes and there is no guarantee that it will be itself an interlaced process.

2.4 Application to multiple scattering

The current section is based on our paper [66]. The main focus is a stochastic approach to the inverse problem of multiple scattering. This is treated throughout Subsections 2.4.1 and 2.4.2.

The inverse problem of multiple scattering aims to infer the properties of complex, *e.g.* geophysical or biological, media by considering multiple scattering of mechanical or electromagnetic waves by these media. In general, this problem can be formulated within the framework of various approximations of the exact equations of radiative transfer [25]. The equations of radiative transfer are concerned with the macroscopic intensity of the wavefield. However, in multiple scattering situations, these equations admit of a straightforward microscopic interpretation. Considering a specific example, we will use this interpretation to introduce a stochastic model of multiple scattering based on rotation compound Poisson processes, which we have developed in Subsection 2.3.2. Within this model, the physical inverse problem of multiple scattering is reformulated as a statistical nonparametric estimation problem, namely the problem of decompounding. Subsection 2.4.1 presents our stochastic model of multiple scattering. The inverse problem of multiple scattering is identified with a special case of the problem of decompounding. In Subsection 2.4.2 this problem is considered independently of its applications. We provide a solution using a characteristic function method and apply this solution to numerical simulations related to Subsection 2.4.1.

Characteristic function methods for nonparametric estimation on compact Lie groups have received special attention in the engineering community. Many applications of such methods are reviewed in [67]. Recent contributions include [76] and [36]. The same methods were studied in the case of the rotation group and other related settings in [32, 33]. An essential difference between the problems treated in these works and the problem of decompounding treated in Subsection 2.4.2 is the fact that decompounding proceeds from indirect observations. We will return to this aspect in Paragraph 2.4.2.2.

2.4.1 The compound Poisson model of multiple scattering

Using rotation compound Poisson processes, which were introduced in Subsection 2.3.2, leads to a model of multiple scattering which is sufficiently precise as well as amenable to statistical treatment. This model will be seen to recover the results of the small angles approximation of radiative transfer. It also allows the formulation of the physical inverse problem of multiple scattering as the statistical nonparametric estimation problem of decompounding. Let us note that a compound Poisson model for the direct problem of multiple scattering was considered in [51]. However, this is based on real-valued compound Poisson processes.

We consider a specific example. The development of Subsection 2.3.2 is converted into the terminology of radiative transfer –see [25]. A scalar plane wave is perpendicularly incident upon a plane parallel multiple scattering layer of thickness H. The global properties of multiple scattering in this layer are determined by the mean free time τ and the mean free path ℓ. Normalizing the velocity of the wave we have $\tau = \ell$. The individual properties of scatterers are determined by the power spectrum of heterogenities which we introduce below. Suppose coordinates and time origin are chosen so that the wave enters the layer at time 0 with direction of propagation s_0 whose coordinates are $[0, 0, 1]$. The geometry of the layer suggests that the wave is transmitted uniformly in every plane section. If $0 \leq z \leq H$ denotes the distance along s_0 into the layer, then the direction of propagation s at any point in the layer is given by s_z with coordinates $[s_z^1, s_z^2, s_z^3]$. Since s_z represents a direction of propagation, we have $s_z \in S^2$ for $0 \leq z \leq H$.

Although s_0 is a determined quantity, s_z for $0 < z \leq H$ should be treated as random. Indeed, the interaction of the wave with the layer takes place in the form of a succession of scattering events. These are understood as interaction of the wave with individual scatterers. Both the time between scattering events and the effect of scatterers on s_z are subject to fluctuations and random physical effects. The distribution of the random variable s_H will be noted I_H. We will identify I_H with the normalized angular pattern of intensity transmitted by the layer. We return below to the validity of this identification. Let us already mention that it will recover the results of the small angles approximation of radiative transfer.

Note N_z the random number of scattering events occurring through the distance $0 \leq z \leq H$. Suppose the n^{th} scattering event takes place at a distance $0 \leq T_n \leq H$. This affects the direction of propagation as follows

$$s_{T_n} = s_{T_n -} X_n \tag{2.113}$$

here X_n is a rotation random variable and we will identify s with a line matrix of coordinates. From (2.113) and the definition of N_z we have –the product is ordered from left to right.

$$s_z = s_0 \left(\prod_{n=0}^{N_z} X_n \right) \tag{2.114}$$

A certain number of standard physical hypotheses can be replaced in (2.114). This will allow for the random product to be exhibited as a zonal invariant rotation compound Poisson process.

Under the condition $\ell \ll H$ it is possible to make the hypothesis that the free path between successive scattering events has an exponential distribution [69]. This allows us to model N_z as a Poisson process with parameter $1/\ell$. Moreover, we suppose the scatterers identical and scattering events independent. This amounts to taking the rotation random variables X_n to be *i.i.d.*. If the additional assumption is accepted that the number of scattering events is independent of the whole outcome of these events then formula (2.114) can be rewritten $0 \leq z \leq H$

$$s_z = s_0 Y_z \tag{2.115}$$

Where Y is a rotation compound Poisson process. It is usual to assume that the random variables X_n have a common probability density p. In the theory of radiative transfer, p is known as the phase function of the layer [25].

In order to simplify p we profit from the physical hypothesis of statistical isotropy. This implies that scattering events in the layer as given by (2.113) are symmetric around the direction of propagation $s_{T_n -}$. Statistical isotropy is a valid assumption in many concrete situations. It is verified by analytical models such as Gaussian and Henyey-Greenstein phase functions, commonly used to describe scattering in geophysical and biological media [34]. Under the hypothesis of statistical isotropy the rotation random variables X_n are zonal invariant and p is given by (2.48)

$$p(\cos\beta) = \sum_{l \geq 0} d_l x_l P_l(\cos\beta)$$

where the coefficients x_l for $l \geq 0$ are said to form the power spectrum of heterogenities [25]. If p is the Henyey-Greenstein phase function then the power spectrum of heterogenities is given by $x_l = g^l$ for ≥ 0 and p can be expressed in the closed form [34, 35]

$$p(\cos\beta) = \frac{1 - g^2}{(1 + g^2 - 2g\cos\beta)^{\frac{3}{2}}} \tag{2.116}$$

The Euler angle $0 \leq \beta \leq \pi$ gives the scattering angle for a single scatterer interacting with the wave whose direction of propagation is s_0. The parameter $g \in [0, 1[$ is called the anisotropy or asymmetry parameter. It can be shown to give the average cosine of the scattering angle β. For the scattering of light waves by water clouds and blood we have respectively $g = 0.85$ and $g = 0.95$, see [35].

Propositions 14 and 15 can be used to give the angular transmitted pattern of intensity I_H in terms of the power spectrum of heterogenities. This is expressed in the following equation (2.117). The hypothesis of statistical isotropy implies that the X_n are zonal invariant. It follows by Proposition 15 that Y_H is zonal invariant. The characteristic function of Y_H is then given by Proposition 14 with nonzero elements determined by Proposition 7 of Paragraph 2.1.2.2. The distribution of s_H is immediately related to that of Y_H due to equation (2.115). Indeed, both distributions depend only on the scattering angle $0 \leq \beta \leq \pi$ between s_0 and s_H. We have

$$\frac{I_H(\beta)}{2\pi} = \sum_{l \geq 0} (2l+1) e^{\frac{H}{\ell}(x_l - 1)} \int_0^\beta P_l(\cos t) \sin t dt \tag{2.117}$$

For the ratio $I_H(\beta)$ of intensity transmitted within a pencil of angle 2β around s_0.

Equation (2.117) is well-known in the small angles approximation of radiative transfer where it is derived under the assumption of strong forward scattering [25]. Mathematically, this translates into a phase function p with a sharp peak around $\beta = 0$. This assumption was not explicitly made in our development of equation (2.117) using rotation compound Poisson processes. However, the identification of I_H with the angular pattern of transmitted intensity implicitly requires for all the intensity of the wave entering the layer to be transmitted. This precludes an important deviation between s_0 and s_H. This identification is justified in situations with strong forward scattering where most of the intensity will be measured in the forward direction. The domain of its validity has been studied using Monte Carlo simulations of the exact equations of radiative transfer developed in [45]. In general, this interpretation of I_H is not valid when scattering angles β appear which are close to the so-called grazing angle $\beta = \pi/2$.

Equation (2.117) relates a directly observable outcome of multiple scattering by the layer to the microscopic properties of the layer. This constitutes an interesting starting point for the formulation of the inverse problem of multiple scattering. Our stochastic model of multiple scattering reformulates this physical inverse problem as a statistical nonparametric estimation problem. The key to this reformulation is the idea that attention should be shifted from the intensity of the wavefield to the direction of propagation characterized as a spherical random variable. Supposing a situation where (2.117) holds, being able to invert it implies access to the power spectrum of heterogenities or alternatively the phase function from direct intensity measurements. This implies inference of physical parameters such as the parameter g of the Henyey-Greenstein phase function or determination of microscopic properties such as the shape of individual scatterers [35].

Our statistical reformulation of this inverse problem consists in considering (2.115) in addition to (2.117). Instead of carrying out measurements of transmitted intensity, we propose to recover the microscopic properties of the layer from observations of s_H. Such observations are equivalent to observations of Y_H. According to the discussion before (2.117), both amount to observations of the scattering angle β between s_0 and s_H. Instead of a physical inverse problem considering only (2.117) we propose a nonparametric estimation problem formulated in terms of the rotation compound Poisson process Y. This is the problem of decompounding which we consider in the following Subsection 2.4.2. Its aim is to estimate the probability density p of the rotation random variables X_n appearing in (2.114) from observations of Y_H. This nonparametric estimation problem can be replaced by a parametric problem if the objective is to infer the power spectrum of heterogenities or the parameter

g of the Henyey-Greenstein phase function (2.116).

2.4.2 The problem of decompounding

We give a precise formulation of the problem of decompounding as identified in the last subsection. Although this problem was motivated by the inverse problem of multiple scattering, it is here considered independently of any applications. In Paragraph 2.4.2.1 a characteristic function method is proposed for the solution of the problem of decompounding. The convergence of the estimates given by this method is proved in Paragraph 2.4.2.2. Finally, Paragraph 2.4.2.3 applies the method of Paragraph 2.4.2.1 to numerical simulations related to the physical situation of Subsection 2.4.1.

In existing literature, *decompounding* refers to a set of nonparametric estimation problems involving scalar compound Poisson processes [8, 72] –the definition of such processes was reminded in (2.89). The main application of these problems is in queuing problems and risk theory [8]. Our paper [66] considered the generalization of these problems to compact Lie groups. According to the general theme of this chapter, the current subsection considers the problem of decompounding for rotation compound Poisson processes. The solution presented here is simply a special case of the general solution considered in [66]. It can be seen as a generalization of the approached used in [72] for scalar compound Poisson processes.

Consider a rotation compound Poisson process Y defined as in (2.90). In particular, λ denotes the parameter of the Poisson process N used to define Y. The problem of decompounding consists in estimation of the common probability density (supposed to exist) of the rotation random variables X_n from observations of the process Y. The unknown probability density will be noted p. Several versions of the problem of decompounding can be stated, depending on the nature of observations made of Y, compare to [66, 72]. We are interested in the following version which corresponds to the analysis of the inverse problem of multiple scattering given in Subsection 2.4.1. We fix $T \geq 0$ and suppose that *i.i.d.* observations $(Z_n)_{n\geq 1}$ are available of a noisy version Z of Y_T. We aim to estimate p from these observations. Z is related to Y_T according to a multiplicative noise model

$$Z = MY_T \tag{2.118}$$

where M and Y_T are independent. By (ii) of Proposition 3 we have for the characteristic function of Z

$$\phi_Z = \phi_M \phi_{Y(T)}$$

The noise model is equivalent to replacing the initial condition $Y_0 = I$ with $Y_0 = M$ which has a general distribution. We consider the case of Brownian noise. The characteristic function of M is then given by (2.103) of Proposition 19

$$\phi_M(l) = e^{-\frac{\sigma^2}{2} l(l+1)} I_l$$

where σ^2 is a variance parameter. Note that this corresponds to a conjugate invariant rotation Brownian motion. In particular, by (ii) of Proposition 6, left and right multiplication of Y_T by the noise M are indifferent, as far as the distribution of Z is concerned. It is possible to interpret the noise model (2.118) in terms of the interlacing construction of Subsection 2.3.4. Using the notation of Proposition 21 we have that $Z_T \overset{d}{=} Z$ if $Ta^2 = \sigma^2$. Accordingly, the observations $(Z_n)_{n\geq 1}$ are equivalent to *i.i.d.* observations of the interlaced process $(Z_t)_{t\geq 0}$ at time T. This process can be seen as a noisy version of the process Y. In relation to the stochastic model of Subsection 2.4.1 we interpret the process $(Z_t)_{t\geq 0}$ as follows. This process includes in addition to the process Y which

models the effect of multiple scattering a conjugate invariant Brownian motion W which we understand as corresponding to local fluctuations taking place between scattering events.

The following Paragraphs 2.4.2.1 and 2.4.2.2 give our solution to the problem of decompounding as stated here. This solution will reflect a difficulty which does not appear in the case of scalar decompounding. Under the conditions of Proposition 16, if the observations $(Z_n)_{n\geq 1}$ correspond to a sufficiently large time T then these observations will be uniformly distributed and will have no memory of the random variables X_n.

2.4.2.1 A characteristic function method

We present a characteristic function method for the problem of decompounding. This is a special case of the method proposed in [66] in the case of compact Lie groups. We proceed from the observations $(Z_n)_{n\geq 1}$ defined according to the noise model (2.118). The unknown density p is given by the following Fourier series

$$p = \sum_{l\geq 0} d_l \operatorname{tr}\left(\phi_X(l) U^{l\dagger}\right) \tag{2.119}$$

where $\phi_X \equiv \phi_{X_1}$. This is simply rewriting formula (2.42) of Paragraph 2.1.2.2. A characteristic function method consists in constructing nonparametric estimates for p from parametric estimates for its Fourier coefficients $\phi_X(l)$ given for $l \geq 0$.

We suppose λ and σ^2 are known. As in Proposition 21, the characteristic function of Z defined in (2.118) is given for $l \geq 0$ by

$$\phi_Z(l) = \exp[T\lambda\phi_X(l) - T\bar{\lambda} I_l] \tag{2.120}$$

where $\bar{\lambda}$ is a constant determined by λ and σ^2. We refer to this transformation $\phi_X \mapsto \phi_Z$ as the compounding transformation. Decompounding will involve local inversion of the compounding transformation. This is clearly related to inversion of the matrix exponential in a neighborhood of $\phi_Z(l)$ for $l \geq 0$. Rather than deal with this problem in general, we make the following simplifying hypothesis.

Hypothesis: X_1 is inverse invariant.

For all $l \geq 0$ we have from (i) of Proposition 5 of Paragraph 2.1.2.2 and by Proposition 15 of Subsection 2.3.2 that both $\phi_X(l)$ and $\phi_Z(l)$ are Hermitian. Moreover, it is clear from (2.120) that $\phi_Z(l)$ is positive definite. Note Log the unique Hermitian matrix logarithm of a Hermitian positive definite matrix. We can now express the inverse of the compounding transformation. From equation (2.120) it follows that

$$\phi_X(l) = \frac{1}{T\lambda} \log[\phi_Z(l)] + (\bar{\lambda}/\lambda) I_l \tag{2.121}$$

Let $l \geq 0$. It follows from the definition (2.32) of the characteristic function that empirical estimates of $\phi_Z(l)$ based on the observations $(Z_n)_{n\geq 1}$ are unbiased and consistent. This is a simple consequence of the strong law of large numbers. See for example [27]. In order to estimate $\phi_X(l)$ using (2.121) it is then important to ensure that the empirical estimates of $\phi_Z(l)$ are asymptotically Hermitian positive definite.

We start by defining the empirical estimates $\hat{\phi}_Z^n(l)$ for $l \geq 0$ and $n \geq 1$

$$\hat{\phi}_Z^n(l) = \frac{1}{2n} \sum_{m=1}^{n} \left(U^l(Z_m) + U^l(Z_m)^\dagger\right)$$

Hermitian symmetrization of empirical estimates is necessary for the application of (2.121). Since it is a projection operation, this symmetrization moreover contributes to a faster convergence of the $\hat{\phi}_Z^n(l)$ to $\phi_Z(l)$.

Continuous dependence of the spectrum of a matrix on its coefficients is a classical result in matrix analysis. Several more or less sophisticated versions of this result exist [41]. For a remarkably straightforward statement see [71]. The spectrum of a complex matrix C will be noted $\lambda(C)$. For each $l \geq 0$ and $n \geq 1$ define the event R_l^n by

$$R_l^n = \{\lambda(\hat{\phi}_Z^n(l)) \subset]0, \infty[\}$$

For $l \geq 0$, the sequence $(R_l^n)_{n\geq 1}$ controls the convergence of the spectra of the empirical estimates $\hat{\phi}_Z^n(l)$. In particular,

$$\mathbb{P}(\cup_{n\geq 1} \cap_{m\geq n} R_l^n) = \lim_n \mathbb{P}(\cap_{m\geq n} R_l^n) = 1$$

Using the events R_l^n we can write down well-defined estimates of ϕ_X. These are noted $\hat{\phi}_X^n(l)$ for $l \geq 0$ and $n \geq 1$

$$\begin{array}{rcll} \hat{\phi}_X^n(l) & = & 0 & \text{on} \quad \Omega - R_l^n \\ \hat{\phi}_X^n(l) & = & \frac{1}{T\lambda} \operatorname{Log}\left[\hat{\phi}_Z^n(l)\right] + \left(\bar{\lambda}/\lambda\right) I_l & \text{on} \quad R_l^n \end{array}$$

This expression gives our parametric estimates for the Fourier coefficients of p. We use them to construct nonparametric estimates based on the Fourier series (2.119). Let $K \geq 0$ and for $l \geq 0$ note

$$f_l = d_l e^{-Kl(l+1)}$$

For $n \geq 1$ and $L \geq 0$ our nonparametric estimate $\hat{p}_L^n$ is given by

$$\hat{p}_L^n = \sum_{l=0}^{L} f_l \operatorname{tr}\left(\hat{\phi}_X^n(l) U^{l\dagger}\right) \tag{2.122}$$

The subscript L corresponds to a cutoff or smoothing parameter. Indeed, infinitely many values of l are excluded from the sum (2.122). When $K > 0$ the coefficients f_l form a convolution mask ensuring that the estimates $\hat{p}_L^n$ can be taken to converge to a smooth probability density, see Paragraph 2.4.2.2.

It is usual to write expressions similar to (2.122) in terms of a rotation invariant kernel. See [32,36]. Such a transformation is not possible here due to the indirect nature of our observations. This is in particular related to the more involved form of the $\hat{\phi}_X^n(l)$ as given above.

2.4.2.2 Convergence of parametric and nonparametric estimates

Here we discuss the convergence of the parametric and nonparametric estimates given in Paragraph 2.4.2.1. Our argument is presented in the form of Propositions 22 and 23 below. Proposition 22 gives the consistency of the parametric estimates $\hat{\phi}_X^n(l)$. Proposition 23 states a subsequent result for the nonparametric estimates $\hat{p}_L^n$.

For Proposition 22 we will need inequalities (2.123) and (2.124). These express stability results for the eigenvalues of Hermitian matrices and for the Hermitian matrix function Log. Let A and B be Hermitian $d \times d$ matrices, for some $d \geq 1$. For $1 \leq i \leq d$ let α_i and β_i be the eigenvalues of A and B respectively. Suppose they are arranged in nondecreasing order. We have

$$\sum_{i=1}^{d} (\beta_i - \alpha_i)^2 \leq |B - A|^2 \tag{2.123}$$

This inequality is known as the Wielandt-Hoffman theorem. In [41], it is stated for A and B real symmetric. The general case of Hermitian A and B can be obtained from this statement using a canonical realification isomorphism.

Suppose A and B are positive definite. For our purpose it is suitable to assume both $\lambda(A)$ and $\lambda(B)$ are contained in an interval $[k,1]$ for some $k>0$. Under this assumption we have the following Lipschitz property

$$|\operatorname{Log}(B)-\operatorname{Log}(A)|\leq\sqrt{d}k^{-2}|B-A| \tag{2.124}$$

In order to obtain (2.124) it is possible to start by expressing $\operatorname{Log}(A)$ as follows

$$\operatorname{Log}(A)=\int_0^1(A-I_d)[t(A-I_d)+I_d]^{-1}dt$$

This expression results from a similar one for the real logarithm applied to each eigenvalue of A. Subtracting the same expression for $\operatorname{Log}(B)$, (2.124) follows by simple calculations.

Proposition 22 *For all $l\geq 0$ we have the limit in probability $\lim_n\hat{\phi}_X^n(l)=\phi_X(l)$.*

Proof: We only need to consider $l>0$. Indeed, $\hat{\phi}_X^n(0)=\phi_X(0)=1$ for all $n\geq 1$. Let $l>0$, for all $n\geq 1$ we have

$$|\hat{\phi}_Z^n(l)|_{op}\leq\frac{1}{2n}\sum_{m=1}^n|U^l(Z_m)|_{op}+|U^l(Z_m)^\dagger|_{op}=1$$

where $|.|_{op}$ is the operator matrix norm. Passing to the limit, it follows from the strong law of large numbers that the same inequality holds for $\phi_Z(l)$. In other words, all the eigenvalues of $\hat{\phi}_X^n(l)$ and of $\phi_X(l)$ are less than unity. Since $\phi_Z(l)$ is positive definite, there exists $k_l>0$ such that $\lambda(\phi_Z(l))\subset[k_l,1]$. For $n\geq 1$ note $\tilde{R}_l^n$ the event

$$\tilde{R}_l^n=\{\lambda(\hat{\phi}_Z^n(l))\subset[k_l/2,1]\}$$

from inequality (2.123) we have

$$\mathbb{P}(\Omega-\tilde{R}_l^n)\leq\mathbb{P}(|\hat{\phi}_Z^n(l)-\phi_Z(l)|>k_l/2)$$

Since $\tilde{R}_l^n\subset R_l^n$, it follows from inequality (2.124) that

$$\mathbb{P}(|\hat{\phi}_X^n(l)-\phi_X(l)|>\varepsilon\cap\tilde{R}_l^n)\leq\mathbb{P}(|\hat{\phi}_Z^n(l)-\phi_Z(l)|>k_l^2\varepsilon/M)$$

for all $\varepsilon>0$, where $M=4\sqrt{d_l}/T\lambda$.

The proof can be completed by a usual application of Chebychev's inequality,

$$\mathbb{P}(|\hat{\phi}_X^n(l)-\phi_X(l)|>\varepsilon)\leq\left(\frac{8+2M^2/\varepsilon^2}{n}\right)\left(\frac{\sqrt{d_l}}{k_l^2}\right)^2 \tag{2.125}$$

for all $\varepsilon>0$.▲

Proposition 23 relies on Proposition 22 and the Peter-Weyl theorem –Theorem 1 of Paragraph 2.1.1.3. It implies the existence of sequences $(\hat{p}_k)_{k\geq 1}$ of nonparametric estimates given by (2.122) converging to p in probability in $L^2(SO(3),\mu)$. Convergence in probability in $L^2(SO(3),\mu)$ means that the following limit in probability holds

$$\lim_k\|\hat{p}_k-p\|=0$$

where $\|.\|$ is the $L^2(SO(3),\mu)$ norm given by the scalar product (2.21) of Paragraph 2.1.1.3

$$\|p\|^2=\langle p,p\rangle=\int_{SO(3)}p^2d\mu$$

Proposition 23 *Putting $K = 0$ in (2.122), we have the limit in probability*

$$\lim_L \lim_n \|\hat{p}_L^n - p\| = 0$$

Proof: For $L \geq 1$ let $p_L \in L^2(SO(3), \mu)$ be given by

$$p_L = \sum_{l=0}^{L} d_l \operatorname{tr}\left(\phi_X(l) U^{l\dagger}\right)$$

By the Peter-Weyl theorem, $\lim_L \|p_L - p\| = 0$. By (2.122) and Proposition 22 we have $\lim_n \|\hat{p}_L^n - p_L\| = 0$ in probability for all $L \geq 1$. This follows since for all $n, L \geq 1$

$$\|\hat{p}_L^n - p_L\|^2 = \sum_{l=0}^{L} d_l |\hat{\phi}_X^n(l) - \phi_X(l)|^2$$

To conclude it is enough to observe that

$$\|\hat{p}_L^n - p\|^2 = \|\hat{p}_L^n - p_L\|^2 + \|p_L - p\|^2 \tag{2.126}$$

for all $n, L \geq 1$.▲

Proposition 22 obtained convergence in probability of the parametric estimates $\hat{\phi}_X^n(l)$ for all $l \geq 0$. These parametric estimates depend only on the observations. In particular, they can be evaluated without any *a priori* knowledge of p. By introducing such knowledge, it is possible to define parametric estimates $\tilde{\phi}_X^n(l)$ converging in the square mean to the same limits $\phi_X(l)$. For $l \geq 0$ and $n \geq 1$ the $\tilde{\phi}_X^n(l)$ are given by

$$\begin{array}{rcll} \tilde{\phi}_X^n(l) & = & 0 & \text{on} \quad \Omega - \tilde{R}_l^n \\ \tilde{\phi}_X^n(l) & = & \frac{1}{T\lambda} \operatorname{Log}\left[\hat{\phi}_Z^n(l)\right] + \left(\bar{\lambda}/\lambda\right) I_{d_l} & \text{on} \quad \tilde{R}_l^n \end{array}$$

where the events $\tilde{R}_l^n$ are as in the proof of Proposition 22 and we assume known *a priori* constants k_l necessary for their definition. As in (2.122), we can define nonparametric estimates $\tilde{p}_L^n$ where for $n, L \geq 1$

$$\tilde{p}_L^n = \sum_{l=0}^{L} f_l \operatorname{tr}\left(\tilde{\phi}_X^n(l) U^{l\dagger}\right)$$

For all $l \geq 0$ and $n \geq 1$ we have

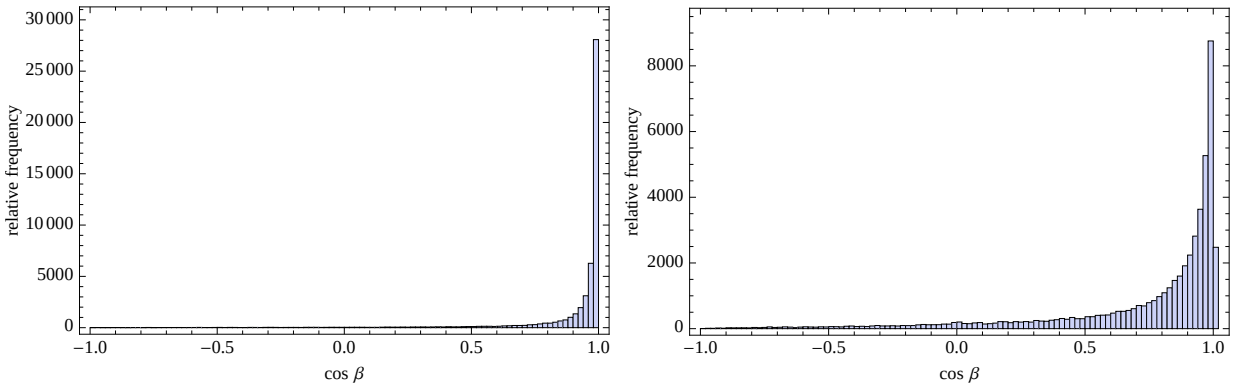
$$\mathbb{E}|\tilde{\phi}_X^n(l) - \phi_X(l)|^2 \leq \frac{M'}{n} \left(\frac{d_l}{k_l^2}\right)^2 \tag{2.127}$$

where M' is a constant depending on the product $T\lambda$. This follows by a reasoning similar to the proof of Proposition 22. Moreover, for all $n, L \geq 1$ we have after putting $K = 0$

$$\mathbb{E}\|\tilde{p}_L^n - p\|^2 \leq \frac{M'}{n} \sum_{l=0}^{L} (d_l^3/k_l^4) + \|p_L - p\|^2 \tag{2.128}$$

for the functions p_l defined in the proof of Proposition 23. This follows from formula (2.126).

We have characterized the convergence of parametric estimates using (2.125) and (2.127) and the convergence of nonparametric estimates using (2.126) and (2.128). We make the following remarks on these formulae. Inequalities (2.125) and (2.127) only give gross bounds for the rate of convergence of parametric estimates. The quality of these bounds improves when the constants k_l are greater,



(a) Histogram of $\cos\beta$ under density p

(b) Histogram of $\cos\beta$ under distribution of $Y(T)$

Figure 2.1: Compounding transformation of p (histograms)

i.e. closer to the value 1. This is equivalent to the $L^2(SO(3),\mu)$ distance between p and the constant function 1 being greater. This last point can be appreciated in relation to the example of figure 2.3 in Paragraph 2.4.2.3. In Proposition 23 and inequality (2.128) we have ignored the role of the constant K. This is discussed in our paper [66].

Inequalities (2.125) and (2.127) describe the convergence of nonparametric estimates in a way similar to the one used in standard works on characteristic function methods [32, 36]. Indeed, the nonparametric estimation error is decomposed into two terms. One is given by the parametric estimation error and the other depends only on p. This second term is given by the convergence of the Fourier series of p. This is determined by the smoothness properties of p. We note the two following differences with [32, 36], both related to the indirect nature of our observations. First, the first and second terms in (2.128) cannot be identified as the "variance" and "bias" of $\tilde{p}_L^n$. Second, (2.128) characterizes the nonparametric estimation error as depending on the whole spectrum of p –through the constants k_l– rather than just its smoothness properties.

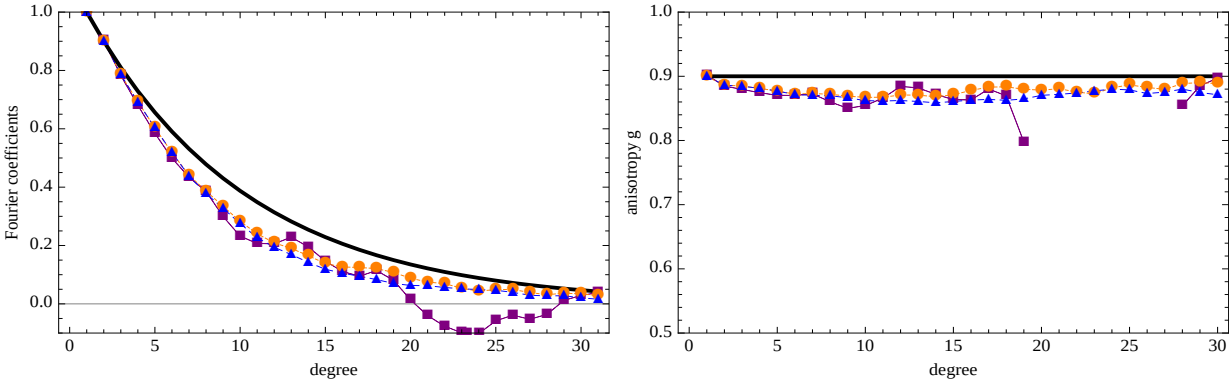
2.4.2.3 Numerical simulations

The characteristic function method of Paragraph 2.4.2.1 is here illustrated for a numerical example. In terms of the stochastic model for multiple scattering given in Subsection 2.4.1, this numerical example corresponds to an application of the problem of decompounding to the inference of the Henyey-Greenstein phase function p and anisotropy parameter g given in (2.116). Our example is of a rotation compound Poisson process Y expressed as in (2.90) of Subsection 2.3.2

$$Y_t = \prod_{n=0}^{N_t} X_n$$

for $t \geq 0$. Where the Poisson process N has parameter $\lambda = 0.3$ and the rotation random variables X_n are zonal invariant with the common probability density p given by expression (2.116). Four values will be considered for the parameter g in this expression: $0.85, 0.9, 0.95$ and 0.99. We will put $T = 10$ for the time at which observations of Y are taken. We simulate a number n of *i.i.d.* observations of $Y(T)$. The following values of n are used: $500, 5000$ and 50000. Note that on average the number $N(T)$ of factors involved in the random product $Y(T)$ is equal to 3.

Before going on, we confirm that the method of Paragraph 2.4.2.1 can be applied for this example. Since the rotation random variables X_n are zonal invariant, they are also inverse invariant. Indeed,



(a) Estimated Fourier coefficients $\hat{x}_l^n$ from decompounding (b) Corresponding estimates $\hat{g}_l^n$ of g (anisotropy parameter)

Figure 2.2: Influence of n ($\square = 5 * 10^2$; $\circ = 5 * 10^3$; $\triangle = 5 * 10^4$)

by (i) of Proposition 7 we have for $l \geq 0$ that the only nonzero element of the matrix $\phi_X(l)$ is $\phi_{00}^l = x_l$ with $x_l \in \mathbb{R}$ –see definition (2.47). It follows that for $l \geq 0$ the matrix $\phi_X(l)$ is diagonal and real. That the X_n are inverse invariant results immediately by (i) of Proposition 5. We have noted before expression (2.116) that $x_l = g^l$ for $l \geq 0$ where $g \in [0, 1[$.

We will present three sets of figures. Figure 2.1 is concerned with the compounding transformation of p. Figure 2.2 illustrates the influence of n on the estimation error for the Fourier coefficients x_l of p and for the parameter g. Figure 2.3 studies the influence of g on the estimation error for the Fourier coefficients x_l given a fixed n. For figures 2.1 and 2.2 we have $g = 0.9$. For figures 2.1 and 2.3 we have $n = 50000$. We now comment on each of these figures.

Figure 2.1 illustrates the relation between the distribution of the X_n as given by the density p and the distribution of $Y(T)$. Both these distributions are studied using histograms. The histogram in figure 2.1(a) is for the cosine of the Euler angle $0 \leq \beta \leq \pi$ associated with the rotation random variable X_1. The histogram in figure 2.1(b) is for the cosine of β associated with $Y(T)$. Figure 2.1 is concerned with the direct compounding transformation rather than the inverse decompounding transformation. It is meant to show the histogram in figure 2.1(b) as function of the one in 2.1(a). As expected, the latter histogram appears as a more heavy-tailed version of the former. This corresponds to the content of Proposition 16 of Subsection 2.3.2. Note also that the dominant value in figure 2.1(b) has moved away from $\beta = 0$.

For figure 2.2, the observations made of $Y(T)$ are used to carry out the decompounding approach of Paragraph 2.4.2.1. The estimation error for the Fourier coefficients x_l and for the parameter g is given graphically for different values of n. Figure 2.2(a) compares the estimated Fourier coefficients of p to their theoretical values $x_l = g^l$. In figure 2.2(b), *a priori* knowledge of the analytical form of the x_l is supposed. This is used to estimate g. A different parametric estimate is obtained from each estimated Fourier coefficient. In figures 2.2(a) and 2.2(b) theoretical values are represented by a thick solid line.

In figure 2.2(a) we have estimated the first $L = 31$ Legendre coefficients for each value of n. Let us note these estimated coefficients $\hat{x}_l^n$ for $0 \leq l < L$ and the corresponding value of n. They can be used to evaluate a nonparametric estimate of p as in formula (2.122). This is done by replacing them in a truncated Fourier series (2.48). We have the nonparametric estimate of p which we note $\hat{p}_L^n$

$$\hat{p}_L^n(\cos\beta) = \sum_{l=0}^{L-1} d_l \hat{x}_l^n P_l(\cos\beta)$$

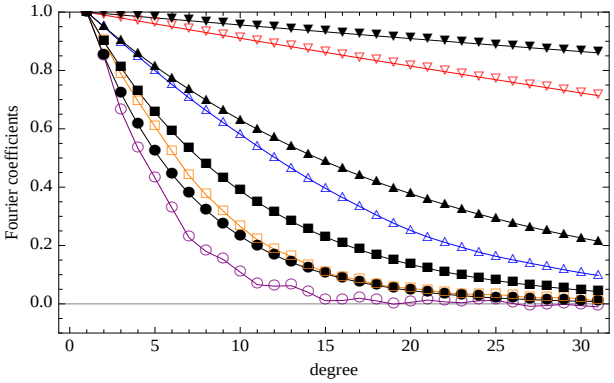


Figure 2.3: Influence of g ($\circ = 0.85$; $\square = 0.9$; $\triangle = 0.95$; $\nabla = 0.99$)

Depending on n, the random nonparametric estimation error from $\hat{p}_L^n$ is given by

$$\sum_{l<L} d_l(\hat{x}_l^n - x_l)^2 + \sum_{l\geq L} d_l x_l^2$$

this is the squared $L^2(SO(3), \mu)$ distance between $\hat{p}_L^n$ and p. In figure 2.2(a) the sum over $l < L$ appears as a weighted quadratic deviation between estimated and theoretical values.

In figure 2.2(b) the estimates $\hat{x}_l^n$ are used to give naive estimates $\hat{g}_l^n$ of g based on the analytical form of the x_l from expression (2.116). The error in each of these estimates $\hat{g}_l^n$ is directly related to the error in the estimate $\hat{x}_l^n$. This latter error is shown for each l and n in figure 2.2(a). The influence of n is not important for small values of l. Visually, the $\hat{x}_l^n$ in figure 2.2(a) agree independently of n for $0 \leq l \leq 5$. For $n = 50000$ the $\hat{x}_l^n$ appear to have a regular dependence on l. For $n = 500$ we have an irregular dependence of the $\hat{x}_l^n$ on l, especially for $l \geq 10$. Moreover, for $l \geq 20$ there appear negative values of $\hat{x}_l^n$, clearly inconsistent with the form $x_l = g^l$. These values do not allow the evaluation of corresponding parametric estimates $\hat{g}_l^n$.

For figure 2.3 we have $n = 50000$. For each value of g we simulated n observations of $Y(T)$ and calculated estimates of the Fourier coefficients of p as for figure 2.2(a). Estimated and theoretical Fourier coefficients are respectively represented by empty and filled-in symbols. It is clear from this figure that the nonparametric estimation error is smaller for larger values of g.

In order to understand this behavior we note that g in (2.116) gives the concentration of p near the value $\beta = 0$. Indeed, when $g = 0$ the function p is constant and the rotation random variables X_n are uniformly distributed. In the limit $g \uparrow 1$ we have that each random variable X_n is almost surely equal to the identity matrix. Conditionally on the event $\{N(T) > 0\}$, the distribution of $Y(T)$ is a mixture of distributions with Henyey-Greenstein density. More precisely, for all $m > 0$ we have the conditional probability density for the Euler angle β associated with $Y(T)$

$$p(\beta|N(T) = m) = \frac{1 - g^{2m}}{(1 + g^{2m} - 2g^m \cos\beta)^{\frac{3}{2}}}$$

In particular, in the limit $g \uparrow 1$ we have that $Y(T)$ is almost surely equal to the identity matrix. Conditionally on $\{N(T) > 0\}$, we have in the limit $g \downarrow 0$ that $Y(T)$ is uniformly distributed.

Let us note that in our example $\mathbb{P}(N(T) > 0) \simeq 0.96$. Figure 2.3 can be understood in light of the above discussion. For greater values of g, observations of $Y(T)$ are concentrated near the identity matrix. This leads to fast convergence of our estimates for the Fourier coefficients of p. For smaller

values of g, observations of $Y(T)$ are more dispersed and the convergence of estimates is slower. In the limit $g \downarrow 0$ the observations are close to uniformly distributed and our approach breaks down due to numerical problems.

2.5 Conclusions

Thic chapter has presented the results of two completed papers [63, 66]. We have dealt with direct applications in wave physics, namely in polarization statistics and multiple scattering. Our motivation was to adapt a consistent probabilistic formalism which could include the more geometric aspects of problems arising in these two fields. This was realized by considering rotation random variables and rotation processes, especially rotation Lévy processes. Consequently, applications to wave physics including both modelling and estimation problems were carried out completely. The main estimation problem dealt with in this chapter is the problem of decompounding, Subsection 2.4.2.

Sections 2.1 and 2.3 offer a self-contained introduction on rotation random variables and rotation Lévy processes. Section 2.1 is mainly devoted to the theory of characteristic functions of rotation random variables. While these are already used in many engineering problems, Section 2.1 presents certain differences with the corresponding literature. Paragraph 2.1.2.2, for example, characterizes the symmetry properties of rotation random variables in terms of their characteristic functions. These properties proved highly important to our applications both in Sections 2.2 and 2.4. Subsection 2.1.4 considers the asymptotic properties of products of *i.i.d.* rotation random variables. This is a classical mathematical problem which we have found to have important implications for estimation problems. Section 2.3 presents a basic introduction to rotation Lévy processes. We have sought to use only the tools of Section 2.1 and some elementary tools of Brownian stochastic calculus. As such, our presentation is based on the examples of rotation compound Poisson processes and rotation Brownian motion. However, in Subsection 2.3.4 we briefly describe the interlacing construction which can be used to give general rotation Lévy processes from rotation compound Poisson processes and rotation Brownian motion.

Section 2.2 is based on our paper [63]. It is concerned with a general modelling problem in polarization statistics and uses the mathematical tools of Section 2.1. The aim is to develop a statistical formalism which generalizes the classical Stokes formalism in optics to higher order statistics. This problem has been considered in several works in the optics literature and different solutions have been proposed, depending on target applications. Section 2.2 takes as its starting point the experiments of Ellis and Dogariu [12, 13]. These experiments show the insufficiency of the classical Stokes formalism which is only based on second order statistics. The formalism developed in Section 2.2 is consistent with these experimental results and also compares favorably to other solutions considered in the literature.

Section 2.4 is based on our paper [66]. It solves the problem of decompounding and presents this problem as a statistical alternative to the physical inverse problem of multiple scattering. The problem of decompounding is a nonparametric estimation problem stated in terms of rotation compound Poisson processes. We solve this problem using a characteristic function method, based on Section 2.1. The decompounding problem presents an essential difficulty in comparison to similar nonparametric estimation problems. Indeed, this problem uses indirect observations. As a result, it involves an additional nonlinearity and does not have the usual invariance properties. In order to overcome this difficulty, application of our characteristic function method to the problem of decompounding requires specific tools from matrix analysis as well as a more detailed statistical treatment. We provide a

theoretical proof of the convergence of our method and discuss it using numerical simulations.

Chapter 3

Filtering of matrix processes under nonlinear constraints

This chapter considers linear filtering problems subject to nonlinear constraints. Filtering is here understood in its wider sense. A signal process is to be inferred or approximated from an observation process. Casting linear filtering problems in matrix form is a standard procedure. It will be seen to lead to a consice expression of nonlinear constraints. While we are interested in continuous time filtering for general matrix processes, we start in Section 3.1 by illustrating the general structure of our approach for discrete time rotation processes –this section is based on our paper [64].

Nonlinear constraints are inherent to a significant class of processes with linear dynamics. These constraints express symmetry considerations related to the nature of target applications. For example, signals subject to orthogonality or unitarity constraints include phase signals, signals from wave physics [63, 66], DNA dynamics [58] and motion capture [38]. A major difficulty arising for such constrained signals is that it is impossible to apply linear signal processing operations while respecting nonlinear constraints. In the case of phase signals, this is dealt with using the extended Kalman filter, as in the example of the PLL –phase locking loop. Geometrically, this is a very simple case and the equations of the extended Kalman filter do not seem to be tractable in the case of, for example, rotation signals.

Most existing work applies so-called global linearization methods for dealing with this difficulty. Applications have included filtering and analysis of variance [18, 75]. In [75], Xavier and Manton explain the setbacks of these methods which attempt to map constrained processes pointwise into unconstrained processes or processes with only linear constraints. Our approach is completely different as it is based on local linearization. As opposed to the pointwise transformations involved in global linearization methods, local linearization is essentially a functional transformation.

Our approach is first presented for the case of discrete time rotation signals in Section 3.1. Local linearization is defined in Subsection 3.1.1 based on a dynamical representation of these signals. Section 3.2 is concerned with continuous time Brownian matrix processes. We consider these processes to give the general setting for the definition of our filtering problems. In this section, local linearization is defined in Subsection 3.2.3. As presented in Section 3.1, our approach to the filtering of matrix processes under nonlinear constraints is the following. One starts from the observation process which is subject to nonlinear constraints. Using local linearization this process as a whole is transformed into a process with only linear constraints. This new process can be treated using linear signal processing operations. An inverse transformation is then applied to recover a process respecting the nonlinear

constraints. This process is expected to infer more directly or to approximate the signal process.

The aim of this chapter is to formulate and justify this approach. Local linearization consists in transforming a process from a constrained domain to a linear domain. The main point in the following will be that this is a stable functional transformation. The stability properties of local linearization imply that the approximation realized by filtering in the linear domain is preserved in the constrained domain. Our approach is formulated and studied in a general setting in Section 3.2. This section requires closer knowledge of Brownian stochastic analysis. The importance of such mathematical techniques to signal processing is well-known. See the survey paper [23] by Haykin *et al.* and the many references therein.

3.1 Filtering of rotation time series

The current section is based on our paper [64]. It illustrates our approach based on local linearization in the special case of discrete time rotation processes. We refer to these as rotation time series. In addition to the difficulty related to nonlinear constraints, rotation time series display problems related to noncommutativity. Local linearization deals with such problems as it preserves the order in time of the samples of a rotation time series –see Paragraph 3.1.1.1 below. Lee and shin [38] and Fang *et al.* [16] have proposed methods based on local linearization for real-time filtering and smoothing applications. They evaluate the performance of these methods through examples. In this section, we aim to gain a theoretical understanding of this aspect by showing that the applied local linearization is stable. This is the object of Paragraph 3.1.1.2.

The plan of the current section is the following. Subsection 3.1.1 defines local linearization and obtains its stability properties. Subsection 3.1.2 gives two examples of our approach to the filtering of rotation time series. Subsection 3.1.3 considers the invariance properties of local linearization. In particular, invariance by time shift is used to consider the notion of a filter for rotation time series. Finally Subsection 3.1.4 prepares the following Section 3.2 by considering an example of a filtering problem for continuous time rotation processes.

3.1.1 Local linearization

Rotation time series are considered to be made up of a fixed number $N \geq 1$ of recordings of rotations $O_n \in SO(3)$ where $1 \leq n \leq N$. We note O the N-tuple of the O_n. In Paragraph 3.1.1.1 local linearization of a rotation time series O is defined in terms of a mapping $O \mapsto o$ where o is a time series of N antisymmetric 3×3 matrices o_n where $1 \leq n \leq N$. Paragraph 3.1.1.2 obtains the stability properties of the inverse mapping $o \mapsto O$.

It will be seen that local linearization has a much wider range of applicability than global linearization. In particular, it can always be applied to rotation time series O arising from a continuous process $Y : \mathbb{R}_+ \to SO(3)$. Global linearization imposes strong conditions on the rotation times series O. This is necessary to avoid the problems of distortion and singularities which do not arise with local linearization. This was discussed in [38] and we return to it in Paragraph 3.1.1.1.

3.1.1.1 Dynamics of rotation time series

Let O be a rotation time series. By considering a dynamical representation of O it will be possible to introduce a local linearization mapping $O \mapsto o$ where o is a time series of antisymmetric 3×3

matrices. Such a representation is given by

$$O_{n+1} = O_n \Omega_{n+1} \tag{3.1}$$

for $1 \leq n \leq N-1$, where $\Omega_{n+1} = O_n^T O_{n+1}$. In the language of solid body kinematics, right multiplication by Ω_{n+1} in (3.1) corresponds to a body rather than to a space representation. Consider the N-tuple Ω where $\Omega_1 = O_1$ and Ω_{n+1} is given by (3.1) for $1 \leq n \leq N-1$. It is clear that O and Ω are equivalent. More precisely, for $1 \leq n \leq N$

$$O_n = \prod_{i=1}^{n} \Omega_i = \Omega_1 \dots \Omega_n \tag{3.2}$$

where the product is ordered from left to right. Due to the fact that the matrix group $SO(3)$ is noncommutative, the order in time of the samples O_n of O takes on an additional importance. This is taken into account in (3.2) as a similar order is assigned to the elements of Ω. The construction of Ω as in (3.1) is the first step of the local linearization mapping $O \mapsto o$.

In order to apply local linearization to the time series O we need to assume it satisfies the following mild regularity condition

$$\varphi_n = \arccos[(\mathrm{tr}(\Omega_n) - 1)/2] < \pi \tag{3.3}$$

for $1 \leq n \leq N$, where arccos is taken so that$0 \leq \varphi_n \leq \pi$. The angle φ_n given in (3.3) is the angle of the rotation Ω_n. Condition (3.3) imposes an upper bound on the angular distance between successive samples O_n and O_{n+1} for $1 \leq n \leq N-1$. Under this condition we have that for $1 \leq n \leq N$ the matrix Ω_n has a uniquely defined matrix logarithm $\omega_n = \log(\Omega_n)$ which is a real antisymmetric matrix. We note ω the N-tuple of the ω_n. For $1 \leq n \leq N$ we have –see [1].

$$\omega_n = \frac{\varphi_n}{2 \sin \varphi_n} (O_n - O_n^T) \tag{3.4}$$

when $\varphi_n \neq 0$ and $\omega_n = 0$ when $\varphi_n = 0$. Local linearization consists in the mapping $O \mapsto o$ where o is a time series of antisymmetric matrices o_n for $1 \leq n \leq N$ given by

$$o_n = \sum_{i=1}^{n} \omega_i \tag{3.5}$$

This is similar to (3.2) with the additive structure of antisymmetric matrices replacing the multiplicative structure of $SO(3)$. It is important to note two properties of the mapping $O \mapsto o$. First, this is a mapping of the whole rotation time series O. This is in the sense that for $1 \leq n \leq N$ the matrix o_n depends on $O_1 \dots O_n$. Second, this mapping does realize the first goal of local linearization. Indeed, the rotation time series O is subject to nonlinear constraints while the time series of antisymmetric matrices o is only subject to linear constraints.

Again, the time series o is equivalent to O since we have

$$O_n = \prod_{i=1}^{n} \exp(o_i - o_{i-1}) \tag{3.6}$$

with the convention $o_0 = 0$. The following Paragraph 3.1.1.2 is concerned with the stability properties of the mapping $o \mapsto O$ of (3.6). We do not impose any conditions on o, apart from being a time series of N antisymmetric matrices. Thus, the transformation $o \mapsto O$ is not invertible, since condition (3.3) is not guaranteed.

Condition (3.3) is essential for methods based on local linearization. It should be noted that this conditions is far less restrictive than what is needed to perform global linearization. Let us admit

that the time series O arises from a continuous time rotation process $Y : \mathbb{R}_+ \to SO(3)$. It is clear that if Y is continuous then (3.3) can always be satisfied by sampling Y with a sufficiently high frequency. This remains true of Y has jumps but that these jumps have angles bounded by some $a < \pi$. Global linearization requires that condition (3.3) be satisfied by each O_n. This is a condition referring directly to Y. On the contrary, we have discussed that condition (3.3) merely refers to the way Y is sampled. Generally speaking, local linearization has a much wider range of applicability than global linearization.

Imposing condition (3.3) on each O_n, global linearization attempts to eliminate nonlinear constraints by individually transforming the sample O_n. A time series of antisymmetric matrices $\tilde{o}_n = \log(O_n)$ is proposed as a linearization of O –see for example [18]. The mapping thus obtained suffers from important distortion problems and is moreover not coordinate invariant. In spite of such setbacks global linearization methods remain quite popular in practice [38]. This is in part due to their simplicity.

3.1.1.2 Stability of local linearization

The current paragraph is concerned with the stability properties of the mapping $o \mapsto O$ defined in (3.6). The mapping $O \mapsto o$ of (3.5) obtains a time series o subject only to linear constraints from the rotation time series O. Linear signal processing operations can be applied to o. It must be ensured that the approximation realized by such operations in the linear domain is preserved by the mapping $o \mapsto O$. The following discussion will ensure that this is indeed the case. However, the mapping $o \mapsto O$ seems to suffer from "drifting" problems due to propagation of errors in formula (3.6). This was pointed out in [38].

Note that the mapping $o \mapsto O$ can be calculated as $o \mapsto \omega \mapsto \Omega \mapsto O$. The first step $o \mapsto \omega$ is an invertible linear transformation

$$\omega_n = o_n - o_{n-1} \tag{3.7}$$

for $1 \leq n \leq N$. The only effect of this step on stability is a scale factor. We consider the two following steps in the following Propositions 24 and 25.

Note that $\omega \mapsto \Omega$ is calculated elementwise, $\Omega_n = \exp(\omega_n)$ for $1 \leq n \leq N$. The stability of this mapping is directly given by that of the matrix exponential. Let ω and ω' be N-tuples of antisymmetric matrices noted ω_n and ω'_n. For $1 \leq n \leq N$ let $\Omega_n = \exp(\omega_n)$ and $\Omega'_n = \exp(\omega'_n)$. We have the following proposition.

Proposition 24 *For $1 \leq n \leq N$ the following order of magnitude holds*

$$\Omega'_n - \Omega_n = \Omega_n \int_0^1 \Omega_n^{-s}(\omega'_n - \omega_n)\Omega_n^s ds + o(|\omega'_n - \omega_n|) \tag{3.8}$$

where $|.|$ is the Euclidean matrix norm and matrix exponents are understood as in $\Omega_s = \exp(s\omega_n)$ for $-1 \leq s \leq 1$.

Proof: Let us fix $1 \leq n \leq N$. The following reasoning is independent of n. Consider the function D defined by

$$D(t) = \exp(t\omega'_n) - \exp(t\omega_n)$$

for $0 \leq t \leq 1$. By direct calculation, we have that for $0 \leq t \leq 1$

$$\frac{d}{dt}D(t) = \omega_n D(t) + (\omega'_n - \omega_n)\exp(t\omega'_n) \tag{3.9}$$

This is a first order linear ordinary differential equation for D. It can be solved by variation of constants which leads to the following formula

$$D(t) = \int_0^t \exp[(t-s)\omega_n](\omega'_n - \omega_n)\exp(s\omega'_n)ds \tag{3.10}$$

Putting $t = 1$ we have with our notation for matrix exponents

$$\Omega'_n - \Omega_n = \Omega_n \int_0^1 \Omega_n^{-s}(\omega'_n - \omega_n)\Omega_n'^s ds \tag{3.11}$$

In order to prove (3.8) we must estimate the following difference

$$\Omega'_n - \Omega_n - \Omega_n \int_0^1 \Omega_n^{-s}(\omega'_n - \omega_n)\Omega_n^s ds$$

Replacing $\Omega'_n - \Omega_n$ from (3.11) we have that this is equal to

$$\Omega_n \int_0^1 \int_0^s \Omega_n^{-s}(\omega'_n - \omega_n)\Omega_n^{s-r}(\omega'_n - \omega_n)\Omega_n'^r dr ds$$

which is clearly of second order in $\omega'_n - \omega_n$. The order of magnitude (3.8) can now be checked using usual norm inequalities.▲

The last step in the mapping $o \mapsto O$ is $\Omega \mapsto O$. This is not an elementwise transformation as each O_n depends on Ω_i for $1 \leq i \leq n$ –see (3.2). Consequently, this step plays a more important role in drifting problems. Retaining the notation of Proposition 24, note O' the time series obtained from the N-tuple Ω' as in (3.2). We have the following proposition.

Proposition 25 *For $1 \leq n \leq N$ the following order of magnitude holds*

$$O'_n - O_n = \sum_{i=1}^{n} \prod_{1 \leq j < i} \Omega_j(\Omega'_i - \Omega_i) \prod_{i < j \leq n} \Omega_j + o(|\Omega' - \Omega|) \tag{3.12}$$

where $|\Omega' - \Omega| = \max_n |\Omega'_n - \Omega_n|$ and the products are ordered from left to right.

Proof: For all $1 \leq n \leq N$ we have the following identity

$$O'_n - O_n = \sum_{i=1}^{n} \prod_{1 \leq j < i} \Omega'_j(\Omega'_i - \Omega_i) \prod_{i < j \leq n} \Omega_j \tag{3.13}$$

In order to prove (3.13), note that the following sum is telescopic and factorize $(\Omega'_i - \Omega_i)$ in each term of the sum

$$O'_n - O_n = \sum_{i=1}^{n} \prod_{1 \leq j \leq i} \Omega'_j \prod_{i < j \leq n} \Omega_j - \prod_{1 \leq j < i} \Omega'_j \prod_{i \leq j \leq n} \Omega_j \tag{3.14}$$

Subtracting (3.12) from (3.13) it is possible to estimate the difference

$$O'_n - O_n - \sum_{i=1}^{n} \prod_{1 \leq j < i} \Omega_j(\Omega'_i - \Omega_i) \prod_{i < j \leq n} \Omega_j$$

This is equal to

$$\sum_{i=1}^{n} \sum_{j=1}^{i-1} \prod_{k=1}^{j-1} \Omega'_k(\Omega'_j - \Omega_j) \prod_{k=j+1}^{i-1} \Omega_k(\Omega'_i - \Omega_i) \prod_{j=i+1}^{n} \Omega_j$$

which is of second order in each $\Omega'_i - \Omega_i$. The proof can be completed as for the last proposition.▲

Relations (3.8) and (3.12) respectively imply that the mappings $\omega \mapsto \Omega$ and $\Omega \mapsto O$ are differentiable. Furthermore, it can be inferred from these relations that these transformation are C^1. Accepting this, we have by composition of C^1 transformations that $o \mapsto O$ is C^1. Applying usual norm inequalities to (3.10) and (3.13) it is possible to obtain the following Lipschitz inequality

$$|O' - O| \leq 3N|o' - o| \tag{3.15}$$

The bound established by this inequality is a gross estimation, especially for larger values of N. We will see in the following Subsection 3.1.2 that it can be considerably improved for certain specific examples.

Condition (3.3) is equivalent to the condition that

$$(1/\sqrt{2})|o_n - o_{n-1}| = (1/\sqrt{2})|\omega_n| < \pi \tag{3.16}$$

for $1 \leq n \leq N$. Under this condition, we have the following result which will be considered again in Subsection 3.1.2. The derivative of the transformation $o \mapsto O$ has full rank –*i.e.* it is an invertible linear transformation– at every o such that (3.3) is verified. We do not prove this result here. It follows from the two following assertions. First, that under (3.16) the derivative of $\omega \mapsto \Omega$ has full rank. This follows from a classical result on the Riemannian geometry of the three sphere $S^3 \subset \mathbb{R}^4$ –see [10]. Second, that the derivative of $\Omega \mapsto O$ has full rank at every Ω. This can be proved directly from (3.12).

3.1.2 Applications

The current subsection gives two examples of our approach to the filtering of rotation time series. A discrete time variant of the so-called problem of geodesic regression is presented in Paragraph 3.1.2.1. Paragraph 3.1.2.2 treats a similar problem with a different noise model. Geodesic regression is central to *principal geodesic analysis* which generalizes the concepts of analysis of variance to rotation data [18, 65]. In relation to principal geodesic analysis, geodesic regression was considered as an optimization problem in our paper [65]. In Paragraph 3.1.2.1, this is discussed in comparison to our current approach based on local linearization.

Paragraph 3.1.2.2 does not use the term *filtering* in the wide sense suggested in the introduction. More precisely, this paragraph refers to filtering in the sense of a causal operation supposing only real-time observations. This is in opposition with the *smoothing* situation of Paragraph 3.1.2.1 which requires full knowledge of the observed time series. The examples of Paragraphs 3.1.2.1 and 3.1.2.2 thus differ by the considered noise models as well as by the nature of observations. Clearly, these two examples are directed towards substantially different applications.

In Paragraphs 3.1.2.1 and 3.1.2.2 a vocabulary is used which is specific to data analysis problems involving orientation and rotation data. See [17, 18]. Consider a differentiable curve $Y : \mathbb{R}_+ \to SO(3)$. For $t \geq 0$ let

$$\omega(t) = Y^T(t)\frac{d}{dt}Y(t) \tag{3.17}$$

Let us remind that $\omega(t)$ is called the body angular velocity corresponding to the rotational motion of a solid body as described by the curve Y. It is straightforward to show that $\omega(t)$ is an antisymmetric matrix. This follows by differentiating the condition $Y^T(t)Y(t) = I$, which holds for $t \geq 0$. We will say that Y is a geodesic if for some antisymmetric matrix ω

$$Y(t) = \exp(\omega t) \tag{3.18}$$

for $t \geq 0$. From the point of view of kinematics, (3.18) is a uniform rotation with angular velocity ω. This is called a geodesic since the curve $Y : \mathbb{R}_+ \to SO(3)$ given in this way is a geodesic for the biinvariant Riemannian metric of $SO(3)$. This vocabulary is more common in the literature. It is adapted to geometric approaches to the analysis of orientation and rotation data, often based on optimization.

The noncommutativity of matrix multiplication is expressed in the case of the matrix group $SO(3)$ by the Baker-Campbell-Hausdorff formula. This will be used in Paragraph 3.1.2.1 to evaluate the error from our approach based on local linearization.

$$\exp(\omega_1)\exp(\omega_2) = \exp\left(\omega_1 + \omega_2 + \frac{1}{2}[\omega_1, \omega_2] + O(|\omega_1|^3 + |\omega_2|^3)\right) \tag{3.19}$$

for all antisymmetric matrices ω_1 and ω_2 and where $O(|\omega_1|^3 + |\omega_2|^3)$ is also antisymmetric. The notation $[\omega_1, \omega_2]$ stands for the commutator of ω_1 and ω_2. That is, $[\omega_1, \omega_2] = \omega_1\omega_2 - \omega_2\omega_1$. Note that when $[\omega_1, \omega_2] = 0$ we have

$$\exp(\omega_1)\exp(\omega_2) = \exp(\omega_1 + \omega_2)$$

as can easily be verified from the fact that $[\omega_1, \omega_2] = 0$ *iff* $\omega_2 = k\omega_1$ for some $k \in \mathbb{R}$.

For the noise model used in Paragraph 3.1.2.1 we have chosen a Gaussian law on $SO(3)$. We say that a rotation random variable N has Gaussian law if there exists a Gaussian vector $B \in \mathbb{R}^3$ of coordinates $[B^1, B^2, B^3]$ such that [17, 54]

$$N = \exp[B^1 J_1 + B^2 J_2 + B^3 J_3] \tag{3.20}$$

where the matrices J_1, J_2, J_3 were given in (2.7) of Paragraph 2.1.1.1. This approximates a law arising from rotation Brownian motion –compare to Paragraph 2.3.3.1– and can be dealt with directly using the Baker-Campbell-Hausdorff formula.

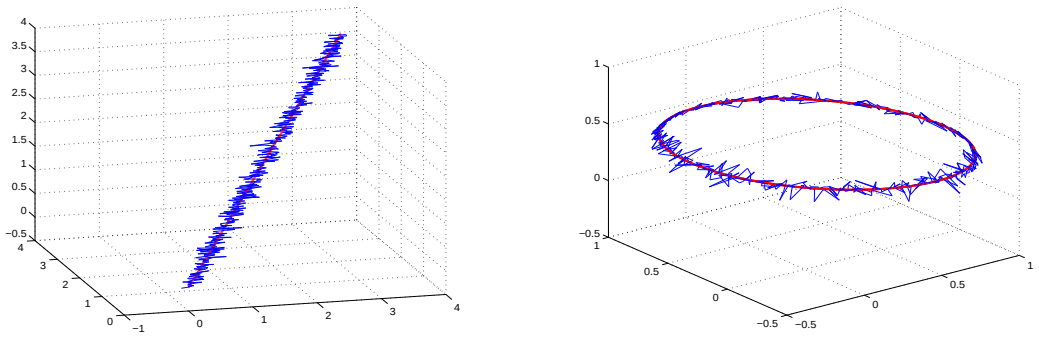
3.1.2.1 Geodesic regression

Here a smoothing problem is addressed. This problem is a discrete time version of the problem of geodesic regression. See [49]. The problem is stated in terms of an observed rotation time series O^* for which there is *a priori* knowledge that O^* is a noisy geodesic. More precisely, it is known that there exists an antisymmetric matrix ω verifying (3.16) such that for all $1 \leq n \leq N$

$$O^*_n = \exp(n\omega)N_n \tag{3.21}$$

where the N_n are independent observations of a Gaussian law on the rotation groups $SO(3)$. See formula (3.20). The time series N models the noise involved in the problem. A multiplicative noise model is considered. The choice of multiplicative noise on the right is an arbitrary one in the current context. The goal here is to estimate ω from an observation of the whole time series O^*. Geodesic regression is the core ingredient in data analysis methods such as principal geodesic analysis. See [18]. Principal geodesic analysis is usually applied using global linearization methods, in which case only approximative results are obtained. Exact principal geodesic analysis has been attempted in our paper [65]. It has a high computational cost. It is possible that local linearization can constitute a good compromise between the two.

In order to follow the local linearization methods prescribed in the introduction, one must start by applying local linearization to O^*. This is only feasible if O^* verifies (3.3). This is indeed the case, with probability 1, since for $1 \leq n \leq N$ the random variable O^*_n has an absolutely continuous



(a) Local linearization of a noisy geodesic

(b) Trajectory of a point by a noisy geodesic

Figure 3.1: Local linearization for geodesic regression

distribution, with respect to the Haar measure of $SO(3)$. This follows from the absolute continuity of the Gaussian law.

Figures 3.1(a) and 3.1(b) illustrate the result of applying local linearization to O^*. In order to obtain O^*, the noise N was generated according to [54]. The blue curve in figure 3.1(a) gives the result of local linearization of O^*. For figure 3.1(a), antisymmetric matrices are identified with vectors in $\mathbb{R}^3$.

$$A \leftrightarrow \hat{A} = [a_1, a_2, a_3] \text{ where } A = a_1 J_1 + a_2 J_2 + a_3 J_3 \tag{3.22}$$

for all antisymmetric matrix A. The vector $\hat{A}$ of coordinates $[a_1, a_2, a_3]$ is directed along the axis of rotation and its length is equal to the angle of rotation of $\exp(A) \in SO(3)$. The red line in figure 3.1(a) gives the vector $\hat{\omega}$ in $\mathbb{R}^3$ corresponding to ω as in (3.22). In figure 3.1(b) the blue and red curves represent the trajectory of a point of the sphere $S^2 \subset \mathbb{R}^3$ under the effect of the successive rotations O_n^* and $\exp(n\omega)$ for $1 \leq n \leq N$, respectively. These trajectories do not completely determine the corresponding rotation time series but they constitute a simple means of visualization. Figure 3.1(a) shows that by applying local linearization to the noisy geodesic O^*, a noisy version of the "straight line" $n\hat{\omega}$ for $1 \leq n \leq N$ is obtained. A simple linear regression then allows the recovery of ω, thus solving the geodesic regression problem. Our goal is to understand this result.

Note Ω^* the rotation time series where $\Omega_1^* = O_1^*$ and for $1 \leq n \leq N-1$ we have $\Omega_{n+1} = {O_n^*}^{-1} O_{n+1}^*$. According the Baker-Campbell-Hausdorff formula (3.19)

$$\begin{aligned} \log(\Omega_1^*) &= \omega + \nu_1 + O(|\nu_1|) \\ \log(\Omega_n^*) &= \omega + \nu_n + O(|\nu_n|) \end{aligned} \tag{3.23}$$

where $\nu_1 = \log(N_1)$ and for $1 \leq n \leq N-1$ we have $\nu_{n+1} = \log(N_n^{-1} N_{n+1})$. The reminder $O(|\nu_n|)$ for $1 \leq n \leq N$ in (3.23) can be estimated more precisely. We have $|O(|\nu_n|)| \leq \pi|\nu_n| + o(|\nu_n|^2)$ for all $1 \leq n \leq N$.

It follows that by applying local linearization to O^* a time series of antisymmetric matrices o^* is obtained that is given by

$$o_n^* = n\omega + \sum_{i=1}^{n} \nu_i + O(|\nu|) \tag{3.24}$$

for $1 \leq n \leq N$. The error term $O(|\nu|)$ in formula (3.24) is small with a high probability. Indeed, the samples of a Gaussian law are near the identity I with a high probability and their logarithms should be small, in the same way. Under this result, formula (3.24) explains what is obtained experimentally

in figure (3.1(a)). Namely, that applying local linearization to the noisy geodesic O^*, a noisy version of the time series with samples $n\omega$ is obtained.

In order to retrieve ω by linear regression, one calculates the estimate ω^* given by

$$\omega_n^* = \frac{\sum_{n=1}^{N} n o_n^*}{\sum_{n=1}^{N} n^2} \tag{3.25}$$

This estimate is optimal in a mean square sense. It incurres an error given by the following formula

$$|\omega - \omega_n^*| = \left| \frac{\sum_{n=1}^{N} n \left(\sum_{i=1}^{n} \nu_i + O(|\nu|)\right)}{\sum_{n=1}^{N} n^2} \right| \tag{3.26}$$

from which it follows

$$|\omega - \omega_n^*| \leq (1+\pi)|\nu| + o(|\nu_n|^2) \tag{3.27}$$

If one uses ω^* to estimate a geodesic $\exp(n\omega^*)$, then by formula (3.10)

$$|\exp(n\omega) - \exp(n\omega^*)| \leq 3(1+\pi)|\nu| + o(|\nu_n|^2) \tag{3.28}$$

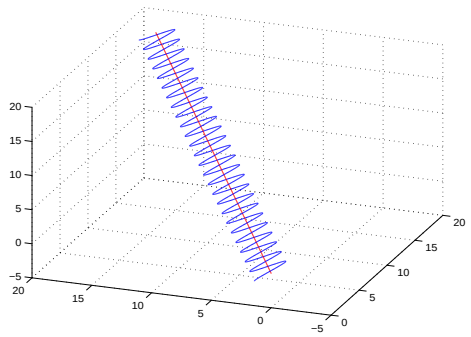
for all $1 \leq n \leq N$. This bound gives an order of greatness of the error made in using local linearization to solve a geodesic regression problem. In the approximation considered here, it is possible to assume that ν has a Gaussian distribution of a certain variance σ^2 corresponding to the dispersion of the samples of N away from I on the group manifold of $SO(3)$. Up to third order terms, the root mean square error from using local linearization is thus roughly $4\pi\sigma$.

A synthesis of the approach of the current subsection is now possible. Methods based on local linearization have been advanced in the literature due to their successful application. These methods are based on the *prescription* of applying local linearization in order to retrieve a linear time series which can be processes in a classical way. This has a clear practical advantage but lacks general foundation. It is not immediately visible that these methods are optimal, even for a limited range of problems, in a usual sense. In the example studied here, it is possible to conclude that local linearization yields satisfactory results thanks to **i)** The statistics of the multiplicative Gaussian noise. In particular that its distribution is concentrated near I. **ii)** The stability of the transformation of equation (3.6). This implies that the error made in the linear domain is propagated to the non linear domain with a reasonabe linear scale factor, inequality (3.28). By contrast to the use of local linearization, it is possible to consider a least square solution to geodesic regression using the derivative of the exponential map as obtained in formula (3.8), noting that this derivative has full rank under condition (3.16) –see Paragraph 3.1.1.2. This optimization-based approach leads to a nonlinear matrix equation. The numerical solution of such an optimal approach is possibly less stable than the application of the suboptimal approach of local linearization.

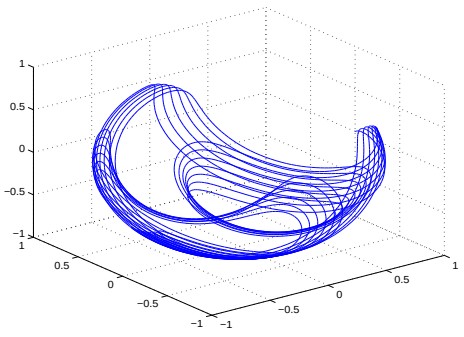
3.1.2.2 Deterministic filtering

The problem considered here is similar to that of Paragraph 3.1.2.1. In both cases, the signal we seek to recover is a geodesic. However, here the problem is considered as a filtering problem, as opposed to the smoothing situation proposed in the last subsection. In other words, the solution proposed here is causal and is compatible with a real-time situation. This notion of causality and the definition of filters for rotation time series are considered in more detail in Subsection 3.1.3. The observed rotation time series is O^* given by

$$O_n^* = \prod_{i=1}^{n} \exp\left(\omega + [\sin(2\pi f n) - \sin(2\pi f(n-1))]\omega'\right) \tag{3.29}$$



(a) Result of local linearization



(b) Trajectory of a point under additive trigonometric noise

Figure 3.2: Additive trigonometric noise

ω and ω' being antisymmetric matrices and f a frequency. Our goal is to recover ω, so that the presence of the trigonometric term in the exponential factors in (3.29) corresponds to noise. This is a different noise model than the one of the last subsection, where the noise affected the observations multiplicatively. Applying local linearization to O^* one obtains a time series o^* given by

$$o^*_n = n\omega + \sin(2\pi f n)\omega' \tag{3.30}$$

from which ω can be recovered after derivation and low pass filtering, both operations expressible in terms of convolution masks –see Section 3.1.3. An example of the current problem is given in figure 3.2. Figure 3.2(a) gives the result of local linearization of O^*, using the same convention as for figure 3.1(a). ω is given by the red line in this figure. Figure 3.2(b) gives the trajectory of a point of S^2 under the rotations O^*_n.

In the current subsection, additive noise affected ω directly, in the linear domain. This is different from the multiplicative noise of the last subsection, which was directly related to the observations. When additive noise appears in the linear domain, local linearization methods have higher performances than in other cases. In fact, they can be –theoretically– exact. This is the case in equation (3.30) where it is possible, in principle, to recover ω exactly. The noise model of the current subsection, that of additive noise in the linear domain, is a realistic one for applications dealing directly with rotations from mechanical motions. In these applications, what is related to fluctuating outside conditions is the "angular velocity" time series o in the linear domain and not the actual motion O. Remarkably, figure 3.2 shows that local linearization of a seemingly complicated time series can yield a simple representation. This shows the potential of local linearization as a method for motion analysis.

3.1.3 Invariance properties

Certain important properties of local linearization can be grouped together under the general name of invariance properties. We here give a general formulation of such properties and consider briefly invariance by left multiplication and by time shift. Using the properties of invariance by time shift, we consider the notion of a filter for rotation time series.

Let T be a mapping of the set of rotation time series such that for all time series O the time series $T(O)$ satisfies condition (3.3). T then defines a mapping t of the set of time series of antisymmetric matrices. Indeed, for all time series o of antisymmetric matrices, let O correspond to o according to

(3.6). Note $t(o)$ the time series of antisymmetric matrices obtained by applying local linearization to $T(O)$. In particular, $t(o)$ verifies (3.16). This defines the required mapping. The relation between T and t is expressed in the following diagram. Upward arrows denote the application of equation (3.6) and downward arrows denote the application of equation (3.5).

$$\begin{array}{ccc} O & \stackrel{T}{\longrightarrow} & T(O) \\ \uparrow & & \downarrow \\ o & \stackrel{t}{\longrightarrow} & t(o) \end{array} \tag{3.31}$$

Invariance refers to a set of cases where a simple relationship is obtained between T and t. We now present invariance by multiplication and time shift. Other cases are considered in [38].

For invariance by multiplication, consider for the mapping T the following transformation. $K \in SO(3)$ is fixed and for all rotation time series O its image $T(O)$ is defined by $T(O)_n = KO_n$ for $1 \leq n \leq N$. It is necessary to limit the domain of T to the set of time series O such that KO_1 verifies (3.3). Note that for all $1 \leq n \leq N-1$ we have

$$T(O)_n^{-1} T(O)_{n+1} = O_n^{-1} K^{-1} K O_{n+1} = \Omega_{n+1} \tag{3.32}$$

Note o the time series obtained from O by local linearization. It follows from equation (3.5) that local linearization of $T(O)$ gives the time series $t(o)$ defined as follows

$$\begin{aligned} t(o)_1 &= \log(KO_1) \\ t(o)_n &= (t(o)_1 - o_1) + o_n \end{aligned} \tag{3.33}$$

In other words, o and $t(o)$ differ by an initial condition. In the same way, (3.32) shows that the dynamics of O is the same as that of $T(O)$. It is important here to emphasize that (3.32) and (3.33) were obtained in the case where T is defined by left and not right multiplication by K. In particular (3.32) obtains because the product in (3.1) is ordered from left to right. Formula (3.33) tells us that local linearization is essentially independent from any choice of initial reference base. This is different from global linearization which always refers to a choice of $K \in SO(3)$ to be used as reference base for projecting data onto some vector space.

Invariance by time shift is an important property in relation to signal processing. It indicates that local linearization is a causal operation. This means that it is compatible with filtering and other real-time applications. See [16]. Loosely speaking, what is to be expressed by this property is that for all $1 \leq n \leq N$ the antisymmetric matrix o_n obtained in (3.5) depends only on the samples O_i with $1 \leq i \leq n$. This follows immediately if one notes that for all $1 < n \leq N$ the rotation Ω_n depends only on O_{n-1} and O_n, while $\Omega_1 = O_1$.

Consider the mapping T where for every rotation time series O verifying (3.3) we have $T(O)_n = O_{n-\tau}$ where $1 \leq \tau$ and where the convention is made that $O_m = I$ for all $m \leq 0$. The mapping t corresponding to T in this case is defined in the same way. For all o verifying (3.16), $t(o)_n = o_{n-\tau}$, with the convention that $o_m = 0$ for all $m \leq 0$. Both T and t are time shift (delay) operators. In other words, the operation of local linearization commutes with time shift.

Let f denote the operation of applying a convolution mask to o. We write $f(o)$ for the time series

$$f(o)_n = \sum_{i=0}^{n} a_i o_{n-i} \tag{3.34}$$

where $a_i \in \mathbb{R}$ for $0 \leq i \leq N$. It is a celebrated fact in signal processing that

$$f(t(o)) = t(f(o)) \tag{3.35}$$

whenever t is a time shift as above. Define the operation F for rotation time series according to the following diagram, which runs in the opposite direction to diagram (3.31).

$$\begin{array}{ccc} O & \xrightarrow{F} & F(O) \\ \downarrow & & \uparrow \\ o & \xrightarrow{f} & f(o) \end{array} \tag{3.36}$$

By following the arrows of diagrams (3.31) and (3.36), it is possible to prove that

$$F(T(O)) = T(F(O)) \tag{3.37}$$

whenever T is a time shift as above. The operation F thus has the same property of invariance by time shift as the convolution mask f. Diagram (3.36) was applied in Paragraph 3.1.2.2, where the operation F was thought of as a filter for rotation time series. This is an adequate terminology if one considers the defining property of a filter to be invariance by time shift. However, there is a difference between operations f and F in that F is more liable to drifting problems.

Of an interest similar to that of multiplication invariance is coordinate invariance. This is a property of a different kind than the two properties we have presented here. What it means is that local linearization is independent of any choice of coordinate system on $SO(3)$ used to express the samples of rotation time series. Certain approaches to the processing of rotations consist in processing their representation in terms of some coordinates, *e.g.* Euler angles. These approaches suffer from distortion problems and singularities. See [68].

3.1.4 Relation to continuous time processes

There exists a close correspondence between the discrete time formalism of the current section and the general continuous time formalism of the following section. Through this correspondence, it will be possible in the following section to formulate our approach based on local linearization for the general setting of continuous time Brownian matrix processes. As a typical example and in order to prepare the following section, we here study a continuous time version of the problem of geodesic regression of Paragraph 3.1.2.1. For the current subsection see [29]. More detailed references are given in the following section. We are now concerned with the two following differential equations

$$dY_t^i = Y_t^i dX_t^i \qquad Y_0^i = I \tag{3.38}$$

Here $i = 1, 2$ and the unknown processes Y^1 and Y^2 are 3×3 matrix processes. The processes X^1 and X^2 are given by

$$X_t^1 = \omega t \qquad X_t^2 = J_t + X_t^1 \tag{3.39}$$

for $t \geq 0$. Where ω is an antisymmetric matrix and J is a Brownian process $J_t = B_t^1 J_1 + B_t^2 J_2 + B_t^3 J_3$ for $[B^1, B^2, B^3]$ the coordinates of a $\mathbb{R}^3$-valued Brownian motion B. For $i = 1$, equation (3.38) is a first order ordinary linear differential equation. Its solution is $Y_t^1 = \exp(\omega t)$ for $t \geq 0$. For $i = 2$, equation (3.38) is a linear stochastic differential equation driven by the diffusion process X^2. This is understood as a Stratonovich equation, so that when $\omega = 0$ we have the equation of rotation Brownian motion, equation (2.94) of Subsection 2.3.3. In general, the solution of (3.38) for $i = 2$ is called rotation Brownian motion with drift. The equivalent Itô equation is the following

$$dY_t^2 = Y_t^2 d\tilde{X}_t^2 \qquad \tilde{X}_t^2 = J_t + \left(X_t^1 + \frac{\mathbb{E}[J_t^2]}{2} \right) \tag{3.40}$$

which is again linear. It follows that equation (3.40) has a unique solution which moreover has almost surely continuous paths defined for all $t \geq 0$. We note this solution $Y^2 = (Y_t^2)_{t\geq 0}$. The problem of geodesic regression can be posed in continuous time as follows. For some fixed $T \geq 0$, a path of the process Y^2 is observed over the interval $0 \leq t \leq T$. From this continuous time set of observations, the aim is to estimate ω or equivalently to recover the process Y^1. As mentioned in Subsection 3.1.2, Y^1 is referred to as a geodesic on $SO(3)$.

The process Y^2 is a rotation process. As for rotation Brownian motion in Paragraph 2.3.3.1, the integration by parts rule of Stratonovich calculus can be used to show that for all $t \geq 0$ we have $Y_t^2 \in SO(3)$ almost surely. In general, this property is not preserved by linear operations. The simplest example is that of the average $\mathbb{E}(Y_t^2)$ for $t \geq 0$. Calculating this average, when it is possible to do so, can be seen as a rudimentary way of obtaining a deterministic process from Y^2. This is a pertinent estimate of the process Y^1 in as much as Y^1 is itself deterministic. Suppose for simplicity that the covariance matrix C of B is of the form $a^2 I$ with $a \in \mathbb{R}$. In this case for $t \geq 0$

$$\tilde{X}_t^2 = J_t + \left(\omega - a^2 I\right) t$$

For $t \geq 0$, the average $\mathbb{E}(Y_t^2)$ can be obtained by discarding the term J_t from the expression of $\tilde{X}_t^2$ in equation (3.40). Indeed, this Brownian term has no influence on the average of Y_t^2. We have

$$\mathbb{E}(Y_t^2) = \exp\left[\left(\omega - a^2 I\right) t\right] \tag{3.41}$$

Clearly, if $t > 0$ then $\mathbb{E}(Y_t^2) \notin SO(3)$ unless $a = 0$. Noting moreover that the dynamics of Y^2 is not stationary we have that averaging of Y_t^2 is not directly applicable for our problem of geodesic regression. We will see in Subsection 3.2.2 how to overcome this difficulty in order to take advantage of the special analytical form of (3.41).

In order to follow the approach already applied for the two examples of Subsection 3.1.2, it is necessary to generalize the local linearization transformations (3.5) and (3.6) of Subsection 3.1.1 to the context of continuous time processes. This is done using the pair of approximations (3.42) below. Let Y denote either one of the processes Y^1 and Y^2. In the same way, let X denote X^1 or X^2. For $n \geq 0$ define processes $Y^n = (Y_t^n)_{t\geq 0}$ and $X^n = (X_t^n)_{t\geq 0}$ as follows[1]

$$X_t^n = \sum_{m\geq 0} \log\left(Y_{t\wedge t_m^n}^{-1} Y_{t\wedge t_{m+1}^n}\right) \qquad Y_t^n = \prod_{m\geq 0} \exp\left(X_{t\wedge t_{m+1}^n} - X_{t\wedge t_m^n}\right) \tag{3.42}$$

where $t_m^n = m/2^n$ for $n, m \geq 0$. We have that the X^n converge to X and the Y^n to Y locally uniformly in the square mean. Starting from the observed process Y^2 it is possible to use (3.42) to approximate X^2. This satisfies the general aim of local linearization. Indeed, while Y^2 is constrained to $SO(3)$ the process X^2 is an antisymmetric matrix process subject to no nonlinear constraints. The approximations (3.42) are similar in their structure to the transformations (3.5) and (3.6).

The approximation result for the Y^n will be given in Subsection 3.2.1 and the reciprocal result for the X^n in Subsection 3.2.3. For $n \geq 0$, the processes X^n can be obtained pathwise from Y. The same is true for the Y^n and X as it is clear in (3.42). It follows that (3.42) can be applied to perform local linearization for the continuous time geodesic regression problem studied here. Indeed, this problem assumes given a path of Y^2 over the interval $0 \leq t \leq T$ and this is sufficient for computation of the processes X^n. These processes approximate X^2 from which ω can be obtained in an elementary way.

[1]We use the notation $a \wedge b = \min\{a, b\}$. Accordingly, in (3.42) we have that for $t \geq 0$ sums and products run only over points $t_m^n < t$ and are in particular finite. This is a concise way of writing integrals and also multiplicative integrals such as the ones considered for rotation Brownian motion in Paragraph 2.3.3.1.

As for the discrete time setting considered in the current section, application of an approach based on local linearization is justified by the invariance and stability properties of the transformations (3.42). In the following section, invariance by time shift will be formulated in terms of a Markov property of Brownian matrix processes, see Subsection 3.2.4. Also in the general case of Brownian matrix processes, Subsection 3.2.5 will consider the stability of the transformation $X \mapsto Y$.

It is possible to study the stability of the transformation $X \mapsto Y$ using integration by parts. This point of view was developed in general in [29, 30]. It is moreover readily generalizable to processes with values in any Lie group. Unfortunately, it falls outside the scope of Section 3.2 and will not be developed in Subsection 3.2.5. We here try to indicate it using a brief example. Consider the deviation between Y^2 and Y^1. By the integration by parts rule of Stratonovich calculus we have for $t \geq 0$

$$d\mathcal{E}_t = Y_t^2 dJ_t {Y_t^1}^T \qquad \mathcal{E}_0 = I \tag{3.43}$$

where $\mathcal{E} = (\mathcal{E}_t)_{t \geq 0}$ is defined by $\mathcal{E}_t = Y_t^2 {Y_t^1}^T$ and we have used the fact that ${Y^1}_t^T = \exp(-\omega t)$. The integral form of (3.43) is the following

$$\mathcal{E}_t = I + \int_0^t Y_s^2 dJ_s {Y_s^1}^T \tag{3.44}$$

This can be used to estimate the deviation between Y^2 and Y^1 in terms of the difference J between X^2 and X^1. The angular distance between Y_t^2 and Y_t^1 is immediately given by $\mathcal{E}_t$. Indeed, This angular distance is the following quantity

$$\arccos\left[(\mathrm{tr}(\mathcal{E}_t) - 1)/2\right] \tag{3.45}$$

which can also be used to measure the deviation between Y^2 and Y^1.

3.2 Filtering of Brownian matrix processes

Here we formulate in general the approach based on local linearization outlined in the last section. This is done for Brownian matrix processes. By gaining a closer understanding of the dynamics of these processes it will be possible to deal with them correctly in statistical estimation problems and in particular to study the performances of our approach. Given $d \geq 1$, we will note in this section $\mathcal{L}(d)$ the space of $d \times d$ real matrices and $\mathcal{GL}(d) \subset \mathcal{L}(d)$ the subset of invertible matrices. We suppose given a complete probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

We have already met two examples of Brownian matrix processes. More precisely, Subsection 2.3.3 studied rotation Brownian motion and Subsection 3.1.4 considered rotation Brownian motion with drift. Both processes were given by Stratonovich linear stochastic differential equations of the form

$$dY_t = Y_t dX_t \qquad Y_0 = I \tag{3.46}$$

where Y is a rotation process and X is a driving process of the form $X_t = X_t^1 J_1 + X_t^2 J_2 + X_t^3 J_3$ for $t \geq 0$ where $[X^1, X^2, X^3]$ are the coordinates of a $\mathbb{R}^3$-valued diffusion process. In this section the driving process X is more generally a $\mathcal{L}(d)$-valued diffusion process for some $d \geq 1$. For $t \geq 0$, the matrix elements of X_t are as follows. The stochastic integral here is an Itô integral.

$$X_t^{ij} = f_{ij}(t) + \int_0^t Q_{ij}(s) dB_s^{ij} \tag{3.47}$$

for $1 \leq i, j \leq d$, where f_{ij} and Q_{ij} are functions $\mathbb{R}_+ \to \mathbb{R}$. The processes B^{ij} are the matrix elements of a $\mathcal{L}(d)$-valued Brownian motion process. For $1 \leq i, j \leq d$, we consider that f_{ij} is continuous and of

finite variation, *e.g.* a C^1 function, and that Q_{ij} is continuous. Under these conditions, it is simply seen that the following Stratonovich equation is well-defined

$$Y_t = I_d + \int_0^t Y_s dX_s \qquad Y_t^{ij} = \delta_{ij} + \sum_{k=1}^{d} \int_0^t Y_s^{ik} dX_s^{kj} \tag{3.48}$$

where I_d is the $d \times d$ identity matrix. We have written equation (3.48) both in matrix form and for each matrix element Y^{ij} where $1 \leq i, j \leq d$. This equation is an integral form of equation (3.46) and only differs from it by the fact that we now consider general matrix processes rather than rotation processes. In Subsection 3.2.3 equations of the form (3.48) will be identified as a general model for Brownian matrix processes. This result is based on a general form of the pair of approximations (3.42) of Subsection 3.1.4.

The solution of (3.48) for a given process X will be noted $\mathcal{E}(X) = (\mathcal{E}(X)_t)_{t\geq 0}$. Subsection 3.2.1 below will show that $\mathcal{E}(X)$ is subject to nonlinear constraints. However, these nonlinear constraints are specified in terms of linear constraints on X. In Subsection 3.2.2 the multiplicative structure of $\mathcal{E}(X)$ is exhibited. This is applied in solving the continuous time problem of geodesic regression discussed in Subsection 3.1.4. Subsections 3.2.3 and 3.2.5 correspond to the main goal of the current section. Subsection 3.2.3 considers the inverse of the transformation $X \mapsto \mathcal{E}(X)$. We will consider this inverse transformation as a local linearization transformation used to overcome nonlinear constraints. The stability of the transformation $X \mapsto \mathcal{E}(X)$ is studied in Subsection 3.2.5. As in Section 3.1, these are used to justify the application of a local linearization based approach. Subsection 3.2.4 gives the invariance properties of the transformation $X \mapsto \mathcal{E}(X)$. A recurrent idea in the following will be the equivalence of the two processes X and $\mathcal{E}(X)$. Proposition 27 of Subsection 3.2.1 states that the distribution of $\mathcal{E}(X)$ is completely determined by that of X. Similarly, Proposition 31 of Subsection 3.2.3 states that the natural filtrations of X and $\mathcal{E}(X)$ are identical –we interpret this result in terms of filtering applications in Subsection 3.2.3.

For a process X as in (3.47) note $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ the natural filtration of X. For $t \geq 0$ we have $\mathcal{F}_t = \sigma\{X_s, 0 \leq s \leq t\}$. Existence and uniqueness of the solution of (3.48) can be shown using the Picard approximation. This approximation for equations of the form (3.48) where X is a continuous semimartingale was given by Karandikar [28]. Another approximation for these equations is by multiplicative integrals. This was considered by the same author [28, 29] and we give it in Subsection 3.2.1 for the case where X is of the form (3.47). A comprehensive theory of multiplicative integration can be found in Emery's paper [14] which includes the case where the driving process X is a discontinuous semimartingale. The Picard approximation states that if X^0 is the constant process $X^0 = I_d$ and for $n \geq 1$ the process $X^n = (X_t^n)_{t\geq 0}$ is given by

$$X_t^n = I_d + \int_0^t X_s^{n-1} dX_s \tag{3.49}$$

then the processes $(X^n)_{n\geq 0}$ converge locally uniformly in the square mean to a solution of (3.48) which is moreover unique. Remember that the definition of local uniform convergence in the square mean was mentioned in equation (2.99), Paragraph 2.3.3.1. It results from this approximation that the solution $\mathcal{E}(X)$ of (3.48) is $\mathcal{F}$-adapted with almost surely continuous paths [28].

Approximation of equation (3.46) by multiplicative integrals will play the more important role in the following. It will be seen in Subsection 3.2.1 that this approximation is more adapted to the study of the nonlinear constraints inherent to (3.48). We have already met an example of approximation by multiplicative integrals in Paragraph 2.3.3.1. This was McKean's approximation of rotation Brownian motion. We also gave a somewhat different approximation by multiplicative integrals for rotation

Brownian motion with drift in Subsection 3.1.4 –see approximations (3.42). We briefly discuss the difference between the two approximations in Subsection 3.2.1. In addition to its important role in the following, approximation by multiplicative integrals has been extensively used in order to implement stochastic optimization or sampling techniques for subspace tracking and estimation applications. See [48, 70] or the recent reference by Grenander and Miller [22].

Note that Subsections 3.2.1 to 3.2.4 in the following present a more mathematical content. In Subsection 3.2.5 we finally formulate our approach to the filtering of Brownian matrix processes and consider its performances. The overall mathematical content of the current section is reflected in the references cited here, all of which belong to the mathematical literature. We do not require more knowledge of nonlinear filtering than is given in a general reference such as [23].

3.2.1 Multiplicative integrals

Solutions of equations of the form (3.48) are a general model for Brownian matrix processes under nonlinear constraints. For a given X of the form (3.47), the current subsection will clarify the constraints verified by the solution $\mathcal{E}(X)$ of (3.48). We have already seen in Paragraph 2.3.3.1 and in Subsection 3.1.4 that solutions of equations of the form (3.48) have their values in the group $SO(3)$. This particular type of nonlinear constraint was discussed in Section 3.1. Proposition 26 below states that $\mathcal{E}(X)$ has its values in a matrix Lie group, in fact a subgroup of $\mathcal{GL}(d)$. This proposition, along with Proposition 30 of Subsection 3.2.3 formulates the following important point. Although the solution $\mathcal{E}(X)$ of equation (3.48) is subject to nonlinear constraints, these constraints can be reduced to linear constraints on the driving process X. Moreover, given any process Y subject to this type of nonlinear constraint, *i.e.* having its values in a matrix Lie group, then we can identify the dynamics of Y as given by an equation of the form (3.48). Let us emphasize that from the point of view of applications the constraints we have described are quite general as they correspond to symmetry considerations important to a wide range of applied problems.

Theorem 3 below establishes the approximation of $\mathcal{E}(X)$ as a multiplicative integral. We apply this theorem in the proof of Propositions 26 and 27. As already noted, the multiplicative integrals considered in Paragraph 2.3.3.1 and in Subsection 3.1.4 are essentially different. The current subsection closes with a comparison of these two approximations. Theorem 3 is here accepted without proof. We refer to [28] for a general discussion of the proof.

To express the multiplicative integral of Theorem 3 we remind the elementary notion of a decreasing sequence of partitions. By a partition of $\mathbb{R}_+$ we mean an increasing sequence $(t_m)_{m\geq 0}$ with $t_0 = 0$ such that $t_m \uparrow \infty$ and that $\sup_{m\geq 1} |t_m - t_{m-1}| < \infty$. A decreasing sequence of partitions of $\mathbb{R}_+$ is said to be given if for all $N \geq 1$ a partition $(t_m^N)_{m\geq 0}$ of $\mathbb{R}_+$ is given and we have $|t^N| \downarrow 0$, where $|t^N| = \sup_{m\geq 1} |t_m^N - t_{m-1}^N|$ for $N \geq 1$.

Theorem 3 *Let X be given by (3.47) and suppose $(t_m^N)_{m\geq 0}$ for $N \geq 1$ gives a decreasing sequence of partitions of $\mathbb{R}_+$. For $N \geq 1$ consider the process Y^N given for $t \geq 0$ by*

$$Y_t^N = \prod_{m\geq 0} \exp(X_{t\wedge t_{m+1}^N} - X_{t\wedge t_m^N}) \tag{3.50}$$

We have that the processes Y^N converge to $\mathcal{E}(X)$ locally uniformly in the square mean.

Note that formula (3.50) generalizes the multiplicative integral in (3.42) of Subsection 3.1.4. Formula (3.52) below generalizes the multiplicative integral (2.100) of Paragraph 2.3.3.1.

We now apply Theorem 3 to Propositions 26 and 27. Proposition 26 states that $\mathcal{E}(X)$ has its values in a matrix Lie group. This amounts to a set of nonlinear constraints satisfied by $\mathcal{E}(X)$, *e.g.*

conditions (2.1) of Subsection 2.1.1. These constraints reduce to linear constraints on X. Proposition 30 of Subsection 3.2.3 will give the converse of Proposition 26. Let us specify our use of the term matrix Lie group in the following. Remember that $\mathcal{GL}(d)$ is an open subset of the vector space $\mathcal{L}(d)$. By a matrix Lie group we mean a subgroup $G \subset \mathcal{GL}(d)$ such that G is a submanifold of $\mathcal{GL}(d)$. We deal with this definition through the following fact –see for instance [46].

Fact: A subgroup $G \subset \mathcal{GL}(d)$ is a matrix Lie group *iff* G is closed in $\mathcal{GL}(d)$.

Let G be a matrix Lie group and note $T_I G$ its Lie algebra. Remember that $T_I G \subset \mathcal{L}(d)$ is a vector space of matrices and that for all $J \in T_I G$ we have that $\exp(J) \in G$. We have the following proposition.

Proposition 26 *If for $t \geq 0$ we have $X_t \in T_I G$ almost surely then for $t \geq 0$ we have $\mathcal{E}(X)_t \in G$ almost surely.*

Proof: We start by proving that $\mathcal{E}(X)_t \in \mathcal{GL}(d)$ almost surely for $t \geq 0$. Note $Y \equiv \mathcal{E}(X)$. We construct a process $\bar{Y} = (\bar{Y}_t)_{t \geq 0}$ such that for $t \geq 0$ we have $Y_t \bar{Y}_t = I_d$. Let $\bar{X}$ be the process where $\bar{X}_t = -X_t^T$ for $t \geq 0$ and define $\bar{Y}_t = \left[\mathcal{E}(\bar{X})_t\right]^T$. It is checked by simple manipulation of equation (3.48) that $\bar{Y}$ satisfies the Stratonovich stochastic differential equation

$$\bar{Y}_t = I_d - \int_0^t dX_s \bar{Y}_s \qquad \bar{Y}_t^{ij} = \delta_{ij} - \sum_{k=1}^{d} \int_0^t \bar{Y}_s^{kj} dX_s^{ik} \tag{3.51}$$

which we have again written both in matrix form and for each matrix element $\bar{Y}^{ij}$ where $1 \leq i, j \leq d$. Using the integration by parts rule of Stratonovich we have for $t \geq 0$

$$Y_t \bar{Y}_t - Y_0 \bar{Y}_0 = \int_0^t Y_s dX_s \bar{Y}_s - \int_0^t Y_s dX_s \bar{Y}_s = 0$$

since $Y_0 \bar{Y}_0 = I_d$ we can conclude that $\mathcal{E}(X)_t \in \mathcal{GL}(d)$ almost surely for $t \geq 0$.

We now prove that for $t \geq 0$ we have $\mathcal{E}(X)_t \in G$ almost surely. Let $(t_m^N)_{m \geq 0}$ for $N \geq 1$ give a decreasing sequence of partitions of $\mathbb{R}_+$. For $t \geq 0$ Theorem 3 states that $\mathcal{E}(X)_t$ is the limit in the square mean of the random variables Y_t^N given by (3.50). It follows that there exists a subsequence $Y_t^{N_k}$ –for $k \geq 1$– such that $\lim_k Y_t^{N_k} = \mathcal{E}(X)_t$ almost surely. Note that for a fixed $t \geq 0$ the product in (3.50) only involves a finite number of factors. It follows that $Y_t^{N_k} \in G$ for $k \geq 1$. From the fact that $\mathcal{E}(X)_t \in \mathcal{GL}(d)$ almost surely and that G is closed in $\mathcal{GL}(d)$ we now have that $\mathcal{E}(X)_t \in G$ almost surely.▲

The following Proposition 27 states that the distribution of $\mathcal{E}(X)$ depends only on that of X. In Paragraph 2.3.3.2, Proposition 18 established the symmetry properties of rotation Brownian motion. This proposition is in fact a special case of Proposition 27. Note that the following proof is similar to that of Proposition 18.

Proposition 27 *Let X^1 and X^2 be processes of the form (3.47) and suppose X^1 and X^2 have identical finite dimensional distributions. We have that $\mathcal{E}(X^1)$ and $\mathcal{E}(X^2)$ have identical finite dimensional distributions.*

Proof: We will only prove that $\mathcal{E}(X^1)_t \stackrel{d}{=} \mathcal{E}(X^2)_t$ for $t \geq 0$. This can be used along with (i) of Proposition 29 of the next subsection in order to give a complete proof of the proposition.

Let $t \geq 0$ and consider for $N \geq 1$ the random products π_1^N and π_2^N given as in (3.50)

$$\pi_1^N = \prod_{m \geq 0} \exp(X^1_{t \wedge t_{m+1}^N} - X^1_{t \wedge t_m^N}) \qquad \pi_2^N = \prod_{m \geq 0} \exp(X^2_{t \wedge t_{m+1}^N} - X^2_{t \wedge t_m^N})$$

Where $(t_m^N)_{m\geq 0}$ give a decreasing sequence of partitions of $\mathbb{R}_+$. Note that each of these products only involves a finite number of factors. More precisely

$$\pi_1^N = \prod_{t_m^N \leq t} \exp(X^1_{t\wedge t^N_{m+1}} - X^1_{t\wedge t^N_m}) \qquad \pi_2^N = \prod_{t_m^N \leq t} \exp(X^2_{t\wedge t^N_{m+1}} - X^2_{t\wedge t^N_m})$$

The factors in each of these products π_1^N and π_1^N are independent. This follows immediately from the fact that the processes X^1 and X^2 have independent increments. Moreover, for all $N \geq 1$ and $m \geq 0$ we have by hypothesis

$$X^1_{t\wedge t^N_{m+1}} - X^1_{t\wedge t^N_m} \stackrel{d}{=} X^2_{t\wedge t^N_{m+1}} - X^2_{t\wedge t^N_m}$$

It follows that for $N \geq 1$ the factors in the products π_1^N and π_2^N which correspond to the same value of m are equal in distribution. Finally, for $N \geq 1$ we have $\pi_1^N \stackrel{d}{=} \pi_2^N$. We now have $\mathcal{E}(X^1)_t \stackrel{d}{=} \mathcal{E}(X^2)_t$, since equality in distribution is preserved by limits in the square mean.▲

Consider the multiplicative integral (2.100) of Paragraph 2.3.3.1. In order to generalize it to equation (3.48) we consider the following processes $\tilde{Y}^N$ for $N \geq 1$. These are based on a linear interpolation of the driving process X of equation (3.48). Using the notation of Theorem 3, let $\Delta X_m^N = X_{t^N_{M+1}} - X_{t^N_m}$ and $\Delta t_m^N = t_{m+1}^N - t_m^N$ for $N \geq 1$ and $m \geq 0$. For $N \geq 1$ and $t \geq 0$ the processes $\tilde{Y}^N$ is given in formula (3.52). We have that these processes converge locally uniformly in the square mean to $\mathcal{E}(X)$.

$$\tilde{Y}_t^N = \prod_{m\geq 0} \exp((\Delta X_m^N / \Delta t_m^N)(t \wedge t_{m+1}^N - t \wedge t_m^N)) \tag{3.52}$$

There are important differences between the approximations (3.50) and (3.52) of equation (3.48). Note first that the construction of the processes $\tilde{Y}^N$ from X is anticipative. For $t_m^N \leq t < t_{m+1}^N$ the value of $\tilde{Y}_t^N$ depends on $X_{t^N_{m+1}}$ through ΔX_m^N. On the contrary, the process Y^N for $N \geq 1$ is clearly $\mathcal{F}$-adapted and can be constructed from X in a causal way. The anticipativeness of the processes $\tilde{Y}^N$ reflects the fact that they are based on a linear interpolation of X between the points of the partition $(t_m^N)_{m\geq 0}$. This construction has the advantage of leading to processes Y^N which satisfy ordinary differential equations and which can be interpreted in a simple way –compare to Paragraph 2.3.3.1. The paths of the processes Y^N inherit the irregular nature of the Brownian paths of X and can only be described in terms of stochastic differential equations.

3.2.2 Multiplicative structure of $\mathcal{E}(X)$

The solution $\mathcal{E}(X)$ of equation (3.48) is a nonstationary process. It will be seen in the current subsection that $\mathcal{E}(X)$ has a multiplicative structure defined in terms its increments $\mathcal{E}(X)_{(s|t)}$ which are $\mathcal{GL}(d)$-valued random variables for $0 \leq s \leq t$. This multiplicative structure corresponds to the definition of Lévy processes given by conditions (L1) to (L4) of Subsection 2.3.1. Under a simple condition on the driving process X, the increments of $\mathcal{E}(X)$ will be found to be stationary. In this way, a stationary structure is identified within the nonstationary process $\mathcal{E}(X)$. This is applied in Paragraph 3.2.2.2 below to give a solution of the problem of geodesic regression presented in Subsection 3.1.4. We start in Paragraph 3.2.2.1 by proving Propositions 28 and 29 which give the main properties of the increments of $\mathcal{E}(X)$.

3.2.2.1 Increments and their properties

In the case of rotation Lévy processes, increments were defined in Subsection 2.3.1 and were required to verify conditions (L1) to (L4). It is possible to start with a similar definition of the increments

of $\mathcal{E}(X)$. It was seen in the proof of Proposition 26 that $\mathcal{E}(X)_t \in \mathcal{GL}(d)$ almost surely for all $t \geq 0$. Accordingly, we can define for $0 \leq s \leq t$

$$\mathcal{E}(X)_{(s|t)} = [\mathcal{E}(X)_s]^{-1}\mathcal{E}(X)_t \tag{3.53}$$

Our main goal in the current paragraph will be to prove Proposition 29. This proposition states that the increments of $\mathcal{E}(X)$ as defined by (3.53) are independent and gives a condition under which they are stationary. Thus, the increments of $\mathcal{E}(X)$ are seen to verify conditions similar to conditions (L1) to (L4) of Subsection 2.3.1. It is not straightforward to establish Proposition 29 from definition (3.53). We will start by proving Proposition 28 which gives a dynamical definition of the increments of $\mathcal{E}(X)$. Proposition 29 will follow directly from Proposition 28.

Before stating Proposition 28, let us note the following. Suppose given a process X as in (3.47). For $s \geq 0$ define the process $X^s = (X^s_t)_{t\geq 0}$ by $X^s_t = X_{s+t} - X_s$. We admit that X^s can be expressed in the form (3.47), so that the process $\mathcal{E}(X^s)$ is defined as above –using for instance the Picard approximation.

Proposition 28 *For all $0 \leq s \leq t$ we have*

$$\mathcal{E}(X)_{(s|t)} = \mathcal{E}(X^s)_{t-s} \tag{3.54}$$

Proof: Let $s \geq 0$ be fixed. Note $Y^1 = (Y^1_t)_{t\geq 0}$ the process given by $Y^1_t = \mathcal{E}(X)_{t\wedge s}$. Let $Y^2 = (Y^2_t)_{t\geq 0}$ be the process $Y^2_t = \mathcal{E}(X^s)_{(t-s)^+}$ –for $a \in \mathbb{R}$ we write $a^+ = \max\{a, 0\}$. We will prove that for $t \geq 0$

$$\mathcal{E}(X)_t = Y^1_t Y^2_t \tag{3.55}$$

By replacing for some $s \leq t$ and comparing to (3.53) the proposition is proved. Let X^1 and X^2 be the processes defined for $t \geq 0$ by $X^1_t = X_{t\wedge s}$ and $X^2_t = X_t - X^1_t$. Note that Y^1 solves the following Stratonovich stochastic differential equation

$$Y^1_t = I_d + \int_0^t Y^1_u X^1_u \tag{3.56}$$

for $t \geq 0$. We will prove that Y^2 solves the additional equation

$$Y^2_t = I_d + \int_0^t Y^2_u X^2_u \tag{3.57}$$

for $t \geq 0$. This follows from the definition of $\mathcal{E}(X^s)$

$$\mathcal{E}(X^s)_t = I_d + \int_0^t \mathcal{E}(X^s)_u dX^s_u$$

by applying the change of time $t \mapsto (t-s)^+$. It is now possible to prove (3.55) by checking that the product on the right hand side of this formula verifies equation (3.48). This is done by integration by parts using (3.56) and (3.57).▲

For Proposition 29 we will say that X of the form (3.47) is stationary if there exists $h \in \mathcal{L}(d)$ such that for $t \geq 0$ we have $f_{ij}(t) = h_{ij}t$ and $Q_{ij}(t) = 1$ for $1 \leq i, j \leq d$.

Proposition 29 *The following hold for all $0 \leq s \leq t$*
(i) *$\mathcal{E}(X)_{(s|t)}$ is independent of $\mathcal{F}_s$.*
(ii) *If X is stationary then $\mathcal{E}(X)_{(s|t)} \overset{d}{=} \mathcal{E}(X)_{t-s}$.*

Proof: (ii) follows using Proposition 27 of Subsection 3.2.1. If X is stationary then for $s \geq 0$ the processes X and X^s have identical finite dimensional distributions. It follows that for $s \leq t$ we have $\mathcal{E}(X)_{t-s} \stackrel{d}{=} \mathcal{E}(X^s)_{t-s}$. By Proposition 28 we have $\mathcal{E}(X)_{(s|t)} = \mathcal{E}(X^s)_{t-s}$.

We prove (i) using Theorem 3. For $u \geq 0$ we have that the the following products π^N converge in the square mean to $\mathcal{E}(X^s)_u$

$$\pi^N = \prod_{t_m^N \leq u} \exp(X^s_{u \wedge t^N_{m+1}} - X^s_{u \wedge t^N_m})$$

where $(t_m^N)_{m \geq 0}$ for $N \geq 1$ give a decreasing sequence of partitions of $\mathbb{R}_+$. Using the fact that X has independent increments we can show that π^N for $N \geq 1$ is independent of $\mathcal{F}_s$. Since independence is preserved by limits in the square mean, we have that $\mathcal{E}(X^s)_u$ is independent of $\mathcal{F}_s$. Now, (i) follows by putting $u = t - s$ and using Proposition 28.▲

To end this paragraph we make the following remark in relation to Subsection 3.2.4. The increments of $\mathcal{E}(X)$ are defined in (3.53) using left multiplication. That is

$$\mathcal{E}(X)_t = \mathcal{E}(X)_s \mathcal{E}(X)_{(s|t)}$$

for $0 \leq s \leq t$. It is possible to define increments using right multiplication by considering the quantities

$$\mathcal{E}(X)_t [\mathcal{E}(X)_s]^{-1} \tag{3.58}$$

for $0 \leq s \leq t$. The choice of expression (3.53) over (3.58) is essential. It is only with this definition that Proposition 29 holds. In Subsection 3.2.4, we will see how to construct a process similar to $\mathcal{E}(X)$ but for which increments defined as in (3.58) verify Proposition 29. This situation is parallel to the discussion in Subsection 2.3.1 of left and right rotation Lévy processes.

3.2.2.2 Application to geodesic regression

The problem of geodesic regression stated in Subsection 3.1.4 can now be considered as a parametric estimation problem. This is made possible by of Proposition 29 of the last paragraph. Let us start by reminding this problem and explaining the role of Proposition 29.

The observed process satisfies the Stratonovich stochastic differential equation (3.38) which we copy here.

$$dY_t = Y_t dX_t \qquad Y_0 = I$$

The driving process X is made up of the unknown angular velocity ω and an additive Brownian noise process J. For $t \geq 0$ we have $J_t = B_t^1 J_1 + B_t^2 J_2 + B_t^3 J_3$ where $[B^1, B^2, B^3]$ are the coordinates of a $\mathbb{R}^3$-valued Brownian motion process B.

$$X = \omega t + J_t$$

Let us remind that ω is an antisymmetric matrix. We can write $\omega = \omega_1 J_1 + \omega_2 J_2 + \omega_3 J_3$. It follows that the driving process X is equivalent to a $\mathbb{R}^3$-valued Brownian motion with drift. Note that in Subsection 3.1.4 we noted Y^1 the signal process and Y^2 the observed process. Accordingly, these two processes were related to driving processes X^1 and X^2 –see equation (3.38). Here we will not consider Y^1 and X^1 directly and so we have written Y and X instead of Y^2 and X^2.

Our observations consist in a trajectory of the process Y taken over an interval $0 \leq t \leq T$, $T \geq 0$ being fixed. For simplicity we will make the hypothesis that the covariance matrix C of B is of the form $C = a^2 I$ where $a \in \mathbb{R}$. This will be essential for formula (3.60) below. For all $t \geq 0$ we have found in formula (3.41) of Subsection 3.1.4

$$\mathbb{E}(Y_t) = \exp\left[\left(\omega - a^2 I\right) t\right] \tag{3.59}$$

As noted before, the stochastic differential equation (3.38) for Y is of the form (3.48). Moreover, it is driven by a stationary process X. According to Proposition 29, the process $Y \equiv \mathcal{E}(X)$ has independent and stationary increments. This is essential to our estimation approach. While the process Y itself is nonstationary and cannot be used in estimation, its increments taken over equal time intervals are *i.i.d.* and offer a standard starting point for estimation problems.

Suppose there exists $N \geq 1$ such that $T \geq N$. Concretely, let us choose in this case $N = \lfloor T \rfloor$ the lower integer part of T. We will be interested in the matrix $\mathbb{E}(Y_1) = M$. This is a constant matrix depending on the unknown ω and on the variance parameter a^2 as in (3.59). These two quantities can be separated by carrying out a polar decomposition of the matrix M. Indeed M can be decomposed into a symmetric positive definite factor P depending only on a^2 and an orthogonal factor O depending only on ω.

$$P = MM^T = e^{-2a^2} I \qquad\qquad O = P^{-\frac{1}{2}} M = \exp(\omega) \tag{3.60}$$

where the square root of a symmetric positive definite matrix is taken as the unique symmetric positive definite matrix square root.

It is now possible to consider an estimation approach similar to that of Paragraph 2.4.2.1. For $1 \leq n \leq N$ let Z_n be the rotation random variable[2]

$$Z_n = Y_{n-1}^T Y_n \equiv Y_{(n-1|n)}$$

We have by Proposition 29 that the Z_n are *i.i.d.* rotation random variables. Consider the following estimate of M

$$\hat{M}_N = \frac{1}{N} \sum_{n=1}^{N} Z_n \tag{3.61}$$

We consider the convergence of the estimates $\hat{M}_N$ when $T \uparrow \infty$. That is, when the time over which the process Y is observed becomes large. In this limit we have by the definition of N that $N \uparrow \infty$. By the strong law of large numbers $\lim_N \hat{M}_N = M$ almost surely.

The problem of estimating ω can now be approached as for the problem of decompounding considered in Section 2.4.2. Formulae (3.59) and (3.60) play a similar role to (2.120) and (2.121) of Paragraph 2.4.2.1, respectively. Indeed, (3.59) relates the unknown ω to the average quantity M while (3.60) gives the inverse transformation leading to an expression of ω. As in Paragraph 2.4.2.1, we are interested in local inversion of (3.59) using (3.60) and the empirical estimates $\hat{M}_N$ which converge to M almost surely. Given the special form of M in (3.59), the following simplification can be made in (3.60). We have that

$$O = \frac{\sqrt{3}}{|M|} M \tag{3.62}$$

For $N \geq 1$ define $\hat{O}_N$ as in (3.63) below. It is clear that $\hat{O}_N$ is well-defined. Moreover, a simple continuity argument shows that $\lim_N \hat{O}_N = O$ almost surely. We consider the random variables $\hat{O}_N$ as our estimates for O which depends immediately on ω. Note that the correspondence $O \leftrightarrow \omega$ is unique when O verifies condition (3.3) of Paragraph 3.1.1.1. If this condition is not verified then ω is only determined up to sign. For $N \geq 1$ we put

$$\hat{O}_N = 0 \text{ if } M_N = 0 \qquad \hat{O}_N = \frac{\sqrt{3}}{|\hat{M}_N|} \hat{M}_N \text{ otherwise} \tag{3.63}$$

[2]remember that since $Y_{n-1} \in SO(3)$ we have $Y_{n-1}^{-1} = Y_{n-1}^T$.

While the convergence of the $\hat{O}_N$ to O is clearly ensured, this is a surprising result. These estimates are calculated according to an apparently naive recepie. Indeed, one starts by averaging the *i.i.d.* observations $Z_1, \ldots, Z_N$ to obtain M_N. The mapping $(Z_1, \ldots, Z_N) \mapsto M_N$ consists in a linear operation applied to observations which are subject to a nonlinear constraint, in particular that $Z_n \in SO(3)$ for $1 \leq n \leq N$. To recapture this nonlinear constraint, one finally performs a renormalization step as in (3.63).

The success of (3.63) is due to the fact that it arises from (3.59) which includes *a priori* knowledge of the distribution of the observations $Z_1, \ldots, Z_N$. It is thus interesting to compare our use of (3.63) to a method which required no *a priori* knowledge on the observations.

Such a purely geometric method can be defined in terms of the center of mass or Fréchet mean of the observations. This is given a precise mathematical formulation by Manton in [43]. Its applications include blind source separation [44] and medical image processing [55]. The center of mass $\tilde{O}_N$ of the observations $Z_1, \ldots, Z_N$ is defined in terms of a nonlinear optimization problem. More precisely, $\tilde{O}_N \in SO(3)$ is a minimiser of the function $f : SO(3) \to \mathbb{R}_+$ given as in [43]

$$f(X) = \frac{1}{2} \sum_{n=1}^{N} d^2(Z_n, X) \tag{3.64}$$

for $X \in SO(3)$. Here, for $1 \leq n \leq N$ the notation $d^2(Z_n, X)$ stands for the squared angular distance between Z_n and X. The definition of the function f in (3.64) implies that $\tilde{O}_N$ estimates O based on a criterion of least square angular distance from the observations $Z_1, \ldots, Z_N$. In [43], it is shows that while $\tilde{O}_N$ always exists it is not always uniquely defined. When it is well-defined, $\tilde{O}_N$ can be calculated using an intrinsic Newton method.

For the geodesic regression problem at hand, our estimates $\hat{O}_N$ have important advantages over the center of mass $\tilde{O}_N$. They are simple to calculate and well-defined while the center of mass can only be calculated using an iterative algorithm and is possibly not well-defined.

3.2.3 Inversion of dynamics

This subsection defines our local linearization transformation for Brownian matrix processes. In Subsection 3.2.5, this will be used in generalizing the filtering approach outlined in Subsection 3.1. The precise form of this transformation will be given in Theorem 4. This theorem identifies for any given Brownian matrix process Y a process X of the form (3.47) such that $Y \equiv \mathcal{E}(X)$. This consists in an inversion of the transformation $X \mapsto \mathcal{E}(X)$ defined by equation (3.48). Formally, Theorems 3 and 4 can be considered in analogy with the discrete time transformations defined in Paragraph 3.1.1.1.

Proposition 30 establishes an important property of local linearization. Suppose the Brownian matrix process Y has its values in a matrix Lie group, so that it is subject to nonlinear constraints. The corresponding process X is of the form (3.47) and subject only to linear constraints. Proposition 31 gives the equivalence of the natural filtrations of Y and X. We will discuss this result in terms of filtering applications. We now give Proposition 30. Although it uses Theorem 4, the first part of this proposition can be proved directly.

We start from the following observation. Let X be a process of the form (3.47). It is straightforward from equation (3.48) that X can be recovered from $\mathcal{E}(X)$. Indeed, for $t \geq 0$

$$X_t = \int_0^t Y_s^{-1} dY_s \tag{3.65}$$

where we have put $Y \equiv \mathcal{E}(X)$. After specifying the class of processes Y for which Proposition 30 will apply, we will simply use the same formula (3.65) to obtain the required process X.

Let the process $Y = (Y_t)_{t\geq 0}$ have its values in $\mathcal{GL}(d)$ and in particular $Y_0 = I_d$. Suppose also Y is a continuous semimartingale. For $0 \leq s \leq t$ define the increment of Y as

$$Y_{(s|t)} = Y_s^{-1} Y_t$$

We assume Y has independent increments. Let $\mathcal{H}$ be the natural filtration of Y. That is, $\mathcal{H}_t = \sigma\{Y_s, s \leq t\}$ for $t \geq 0$. For $0 \leq s \leq t$ we have that $Y_{(s|t)}$ is independent of $\mathcal{H}_s$. The following process $X = (X_t)_{t\geq 0}$ is well-defined

$$X_t = \int_0^t Y_s^{-1} dY_s \tag{3.66}$$

Proposition 30 *Let Y be as above and X be given by (3.66). Let G be a matrix Lie group with Lie algebra $T_I G$. The following hold*
(i) *X is of the form (3.47) and $Y \equiv \mathcal{E}(X)$.*
(ii) *If for $t \geq 0$ we have $Y_t \in G$ almost surely then for $t \geq 0$ we have $X_t \in T_I G$ almost surely.*

Proof: We start by proving (i). The proof of (ii) uses Theorem 4 below.

From its definition (3.66), X is a continuous semimartingale. By equation (3.48) it follows that for $t \geq 0$

$$Y_t = I + \int_0^t Y_s Y_s^{-1} dY_s = I + \int_0^t Y_s dX_s$$

this proves that $Y \equiv \mathcal{E}(X)$. In order to show that X is of the form (3.47) it is enough to show that X has independent increments. It is clear from (3.66) that X is $\mathcal{H}$-adapted. For $0 \leq s \leq t$ we have the limit in probability

$$X_t - X_s = \int_s^t Y_s^{-1} dY_s = \lim_n \sum_{m=1}^{M} Y_{t_{m-1}^n}^{-1} (Y_{t_m^n} - Y_{t_{m-1}^n})$$

where the step of the partitions $s = t_1^n < \ldots < t_M^n = t$ of the interval $[s, t]$ defined for $n \geq 1$ decreases to zero with n. It is possible to transform the sum in the last expression to obtain

$$X_t - X_s = \lim_n \sum_{m=1}^{M} Y_{(t_{m-1}^n | t_m^n)} - I \tag{3.67}$$

By hypothesis for all $n \geq 1$ and $m \geq 0$ we have that $Y_{(t_{m-1}^n | t_m^n)}$ is independent of $\mathcal{H}_{t_m^n}$. By using the fact that independence is preserved by limits in probability we can conclude that $X_t - X_s$ is independent of $\mathcal{H}_s$. It follows in particular that X has independent increments. Since X is a continuous process with independent increments, we have that it is of the form (3.47). This is an immediate result of the Lévy-Itô representation –see [27].

The proof of (ii) follows immediately from Theorem 4. For $t \geq 0$ we have by this theorem the limit in probability $\lim_N X_t^N = X_t$. Here the processes X^N for $N \geq 1$ are given in expression (3.68). It is clear that for $N \geq 1$ we have $X_t^N \in T_I G$ almost surely. It follows that for $t \geq 0$ we have $X_t \in T_I G$ almost surely.▲

The following Proposition 31 states that the natural filtrations of the processes X and $\mathcal{E}(X)$ are identical. While the proof of this proposition is simple, it is of general significance to the approach presented in the current chapter. Given a process X of the form (3.47), it is clear that the processes X and $\mathcal{E}(X)$ are of different geometric natures. While $\mathcal{E}(X)$ is subject to nonlinear constraints X is only subject to linear constraints. Proposition 31 tells us that in spite of this difference, X and $\mathcal{E}(X)$ have the same content in terms of information. In particular, filtering with respect to observations of either X or $\mathcal{E}(X)$ leads to the same posterior knowledge.

Proposition 31 *In the notation of Proposition 30, let $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ and $\mathcal{H} = (\mathcal{H}_t)_{t\geq 0}$ be the natural filtrations of X and Y respectively. We have that $\mathcal{F}_t = \mathcal{H}_t$ for $t \geq 0$.*

Proof: The proof is immediate. By (i) of Proposition 30 we have $Y \equiv \mathcal{E}(X)$. It follows that Y is $\mathcal{F}$-adapted. In particular, for $t \geq 0$ we have $\mathcal{H}_t \subset \mathcal{F}_t$. By (3.66) we have that X is $\mathcal{H}$-adapted so that for $t \geq 0$ we have $\mathcal{F}_t \subset \mathcal{H}_t$. Finally, $\mathcal{F}_t = \mathcal{H}_t$ for $t \geq 0$.▲

We finally give Theorem 4. In the current continuous time setting, local linearization is given as a limiting process. Such a limiting process has already been stated in the special case of formulae (3.42) of Subsection 3.1.4. While the use of a limiting process raises an additional technical difficulty, the corresponding local linearization transformation retains the desirable properties of local linearization as discussed in Section 3.1. In order to state Theorem 4, we need the following notation. Let $\log : \mathcal{GL}(d) \to \mathcal{L}(d)$ be the mapping where for $y \in \mathcal{GL}(d)$ we have $\log(y) = x$ where x is the unique $x \in \mathcal{L}(d)$ such that $y = \exp(x)$ when $|I_d - y| < 1$ and $\log(y) = 0$ otherwise.

Theorem 4 *In the notation of Proposition 31, let $(t_m^N)_{m\geq 0}$ for $N \geq 1$ give a decreasing sequence of partitions of $\mathbb{R}_+$, we have that the following processes $X^N = (X_t^N)_{t\geq 0}$ are continuous $\mathcal{H}$-adapted and converge locally uniformly in probability to X.*

$$X_t^N = \sum_{m\geq 0} \log\left[Y_{(t_m^N \wedge t | t_{m+1}^N \wedge t)}\right] \tag{3.68}$$

We do not detail the proof of Theorem 4 here. It can be seen to follow by a development of formula (3.67) in the proof of Proposition 30. Indeed, from the definition of X in (3.66) we have for $t \geq 0$

$$X_t = \lim_N \sum_{m\geq 0} Y_{t_m^N}^{-1}(Y_{t_{m+1}^N} - Y_{t_m^N}) = \lim_N \sum_{m\geq 0} Y_{(t_m^N \wedge t | t_{m+1}^N \wedge t)} - I$$

Remember now that for $y \in \mathcal{GL}(d)$ such that $|I - y| < 1$ we have the following approximation.

$$\log(y) = (y-1) + O(|y-1|^2) \tag{3.69}$$

Replacing in the expression for X_t we have

$$X_t = \lim_N \sum_{m\geq 0} \log\left[Y_{(t_m^N \wedge t | t_{m+1}^N \wedge t)}\right] + \lim_N \sum_{m\geq 0} O(|Y_{(t_m^N \wedge t | t_{m+1}^N \wedge t)} - I|^2)$$

The processes defined for $N \geq 1$ by the second sum on the right hand side converge locally uniformly in probability to the zero process when $N \uparrow \infty$.

3.2.4 Invariance properties

In the above development on matrix Brownian processes, certain choices were made in an arbitrary fashion. By adopting formula (3.53) in Subsection 3.2.2 we considered left rather than right increments. In general, we have also worked with an arbitrary time parameter attached to the driving process X of equation (3.48). In the current subsection, we will try to apprehend these choices by considering related invariance properties.

In relation to our general aim of using local linearization for the filtering of Brownian matrix processes, invariance properties will clarify the range of applicability of local linearization as defined in the last Subsection 3.2.3. Proposition 32 below will state the invariance by multiplication of local linearization. It is seen that our choice of formula (3.53) leads to invariance by left multiplication.

Invariance by right multiplication obtains when increments can be defined as in formula (3.58). Invariance by time shift is considered in Proposition 33 in the form of a Markov property of Brownian matrix processes.

For a process X of the form (3.47), $Y \equiv \mathcal{E}(X)$ has the property of invariance by left multiplication. Given any random variable K which is $\mathcal{F}_0$-measurable and has its values in $\mathcal{GL}(d)$, the process $Z \equiv (K\mathcal{E}(X)_t)_{t\geq 0}$ is the unique solution of the following Stratonovich stochastic differential equation

$$Z_t = K + \int_0^t Z_s dX_s \tag{3.70}$$

The left increments of Z are the same as those of Y and have the same properties given by Proposition 29. Indeed, note that for all $0 \leq s \leq t$ we have

$$Z_s^{-1} Z_t = Y_s^{-1} Y_t = \mathcal{E}(X)_{(s|t)}$$

Local linearization as in formula (3.68) leads to the same process X when applied to either one of the processes Y or Z. We thus have that local linearization is invariant by left multiplication for processes whose increments are defined by formula (3.53).

In order to obtain a similar property of invariance by right multiplication we consider the following process $\bar{Y}$ where for $t \geq 0$ we have

$$\bar{Y}_t = \left[\mathcal{E}(X^T)_t\right]^T \tag{3.71}$$

which is defined in the same way as Y, since the process $X^T = (X_t^T)_{t\geq 0}$ is of the form (3.47). For the process $\bar{Y}$ we consider right increments given by formula (3.58). More precisely, for $0 \leq s \leq t$ the increment $\bar{Y}_{(s|t)}$ is defined as

$$\bar{Y}_{(s|t)} = \bar{Y}_t \bar{Y}_s^{-1} \tag{3.72}$$

For the process $\bar{Y}$, we have the following proposition. It states that $\bar{Y}$ is invariant by right multiplication and can be proved immediately from (3.71).

Proposition 32 *The following hold*
(i) *Let K be as in (3.70). The process $\bar{Z} \equiv (\bar{Y}_t K)_{t\geq 0}$ is the unique solution of the following Stratonovich stochastic differential equation*

$$\bar{Z}_t = K + \int_0^t dX_s \bar{Z}_s \tag{3.73}$$

The straightforward relation (3.71) between Y and $\bar{Y}$ allows for the properties of Y seen in Subsections 3.2.2 and 3.2.3 to be proved for $\bar{Y}$. We have by (3.71) that for all $0 \leq s \leq t$

$$\bar{Y}_{(s|t)} = [\mathcal{E}(X^T)_{(s|t)}]^T$$

It follows from this identity that the right increments of $\bar{Y}$ verify Proposition 29. More precisely, we have for all $0 \leq s \leq t$
(i) $\bar{Y}_{(s|t)}$ is independent of $\mathcal{F}_s$.
(ii) If X is stationary then $\bar{Y}_{(s|t)} \stackrel{d}{=} \bar{Y}_{t-s}$.
Formula (3.68) can be used immediately to apply local linearization to $\bar{Y}$. It is enough to replace the left increments in this formula by right increments as defined in (3.72). Retaining the notation of Theorem 4 we have the processes X^N given for $t \geq 0$ below converge locally uniformly in the square mean to X

$$X_t^N = \sum_{m\geq 0} \log\left[\bar{Y}_{(t_m^N \wedge t | t_{m+1}^N \wedge t)}\right]$$

where the increments of $\bar{Y}$ are given by (3.72). Note again that local linearization leads to the same process X when applied to $\bar{Y}$ or to any process of the form $\bar{Z}$ as in Proposition 32. This expresses the invariance by right multiplication of local linearization.

The following Proposition 33 considers the invariance by time shift of local linearization. Note that item (i) of this proposition can be interpreted as follows. A time shift in the driving process X leads to a similar time shift in the process $\mathcal{E}(X)$. For the driving process X time shift is defined in terms of the linear structure of $\mathcal{L}(d)$, while for $\mathcal{E}(X)$ it is defined in terms of the multiplicative structure of $\mathcal{GL}(d)$.

Proposition 33 *Suppose X is of the form (3.47) and note $Y \equiv \mathcal{E}(X)$. Consider the process $Y^s = (Y_t^s)_{t\geq 0}$ where $Y_t^s = Y_{(s|s+t)}$. The following hold*
(i) *$Y^s \equiv \mathcal{E}(X^s)$ where $X^s = (X_t^s)_{t\geq 0}$ is given by $X_t^s = X_{s+t} - X_s$.*
(ii) *Y^s is independent of $\mathcal{F}_s$.*
(iii) *If X is stationary then Y and Y^s have identical finite dimensional distributions.*

Proof: Note that (i) follows immediately from formula (3.54) of Proposition 28. Moreover, (iii) follows from (ii) of Proposition 29. Indeed, when X is stationary then X and X^s have identical finite dimensional distributions, for all $s \geq 0$. The proof of (ii) follows using (i) of Proposition 29.

In order to prove (ii) we must prove that for all $N \geq 1$ and all $0 \leq t_1 \leq \ldots \leq t_N$ we have that the random variable $V = [Y_{t_1}^s, \ldots, Y_{t_N}^s]$ which takes its values in $[\mathcal{GL}(d)]^N$ is independent of $\mathcal{F}_s$. It is clear that for all such N and $t_1, \ldots, t_N$ the $[\mathcal{GL}(d)]^N$-valued random variable $W = [Y_{(t_1|t_2)}^s, \ldots, Y_{(t_{N-1}|t_N)}^s]$ is independent of $\mathcal{F}_s$. Indeed, this can be seen to follow by (i) of Proposition 29. To complete the proof, we express V in the form $V = f(W)$ where f is a continuous mapping of $[\mathcal{GL}(d)]^N$. To this effect, let f be given for $w = [v_1, \ldots v_N] \in [\mathcal{GL}(d)]^N$ by $v = f(w) = [v_1, \ldots v_n]$ where for $1 \leq n \leq N$

$$v_n = \prod_{i=1}^{n} w_i$$

where the product is ordered from left to right.▲

Let us finally note that the invariance properties given in the current subsection can be seen as continuous time counterparts of the properties given in Subsection 3.1.3 for the case of discrete time rotation processes. The role of these properties in filtering applications can be discussed as in Subsection 3.1.3 –compare to [64] where we discuss continuous time rotation processes.

3.2.5 Stability properties

The current subsection formulates our approach to the filtering of matrix processes under nonlinear constraints. This is done within the framework of Brownian matrix processes, which was developed in Subsections 3.2.1 to 3.2.4 above. As explained in the introduction of the current chapter, we seek to study for Brownian matrix processes a similar use of local linearization to what was done for discrete time rotation processes in Section 3.1. Local linearization is now implemented according to Theorem 4 of Subsection 3.2.3. The stability properties related to our local linearization transformation play the main role in justifying our approach.

Our main results is Proposition 35. This states the stability of the transformation $X \mapsto \mathcal{E}(X)$. Of course, this proposition is intended in the context of our application of local linearization which we start by describing more precisely. Let $d \geq 1$ and suppose $G \subset \mathcal{GL}(d)$ is a matrix Lie group and note $T_I G$ its Lie algebra. Consider a filtering problem that can be stated as follows. A signal process Y^1 is assumed, to which we do not have direct access. On the other hand, we have observations of

another process Y^2 which we call the observation process. Both these processes have their values in G and our problem is to infer the signal process Y^1 using the information available in the observation process Y^2. A solution consists in a process $\hat{Y}^1$ verifying the following two requirements.

- $\hat{Y}^1$ should be *optimal*. This process is constructed only from available observations of Y^2. It should make complete use of the information available in these observations.
- $\hat{Y}^1$ verifies the same nonlinear *constraints* as Y^1 and Y^2. Namely, it should have its values in G.

In general, these two requirements are not compatible. It is well known from the theory of optimal filtering that the process $\hat{Y}^1$ verifying the first requirement of optimality is the conditional expectation process [23]

$$\hat{Y}^1_t = \mathbb{E}[Y^1|\mathcal{F}_t] \tag{3.74}$$

for $t \geq 0$, where $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ is the natural filtration of Y^2. This process is moreover the best approximation of Y^1 in the sense of the pointwise square mean. Since conditional expectation is a linear operation, it is clear that $\hat{Y}^1$ does not in general verify the second requirement of nonlinear constraints.

Suppose Y^1 and Y^2 verify the conditions of Proposition 30. In this case we know that there exist processes X^1 and X^2 of the form (3.47) with values in T_IG such that $Y^i \equiv \mathcal{E}(X^i)$ for $i = 1, 2$. Moreover, X^1 and X^2 are given by Theorem 4. It seems desirable to deal with the two requirements stated above through the process X^2 rather than the observation process Y^2. Note that

- By Proposition 31 the natural filtration of X^2 is also equal to $\mathcal{F}$. In other words, X^2 contains the same information as Y^2.
- X^2 is only subject to linear constraints.

The main idea of our approach is to seek $\hat{Y}^1$ of the form $\mathcal{E}(\hat{X}^1)$ where $\hat{X}^1$ is a process of the form (3.47) with values in T_IG. By proposition 26 we have that $\hat{Y}^1$ then has its values in G, *i.e.* it verifies our second requirement related to nonlinear constraints. The process $\hat{X}^1$ is to be constructed from X^2 and should be a good approximation of X^1. We can describe this approach using a diagram similar to (3.36) of Subsection 3.1.3.

$$\begin{array}{ccc} Y^2 & \xrightarrow{F} & \hat{Y}^1 \approx Y^1 \\ \downarrow & & \uparrow \\ X^2 & \xrightarrow{f} & \hat{X}^1 \approx X^1 \end{array} \tag{3.75}$$

Here, downward arrows correspond to the application of Theorem 4 while upward arrows correspond to the application of Theorem 3. Theorem 4 is thought of as giving a local linearization transformation. Note that both these theorems involve a limiting process and thus we do not have X^2 or $\hat{Y}^1$ exactly. We will neglect this difficulty in the following. Let us just mention that Karandikar [28] discusses how the convergence rate in these theorems can be improved to an exponential rate.

We have formulated our use of local linearization in the filtering of matrix processes under nonlinear constraints. The resulting approach is not optimal in any clear sense and we have not yet given it any justification. At this stage, this approach is only a plausible *prescription* for approximating Y^1 given Y^2. It is precisely the stability of the transformation $X \mapsto \mathcal{E}(X)$ that justifies our approach. Assume we have the necessary *a priori* knowledge on X^1 and X^2 to realize a good approximation $\hat{X}^1 \approx X^1$ as in diagram (3.75). Proposition 35 ensures that this approximation will be preserved by application of the upward arrow in the diagram and we should obtain an accordingly acceptable approximation

$\hat{Y}^1 \approx Y^1$. We will see however that we have –as in Paragraph 3.1.1.2 for the case of rotation time series– the appearance of drifting problems. As a result, obtaining a good approximation $\hat{Y}^1 \approx Y^1$ may turn out to be highly costly even if possible in principle.

We now turn to Proposition 35. This proposition states the stability of the transformation $X \mapsto \mathcal{E}(X)$. It is not intended to give a sharp estimate of this stability but rather to provide a qualitative appreciation of the roles of stability and drifting as mentioned above. Mathematically, Proposition 35 is concerned with the stability of the stochastic differential equation (3.48) with respect to the driving process X. The statement of this proposition is chosen in view of obtaining a straightforward proof and is thus somewhat restricted. It should be noted that a variety of more general stability results of the same type are known in the literature. See the comprehensive book by Protter [57].

Let X^1 and X^2 be any two processes of the form (3.47). Proposition 35 gives an upper bound on the distance between $Y^1 \equiv \mathcal{E}(X^1)$ and $Y^2 \equiv \mathcal{E}(X^2)$ in terms of the distance between X^1 and X^2. For the distance between $\mathcal{E}(X^1)$ and $\mathcal{E}(X^2)$ we have in mind the square mean uniform distance. The adequate distance between X^1 and X^2 is precised in the intermediate Proposition 34 below. The matrix elements of the processes X^1 and X^2 are as follows

$$X^{1\,ij}_t = f^1_{ij}(t) + \int_0^t Q^1_{ij}(s) dB^{ij}_s \qquad X^{2\,ij}_t = f^2_{ij}(t) + \int_0^t Q^2_{ij}(s) dB^{ij}_s \tag{3.76}$$

for $1 \leq i, j \leq d$ and $t \geq 0$. Here, the functions f^1_{ij}, f^2_{ij} and Q^1_{ij}, Q^2_{ij} are as in (3.47). In order to simplify the proof of Proposition 35 the following two assumptions are made. First, note that expressions (3.76) impose that X^1 and X^2 are defined in terms of the same $\mathcal{L}(d)$-valued Brownian motion B. This restricts the range of problems for which we can consider the approach of diagram (3.75) but is for instance sufficient for denoising problems. Second, although the stochastic integral in equation (3.48) is a Stratonovich integral, all stochastic integrals appearing in Propositions 34 and 35 will be treated as Itô integrals. To see that this can be done without any loss of generality, note the following. Transforming equation (3.48) into an Itô equation does not change its general form but simply adds a new term to the driving process X. This new term is an increasing function and can be absorbed into the functions f_{ij} of formula (3.47).

The following Proposition 34 is a technical lemma to be used in the proof of Proposition 35. Readers who prefer to skip its proof may do so after familiarizing themselves with its statement. For all process X of the form (3.47), Proposition 34 exhibits an increasing function $A_X : \mathbb{R}_+ \to \mathbb{R}_+$ which dominates stochastic integrals with respect to X. This proposition introduces the following notation. Let Z be a $\mathcal{L}(d)$-valued continuous process. For $T \geq 0$ we note Z^*_T the $\mathcal{L}(d)$-valued random variable whose matrix elements are $Z^{*\,ij}_T = \sup_{t \leq T} |Z^{ij}_T|$.

Proposition 34 *Let X be a process of the form (3.47) and note $\mathcal{F}$ its natural filtration. There exists an increasing function $A_X : \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $\mathcal{L}(d)$-valued process Y which is continuous, square integrable and $\mathcal{F}$-adapted we have*

$$\mathbb{E}|Z^*_T|^2 \leq \int_0^T \mathbb{E}|Y_t|^2 dA_X(t) \tag{3.77}$$

for $T \geq 0$. Where Z is the stochastic integral process, $Z_t = \int_0^t Y_s dX_s$ for $t \geq 0$.

Proof: In the proof, the notation of (3.47) is used. For $T \geq 0$ we have by direct calculation

$$\mathbb{E}|Z^*_T|^2 \leq 2d \sum_{i,j,k=1}^{d} \mathbb{E} \sup_{t \leq T} \left| \int_0^t Y^{ik}_s df_{kj}(s) \right|^2 + \mathbb{E} \sup_{t \leq T} \left| \int_0^t Y^{ik}_s Q_{kj}(s) dB^{kj}_s \right|^2 \tag{3.78}$$

For the first term under the sum we have by the Cauchy-Schwarz inequality

$$\left|\int_0^T |Y_t^{ik}| d|f_{kj}|(t)\right|^2 \leq \int_0^\infty \frac{d|f_{kj}|(t)}{(1+|f_{kj}|^2(t))} \int_0^T |Y_t^{ik}|^2 (1+|f_{kj}|^2(t)) d|f_{kj}|(t)$$

where $|f_{kj}| : \mathbb{R}_+ \to \mathbb{R}_+$ denotes the variation of the function f_{kj}. It follows that

$$\mathbb{E} \sup_{t \leq T} \left|\int_0^t Y_s^{ik} df_{kj}(s)\right|^2 \leq \frac{\pi}{2} \int_0^T \mathbb{E}|Y_t^{ik}|^2 (1+|f_{kj}|^2(t)) d|f_{kj}|(t)$$

For the second term we have by Doob's quadratic martingale inequality

$$\mathbb{E} \sup_{t \leq T} \left|\int_0^t Y_s^{ik} Q_{kj}(s) dB_s^{kj}\right|^2 \leq 4\mathbb{E}|B_1|^2 \int_0^T \mathbb{E}|Y_t^{ik}|^2 Q_{kj}^2(t) dt$$

Replacing these two inequalities in (3.78) and summing over i, j, k we have that (3.77) holds for the following function A_X

$$A_X(t) = 12d^2 \sum_{i,j=1}^d \left[|f_{ij}|(t) + \frac{|f_{ij}|^2(t)}{3} + \mathbb{E}|B_1|^2 \int_0^t Q_{ij}^2(s) ds\right] \tag{3.79}$$

for $t \geq 0$, which is indeed a positive increasing function.▲

We now give Proposition 35. For the processes X^1 and X^2 of (3.76), note X the process where $X_t = X_t^1 - X_t^2$ for $t \geq 0$. It follows from (3.76) that X is also of the form (3.47). Let A_X be the positive increasing function associated to X as in Proposition 34. Proposition 35 will bound the uniform mean square distance between Y^1 and Y^2 in terms of the function A_X. Note that this function is a strong measure of the distance between X^1 and X^2. Indeed, it follows from Proposition 34 that

$$\mathbb{E}|X_T^*|^2 \leq A_X(T)$$

for all $T \geq 0$. After proving Proposition 35, we will discuss it in terms of the approach of diagram (3.75).

Proposition 35 *Note $D = (D_t)_{t \geq 0}$ the process where $D_t = Y_t^1 - Y_t^2$. For $T \geq 0$ we have the following inequality*

$$\mathbb{E}|D_T^*|^2 \leq 4d \exp[2A_{X^1}(T) + 2A_{X^2}(T)] A_X(T) \tag{3.80}$$

where A_{X^1} and A_{X^2} are the positive increasing functions associated to X^1 and X^2 as in Proposition 34.

Proof: Before considering the process D, let us start by obtaining bounds on the processes Y^1 and Y^2. We do this using Proposition 34. Letting $i = 1, 2$, and applying Proposition 34 to equation (3.48) for Y^i we have for $T \geq 0$

$$\mathbb{E}|{Y^i}_T^*|^2 \leq 2d + 2 \int_0^T \mathbb{E}|{Y^i}_t^*|^2 dA_{X^i}(t)$$

It results by elementary calculus

$$\mathbb{E}|{Y^i}_T^*|^2 \leq 2d \exp[2A_{X^i}(T)] \tag{3.81}$$

Turning to the process D, note that we can write[3]

$$D_T = \int_0^T Y_t^1 dX_t + \int_0^T D_t dX_t^2$$

[3]These two integrals are well-defined since all processes are adapted to the natural filtration of the underlying Brownian motion B of (3.76). For the same reason, it is possible to use Proposition 34 for these integrals.

for $T \geq 0$. By applying Proposition 34 to both integrals we have

$$\mathbb{E}|D_T^*|^2 \leq 4d \exp[2A_{X^1}(T)]A_X(T) + 2\int_0^T \mathbb{E}|D_T^*|^2 dA_{X_2}(T)$$

where we have used inequality (3.81) for the first term. It now follows as for (3.81) that

$$\mathbb{E}|D_T^*|^2 \leq 4d \exp[2A_{X^1}(T)] \exp[2A_{X^2}(T)]A_X(T)$$

which completes the proof.▲

Let us now return to the notation of diagram (3.75). We wish to discuss the approximation $\hat{Y}^1 \approx Y^1$ using Proposition 35. In order to do so we must assume that the processes $\hat{X}^1$ and X^1 verify the conditions of this proposition. We emphasize that while these conditions contain strong restrictions they were only introduced to simplify the proof. The general behavior which we describe remains effective in a considerably more general range of problems.

Note $E = (E_t)_{t\geq 0}$ the process $E_t = \hat{Y}_t^1 - Y_t^1$. This process gives the error arising from our use of diagram (3.75) to approximate Y^1 from Y^2. Note also $e = (e_t)_{t\geq 0}$ the process $e_t = \hat{X}_t^1 - X_t^1$. According to the diagram, we control directly the process e as we construct the approximation $\hat{X}^1 \approx X^1$. The performance of our approach depends on the possibility of dominating the process E using only the process e. Proposition 35 provides the following bound (3.82) on E. Let $A_{\hat{X}^1}, A_{X_1}$ and A_e the positive increasing functions associated to the processes $\hat{X}^1, X^1$ and e as in Proposition 34. We rewrite inequality (3.80) for these processes.

$$\mathbb{E}|E_T^*|^2 \leq 4d \exp[2A_{\hat{X}^1}(T) + 2A_{X^1}(T)]A_e(T) \tag{3.82}$$

Inequality (3.82) illustrates the roles of stability and drifting in determining the performances of the approach of diagram (3.75). Note that this can be rewritten

$$\mathbb{E}|E_T^*|^2 \leq K(T)A_e(T)$$

for $T \geq 0$. Clearly, the error E is dominated by the function A_e associated with the process e. Fixing $T \geq 0$, any improvement in the approximation $\hat{X}^1 \approx X^1$ thus leads to an improvement in the approximation $\hat{Y}^1 \approx Y^1$, modulated by a multiplicative factor K. Drifting is precisely related to the exponential growth of the factor K in relation with the time T. For larger values of T, in order to obtain a fixed performance of the approximation $\hat{Y}^1 \approx Y^1$ a much higher quality of the approximation $\hat{X}^1 \approx X^1$ might be required. We have described drifting problems in the discrete time setting of Paragraph 3.1.1.2. Here, we have used Proposition 35 to display these problems mathematically in the case of Brownian matrix processes. Thus, drifting appears a general limitation to the performances of methods based on local linearization.

3.3 Conclusions

In comparison to Chapter 2, the current chapter is of a more exploratory nature. The aim is to give a general formulation of the use of local linearization for filtering under nonlinear constraints and to examine the resulting performances. Our interest in local linearization is due to its consistent and rich mathematical structure in addition to its relative simplicity of use. In the literature, filtering problems involving nonlinear constraints are either solved using *ad hoc* global linearization methods or complicated nonlinear optimization techniques. While there have been isolated instances of applications of local linearization in filtering, *e.g.* in motion capture, we do not know of any general study of such

applications. From the outset, we identify the stability and invariance properties of local linearization as essential to its successful application and wish to gain a theoretical understanding of their role. Section 3.1 takes a practical approach to this goal, while Section 3.2 attempts a more general and theoretical study.

Section 3.1 is based on our paper [64]. This Section 3.1 restricts itself to rotation time series for which it develops in detail the stability properties of local linearization. These properties are applied to the study of two examples of filtering problems. In particular, the problem of geodesic regression in considered in Paragraph 3.1.2.1. We had considered this same problem from an optimization point of view in the paper [65]. Although [65] considers geodesic regression in relation to a data analysis application, the papers [64] and [65] provide a first basis for the comparison of our local linearization based approach to optimization methods. Paragraph 3.1.1.2 includes the main mathematical development of this section. The stability of local linearization is given in the form of both global Lipschitz properties and local differentiability results. This development is a necessary step for any systematic study of local linearization based methods for rotation time series.

Section 3.2 attempts to generalize the approach of Section 3.1 to Brownian matrix processes. These processes are described by a class of linear stochastic differential equations of which we had already met an example in Chapter 2 –namely, rotation Brownian motion in Subsection 2.3.3. Brownian matrix processes constitute a general model for processes with values in matrix Lie groups which have independent increments and continuous paths. Using their defining stochastic differential equations, Subsection 3.2.3 naturally introduces a local linearization transformation for Brownian matrix processes. Section 3.2 is devoted to providing a self-contained background on the underlying mathematical formalism. In doing so, we establish the main results which allow the formulation of our approach in Subsection 3.2.5. Concrete application and a more detailed study of this approach remain open goals for future work.

In terms of this approach, Propositions 26 and 30 establish the use of local linearization in dealing with nonlinear constraints. Proposition 29 characterizes the properties of independence and stationarity of the increments of Brownian matrix processes. As seen in Paragraph 3.1.2.1, these properties are important to estimation problems stated in terms of these processes. Proposition 35 is used in Subsection 3.2.5 to illustrate the performances of our approach based on local linearization. Although we only provide a general discussion, Proposition 35 already clarifies the importance of the stability properties related to local linearization and allows us to describe the role of drifting problems.

Let us end by noting a remarkable result obtained in Paragraph 3.1.2.1. In the presence of a nontrivial noise model, this paragraph shows that a problem of averaging on the rotation group can be solved using a straightforward analytical formula. In current literature, this type of averaging problem is considered through optimization techniques and its solution is obtained using a gradient algorithm. We have obtained this considerable simplification thanks to our formalism of Brownian matrix processes.

Appendix A

Résumé en Français

La présente annexe contient un résumé en Français des résultats les plus importants de cette thèse. Ainsi, nous reprenons en les condensant les parties correspondantes dans le corps de la thèse. Les sections A.2 et A.3 ci-dessous sont respectivement issues des chapitres 2 et 3. La section A.1 commence par rappeler l'objectif général de la thèse ainsi que les problèmes les plus importants qu'elle a résolu. C'est en fait une traduction directe de l'introduction Anglaise.

A.1 Introduction

Cette thèse considère des problèmes d'estimation et de filtrage posés en termes de processus ayant leurs valeurs dans des groupes de Lie matriciels. Notre intérêt pour ces processus est motivé par des applications directes mais aussi par leur importance générale pour la modélisation. La thèse a été divisée en deux chapitres. Le chapitre 2 se restreint aux processus ayant leurs valeurs dans le groupe des rotations. Nous parlons, pour plus de concision, de processus de rotation. Le problème principal résolu dans ce chapitre est le problème du *decompounding*[1], qui est un problème d'estimation non paramétrique. La solution en est réalisée dans la sous-section 2.4.2. Le chapitre 3 considère de façon plus générale les processus avec leurs valeurs dans un groupe de Lie matriciel quelconque. Ce chapitre étudie une nouvelle approche du filtrage de ces processus, basée sur la notion de linéarisation locale. La formulation précise de cette approche est réalisée à la sous-section 3.2.5.

Le chapitre 2 est basé sur nos deux articles [63,66]. Il est consacré à la partie du travail de thèse qui relève d'applications directes, notamment en physique des ondes. La section 2.2 présente un nouveau formalisme pour la polarisation statistique [63]. La section 2.4 pose le problème du *decompounding* en tant qu'alternative statistique au problème inverse de la diffusion multiple. Sous une forme plus générale, ce problème fut considéré et résolu dans [66] –A notre connaissance, il s'agit de la première fois que ce problème est considéré.

Le outils mathématiques nécessaires aux sections 2.2 et 2.4 sont donnés aux sections 2.1 et 2.3. Nous devons essentiellement présenter la théorie des fonctions caractéristiques de variables aléatoires de rotation. L'utilisation de ces fonctions caractéristiques connait récemment une popularité croissante, tout particulièrement pour l'estimation non paramétrique –voir [67] et la discussion faite à la section 2.1. Notre présentation en introduit certains nouveaux aspects importants aux applications des sections 2.2 et 2.4. Par exemple, le paragraphe 2.1.2.2 caractérise les propriétés de symétrie des variables aléatoires de rotation, qui sont à leur tour un ingrédient principal dans notre solution du

[1] Nous ne connaisons pas de nom Français pour ce problème et garderons tout le long son nom Anglais.

problème du *decompounding*. La section 2.3 étudie une classe importante de processus de rotation: Les processus de Lévy de rotation.

La section 2.2 considère un problème d'importance récente en polarisation statistique. Le formalisme de Stokes classique en polarisation est basé seulement sur les statistiques d'ordre deux du champ d'onde optique et notre but est de le généraliser aux statistiques d'ordre supérieur. Les expériences d'Ellis et Dogariu que nous décrivons à la section 2.2.2 peuvent être vues comme notre motivation principale. A la sous-section 2.2.3 nous proposons un nouveau formalisme qui généralise le formalisme de Stokes classique aux statistiques d'ordre supérieur. Ce formalisme réussit à clarifier les contradictions levées par les expériences de la sous-section 2.2.2. Nous comparons ce formalisme à d'autres tentatives faites dans la littérature pour inclure les statistiques d'ordre supérieur dans la description de la polarisation. La section 2.2.4 utilise notre formalisme afin d'étudier le problème physique de la dépolarisation.

La section 2.4 pose et résout le problème du *decompounding*. La sous-section 2.4.1 discute de ce problème en tant qu'alternative au problème inverse de la diffusion multiple. La sous-section 2.4.2 étudie le problème du *decompounding* en lui même. C'est un problème d'estimation non paramétrique posé en termes des processus de Poisson composés de rotation et des processus entrelacés, qui ont été introduits aux sous-sections 2.3.2 et 2.3.4. Nous proposons une solution du problème du *decompounding* utilisant une méthode de fonction caractéristique. Alors que l'utilisation de telles méthodes pour l'estimation paramétrique des variables aléatoires de rotation est déjà bien établie, le problème du *decompounding* présente une nouvelle difficulté essentielle en vue de ces méthodes. Plus précisément, ce problème part d'observations indirectes et sa solution exige en conséquence des outils spécifiques en analyse matricielle ainsi qu'un traitement probabiliste plus détaillé. Le paragraphe 2.4.2.2 fournit les preuves mathématiques pour la convergence de notre méthode de fonction caractéristique. Le paragraphe 2.4.2.3 illustre cette convergence à l'aide de simulations numériques.

Le chapitre 3 est d'une nature plus exploratoire que le chapitre 2. Ce chapitre considère les processus ayant leurs valeurs dans un groupe de Lie matriciel quelconque. Dans ce cas général, il n'existe pas de définition utile de la fonction caractéristique comme au chapitre 2. Ainsi, les processus en question sont considérés à l'aide des équations différentielles stochastiques qui les définissent. Nous nous intéressons aux processus à valeurs dans les groupes de Lie matriciels en tant que modèle général pour les processus sous contraintes non linéaires, ces contraintes exprimant des considérations de symétrie. De tels processus apparaissent dans une grande variété de domaines : Capture de mouvement, dynamique de l'ADN, imagerie médicale. Nous nous adressons au problème appliqué du filtrage sous contraintes non linéaires. La difficulté principale à surmonter est l'impossibilité d'appliquer les opérations de traitement du signal linéaires tout en respectant les contraintes non linéaires.

En pratique, cette difficulté est souvent contournée à l'aide des méthodes dite de linéarisation globale. De telles méthodes permettent une implémentation facile mais souffrent de plusieurs limitations –ceci a été discuté par Xavier et Manton [75] et Lee et Shin [38]. Motivés par la réussite de son application en capture de mouvement [16, 38], nous nous intéressons à la notion de linéarisation locale. Le but du chapitre 3 est de donner une formulation générale de l'utilisation de la linéarisation locale pour le filtrage sous contraintes non linéaires et de justifier les performances résultantes. La section 3.1 est basée sur notre article [64] et suit une approche pratique de ce but. La linéarisation locale est considérée pour le cas spécial des processus de rotation en temps discret. Nous étudions, pour ce cas précis, les propriétés de stabilité et d'invariance de la linéarisation locale. A la sous-section 3.1.2 nous donnons deux exemples numériques de problèmes de filtrage pour lesquels nous présentons notre méthode en comparaison avec d'autres travaux récents, notamment basés sur l'optimisation.

La section 3.2 correspond au but principal de ce chapitre. Nous considérons des processus qui peuvent être représentés comme solutions d'équations stochastiques différentielles linéaires de la forme suivante.

$$dY_t = Y_t dX_t \qquad Y_0 = I_d$$

Notre processus inconnu Y et le processus dirigeant X ont leurs valeurs dans l'espace des matrices $d \times d$ réelles, pour un certain $d \geq 1$. De plus, X est un processus Brownien à accroissements indépendants et l'équation est considérée comme un équation de Stratonovich. Pour la condition initiale, I_d dénote la matrice identité $d \times d$. Nous appelons les processus de la forme Y des processus matriciels Brownien. La section 3.2 développe le formalisme mathématique nécessaire à la formulation correcte et générale de l'utilisation de la linéarisation locale pour les processus matriciels Brownien.

La sous-section 3.2.1 montre que les processus matriciels Browniens ont de façon naturelle leurs valeurs dans les groupes de Lie matriciels. Ainsi, ils vérifient le type général des contraintes non linéaire que nous voulons exprimer. A la sous-section 3.2.3, la correspondance $Y \mapsto X$ pour un processus matriciel Brownien Y défini comme ci-dessus est considérée. C'est une transformation bien définie dont nous montrons qu'elle transforme le processus Y sujet à des contraintes non linéaires en le processus X sujet uniquement à des contraintes linéaires. La linéarisation locale d'une processus matriciel Brownien Y consiste précisément en l'application de la transformation $Y \mapsto X$. Elle peut être calculée de façon causale –en se servant seulement des valeurs courantes de Y– et élimine les contraintes non linéaires imposées sur Y. La sous-section 3.2.5 donne finalement la formulation précise de notre approche du filtrage des processus matriciels Brownien et énonce les propriétés de stabilité sous-jacentes. Ces propriétés constituent la justification principale de notre approche. Il est important de rappeler que la chapitre 3 a le but limité de formuler et d'étudier de façon générale l'utilisation de la linéarisation locale pour les processus matriciels Browniens. Bien que nous réalisions ce but à la section 3.2, l'application concrète et l'étude détaillée de notre approche reste un objectif ouvert pour des travaux futurs.

A.2 Problème du *decompounding*

Le problème principal résolu dans le chapitre 2 est le problème du *decompounding*. Nous allons ici poser ce problème et en présenter la solution qui a été proposée à la sous-section 2.4.2. En passant nous rappellerons au A.2.1 quelques bases sur les fonctions caractéristiques des variables aléatoires de rotation. Au A.2.2 nous reverrons rapidement la définition des processus de Poisson composés de rotation. C'est là un exemple de processus de Lévy de rotation et nous renvoyons à la section 2.3 pour une étude plus complète de ce type de processus. A la section 2.4 le problème du *decompounding* a été appliqué au problème inverse de la diffusion multiple –voir plus précisément la sous-section 2.4.1. Nous n'allons pas ici rendre compte de cet aspect, et le résumé suivant est consacré au problème du *decompounding* en lui même.

A.2.1 Fonctions caractéristiques des variables aléatoires de rotation

Les fonctions caractéristiques de variables aléatoires scalaires ou vectorielles sont définies grâce à la transformée de Fourier classique. Leur généralisation aux variables aléatoires de rotation se fait grâce à l'analyse de Fourier sur le groupe des rotations. Notre présentation sur les fonctions caractéristiques est basée sur [21, 40] et reprend aussi certains travaux récents [36, 76].

Le groupe des rotations de l'espace est identifié avec le groupe matriciel $SO(3)$. Une matrice 3×3 réelle R appartient à $SO(3)$ si elle vérifie les deux conditions suivantes

$$RR^T = I \qquad \det R = 1 \tag{A.1}$$

où I est la matrice identité. Une variable aléatoire de rotation X est une matrice aléatoire 3×3 réelle tel que en plus X appartient à $SO(3)$. Par densité de probabilité d'une variable aléatoire de rotation X, nous entendons toujours une densité de probabilité carré intégrable par rapport à la mesure de Haar de $SO(3)$. La mesure de Haar est notée μ et sa forme explicite est donnée au paragraphe 2.1.1.3. C'est l'unique mesure normalisée sur $SO(3)$ qui soit invariante par multiplication. Si p est la densité de probabilité de X alors p est carré intégrable par rapport à μ et pour toute fonction continue $h : SO(3) \to \mathbb{C}$

$$\mathbb{E}[h(X)] = \int_{SO(3)} phd\mu$$

Ici $\mathbb{E}$ est l'espérance par rapport à une probabilité sous-jacente $\mathbb{P}$.

Pour définir la fonction caractéristique de X nous devons introduire les représentations irréductibles du groupe $SO(3)$ et énoncer le théorème de Peter-Weyl. C'est le théorème 5 ci-dessous. Ces représentations forment en faite une famille d'homomorphismes U^l indexée par $l \geq 0$. Pour $l \geq 0$ nous avons une application $R \mapsto U^l(R)$ continue et définie pour $R \in SO(3)$ vérifiant la propriété suivante. Pour $R \in SO(3)$ l'image $U^l(R)$ est une matrice $d_l \times d_l$ complexe. Ici $d_l = 2l + 1$, par convention. De plus, pour tout $R_1, R_2, R \in SO(3)$ nous avons

$$U^l(R_1R_2) = U^l(R_1)U^l(R_2) \qquad [U^l(R)]^{-1} = U^l(R^T) = [U^l(R)]^\dagger \tag{A.2}$$

où † dénote la transposée Hermitienne. La première identité est la propriété d'homomorphisme de U^l, alors que la deuxième implique que U^l est à valeurs unitaires. Les formes explicites des applications U^l, pour $l \geq 0$, sont données au paragraphe 2.1.1.3, au moyen des angles d'Euler.

Bien entendu, étant donné $l \geq 0$ nous pouvons considérer d_l^2 fonctions sur $SO(3)$ et à valeurs complexes données par les éléments de la matrice $U^l(R)$. Il est habituel d'indexer ces fonctions par $-l \leq m, n \leq l$ en les notant $U^l_{mn} : SO(3) \to \mathbb{C}$. Le théorème de Peter-Weyl nous dit que la famille de fonctions ainsi définie et indexée par $l \geq 0$ et $-l \leq m, n \leq l$ est complète dans l'espace des fonctions carré intégrables par rapport à la mesure de Haar. La série (A.3) est dite série de Fourier de la fonction h en question.

Theorem 5 *Pour toute fonction h carré intégrable par rapport à μ la série suivante (A.3) converge vers h. Si h est continue, alors (A.3) converge vers h uniformément.*

$$h = \sum_{l \geq 0} d_l \operatorname{tr}\left(\hat{h}^l {U^l}^\dagger\right) = \sum_{l \geq 0} d_l \sum_{m,n=-l}^{l} \hat{h}^l_{mn} U^{l^*}_{mn} \tag{A.3}$$

où tr *dénote la trace d'une matrice.*

Bien que le théorème de Peter-Weyl n'intervienne pas dans la définition de la fonction caractéristique d'une variable aléatoire de rotation, elle est à la base de ses propriétés les plus importantes. Soit X une variable aléatoire de rotation. La fonction caractéristique de X est la suite $\phi_X = \{\phi_X(l)\}_{l \geq 0}$ de matrices $d_l \times d_l$ complexes données par

$$\phi_X(l) = \mathbb{E}[U^l(X)] \tag{A.4}$$

La proposition 36 suivante condense les propositions 2 et 3 du paragraphe 2.1.2.1. De même, la proposition 37 reprend les propositions 5 et 6 du paragraphe 2.1.2.2. Les démonstrations des propositions 36 et 37 se font de façon directe grâce au théorème de Peter-Weyl et aux propriétés (A.2).

La proposition 36 rappelle la relation entre les fonctions caractéristiques et les notions de convolution et de convergence en distribution. C'est une généralisation de propriétés classiques pour les variables aléatoires scalaires. Rappelons qu'une suite $(X_n)_{n\geq 1}$ de variables aléatoires de rotation et dite convergente en distribution vers une variable aléatoire X si pour toute fonction continue $h : SO(3) \to \mathbb{C}$ nous avons

$$\lim_n \mathbb{E}(h(X_n)) = \mathbb{E}(h(X))$$

Cette propriété s'écrit $X_n \xrightarrow{d} X$.

Proposition 36 *Nous avons les propriétés suivantes*
(i) *Soient X et Y des variables aléatoires de rotation indépendantes et $Z = XY$. Nous avons pour $l \geq 0$*

$$\phi_Z(l) = \phi_X(l)\phi_Y(l)$$

(i) *Une suite $(X_n)_{n\geq 1}$ de variables aléatoires de rotation converge en distribution vers une variable aléatoire X ssi pour tout $l \geq 0$*

$$\lim_n \phi_{X_n}(l) = \phi_X(l)$$

Afin de résoudre le problème du *decompounding* nous imposerons certaines conditions de symétrie sur les variables aléatoires considérées. Nous détaillons ces propriétés ici en utilisant les fonctions caractéristiques, l'analyse suivante est adaptée à partir de [40].

Nous dirons qu'une variable aléatoire de rotation X est invariante pas inversion si $X \overset{d}{=} X^T$. De même, nous dirons que X est invariante par conjugaison si pour tout $R \in SO(3)$ nous avons $X \overset{d}{=} RXR^T$. Comme d'habitude, $\overset{d}{=}$ dénote l'égalité en distribution. Rappelons que pour deux variables aléatoires de rotation X et Y nous avons $X \overset{d}{=} Y$ *ssi* $\phi_X = \phi_Y$. Cette caractérisation est un résultat immédiat du théorème de Peter-Weyl. Elle peut être utilisée pour démontrer la proposition suivante.

Proposition 37 *Soient X et Y des variables aléatoires de rotation indépendantes. Nous avons les propriétés suivantes*
(i) *X est invariante par inversion ssi pour tout $l \geq 0$ nous avons que $\phi_X(l)$ est Hermitienne.*
(ii) *Si X est invariante par inversion et $X_1, \ldots, X_n$ sont des copies indépendantes de X alors le produit $X_1 \ldots X_n$ est invariant par inversion.*
(iii) *X est invariante par conjugaison ssi pour tout $l \geq 0$ nous avons $\phi_X(l) = a_l I_l$ où $a_l \in \mathbb{R}$ et I_l est la matrice identité $d_l \times d_l$.*
(iv) *Si X et Y sont invariantes par conjugaison alors XY est invariante par conjugaison.*
(v) *X est invariante par conjugaison ssi $XY \overset{d}{=} YX$.*

A.2.2 Processus de Poisson composés

Les processus de Poisson composés de rotation généralisent naturellement les processus de Poisson composés scalaires. Ils ont été introduits dans [2]. Rappelons d'abord la définition d'un processus de Poisson composé scalaire. Soit $N = (N_t)_{t\geq 0}$ un processus de Poisson de paramètre $\lambda > 0$. Soient $(x_n)_{n\geq 1}$ des variables aléatoires réelles et *i.i.d.*. Si la famille $(x_n)_{n\geq 1}$ est elle même indépendante de

N alors le processus y suivant est un processus de Poisson composé

$$y_t = \sum_{n=0}^{N_t} x_n$$

où nous posons $x_0 = 0$. Les processus de Poisson de rotation se définissent de façon identique. Replaçons la famille $(x_n)_{n\geq 1}$ par une famille $(X_n)_{n\geq 1}$ de variables aléatoires de rotation *i.i.d.* et qui soit elle aussi indépendante de N. Le processus Y suivant est un processus de Poisson composé de rotation

$$Y_t = \prod_{n=0}^{N_t} X_n$$

Par convention, le produit est ici ordonné de gauche à droite et $X_0 = I$. Il est possible d'obtenir un processus de Poisson composé *droit* en considérant le processus Y^T. Les deux notions sont équivalentes. Voir [2, 40].

Nous résumons maintenant les propriétés de symétrie de la variable aléatoire Y_t pour un $t \geq 0$ fixé. Notons d'abord que Y_t n'a pas de densité de probabilité par rapport à μ. En effet, quelque soit t nous avons $\mathbb{P}(Y_t = I) \leq \mathbb{P}(N_t = 0) = e^{-\lambda t}$. Il suit que Y_t a un atome en I. En l'absence de densité de probabilité, nous étudions Y_t à l'aide de sa fonction caractéristique. Ceci est fait dans la proposition 38 suivante qui est une généralisation immédiate de la formule connue pour les processus de Poisson composés scalaires.

Proposition 38 *Soit le processus $Y = (Y_t)_{t\geq 0}$ défini comme ci-dessus. Pour $t \geq 0$ nous avons*

$$\phi_t(l) = \exp[\lambda t(\phi(l) - I_l)] \tag{A.5}$$

pour $l \geq 0$, où $\phi_t \equiv \phi_{Y_t}$ et $\phi \equiv \phi_{X_1}$.

Démonstration : La démonstration est basée sur un calcul classique. L'idée est de conditionner sur les valeurs de N_t. Comme N et les $(X_n)_{n\geq 1}$ sont indépendants, on peut écrire pour $l \geq 0$

$$\phi_t(l) = e^{-\lambda t} \sum_{n\geq 0} \frac{(\lambda t)^n}{n!} \mathbb{E} \prod_{m=0}^{n} U^l(X_m)$$

Comme les $(X_n)_{n\geq 1}$ sont *i.i.d.* il est possible de remplacer

$$\mathbb{E} \prod_{m=0}^{n} U^l(X_m) = \prod_{m=0}^{n} \mathbb{E}(U^l(X_m)) = \phi(l)^n$$

la proposition s'obtient en réarrangeant la somme.▲

En se servant des propositions 37 et 38 nous avons la proposition suivante. Elle stipule que pour $t \geq 0$ les propriétés de symétrie de Y_t sont les mêmes que celles des X_n.

Proposition 39 *Soit le processus $Y = (Y_t)_{t\geq 0}$ défini comme ci-dessus. Pour $t \geq 0$ nous avons*
(i) *Si X_1 est invariante par inversion alors Y_t est invariante par inversion.* (ii) *Si X_1 est invariante par conjugaison alors Y_t est invariante par conjugaison.*

Pour finir notre présentation des processus de Poisson composés de rotation, donnons la proposition 40. Cette proposition concerne l'uniformisation de la distribution de Y_t quand $t \uparrow \infty$. Cette propriété est similaire au comportement des produits $X_1 \dots X_n$ quand $n \uparrow \infty$ qui a été étudié à la sous-section 2.1.4. Nous disons qu'une variable aléatoire de rotation X est supportée par un ensemble mesurable

$B \subset SO(3)$ si $\mathbb{P}(X \in B) = 1$. Si X et X' sont des variables aléatoires de rotation telles que $X \stackrel{d}{=} X'$ alors X est supportée par B *ssi* X' est supportée par B. Pour la proposition 40, U sera une variable aléatoire de rotation de densité de probabilité constante égale à 1. Autrement dit, U est uniformément distribuée sur $SO(3)$.

Proposition 40 *Si X_1 n'est supportée par aucun sous-groupe propre fermé G de $SO(3)$ ou classe RG, $R \in SO(3)$ d'un tel sous-groupe alors $Y_t \stackrel{d}{\to} U$ quand $t \uparrow \infty$.*

La démonstration de la proposition 40 est immédiate en utilisant le (ii) de la proposition 36 et la proposition 38.

A.2.3 Le problème du *decompounding*

Dans la littérature existante le mot *decompounding* fait référence à un ensemble de problèmes d'estimation non paramétrique concernant les processus de Poisson composés scalaires [8, 72]. L'application principale de ces problèmes concerne les problèmes de fils d'attente et à la théorie du risque [8]. Nous allons formuler le problème du *decompounding* pour les processus de Poisson composés de rotation et en présenter une solution complète. Cette solution est la même que nous avons donné dans [66]. Elle est basée sur une méthode de fonction caractéristique.

Soit Y un processus de Poisson composé de rotation défini comme au A.2.2. En particulier, λ est le paramètre du processus de Poisson N définissant Y. Le problème du *decompounding* consiste en l'estimation de la densité de probabilité commune des variables aléatoires de rotation X_n. Nous supposons que cette densité existe en effet et la notons p. L'estimation doit se faire à partir d'observations du processus Y. Plusieurs versions du problème du *decompounding* existent, selon le type d'observations faites du processus Y, voir [66]. Nous sommes intéressés par la version suivante qui correspond au problème physique de la sous-section 2.4.1. Nous fixons $T \geq 0$ et supposons données des observations *i.i.d* $(Z_n)_{n\geq 1}$ d'une version bruitée Z de Y_T. Notre but est d'estimer p à partir des $(Z_n)_{n\geq 1}$. Z est lié à Y par un modèle de bruit multiplicatif

$$Z = MY_T \tag{A.6}$$

où M et Y_T sont indépendantes. Par le (i) de la proposition 36 nous avons la fonction caractéristique de Z

$$\phi_Z = \phi_M \phi_{Y(T)}$$

Ce modèle de bruit est équivalent à une condition initiale $Y_0 = M$ au lieu de $Y_0 = I$. Nous considérons le cas d'un bruit Brownien. La fonction caractéristique de M est donné comme à la proposition 19 de la sous-section 2.3.3

$$\phi_M(l) = \exp -\frac{\sigma^2}{2} l(l+1) I_l$$

où σ^2 est un paramètre de variance. Ceci correspond à un mouvement Brownien invariant par conjugaison. En particulier, grâce au (v) de la proposition 37, la multiplication à gauche ou à droite par M a le même effet quant à la distribution de Z. Il est aussi possible de faire une interprétation de ce modèle de bruit basée sur la construction d'entrelacement décrite à la sous-section 2.3.4.

Nous allons donner dans A.2.4 et A.2.5 notre solution du problème du *decompounding* posé ici. Cette solution reflète une difficulté qui n'existe pas pour les processus de Poisson composés scalaires. Sous les conditions de la proposition 40, si les observations $(Z_n)_{n\geq 1}$ correspondent à un temps T suffisamment grand alors ces observations seront uniformément distribuées et donc sans mémoire des variables aléatoires de rotation X_n.

A.2.4 Une méthode de fonction caractéristique

Nous présentons une méthode de fonction caractéristique pour le problème du *decompounding*. La solution procède des observations $(Z_n)_{n\geq 1}$ définies par le modèle de bruit (A.6). La densité inconnue p est donnée par la série de Fourier suivante

$$p = \sum_{l\geq 0} d_l \operatorname{tr}\left(\phi_X(l) U^{l\dagger}\right) \tag{A.7}$$

où $\phi_X \equiv \phi_{X_1}$. Une méthode de fonction caractéristique consiste en la construction d'une estimée non paramétrique de p à partir d'estimées paramétriques pour les $\phi_X(l)$, $l \geq 0$.

Nous supposons connus λ et σ^2. La proposition 38 avec le (i) de la proposition 36 donnent pour la fonction caractéristique de Z définie dans (A.6)

$$\phi_Z(l) = \exp[T\lambda\phi_X(l) - T\bar{\lambda} I_l] \tag{A.8}$$

pour $l \geq 0$. Ici $\bar{\lambda}$ est une constante déterminée par λ et σ^2. Nous appelons cette transformation $\phi_X \mapsto \phi_Z$ la transformation de *compounding*[2]. Le *decompounding* consistera en une inversion locale de la transformation de compounding. Ceci est clairement lié à l'inversion locale de l'exponentielle de matrice au voisinage de $\phi_Z(l)$ pour $l \geq 0$. Plutôt que de traiter ce problème de façon générale, nous le simplifions en faisant l'hypothèse suivante

Hypothèse: X_1 est invariante par inversion.

Pour tout $l \geq 0$ nous avons par le (i) de la proposition 37 et par la proposition 15 que $\phi_X(l)$ et $\phi_Z(l)$ sont toutes les deux Hermitiennes. De plus, il est claire par (A.8) que $\phi_Z(l)$ est définie positive. Nous noterons Log l'unique logarithme de matrice Hermitien d'une matrice Hermitienne définie positive. Nous pouvons maintenant exprimer l'inverse de la transformation de *compounding*. De l'équation (A.8) il suit que

$$\phi_X(l) = \frac{1}{T\lambda} \log[\phi_Z(l)] + (\bar{\lambda}/\lambda) I_l \tag{A.9}$$

Soit $l \geq 0$. Il suit de la définition (A.4) de la fonction caractéristique que les estimées empiriques de $\phi_Z(l)$ basées sur les observations $(Z_n)_{n\geq 1}$ sont non biaisées et consistentes. Ceci est une simple conséquence de la loi forte des grands nombres. Voir par exemple [27]. Afin d'estimer $\phi_X(l)$ à partir de (A.9) il est alors important de s'assurer que les estimées empiriques de $\phi_Z(l)$ soient asymptotiquement Hermitiennes définies positives.

Commençons en définissant les estimées empiriques $\hat{\phi}_Z^n(l)$ pour $l \geq 0$ et $n \geq 1$

$$\hat{\phi}_Z^n(l) = \frac{1}{2n} \sum_{m=1}^{n} \left(U^l(Z_m) + U^l(Z_m)^\dagger\right)$$

La symétrisation Hermitienne des estimées empiriques est nécessaire pour l'application de (A.9). Comme c'est en plus une opération de projection, elle contribue à une convergence plus rapide des $\hat{\phi}_Z^n(l)$ vers $\phi_Z(l)$.

La dépendance continue du spectre d'une matrice par rapport à ses coefficients est un résultat classique en analyse matricielle. Plusieurs énoncés plus ou moins sophistiqués de ce résultat existent [41]. Pour une version particulièrement simple, voir [71]. Le spectre d'une matrice complexe C sera noté $\lambda(C)$. Pour chaque $l \geq 0$ et $n \geq 1$ définissons l'évènement R_l^n par

$$R_l^n = \{\lambda(\hat{\phi}_Z^n(l)) \subset]0, \infty[\}$$

[2]Encore une fois, par défaut d'un terme Français connu.

Pour $l \geq 0$, la séquence $(R_l^n)_{n\geq 1}$ contrôle la convergence des spectres des estimées empiriques $\hat{\phi}_Z^n(l)$. En particulier,

$$\mathbb{P}(\cup_{n\geq 1} \cap_{m\geq n} R_l^n) = \lim_n \mathbb{P}(\cap_{m\geq n} R_l^n) = 1$$

En utilisant les évènements R_l^n nous pouvons écrire des estimées bien définies de ϕ_X. Nous les notons $\hat{\phi}_X^n(l)$ pour $l \geq 0$ et $n \geq 1$

$$\begin{array}{rcll} \hat{\phi}_X^n(l) & = & 0 & \text{sur} \quad \Omega - R_l^n \\ \hat{\phi}_X^n(l) & = & \frac{1}{T\lambda} \operatorname{Log}\left[\hat{\phi}_Z^n(l)\right] + \left(\bar{\lambda}/\lambda\right) I_l & \text{sur} \quad R_l^n \end{array}$$

Cette expression donne nos estimées paramétriques pour les coefficients de Fourier de p. Nous les utilisons pour construire des estimées non paramétriques sous la forme de séries de Fourier (A.7). Soit $K \geq 0$ et notons pour $l \geq 0$

$$f_l = d_l e^{-Kl(l+1)}$$

Pour $n \geq 1$ et $L \geq 0$ nos estimées non paramétriques $\hat{p}_L^n$ sont données par

$$\hat{p}_L^n = \sum_{l=0}^{L} f_l \operatorname{tr}\left(\hat{\phi}_X^n(l) U^{l\dagger}\right) \tag{A.10}$$

L'indice L correspond à un paramètre de coupure ou de lissage. En effet, un nombre infini de valeurs de l est exclu de la somme (A.10). Lorsque $K > 0$ les coefficients f_l forment un masque de convolution assurant que les estimées $\hat{p}_L^n$ convergeront vers une densité de probabilité différentiable. Nous préciserons davantage le rôle de K au A.2.5.

Il est usuel d'écrire les expressions similaires à (A.10) en termes de noyaux invariants par rotation. Voir [32, 36]. Une telle transformation n'est pas possible ici à cause de la nature indirecte de nos observations. Ceci est en particulier lié à la forme plus compliquée des $\hat{\phi}_X^n(l)$ ci-dessus.

A.2.5 Convergence des estimées paramétriques et non paramétriques

Nous allons ici discuter de la convergence des estimées paramétriques et non paramétriques définies au A.2.4. Notre argument se présentera sous la forme des propositions 41 et 42 ci-dessous. La proposition donne la consistance des estimées paramétriques $\hat{\phi}_X^n(l)$. La proposition 23 en conclut le résultat correspondant pour les estimées non paramétriques $\hat{p}_L^n$.

Pour la proposition 41 nous aurons besoin des inégalités (A.11) et (A.12). Ces deux inégalités expriment des résultats de stabilité pour les valeurs propres de matrices Hermitiennes et pour la fonction de matrice Hermitienne Log. Soient A et B des matrices Hermitiennes $d \times d$, pour un certain $d \geq 1$. Pour $1 \leq i \leq d$ soit α_i et β_i les valeurs propres de A et B respectivement. Supposons qu'elles soient rangées dans un ordre croissant. Nous avons

$$\sum_{i=1}^{d} (\beta_i - \alpha_i)^2 \leq |B - A|^2 \tag{A.11}$$

où $|.|$ est la norme Euclidienne pour les matrices. C'est inégalité est connue sous le nom du théorème de Wielandt-Hoffman. Dans [41], elle est donnée pour A et B réelles et symétriques. Le cas général de A et B Hermitiennes peut être obtenu de ce résultat en utilisant un isomorphisme de réelification.

Supposons que A et B soient définies positives. Pour notre but il sera convenable de supposer que $\lambda(A)$ et $\lambda(B)$ sont tous les deux contenus dans un intervalle $[k, 1]$ pour un certain $k > 0$. Nous avons alors la propriété de Lipschitz suivante

$$|\operatorname{Log}(B) - \operatorname{Log}(A)| \leq \sqrt{d} k^{-2} |B - A| \tag{A.12}$$

Pour obtenir (A.12) il est possible d'exprimer $\mathrm{Log}(A)$ comme suit

$$\mathrm{Log}(A) = \int_0^1 (A - I_d)[t(A - I_d) + I_d]^{-1} dt$$

Cette expression résulte d'une expression similaire pour le logarithme réel appliqué à chaque valeur propre de A. En soustrayant la même expression pour $\mathrm{Log}(B)$, (A.12) s'obtient par des calculs simples.

Proposition 41 *Pour tout $l \geq 0$ nous avons la limite en probabilité $\lim_n \hat{\phi}_X^n(l) = \phi_X(l)$.*

Démonstration : Il suffit de considérer $l > 0$. En effet, $\hat{\phi}_X^n(0) = \phi_X(0) = 1$ pour tout $n \geq 1$. Soit $l > 0$, pour tout $n \geq 1$ nous avons

$$|\hat{\phi}_Z^n(l)|_{op} \leq \frac{1}{2n} \sum_{m=1}^{n} |U^l(Z_m)|_{op} + |U^l(Z_m)^\dagger|_{op} = 1$$

où $|.|_{op}$ est la norme d'opérateur pour les matrices. Passant à la limite, il suit de la loi forte des grands nombres que la même inégalité est vérifiée pour $\phi_Z(l)$. En d'autres termes, toutes les valeurs propres de $\hat{\phi}_X^n(l)$ et de $\phi_X(l)$ sont inférieures à l'unité. Comme $\phi_Z(l)$ est définie positive, il existe $k_l > 0$ tel que $\lambda(\phi_Z(l)) \subset [k_l, 1]$. Pour $n \geq 1$ notons $\tilde{R}_l^n$ l'évènement

$$\tilde{R}_l^n = \{\lambda(\hat{\phi}_Z^n(l)) \subset [k_l/2, 1]\}$$

Par l'inégalité (A.11) nous avons

$$\mathbb{P}(\Omega - \tilde{R}_l^n) \leq \mathbb{P}(|\hat{\phi}_Z^n(l) - \phi_Z(l)| > k_l/2)$$

Comme $\tilde{R}_l^n \subset R_l^n$, il suit de l'inégalité (A.12) que

$$\mathbb{P}(|\hat{\phi}_X^n(l) - \phi_X(l)| > \varepsilon \cap \tilde{R}_l^n) \leq \mathbb{P}(|\hat{\phi}_Z^n(l) - \phi_Z(l)| > k_l^2 \varepsilon / M)$$

pour tout $\varepsilon > 0$, où $M = 4\sqrt{d_l}/T\lambda$.

La preuve peut être complétée par une application usuelle de l'inégalité de Chebychev,

$$\mathbb{P}(|\hat{\phi}_X^n(l) - \phi_X(l)| > \varepsilon) \leq \left(\frac{8 + 2M^2/\varepsilon^2}{n}\right) \left(\frac{\sqrt{d_l}}{k_l^2}\right)^2 \tag{A.13}$$

pour tout $\varepsilon > 0$.▲

La proposition 42 utilise la proposition 41 et le théorème de Peter-Weyl –théorème 5 du A.2.1. Elle implique l'existence de suites $(\hat{p}_k)_{k \geq 1}$ d'estimées non paramétriques données par (A.10) convergeant vert p en probabilité dans l'espace des fonctions carré intégrables sur $SO(3)$. Cette convergence en probabilité signifie que la limite suivante en probabilité a lieu

$$\lim_k \|\hat{p}_k - p\| = 0$$

où $\|.\|$ est la norme

$$\|p\|^2 = \int_{SO(3)} p^2 d\mu$$

Proposition 42 *En remplaçant $K = 0$ dans (A.10), nous avons la limite en probabilité*

$$\lim_L \lim_n \|\hat{p}_L^n - p\| = 0$$

Démonstration : Pour $L \geq 1$ soit p_L la fonction suivante

$$p_L = \sum_{l=0}^{L} d_l \operatorname{tr}\left(\phi_X(l) U^{l\dagger}\right)$$

D'après le théorème de Peter-Weyl, $\lim_L \|p_L - p\| = 0$. Par (A.10) et la proposition 41 nous avons $\lim_n \|\hat{p}_L^n - p_L\| = 0$ en probabilité pour tout $L \geq 1$. Ceci résulte du fait que pour tout $n, L \geq 1$

$$\|\hat{p}_L^n - p_L\|^2 = \sum_{l=0}^{L} d_l |\hat{\phi}_X^n(l) - \phi_X(l)|^2$$

Pour conclure il suffit d'observer que

$$\|\hat{p}_L^n - p\|^2 = \|\hat{p}_L^n - p_L\|^2 + \|p_L - p\|^2 \tag{A.14}$$

pour tous $n, L \geq 1$.▲

Nous avons caractérisé la convergence des estimées paramétriques en utilisant (A.13) et la convergence des estimées non paramétriques en utilisant (A.14). Faisons les remarques suivantes sur ces formules. L'inégalité (A.13) donne seulement une estimation grossière pour le taux de convergence des estimées paramétriques. La qualité de ces estimations est meilleur quand les constantes k_l sont plus grandes, c'est à dire proches de la valeur 1. Ceci est équivalent à une plus grande distance au sens de la norme $\|.\|$ entre p et la fonction constante 1. Il faut interpréter ce dernier point en relation avec la figure 2.3 du paragraphe 2.4.2.3. Dans la proposition 42 nous avons ignoré le rôle de la constante K. Il est discuté dans notre papier [66].

A.3 Filtrage des processus matriciels sous contraintes non linéaires

Cette section correspond au chapitre 3 dans le corps de la thèse. Rappelons que ce chapitre est constitué des deux sections, 3.1 et 3.2. La section 3.2 correspond au but principal de ce chapitre, alors que la 3.1 ne fait que l'illustrer. Pour cette raison, nous avons choisi de reproduire la section 3.1.

Cette section donne la formulation générale de l'approche basée sur la linéarisation locale. Ceci est fait pour les processus matriciels Browniens. En réalisant une compréhension plus poussée de la dynamique de ces processus, il sera possible de les considérer de façon correcte dans les problèmes d'estimation et en particulier d'étudier les performances de notre approche. Fixons $d \geq 1$, nous allons noter $\mathcal{L}(d)$ l'espace des matrices réelles $d \times d$ et $\mathcal{GL}(d) \subset \mathcal{L}(d)$ le sous-ensemble des matrices inversibles. Nous supposons donné un espace de probabilité complet $(\Omega, \mathcal{A}, \mathbb{P})$. Considérons une équation différentielle stochastique linéaire de Stratonovich de la forme suivante

$$dY_t = Y_t dX_t \qquad Y_0 = I \tag{A.15}$$

où le processus inconnu Y et le processus dirigeant X sont à valeurs dans $\mathcal{L}(d)$ et I_d est la matrice identité dans $\mathcal{L}(d)$. Pour $t \geq 0$ les éléments de matrice de X_t sont comme suit, l'intégrale étant une intégrale d'Itô

$$X_t^{ij} = f_{ij}(t) + \int_0^t Q_{ij}(s) dB_s^{ij} \tag{A.16}$$

pour $1 \leq i, j \leq d$, où f_{ij} et A_{ij} sont des fonctions $\mathbb{R}_+ \to \mathbb{R}_+$. Les processus B^{ij} sont les éléments de matrice d'une mouvement Brownien B à valeurs dans $\mathcal{L}(d)$. Pour $1 \leq i, j \leq d$ nous considérons que

f_{ij} est continue et à variations finies, par exemple une fonction C^1, et que Q_{ij} est continue. Avec ces conditions, il est directe de voir que l'équation (A.15) est bien définie. Nous la recopions sous forme intégrale et pour les éléments de matrice Y^{ij}, où $1 \leq i, j \leq d$.

$$Y_t = I_d + \int_0^t Y_s dX_s \qquad Y_t^{ij} = \delta_{ij} + \sum_{k=1}^{d} \int_0^t Y_s^{ik} dX_s^{kj} \tag{A.17}$$

La solution de (A.17) pour un processus donnée X de la forme (A.16) est notée $\mathcal{E}(X) = (\mathcal{E}(X)_t)_{t\geq 0}$. Au A.3.1 nous montrerons que $\mathcal{E}(X)$ est sujet à des contraintes non linéaires. Cependant, ces contraintes non linéaires se réduisent à des contraintes linéaires sur X. La structure multiplicative de $\mathcal{E}(X)$ est exhibée au A.3.2. Les A.3.3 et A.3.4 correspondent au but principal de cette section. Le A.3.3 considère l'inversion de la transformation $X \mapsto \mathcal{E}(X)$. Nous utiliserons cette transformation inverse comme transformation de linéarisation locale utilisée pour se débarrasser des contraintes non linéaires. La stabilité de la transformation $X \mapsto \mathcal{E}(X)$ est considérée au A.3.4. Ce sont ces propriétés de stabilité et d'invariance qui justifieront l'utilisation de la linéarisation locale. Une idée récurrente par la suite sera l'équivalence entre les processus X et $\mathcal{E}(X)$. La proposition 44 du A.3.1 stipule que la distribution de $\mathcal{E}(X)$ est complètement déterminée par celle de X. De façon similaire, la proposition 48 montre que le contenu en information des processus $\mathcal{E}(X)$ et X est en fait le même.

Pour un processus X de la forme (A.16) notons $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ la filtration naturelle de X. L'existence et l'unicité de $\mathcal{E}(X)$ peuvent être montrées à l'aide de l'approximation de Picard [28]. L'équation (A.17) peut également être approximée à l'aide d'intégrales multiplicatives [14]. L'approximation de Picard est définie de façon itérative. Étant donné le processus constant $X^0 = I_d$ soit pour $n \geq 1$ le processus X^n donné pour $t \geq 0$ comme suit

$$X_t^n = I_d + \int_0^t X_s^{n-1} dX_s \tag{A.18}$$

alors les processus $(X^n)_{n\geq 1}$ convergent localement uniformément en moyenne quadratique vers $\mathcal{E}(X)$. Il résulte de cette approximation que $\mathcal{E}(X)$ est un processus $\mathcal{F}$-adapté et continu [28].

Un rôle plus important sera par la suite dévolu à l'approximation de l'équation (A.17) par des intégrales multiplicatives. Cette approximation est plus adaptée à l'étude des contraintes non linéaires imposées à $\mathcal{E}(X)$, comme nous verrons au A.3.1. En plus de leur importance pour la suite, les intégrales multiplicatives ont été utilisées dans la littérature pour implémenter des techniques d'optimisation stochastique et d'échantillonnage pour des applications de poursuite et d'estimation de sous-espaces. Voir le livre récent de Miller et Grenander [22].

A.3.1 Intégrales multiplicatives

Les solutions d'équations de la forme (A.17) constituent un modèle général des processus matriciels Browniens sous contraints non linéaires. Pour un X donné de la forme (A.16), nous allons préciser les contraintes vérifiées par la solution $\mathcal{E}(X)$ de (A.17). La proposition 43 ci-dessous stipule que $\mathcal{E}(X)$ a ses valeurs dans un groupe de Lie matriciel, en effet un sous-groupe de $\mathcal{GL}(d)$. Cette proposition, avec la proposition 47 du A.3.3, permet de formuler un résultat important. Bien que $\mathcal{E}(X)$ soit sujet à des contraintes non linéaires, ces contraintes se réduisent à des contraintes linéaires sur X. De plus, étant donné un processus matriciel Brownien Y sujet au même type de contraintes non linéaires, c'est à dire à valeurs dans un groupe de Lie matriciel, nous pouvons identifier la dynamique de Y comme étant donnée par une équation de la forme (A.17). Précisons que du point de vue des applications les contraintes que nous considérons sont d'une grande généralité. En effet, elles correspondent à des considérations de symétrie connues dans une grande variété de problèmes appliqués.

Le théorème 6 ci-dessous établit l'approximation de $\mathcal{E}(X)$ par des intégrales multiplicatives. Il sera appliqué à la démonstration des propositions 43 et 44. Ce théorème est ici accepté sans démonstration. Une discussion générale de la démonstration est faite dans [28]. Rappelons avant d'énoncer le théorème 6 ce que nous entendons par suite décroissante de subdivisions. Une subdivision de $\mathbb{R}_+$ est une suite croissante $(t_m)_{m\geq 0}$ où $t_0 = 0$ et de plus $t_m \uparrow \infty$ et $\sum_{m\geq 1} |t_m - t_{m-1}| < \infty$. Une suite décroissante de subdivisions de $\mathbb{R}_+$ est donnée si pour tout $N \geq 1$ une subdivision $(t_m^N)_{m\geq 0}$ de $\mathbb{R}_+$ est donnée et nous avons $|t^N| \downarrow 0$, où $|t^N| = \sum_{m\geq 1} |t_m^N - t_{m-1}^N|$ pour $N \geq 1$.

Theorem 6 *Soit X donné par (A.16) et $(t_m^N)_{m\geq 0}$ qui donne pour $N \geq 1$ une suite décroissante de subdivisions de $\mathbb{R}_+$. Pour $N \geq 1$ soit Y^N le processus tel que pour $t \geq 0$ nous avons*

$$Y_t^N = \prod_{m\geq 0} \exp(X_{t\wedge t_{m+1}^N} - X_{t\wedge t_m^N}) \tag{A.19}$$

Alors les processus Y^N convergent vers $\mathcal{E}(X)$ localement uniformément en moyenne quadratique.

Appliquons ce théorème aux propositions 43 et 44. La proposition 43 stipule que $\mathcal{E}(X)$ a ses valeurs dans un groupe de Lie matriciel. Ceci est équivalent à un ensemble de contraintes non linéaires imposées à $\mathcal{E}(X)$. La proposition 47 du A.3.3 est la contraposée de cette proposition. Rappelons la notion de groupe de Lie matriciel. C'est un sous-groupe $G \subset \mathcal{GL}(d)$ tel que G soit une sous-variété de $\mathcal{GL}(d)$ dans l'espace Euclidien $\mathcal{L}(d)$. Il nous suffira de connaitre la caractérisation suivante [46]: Un sous-groupe $G \subset \mathcal{GL}(d)$ est un groupe de Lie matriciel *ssi* G est fermé dans $\mathcal{GL}(d)$.

Soit maintenant G un groupe de Lie matriciel et $T_I G$ son algèbre de Lie. Rappelons que $T_I G \subset \mathcal{L}(d)$ est un espace vectoriel de matrices et que pour tout $J \in T_I G$ nous avons $\exp(J) \in G$. Nous avons la proposition suivante.

Proposition 43 *Si pour $t \geq 0$ nous avons $X_t \in T_I G$ presque sûrement alors pour $t \geq 0$ nous avons $\mathcal{E}(X)_t \in G$ presque sûrement.*

Démonstration : Commençons par montrer que $\mathcal{E}(X)_t \in \mathcal{GL}(d)$ presque sûrement pour $t \geq 0$. Utilisons la notation $Y \equiv \mathcal{E}(X)$. Nous allons construire un processus $\bar{Y} = (\bar{Y}_t)_{t\geq 0}$ tel que pour $t \geq 0$ nous avons $Y_t \bar{Y}_t = I_d$. Soit $\bar{X}$ le processus tel que $\bar{X}_t = -X^T$ pour $t \geq 0$ et posons $\bar{Y}_t = \left[\mathcal{E}(\bar{X})_t\right]^T$. Il peut être vérifié par des manipulations simples de l'équation (A.17) que $\bar{Y}$ satisfait à l'équation de Stratonovich suivante

$$\bar{Y}_t = I_d - \int_0^t dX_s \bar{Y}_s \qquad \bar{Y}_t^{ij} = \delta_{ij} - \sum_{k=1}^d \int_0^t \bar{Y}_s^{kj} dX_s^{ik} \tag{A.20}$$

que nous avons encore une fois écrit sous forme matricielle et pour les éléments de matrice $\bar{Y}^{ij}$ pour $1 \leq i, j \leq d$. Par intégration par partie nous avons

$$Y_t \bar{Y}_t - Y_0 \bar{Y}_0 = \int_0^t Y_s dX_s \bar{Y}_s - \int_0^t Y_s dX_s \bar{Y}_s = 0$$

Comme $Y_0 \bar{Y}_0 = I_d$ nous avons que $\mathcal{E}(X) \in \mathcal{GL}(d)$ presque sûrement pour $t \geq 0$.

Montrons maintenant que pour $t \geq 0$ nous avons $\mathcal{E}(X) \in G$ presque sûrement. Soient Y^N pour $N \geq 1$ des processus de la forme (A.19). Il suit du théorème 6 que $\mathcal{E}(X)_t$ est la limite en moyenne quadratique des variables aléatoires Y_t^N. Notons que pour tout $t \geq 0$ le produit (A.19) contient seulement un nombre fini de facteurs. Il suit que $Y_t^N \subset G$, chacun de ces facteurs étant dans G. Par passage à la limite, et comme $\mathcal{E}(X)_t \in \mathcal{GL}(d)$ et G est fermé dans $\mathcal{GL}(d)$, nous avons que $\mathcal{E}(X)_t \in G$ presque sûrement.▲

La proposition 44 suivante stipule que la distribution de $\mathcal{E}(X)$ ne dépend que de celle de X. La démonstration est une application directe du théorème 6. Et nous l'omettons ici.

Proposition 44 *Soient X^1 et X^2 des processus de la forme (A.16) et supposons que X^1 et X^2 aient les mêmes distributions en dimension finie. Nous avons alors que $\mathcal{E}(X^1)$ et $\mathcal{E}(X^2)$ ont les mêmes distributions en dimension finie.*

A.3.2 Structure multiplicative de $\mathcal{E}(X)$

La solution $\mathcal{E}(X)$ de l'équation (A.17) est un processus non stationnaire. Nous allons ici voir que $\mathcal{E}(X)$ possède une structure multiplicative définie en termes de ses incréments $\mathcal{E}(X)_{(s|t)}$ qui sont pour $0 \leq s \leq t$ des variables aléatoires dans $\mathcal{GL}(d)$. La structure multiplicative de $\mathcal{E}(X)$ correspond à la définition des processus de Lévy grâce aux conditions (L1) à (L4) de la section 2.3. Sous une condition simple sur X, les incréments de $\mathcal{E}(X)$ sont stationnaires. Ainsi, une structure stationnaire est identifiée au sein du processus non stationnaire $\mathcal{E}(X)$. Ceci est d'une grande valeur pour les problèmes d'estimation à partir d'observations des trajectoires de $\mathcal{E}(X)$. Les propositions 45 et 46 donnent les propriétés principales des incréments de $\mathcal{E}(X)$.

Nous avons vu à la démonstration de la proposition 43 que $\mathcal{E}(X)_t \in \mathcal{GL}(d)$ presque sûrement pour tout $t \geq 0$. Il est donc possible de définir pour $0 \leq s \leq t$

$$\mathcal{E}(X)_{(s|t)} = [\mathcal{E}(X)_s]^{-1}\mathcal{E}(X)_t \tag{A.21}$$

notre but principal sera ici de démontrer la proposition 46. Cette proposition affirme que les incréments de $\mathcal{E}(X)$ donnés par (A.21) sont indépendants et précise une condition sur X garantissant leur stationnarité. Ainsi, les incréments de $\mathcal{E}(X)$ vérifient des conditions similaires aux (L1) à (L4) de la section 2.3. Il n'est pas aisé d'établir la proposition 46 à partir de la définition A.21 des incréments. Nous commençons dans la proposition 45 par donner une définition dynamique de ces incréments. La proposition 46 pourra alors suivre directement.

Pour la proposition 45 nous avons besoin de la notation suivante. Soit donné X de la forme (A.16). Pour $s \geq 0$ nous définissons le processus $X^s = (X^s_t)_{t\geq 0}$ où $X^s_t = X_{s+t} - X_s$. Nous admettons que X^s peut s'écrire sous la forme (A.16), de façon que le processus $\mathcal{E}(X^s)$ soit défini comme ci-dessus.

Proposition 45 *Pour tout $0 \leq s \leq t$ nous avons*

$$\mathcal{E}(X)_{(s|t)} = \mathcal{E}(X^s)_{t-s} \tag{A.22}$$

Démonstration : Soit $s \geq 0$. Notons $Y^1 = (Y^1_t)_{t\geq 0}$ le processus donné par $Y^1_t = \mathcal{E}(X)_{t\wedge s}$. Soit $Y^1 = (Y^2_t)_{t\geq 0}$ le processus où $Y^2_t = \mathcal{E}(X^s)_{(t-s)^+}$. Nous allons montrer pour $t \geq 0$ que

$$\mathcal{E}(X)_t = Y^1_t Y^2_t \tag{A.23}$$

En remplaçant pour $s \leq t$ et en comparant à (A.21) la proposition sera ainsi prouvée. Soient X^1 et X^2 les processus définis pour $t \geq 0$ par $X^1_t = X_{t\wedge s}$ et $X^2_t = X_t - X^1_t$. Notons que Y^1 résout l'équation de Stratonovich suivante

$$Y^1_t = I_d + \int_0^t Y^1_u X^1_u \tag{A.24}$$

pour $t \geq 0$. Nous allons montrer que Y^2 résout l'équation associée

$$Y^2_t = I_d + \int_0^t Y^2_u X^2_u \tag{A.25}$$

pour $t \geq 0$. Ceci résulte en effet de la définition de $\mathcal{E}(X^s)$

$$\mathcal{E}(X^s)_t = I_d + \int_0^t \mathcal{E}(X^s)_u dX^s_u$$

en appliquant le changement de temps $t \mapsto (t-s)^+$. Il est maintenant possible de montrer (A.23) en vérifiant que le produit du membre droit de cette formule est bien une solution de l'équation (A.17). Ceci se fait par intégration par partie et en utilisant (A.24) et (A.25).▲

Pour la proposition 46 nous dirons que X de la forme (A.16) est stationnaire s'il existe $h \in \mathcal{L}(d)$ tel que pour $t \geq 0$ nous avons $f_{ij}(t) = h_{ij}t$ et $Q_{ij}(t) = 1$ for $1 \leq i, j \leq d$.

Proposition 46 *Pour tout $0 \leq s \leq t$ nous avons les propriétés suivantes*
(i) *$\mathcal{E}(X)_{(s|t)}$ est indépendant de $\mathcal{F}_s$.*
(ii) *Si X est stationnaire alors $\mathcal{E}(X)_{(s|t)} \stackrel{d}{=} \mathcal{E}(X)_{t-s}$.*

Démonstration : (ii) suit de la proposition 44 du A.3.1. Si X est stationnaire alors pour tout $s \geq 0$ les processus X et X^s ont les mêmes distributions en dimension finie. Il suit que pour $s \leq t$ nous avons $\mathcal{E}(X)_{t-s} \stackrel{d}{=} \mathcal{E}(X^s)_{t-s}$. Par la proposition 45 nous avons que $\mathcal{E}(X)_{(s|t)} = \mathcal{E}(X^s)_{t-s}$.

Nous allons démontrer (i) à l'aide du théorème 6. Pour $u \geq 0$ nous avons que les produits π^N suivants convergent en moyenne quadratique vers $\mathcal{E}(X^s)_u$

$$\pi^N = \prod_{t_m^N \leq u} \exp(X^s_{u \wedge t^N_{m+1}} - X^s_{u \wedge t^N_m})$$

où $(t_m^N)_{m \geq 0}$ donne pour $N \geq 1$ une suite décroissante de subdivisions de $\mathbb{R}_+$. Comme X est à accroissements indépendants il est possible de montrer que π^N pour $N \geq 1$ est indépendant de $\mathcal{F}_s$. Cette propriété est préservée par les limites en moyenne quadratique et il suit que $\mathcal{E}(X^s)_u$ est indépendant de $\mathcal{F}_s$. Il suffit de prendre $u = t - s$ et d'utiliser la proposition 45 pour compléter la démonstration.▲

A.3.3 Inversion de la dynamique

Nous définissons ici notre transformation de linéarisation locale pour les processus matriciels Browniens. Au A.3.4, nous l'utiliserons pour formuler notre approche pour le filtrage de ces signaux. La forme précise de cette transformation est donnée au théorème 7. Ce théorème identifie pour tout processus matriciel Brownien Y un processus X de la forme (A.16) tel que $Y \equiv \mathcal{E}(X)$. Ainsi est réalisée une inversion de la transformation $X \mapsto \mathcal{E}(X)$ définie par l'équation (A.17).

La proposition 47 établit une propriété importante de la linéarisaion locale. Supposons que le processus matriciel Brownien Y ait ses valeurs dans un groupe de Lie matriciel, de façon à ce qu'il soit sujet à des contraintes non linéaires. Le processus X correspondant est alors de la forme (A.16) et n'est sujet qu'à des contraintes linéaires. La proposition 48 donne l'équivalence des filtrations naturelles de Y et X. Ce résultat sera discuté en termes d'applications de filtrage. Donnons maintenant la proposition 47. Bien qu'il utilise le théorème 7, ce théorème peut dans sa première partie être démontré directement.

Partons de l'observation suivante. Si X est de la forme (A.16), il est claire d'après l'équation (A.17) que X peut être retrouvé à partir de $\mathcal{E}(X)$. En effet, pour $t \geq 0$

$$X_t = \int_0^t Y_s^{-1} dY_s \tag{A.26}$$

où nous avons mis $Y \equiv \mathcal{E}(X)$. Après avoir précisé la classe de processus Y pour lesquels la proposition 47 aura lieu, nous utiliserons la même formule (A.26) pour obtenir le processus X désiré.

Soit le processus $Y = (Y_t)_{t \geq 0}$ ayant ses valeurs dans $\mathcal{GL}(d)$, en particulier $Y_0 = I_d$. Nous supposons que Y est de plus une semimartingale continue. Pour $0 \leq s \leq t$ définissons l'incrément de Y comme

suit

$$Y_{(s|t)} = Y_s^{-1} Y_t$$

Nous supposons que Y a des incréments indépendants. Soit $\mathcal{H}$ la filtration naturelle de Y. Pour $0 \leq s \leq t$ nous avons que $Y_{(s|t)}$ est indépendant de $\mathcal{H}_s$. Le processus X suivant est bien défini

$$X_t = \int_0^t Y_s^{-1} dY_s \tag{A.27}$$

Proposition 47 *Soit Y comme ci-dessus et X donné par (A.27). Soit G un groupe de Lie matriciel et $T_I G$ son algèbre de Lie. Nous avons les propriétés suivantes*
(i) *X est de la forme (A.16) et $Y \equiv \mathcal{E}(X)$.*
(ii) *Si pour $t \geq 0$ nous avons $Y_t \in G$ presque sûrement alors pour $t \geq 0$ nous avons $X_t \in T_I G$ presque sûrement.*

Démonstration : Nous ne donnons que la démonstration de (i). Celle de (ii) utilise le théorème 7 ci-dessous. Elle est analogue à la dernière partie de la démonstration de la proposition 43.

De part sa définition (A.27), X est une semimartingale continue. Par l'équation (A.17) il suit que pour $t \geq 0$

$$Y_t = I + \int_0^t Y_s Y_s^{-1} dY_s = I + \int_0^t Y_s dX_s$$

ceci montre que $Y \equiv \mathcal{E}(X)$. Afin de montrer que X est de la forme (A.16) il suffit de montrer que X est à accroissements indépendants. I est claire de (A.27) que X est $\mathcal{H}$-adapté. Pour $0 \leq s \leq t$ nous avons la limite en probabilité

$$X_t - X_s = \int_s^t Y_s^{-1} dY_s = \lim_n \sum_{m=1}^{M} Y_{t_{m-1}^n}^{-1} (Y_{t_m^n} - Y_{t_{m-1}^n})$$

où le pas de la partition $s = t_1^n < \ldots < t_M^n = t$ de l'intervalle $[s,t]$ définie pour $n \geq 1$ décroit vers 0 avec n. Il est possible de transformer la somme dans la dernière expression pour obtenir

$$X_t - X_s = \lim_n \sum_{m=1}^{M} Y_{(t_{m-1}^n | t_m^n)} - I \tag{A.28}$$

Par hypothèse, pour $n \geq 1$ et $m \geq 0$ nous avons que $Y_{(t_{m-1}^n | t_m^n)}$ est indépendant de $\mathcal{H}_{t_m^n}$. En utilisant le fait que l'indépendance est préservée par les limites en probabilité nous pouvons conclure que $X_t - X_s$ est indépendant de $\mathcal{H}_s$. En particulier, X est donc à accroissements indépendants. Il suit donc de la décomposition de Lévy-Itô que X est de la forme (A.16). Voir [27].▲

La proposition 48 suivante stipule que l'es filtration naturelles des processus X et $\mathcal{E}(X)$ sont identiques. Alors que sa preuve est simple, cette proposition est d'une signification générale pour notre approche. Étant donné un processus X de la forme (A.16), il est claire que les processus X et $\mathcal{E}(X)$ sont de natures géométriques différentes. Alors que $\mathcal{E}(X)$ est sujet à des contraintes non linéaires X est seulement sujet à des contraintes linéaires. La proposition 48 nous dit qu'en dépit de cette différence, X et $\mathcal{E}(X)$ ont le même contenu en information. En particulier, le filtrage par rapport à des observations de X ou de $\mathcal{E}(X)$ donne lieu à la même connaissance *a posteriori*.

Proposition 48 *Dans la notation de la proposition 47, soient $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ et $\mathcal{H} = (\mathcal{H}_t)_{t \geq 0}$ les filtrations naturelles de X et de Y respectivement. Nous avons que $\mathcal{F}_t = \mathcal{H}_t$ pour $t \geq 0$.*

Passons maintenant au théorème 7. Pour nos processus qui sont définis en temps continu, la linéarisation locale est donnée par un passage à la limite. Pour énoncer ce théorème, nous avons besoin de la notation suivante. Soit $\log : \mathcal{GL}(d) \to \mathcal{L}(d)$ l'application qui associe à $y \in \mathcal{GL}(d)$ l'unique x tel que $\exp(x) = y$ dans le cas où $|I_d - y| < 1$ et qui lui associe 0 sinon. Pour la démonstration du théorème 7 voir [30].

Theorem 7 *Dans la notation de la proposition 48, supposons que $(t_m^N)_{m\geq 0}$ pour $N \geq 1$ donne une suite décroissante de subdivisions de $\mathbb{R}_+$, nous avons que les processus suivants $X^N = (X_t^N)_{t\geq 0}$ sont continus, $\mathcal{H}$-adaptés et convergent localement uniformément en probabilité vers X.*

$$X_t^N = \sum_{m\geq 0} \log\left[Y_{(t_m^N \wedge t | t_{m+1}^N \wedge t)}\right] \tag{A.29}$$

A.3.4 Propriétés de stabilité

Nous allons finalement donner une formulation générale de notre approche du filtrage des processus matriciels sous contraintes non linéaires. Ceci sera fait pour les processus matriciels Browniens dont les propriétés ont été développées ci-dessus. La transformation linéarisation locale est implémentée grâce au théorème 7 et ce sont les propriétés de stabilité liées à cette transformation qui justifieront notre approche.

Notre résultat principal est la proposition 50. Cette proposition obtient la stabilité de la transformation $X \mapsto \mathcal{E}(X)$. Bien entendu, cette proposition est à comprendre dans le contexte de notre application de la linéarisation locale que nous commençons ici par décrire. Fixons $d \geq 1$ et soit $G \subset \mathcal{GL}(d)$ un groupe de Lie matriciel avec $T_I G$ son algèbre de Lie. Soit le problème de filtrage suivant. Un processus signal Y^1 est supposé auquel nous n'avons pas accès directement. Au contraire, nous avons des observations d'un autre processus Y^2 que nous appelons le processus observation. Ces deux processus ont leurs valeurs dans G et notre problème est d'inférer le processus signal Y^1 en utilisant l'information disponibles grâce au processus observation Y^2. Une solution consiste en un processus $\hat{Y}^1$ vérifiant les deux conditions suivantes.

- $\hat{Y}^1$ doit être optimal. Ce processus se construit uniquement à partir des observations disponibles de Y^2. Il devra en faire l'usage le plus complet possible.
- $\hat{Y}^1$ vérifie les mêmes contraintes non linéaires que Y^1 et Y^2. Notamment, il a ses valeurs dans G.

En général, ces deux conditions ne sont pas compatibles. Il est bien connu en théorie du filtrage optimal que le processus $\hat{Y}^1$ vérifiant la première condition d'optimalité est le processus d'espérance conditionnelle [23]

$$\hat{Y}_t^1 = \mathbb{E}[Y^1 | \mathcal{F}_t] \tag{A.30}$$

pour $t \geq 0$, où $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ est la filtration naturelle de Y^2. Ce processus est de plus la meilleure approximation de Y^1 au sens de la moyenne quadratique. Comme l'espérance conditionnelle est une opération linéaire, il est claire que $\hat{Y}^1$ ne satisfait pas en général à la deuxième condition de contraintes non linéaires.

Supposons que Y^1 et Y^2 vérifient les conditions de la proposition 47. Dans ce cas nous savons qu'il existe des processus X^1 et X^2 de la forme (A.16) à valeurs dans $T_I G$ tels que $Y^i = \mathcal{E}(X^i)$ pour $i = 1, 2$. De plus X^1 e X^2 sont donnés par le théorème 7. Il parait désirable de répondre aux deux conditions ci-dessus en considérant le processus X^2 plutôt que le processus d'observation Y^2. Faisons les deux remarques suivantes

- Par la proposition 48 la filtration naturelle de X^2 est aussi égale à $\mathcal{F}$. En d'autres termes, X^2 contient la même information que Y^2.

- X^2 est seulement sujet à des contraintes linéaires.

L'idée principale de notre approche est de chercher $\hat{Y}^1$ sous la forme $\mathcal{E}(\hat{X}^1)$ où $\hat{X}^1$ est un processus de la forme (A.16) à valeurs dans $T_I G$. Par la proposition 43 nous aurons que $\hat{Y}^1$ a ses valeurs dans G et vérifie donc la deuxième condition liée aux contraintes non linéaires. Le processus $\hat{X}^1$ sera construit à partir de X^2 et devra être une bonne approximation de X^1. Nous pouvons décrire cette approche à l'aide du diagramme suivant

$$\begin{array}{ccc} Y^2 & \xrightarrow{F} & \hat{Y}^1 \approx Y^1 \\ \downarrow & & \uparrow \\ X^2 & \xrightarrow{f} & \hat{X}^1 \approx X^1 \end{array} \tag{A.31}$$

Ici, les flèches descendantes correspondent à l'application du théorème 7 et les flèches montantes à l'application du théorème 6. Le théorème 7 donne notre transformation de linéarisation locale. Notons que ces deux théorèmes contiennent des passages à la limite et nous n'avons ainsi pas les processus X^2 et $\hat{Y}^1$ de façon exacte. Pour l'instant, nous négligeons cette difficulté. Il est discuté dans [28] comment la convergence dans ces deux théorèmes peut être améliorée jusqu'à atteindre un taux exponentiel.

Nous avons formulé l'utilisation de la linéarisation locale pour le filtrage des processus matriciels sous contraintes non linéaires. Il n'est pas clair que l'approche ainsi définie soit optimale en un quelconque sens habituel et nous ne l'avons toujours pas justifiée. A ce stade, ce n'est donc qu'une *préscription* plausible pour approximer Y^1 à partir de Y^2. C'est précisément la stabilité de la transformation $X \mapsto \mathcal{E}(X)$ qui justifiera notre approche. En effet, supposons que nous ayons les connaissances *a priori* nécessaires sur X^1 et X^2 pour obtenir une bonne approximation $\hat{X}^1 \approx X^1$ comme dans le diagramme (A.31). La proposition 50 nous garantit que cette approximation sera conservée par l'application des flèches montantes et nous devrions ainsi obtenir une approximation acceptable $\hat{Y}^1 \approx Y^1$. Nous verrons cependant que nous avons des problèmes de dérive. Plus précisément, une bonne approximation $\hat{Y}^1 \approx Y^1$ pourrait s'avérer très coûteuse, même si elle est possible en principe.

Donnons maintenant la proposition 50. Cette proposition décrit la stabilité de la transformaton $X \mapsto \mathcal{E}(X)$. Sans donner une estimation très forte de cette stabilité, elle permet une compréhension qualitative des rôles de la stabilité et de la dérive mentionnés ci-dessus. D'un point de vu mathématique la proposition 50 est concernée par la stabilité de l'équation (A.17) par rapport au processus dirigeant X. L'énoncé de cette proposition a été choisi afin de rendre accessible sa démonstration et il est naturel qu'il paraisse relativement restreint –voir sous-section 3.2.5. Remarquons que des résultats de stabilité du même type mais beaucoup plus généraux sont connus dans la littérature. Voir en particulier [57].

Soient X^1 et X^2 deux processus quelconques de la forme (A.16). La proposition 50 borne par le haut la distance entre $Y^1 \equiv \mathcal{E}(X^1)$ et $Y^2 \equiv \mathcal{E}(X^2)$ à l'aide de la distance entre X^1 et X^2. Pour la distance entre $\mathcal{E}(X^1)$ et $\mathcal{E}(X^2)$ nous envisageons la distance uniforme en moyenne quadratique. La distance entre X^1 et X^2 est précisée par la proposition intermédiaire 49 ci-dessous. Les éléments de matrice des processus X^1 et X^2 sont comme suit

$$X^{1}{}_t^{ij} = f_{ij}^1(t) + \int_0^t Q_{ij}^1(s) dB_s^{ij} \qquad X^{2}{}_t^{ij} = f_{ij}^2(t) + \int_0^t Q_{ij}^2(s) dB_s^{ij} \tag{A.32}$$

pour $1 \leq i, j \leq d$ et $t \geq 0$. Les fonctions f_{ij}^1, f_{ij}^2 et Q_{ij}^1, Q_{ij}^2 sont comme dans (A.16). Afin de simplifier la démonstration de la proposition 50 les deux hypothèses suivantes sont faites. Premièrement, les

expressions (A.32) imposent que X^1 et X^2 sont définis pour le même mouvement Brownien B dans $\mathcal{L}(d)$. Cela place une restriction sur les problèmes envisageables mais reste par exemple suffisant pour les problèmes de débruitage. Deuxièmement, bien que l'intégrale stochastique dans l'équation (A.17) soit une intégrale de Stratonovich, toutes les intégrales stochastiques dans les propositions 49 et 50 seront traitées comme des intégrales d'Itô. Pour voir que nous pouvons le faire sans perdre de généralité, faisons la remarque suivante. Le fait de transformer l'équation (A.17) en équation d'Itô ne change pas sa forme générale mais rajoute simplement un terme correctif au processus dirigeant X. Ce nouveau terme est une fonction croissante qui pourra être absorbée dans les fonctions f_{ij} de la formule (A.16).

La proposition 49 est un lemme technique utilisé dans la preuve de la proposition 50. Pour tout processus X de la forme (A.16), la proposition 49 exhibe une fonction croissante $A_X : \mathbb{R}_+ \to \mathbb{R}_+$ qui domine les intégrales stochastiques par rapport à X. Pour cette proposition nous avons la notation suivante. Soit Z un processus continu à valeurs dans $\mathcal{L}(d)$. Pour $T \geq 0$ notons Z_T^* la variable aléatoire dans $\mathcal{L}(d)$ dont les éléments de matrice sont $Z^{*ij}_T = \sup_{t \leq T} |Z_T^{ij}|$. Pour les démonstrations des propositions 49 et 50, voir la sous-section 3.2.5.

Proposition 49 *Soit X un processus de la forme (A.16) dont la filtration naturelle est notée $\mathcal{F}$. Il existe une fonction croissante $A_X : \mathbb{R}_+ \to \mathbb{R}_+$ telle que pour tout processus Y à valeurs dans $\mathcal{L}(d)$ qui est continu, carré intégrable et $\mathcal{F}$-adapté nous ayons*

$$\mathbb{E}|Z_T^*|^2 \leq \int_0^T \mathbb{E}|Y_t|^2 dA_X(t) \tag{A.33}$$

pour $T \geq 0$. Ici Z est le processus d'intégrale stochastique, $Z_t = \int_0^t Y_s dX_s$ pour $t \geq 0$.

Pour les processus X^1 et X^2 de (A.32), soit X le processus où $X_t = X_t^1 - X_t^2$ pour $t \geq 0$. Il suit de (A.32) que X est également de la forme (A.16). Soit A_X la fonction positive croissante associée à X comme dans la proposition 49. La proposition 50 donne une borne supérieure sur la distance uniforme en moyenne quadratique entre Y^1 et Y^2 au moyen de la fonction A_X. Cette fonction est une mesure forte de la distance entre X^1 et X^2. En effet, nous avons par la proposition 49

$$\mathbb{E}|X_T^*|^2 \leq A_X(T)$$

pour tout $T \geq 0$. Nous allons donner l'énoncé de la proposition 35 et la discuter en termes de l'approche du diagramme (A.31).

Proposition 50 *Soit $D = (D_t)_{t \geq 0}$ le processus où $D_t = Y_t^1 - Y_t^2$. Pour $T \geq 0$ nous avons l'inégalité suivante*

$$\mathbb{E}|D_T^*|^2 \leq 4d \exp[2A_{X^1}(T) + 2A_{X^2}(T)] A_X(T) \tag{A.34}$$

où A_{X^1} et A_{X^2} sont les fonctions positives croissantes associées à X^1 et X^2 comme dans la proposition 34.

Pour notre discussion nous reprenons l'inégalité (A.34) avec la notation du diagramme (A.31). Nous voulons discuter de l'approximation $\hat{Y}^1 \approx Y^1$ d'après la proposition 50. Nous devons donc supposer que les processus $\hat{X}^1$ et X^1 vérifient les conditions de cette proposition. Rappelons que les restrictions contenues dans ces conditions ont été introduites dans l'unique but de simplifier la démonstration faite à la sous section 3.2.5. Le comportement que nous allons décrire reste vrai pour des situations beaucoup plus générales.

Soit $E = (E_t)_{t\geq 0}$ le processus $E_t = \hat{Y}^1 - Y^1$. Ce processus donne l'erreur due à notre utilisation du diagramme (A.31) pour approximer Y^1 à partir de Y^2. Soit de plus $e = (e_t)_{t\geq 0}$ le processus $e_t = \hat{X}^1 - X^1$. D'après le diagramme nous contrôlons directement le processus e en construisant l'approximation $\hat{X}^1 \approx X^1$. Les performances de notre approche dépendent de la possibilité de dominer le processus E en utilisant le processus e. La proposition 50 fournit la borne suivante (A.35) sur le processus E. Soient $A_{\hat{X}^1}, A_{X_1}$ et A_e les fonctions positives croissantes associées aux processus $\hat{X}^1, X^1$ et e comme dans la proposition 34. Nous pouvons récrire l'inégalité (A.34) pour ces processus.

$$\mathbb{E}|E_T^*|^2 \leq 4d \exp[2A_{\hat{X}^1}(T) + 2A_{X^1}(T)]A_e(T) \tag{A.35}$$

Nous avons ici une illustration des rôles de la stabilité et de la dérive en ce qui concerne les performances de l'approche du diagramme (A.31). La dernière inégalité peut être récrite

$$\mathbb{E}|E_T^*|^2 \leq K(T)A_e(T)$$

pour $T \geq 0$. Il est clair que l'erreur E est domniée par la fonction A_e associée au processus e. Pour $T \geq 0$ fixé, toute amélioration dans l'approximation $\hat{X}^1 \approx X^1$ conduit à une amélioration dans l'approximation $\hat{Y}^1 \approx Y^1$, modulée par un facteur multiplicatif K. Les problèmes de dérive sont précisément liés à la croissance exponentielle du facteur K en relation avec le temps T. Pour des valeurs plus grandes de T, afin d'obtenir un niveau donné de performance de l'approximation $\hat{Y}^1 \approx Y^1$ une qualité beaucoup plus grande de l'approximation $\hat{X}^1 \approx X^1$ pourrait être requise. Les problème de dérive sont ainsi de façon générale une limitation fondamentale des méthodes basées sur la linéarisation locale.

A.4 Conclusions

De façon générale, cette thèse a été guidée par une approche donnant une plus grande place aux outils mathématiques. L'utilisation de l'analyse de Fourier et du calcul stochastique a permis d'étudier d'un point de vue intrinsèque les processus ayant leurs valeurs dans les groupes de Lie matriciels. Nous avons ainsi réalisé une compréhension plus poussée de la dynamique de ces processus et une description plus détaillée de leurs statistiques. C'était là notre point de départ pour résoudre des problèmes d'estimation paramétrique et non paramétrique et pour formuler une méthode de filtrage spécifique à la structure de ces processus.

Au chapitre 2 nous avons résolu le problème du *decompounding*. C'est un problème d'estimation non paramétrique posé en termes des processus de Poisson composés de rotation. Notre solution est basée sur une méthode de fonction caractéristique et a du faire face à une nouvelle difficulté essentielle à ce problème. En effet, partant d'observations indirectes une méthode de fonction caractéristique perd ses propriétés de linéarité et d'invariance. En plus de la solution du problème du *decompounding*, le chapitre 2 a présenté des applications concrètes en polarisation statistique et en diffusion multiple. Ce chapitre s'est borné aux processus de rotation pour lesquels nous disposons des outils puissants de l'analyse de Fourier.

Le chapitre 3 avait pour but de formuler de façon générale et de commencer à étudier l'utilisation de la linéarisation locale pour le filtrage des processus à valeurs dans les groupes de Lie matriciels. A la section 3.1, nous avons appliqué avec succès la linéarisation locale au filtrage des séries temporelles de rotation. La section 3.2 a établi les résultats théoriques nécessaires à une formulation générale de notre approche. Une étude plus détaillée de cette méthode reste un objectif pour des travaux futurs.

Bibliography

[1] S. L. Altmann. *Rotations, quaternions and double groups.* Dover publications, 2005.

[2] D. Applebaum. Compound Poisson processes and Lévy processes in groups and symmetric spaces. *Journal of Theoretical Probability*, 13, 2000.

[3] D. Applebaum. *Lévy processes and Stochastic calculus.* Cambridge University Press, 2005.

[4] D. Applebaum and A. Estrade. Isotropic lévy processes on riemannian manifolds. *The annals of probability*, 2000.

[5] A. O. Barut and R. Raczka. *Theory of group representations and applications.* World scientific publishing, 1986.

[6] D. Bebbington, L. Carrea, and E. Krogager. Geometric polarimetry – Part I: Spinors and wave states. *IEEE transactions on geoscience and remote sensing*, 2009.

[7] C. Brosseau. *Fundamentals of polarized light: A statistical approach.* John Wiley & Sons Inc., 1998.

[8] B. Buchmann and R. Grübel. Decompounding: An estimation problem for Poisson random sums. *The annals of statistics*, 2003.

[9] S. Chandrasekhar. *Radiative transfer.* Dover publications, 1960.

[10] I. Chavel. *Riemannian geometry, a modern perspective.* Cambridge university press, 2006.

[11] P. Debye. *Polar molecules.* Dover publications, 1929.

[12] J. Ellis and A. Dogariu. Differentiation of globally unpolarized complex random fields. *Journal of the optical society of america*, 2004.

[13] J. Ellis and A. Dogariu. Discimination of globally unpolarized fields through stokes vector element correlations. *Journal of the optical society of america*, 2005.

[14] M. Emery. Stabilité des solutions des équations différentielles stochastiques: Application aux intégrales multiplicative stochastiques. *Z. Wahrsch. Verw. Gebeite.*, 1978.

[15] A. Estrade. Exponentielle stochastique et intégrale multiplicative discontinue. *Annales de l'institut Henri Poicaré (B), probabilités et statistiques*, 1992.

[16] Y. C. Fang, C. C. Hsieh, C. C. Kim, J. J. Chang, and T. C. Wooll. Real time motion fairing with unit quaternions. *Computer-aided design*, 1998.

[17] P. T. Fletcher, S. Joshi, C. Lu, and S. Pizer. Gaussian distributions on Lie groups and their application to statistical shape analysis. In *IPMI*. 2003.

[18] P. T. Fletcher, S. Joshi, C. Lu, and S. M. Pizer. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on medical imaging*, 2004.

[19] J. W. Goodman. Statistical properties of laser speckle patterns. In J. C. Dainty, editor, *Laser speckle and related phenomena*. Springer-Verlag, 1975.

[20] J. P. Gordon and H. Kogelnik. Pmd fundamentals: Polarization mode dispersion in optical fibers. *Proceedings of the national academy of science of the United States of America*, 2000.

[21] U. Grenander. *Probabilities on algebraic structures*. John Wiley & Sons Inc., 1963.

[22] U. Grenander and M. Miller. *Pattern theory: From representation to inference*. Oxford university press, 2007.

[23] S. Haykin, P. Yee, and E. Debrez. Optimum nonlinear filtering. *IEEE Transactions on signal processing*, 1997.

[24] H. Heyer. *Probability measures on locally compact groups*. Springer-Verlag, 1977.

[25] A. Ishimaru. *Wave propagation and scattering in random media, Vol.1,2*. Academic Press, 1978.

[26] E. Jakeman. Polarization characteristics of non gaussian scattering by small particles. *Waves in random media*, 1995.

[27] O. Kallenberg. *Foundations of modern probability*. Springer-Verlag, 2002.

[28] R. L. Karandikar. A.s. approximation results for multiplicative stochastic integrals. *Séminaire de probabilités (Strasbourg)*, 1982.

[29] R. L. Karandikar. Girsanov type formula for a lie group valued brownian motion. *Séminaire de probabilités (Strasbourg)*, 1983.

[30] R. L. Karandikar. Multiplicative decomposition of nonsingular matrix valued semimartingales. *Séminaire de probabilités (Strasbourg)*, 1991.

[31] K. Kim, L. Mandel, and E. Wolf. Relationship between jones and mueller matrices for random media. *Journal of the optical society of America*, 1987.

[32] P. T. Kim and J. Y. Koo. Optimal spherical deconvolution. *Journal of multivariate analysis*, 2002.

[33] P.T. Kim. Deconvolution density estimation on SO(N). *The annals of statistics*, 1998.

[34] L. Klimes. Correlation functions of random media. *Pure and applied geophysics*, 2002.

[35] A. A. Kokhanovsky. Small angle approximations of the radiative transfer theory. *Journal of physics D*, 1997.

[36] J.Y. Koo and P.T. Kim. *IEEE*.

[37] L. D. Landau and E. M. Lifchitz. *Physique statistique*. MIR-Ellipses, 1994.

[38] J. Lee and S. Y. Shin. General construction of time domaine filters for orientation data. *IEEE Transactions on visualisation and computer graphics*, 2002.

[39] M. Liao. Lévy processes and Fourier analysis on compact Lie groups. *The annals of probability*, 2004.

[40] M. Liao. *Lévy processes on Lie groups*. Cambridge University Press, 2004.

[41] C. F. Van Loan and G. H. Golub. *Matrix computations*. The John Hopkins University Press, 1989.

[42] A. Luis. Degree of polarization in quantum optics. *Physical review A*, 2002.

[43] J. H. Manton. A globally convergence algorithm for computing the center of mass on compact lie groups. In *ICARCV*, 2004.

[44] J. H. Manton. A centroid approach to the joint approximate diagonalization problem: The real symmetric case. *Digital signal processing*, 2006.

[45] L. Margerin, M. Campillo, and B. Van Tiggelin. Monte carlo simulation of multiple scattering of elastic waves. *Journal geophysical research*, 2000.

[46] J. E. Marsden and T. S. Ratiu. *Introduction to mechanics and symmetry*. Springer-Verlag, 1998.

[47] H. P. McKean. Brownian motion on the 3-dimensional rotation group. *Mem. Col. Sci. Kyoto, 33*, 1960.

[48] M. I. Miller, A. Srivastave, and U. Grenander. Conditional expectation estimation via jump-diffusion processes in multiple target tracking/recognition. *IEEE tranactions on signal processing*, 1995.

[49] M. Moakher. Means and averaging in the group of rotations. *SIAM J. matrix anal. appl.*, 2002.

[50] M. A. Naimark. *Les représentations linéaires du groupe de Lorentz*. Dunod, 1962.

[51] X. Ning, L. Papiez, and G. Sandinson. Compound poisson process method for the multiple scattering of charged particles. *Physical review E*, 1995.

[52] J. Ohtsubo and T. Asakura. Measurement of surface roughness using speckle patterns with non gaussian statistics. *Optics Communications*, 1978.

[53] E. M. Ortiz, F. Gonzales, and F. Moreno. Intensity statistics of the light scattered from particulate surfaces: Interacting particles and non gaussian effects. *Optics Communications*, 2000.

[54] X. Pennec. Probabilities and statistics on riemannian manifolds: Basic tools for geometric measurements. Technical report, INRIA-Projet Epidaure, 1999.

[55] X. Pennec. Bi-invariant means in lie groups. applications to left-invariant polyaffine transformations. Technical report, INRIA-Projet Epidaure, 2006.

[56] F. Perrin. Etude mathématique du mouvement brownien de rotation. *Annales scientifiques de l'ENS*, 1928.

[57] Ph. Protter. *Stochastic integration and differential equations*. Springer-Verlag, 2005.

[58] V. Rossetto-Giaccherino. *Mécanique statistique des systèmes sous contraintes: Topologie de l'ADN et simulations électrostatiques*. PhD thesis, Université Paris 6 –Laboratoire de physico-chimie théorique, 2002.

[59] B. Ruffing and J. Fleischer. Spectral correlation of partially or fully developed speckle patterns generated by rough surfaces. *Journal of the optical society of America*, 1985.

[60] Ph. Réfrégier. Polarization degree of optical waves with non gaussian probability density functions: Kullback relative entropy-based approach. *Optics letters*, 2005.

[61] Ph. Réfrégier, J. Fade, and M. Roche. Estimation precision of the degree of polarization from a single intensity image. *Optics Letters*, 2007.

[62] Ph. Réfrégier and F. Goudail. Kullback relative entropy and characterization of partially polarized optical waves. *Journal of the optical society of America*, 2006.

[63] S. Said and N. Le Bihan. Higher order statistics of stokes parameters in a random birefringent medium. *Waves in random and complex media*, 2008.

[64] S. Said, N. Le Bihan, and S. J. Sangwine. A stability approach to the analysis of rotation time series. In *SYSID*, 2009.

[65] S. Said, N. Courty, N. Le Bihan, and S. J. Sangwine. Exact principal geodesic analysis for data on SO(3). In *EUSIPCO*. 2007.

[66] S. Said, C. Lageman, N. Le Bihan, and J. H. Manton. Decompounding on compact Lie groups. *IEEE transactions on information theory*, 2009.

[67] G. S.Chirikjian and A. B. Kyatkin. *Engineering applications of noncommutative harmonic analysis*. CRC Press, 2000.

[68] K. Schoemake. Animating rotation with quaternion curves. In *SIGGRAPH*. 1985.

[69] P. Sheng. *Wave scattering, localization and mesoscopic phenomena*. Academic press, 1995.

[70] A. Srivastave. A bayesian approach to geometric subspace estimation. *IEEE*, 2000.

[71] J. D. Uherka and A. M. Sergott. On the continuous dependence of the roots of a polynomial on its coefficients. *The American mathematical monthly*, 1977.

[72] B. van Es, S. Gugushvili, and P. Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, (3), 2007.

[73] A. Vanucci and A. Bononi. Statistical characterization of the jones matrix of long fibers affected by polarization mode dispersion (pmd). *Journal of lightwave technology*, 2002.

[74] N. Ja. Vilenkin. *Fonctions spéciales et théorie de la représentation des groupes*. Dunod, 1969.

[75] S. Xavier and J. H. Manton. On the generalization of AR processes to riemannian manifolds. In *ICASSP*. 2006.

[76] B. Yazici. Stochastic deconvolution over groups. *IEEE transactions on information theory*, 2004.