Decompounding on Compact Lie Groups

Salem Said, Christian Lageman, Nicolas Le Bihan, and Jonathan H. Manton, *Senior Member, IEEE*

Abstract—Noncommutative harmonic analysis is used to solve a nonparametric estimation problem stated in terms of compound Poisson processes on compact Lie groups. This problem of *decompounding* is a generalization of a similar classical problem. The proposed solution is based on a characteristic function method. The treated problem is important to recent models of the physical inverse problem of multiple scattering.

Index Terms—Compact Lie groups, compound Poisson processes, multiple scattering, noncommutative harmonic analysis, nonparametric estimation.

I. INTRODUCTION

THIS paper studies the following nonparametric estimation problem. Let $(X_n)_{n \geq 1}$ be *i.i.d.* G-valued random variables for some group G, and let e denote the identity element of G. For example, G might be the group of 3×3 orthogonal matrices, in which case each X_n would be a random 3×3 orthogonal matrix and e would be the 3×3 identity matrix. The process

$$Y(t) = \prod_{n=0}^{N(t)} X_n, \quad X_0 = e$$

where $N = (N(t))_{t \geq 0}$ is a Poisson process with parameter $\lambda > 0$, is called a G-valued compound Poisson process. If G is not commutative, the above products are taken to be ordered from left to right, and $Y(t)$ is called a *left* compound Poisson process. It is assumed that the random variables X_n and $N(t)$ are independent of each other, and for simplicity, it is further assumed that the Poisson parameter λ is known. The general problem is to estimate the distribution of the X_n given partial observations of one or more realisations of the compound Poisson process $Y(t)$. Of specific interest, is the case when multiple realisations of $Y(T)$ are available, for some fixed time instant $T > 0$.

S. Said and N. Le Bihan are with the GIPSA-Lab, Department of Images and Signal, Grenoble, France (e-mail: salem.said@gipsa-lab.grenoble-inp.fr; Bihan@gipsa-lab.grenoble-inp.fr).

C. Lageman is with the Department of Mathematics, University of Würzburg, Am Hubland, 97074 Würzburg, Germany (e-mail: christian.lageman@mathematik.uni-wuerzburg.de).

J. H. Manton is with the University of Melbourne, Department of Electrical and Electronic Engineering, Victoria 3010, Australia (e-mail: jmanton@unimelb.edu.au).

The real numbers form a group, with addition being the group operation. Choosing G to be this group results in the ordinary compound Poisson process $y(t) = \sum_{n=0}^{N(t)} x_n$ where $x_0 = 0$ and x_n for $n \geq 1$ are real-valued *i.i.d.* random variables. Estimating the distribution of the x_n is known as decompounding and has been well-studied [1], [2]. In the present paper, decompounding techniques are extended to the case when G is a noncommutative group. This new case can not be obtained trivially and requires ideas from noncommutative harmonic analysis. Although group-valued compound Poisson processes were introduced by Applebaum in [3], the corresponding decompounding problem has not been addressed in generality before.

This paper contributes to the relatively recent trend consisting in the application of noncommutative harmonic analysis (*i.e.*, harmonic analysis on groups) to estimation and inverse problems. It addresses a nonparametric estimation problem stated in terms of compound Poisson processes on compact Lie groups. We refer to this as the problem of *decompounding* on compact Lie groups, since it directly generalizes the classical problem of decompounding for scalar processes. This generalization is mathematically natural and is motivated by the physical inverse problem of multiple scattering. In particular, this paper also contributes to the modelling of multiple scattering using compound Poisson processes.

Compound Poisson processes model the accumulation of rare events. As such, scalar compound Poisson processes are important tools in queuing and traffic problems and in risk theory. The classical problem of decompounding arises in the context of these processes. A functional approach to this problem is given by Buchmann and Grübel [1]. A characteristic function method is studied by van Es *et al.* [2]. The applications of decompounding in queuing problems and risk theory are referenced in [1]. We extend this problem by considering decompounding on compact Lie groups. We approach this new problem by using noncommutative harmonic analysis to generalize the above mentioned method of [2].

The important potential which noncommutative harmonic analysis holds for engineering problems is well illustrated in the book of Chirikjian and Kyatkin [4]. Its importance to nonparametric estimation stems from the fact that it leads to the successful generalization of the highly important concept of characteristic function in probability. In mathematical research, this generalization was pioneered by Grenander [5] and extensively developed by Heyer [6]. It has received special attention in the engineering community. See Yazici [7] and the papers by Kim *et al.* [8]–[11].

The paper is organized as follows. Section II sets down the necessary background in harmonic analysis and characteristic functions on compact Lie groups. Section III introduces compound Poisson processes on compact Lie groups. In Section IV, we state the decompounding problem for these processes and

present our approach based on noncommutative harmonic analysis. In Section V, we propose a model for multiple scattering based on compound Poisson processes on the rotation group $SO(3)$. Within this model, decompounding appears as a physical inverse problem. We apply our approach as described in Section IV to this problem using numerical simulations.

II. CHARACTERISTIC FUNCTIONS ON COMPACT LIE GROUPS

Characteristic functions of scalar and vector-valued random variables are defined using the usual Fourier transform. Their extension to random variables with values on compact Lie groups owes to the tools of harmonic analysis on these groups. Our presentation of characteristic functions is adapted from [5], [12]. Harmonic analysis on compact Lie groups is presented in more detail in recent papers [8], [7]. More thorough classical references thereon include [13], [14].

Let G be a compact connected Lie group with identity e. We denote by μ the biinvariant normalized Haar measure on G. Hilbert spaces of square integrable (with respect to μ) complex and real-valued functions on G are denoted $L^2(G, \mathbb{C})$ and $L^2(G, \mathbb{R})$. A representation of G is a continuous homomorphism $\pi : G \rightarrow GL(V)$ with V a complex Hilbert space and $GL(V)$ the group of invertible bounded linear maps of V. It is called irreducible if any G-invariant subspace of V is trivial i.e., equals $\{0\}$ or V. Two representations $\pi_i : G \rightarrow GL(V_i)$-with $i = 1, 2$-are called equivalent if there exists an invertible bounded linear map $L : V_1 \rightarrow V_2$ such that $L \circ \pi_1(g) = \pi_2(g) \circ L$ for all $g \in G$. Using this relation, the set of irreducible representations of G is partitioned into equivalence classes.

The central result of harmonic analysis on compact groups is the Peter–Weyl theorem. For the current context, it can be stated as follows. Let $\mathrm{Irr}(G)$ be the set of equivalence classes of irreducible representations of G. $\mathrm{Irr}(G)$ is a countable set. If $\delta \in \mathrm{Irr}(G)$ then we have the two following facts. All representations of the class δ have the same finite dimension d_δ. There exists in this class a unitary representation U^δ. Choosing one such representation we can suppose that $U^\delta : G \rightarrow SU(\mathbb{C}^{d_\delta})$ with $SU(\mathbb{C}^{d_\delta})$ the group of special unitary $d_\delta \times d_\delta$ matrices. We distinguish the unit representation $\delta_0 \in \mathrm{Irr}(G)$ where $U^{\delta_0}(g) = 1$ for all $g \in G$. With this choice being fixed, we can state the Peter–Weyl theorem.

Theorem 1 (Peter–Weyl): The functions $d_\delta^{1/2} U_{ij}^\delta$ taken for $\delta \in \mathrm{Irr}(G)$ and $i, j = 1, \ldots, d_\delta$ form an orthonormal basis of $L^2(G, \mathbb{C})$.

Note that U_{ij}^δ is the usual notation for the matrix elements of U^δ. For all $f \in L^2(G, \mathbb{C})$ the theorem gives the Fourier pair

$$A_\delta = \int f(g) U^\delta(g)^\dagger d\mu(g) \tag{1}$$

$$f(g) = \sum_{\delta \in \mathrm{Irr}(G)} d_\delta \mathrm{tr}\left(A_\delta U^\delta(g)\right) \tag{2}$$

where $\dagger$ denotes the Hermitian conjugate and tr the trace. The Fourier series (2) converges in $L^2(G, \mathbb{C})$.

Consider the example $G = S^1$. It is possible to make the identification $\delta = 0, 1, \ldots$. Then $U^\delta(z) = z^\delta$ for $z \in S^1$.

Writting $z = e^{i\theta}$ for some $\theta \in [0, 2\pi]$, this gives the classical Fourier expansion of periodic functions.

We consider random objects and in particular G-valued random variables defined on a suitable probability space $(\Omega, \mathcal{A}, \mathbb{P})$. When referring to the probability density of such a random variable X, we mean a probability density $p_X \in L^2(G, \mathbb{R})$ with respect to μ. The characteristic function of a G-valued random variable is defined as follows. Compare to [5].

Definition 1: Let X be a G-valued random variable. The characteristic function of X is the map ϕ_X given by

$$\delta \mapsto \phi_X(\delta) = \mathbb{E}(U^\delta(X)) \quad \delta \in \mathrm{Irr}(G).$$

Here, $\mathbb{E}$ stands for expectation on the underlying probability space. For all $\delta \in \mathrm{Irr}(G)$, the expectation in the definition is finite since U^δ has unitary values. When X has a probability density p_X its characteristic function gives the Fourier coefficients of p_X as in (1). We have

$$\phi_X(\delta) = \mathbb{E}(U^\delta(X)) = \int p(g) U^\delta(g) d\mu(g) \quad \delta \in \mathrm{Irr}(G).$$

The following proposition 1 recalls the relation between characteristic functions and the concepts of convolution and convergence in distribution. It is a generalization of classical properties for scalar random variables. Remember that a sequence $(X_n)_{n \geq 1}$ of G-valued random variables is said to converge in distribution to a random variable X if for all real-valued continuous function f on G, we have

$$\lim_n \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)).$$

The proof of proposition 1 is straightforward. See [5].

Proposition 1: The following two properties hold.
1) Let X and Y be independent G-valued random variables and let $Z = XY$. We have for all $\delta \in \mathrm{Irr}(G)$

$$\phi_Z(\delta) = \phi_X(\delta)\phi_Y(\delta).$$

2) A sequence $(X_n)_{n \geq 1}$ of G-valued random variables converges in distribution to a random variable X *iff* for all $\delta \in \mathrm{Irr}(G)$

$$\lim_n \phi_{X_n}(\delta) = \phi_X(\delta).$$

In order to solve our estimation problem in Section IV we will require random variables to have certain symmetry properties. We deal with these properties here. The following analysis draws on Liao [12], [15].

We will say that a G-valued random variable X is inverse invariant if $X \overset{d}{=} X^{-1}$. We will say that it is conjugate invariant if for all $k \in G$ we have that $X \overset{d}{=} kXk^{-1}$. As usual $\overset{d}{=}$ denotes equality in distribution. The following proposition 2 characterizes these two symmetry properties in terms of characteristic functions. It will be important to remember that for any two G-valued random variables X and Y we have $X \overset{d}{=} Y$ *iff*

$\phi_X = \phi_Y$. This results from the completeness of the basis given by the U^δ as stated in the Peter–Weyl theorem [5].

Proposition 2: The following properties hold.

1) X is inverse invariant *iff* for all $\delta \in \mathrm{Irr}(G)$ we have that $\phi_X(\delta)$ is Hermitian.

2) Let X be inverse invariant. If $X_1, \ldots, X_n$ are independent copies of X then the product $X_1 \ldots X_n$ is inverse invariant.

3) X is conjugate invariant *iff* for all $\delta \in \mathrm{Irr}(G)$ we have that $\phi_X(\delta) = a_\delta I_{d_\delta}$ where $a_\delta \in \mathbb{C}$ and I_{d_δ} is the $d_\delta \times d_\delta$ identity matrix.

4) If X and Y are independent and conjugate invariant then XY is conjugate invariant.

5) X is conjugate invariant *iff* for all G-valued random variable Y independent of X we have $XY \stackrel{d}{=} YX$.

Proof:

1) Note that for all $\delta \in \mathrm{Irr}(G)$ we have by the homomorphism property of U^δ and the fact that it has unitary values

$$\phi_{X^{-1}}(\delta) = \mathbb{E}(U^\delta(X^{-1})) = \mathbb{E}(U^\delta(X))^\dagger = \phi_X(\delta)^\dagger.$$

2) This follows from 1 of proposition 2 and 1 of proposition 1, since the powers of a Hermitian matrix are Hermitian.

3) Note that for all $k \in G$ we have that $X \stackrel{d}{=} kXk^{-1}$ *iff* for all $\delta \in \mathrm{Irr}(G)$

$$\mathbb{E}(U^\delta(X)) = \mathbb{E}(U^\delta(kXk^{-1})) = U^\delta(k)\mathbb{E}(U^\delta(X))U^\delta(k)^\dagger$$

identifying ϕ_X on both sides, this becomes

$$\phi_X(\delta) = U^\delta(k)\phi_X(\delta)U^\delta(k)^\dagger.$$

If this relation is verified for all $k \in G$ then $\phi_X(\delta)$ is a multiple of I_{d_δ}. This follows by Schur's lemma [13].

4) This follows from 3 of proposition 2 and 1 of proposition 1.

5) The *if* part follows by setting $Y = k \in G$ for arbitrary k. The *only if* part follows from 3 of proposition 2 and 1 of proposition 1. ∎

Article 1 of proposition 2 motivates a practical recipe for generating inverse invariant random variables from general random variables. Let X and Y be G-valued random variables. Suppose X and Y are independent with $Y \stackrel{d}{=} X^{-1}$. It can be verified by 1 of proposition 2 that $XY \stackrel{d}{=} YX$ and that both these products are inverse invariant. In practice, if we have generated X then we can immediately generate Y as above. In this way, an inverse invariant XY or YX is generated from X.

III. COMPOUND POISSON PROCESSES

Compound Poisson processes on groups naturally generalize scalar compound Poisson processes. They are introduced by Applebaum in [3]. Let us start by recalling the definition of scalar 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.* $\mathbb{R}$-valued random variables. Suppose the family $(x_n)_{n \geq 1}$

is itself independent of N. The following process y is said to be a compound Poisson process

$$y(t) = \sum_{n=0}^{N(t)} x_n$$

G-valued 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.* G-valued random variables and suppose as before that the family $(X_n)_{n \geq 1}$ is independent of N. The following process Y is said to be a G-valued left compound Poisson process

$$Y(t) = \prod_{n=0}^{N(t)} X_n.$$

We understand that products are ordered from left to right. It is possible to obtain a right compound Poisson process by considering $Y(t)^{-1}$ instead. Thus, the two concepts are equivalent. See [12] and [3].

Before going on, we make the following remark on the above definition of compound Poisson processes. This definition was stated for G a compact connected Lie group. This topological and manifold structure of G is not necessary for the definition, which can be stated in its above form for any group with a measurable space structure. The compact group structure of G allows us to use the Peter–Weyl theorem and characteristic functions. The connected Lie group structure allows the introduction of Brownian noise in Section IV, see [12].

We wish to 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 $\mathbb{P}(Y(t) = e) \geq \mathbb{P}(N(t) = 0) = e^{-\lambda t}$. It follows that $Y(t)$ has an atom at e. In the absence of a probability density, we study $Y(t)$ for $t \geq 0$ using its characteristic function. This is given in the following Proposition 3 which can be seen to immediately extend the well known formula for scalar compound Poisson processes. This proposition follows [12], [3].

Proposition 3: For all $t \geq 0$ the characteristic function $\phi_{Y(t)}$ of $Y(t)$ is given by

$$\phi_{Y(t)}(\delta) = \exp(\lambda t(\phi_X(\delta) - I_{d_\delta})) \qquad (3)$$

for $\delta \in \mathrm{Irr}(G)$, where $\phi_X \equiv \phi_{X_1}$.

Proof: Let $t \geq 0$. $\phi_{Y(t)}$ can be calculated by conditioning over the values of $N(t)$. Using the independence of N and $(X_n)_{n \geq 1}$ we have for $\delta \in \mathrm{Irr}(G)$

$$\phi_{Y(t)}(\delta) = e^{-\lambda t} \sum_{n \geq 0} \frac{(\lambda t)^n}{n!} \mathbb{E} \prod_{m=0}^{n} U^\delta(X_m).$$

Using the fact that $(X_n)_{n \geq 1}$ are *i.i.d.* it is possible to replace

$$\mathbb{E} \prod_{m=0}^{n} U^\delta(X_m) = \prod_{m=0}^{n} \mathbb{E}(U^\delta(X_m)) = \phi_X(\delta)^n$$

the proposition follows by rearranging the sum. ∎

Combining Propositions 3 and 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.

Proposition 4: For all $t \geq 0$ we have:
1) if X_1 is inverse invariant, then so is $Y(t)$;
2) if X_1 is conjugate invariant, then so is $Y(t)$.

We end this section with Proposition 5. It gives a property of uniformization of the distribution of $Y(t)$ as $t \uparrow \infty$. This is similar to the behavior of the products $X_1 \ldots X_n$ for $n \uparrow \infty$, see [5]. For a more general version of Proposition 5 see [12], [15]. We say that a G-valued random variable X is supported by a measurable subset S of G if $\mathbb{P}(X \in S) = 1$. If X and X' are G-valued random variables with $X \stackrel{d}{=} X'$ then X is supported by S *iff* X' is supported by S. In Proposition 5, U is a G-valued random variable with probability density identically equal to 1. That is, U is uniformly distributed on G.

Proposition 5: If X_1 is not supported by any closed proper subgroup S of G or coset gS, $g \in G$ of such a subgroup then $Y(t)$ converges in distribution to U as $t \uparrow \infty$.

Proof: Under the conditions of the proposition we have for all $\delta \neq \delta_0$ that the eigenvalues of $\phi_X(\delta)$ are all < 1 in modulus [5]. It follows that the eigenvalues of $\phi_X(\delta) - I_{d_\delta}$ all have negative real parts. Thus, when $\delta \neq \delta_0$ we have by (3) that $\phi_{Y(t)}(\delta) \to 0$ as $t \uparrow \infty$. Moreover, it is immediate that $\phi_{Y(t)}(\delta_0) = 1$ for $t \geq 0$. We conclude using 2 of Proposition 1. Note that [13]

$$\phi_U(\delta) = \int U^\delta(g) d\mu(g) = 0 \quad \delta \neq \delta_0$$

and $\phi_U(\delta_0) = 1$ trivially. ∎

IV. DECOMPOUNDING

In existing literature, *decompounding* refers to a set of nonparametric estimation problems involving scalar compound Poisson processes [1], [2]. In this section, we will consider the generalization of these problems to compound Poisson processes on compact Lie groups. The new problems can be stated in the notation of Section III. We refer to them also as decompounding problems. As in the scalar case, they consist in estimation of the common probability density (supposed to exist) of the random variables X_n from observations of the process Y. The unknown common probability density of the X_n will be denoted p. We are unaware of any work on similar problems for vector-valued compound Poisson processes. Our consideration of compact Lie groups is motivated by the applications presented in Section V.

A. Typology of Decompounding Problems

Several decompounding problems can be stated, depending on the nature of the observations made of Y [2]. Decompounding is performed from *high-frequency* observations if an individual trajectory of the process Y is observed over time intervals $[0, T]$ where $T \uparrow \infty$. It is performed from *low-frequency* observations if *i.i.d.* observations are made of the random variable $Y(T)$ for a fixed $T \geq 0$.

Decompounding from high- and low-frequency observations lead to different difficulties. For high-frequency observations, the problem is greatly simplified if the assumption is made that X_n does not take the value e, for any $n \geq 1$. With probability 1, a trajectory of N has infinitely many jumps over $t \geq 0$. Under the assumption we have made, all these jumps correspond to jumps of Y which we do observe. The jumps of Y then give *i.i.d.* observations of X_1 and the average time between these jumps is $1/\lambda$. In particular, it is important for high-frequency observations to take the limit $T \uparrow \infty$.

Low frequency observations do not give direct access to λ. In scalar decompounding from low-frequency observations, λ is often assumed to be known [1], [2]. In the context of a compact group G, Proposition 5 leads to a difficulty that does not appear in scalar decompounding. Under the conditions of this proposition, if low-frequency observations are made at a sufficiently large time T then these observations will be uniformly distributed on G and will have no memory of the random variables X_n.

A third intermediate type of observations is possible. It is possible to make observations of an individual trajectory of Y at regular time intervals $T, 2T, \ldots$. This is in fact equivalent to low-frequency distributions. Remember that N is a Lévy process, *i.e.*, has independent stationary increments. Moreover we have that the $(X_n)_{n \geq 1}$ are *i.i.d.* Using this, it is possible to prove that the G-valued random variables

$$Y(T), Y(T)^{-1} Y(2T), Y(2T)^{-1} Y(3T) \ldots$$

are *i.i.d.*. Thus, our observations are *i.i.d.* observations of $Y(T)$. This remark refers to the fact that Y is a left Lévy process in G [12]. We do not develop this here.

B. Noise Model for Low Frequency Observations

We will consider decompounding from low-frequency observations. $T \geq 0$ is fixed and *i.i.d.* observations $(Z_n)_{n \geq 1}$ of a noisy version Z of $Y(T)$ are available. Z is given by Y corrupted by multiplicative noise. We have the noise model

$$Z = MY(T) \tag{4}$$

where M is independent of Y. By 1 of Proposition 1 we have for the characteristic function of Z

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

The noise model is equivalent to having an initial value $Y(0) = M$ with a general distribution. We consider the case of Brownian noise. The characteristic function of M is then given by [12], [8]

$$\phi_M(\delta) = \exp\left(-\lambda_\delta \frac{\sigma^2}{2}\right) I_{d_\delta}$$

where σ^2 is a variance parameter and for $\delta \in \mathrm{Irr}(G)$ the constant λ_δ is the corresponding eigenvalue of the Laplace–Beltrami operator. In particular, $\lambda_{\delta_0} = 0$ and $\lambda_\delta > 0$ for $\delta \neq \delta_0$. It is clear from 3 of Proposition 2 that M is conjugate invariant. It follows by 4 of Proposition 2 that, as far as the distribution of Z is concerned, left and right multiplication of $Y(T)$ by the noise M are indifferent.

It is possible to construct a G-valued process ζ such that $Z \stackrel{d}{=} \zeta(T)$. The corresponding construction is well known in

the theory of group-valued Lévy processes and is referred to as interlacing [3], [12]. Here we only state this construction. Let W be a Brownian motion on G independent of N and with variance parameter $\bar{\sigma}^2$. This is a process with continuous paths and independent stationary increments. Moreover, $W(0) = e$ and for $\delta \in \mathrm{Irr}(G)$

$$\phi_{W(t)}(\delta) = \exp\left(-\lambda_\delta \frac{\bar{\sigma}^2}{2} t\right) I_{d_\delta}.$$

Let $T_0 = 0$ and suppose $(T_n)_{n \geq 1}$ are the jump times of N. The interlaced process ζ is defined as follows. We have $\zeta(0) = e$. For $t > 0$ and $n \geq 1$ we have

$$\zeta(t) = \zeta(T_{n-1})W(T_{n-1})^{-1}W(t) \quad \text{on} \quad \{T_{n-1} \leq t < T_n\}$$

where the following formula holds at each time T_n (here, $\zeta(T_n-)$ denotes the left limit at T_n):

$$\zeta(T_n) = \zeta(T_n-)X_n.$$

This definition is sufficient, since $T_n \uparrow \infty$ almost surely. The term interlacing comes from the fact that the trajectories of ζ are obtained by introducing the jumps of Y into the trajectories of W as these jumps occur. The trajectories of W are thus interlaced with the jumps of Y.

For $t \geq 0$ the characteristic function of $\zeta(t)$ is given by

$$\phi_{\zeta(t)}(\delta) = \exp\left(t\lambda\phi_X(\delta) - tI_{d_\delta}\left(\lambda + \frac{\lambda_\delta \bar{\sigma}^2}{2}\right)\right) \quad (5)$$

for $\delta \in \mathrm{Irr}(G)$. It follows that we have $Z \overset{d}{=} \zeta(T)$ if $T\bar{\sigma}^2 = \sigma^2$.

Although we do not deal with the case of high-frequency observations, we would like to end this subsection with a remark on the role of noise in this case. The trajectories of the interlaced process ζ are noisy versions of the trajectories of Y. However, these trajectories have the same jumps as the trajectories of Y. In this sense, high-frequency observations are unaltered by noise.

C. Characteristic Function Method

We present a characteristic function method for decompounding from low-frequency observations. This method extends a similar one considered in [2]. In carrying out this extension, we are guided by the properties of characteristic functions on G presented in Section II. Our observations $(Z_n)_{n \geq 1}$ and noise model (4) were described in Section IV-B. We aim to estimate the common density p of the X_n. A characteristic function method consists in constructing nonparametric estimates for p from parametric estimates for its Fourier coefficients $\phi_X(\delta)$ given for $\delta \in \mathrm{Irr}(G)$. See [8].

We suppose that λ and σ^2 are known. Equation (5) can be copied as follows:

$$\phi_Z(\delta) = \exp\left(T\lambda\phi_X(\delta) - T\bar{\lambda}I_{d_\delta}\right) \quad \delta \in \mathrm{Irr}(G) \quad (6)$$

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 $\phi_Z(\delta)$ for all $\delta \in \mathrm{Irr}(G)$. Rather than deal with this problem in general, we make the following simplifying assumption.

Assumption: X_1 is inverse invariant.

For all $\delta \in \mathrm{Irr}(G)$ we have by applying 1 of Proposition 2 and (6) to this assumption that $\phi_Z(\delta)$ is Hermitian 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 (6), it follows that

$$\phi_X(\delta) = \frac{1}{T\lambda}\mathrm{Log}\left[\phi_Z(\delta)\right] + \left(\bar{\lambda}/\lambda\right)I_{d_\delta} \quad \delta \in \mathrm{Irr}(G). \quad (7)$$

Let $\delta \in \mathrm{Irr}(G)$. It follows from definition 1 that empirical estimates of $\phi_Z(\delta)$ 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 [16]. In order to estimate $\phi_X(\delta)$ using (7), it is then important to ensure that the empirical estimates of $\phi_Z(\delta)$ are asymptotically Hermitian positive definite.

We start by defining the empirical estimates $\hat{\phi}_Z^n(\delta)$ for $\delta \in \mathrm{Irr}(G)$ and $n \geq 1$

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

Hermitian symmetrization of empirical estimates is necessary for the application of (7). Since it is a projection operator, this symmetrization moreover contributes to a faster convergence of the $\hat{\phi}_Z^n(\delta)$ to $\phi_Z(\delta)$.

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 [17]. For a remarkably straightforward statement, see [18]. For a complex matrix C we will denote $\lambda(C)$ its spectrum. For each $\delta \in \mathrm{Irr}(G)$ and $n \geq 1$ define the event R_δ^n by

$$R_\delta^n = \left\{\lambda\left(\hat{\phi}_Z^n(\delta)\right) \subset]0, \infty[\right\}.$$

For $\delta \in \mathrm{Irr}(G)$, the sequence $(R_\delta^n)_{n \geq 1}$ controls the convergence of the spectra of the empirical estimates $\hat{\phi}_Z^n(\delta)$. In particular, since $\lim_n \hat{\phi}_Z^n(\delta) = \phi_Z(\delta)$ almost surely

$$\mathbb{P}\left(\cup_{n \geq 1} \cap_{m \geq n} R_\delta^m\right) = \lim_n \mathbb{P}\left(\cap_{m \geq n} R_\delta^m\right) = 1.$$

Using the events R_δ^n we can write down well defined estimates of ϕ_X. These are denoted $\hat{\phi}_X^n(\delta)$ for $\delta \in \mathrm{Irr}(G)$ and $n \geq 1$

$$\hat{\phi}_X^n(\delta) = 0 \quad \text{on} \quad \Omega - R_\delta^n$$
$$\hat{\phi}_X^n(\delta) = \frac{1}{T\lambda}\mathrm{Log}\left[\hat{\phi}_Z^n(\delta)\right] + \left(\bar{\lambda}/\lambda\right)I_{d_\delta} \quad \text{on} \quad R_\delta^n.$$

This expression gives our parametric estimates for the Fourier coefficients of p. We use them to construct nonparametric estimates based on an expression of the form (2). Let $(\Gamma_l)_{l \geq 1}$ be an increasing sequence of finite subsets $\Gamma_l \subset \mathrm{Irr}(G)$ with the limit $\cup_{l \geq 1}\Gamma_l = \mathrm{Irr}(G) - \{\delta_0\}$. Let $K \geq 0$ and for each $\delta \in \mathrm{Irr}(G)$ define

$$f_\delta = d_\delta e^{-K\lambda_\delta}.$$

For $n \geq 1$ and $l \geq 1$, our nonparametric estimate $\hat{p}_l^n$ is given by

$$\hat{p}_l^n(g) = 1 + \sum_{\delta \in \Gamma_l} f_\delta \mathrm{tr}\left(\hat{\phi}_X^n(\delta) U^\delta(g)^\dagger\right) \quad g \in G. \quad (8)$$

The subscript $l \geq 1$ corresponds to a cutoff or smoothing parameter. Indeed, infinitely many representations are excluded from the sum over Γ_l. A more complete expression of this fact appears in [8]. When $K > 0$ the coefficients f_δ form a convolution mask ensuring that the estimates $\hat{p}_l^n$ can be taken to converge to a smooth probability density. We make this more precise in Section IV-D.

It is usual to rewrite expressions similar to (8) in terms of a group invariant kernel. See [8] and [9]. 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 $\hat{\phi}_X^n(\delta)$ as given above.

D. Convergence of Parametric and Nonparametric Estimates

Here, we discuss the convergence of the parametric and nonparametric estimates given in Section IV-C. Our argument is presented in the form of Propositions 6 and 7 below. Proposition 6 gives the consistency of the parametric estimates $\hat{\phi}_X^n(\delta)$. Proposition 7 states a subsequent result for the nonparametric estimates $\hat{p}_l^n$.

For Proposition 6, we will need inequalities (9) and (10). 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

$$\sum_{i=1}^{d} (\beta_i - \alpha_i)^2 \leq |B - A|^2 \quad (9)$$

where $|\cdot|$ is the Euclidean matrix norm. This inequality is known as the Wielandt–Hoffman theorem. In [17], 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:

$$|\mathrm{Log}(B) - \mathrm{Log}(A)| \leq \sqrt{d}k^{-2}|B - A|. \quad (10)$$

In order to obtain (10), it is possible to start by expressing $\mathrm{Log}(A)$ as follows:

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

This expression results from a similar one for the real logarithm applied to each eigenvalue of A. Subtracting the same expression for $\mathrm{Log}(B)$, (10) follows by simple calculations.

Proposition 6: For all $\delta \in \mathrm{Irr}(G)$ we have the limit in probability $\lim_n \hat{\phi}_X^n(\delta) = \phi_X(\delta)$.

Proof: We only need to consider $\delta \neq \delta_0$. Indeed, $\hat{\phi}_X^n(\delta_0) = \phi_X(\delta_0) = 1$ for all $n \geq 1$. Let $\delta \neq \delta_0$, for all $n \geq 1$, we have

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

where $|\cdot|_{op}$ is the operator matrix norm. Passing to the limit, we have the same inequality for $\phi_Z(\delta)$. It follows that all eigenvalues of $\hat{\phi}_Z^n(\delta)$ or $\phi_Z(\delta)$ are ≤ 1. Since $\phi_Z(\delta)$ is positive definite, there exists $k_\delta > 0$ such that $\lambda(\phi_Z(\delta)) \subset [k_\delta, 1]$. For $n \geq 1$, let $\tilde{R}_\delta^n$ be the event

$$\tilde{R}_\delta^n = \left\{\lambda\left(\hat{\phi}_Z^n(\delta)\right) \subset [k_\delta/2, 1]\right\}.$$

From inequality (9), we have

$$\mathbb{P}(\Omega - \tilde{R}_\delta^n) \leq \mathbb{P}\left(\left|\hat{\phi}_Z^n(\delta) - \phi_Z(\delta)\right| > k_\delta/2\right).$$

Since $\tilde{R}_\delta^n \subset R_\delta^n$, it follows from inequality (10) that

$$\mathbb{P}\left(\left|\hat{\phi}_X^n(\delta) - \phi_X(\delta)\right| > \varepsilon \cap \tilde{R}_\delta^n\right)$$
$$\leq \mathbb{P}\left(\left|\hat{\phi}_Z^n(\delta) - \phi_Z(\delta)\right| > k_\delta^2 \varepsilon/L\right)$$

for all $\varepsilon > 0$, where $L = 4\sqrt{d_\delta}/T\lambda$.

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

$$\mathbb{P}\left(\left|\hat{\phi}_X^n(\delta) - \phi_X(\delta)\right| > \varepsilon\right) \leq \left(\frac{8 + 2L^2/\varepsilon^2}{n}\right)\left(\frac{\sqrt{d_\delta}}{k_\delta^2}\right)^2 \quad (11)$$

for all $\varepsilon > 0$. ∎

Proposition 7 relies on Proposition 6 and the Peter–Weyl theorem. It implies the existence of sequences $(\hat{p}_k)_{k \geq 1}$, of nonparametric estimates given by (8), converging to p in probability in $L^2(G, \mathbb{C})$ with any prescribed rate of convergence. Convergence in probability in $L^2(G, \mathbb{C})$ means that the following limit in probability holds:

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

where $\|\cdot\|$ is the $L^2(G, \mathbb{C})$ norm. It is clear from (8) that for all $k \geq 1$, we have $\hat{p}_k \in L^2(G, \mathbb{C})$. In order to obtain nonparametric estimates in $L^2(G, \mathbb{R})$ and converging to p in the same sense, it is enough to consider the real parts of the $\hat{p}_k$. The following proof of Proposition 7 uses Plancherel's formula as in [8].

Proposition 7: Putting $K = 0$ in (8), we have the limit in probability

$$\lim_l \lim_n \|\hat{p}_l^n - p\| = 0.$$

Proof: For $l \geq 1$, let $p_l \in L^2(G, \mathbb{C})$ be given by

$$p_l(g) = 1 + \sum_{\delta \in \Gamma_l} \mathrm{tr}\left(\phi_X(\delta) U^\delta(g)^\dagger\right)$$

for $g \in G$. By the Peter–Weyl theorem, $\lim_l \|p_l - p\| = 0$. By (8) and Proposition 6, we have $\lim_n \|\hat{p}_l^n - p_l\| = 0$ in probability for all $l \geq 1$. The proposition follows by observing that

$$\|\hat{p}_l^n - p\|^2 = \|\hat{p}_l^n - p_l\|^2 + \|p_l - p\|^2 \qquad (12)$$

for all $n, l \geq 1$. ∎

Proposition 6 obtained convergence in probability of the parametric estimates $\hat{\phi}_X^n(\delta)$ for all $\delta \in \mathrm{Irr}(G)$. 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 $\tilde{\phi}_X^n(\delta)$ converging in the square mean to the same limits $\phi_X(\delta)$. For $\delta \in \mathrm{Irr}(G)$ and $n \geq 1$, the $\tilde{\phi}_X^n(\delta)$ are given by

$$\tilde{\phi}_X^n(\delta) = 0 \ \text{ on } \ \Omega - \tilde{R}_\delta^n$$
$$\tilde{\phi}_X^n(\delta) = \frac{1}{T\lambda} \mathrm{Log} \left[\hat{\phi}_Z^n(\delta) \right] + \left(\bar{\lambda}/\lambda \right) I_{d_\delta} \ \text{ on } \ \tilde{R}_\delta^n$$

where the events $\tilde{R}_\delta^n$ are as in the proof of Proposition 6, and we assume known *a priori* constants k_δ necessary for their definition. As in (8), we can define nonparametric estimates $\tilde{p}_l^n$ where for $n, l \geq 1$

$$\tilde{p}_l^n(g) = 1 + \sum_{\delta \in \Gamma_l} f_\delta \mathrm{tr} \left(\tilde{\phi}_X^n(\delta) U^\delta(g)^\dagger \right) \quad g \in G.$$

For all $\delta \in \mathrm{Irr}(G)$ and $n \geq 1$, we have

$$\mathbb{E} \left| \tilde{\phi}_X^n(\delta) - \phi_X(\delta) \right|^2 \leq \frac{L'}{n} \left(\frac{d_\delta}{k_\delta^2} \right)^2 \qquad (13)$$

where L' is a constant depending on the product $T\lambda$. This follows by a reasoning similar to the proof of Proposition 6. Moreover, for all $n, l \geq 1$, we have after putting $K = 0$

$$\mathbb{E} \|\tilde{p}_l^n - p\|^2 \leq \frac{L'}{n} \sum_{\delta \in \Gamma_l} (d_\delta^3 / k_\delta^4) + \|p_l - p\|^2 \qquad (14)$$

for the functions p_l defined in the proof of Proposition 7. This follows from Plancherel's formula in (12).

We have characterized the convergence of parametric estimates using (11) and (13) and the convergence of nonparametric estimates using (12) and (14). We make the following remarks on these formulae. Inequalities (11) and (13) only give gross bounds for the rate of convergence of parametric estimates. The quality of these bounds improves when the constants k_δ are greater, *i.e.*, closer to the value 1. This is equivalent to the $L^2(G, \mathbb{R})$ distance between p and the uniform density being greater. This last point can be appreciated in relation to the example of Fig. 3 in Section V-C.

Equations (12) and (14) describe the convergence of nonparametric estimates in a way similar to the one used in [8]. 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 [8], both related to the indirect nature of our observations. First, the first and second terms in (14) can not be identified as the "variance" and "bias" of $\tilde{p}_l^n$. Second, (14) characterizes the nonparametric estimation error as depending on the whole spectrum of p—through the constants k_δ—rather than just its smoothness properties.

We finally return to the role of the parameter K introduced in (8). For simplicity, we have put $K = 0$ for Proposition 7 and inequality (14). Let $K > 0$. The following function $p_K \in L^2(G, \mathbb{R})$ is an infinitely differentiable probability density, compare to [12], [8]

$$p_K(g) = 1 + \sum_{\delta \neq \delta_0} f_\delta \mathrm{tr}(A_\delta U^\delta(g)^\dagger). \qquad (15)$$

Using the same K in (8) and proceeding as for proposition 7, it is possible to obtain the limit in probability

$$\lim_l \lim_n \|\hat{p}_l^n - p_K\| = 0.$$

A similar limit also holds for the $\tilde{p}_l^n$. Note that in addition to being smooth, p_K can be chosen arbitrarily close to p in $L^2(G, \mathbb{R})$ for $K > 0$ small enough.

V. DECOMPOUNDING ON $SO(3)$ AND MULTIPLE SCATTERING

This section fulfills two goals. First, it summarizes recent use of compound Poisson processes on the rotation group $SO(3)$ in the modelling of multiple scattering and introduces decompounding on $SO(3)$ as a physical inverse problem. Second, it illustrates the characteristic function method presented in Section IV-C by applying it to a numerical example of decompounding on $SO(3)$. Nonparametric estimation on the rotation group $SO(3)$ has received special attention [11], [9]. It is important to many concrete applications and constitutes a privileged starting point for generalization to compact groups.

A. Compound Poisson Model for Multiple Scattering

Many experimental and applied settings aim to infer the properties of complex, *e.g.,* geophysical or biological, media by considering multiple scattering of mechanical or electromagnetic waves by these media. Inference problems arising in this way are formulated as physical inverse problems within the framework of various approximations of the exact equations of radiative transfer. See [19]–[21].

A compound Poisson model for the direct problem of multiple scattering was considered by Ning *et al.* [22]. It is based on a $\mathbb{R}$-valued compound Poisson process. Consideration of compound Poisson processes on $SO(3)$ leads to a model of multiple scattering which is sufficiently precise as well as amenable to statistical treatment. This model extends the validity of the small angles approximation of radiative transfer. It also allows the formulation of the physical inverse problem of multiple scattering as a statistical nonparametric estimation problem.

We give an example expanding the above discussion. The development of Section III is converted into the terminology of radiative transfer, see [23]. Certain usual results in harmonic analysis on $SO(3)$ are here referred to freely. They are set down in a precise form in Section V-B.

A scalar plane wave is perpendicularly incident upon a plane parallel multiple scattering layer of thickness H. The velocity of the wave in the layer is normalized so that we have $\tau = \ell$ for the mean free time τ and mean free path ℓ. Coordinates and time origin are chosen so that the wave enters the layer at time 0 with direction of propagation $s(0) = (0, 0, 1)$. After time t in the layer this direction of propagation becomes $s(t) = (s^1(t), s^2(t), s^3(t))$. This is considered to be a random variable with values on the unit sphere $S^2 \subset \mathbb{R}^3$. The distribution of the random variable $s(H)$ is denoted I_H. It is identified with the normalized angular pattern of intensity transmitted by the layer. We return below to the validity of this identification.

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 present at random emplacements throughout the layer. The random number of scattering events up to time $0 \le t \le H$ will be denoted $N(t)$. Suppose the nth scattering event takes place at the time $0 \le T_n \le H$. This affects the direction of propagation as follows:

$$s(T_n) = s(T_n-)X_n. \quad (16)$$

Here, X_n is a random variable with values in $SO(3)$. It is identified with a random orthogonal matrix. Formula (16) is understood as a matrix equality where $s(T_n)$ and $s(T_n-)$ are line vectors. From (16) and the definition of $N(t)$ we can write for $0 \le t \le H$

$$s(t) = s(0) \left(\prod_{n=0}^{N(t)} X_n \right). \quad (17)$$

A certain number of standard physical hypotheses can be incorporated in (17). This will allow for the random product therein to be exhibited as a conjugate invariant compound Poisson process on $SO(3)$.

Under the condition $\ell \ll H$ it is possible to make the hypothesis that the time between successive scattering events has an exponential distribution [21]. This allows us to model $N(t)$ as a Poisson process with parameter $1/\ell$. Moreover, we suppose the scatterers identical and scattering events independent. This amounts to taking the $SO(3)$-valued 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 (17) can be rewritten for $0 \le t \le H$ as

$$s(t) = s(0)Y(t) \quad (18)$$

where Y is a (left) compound Poisson process on $SO(3)$ with parameter $1/\ell$. It is usual to assume that the random variables X_n have a common probability density p. For homogeneity with Section IV, we mention that p is a square integrable probability density with respect to the Haar measure of $SO(3)$. In the theory of radiative transfer, p is known as the phase function of the layer [23].

In order to simplify the Fourier series of p to a Legendre series (22) we profit from the physical hypothesis of statistical isotropy. This implies that scattering events in the layer as given by (16) are symmetric around the direction of propagation

$s(T_n-)$. Statistical isotropy is a valid assumption in a plurality of 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 [24].

Under the hypothesis of statistical isotropy the phase function p is a zonal function in the sense precised in Section V-B. It admits a Legendre series (22) wherein the coefficients a_δ for $\delta \in \mathbb{N}$ are said to form the associated power spectrum of heterogenities [23]. If p is the Henyey–Greenstein phase function then the power spectrum of heterogenities is given by $a_\delta = g^\delta$ for $\delta \in \mathbb{N}$ and p can be expressed in the closed form [24], [25]

$$p(\cos\theta) = \frac{1 - g^2}{(1 + g^2 - 2g\cos\theta)^{\frac{3}{2}}}. \quad (19)$$

In this formula the variable $\theta \in [0, \pi]$ refers to the scattering angle from an individual scatterer. It is given a mathematical definition in formula (22) of Section V-B. 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 θ. For the scattering of light waves by water clouds and blood, we have respectively $g = 0.85$ and $g = 0.95$, see [25].

Proposition 3 of Section III can be used to give the angular pattern of transmitted intensity I_H in terms of the power spectrum of heterogenities. This is expressed in the following (20). This relates the directly observable outcome of multiple scattering in the layer to the constitutive microscopic properties of the layer, typically quite difficult to ascertain directly. Replacing in Proposition 3 the definition of the process Y of (18) and using the Legendre series (22) of p we have

$$\frac{I_H(\theta)}{2\pi} = \sum_{\delta \ge 0} (2\delta + 1) e^{\frac{H}{\ell}(a_\delta - 1)} \int_0^\theta P_\delta(\cos\xi) \sin\xi d\xi \quad (20)$$

where $I_H(\theta)$ is the ratio of intensity transmitted within a pencil of angle 2θ around $s(0)$.

Equation (20) is well known in the small angles approximation of radiative transfer where it is derived under the assumption of strong forward scattering [23]. Mathematically, this translates into a phase function p with a sharp peak around $\theta = 0$. Our probabilistic development of (20) does not explicitly make this assumption. 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 $s(H)$.

Equation (20) is an interesting starting point for the formulation of the physical inverse problem of multiple scattering. Supposing a situation where this equation 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 [25].

Our use of compound Poisson processes on $SO(3)$ to model multiple scattering lead to the probabilistic counterpart (18) of (20). In relation to (18), the physical inverse problem inherent

to (20) is reformulated as a statistical estimation problem. This appears as the problem of decompounding on $SO(3)$ or some related parametric estimation problem. A crucial difference between the two approaches is that they proceed from different types of data.

Suppose the distribution of $s(0)$ is known and symmetric around $(0, 0, 1)$—this is the case in many experimental settings. Instead of carrying out measurements of transmitted intensity, it is possible to make observations of $s(H)$. Under the hypothesis of statistical isotropy these observations of $s(H)$ are equivalent to observations of $Y(H)$. If our objective is to estimate the phase function p then we have to deal with decompounding on $SO(3)$ from low-frequency observations of Y. In many cases, we could be interested in the power spectrum of heterogenities or some related physical parameters. We then have to deal with a parametric estimation problem.

B. Harmonic Analysis on $SO(3)$

We here make a short digression on harmonic analysis on $SO(3)$ in order to clarify the references made to this subject in Section V-A and to prepare for Section V-C. $SO(3)$ is often used as the archetype compact connected Lie group. Essentially, we will specify the Peter–Weyl theorem as stated in Section II to the case $G = SO(3)$. For the following, see [9] or the more detailed account in [4].

We use the notation of Section II. In particular, μ denotes the Haar measure of $SO(3)$. It is possible to identify $\mathrm{Irr}(SO(3)) = \mathbb{N}$ so that $d_\delta = 2\delta + 1$ for each $\delta \in \mathrm{Irr}(SO(3))$. With this identification, the most current choice of functions $U^\delta : SO(3) \to SU(d_\delta)$ can be given in analytical form using the parameterization of $SO(3)$ by Euler angles.

The ZYZ Euler angles $\varphi, \psi \in [0, 2\pi]$ and $\theta \in [0, \pi]$ are well defined coordinates only on a subset of $SO(3)$. This is, however, a dense subset in the Euclidean topology of $SO(3)$ and has Haar measure equal to 1. Let $p : SO(3) \to \mathbb{C}$. If p is continuous or $p \in L^2(SO(3), \mathbb{C})$ it follows that p can be identified with a function of the Euler angles $p \equiv p(\varphi, \theta, \psi)$. The chosen functions U^δ are extended by continuity from the following expression for their matrix elements

$$U^\delta_{ab}(\varphi, \theta, \psi) = e^{-\mathrm{i}a\varphi} d^\delta_{ab}(\cos \theta) e^{-\mathrm{i}b\psi} \tag{21}$$

for $\delta \in \mathrm{Irr}(SO(3))$ and $-\delta \le a, b \le \delta$. The notation d^δ_{ab} is used for the real-valued Wigner d-functions, which can be given in terms of the Jacobi polynomials. For $\delta \in \mathrm{Irr}(SO(3))$ we have $d^\delta_{00} = P_\delta$ the Legendre polynomial of order δ.

The Haar measure μ is expressed in the coordinates (φ, θ, ψ) as follows:

$$d\mu(\varphi, \theta, \psi) = \frac{1}{8\pi^2} \sin \theta d\varphi d\theta d\psi.$$

Suppose a function $p \in L^2(SO(3), \mathbb{C})$ is expressed in the form $p(\varphi, \theta, \psi)$. In order to obtain its Fourier coefficients, it is enough to replace the above expressions for the functions U^δ and μ in formula (1). This formula then reduces to a triple integral. By the Peter–Weyl theorem, the Fourier coefficients of p give rise to a Fourier series approximating p in $L^2(SO(3), \mathbb{C})$.

The class of zonal functions on $SO(3)$ arises in relation to the hypothesis of statistical isotropy mentioned in Section V-A.

We will say that a function $p \in L^2(SO(3), \mathbb{C})$ is zonal if $p \equiv p(\theta)$. That is, if the expression of p in the coordinates (φ, θ, ψ) depends only on θ. Zonal functions form a closed subspace of $p \in L^2(SO(3), \mathbb{C})$. If p is a zonal function then its Fourier series reduces to a Legendre series

$$p(\theta) = \sum_{\delta \ge 0} (2\delta + 1) a_\delta P_\delta(\cos \theta) \tag{22}$$

where for $\delta \ge 0$ the Legendre coefficient a_δ is given by

$$a_\delta = \frac{1}{2} \int_0^\pi p(\theta) P_\delta(\cos \theta) \sin \theta d\theta. \tag{23}$$

Identities (22) and (23) can be found as follows. Let p be a zonal function. For $\delta \in \mathrm{Irr}(SO(3))$ let A_δ be the Fourier coefficients of p obtained by taking $p = f$ in (1). The matrix elements of each A_δ are denoted A^{ab}_δ for $-\delta \le a, b \le \delta$. For all δ, a, b as above we have that A^{ab}_δ is given by

$$\frac{1}{8\pi^2} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} e^{\mathrm{i}b\varphi} p(\theta) d^\delta_{ba}(\cos \theta) e^{\mathrm{a}\psi} \sin \theta d\varphi d\theta d\psi$$

which follows using (1). Thus, for all $\delta \in \mathrm{Irr}(SO(3))$ we have that $A^{ab}_\delta \ne 0$ only if $a = b = 0$. In other words, the matrix A_δ contains at most one nonzero element. This is the diagonal element $A^{00}_\delta = a_\delta$ given by identity (23). Identity (22) follows by constructing the Fourier series of p as in (2).

C. Numerical Simulations

Here, we will illustrate the characteristic function method of Section IV-C by applying it to a numerical example of decompounding on $SO(3)$. Within this example we will consider a parametric estimation problem related to a physical inverse problem as described in Section V-A. Our example is of a compound Poisson process Y on $SO(3)$. As in Section V-A, $SO(3)$-valued random variables are identified with random orthogonal matrices. For $t \ge 0$

$$Y(t) = \prod_{n=0}^{N(t)} X_n$$

where the Poisson process N has parameter $\lambda = 0.3$ and the random variables X_n have a common probability density p given by expression (19). Four values will be considered for the parameter g in this expression: 0.85, 0.9, 0.95, and 0.99. We will put $T = 10$. 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 Section IV-C can be applied for this example. In other words, that the X_n with the proposed density p are inverse invariant. This follows from the development after identities (22) and (23). Indeed, the matrices A_δ obtained for p are diagonal with exactly one nonzero diagonal element $a_\delta = g^\delta$. Since g is real, we have that A_δ is Hermitian for all $\delta \in \mathrm{Irr}(SO(3))$. Inverse invariance follows by 1 of Proposition 2.

We will present three sets of figures. Fig. 1 is concerned with the compounding transformation of p. Fig. 2 illustrates the influence of n on parametric and nonparametric estimation errors.
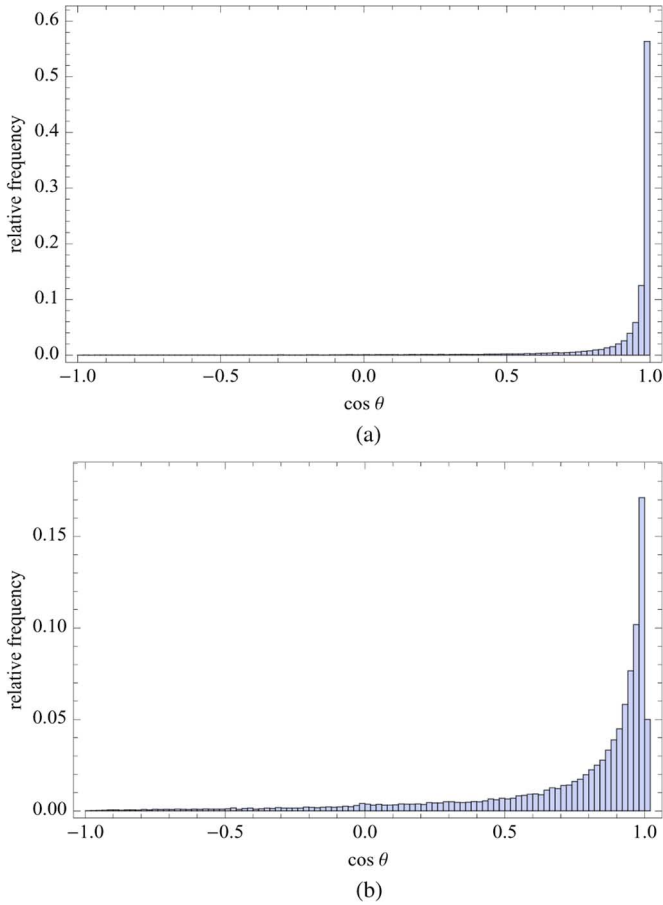
Fig. 1. Compounding transformation of p (histograms). (a) Histogram of $\cos\theta$ under density p. (b) Histogram of $\cos\theta$ under distribution of $Y(T)$.

Fig. 2. Influence of n ($\square = 5*10^2$; $\circ = 5*10^3$; $\triangle = 5*10^4$). (a) Estimated Legendre coefficients $\hat{a}_\delta^n$ from decompounding. (b) Corresponding estimates $\hat{g}_\delta^n$ of g (anisotropy parameter).

Fig. 3 studies the influence of g on the nonparametric estimation error for fixed n. For Figs. 1 and 2, we have $g = 0.9$. For Figs. 1 and 3, we have $n = 50000$. We now comment on each of these figures.

Fig. 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 Fig. 1(a) is for the cosine of the Euler angle $\theta \in [0, \pi]$ associated with the random variable X_1. The histogram in Fig. 1(b) is for the cosine of θ associated with $Y(T)$.

Fig. 1 is concerned with the direct compounding transformation rather than the inverse decompounding transformation. It is meant to show the histogram in Fig. 1(b) as function of the one in Fig. 1(a). As expected, the latter histogram appears as a wider version of the former. This corresponds to the content of Proposition 5 of Section III. Note also that the dominant value in Fig. 1(b) has moved away from $\theta = 0$.

For Fig. 2, the observations made of $Y(T)$ are used to carry out the decompounding approach of Section IV-C. Parametric and nonparametric estimation errors are given graphically for different values of n. Fig. 2(a) compares the estimated Legendre coefficients of p to their theoretical values $a_\delta = g^\delta$ for $\delta \geq 0$. In Fig. 2(b), *a priori* knowledge of the analytical form of the a_δ is supposed. This is used to estimate g. A different parametric estimate is obtained from each estimated Legendre coefficient.
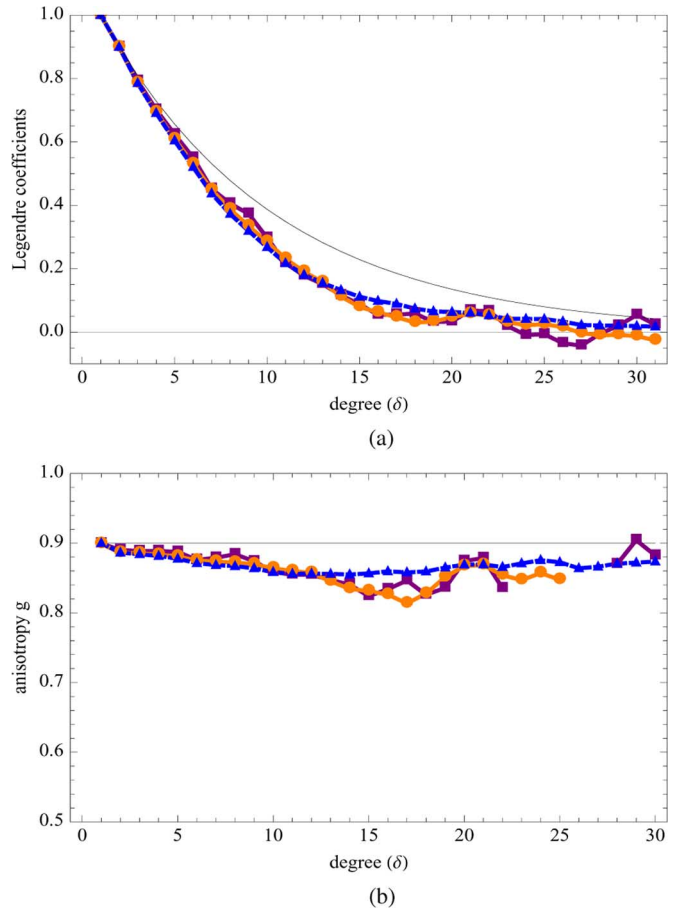
In Fig. 2(a) and (b), theoretical values are represented by a solid line.

In Fig. 2(a), we have the estimated first $l = 31$ Legendre coefficients for each value of n. Let us call these coefficients $\hat{a}_\delta^n$ for $0 \leq \delta \leq l$ and the corresponding value of n. They can be used to evaluate a nonparametric estimate of p as in formula (8). This is done by replacing them in a truncated Legendre series (22). We have the nonparametric estimate of p which we denote $\hat{p}_l^n$

$$\hat{p}_l^n(\theta) = 1 + \sum_{\delta=1}^{l-1}(2\delta + 1)\hat{a}_\delta^n P_\delta(\cos\theta)$$

where for all values of n we have that $\hat{a}_0^n = a_0 = 1$. Depending on n, the random nonparametric estimation error from $\hat{p}_l^n$ is given by

$$\sum_{\delta < l}(2\delta + 1)\left(\hat{a}_\delta^n - a_\delta\right)^2 + \sum_{\delta \geq l}(2\delta + 1)a_\delta^2$$

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

In Fig. 2(b), the estimates $\hat{a}_\delta^n$ are used to give naive estimates $\hat{g}_\delta^n$ of g based on the analytical form of the a_δ. The error in each
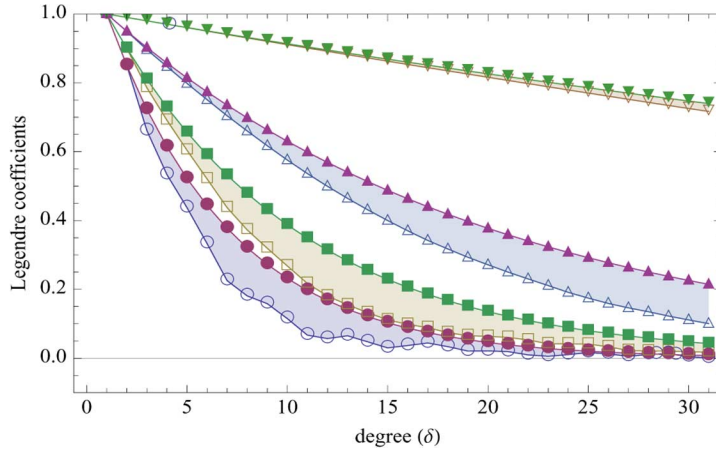
Fig. 3. Influence of g ($\circ = 0.85; \square = 0.9; \triangle = 0.95; \nabla = 0.99$).

of these estimates $\hat{g}_\delta^n$ is directly related to the error in the estimate $\hat{a}_\delta^n$. This latter error is shown for each δ and n in Fig. 2(a). The influence of n is not important for small values of δ. Visually, the $\hat{a}_\delta^n$ in Fig. 2(a) agree independently of n for $0 \le \delta \le 5$. For $n = 50000$, the $\hat{a}_\delta^n$ appear to have a regular dependence on δ. For $n = 5000$ and $n = 500$, we have an irregular dependence of the $\hat{a}_\delta^n$ on δ, especially for $\delta \ge 20$. Moreover, for $\delta \ge 25$, we have negative values of $\hat{a}_\delta^n$, clearly inconsistent with the form $a_\delta = g^\delta$. These values do not allow the evaluation of corresponding parametric estimates $\hat{g}_\delta^n$.

Let us remind that g is an important parameter in multiple scattering applications. For multiple scattering media with Henyey–Greenstein phase function (19), g is the main parameter characterizing the scattering process. Its estimation from observations as the ones described in Section V-A is equivalent to a physical inverse problem. This leads to the physical interpretation of the parametric estimation problem represented in Fig. 2(b).

For Fig. 3, we have $n = 50000$. For each value of g, we simulated n observations of $Y(T)$ and calculated estimates of the Legendre coefficients of p as for Fig. 2(a). Estimated and theoretical Legendre 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. Estimation of the Legendre coefficients is virtually exact for $g = 0.99$.

In order to understand this behavior, we note that g in (19) gives the concentration of p near the value $\theta = 0$. Indeed, when $g = 0$ the function p is constant and the random variables X_n are uniformly distributed on $SO(3)$. 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 $n > 0$ we have the conditional probability density for the Euler angle θ associated with $Y(T)$

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

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 on $SO(3)$.

Let us note that in our example $\mathbb{P}(N(T) > 0) \simeq 0.96$. Fig. 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 Legendre 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 on $SO(3)$ and our approach breaks down due to numerical problems.

VI. CONCLUSION

Nonparametric estimation on compact Lie groups, especially using characteristic function methods, is by now a relatively familiar topic in relation to several engineering applications. It has received comprehensive treatment in the case where estimation is carried out directly from some stationary process. That is, from *i.i.d.* observations of a group-valued random variable. This paper has applied a characteristic function method to the problem of decompounding on compact Lie groups. For this problem, nonparametric estimation is required from indirect observations defined in terms of a nonstationary process.

A first approach of decompounding on compact Lie groups was given. It was guided by existing characteristic function methods for the classical problem of decompounding. These methods were transposed directly to the setting of harmonic analysis on compact Lie groups. Under a suitable symmetry assumption, treatment of the indirect nature of observations was simplified. The ensuing nonparametric estimation error was characterized as depending on the whole spectrum of the target density rather than just its smoothness class. In some aspects, our approach of decompounding on compact Lie groups might appear summary. We hope, however, that is will attract attention to various problems of the statistics of nonstationary stochastic processes on groups.

This paper also discussed the importance of decompounding on $SO(3)$ to the physical inverse problem of multiple scattering. Under a probabilistic interpretation of the theory of radiative transfer, models based on compound Poisson processes on $SO(3)$ were found consistent with the results of the

small angles approximation of radiative transfer. The possibility of reformulating physical inverse problems of multiple scattering as parametric or nonparametric statistical estimation problems was discussed. The statistical nature of this new point of view seems desirable given the high complexity of multiple scattering situations. In practice, it might require considerably more elaborate measurements.

REFERENCES

[1] B. Buchmann and R. Grübel, "Decompounding: An estimation problem for Poisson random sums," *Ann. Statist.*, vol. 31, no. 4, pp. 1054–1074, 2003.

[2] B. van Es, S. Gugushvili, and P. Spreij, "A Kernel type nonparametric density estimator for decompounding," *Bernoulli*, vol. 13, no. 3, pp. 672–694, 2007.

[3] D. Applebaum, "Compound Poisson processes and Lévy processes in groups and symmetric spaces," *J. Theoret. Probab.*, vol. 13, no. 2, pp. 383–425, 2000.

[4] G. Chirikjian and A. Kyatkin, *Engineering Applications of Noncommutative Harmonic Analysis*. Boca Raton, FL: CRC, 2000.

[5] U. Grenander, *Probabilities on Algebraic Structures*. Hoboken, NJ: Wiley, 1963.

[6] H. Heyer, *Probability Measures on Locally Compact Groups*. New York: Springer Verlag, 1977.

[7] B. Yazici, "Stochastic deconvolution over groups," *IEEE Trans. Inf. Theory*, vol. 50, pp. 494–510, 2004.

[8] J.-Y. Koo and P. Kim, "Asymptotic minimax bounds for stochastic deconvolution over groups," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 289–298, Jan. 2008.

[9] P. T. Kim and J.-Y. Koo, "Optimal spherical deconvolution," *J. Multivariate Anal.*, vol. 80, pp. 21–42, 2002.

[10] P. Kim and D. Richards, "Deconvolution density estimation on compact Lie groups," in *Algebraic Methods in Statistics and Probability*. Providence, RI: AMS, 2001, pp. 155–171.

[11] P. Kim, "Deconvolution density estimation on $SO(N)$," *The Ann. Statist.*, vol. 26, no. 3, pp. 1083–1102, 1998.

[12] M. Liao, *Lévy Processes on Lie Groups*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[13] T. Bröcker and T. tom Dieck, *Representations of Compact Lie Groups*. New York: Springer, 1985.

[14] J. Duistermaat and J. Kolk, *Lie Groups*. New York: Springer Verlag, 2000.

[15] M. Liao, "Lévy processes and Fourier analysis on compact Lie groups," *Ann. Probab.*, vol. 32, no. 2, pp. 1553–1573, 2004.

[16] O. Kallenberg, *Foundations of Modern Probability*. New York: Springer Verlag, 2002.

[17] C. Van Loan and G. Golub, *Matrix Computations*. Baltimore, MD: John Hopkins Univ. Press, 1989.

[18] J. Uherka and A. Sergott, "On the continuous dependence of the roots of a polynomial on its coefficients," *Amer. Math. Monthly*, vol. 84, no. 5, pp. 368–370, 1977.

[19] H. Sato and M. Fehler, *Seismic Wave Propagation and Scattering in the Heterogeneous Earth*. New York: Springer, 1998.

[20] R. Xu, *Particle suspensions: Light scattering methods*. Norwell, MA: Kluwer, 2002.

[21] P. Sheng, *Wave Scattering, Localization and Mesoscopic Phenomena*. New York: Academic, 1995.

[22] X. Ning, L. Papiez, and G. Sandinson, "Compound Poisson process method for the multiple scattering of charged particles," *Phys. Rev. E*, vol. 52, no. 5, pp. 5621–5633, 1995.

[23] A. Ishimaru, *Wave Propagation and Scattering in Random Media*. New York: Academic, 1978, vol. 1,2.

[24] L. Klimes, "Correlation functions of random media," *Pure Appl. Geophys.*, vol. 159, pp. 1811–1831, 2002.

[25] A. Kokhanovsky, "Small angle approximations of the radiative transfer theory," *J. Phys. D*, vol. 30, pp. 2837–2840, 1997.

Salem Said recently completed the Ph.D. thesis at the Institut polytechnique de Grenoble, France. The subject of his thesis is "Estimation and filtering of processes in matrix Lie groups".

His general research interests include stochastic filtering and stochastic processes in manifolds. His research activity also includes stochastic modelling in polarization optics and multiple scattering.

Christian Lageman received the Dipl.-Inf. degree in computer science and the Dipl.-Math. and Dr. rer. nat. degrees in mathematics from the University of Würzburg, Germany, in 2001, 2002, and 2007, respectively.

From 2007 to 2008, he worked as a Research Fellow at the Centre for Mathematics and its Applications, The Australian National University, Canberra, Australia. From 2008 to 2009, he was on a postdoctoral fellowship at the Department of Electrical Engineering and Computer Science, University of Liège, Belgium. Since September 2009, he has been with the Department of Mathematics, University of Würzburg, Germany. His current research interests are in nonlinear control theory, optimization, and estimation on manifolds.

Nicolas Le Bihan was born in 1974 in Morlaix, France. He received the B.Sc. degree in physics from the Université de Bretagne Occidentale (UBO, Brest), France, in 1997, and the M.Sc. and Ph.D. degrees in signal processing, respectively in 1998 and 2001, both from the Institut Polytechnique de Grenoble (Grenoble INP), France.

Since 2002, he has been Chargé de Recherche at the Centre National de la Recherche Scientifique (CNRS) and is working with the Department of Images and Signals at the GIPSA-Lab (CNRS UMR 5083) in Grenoble, France.

His research interests include statistical signal processing on groups and manifolds and its applications in polarized wave physics, waves in disordered media, and geophysics.

Jonathan H. Manton (S'95–M'02–SM'03) received the B.S. (mathematics) and B.Eng. (electrical) degrees in 1995 and the Ph.D. degree in 1998, all from the University of Melbourne, Australia.

From 1998 to 2004, he was with the Department of Electrical and Electronic Engineering at the University of Melbourne. During that time, he held a Postdoctoral Research Fellowship then subsequently a Queen Elizabeth II Fellowship, both from the Australian Research Council. In 2005, he became a full Professor in the Department of Information Engineering, Research School of Information Sciences and Engineering (RSISE) at the Australian National University. From July 2006 till May 2008, he was on secondment to the Australian Research Council as Executive Director, Mathematics, Information and Communication Sciences. Currently, he holds a distinguished Chair at the University of Melbourne with the title Future Generation Professor. He is also an Adjunct Professor in the Mathematical Sciences Institute at the Australian National University. His traditional research interests range from pure mathematics (e.g., commutative algebra, algebraic geometry, differential geometry) to engineering (e.g., signal processing, wireless communications). Recently though, led by a desire to participate in the convergence of the life sciences and the mathematical sciences, he has commenced learning neuroscience. He also has extensive experience in software development.