Particle Filtering with Observations in a Manifold: A Proof of Convergence and Two Worked Examples

SALEM SAID
JONATHAN H. MANTON

Particle filtering is currently a popular numerical method for solving stochastic filtering problems. This paper outlines its application to continuous time filtering problems with observations in a manifold. Such problems include a variety of important engineering situations and have received independent interest in the mathematical literature. The paper begins by stating a general stochastic filtering problem where the observation process, conditionally on the unknown signal, is an elliptic diffusion in a differentiable manifold. Using a geometric structure (a Riemannian metric and a connection) which is adapted to the observation model, it expresses the solution of this problem in the form of a Kallianpur-Striebel formula. The paper proposes a new particle filtering algorithm which implements this formula using sequential Monte Carlo strategy. This algorithm is based on an original use of the concept of connector map, which is here applied for the first time in the context of filtering problems. The paper proves the convergence of this algorithm, under the assumption that the underlying manifold is compact, and illustrates it with two worked examples. In the first example, the observations lie in the special orthogonal group $SO(3)$. The second example is concerned with the case of observations in the unit sphere S^2.

Manuscript received February 28, 2016; released for publication March 11, 2016.

Refereeing of this contribution was handled by Gerhard Kurz.

Authors' addresses: S. Said, CNRS, Laboratoire IMS (UMR 5218), Université de Bordeaux; J. Manton, The University of Melbourne, Department of Electrical and Electronic Engineering.

1557-6418/16/$17.00 © 2016 JAIF

1. INTRODUCTION

The language of differential geometry is increasingly being used by engineers. This is due to the realisation that, in many concrete situations, the most natural mathematical model involves a nontrivial differentiable manifold. Key differential geometric concepts (tangent vectors, geodesics, etc.) have allowed for tasks such as optimisation and stochastic modeling to be approached in a unified and intuitive way [1].

This trend has lead to several variants of "particle filtering on manifolds" proposed in the literature, (for example, [2–9]). These involve discrete time filtering problems where either the unknown signal or the observation process lie in a differentiable manifold. The current paper outlines the application of particle filtering to continuous time filtering problems where the observation process takes its values in a differentiable manifold, and the unknown signal is a hidden Markov process. These are essentially different from problems where the unknown signal is a diffusion in a differentiable manifold and the observation process follows a classical additive white noise model; see Section 2.

The choice of continuous time over discrete time has technical and modeling advantages. It allows for the machinery of stochastic calculus and differential geometry to be applied. Also, the immense majority of physical models do not immediately discretise the time variable. Concretely, though, the difference between continuous and discrete time is a convenience. The question is to solve-then-discretise or discretise-then-solve. The final product in this paper, the algorithm of Section 4, is in discrete time and can be compared to any other algorithm developed directly in discrete time.

Predominantly, the differentiable manifolds appearing in engineering are classical matrix Lie groups and their symmetric spaces. The general stochastic filtering problem stated in Section 2 starts from an abstract differentiable manifold. All subsequent constructions are stated at this level of generality. It is hoped this will have the advantage of providing a deeper overall understanding. When (as a special instance of the general problem) the manifold is specified to be a matrix manifold, this allows for the complexity of differential geometric constructions to be reduced. This will be discussed again, using two concrete examples, in Section 5.

A subtheme of this paper is to realise a transfer of knowledge, from the mathematical field of stochastic differential geometry to the engineering community. Stochastic differential geometry was pioneered by Schwartz and Meyer. Among the fundamental texts, it is impossible to ignore the elegant and comprehensive account by Emery [10]. A more recent highly readable textbooks is the one by Hsu [11]. A recent account, written specifically for engineers, is [12].

Stochastic differential geometry begins with a manifold equipped with a connection. In differential geometry, connections are introduced to distinguish in an invariant way those differentiable curves which are geodesics; in other words, zero acceleration lines. Stochastic differential geometry uses a connection to distinguish those pathwise continuous processes which are martingales; in other words zero drift processes.

For the purpose of filtering, it is very useful to think in terms of *antidevelopment*. The antidevelopment of a differentiable curve in a manifold is a differentiable curve in a Euclidean space, which could be identified with the tangent space at the base point of the curve. This can be visualised when the manifold is a two dimensional surface embedded in physical space, for instance a sphere. The antidevelopment of a differentiable curve drown on the sphere is the trace that it leaves on a tangent plane while rolling without slipping (one could imagine the curve is drawn in ink so it leaves a trace on the plane). This visualisation is helpful for intuition, but it is important to keep in mind the notion of antidevelopment depends on a choice of connection. The picture of rolling without slipping corresponds to the connection inherited from the ambient space. Roughly speaking, the relation between a connection and the corresponding antidevelopment is that antidevelopment of a geodesic in a manifold is a straight line in Euclidean space. Similarly, antidevelopment of a martingale in the manifold is a martingale in Euclidean space. Required background from stochastic differential geometry is presented in Section 3.

In terms of stochastic filtering with observations in a manifold, the main problem considered in this paper, antidevelopment plays an essential role. While the observation process is a diffusion in a given differentiable manifold, its antidevelopment process takes its values in a Euclidean space. Regardless of the choice of connection, there is no loss of information in replacing the observation process by its atidevelopment. Moreover, see Proposition 1 of Section 3, an adequate choice of connection (roughly, one which is adapted to the observation model) leads to an antidevelopment process which depends on the unknown signal through a classical additive white noise model. Thus, antidevelopment can be thought of as a preprocessing, reducing the initial problem to a classical filtering problem defined by an additive white noise model. This idea of reduction was used in the engineering literature by Lo [13], in the case of observations in a matrix Lie group. In a recent paper [14], the authors were able to extend it to the general case of observations in a differentiable manifold.

The role of stochastic differential geometry in problems of stochastic filtering with observations in a manifold was exploited more extensively in the mathematical literature. Several authors have used it in deriving Zakai or filtering equations for these problems [15–17]. To the author's knowledge, on the other hand, few papers have been devoted to their numerical solution. Note, however, the paper by Pontier [18], which will be quite important in the following. This proves the convergence of a discrete time filter based on uniformly sampled observations to the solution of the continuous time problem.

The filtering problem is to compute the conditional distribution of the unknown signal given the observations. In section 4, the solution of this problem is expressed in the form of a Kallianpur-Striebel formula (see Proposition 4). This has a structure quite similar to that of the classical Kallianpur-Striebel formula. It is a Bayes formula where the prior distribution of the unknown signal is given by its Markov nature, i.e., initial distribution and transition kernel, and the likelihood function is an exponential functional of the observation process.

The proposed particle filtering algorithm implements the Kallianpur-Striebel formula using sequential Monte Carlo. There are at least two motivations for applying a sequential Monte Carlo approach. First, the fact that it is suitable for real time situations. Second, the computational stability which it provides in dealing with noisy observations over a longer time.

The algorithm follows sequential Monte Carlo strategy of sequential importance sampling with resampling. At a higher level, its main steps are exactly the same as for usual particle filtering. A fixed number of particles is used throughout. The interval of observation is subdivided into subintervals of equal length. Over each subinterval, the particles are propagated without interaction according to (an approximation of) the unknown signal's transition kernel. They are subsequently given new weights corresponding to their likelihood and resampled to eliminate particles with lower weight. For a rigorous general discussion, see Del Moral's monograph [19].

The specific role of stochastic differential geometry only appears in the computation of particle weights. This requires the use of *connector maps*. Under the assumption that the underlying manifold is compact, Proposition 3 of Section 3 shows that connector maps yield "increments" which are approximately normally distributed tangent vectors, conditionally on the unknown signal. The computation of weights then takes place as in the presence of additive normally distributed noise. Simply put, the proposed algorithm is a classical particle filter which employs the geometric concept of connector maps to locally linearise the observation process.

Proposition 5 of Section 4 states the convergence of this discrete algorithm to the solution of the continuous time problem. The convergence takes place as the number of particles tends to infinity and the length of of each subdivision interval tends to zero. It depends on the quality of the approximation being used for the unknown signal transition kernel and also on the one described in Proposition 3. As several approximations are involved, the compromise between complexity and performance should be based on the most difficult one to realise. This leads to a kind of bottleneck effect. See discussion at the end of Section 4.

Section 5 presents two examples illustrating the implementation and performance of the proposed particle filtering algorithm. In the first example, see 5.1, the observation process is a left invariant diffusion with values in the special orthogonal group $SO(3)$; conditionally on the unknown signal. The fact that the observation model is compatible with the Lie group structure of $SO(3)$ leads to a certain simplification of the notions of antidevelopment and connector maps. Roughly speaking, these just amount to application of the group logarithm map (i.e., the matrix logarithm) and this can moreover be replaced by a linear approximation, which reduces computational complexity.

In the example of 5.2, the observation process has its values in the unit sphere S^2. Here, a detailed discussion of the mathematical concepts introduced in the paper is provided and, as in the first example, it is shown how the particle filtering algorithm can be implemented in a way that reduces the complexity of geometric constructions.

These two examples are representative of problems where the observation process lies in a classical matrix Lie group or in a related symmetric space. With some adjustment, they could be extended in a straightforward way to deal with general matrix Lie groups and their symmetric spaces. This has not been possible in the current paper, essentially for a reason of space. However, for the case where the observation process takes its values in a Stiefel manifold, the reader is referred to [20].

The particle filtering algorithm proposed in this paper seems entirely new in the literature. It is encouraging that, building directly on existing results (from [14, 18] and [21]), the algorithm can be described and its convergence proved. However, the current treatment still suffers from certain drawbacks which should be addressed in future work. For instance, the restriction to compact manifolds is quite artificial and it should be possible to drop it with some additional care. Also, the convergence result of Proposition 5 does not explicitly provide a rate of convergence (see discussion after the proposition).

2. GENERAL FILTERING PROBLEM

To state a filtering problem, it is sufficient to define the signal and observation models [22]. In the following, the unknown signal will be a hidden Markov process x with values in some Polish (i.e., complete separable metric) space $(S, \mathcal{S})$. Concretely, in most cases, this space S is either a finite set or a Euclidean space. The observation process Y will be a diffusion in a differentiable manifold $\mathcal{M}$ of dimension d (see [23] for required background in differential geometry). Besides precisely giving the definition of x and Y, this section aims to put the resulting filtering problem into perspective. This is done by comparing it to other filtering problems, both classical and involving an observation process with values in a manifold, and by discussing how it can simulated numerically.

Both x and Y are defined in continuous time. One begins by introducing a complete probability space $(\Omega, \mathcal{A}, \mathbb{P})$ on which x and Y are defined.

The unknown signal x is a time invariant Markov process. Let $C_b(S)$ denote the space of bounded continuous functions $\varphi : S \to \mathbb{R}$. The generator of x is an operator A with domain $D(A)$, a dense subspace of $C_b(\varphi)$ which contains the constant function 1. The generator A is assumed to verify $A1 = 0$ and, for $\varphi \in D(A)$,

$$d\varphi(x_t) = A\varphi(x_t)dt + dM_t^{\varphi}, \tag{1}$$

where $t \geq 0$ is the time variable. The meaning of the above equation is that

$$\varphi(x_t) - \varphi(x_0) - \int_0^t A\varphi(x_s)ds = M_t^{\varphi},$$

where M^{φ} is a martingale adapted to the augmented natural filtration of x, denoted $\mathcal{X}$ [24]. The distribution of x is determined by its generator A and its initial distribution μ; i.e., μ is the distribution of the initial value x_0.

Two typical examples of the above definition are when x is a diffusion in some manifold $\mathcal{N}$ of dimension l, possibly $\mathcal{N} = \mathbb{R}^l$, and when x in a finite state Markov process. The first example may appear when the problem is applied to tracking the pose of a rigid body. The second is generally considered within the framework of change detection. It is usual to refer to x as a hidden Markov process as it is unknown and only observed through Y.

In order to define the observation process Y, assume $\mathcal{M}$ is a C^4 manifold. The signal x is encoded through the sensor function $H : S \times \mathcal{M} \to T\mathcal{M}$, where $T\mathcal{M}$ denotes the tangent bundle of $\mathcal{M}$. For $s \in S$, it is required the application $p \mapsto H(s, p)$ is a C^1 vector field on $\mathcal{M}$. That is, $H(s, p) \in T_p\mathcal{M}$ where $T_p\mathcal{M}$ is the tangent space to $\mathcal{M}$ at p. Observation noise is introduced as follows. Let $(\Sigma_r; r = 1, \ldots, m)$ be C^2 vector fields on $\mathcal{M}$ and B a standard Brownian motion in $\mathbb{R}^m$, which is independent from x. The observation process Y is assumed nonexplosive (i.e., for $p \in \mathcal{M}$, if $Y_0 = p$ then Y_t is defined with values in $\mathcal{M}$ for $t \geq 0$) and satisfying [25],

$$dY_t = H(x_t, Y_t)dt + \Sigma_r(Y_t) \circ dB_t^r. \tag{2}$$

Here and in all the following, summation convention is understood, (that is, a sum is understood over any repeated subscripts or superscripts). Since the Σ_r are vector fields on $\mathcal{M}$, the $\Sigma_r(Y_t)$ are tangent vectors to $\mathcal{M}$ at Y_t. Moreover, the circle is the usual notation for the Stratonovich differential [26]. The filtering problem defined by (1) and (2) is more general than a classical filtering problem where Y depends on x through an additive white noise model. In (2) the observation noise B is "carried" by the vector fields Σ_r. In other words, it is introduced in a way which depends on the current observation. The model (2) reduces to a classical one when $\mathcal{M} = \mathbb{R}^d$ and i) $m = d$ with $\Sigma_r(p) = e_r$, where

$(e_r; r = 1, \ldots, d)$ is a canonical basis of $\mathbb{R}^d$; ii)$H(s,p) = H(s)$ is given by a function $H : S \to \mathbb{R}^d$.

Intuitively, the classical problem is a limit of problems where the observation process is sampled at times $k\delta$ where $k \in \mathbb{N}$ and δ the sampling step size. Then, since Y has values in $\mathbb{R}^d$, it is possible to consider discrete observations

$$\text{(for observations in } \mathbb{R}^d) \quad \Delta Y_k = Y_{(k+1)\delta} - Y_{k\delta}. \tag{3}$$

Each one of these increments is normally distributed conditionally on x. Therefore, the corresponding likelihood function is known analytically. This last aspect is what characterises a classical filtering problem. It is unaffected if x is a diffusion in a manifold $\mathcal{N}$. In practice, whatever the state space S, what matters is the ability to simulate the sample path distribution of x with sufficient precision.

The chief difference between the classical filtering problem, and the problem of filtering with observations in a manifold, given by the observation model (2), is that no formula similar to (3) is immediately available in the latter case. As a result, it is difficult to arrive at an analytic expression of the likelihood function. As the standard analysis, based on the assumption that Y depends on x through an additive white noise model, does not apply to the observation model (2), it becomes necessary to search for a well-defined generalisation of (3). This generalisation, will be introduced as of the beginning of Section 3, in the form of equation (6), based on an original use of the concept of *connector map*. The systematic use of connector maps constitutes the main ingredient of the new particle filtering algorithm proposed in the present paper.

In the mathematical literature, problems with observations in a manifold have been stated in two different forms. In [14, 15] the problem statement is the same as above, with the additional requirement that Y is an elliptic diffusion. In fact, this same requirement will be imposed in Section 3. In [16, 17], the diffusion Y is specified in terms of its horizontal lift. The difference between the two settings is that, in the latter one, there is a choice of connection already included in the problem statement. In the setting of (2), a metric and connection need to be constructed from the vector fields Σ_r. The current paper differs from [15] in the way this construction is defined. See Section 3.

In a real world application, the issue of how to simulate Y numerically is irrelevant. Indeed, Y is itself observed through some measurement device. However, in order to carry out a computer experiment, it is necessary to simulate Y given the model (2). In many cases, $\mathcal{M}$ is embedded in a higher dimensional Euclidean space, $\mathcal{M} \subset \mathbb{R}^N$, and the vector fields H and Σ_r are restrictions of complete vector fields defined on all of $\mathbb{R}^N$. In particular, this is true for all the examples in Section 5. Then, numerical simulation of Y is a matter of solving a stochastic differential equation. It is possible to use Euler, Milstein or a higher order stochastic Taylor scheme, according to desired precision [27]. All these numerical schemes will suffer from the same problem of producing an approximation which "falls off" the embedded manifold. A simple way of dealing with this is to project back onto the manifold once the approximation has been computed. When $\mathcal{M}$ is presented as an abstract manifold, a generally applicable numerical scheme is the McShane approximation. This approximates Y by processes Y^δ where δ is a discretisation step size. On each interval $[k\delta, (k+1)\delta[$ for $k \in \mathbb{N}$, Y^δ is the solution of an ordinary differential equation,

$$\dot{Y}_t^\delta = H(x_t, Y_t^\delta) + \Sigma_r(Y_t)\Delta B_k^r, \tag{4}$$

where the dot denotes differentiation with respect to time and $\Delta B_k = \delta^{-1}(B_{(k+1)\delta} - B_{k\delta})$. When $\mathcal{M}$ is a compact Riemannian manifold and Y_0^δ is taken to be the same as Y_0, the rate of local uniform convergence of the paths of Y^δ to those of Y is given in [28],

$$\mathbb{P}(\sup_{t \leq T} d(Y_t^\delta, Y_t) > \varepsilon) = O(\delta), \tag{5}$$

where d is the Riemannian distance and $\varepsilon > 0$. Note that this depends on T so the same precision ε for larger T requires smaller δ.

The numerical approximation (4) gives an intuitive interpretation of the stochastic differential equation (2), whose rigorous definition may be found in [25]. Roughly, this approximation states that, over short time intervals of the form $[t, t+\delta]$, the observation process Y moves along the integral curves of a random vector field $H + \Sigma_r \Delta B^r$. This motion has a deterministic component H, which depends on the unknown signal, and a stochastic or "noise" component $\Sigma_r \Delta B^r$. These two components are not observed independently, but rather only through the observations Y_t and $Y_{t+\delta}$. From the point of view of filtering, it is natural to search for a transformation which takes these observations Y_t and $Y_{t+\delta}$ to a random vector of the form "H+normally distributed noise." This is the starting point of the following section.

3. STOCHASTIC DIFFERENTIAL GEOMETRY

The aim of this section is to generalise (3) so that it can be applied for the observation model (2). This is done by adopting the following point of view. In (3), ΔY_k is a tangent vector to $\mathcal{M} = \mathbb{R}^d$ at $Y_{k\delta}$, determined from two successive samples $Y_{k\delta}$ and $Y_{(k+1)\delta}$. Accordingly, in the general case of (2), a mapping $I : \mathcal{M} \times \mathcal{M} \to T\mathcal{M}$ will be constructed with the following properties. First, $I(p_1, p_2) \in T_{p_1}\mathcal{M}$ for all $p_1, p_2 \in \mathcal{M}$. Second, if ΔY_k is defined as

$$\Delta Y_k = I(Y_{k\delta}, Y_{(k+1)\delta}), \tag{6}$$

then, in the limit $\delta \downarrow 0$, the distribution of ΔY_k is approximately normal, conditionally on the unknown signal x. Such a mapping I is called a connector map, and it turns out that, it is always possible to choose a connector map

I which verifies the required properties. This is stated precisely in Proposition 3, which is the main result in this section.

Proposition 3 requires some basic concepts from stochastic differential geometry. These are recovered in 3.1 which, in particular, gives the definition of the antidevelopment process y of Y. The choice of connection leading to I is given in 3.2. Finally, Proposition 3 is stated and discussed in 3.3. Here, 3.1 and 3.2 are based on [14].

3.1. Stochastic antidevelopment

In the filtering problem defined by (1) and (2), the available information is a path of the observation process Y, taken over some time interval $t \leq T$. An observer is only able to compute functionals of the process Y. These are processes, in practice real or vector valued, adapted to the augmented natural filtration of Y; which is denoted $\mathcal{Y}$. Two classes of such functionals are the building blocks for the following, the Stratonovich and Itô integrals along Y. The notion of antidevelopment is itself defined using these two kinds of integrals.

Assume a Riemannian metric $\langle \cdot, \cdot \rangle$ and a compatible connection ∇ are defined on $\mathcal{M}$. In the current paragraph, these are not specified and can be chosen arbitrarily.

To define the integrands involved in Stratonovich and Itô integrals along Y, let $\mathcal{F}$ be the filtration where $\mathcal{F}_t = \mathcal{X}_\infty \vee \mathcal{B}_t$. Here, $\mathcal{X}_\infty = \vee_{t \geq 0} \mathcal{X}_t$ and $\mathcal{B}$ is the augmented natural filtration of B. A stochastic integrand θ is an $\mathcal{F}$-adapted process with values in $T^*\mathcal{M}$ which is above Y; this means $\theta_t \in T^*_{Y_t}\mathcal{M}$ for $t \geq 0$.

Let $(\theta_t, \cdot)$ denote the application of the linear form θ_t. The Stratonovich and Itô integrals of θ along Y are real valued $\mathcal{F}$-adapted processes with the following differentials,

$$(\theta_t, \circ dY) = (\theta_t, H)dt + (\theta_t, \Sigma_r) \circ dB^r_t, \tag{7}$$

$$(\theta_t, dY) = \{(\theta_t, H) + (1/2)(\theta_t, \nabla_{\Sigma_r}\Sigma_r)\}dt + (\theta_t, \Sigma_r)dB^r_t. \tag{8}$$

The Stratonovich differential does not involve the chosen connection. On the other hand, the connection appears explicitly in the Itô differential. Note that Y is called a ∇-martingale if $H + (1/2)\nabla_{\Sigma_r}\Sigma_r \equiv 0$. In this case, (θ, dY) is a martingale differential, as is the case for a usual Itô differential.

Using the Riemannian metric $\langle \cdot, \cdot \rangle$, it is possible to define Stratonovich and Itô integrals of vector fields. A vector field along Y is a process E, $\mathcal{F}$-adapted with values in $T\mathcal{M}$ and which is above Y; in the sense that $E_t \in T_{Y_t}\mathcal{M}$ for $t \geq 0$. A corresponding process θ in $T^*\mathcal{M}$ is then given by $(\theta_t, \cdot) = \langle E_t, \cdot \rangle$. The resulting differentials of (7) and (8) are written $\langle E_t, \circ dY \rangle$ and $\langle E_t, dY \rangle$.

In order to formulate the notion of antidevelopment, it is necessary to define what it means for E to be parallel (i.e., along Y). If the paths of Y were differentiable, this would have the usual meaning from differential geometry. While this is not the case, due the presence of Brownian terms, it is still possible to introduce a stochastic covariant derivative of E and require this to vanish. Thus E is said to be parallel if

$$\nabla_{\circ dY} E_t = 0, \tag{9}$$

where the left hand side is the stochastic covariant derivative. In order to give this a precise meaning, consider first the case where G is a C^1 vector field on $\mathcal{M}$ and $E = G(Y_t)$. Assuming the usual properties of a connection, (2) would give

$$\nabla_{\circ dY} G(Y_t) = \nabla_H G(Y_t)dt + \nabla_{\Sigma_r} G(Y_t) \circ dB^r_t. \tag{10}$$

This is extended to a general vector field E along Y by the following transformation. For a C^2 function f on $\mathcal{M}$, let $\nabla^2 f$ be the Hessian of f with respect to ∇. By definition, $\nabla_H G f = HGf - \nabla^2 f(H, G)$ and similarly for $\nabla_{\Sigma_r} G$. Now, $(\theta_t, \cdot) = \nabla^2 f(\cdot, E_t)$ is a process in $T^*\mathcal{M}$ as in (7). Thus, (10) can be generalised by writing

$$\nabla_{\circ dY} E_t f = dE_t f - \nabla^2 f(\circ dY, E_t), \tag{11}$$

which is taken as the definition of the stochastic covariant derivative. In integral notation, (11) defines a vector field $\int_0^t \nabla_{\circ dY} E_s$ along Y. For any C^2 function f on $\mathcal{M}$, this verifies

$$\int_0^t \nabla_{\circ dY} E_s f = E_t f - E_0 f - \int_0^t \nabla^2 f(\circ dY, E_s).$$

Recall the connection ∇ is compatible with the metric $\langle \cdot, \cdot \rangle$. Again, if the usual properties of a connection were assumed, one would expect the following. If E, K are vector fields along Y, then

$$d\langle E_t, K_t \rangle = \langle \nabla_{\circ dY} E_t, K_t \rangle + \langle E_t, \nabla_{\circ dY} K_t \rangle. \tag{12}$$

The proof that this identity indeed holds true is here omitted. The main point is that it is possible to perform calculations involving $\nabla_{\circ dY}$ jut like in differential geometry, by treating $\circ dY$ as the tangent vector to the process Y.

Now, it is possible to define the antidevelopment of Y. This is a process y which has values in $\mathbb{R}^d$. A parallel orthonormal frame is a family $(E^i) \equiv (E^i; i = 1, \ldots, d)$ of vector fields along Y which verify (9) and $\langle E^i_0, E^j_0 \rangle = \delta_{ij}$. In this case, it follows from (12) that $\langle E^i_t, E^j_t \rangle = \delta_{ij}$ for $t \geq 0$. Given a parallel orthonormal frame (E^i), the process y is defined by

$$y^i_t = \int_0^t \langle E^i_s, \circ dY \rangle. \tag{13}$$

It seems this definition is arbitrary, due to the issue of uniqueness of a parallel orthonormal frame. However, the classical uniqueness result for linear stochastic differential equations can be used to show y is essentially unique. In fact, (E^i) is determined by (E_0^i) so that different choices of (E^i) only amount to y being multiplied by an orthogonal matrix.

Stratonovich and Itô integrals along Y can be written as classical Stratonovich and Itô integrals with respect to y. For the Stratonovich integral, this is a straightforward result of (13). For θ as in (7),

$$(\theta_t, \circ dY) = \theta_i(t) \circ dy_t^i, \tag{14}$$

where $\theta_i(t) = (\theta_t, E_t^i)$. This follows by the chain rule of Stratonovich calculus, since $(\theta_t, \cdot) = \theta_i(t)\langle E_t^i, \cdot\rangle$. Recall that (E^i) is a basis of the the tangent space to $\mathcal{M}$ at Y_t.

The case of the Itô integral is slightly more involved. Here, it is necessary to realise that (13) can also be written

$$y_t^i = \int_0^t \langle E_s^i, dY\rangle. \tag{15}$$

In other words, since the vector fields E^i are parallel, there is no difference between their Stratonovich and Itô integrals. Expressing (13) as in (7), (the arguments of H and Σ_r are dropped for space)

$$\langle E_t^i, \circ dY\rangle = \langle E_t^i, H\rangle + \langle E_t^i, \Sigma_r\rangle \circ dB_t^r.$$

In order to change the Stratonovich differential into an Itô differential, note from (9) and (12)

$$d\langle E_t^i, \Sigma_r\rangle = \langle E_t^i, \nabla_{\circ dY_t}\Sigma_r\rangle,$$

and from (10)

$$d\langle E_t^i, \Sigma_r\rangle = \langle E_t^i, \nabla_H \Sigma_r\rangle dt + \langle E_t^i, \nabla_{\Sigma_v}\Sigma_r\rangle \circ dB_t^v.$$

From this, it finally follows

$$\langle E_t^i, \Sigma_r\rangle \circ dB_t^r = \langle E_t^i, \Sigma_r\rangle dB_t^r + (1/2)\langle E_t^i, \nabla_{\Sigma_r}\Sigma_r\rangle dt, \tag{16}$$

so that (15) follows from (8). Given (15), it is possible to write

$$(\theta_t, dY) = \theta_i(t) dy_t^i, \tag{17}$$

using a reasoning similar to the one that lead to (14).

At the beginning of this paragraph, it was claimed that Stratonovich and Itô integrals along Y can be used to obtain functionals of this process. Even when the process θ is $\mathcal{Y}$-adapted, this is not clear from (7) and (8). For example, the right hand side in each of these two formulae contains a stochastic integral with respect to B. However, there is no reason to believe that $\mathcal{B}$ is the same as $\mathcal{Y}$. Indeed, the definition (2) of Y involves both x and B.

It is possible to write (7) and (8) in an alternative form, which makes it evident that the resulting integrals are $\mathcal{Y}$-adapted as soon as θ is $\mathcal{Y}$-adapted. To do so, assume $\mathcal{M}$ is embedded in a higher dimensional Euclidean space, $\mathcal{M} \subset \mathbb{R}^N$. There is little loss of generality in this assumption. Whitney's theorem asserts such an embedding exists if $\mathcal{M}$ is C^∞, connected and paracompact [10]. Such conditions are always verified in practice. Also, when $\mathcal{M}$ is embedded in $\mathbb{R}^N$, it is possible to assume the vector fields H and Σ_r are restrictions of vector fields defined on all of $\mathbb{R}^N$. Under these two assumptions, let $\eta^1, \ldots, \eta^N$ be canonical (i.e., rectangular) coordinates on $\mathbb{R}^N$ and write $\theta(t) = \theta_\alpha(t) d\eta^\alpha$ where $\alpha = 1, \ldots, N$. Replacing in (7) and using the chain rule of Stratonovich calculus,

$$(\theta_t, \circ dY) = \theta_\alpha(t) \circ dY_t^\alpha, \tag{18}$$

where Y^α are the coordinates of Y_t. In other words, the Stratonovich integral (7) is just the classical Stratonovich integral. Moving on to (8), a similar formula will be shown. Note that the coordinates η^α can be thought of as C^2 functions on $\mathcal{M}$. Using the notation $\nabla^2\eta^\alpha(\Sigma_r, \Sigma_v) = H_{rv}^\alpha$,

$$(\theta_t, dY) = \theta_\alpha(t)\{dY_t^\alpha - (1/2)H_{rr}^\alpha(Y_t)dt\}. \tag{19}$$

Thus, the Itô integral (8) is the sum of the classical Itô integral, corresponding to $\theta_\alpha(t)dY_t^\alpha$, and of a correction term depending on the connection ∇. It is possible that the latter correction term vanishes identically, so the Itô integral (8) is the same as the classical Itô integral. This is for instance the case when the connection ∇ is the one which $\mathcal{M}$ inherits from $\mathbb{R}^N$. From the last two formulae, it is seen that the Itô and Stratonovich integrals are $\mathcal{Y}$-adapted as soon as the process θ is $\mathcal{Y}$-adapted. This follows from the same property for classical Itô and Stratonovich integrals.

The proof of (19) is as follows. Assume in (7) that $\theta_t = \tau(Y_t)$ for some differential form τ on $\mathcal{M}$. Passing from Stratonovich to Itô differentials,

$$(\theta_t, \Sigma_r) \circ dB_t^r = (\theta_t, \Sigma_r)dB_t^r + (1/2)\Sigma_r(\theta_t, \Sigma_r)dt.$$

Using the fact that ∇ is compatible with $\langle\cdot,\cdot\rangle$ and comparing to (8), the following general rule is found,

$$(\tau(Y_t), \circ dY) = (\tau(Y_t), dY) + (1/2)(\nabla_{\Sigma_r}\tau(Y_t), \Sigma_r)dt. \tag{20}$$

In order to obtain (19), it is enough to apply (20) with $\tau = d\eta^\alpha$ and recall the definition of $\nabla^2\eta^\alpha$, (as cited before (11)).

In many cases, the embedding $\mathcal{M} \subset \mathbb{R}^N$ is known explicitly. For instance, $\mathcal{M}$ is often directly defined as an embedded submanifold in Euclidean space. Then, formulae (18) and (19) can be used to compute the Itô and Stratonovich integrals. This is useful since the properties of classical stochastic integrals become available. This situation will apply throughout the examples of Section 5.

3.2. Le Jan-Watanabe connection

This paragraph describes the "right" choice for the metric $\langle\cdot,\cdot\rangle$ and the connection ∇. When this choice is used to evaluate (13), the resulting antidevelopment y of Y depends on the unknown signal x through a classical additive white noise model. This reduces the initial filtering problem defined by (1) and (2) to a classical filtering problem. The precise statement is given in Proposition 1.

A usual simplifying assumption imposed on Y in the literature is that, conditionally on x, it is an elliptic diffusion in $\mathcal{M}$. This means that, for each $p \in \mathcal{M}$, the subspace of $T_p\mathcal{M}$ spanned by the $\Sigma_r(p)$ is equal to $T_p\mathcal{M}$. Under this assumption, elementary linear algebra implies there exists a unique Riemannian metric $\langle \cdot,\cdot \rangle$ such that

$$\langle E,K \rangle = \langle E,\Sigma_r(p)\rangle \langle K,\Sigma_r(p)\rangle, \tag{21}$$

for $E,K \in T_p\mathcal{M}$. This will be the choice of metric made in this following. In [15], the corresponding Levi-Civita connection is used in (13). Here, a different connection is used. Namely, the connection ∇ is taken to be the Le Jan-Watanabe connection. Based on [29], this is here defined as follows.

Let E be any C^1 vector field on $\mathcal{M}$. It results from (21) that this can be written $E = E^r \Sigma_r$, where $E^r = \langle E, \Sigma_r \rangle$. For $K \in T_p\mathcal{M}$, let

$$\nabla_K E = (KE^r)\Sigma_r(p). \tag{22}$$

This defines a connection ∇ compatible with $\langle \cdot,\cdot \rangle$. Precisely, for C^1 vector fields E,G on $\mathcal{M}$,

$$\begin{aligned} K\langle E,G \rangle &= (KE^r)G^r + E^r(KG^r) \\ &= \langle \nabla_K E, G\rangle + \langle E, \nabla_K G\rangle, \end{aligned} \tag{23}$$

where (21) and (22) have been used.

With regard to the proof of Propostion 1, the only necessary property of ∇ is

$$\nabla_{\Sigma_r}\Sigma_r = 0. \tag{24}$$

To obtain (24), replace (21) in (22). Since Σ_r is a derivation,

$$\begin{aligned} &\nabla_{\Sigma_r}\Sigma_r = \\ &\Sigma_r\langle \Sigma_r,\Sigma_v\rangle \Sigma_v = \\ &\Sigma_r\langle \Sigma_r,\Sigma_w\rangle\langle \Sigma_v,\Sigma_w\rangle \Sigma_v + \langle \Sigma_r,\Sigma_w\rangle \Sigma_r\langle \Sigma_v,\Sigma_w\rangle\Sigma_v. \end{aligned}$$

A simplification of the third line gives

$$\begin{aligned} &\nabla_{\Sigma_r}\Sigma_r = \\ &\Sigma_r\langle \Sigma_r,\Sigma_w\rangle \Sigma_w + \Sigma_w\langle \Sigma_w,\Sigma_v\rangle\Sigma_v = \\ &\nabla_{\Sigma_r}\Sigma_r + \nabla_{\Sigma_w}\Sigma_w, \end{aligned}$$

which immediately leads to (24).

PROPOSITION 1. *Let (E^i) be a parallel orthonormal frame and y given by (13), where the connection ∇ is defined by (22). Also, let $\bar{\mathcal{Y}}$ be the augmented natural filtration of y. Then, y has its values in $\mathbb{R}^d$. Moreover, for $t \geq 0$, $\bar{\mathcal{Y}}_t = \mathcal{Y}_t$ and*

$$dy_t = h_t dt + d\beta_t, \quad h_t^i = \langle E_t^i, H\rangle, \tag{25}$$

where β is a Brownian motion in $\mathbb{R}^d$ which is independent from x.

PROOF By definition, y has its values in $\mathbb{R}^d$. The proof of the second claim, i.e., $\bar{\mathcal{Y}}_t = \mathcal{Y}_t$, is identical to that of the analogous claim made in [15], (Theorem IV.3 on page 135). More generally, this claim holds for any connection ∇ compatible with $\langle \cdot,\cdot \rangle$.

In order to obtain (25) define first

$$\beta_t^i = \int_0^t \langle E_s^i, \Sigma_r\rangle dB_s^r. \tag{26}$$

Clearly, β is an $\mathcal{F}$-local martingale. Moreover, (21) implies

$$\int_0^t \langle E_s^i,\Sigma_r\rangle\langle E_s^j,\Sigma_r\rangle ds = \delta_{ij}\int_0^t ds.$$

By Lévy's characterisation, β is an $\mathcal{F}$-Brownian motion. Since $\mathcal{F}_0 = \mathcal{X}_\infty$, it follows β is independent from x.

Recall that y can be computed from (15). Replacing the definition (8) of this integral, it follows

$$dy_t^i = \{h_t^i + (1/2)\langle E_t^i, \nabla_{\Sigma_r}\Sigma_r\rangle\}dt + d\beta_t^i.$$

However, (24) states the second term is identically zero. This completes the proof of (25).

3.3. Connector maps

This paragraph is devoted to Proposition 3, which is the main result in the current section. This theorem states that the mapping I to be used in (6) is a connector map which verifies certain conditions, expressed in terms of the Le Jan-Watanabe connection introduced in (22) of the previous paragraph. Roughly, such a connector map becomes the discrete counterpart of antidevelopment. This is the content of Proposition 2 below.

As in 3.1, let $\langle \cdot,\cdot \rangle$ be a Riemannian metric and ∇ a compatible connection. Consider first the definition of a geodesic connector. Recall that, for each $p \in \mathcal{M}$ and $K \in T_p\mathcal{M}$ there exists a unique geodesic curve $\gamma :]-\epsilon,\epsilon[\to \mathcal{M}$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = K$, where $\epsilon > 0$. A geodesic curve is one satisfying the geodesic equation

$$\nabla_{\dot{\gamma}}\dot{\gamma}(t) = 0. \tag{27}$$

If the manifold $\mathcal{M}$ is complete for the connection ∇, then any geodesic curve γ can be extended to all $t \in \mathbb{R}$, (that is, ϵ can be taken arbitrarily large). Assuming this is the case, consider the exponential mapping $\exp : T\mathcal{M} \to \mathcal{M}$. For p and K as above, it is usual to write $\exp(K) = \exp_p(K)$ in order to distinguish the base point p. By definition, $\exp_p(K) = \gamma(1)$.

The exponential mapping is locally invertible. If two points $p,q \in \mathcal{M}$ are close enough to each other, then there exists a unique geodesic γ as above such that $\gamma(1) = q$. It is suitable to write $K = \log_p(q)$, since then "log is the inverse of exp." Note that both $\exp_p$ and $\log_p$ are C^2 mappings [23].

Let $d(\cdot,\cdot)$ denote the Riemannian distance associated to $\langle \cdot,\cdot \rangle$. It will be assumed that $\mathcal{M}$ has a strictly positive radius of injectivity for the connection ∇. That is, there exists $r > 0$ such that for all $p,q \in \mathcal{M}$ the inequality $d(p,q) < r$ implies $\log_p(q)$ is well defined. The two assumptions of completeness of $\mathcal{M}$ with respect to ∇ and of a strictly positive radius of injectivity are note very stringent. In particular, they are verified whenever $\mathcal{M}$ is compact.

A geodesic connector is a mapping I defined following [15]. Let $u < r$ and $\phi : \mathbb{R}_+ \to [0,1]$ a decreasing C^∞ function such that $\phi(d) = 1$ if $d \leq u$ and $\phi(d) = 0$ if $d \geq r$. Let I be given by

$$I(p,q) = \phi(d(p,q))\log_p(q). \tag{28}$$

Now, I is essentially intended to be the mapping $\log_p$ where p is the first argument. The function ϕ is only introduced as a cutoff, to avoid points q which lie beyond the injectivity radius.

In the following, let $\|\cdot\|$ denote Riemannian length, (e.g., $\|G\|^2 = \langle G,G\rangle$), and $\mathbb{E}$ expectation with respect to $\mathbb{P}$.

PROPOSITION 2. *Assume $\mathcal{M}$ is compact. Let G be a vector field along Y such that $t \mapsto \|G_t\|$ is bounded. Then, for $t \geq 0$, the Itô integral $R = \int_0^t \langle G_s, dY\rangle$ is square integrable (that is, $\mathbb{E}|R|^2 < \infty$). Moreover, if for $\delta > 0$*

$$R_\delta = \sum_{k\delta < t} \langle G_{k\delta}, I(Y_{k\delta}, Y_{(k+1)\delta})\rangle, \tag{29}$$

then $\mathbb{E}|R_\delta - R|^2 \to 0$ as $\delta \downarrow 0$.

Proposition 2, (exactly the same as Theorem 3.4.2 on page 55 in [30]), generalises the definition of a classical Itô integral as a limit in the square mean of Riemann sums. In the classical definition, the increments of the integrating process are given by (3). Proposition 2 shows that for an Itô integral as in (8), at least when the manifold $\mathcal{M}$ is compact, the limit continues to hold if the classical increments are replaced by those obtained from a geodesic connector. That is, from (6) where I is given by (28). Note that (3) is truely a special case of (28), since geodesics in $\mathcal{M} = \mathbb{R}^d$ are just straight lines.

The use of geodesic connectors seems natural from a theoretical point of view. However, the mapping I of (28) may be quite difficult to compute. It is important to note that Proposition 2 continuous to hold, all other hypotheses being the same, if I is any mapping with the same property $I(p_1,p_2) \in T_{p_1}\mathcal{M}$, as long as (this is proved in Appendix A)

(I1) I is jointly C^2 (as a mapping $\mathcal{M} \times \mathcal{M} \to T\mathcal{M}$)

(I2) $dI(p,p)(V) = V$ for all $V \in T_p\mathcal{M}$

(I3) $\nabla^2 I(p,p)(V,V) = 0$ for all $V \in T_p\mathcal{M}$

Here, dI and $\nabla^2 I$ are the derivative and the Hessian of I with respect to its second argument. When the first argument p is fixed, I has its values in a fixed vector space $T_p\mathcal{M}$. Therefore, it is possible to speak of its derivative and Hessian. Conditions (**I1-I3**) are verified by the mapping of (28). Intuitively, any mapping I which verifies them is "geodesic connector up to second order."

It is now possible to state Proposition 3.

PROPOSITION 3. *Assume $\mathcal{M}$ is compact and the connection ∇ is given by (22). Moreover, let y be given by (13). For $\delta > 0$, let ΔY_k be given by (6) and $\Delta y_k = y_{(k+1)\delta} - y_{k\delta}$ where $k \in \mathbb{N}$. Then,*

$$\mathbb{E}|\Delta y_k^i - \Delta Y_k^i|^2 = o(\delta), \tag{30}$$

where $\Delta Y_k^i = \langle E_{k\delta}^i, \Delta Y_k\rangle$. In particular, as $\delta \downarrow 0$, the distribution of the vector ΔY_k^i conditionally on $\mathcal{F}_{k\delta}$ is approximately normal. Precisely, as $\delta \downarrow 0$,

$$\mathcal{L}\left\{\delta^{-1/2}(\Delta Y_k^i - \delta \times h_{k\delta}^i)\right\} \longrightarrow N_d. \tag{31}$$

Here, $\mathcal{L}$ denotes the law of the random vector whose components are in brackets and N_d denotes a standard normal distribution in $\mathbb{R}^d$.

PROOF The order of magnitude given in equation (30) is essentially a result of Proposition 2.

For (31), note from Proposition 1

$$\Delta y_k^i = \int_{k\delta}^{(k+1)\delta} h_t^i dt + \{\beta_{(k+1)\delta}^i - \beta_{k\delta}^i\},$$

but then, as is clear by bounded convergence,

$$\mathcal{L}\{\delta^{-1/2}(\Delta y_k^i - \delta \times h_{k\delta}^i)\} \longrightarrow N_d.$$

The proposition follows from (30) by noting that $h_{k\delta}$ is measurable wit respect to $\mathcal{F}_{k\delta}$ and β is an $\mathcal{F}$-Brownian motion, so $\beta_{(k+1)\delta} - \beta_{k\delta}$ is independent of $\mathcal{F}_{k\delta}$ with the required covariance.

4. PARTICLE FILTERING ALGORITHM

This section is concerned with the numerical solution of the filtering problem defined by (1) and (2). In 4.1, Proposition 4 expresses the solution in closed form using a Kallianpur-Striebel formula. In 4.2, a particle filtering algorithm is proposed in order to numerically implement this formula. The convergence of this algorithm to the solution of the original filtering problem, (i.e., that of (1) and (2)), is the subject of Proposition 5 in 4.3.

4.1. Kallianpur-Striebel formula

A Kallianpur-Striebel formula may be though of as an abstract Bayes formula which expresses in closed form the solution of a continuous time filtering problem. For classical filtering problems, where the observation process depends on the unknown signal through an additive white noise model, the Kallianpur-Striebel formula is quite well known [22]. The idea in this section is to profit from Proposition 1 in order to obtain a similar formula for the problem defined by (1) and (2).

The following Proposition 4 uses the notion of a *copy* of the unknown signal independent from the observation process. This means a process $\tilde{x}$ with the same law as the signal x, given by (1), but which is independent from the observation process Y.

Let π_t be the conditional distribution of x_t, given past observations $\mathcal{Y}_t$. The aim of the filtering problem is precisely to find $\pi_t(\varphi)$ for any function $\varphi \in C_b(S)$. By definition, π is a *càdlàg* process with values in the space of probability measures on $(S, \mathcal{S})$ and such that

$$\pi_t(\varphi) = \mathbb{E}[\varphi(x_t) \mid \mathcal{Y}_t]. \tag{32}$$

From this, it is clear that the process π is $\mathcal{Y}$-adapted. Thus it should be possible to write it down in terms of some functional of the process Y (see the discussion at the beginning of 3.1).

PROPOSITION 4. *Let $\tilde{x}$ be a copy of x independent from Y. Let $\tilde{H}$ be the process where $\tilde{H}_t = H(\tilde{x}_t, Y_t)$. For $\varphi \in C_b(S)$,*

$$\pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(1)}, \quad \rho_t(\varphi) = \mathbb{E}[\varphi(\tilde{x}_t) L_t \mid \mathcal{Y}_\infty], \tag{33}$$

where $\mathcal{Y}_\infty = \vee_{t \geq 0} \mathcal{Y}_t$ and

$$L_t = \exp\left(\int_0^t \langle \tilde{H}_s, dY \rangle - (1/2) \int_0^t \|\tilde{H}_s\|^2 ds \right). \tag{34}$$

Here the Itô differential $\langle \tilde{H}_s, dY \rangle$ is computed using the Le Jan-Watanabe connection defined by (22).

PROOF Recall that Proposition 1 states $\mathcal{Y} = \bar{\mathcal{Y}}$, where $\bar{\mathcal{Y}}$ is the augmented natural filtration of y. If y is computed using the Le Jan-Watanabe connection defined by (22), then y depends on x using a classical additive white noise model (25). The corresponding Kallianpur-Striebel formula is the same as above, but with L_t given by (this is the classical formula [22])

$$L_t = \exp\left(\int_0^t \tilde{h}_i(s) dy_s^i - \int_0^t \tilde{h}_i(s)\tilde{h}_i(s) ds \right).$$

Here, $\tilde{h}_i(s) = \langle E_s^i, \tilde{H} \rangle$, as in (25). Using the fact that $\mathcal{Y} = \bar{\mathcal{Y}}$, the Proposition follows by (17) and (32).

4.2. Algorithm description

Numerical implementation of the Kallianpur-Striebel formula, i.e., of (33) in Proposition 4, can proceed in at least two ways [22]. First, it is possible to derive from this formula the corresponding Zakai equation, whose solution can be attempted using a finite difference or spectral method. The Zakai equation is a stochastic differential equation satisfied by the unnormalised distribution ρ. It was indeed found in [14], but it seems its solution has not been specifically considered in the literature. Second, thinking of the Kallianpur-Striebel formula as an abstract Bayes formula, it is possible to implement it using sequential Monte Carlo strategy. Here, this latter option is pursued. The result takes the form of a particle filtering algorithm.

The proposed particle filtering algorithm and its convergence are easily understood in the framework of a discretised version of the original filtering problem of (1) and (2). Note that, in (33), $\tilde{x}$ and Y are taken to be independent. It is then natural to think of $\tilde{x}$ as a computer simulated version of x. However, this is just an idealisation. A computer experiment can only yield an approximation of the model (1). Such an approximation is obtained for a given discretisation step size $\delta > 0$. It is a process denoted x^δ with its value at time $k\delta$ written $x^\delta_{k\delta} = x^\delta_k$. The particle filtering algorithm will be convergent when x^δ verifies the following conditions

- (**x1**) $\{x^\delta_k; k \in \mathbb{N}\}$ is a time invariant Markov chain
- (**x2**) x^δ_0 has the same distribution as x_0 (noted μ)
- (**x3**) $\{x^\delta_k\}$ has strong order of convergence 0.5
- (**x4**) the transition kernel $q(s, ds')$ of $\{x^\delta_k\}$ is Feller

Conditions (**x1-x2**) are necessary for the statement of the algorithm. Conditions (**x3-x4**) do not appear in the algorithm, but are required in showing its convergence. The meaning of (**x3**) is

$$\mathbb{E}[d_S^2(x_t, x^\delta_k)] = O(\delta), \tag{35}$$

for k such that $t \in [k\delta, (k+1)\delta[$, where $d_S(\cdot,\cdot)$ is the metric in the state space S. The Feller condition (**x4**) is that for all $\varphi \in C_b(S)$, if $q\varphi$ is the function

$$q\varphi(s) = \int_S q(s, ds')\varphi(s'),$$

then $q\varphi \in C_b(S)$. Recall that $q(s, ds')$ is the transition probability for going from s to the "neighborhood" ds'.

If x is a diffusion in $\mathbb{R}^l$ with bounded Lipschitz coefficients and x_0 square integrable, then the first order Euler scheme gives an approximation x^δ verifying (**x1-x4**). In fact, the assumptions on the coefficients of x can be weakened to those of [27], (Exercise 9.6.3 on page 326). Of course, in this case the metric $d_S(\cdot,\cdot)$ is just the Euclidean metric of $\mathbb{R}^l$.

If x is a diffusion in a compact Riemannian manifold $\mathcal{N}$, then conditions (**x1-x4**) are verified when x is approximated using successive geodesic steps.

In a purely heuristic way, consider now a discrete time filtering problem where the unknown signal is the Markov chain $\{x^\delta_k\}$ and the observations are the sequence of increments $\{\Delta Y_k\}$. Here, ΔY_k is obtained from (6) where the connection ∇ is that of (22).

Based on Proposition 3 and on (35), consider the likelihood for the observations ΔY_k to be

$$l(x^\delta_k, \Delta Y_k) = \exp\left(\langle H^\delta_k, \Delta Y_k \rangle - \frac{\delta}{2} \|H^\delta_k\|^2 \right), \tag{36}$$

where $H^\delta_k = H(x^\delta_k, Y_{k\delta})$. Applying the usual Bayes formula for a discrete time filtering problem obtained in this way, the resulting conditional distribution given the first $M + 1$ observations is

$$\pi^\delta_M(\varphi) = \frac{\rho^\delta_M(\varphi)}{\rho^\delta_M(1)}, \quad \rho^\delta_M(\varphi) = \mathbb{E}[\varphi(\tilde{x}^\delta_M) L^\delta_M \mid \mathcal{Y}_\infty], \tag{37}$$

for $\varphi \in C_b(S)$, where $\{\tilde{x}_k^\delta\}$ is a copy of $\{x_k^\delta\}$ independent from Y and

$$L_M^\delta = \prod_{k=0}^{M} l(\tilde{x}_k^\delta, \Delta Y_k). \tag{38}$$

This is really just Bayes formula in discrete time. From (37),

$$\rho_M^\delta(\varphi) = \rho_{M-1}^\delta(\varphi_{M-}), \tag{39}$$

where

$$\varphi_{M-}(s) = \int_S l(s', \Delta Y_M)\varphi(s')q(s, ds'). \tag{40}$$

This is the usual "prediction-measurement update" formula, where prediction is according to the transition kernel $q(s,ds')$ and measurement update is based on the likelihood $l(s', \Delta Y_M)$.

It is important to note that π_M^δ is a true probability measure on S conditionally on the increments $\Delta Y_0, \ldots, \Delta Y_M$. In particular, it makes sense to speak of sampling from π_M^δ once these increments are given. This is the aim of the proposed particle filtering algorithm.

The algorithm has as its input the sequence of increments $\{\Delta Y_k\}$ and is parameterised by the number of particles N. The output after $M+1$ observations is a family of N particles $\hat{x}_M^i, \ldots, \hat{x}_M^N \in S$ which give a Monte Carlo approximation $\hat{\pi}_M^\delta$ of π_M^δ. This is

$$\hat{\pi}_M^\delta(\varphi) = (1/N)\sum_{i=1}^{N} \varphi(\hat{x}_M^i). \tag{41}$$

The implementation is the following,

- when ΔY_0 is available
  1 generate *i.i.d.* particles $\tilde{x}_0^1, \ldots, \tilde{x}_0^N \sim \mu$
  2 compute normalised weights, $w_0^i \propto l(\tilde{x}_0^i, \Delta Y_0)$
  3 generate $(n_0^1, \ldots, n_0^N) \sim \text{multinomial}(w_0^1, \ldots, w_0^N)$ and replace $\tilde{x}_0^i$ by n_0^i particles with same value
  4 relabel new particles $\hat{x}_0^1, \ldots, \hat{x}_0^N$; set $w_0^i = 1/N$
- when ΔY_k is available $(k \geq 1)$
  1 generate particles $\tilde{x}_k^i \sim q(\hat{x}_{k-1}^i, ds)$
  2 compute normalised weights, $w_k^i \propto w_{k-1}^i l(\tilde{x}_k^i, \Delta Y_k)$
  3 generate $(n_k^1, \ldots, n_k^N) \sim \text{multinomial}(w_k^1, \ldots, w_k^N)$ and replace $\tilde{x}_k^i$ by n_k^i particles with same value
  4 relabel new particles $\hat{x}_k^1, \ldots, \hat{x}_k^N$; set $w_0^i = 1/N$

These steps are very much the same as in the classical bootstrap filter. The geometry of the observation process Y only appears through the use of a connector map I, which provides the increments ΔY_k.

The bootstrap filter is the simplest, but the least robust, particle filtering algorithm and there are many improvements upon it known in the literature. These can all be implemented in an equally direct way once the ΔY_k have been obtained.

To summarise, a connector map I leads to increments ΔY_k with approximate likelihood given by (36). Once this situation is accepted, it can be replaced into any suitable algorithm. It is possible to say that the connector map serves to linearise the observation process Y locally, i.e., in the neighborhood of each sample $Y_{k\delta}$ as in (6).

4.3. Convergence

The convergence of the particle filtering algorithm proposed in the previous paragraph is here given in Proposition 5. Precisely, what is meant by this is the convergence of $\hat{\pi}_M^\delta$ to π_t as the step size δ goes to zero and the number of particles N goes to infinity, when M is taken of the order of t/δ. That is, the order of the number of increments ΔY_k which can be constructed from (6) up to time t.

An equally important question, not dealt with here, is the convergence as t goes to infinity of the conditional distribution π_t or of its Monte Carlo approximation obtained from the particle filtering algorithm. This is related to the eventual ergodicity or mixing properties of the unknown signal x.

Proposition 5 is based on two lemmas, which are first given. Lemma 1 states the convergence of π_M^δ to π_t. Lemma 2 states the convergence of $\hat{\pi}_M^\delta$ to π_M^δ for any given value of δ. This latter limit is not shown to be uniform in δ. Thus, in their form stated below, Lemmas 1 and 2 cannot be used to justify an approach where δ is taken proportional to $1/N$ (or some other function of N which converges to zero as N goes to infinity) and $\hat{\pi}_M^\delta$ is then computed for a large value of N.

It is possible to say that Proposition 5 only provides, in a satisfactory way, the consistency of the approximation $\hat{\pi}_M^\delta$. That is, the fact that it is possible to choose δ and N to make this approximation arbitrarily close to π_t.

For Lemma 1, the two following conditions are required.

(H1) $\|H\|$ is bounded (as a function $S \times \mathcal{M} \to \mathbb{R}_+$)

(H2) $\|H(s,p) - H(s',p)\| \leq C d_S(s,s')$ for all $p \in \mathcal{M}$, where the constant C does not depend on p

These are quite strong restrictions, however they allow for straightforward proofs. Replacing them by weaker conditions may lead to convergence in the square mean being replaced by convergence in probability, (see the statement of the lemma), but it would not introduce any more fundamental changes.

Assumption (**H2**) means that H is a Lipschitz continuous application from S to the space of continuous vector fields on $\mathcal{M}$, this latter space being equipped with its topology of uniform convergence with respect to the Riemannian metric.

LEMMA 1. *Assume $\mathcal{M}$ is compact and conditions (*__x1-x3__*), (*__H1-H2__*) hold. Let $\varphi \in C_b(S)$ be Lipschitz continuous, such that $\varphi \in D(A)$ and $A\varphi \in C_b(S)$, (recall the notation of (1)). Then, if M is the integer part of t/δ,*

$$\mathbb{E}|\pi_M^\delta(\varphi) - \pi_t(\varphi)|^2 = O(\delta).$$

PROOF The proof is identical to the one in [15], (Proposition 2.1 on page 292). Here, the main steps are indicated.

Note first that it is possible to consider $M\delta = t$. Indeed, the assumption that $\varphi \in D(A)$ and $A\varphi \in C_b(S)$ guarantees

$$\mathbb{E}|\pi_t(\varphi) - \pi_{M\delta}(\varphi)|^2 = O(\delta).$$

This is since π_t verifies the filtering equation given in [14], which has bounded coefficients under condition (**H1**).

The proof follows from the identity

$$\pi_M^\delta(\varphi) - \pi_t(\varphi) = \mathbb{E}[\varphi(x_M^\delta) - \varphi(x_t) \mid \mathcal{Y}_\infty] + \mathbb{E}[(A-1)(\varphi(x_M^\delta) - \pi_M^\delta(\varphi)) \mid \mathcal{Y}_\infty], \tag{42}$$

where $A = L_M^\delta / L_t$. This can be proved by using the Kallianpur-Striebel formula (33) to express the conditional expectations and then by developing the products.

Since φ is Lipschitz continuous, the square mean of the first term is bounded by

$$\mathbb{E}|\varphi(x_M^\delta) - \varphi(x_t)|^2 \le C\mathbb{E}[d_S^2(x_t, x_k^\delta)] = O(\delta), \tag{43}$$

where the second inequality uses (**x3**).

The bound for the second term is more delicate. The idea is to note

$$L_M^\delta = \exp\left(\sum_{k\delta<t} \langle H_k^\delta, \Delta Y_k\rangle - (\delta/2)\sum_{k\delta<t} \|H_k^\delta\|^2\right) \tag{44}$$

and compare this to L_t, given by (34), using Proposition 2. The detailed development requires condition (**H2**) and gives

$$\mathbb{E}|(A-1)(\varphi(x_M^\delta) - \pi_M^\delta(\varphi))|^2 \le 4\|\varphi\|^2 \mathbb{E}(A-1)^2 = O(\delta), \tag{45}$$

where $\|\varphi\|$ is the supremum of $|\varphi(s)|$ over $s \in S$.

The proof is completed by applying Minkowski's inequality, (43) and (45).

LEMMA 2. *Assume conditions (***x1-x2***), (***x4***) and (***H1***) hold. For all $\varphi \in C_b(S)$ and any values of δ and M*

$$\mathbb{E}|\hat{\pi}_M^\delta(\varphi) - \pi_M^\delta(\varphi)|^2 \to 0 \quad \text{as} \quad N \uparrow \infty.$$

PROOF It is clear from (41) that $\hat{\pi}_M^\delta(\varphi)$ is bounded. Note, moreover, that $\pi_M^\delta(\varphi)$ is square integrable. By dominated convergence, in order to show the Lemma, it is enough to show that $\hat{\pi}_M^\delta(\varphi)$ converges to $\pi_M^\delta(\varphi)$ almost surely.

Almost sure convergence is a direct application of a general theorem from [21], (Theorem 1, on page 742). This requires that the transition kernel be Feller, which is condition (**x4**), and that the likelihood function is continuous, bounded and strictly positive.

Here, the likelihood function is the one corresponding to (36). That is

$$l(s, \Delta Y_{k\delta}) = \exp\left(\langle H(s, Y_{k\delta}), \Delta Y_k\rangle - \frac{\delta}{2}\|H(s, Y_{k\delta})\|^2\right).$$

That this is continuous, as a function of s, and strictly positive is immediate. Boundedness follows since the second term under the exponential is negative and

$$\langle H(s, Y_{k\delta}), \Delta Y_k\rangle \le r\|H(s, Y_{k\delta})\|.$$

By Cauchy-Schwarz inequality, where r is as in the definition (28) of the mapping I.

Finally, it is possible to conclude by condition (**H1**).

Now, it is possible to state Proposition 5 which combines Lemmas 1 and 2.

PROPOSITION 5. *Assume that $\mathcal{M}$ is compact and that conditions (***x1-x4***), (***H1-H2***) hold. Let φ be as in Lemma 1. Then, if M is the integer part of t/δ*

$$\lim_{\delta\downarrow 0} \lim_{N\uparrow\infty} \mathbb{E}|\hat{\pi}_M^\delta(\varphi) - \pi_t(\varphi)|^2 = 0.$$

PROOF The conditions of Lemmas 1 and 2 are united. By Lemma 1

$$\lim_{\delta\downarrow 0} \mathbb{E}|\pi_M^\delta(\varphi) - \pi_t(\varphi)|^2 = 0,$$

where the expression under the limit does not depend on N. By Lemma 2

$$\lim_{N\uparrow\infty} \mathbb{E}|\hat{\pi}_M^\delta(\varphi) - \pi_M^\delta(\varphi)|^2 = 0,$$

for any values of δ and M. Thus, the proposition follows by adding together these two limits and applying Minkowski's inequality.

Proposition 5 does not explicitly provide the rate at which $\hat{\pi}_M^\delta$ converges to π_t. Obtaining this rate of convergence requires a deeper analysis than provided here (in Lemmas 1 and 2). This can be carried out on the basis of the corresponding analysis for a classical filtering problem [22], (Chapter 9), but still has not been pursued in the literature. Clearly, this situation represents an important drawback for practical application.

The information lacking from Proposition 5, i.e., the rate of convergence, can eventually be recovered on a case by case basis. For the convergence of π_M^δ to π_t, Lemma 1 can be used to obtain the precise rate of convergence, (by expressing the various constants appearing in the proof). Then, for any required values of δ and M, the problem is to find a number of particles N sufficiently large for a given precision. It is well known that $\hat{\pi}_M^\delta$ converges to π_M^δ at a rate of the order of $1/N$ but where the involved constants depend on the observations $\{\Delta Y_k\}$. For an individual realisation of the observations, this can be made precise either through additional calculation or through computer experiments, based on the specific model being studied, (that is, on a particular instance of equations (1) and (2)).

Regarding the convergence of π_M^δ to π_t, stated in Lemma 1, it is useful to make the following remark. The order of convergence in condition (**x3**) can often be improved. For example [27], if x is a diffusion in $\mathbb{R}^l$, using the Milstein approximation instead of a first order Euler scheme gives strong order 1 instead of 0.5. However, this is not enough to improve the overall order of magnitude given in the lemma. As can be seen from the proof, this order of magnitude involves both (43) and (45). While improving the order of convergence in condition (**x3**) will accordingly improve the order of magnitude in (43), it has no similar effect on (45). The computer experiments presented in the following section show that the proposed particle filtering algorithm, when implemented with adequate values of δ and N, performs in a sensibly satisfactory way for the chosen examples.

5. EXAMPLES: OBSERVATIONS IN $SO(3)$ AND S^2

The stochastic filtering problem stated in Section 2 is of a quite general form. By specifying the state space S, the manifold $\mathcal{M}$ and the various objects appearing in (1) and (2), it is possible to recover a wide range of problems. As already mentioned, these include the classical ones defined by an additive white noise model.

Several natural and important observation models also arise as special cases of the problem of Section 2. Of particular interest in engineering is the case where the manifold $\mathcal{M}$ is a classical matrix manifold, (that is, a matrix Lie group or a related symmetric space), and the observation process Y, conditionally on the unknown signal x, is an invariant diffusion.

In signal and image processing, classical matrix manifolds such as Stiefel and Grassmann manifolds are of widespread use and importance. Therefore, it is natural to consider observation models compatible with their underlying structure. Roughly speaking, these are exactly the ones involving invariant diffusions.

To explain and optimise the implementation and performance of the particle filtering algorithm of Section 4 in the special case of classical matrix manifolds should be one of the main objectives for the present work. Mainly for a reason of space, the current section has a more modest scope dealing only with two individual examples.

In 5.1, the problem is considered where the observation process Y is an invariant diffusion in the special orthogonal group $SO(3)$. In 5.2 a similar problem is studied but where Y has its values in the unit sphere S^2. These two examples serve as case studies. They show how, when faced with a problem of the kind given by (1) and (2), to carry out the various steps leading to a successful implementation of the particle filtering algorithm of Section 4. These include, at least, specification of the metric (21) and the connection (22), choice of the connector map I to be replaced in (6) and choice of the approximation $\{x_k^\delta\}$ of the signal x.

It should be noted none of these steps is known a priori, just by knowing the manifold $\mathcal{M}$. They are all carried out based on (1) and (2). In particular, the geometric structure given by the metric and connection of (21) and (22) is adapted to the observation model.

5.1. Observations in $SO(3)$

The first example considers the case where the observation process Y, conditionally on the unknown signal x, is a left invariant diffusion in the special orthogonal group $SO(3)$. This is the set of 3×3 real matrices g which are orthogonal and have unit determinant. That is,

$$g^{-1} = g^\dagger, \quad \det(g) = 1, \tag{46}$$

where † denotes matrix transpose. As a subset of the vector space $\mathbb{R}^{3\times 3}$ (space of 3×3 real matrices), $SO(3)$ is connected and compact. Furthermore, it is closed under matrix multiplication and inversion. Thus, $SO(3)$ is a compact connected Lie group [31]. The 3×3 identity matrix is denoted e; clearly $e \in SO(3)$.

Let $\mathfrak{so}(3)$ denote the subspace of $\mathbb{R}^{3\times 3}$, consisting of all antisymmetric matrices. That is, matrices $\sigma \in \mathbb{R}^{3\times 3}$ such that $\sigma + \sigma^\dagger = 0$. It can be shown that, for $\sigma \in \mathbb{R}^{3\times 3}$ and $\gamma(t) = \exp(t\sigma)$ where $t \in \mathbb{R}$, $\gamma(t) \in SO(3)$ for all $t \in \mathbb{R}$ if and only if $\sigma \in \mathfrak{so}(3)$. By definition, this means that $\mathfrak{so}(3)$ is the Lie algebra of the Lie group $SO(3)$. For this and other facts on compact Lie groups used in the following, see [31].

It is not surprising that, being defined by the differentiable constraints (46), $SO(3)$ is a differentiable manifold. Moreover, $\mathfrak{so}(3)$ can be identified with the tangent space $T_e SO(3)$.

The special orthogonal group $SO(3)$ is quite important in many applications. A matrix $g \in SO(3)$ defines an orientation preserving rotation in $\mathbb{R}^3$. Furthermore, $SO(3)$ is often thought of as the typical example of a nontrivial compact connected Lie group. The presentation in the rest of this paragraph generalises to any compact connected Lie group with very minor changes.

In terms of the general filtering problem of Section 2, the example considered here makes no restriction on the signal model (that is, on (1)). The observation model (2) is specified in the following way.

The sensor function H and the vector fields Σ_r are defined in terms of left invariant vector fields on $SO(3)$. For each $\sigma \in \mathfrak{so}(3)$, there is a corresponding vector field Σ on $SO(3)$ where,

$$\Sigma(g) = g\sigma, \quad g \in SO(3). \tag{47}$$

This means that for each $g \in SO(3)$, there exists some differentiable curve $\gamma_g :]-\epsilon, \epsilon[\to SO(3)$ such that $\gamma_g(0) = g$ and $\dot{\gamma}_g(0) = \Sigma(g)$. In fact, it is quite straightforward to obtain such a curve. First, let γ_e be the curve $\gamma_e(t) = \exp(t\sigma)$. This is defined for all real t and has its values in $SO(3)$ as mentioned above. By elementary properties of the matrix exponential, $\gamma_e(0) = e$ and $\dot{\gamma}_e(0) = \sigma$. To obtain the curve γ_g for any $g \in SO(3)$, it is enough to put $\gamma_g(t) = g\gamma_e(t)$. The meaning of the name "left invariant vector field" is precisely that γ_g can be obtained from γ_e by left multiplication.

Note that σ and Σ define each other uniquely. Any left invariant vector field Σ is of the form (47) where $\sigma = \Sigma(e)$.

It is required that for all $s \in S$, the application $g \mapsto H(s,g)$ is a left invariant vector field. It follows that there exists a mapping $h : S \to \mathfrak{so}(3)$ such that for $g \in SO(3)$

$$H(s,g) = gh(s). \tag{48}$$

The vector fields Σ_r are also taken to be left invariant vector fields. Let $\sigma_1, \sigma_2, \sigma_3$ be any basis of $\mathfrak{so}(3)$. This basis being fixed, let $\Sigma_1, \Sigma_2, \Sigma_3$ be the corresponding left invariant vector fields as in (47). In other words,

$$\Sigma_r(g) = g\sigma_r. \tag{49}$$

Note that the number of vector fields Σ_r is here equal to 3, the dimension of $SO(3)$ (which is also the dimension of $\mathfrak{so}(3)$). This was not assumed in the general problem of Section 2.

Now that H and $\Sigma_1, \Sigma_2, \Sigma_3$ have been defined, it is possible to write down the observation model (2). From (48) and (49), a simple rearrangement shows

$$dY_t = Y_t \circ \{h(x_t)dt + d\hat{B}_t\}. \tag{50}$$

Here $\hat{B}$ is a process with values in $\mathfrak{so}(3)$ defined from a Brownian motion B in $\mathbb{R}^d$, which is independent from x, by

$$\hat{B}_t = B_t^1\sigma_1 + B_t^2\sigma_2 + B_t^3\sigma_3. \tag{51}$$

The reader should be immediately aware equation (50) is just a linear matrix stochastic differential equation. It can be understood for each matrix element after writing down the usual formula for matrix product. A process Y satisfying an equation of this form, when x is assumed known, is called a left invariant diffusion in $SO(3)$. An alternative way of writing equation (50) involves the vector fields $\Sigma_1, \Sigma_2, \Sigma_3$. This is

$$dY_t = \Sigma_r(Y) \circ \{h^r(x_t)dt + dB_t^r\}, \tag{52}$$

where $h = h^r\sigma_r$ for some functions $h^r : S \to \mathbb{R}$. note the differentiability conditions of Seciton 2 are verified. Indeed, as functions of g, $H(s,g)$ and $\Sigma_r(g)$ are linear and therefore smooth (C^∞).

In [13], Lo considered the observation model (50) by itself (but for a general matrix Lie group, not just $SO(3)$). It was proposed that this can be reduced to a classical, additive white noise model by the following simple transformation

$$y_t = \int_0^t Y_s^{-1} \circ dY_s. \tag{53}$$

By the chain rule of Stratonovich calculus, it is clear that

$$dy_t = h(x_t)dt + d\hat{B}_t. \tag{54}$$

This result is strikingly similar to (25) in Proposition 1. It is now shown that, in effect, it is a special case of that proposition.

Following the approach of 3.2, the Le Jan-Watanabe connection is now introduced. Note first the condition of ellipticity is here verified. In fact, a sharper result holds since for each $g \in SO(3)$ the vectors $\Sigma_1(g), \Sigma_2(g), \Sigma_3(g)$ form a basis of $T_gSO(3)$.

Definition (21) amounts to introducing a Riemannian metric on $SO(3)$ such that this basis is orthonormal,

$$\langle \Sigma_r(g), \Sigma_v(g) \rangle = \delta_{rv}. \tag{55}$$

Using this metric, the Le Jan-Watanabe connection is defined by (22) which gives

$$\nabla_{\Sigma_r}\Sigma_v = 0. \tag{56}$$

This immediately implies (24). In fact, for any $g \in SO(3)$ and tangent vector $K \in T_gSO(3)$, it follows by linearity that $\nabla_K\Sigma_r = 0$. Here, one says the vector fields Σ_r form a global parallel frame.

For the following, it is important to note the metric (55) is left invariant. If $g \in SO(3)$ and $E, K \in T_gSO(3)$,

$$\langle E, K \rangle = \langle g^{-1}E, g^{-1}K \rangle. \tag{57}$$

That is, the left hand side is computed in the tangent space $T_gSO(3)$ and the right hand side in $T_eSO(3)$, which is $\mathfrak{so}(3)$. This can be shown by putting $\eta = g^{-1}E \in \mathfrak{so}(3)$ and $\kappa = g^{-1}K \in \mathfrak{so}(3)$ and considering the corresponding left invariant vector fields as in (47). It is then a straightforward result of (55).

Due to (57), the Riemannian metric (55) is completely determined by the basis $\sigma_1, \sigma_2, \sigma_3$. This can be chosen in a completely arbitrary way. It is clear that the following matrices form a basis of $\mathfrak{so}(3)$

$$\omega_1 = \begin{pmatrix} & & \\ & & -1 \\ & 1 & \end{pmatrix} \omega_2 = \begin{pmatrix} & & 1 \\ & & \\ -1 & & \end{pmatrix} \omega_3 = \begin{pmatrix} & -1 & \\ 1 & & \\ & & \end{pmatrix}.$$

The general form of the basis $\sigma_1, \sigma_2, \sigma_3$ is therefore

$$\sigma_r = b_{vr}\omega_v, \tag{58}$$

where b is an invertible matrix.

It is possible that the Riemannian metric (55) will not be biinvariant (i.e., both left and right invariant). This is the case if and only if the matrix b is orthogonal. In practice, there is no reason why this should be the case. In rigid body mechanics [32], an orthogonal matrix b may be chosen when studying the motion of a spherically symmetric body. In general, b is given by the inertia matrix of the body, reflecting its shape and mass distribution. This can be far from spherical symmetry (consider an airplane). The metric (55) is introduced based on the observation model (52). In other words, it is adapted to the observation model.

Furthermore, the connection defined by (56) is known as the Cartan-Schouten (−)-connection. Since it has nonzero torsion, it is not the Levi-Civita connection of any Riemannian metric (in particular, of the metric (55)). When the matrix b is orthogonal, the Levi-Civita connection of the metric (55) is known as the Cartan-Schouten (0)-connection (see [23, 33]).

In order to apply Proposition 1, it is necessary to compute a parallel orthonormal frame (E^i). This turns out to be especially simple. Applying (10) and (56),

$$\nabla_{\circ dY}\Sigma_r(Y_t) = 0.$$

Therefore, by (9) and (12),

$$d\langle E^i, \Sigma_r\rangle = 0.$$

Since both (E^i) and (Σ_r) are orthonormal families, this implies the existence of an orthogonal matrix $a = (a^{ir})$ such that

$$E^i_t = a^{ir}\Sigma_r(Y_t).$$

Moreover, by a choice of initial condition, it is possible to take $a = e$ identity matrix.

Now, (13) gives

$$y^i_t = a^{ir}\int_0^t \langle \Sigma_r(Y_s), \circ dY\rangle. \tag{59}$$

To evaluate this Stratonovich integral, it is possible to apply (18). Note that this states the Stratonovich integral is the same as a classical Stratonovich integral. Applying this prescription is easier than changing the current notation to that of (18) and then changing back. In short, it follows form (49) and (57),

$$y^i_t = \langle a^{ir}\sigma_r, \int_0^t Y_s^{-1}\circ dY_s\rangle.$$

This shows that (54) is the same as (59) up to a change of basis. Thus (54) is indeed a special case of (25), Proposition 1. It is interesting to note the Brownian motion β appearing in Proposition 1 turns out, in the present case, to be the same as the original Brownian motion B.

The transformation (54) taking Y into y can be extended to any Lie group, not just compact and matrix. It is known as the Lie group stochastic logarithm and was generalised extensively by Estrade [34]. That is coincides with a stochastic antidevelopment was first pointed out in [35].

Going on with the programme of applying Section 4 in the current example, consider now the construction of connector maps. Recall this was described in (6) and (28). Moreover, it can be implemented using any mapping I verifying conditions (**I1-I3**).

As in 3.3, the starting point is the notion of geodesic. Finding the geodesics of the connection (56) is straightforward. Indeed, the definition of this connection suggests geodesics are precisely the flow lines of left invariant vector fields. These are the curves of the form γ_g discussed after (47). This is easily checked to be the case by replacing (56) in the geodesic equation (27).

Thus, the mapping $\exp : TSO(3) \to SO(3)$ is related in a simple way to the matrix exponential. For $g \in SO(3)$ and $K \in T_g SO(3)$,

$$exp_g(K) = g\exp(g^{-1}K). \tag{60}$$

Indeed, letting $\sigma = g^{-1}K$, it follows from (47) that $\Sigma(g) = K$. Then, it is clear the right hand side is $\gamma_g(1)$.

To specify the mapping I of (28) to the current context, an estimate of the radius of injectivity of the connection (56) would be needed. In (28), this is the role played by r which is needed in constructing the cut off function ϕ. With the considered geometry, the group $SO(3)$ is a manifold of constant (strictly) positive curvature. The radius of injectivity is then known from Riemannian geometry [23].

Here, a less elegant but simpler approach is taken. For each $g_1 \in SO(3)$ and $K \in T_{g_1}SO(3)$, let $g_2 = \exp_{g_1}(K)$. If $|g_1^{-1}g_2 - e| < 1$, where $|\cdot|$ stands for the Euclidean matrix norm, then $g_1^{-1}g_2$ has a unique matrix logarithm. Denote this $\log(g_1^{-1}g_2)$. Then by (60),

$$K = g_1\log(g_1^{-1}g_2).$$

For any $g_1, g_2 \in SO(3)$ verifying $|g_1^{-1}g_2 - e| < 1$, define

$$\log_{g_1}(g_2) = g_1\log(g_1^{-1}g_2). \tag{61}$$

According to (28), when it is possible, $I(g_1, g_2)$ should coincide with $\log_{g_1}(g_2)$. To take into account couples g_1, g_2 for which this expression is not well defined, let $\phi : SO(3)\times SO(3) \to [0,1]$ be a C^∞ function such that $\phi(g_1, g_2) = 0$ if $|g_1^{-1}g_2 - e| \geq 1$ and $\phi(p,q) = 1$ if $|g_1^{-1}g_2 - e| \leq 1 - \lambda$. Here, $0 < \lambda < 1$ is fixed. Now, I can be defined as

$$I(g_1, g_2) = \phi(g_1, g_2)\log_{g_1}(g_2). \tag{62}$$

Computing a matrix logarithm, even for 3×3 matrices, is a relatively involved task. It is possible to propose an alternative mapping I, which does not involve a matrix logarithm. Recall the first order Taylor expansion of the matrix logarithm at e,

$$\log(g_1^{-1}g_2) = g_1^{-1}g_2 - e + O(|g_1^{-1}g_2 - e|^2). \tag{63}$$

Let Π be a linear projection from $\mathbb{R}^{3\times 3}$ to $\mathfrak{so}(3)$. For instance, $\Pi(\sigma) = (1/2)[\sigma + \sigma^\dagger]$ associates to the matrix σ its antisymmetric part. Instead of (62), it is possible to use

$$I(g_1, g_2) = g_1\Pi(g_1^{-1}g_2 - e). \tag{64}$$

Of course, it is an abuse of notation to call both mappings (62) and (64) by the same name I. Still, this is done since they serve the same purpose. Roughly, the difference between (62) and (64) is that in the latter expression the matrix logarithm is replaced by the first term in its Taylor expansion. Moreover, in this same expression (64), there is no need for a cut off factor since all operations are well defined for $g_1, g_2 \in SO(3)$.

Recall that any chosen mapping I is here required to verify conditions (**I1-I3**). It is straightforward to show this is the case for the mapping (64).

That condition (**I1**) is verified, follows from smoothness of matrix inversion and multiplication; in addition to the linear operation Π. To verify conditions (**I1-I2**), note that in the same notation used to state these conditions,

$$dI(g_1, g_2)(\Sigma_r) = g_1 \Pi(g_1^{-1} \Sigma_r(g_2)),$$

and, using (56) and the definition of the Hessian,

$$\nabla^2 I(g_1, g_2)(\Sigma_r, \Sigma_v) = \Sigma_r \Sigma_v I(g_1, g_2) - \nabla_{\Sigma_r} \Sigma_v I(g_1, g_2) = 0.$$

These simply follow from the fact that (64) is linear in g_2. The latter formula immediately gives condition (**I3**), since $\Sigma_1(g_1), \Sigma_2(g_1), \Sigma_3(g_1)$ is a basis of $T_{g_1}SO(3)$. For condition (**I1**), note by the same argument

$$dI(g_1, g_1)(K) = g_1 \Pi(g_1^{-1} K),$$

for all $K \in T_{g_1}SO(3)$. However, $\kappa = g_1^{-1}K \in \mathfrak{so}(3)$, so that $\Pi(\kappa) = \kappa$. Then,

$$dI(g_1, g_1)(K) = g_1 \kappa = K,$$

which is condition (**I1**). The mapping I of (64) is especially easy to compute. Thanks to (46), the matrix inverse is the same as the transpose.

All that is needed in order to apply the particle filtering algorithm of Section 4 to the current example is the mapping I. At each step of the algorithm, as descried in 4.2, instructions 1,3 and 4 do not involve the observation process Y. This only appears through the increments ΔY_k in instruction 2. For this instruction, each ΔY_k is computed using I as in (6) and replaced in (36) in order to find particle weights.

To illustrate the above discussion, a computer experiment is now presented. For the signal model (1), the experiment considers $S = \mathbb{R}^3$. The unknown signal x is taken to be an Ornstein-Uhlenbeck process,

$$dx_t = -\nu x_t dt + dv_t, \quad x_0 = (0,0,0), \tag{65}$$

where $\nu > 0$ and v is a Brownian motion in $\mathbb{R}^3$ with variance parameter $\sigma^2 = 0.5$. As discussed in 4.2, the first step is to choose an approximation $\{x_k^\delta\}$ of x. In the following, only one value of δ is considered, $\delta = 0.1$. The x_k^δ are constructed using a first order Euler scheme, which verifies conditions (**x1-x4**).

The function $h : \mathbb{R}^3 \to \mathfrak{so}(3)$ is taken to be a linear isomorphism, mapping the canonical basis of $\mathbb{R}^3$ to the basis $\omega_1, \omega_2, \omega_3$ of $\mathfrak{so}(3)$. In other words,

$$h(s) = s^1 \omega_1 + s^2 \omega_2 + s^3 \omega_3, \tag{66}$$

for each $s = (s^1, s^2, s^3)$ in $\mathbb{R}^3$. It should be noted that, with this choice for h, the sensor function H of (48) verifies condition (**H2**) but does not verify condition (**H1**).

It was stated in 4.3 that conditions (**H1-H2**) are not essential for the overall behaviour of the particle filtering algorithm. The missing condition here is condition (**H1**), which requires the sensor function H to be bounded. It can be shown by (48) and (57) that

$$\|H(s,g)\|^2 = \|h(s)\|^2 = |s|^2,$$

where, as before, $|\cdot|$ denotes the Euclidean norm. Thus, condition (**H1**) is equivalent to the condition that the function h be bounded. It is clear this does not hold for the function h in (66). However, it should be noted that the unknown signal x of (65) is a normal process. That is, the distribution of x_t is normal for each $t \geq 0$. Thus, the distribution of $h(x_t)$ has exponentially decreasing tales and finite moments of all orders. This compensates for condition (**H1**) being dropped, since $\|h(x_t)\|^2 = |x_t|^2$ and this has an exponentially small probability of being large.

In order to simulate a trajectory of the observation process Y, a slight modification of the McShane approximation (4) was used. Applied to (50), the McShane approximation gives an approximating process Y^δ which satisfies a linear ordinary differential equation

$$\dot{Y}_t^\delta = Y_t^\delta \{h(x_t) + \Delta \hat{B}_k\}, \tag{67}$$

on each interval $[k\delta, (k+1)\delta[$; for $k \in \mathbb{N}$ and where, as in (4), $\Delta \hat{B}_k = \delta^{-1}(\hat{B}_{(k+1)\delta} - \hat{B}_{k\delta})$.

Equation (67) is a linear ordinary differential equation with time dependent coefficients. This is due to the presence of $h(x_t)$ which contains the unknown signal. Without changing the convergence rate (5), it is possible to consider another approximating proces $\bar{Y}^\delta$ which satisfies an equation with piecewise constant coefficients. This has the advantage of having a straightforward analytical solution. In effect, if $\bar{Y}^\delta$ satisfies the equation

$$\dot{\bar{Y}}_t^\delta = \bar{Y}_t^\delta \{h(x_{k\delta}) + \Delta \hat{B}_k\}, \tag{68}$$

on each interval $[k\delta, (k+1)\delta[$ for $k \in \mathbb{N}$, then

$$\bar{Y}_t^\delta = \bar{Y}_{k\delta}^\delta \exp[(t - k\delta)(h(x_{k\delta}) + \Delta \hat{B}_k)], \tag{69}$$

on the interval $[k\delta, (k+1)\delta[$. Equation (68) is the same as (67), but with $h(x_t)$ replaced by $h(x_{k\delta})$.

Formula (69) is the one used in simulating a trajectory of Y. The experiment was carried out with $b = e$ in (58). In this case (69) admits a simple interpretation since, for $s \in \mathbb{R}^3$, if $\gamma(t) = \exp[th(s)]$ then γ represents a uniform rotation with angular velocity s. Thus, $\bar{Y}^\delta$ consists in a sequence of successive uniform rotations where the angular velocity is $x_{k\delta} + \Delta B_k$ on the interval $[k\delta, (k+1)\delta[$.

Once a trajectory of Y has been simulated, the particle filtering algorithm of Section 4 can be applied immediately. The algorithm has as its input the sequence of increments $\{\Delta Y_k\}$. Here, these are obtained using the mapping I of (64). With the choice of projection Π

discussed above, $\Pi(\sigma)$ equal to the antisymmetric part of σ, this mapping becomes

$$I(g_1, g_2) = g_1[(1/2)(g_1^\dagger g_2 - g_2^\dagger g_1)], \qquad (70)$$

where (46) was used in order to avoid matrix inversion. In the particle filtering algorithm, the mapping I is only used in instruction 2 which requires evaluating the likelihood (36) (computation of normalised weights). Here, it is important to note the factor g_1 before the bracket in the right hand side of (70) can be overlooked. This is seen by replacing the invariance property (57) and definition (70) in (36). As a result, an unnecessary matrix multiplication can be avoided.

Figure 1 below is now used to illustrate the performance of the particle filtering algorithm. With the values of δ and σ^2 mentioned above, this figure shows the estimation error $|\hat{x}_t - x_t|$ for $t \in [1, T]$ where $T = 10$. Here, in the notation of (41), $\hat{x}_t$ is the estimate

$$\hat{x}_t = (1/N)\sum_{i=1}^{N} \hat{x}_M^i, \qquad (71)$$

where M is the integer part of t/δ. This is the expectation of the Monte Carlo approximation $\hat{\pi}_M^\delta$ obtained from the particle filtering algorithm. Since $\hat{\pi}_M^\delta$ converges to π_t, (as described in Proposition 5 of 4.3), $\hat{x}_t$ should also converge to the expectation of π_t, say x_t^*. That x_t^* exists follows from the fact that x_t is normal and thus square integrable, since π_t is defined in (32) as the conditional distribution of x_t given past observations.

The notation $\hat{x}_t$ obscures dependence on δ and N. However, the values of δ and N essentially control the estimation error when (65) is determined; i.e., when ν is given. Figure 1 considers $\nu = 1$ and $\nu = 0.5$. The estimation error in the present case can be understood as combining a bias and a variance. By its very definition, x_t^* is an unbiased estimator of x_t, whereas $\hat{x}_t$ is constructed as an approximation of x_t^*. The bias is then the difference (in the square mean) between $\hat{x}_t$ and x_t^*. The convergence result given in this paper, Proposition 5, is only concerned with this difference and does not say anything about the variance part of the estimation error.

The variance part of the error turns out to be the variance of the conditional distribution π_t. This is due to the following important remark.

The filtering problem considered here (with x given by (65) and Y by (50)) has a finite dimensional solution. This is quite similar to a Kalman-Bucy filter. Proposition 1 states that the conditional distribution π_t given past observations of Y is the same as given past observations of y, (in the proposition, this is the statement that $\bar{\mathcal{Y}}_t = \mathcal{Y}_t$). Now, y is here given by (54). Replacing (66) for the function h shows it is just a linear additive white noise model. Moreover, it is clear from that

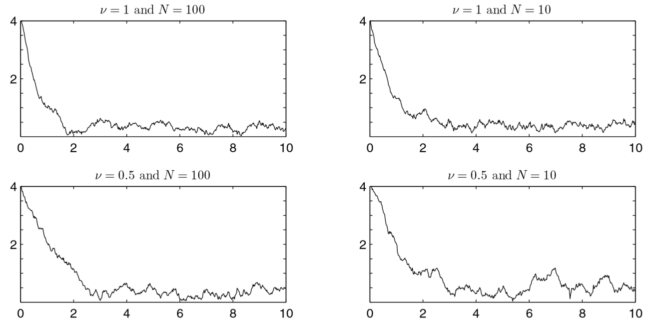


Fig. 1. Influence of ν and N on estimation errors

equation (65) satisfied by x is a linear stochastic differential equation. In particular, as already mentioned, x is normal.

Thus, the conditional distribution π_t is a normal distribution with mean x_t^* and covariance $P_t e$ where $P_t \geq 0$, (recall e is the 3×3 identity matrix). Moreover, x^* and P satisfy the Kalman-Bucy equations [36]

$$dx_t^* = -\nu x_t^* + P_t[dz_t - x_t^* dt], \qquad (72)$$

and

$$\dot{P}_t = -P_t^2 - 2\nu P_t + \sigma^2, \qquad (73)$$

where z is a process with values in $\mathbb{R}^3$ whose coordinates are $z_t^r = \langle y_t, \omega_r \rangle$.

The conditional distribution π_t is the exact solution of the current filtering problem. It can be computed by integrating the Kalman-Bucy equations (72) and (73). This can be done in a standard way after replacing dz_t in terms of $dy_t = Y_t^\dagger \circ dY_t$, which is (53). On the other hand, $\hat{\pi}_M^\delta$ is an approximation of π_t. In addition to being unbiased, x_t^* is an optimal estimator of x_t in the sense of mean square error. As $\hat{x}_t$ converges to x_t^*, the bias part of the estimation error disappears and the optimal error P_t is achieved in the limit.

From a practical point of view, there is no need here to implement a particle filter. In fact, it is even computationally less expensive to integrate (72) and (73). However, as it is well know, the existence of finite dimensional solutions is the exception rather than the rule in real situations. The connection between the present example and rigid body mechanics has already been mentioned. If Y is used to represent the pose of a rigid body, then x is the angular velocity. The filtering problem appears as the problem of tracking angular velocity based on observations of the pose alone. In general, the angular velocity of a rigid body satisfies Euler's equation of rigid body mechanics, which is far more complicated than (65) and in particular nonlinear [32]. Thus, when a realistic model is used for x, using a particle algorithm or some other approximation becomes indispensable.

Before going on to the next example, note the behavior of the estimation error in Figure 1. While $x_0 = (0,0,0)$, the particles were initialised at $(4,0,0)$. Since $\nu > 0$, this initial error is quickly overcome. However, the fact that $\nu > 0$ leads to nonzero asymptotic variance. As t goes to infinity P_t goes to $P_\infty = -\nu + (\sigma^2 + \nu^2)^{1/2}$. Thus, it is not possible to track x exactly. This problem appears since, when $\nu > 0$, the process x is ergodic. In particular, its asymptotic distribution is normal with mean 0 and variance $\sigma^2/2\nu$. This imposes a fundamental limit on estimation error.

5.2. Observations in S^2

In this second example, the observation process Y lies in the unit sphere S^2. Of course, S^2 is the set of $p \in \mathbb{R}^3$ such that $|p| = 1$. In comparison with the previous example, the current one will raise several additional difficulties. Roughly, these are due to the fact that S^2 is not a Lie group but a symmetric space of the compact Lie group $SO(3)$.

The manifold structure of S^2 is most easily understood as inherited from $\mathbb{R}^3$. Precisely, S^2 is a compact embedded submanifold of $\mathbb{R}^3$. For $p \in S^2$, the tangent space T_pS^2 is the subspace of $\mathbb{R}^3$ consisting of those vectors K such that $(K,p) = 0$. Here, $(\cdot,\cdot)$ denotes the standard Euclidean scalar product in $\mathbb{R}^3$. In particular, $(p,p) = |p|^2$.

As in the previous example, no restriction is made on the signal model (1). The sensor function H and the vector fields Σ_r are defined in terms of the action of $SO(3)$ on S^2. This is now briefly discussed.

For each $p \in S^2$, consider the two following linear mappings. First, the orthogonal projection $\Pi_p : \mathbb{R}^3 \to T_pS^2$. This is defined by

$$\Pi_p(v) = v - (p,v)p = p \times v \times p, \qquad (74)$$

for $v \in \mathbb{R}^3$, where $\times$ denotes vector product.

The second mapping is $\Sigma_p : \mathbb{R}^3 \to T_pS^2$ defined as follows. Let $\sigma_1, \sigma_2, \sigma_3$ be as in (58), with $b = e$. In other words, $\sigma_r = \omega_r$. Also, let $\sigma(v) = v^r\sigma_r$ where $v \in \mathbb{R}^3$ (this was called $h(v)$ in (66)). The mapping Σ_p is given by

$$\Sigma_p(v) = \sigma(v)p = v \times p. \qquad (75)$$

This is related to the action of $SO(3)$ on S^2 in a simple way. Note that $\sigma(v) \in \mathfrak{so}(3)$ for $v \in \mathbb{R}^3$. If $\gamma(t) = \exp(t\sigma(v))$ for $t \in \mathbb{R}$ then $\gamma(t) \in SO(3)$ and

$$\Sigma_p(v) = \left.\frac{d}{dt}\right|_{t=0} \gamma(t)p.$$

In other words, $\Sigma_p(v)$ is the velocity of the point p when it is in uniform rotation with angular velocity v.

Unlike Π_p, the mapping Σ_p is not a projection. However, both mappings are surjective and have the same kernel,

$$\mathrm{Ker}(\Pi_p) = \mathrm{Ker}(\Sigma_p) = N_pS^2,$$

where N_pS^2 is the normal space to S^2 at p. This is a one dimensional subspace of $\mathbb{R}^3$ consisting of vectors λp where $\lambda \in \mathbb{R}$.

Using the mappings Π_p and Σ_p, the tangent bundle of the sphere can be described in a covariant way. For any $p,q \in S^2$ there exists $k \in SO(3)$ such that $kp = q$. In fact, there are an infinity of such k. The following relations hold

$$\Pi_{kp}(kv) = k\Pi_p(v), \quad \Sigma_{kp}(kv) = k\Sigma_p(v), \qquad (76)$$

and can also be written

$$\Pi_q(v) = (k\Pi_p k^{-1})(v), \quad \Sigma_q(v) = (k\Sigma_p k^{-1})(v). \qquad (77)$$

Returning to the filtering problem, a general observation model where the observations lie on S^2 can be defined using either the mappings Π_p or Σ_p for each $p \in S^2$. It is preferable to use Σ_p, since it is immediately related to the action of $SO(3)$ on S^2.

The sensor function H will be assumed of the following form,

$$H(s,p) = \Sigma_p(h(s)), \qquad (78)$$

where $h : S \to \mathbb{R}^3$. The vector fields Σ_r will be defined by

$$\Sigma_r(p) = \Sigma_p(e_r), \quad r = 1,2,3, \qquad (79)$$

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

As in the previous example, all the differentiability conditions of Section 2 hold, since the operations used to define H and Σ_r are linear. Moreover, the condition of ellipticity, required to introduce the metric (21) and the Le Jan-Watanabe connection (22), is verified. This is because, by construction, the mapping Σ_p is surjective for each $p \in S^2$. Replacing the definition (75) of Σ_p in (2) gives the observation model

$$dY_t = -Y_t \times \{h(x_t)dt + \circ dB_t\}, \qquad (80)$$

where B is a standard Brownian motion in $\mathbb{R}^3$.

In the current example, the number of vector fields Σ_r is equal to 3 whereas the dimension of S^2 is equal to 2. As a result, there is no simple formula similar to (53) that can be used to find the antidevelopment process y of Y. Rather, it is necessary to consider a parallel frame along Y. This is done after introducing the metric (21) and the Le Jan-Watanabe connection (22). It turns out these are the same as the Riemannian metric that S^2 inherits from $\mathbb{R}^3$ and its associated Levi-Civita connection [29].

It is straightforward to show the Euclidean scalar product $(\cdot,\cdot)$ verifies (21). Let $p \in S^2$ and $K \in T_pS^2$. Note that

$$(K, \Sigma_r(p)) = (p \times K)^r,$$

where the right hand side is the r^{th} component of $p \times K$ in the canonical basis e_1, e_2, e_3. Evaluating the right hand side of (21) (with $K = E$) gives

$$(K, \Sigma_r(p))(K, \Sigma_r(p)) = |p \times K|^2.$$

But this is

$$(p,p)(K,K) - (K,p)^2 = (K,K),$$

since $(p,p) = |p|^2 = 1$ and $(K,p) = 0$. Thus, the metric (21) is the same as $(\cdot,\cdot)$.

For the connection (22), note that by definition this is given by

$$\nabla_K E = K(E, \Sigma_r)\Sigma_r(p),$$

for $K \in T_pS^2$ and any C^1 vector field E on S^2. This can also be written

$$\nabla_K E = \Sigma_p((p \times KE)),$$

where KE is the vector $KE = (KE^1, KE^2, KE^3)$. The last formula follows from (75) and (79) by linearity. If (75) is applied to it again, then it follows

$$\nabla_K E = p \times KE \times p = \Pi_p(KE),$$

which is the definition of the Levi-Civitation connection associated to $(\cdot,\cdot)$; see [23].

In order to construct a parallel frame along Y, let $E_0^1, E_0^2 \in T_{Y_0}S^2$ be orthonormal. Also, let E^1, E^2 be vector fields along Y solving the equation of parallel transport (9). Here, this reads

$$\nabla_{\circ dY} E_t^i = \Pi_{Y_t}(dE_t^i) = 0,$$

for $i = 1,2$ and with initial conditions E_0^1, E_0^2. Another way of writing this equation, based on (74), is

$$dE_t^i = -Y_t(E_t^i, \circ dY_t). \tag{81}$$

To obtain this, it is enough to replace in (74) the fact that

$$d(E_t^i, Y_t) = (\circ dE_t^i, Y_t) + (E_t^i, \circ dY_t) = 0.$$

If $Y_0, E_0^1, -E_0^2$ is a positively oriented orthonormal basis in $\mathbb{R}^3$, then $Y_t, E_t^1, -E_t^2$ will have this same property. It is now assumed this is the case.

From its definition (13), the antidevelopment process y of Y has its values in $\mathbb{R}^2$ and is given by

$$y_t^i = \int_0^t (E_s^i, \circ dY_s). \tag{82}$$

Let z be the process with values in $\mathbb{R}^3$,

$$z_t = h(x_t)dt + dB_t. \tag{83}$$

It follows from (80) that

$$dy_t^i = -(E_t^i, Y_t \times \circ dz_t) = -(E_t^i \times Y_t, \circ dz_t).$$

Given the chosen orientation for Y_t, E_t^1, E_t^2, this yields

$$dy_t^i = (E_t^j, \circ dz_t). \tag{84}$$

From this, it is possible to recover (25) of Proposition 1. Namely,

$$dy_t^i = h_t^i dt + d\beta_t, \tag{85}$$

where $h_t^i = (E_t^i, h(x_t))$ and β is a standard Brownian motion in $\mathbb{R}^2$.

From (84),

$$dy_t^i = h_t^i dt + (E_t^j, \circ dB_t).$$

Thus, it is enough to show

$$(E_t^j, \circ dB_t) = (E_t^j, dB_t) = d\beta_t^i.$$

That the Stratonovich differential can be replaced by an Itô differential follows from

$$(E_t^j, \circ dB_t) = (E_t^j, dB_t) + \tfrac{1}{2}(dE_t^j, dB_t),$$

where the last term denotes quadratic covariation. From (80) and (81), this is

$$(dE_t^j, dB_t) = -(Y_t, dB_t)(E_t^j, Y_t \times dB_t) = (E_t^j, Y_t \times Y_t)dt,$$

which is identically zero. That β is a Brownian motion follows from the fact that E^1, E^2 are orthonormal.

Equation (80) can be rewritten in terms of the parallel frame E^1, E^2 and the antidevelopment process y. Replacing in (80) the fact that Y, E^1, E^2 is an orthonormal basis, it follows from (82) that

$$dY_t = E_t^1 \circ dy_t^1 + E_t^2 \circ dy_t^2. \tag{86}$$

Similarly, (81) can be rewritten using (82),

$$dE_t^i = -Y_t \circ dy_t^i. \tag{87}$$

Now (86) and (87) form a system of linear stochastic differential equations which can be solved knowing the antidevelopment y. This shows that Y can be obtained if y is known. This is in spite of the fact that y has its values in $\mathbb{R}^2$ while Y has its values in $\mathbb{R}^3$.

In order to apply the particle filtering algorithm of Section 4 to the current example, it is enough to specify a mapping I verifying conditions (**I1-I3**) of 3.3. Two such mappings are now considered. First, recall that I can be given by (28) as a geodesic connector. Here, the connection ∇ is the Levi-Civita connection corresponding to the Euclidean scalar product $(\cdot,\cdot)$. Thus, geodesics are to be understood in the usual meaning of large circles. Accordingly, for $p \in S^2$ and $K \in T_pS^2$ with $|K| \neq 0$,

$$\exp_p(K) = \cos|K|p + \sin|K|(K/|K|). \tag{88}$$

Moreover, when $p, q \in S^2$ and $(p,q) \neq \pm 1$,

$$\log_p(q) = \arcsin|\Pi_p(q)|(\Pi_p(q)/|\Pi_p(q)|). \tag{89}$$

Using this last formula, it is possible to implement (28). However, this involves several nonlinear operations. Another, simpler, mapping I can be guessed from (89). Consider the following

$$I(p,q) = \Pi_p(q). \tag{90}$$

This is the first order approximation of $\log_p(q)$ and is well defined for any $p, q \in S^2$. To see that it verifies conditions (**I1-I3**) note that

$$d\Pi_p(q)(V) = \Pi_p(V),$$

for all $V \in T_qS^2$. Thus, if $V \in T_pS^2$,

$$d\Pi_p(p)(V) = V,$$

which is condition (**I2**). Condition (**I3**) reads

$$\nabla^2 \Pi_p(p)(V,V) = 0.$$

By definition of the connection ∇, the left hand side is the projection on T_pS^2 of the acceleration at $t = 0$ of the geodesic

$$\gamma_p(t) = \cos(t)p + \sin(t)(V/|V|),$$

but this acceleration is equal to $-p$, so its projection on T_pS^2 is zero. Finally, condition (**I1**) is easily verified from (74).

The particle filtering algorithm of Section 4 can be applied as before. In particular, instruction 2 is carried out by replacing (78) and (90) into (36) in order to compute particle weights.

Precisely, the likelihood function based on successive samples $Y_{k\delta}$ and $Y_{(k+1)\delta}$ becomes

$$l(s) = \exp\left((h(s) \times Y_{k\delta}, Y_{(k+1)\delta}) - \frac{\delta}{2}|h(s) \times Y_{k\delta}|^2 \right). \quad (91)$$

By (78) and (90), the first term under the exponential is

$$(H(s,Y_{k\delta}), Y_{(k+1)\delta}) = (H(s,Y_{k\delta}), I(Y_{k\delta}, Y_{(k+1)\delta})),$$

just like in (36). The second term can also be found from (78).

The above discussion is now illustrated with a computer experiment. For the signal model, consider $S = \mathbb{R}^3$. The signal is simply a constant $x^* = (0,0,1)$. The function h of (78) is taken to be the identity function $h(s) = s$ for $s \in \mathbb{R}^3$. Just like in the previous example, it can be noted that condition (**H1**) is not verified. Again, since x is normal this condition can be overlooked.

With the function h chosen in this way, the observation model (80) becomes

$$dY_t = -Y_t \times \{x' + \circ dB_t\}. \quad (92)$$

A trajectory of the observation process Y can be simulated using a formula similar to (68). Precisely, consider an approximating process $\bar{Y}^\delta$ where, (letting $\hat{x}^* = \hat{x}^r \sigma_r$),

$$\bar{Y}^\delta_t = \exp[(t - k\delta)(\hat{x}^* + \Delta \hat{B}_k)]\bar{Y}^\delta_{k\delta}, \quad (93)$$

on each interval $[k\delta(k+1)\delta[$. The value of δ used here was $\delta = 0.1$, (the same as in the previous example).

In order to ensure the process $\bar{Y}^\delta$ has its values in $S^2 \subset \mathbb{R}^3$, it is enough to take $\bar{Y}^\delta_0 \in S^2$. This was chosen to be $\bar{Y}^\delta_0 = (0,0,1)$. The product appearing in (93) is between the matrix exponential on the left, which belongs to $SO(3)$, and the vector $\bar{Y}^\delta_{k\delta}$ on the right which belongs to S^2.

Unlike the case of the previous example, (see (72) and (73)), there is here no known finite dimensional solution for optimal estimation of x^*. Application of particle filtering or of some other approximate solution is thus necessary.

Figure 2 below shows the distribution of $N = 1000$ particles in the (x_1, x_2) and (x_1, x_3) planes at times $T =$ 1.5 (top row) and $T = 3$ (bottom row). Here, (x_1, x_2, x_3) are canonical coordinates in the basis e_1, e_2, e_3. The position of x^* is designated by a + and the estimate $\hat{x}_t$ (arithmetic average of the particles as in (71)) by $\circ$. The particles were initially generated from a normal distribution μ with mean $(0.5, 0.5, 1)$ and variance 1. A large value of N was chosen for visualisation. It is possible to use $N = 100$ with a similar performance.

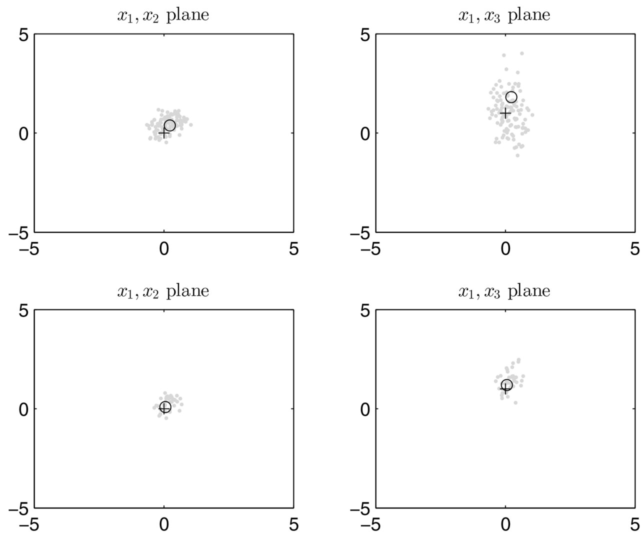


Fig. 2. Particles distribution (grey); estimate ($\circ$); true value (+)

Figure 2 shows the particle filtering algorithm of Section 4 is able to recover x^* within a relatively short time. It is interesting to note the larger variability of the particle distribution in the x_3 direction, apparent in the right column of the figure. This is because $Y_0 = (0,0,1)$ (the same as $\bar{Y}^\delta_0$) so that, initially, the component of x^* along e_3 has no effect on the position of Y.

Due to the presence of noise B, the observation process Y rapidly explores a large area of S^2 (theoretically, Y is a recurrent process in S^2). This allows for the initial ambiguity in the x_3 direction to be overcome.

In the absence of noise, Y_t rotates uniformly around x^*. If x^* and Y_0 are parallel, $Y_t = Y_0$ for all $t \geq 0$. Then, Y contains no information regarding the magnitude of x^*. Otherwise, x^* can be recovered after an arbitrarily short time (knowing the model (92)). Such a situation cannot arise in the general case of noisy observations, since Y_t and x^* do not remain parallel.

The computer experiment presented here shows that the particle filtering algorithm of Section 4 is able to successfully handle a filtering problem which is not solvable by classical methods.

6. CONCLUSION

This paper considered continuous time filtering problems where the observation process, conditionally on the unknown signal, is an elliptic diffusion in a differentiable manifold. In order to numerically solve filtering problems of this kind, the paper proposed a particle filtering algorithm which it also proved to be convergent under some additional technical conditions. Roughly, this algorithm combines the well known sequential Monte Carlo structure of a classical particle filter with the geometric construction of connector maps, used to locally linearise the observation process. To the author's knowledge, the proposed algorithm is entirely new in the literature.

The filtering problems considered in the paper are of a very general form. While this may have lead to some unnecessary abstraction, it also has clear advantages. When dealing with an applied problem, greater generality in mathematical formulation allows additional freedom in choosing a realistic observation model which includes sufficient a priori knowledge of the target application. Also, since most physical phenomena are naturally described in continuous time, the fact of starting from a continuous time formulation should accommodate the majority of physical models.

The particle filtering algorithm proposed in this paper leaves several choices open to the user wishing to implement it. These include the choice of an approximation of the hidden Markov structure and the choice of a connector map for local linearisation. This gives additional adaptability and allows for the trade-off between complexity and performance to be optimised according to applications. In any case, the paper gave precise conditions which the chosen implementation should satisfy in order to produce a consistent numerical solution.

This paper was only a first effort in the new direction of particle filtering with observations in a manifold. It was aimed at laying down a rigorous and adaptable general framework. Hopefully, additional papers strengthening convergence results and exploring in detail important engineering applications will be shortly submitted.

APPENDIX A

In 3.3, Proposition 2 was cited from [30]. However, soon after, a more general claim was made without proof. Namely, that Proposition 2 continues to hold if the mapping I of (28) is replaced by any other mapping which verifies conditions (**I1-I3**).

The proof of this claim is a repetition of the one in [30], but does not seem to have been given explicitly in the literature. For completeness, it is here provided.

In preparation, consider the following generalised Itô formula. Let f be a C^2 function on M and replace $\tau = df$ in (20). This gives

$$df(Y_t) = (df, dY_t) + (1/2)\nabla^2 f(Y_t)(\Sigma_r, \Sigma_r)dt. \qquad (94)$$

Let $I : \mathcal{M} \times \mathcal{M} \to T\mathcal{M}$ be any mapping which verifies conditions (**I1-I3**). For $\delta > 0$ and any $k \in \mathbb{N}$, let

$$I_k(q) = I(Y_{k\delta}, q) \quad q \in \mathcal{M}.$$

Conditionally on $Y_{k\delta} = p$, this is a C^2 function on M with values in $T_p\mathcal{M}$. This is by condition (**I1**). In [37], it was shown that the Itô formula (94) can be applied so

$$I(Y_{k\delta}, Y_{(k+1)\delta}) = \int_{k\delta}^{(k+1)\delta} (dI_k, dY_t) + (1/2)\int_{k\delta}^{(k+1)\delta} \nabla^2 I_k(Y_t)(\Sigma_r, \Sigma_r)dt.$$

Here dI_k and $\nabla^2 I_k$ denote differentiation component by component of the vector valued function I_k after an arbitrary choice of basis.

It will be useful to rewrite this using (17),

$$I^i(Y_{k\delta}, Y_{(k+1)\delta}) = \int_{k\delta}^{(k+1)\delta} dI_k^{ij}(Y_t)dy_t^j + (1/2)\int_{k\delta}^{(k+1)\delta} \nabla^2 I_k^i(Y_t)dt,$$

where $I_k^i = \langle I_k, E^i\rangle$ and

$$dI_k^{ij} = (dI_k^i, E^j), \quad \nabla^2 I_k^i = \nabla^2 I_k^i(E^j, E^j).$$

On the other hand

$$y^i_{(k+1)\delta} - y^i_{k\delta} = \int_{k\delta}^{(k+1)\delta} dI_k^{ij}(Y_{k\delta})dy_t^j + (1/2)\int_{k\delta}^{(k+1)\delta} \nabla^2 I_k^i(Y_{k\delta})dt.$$

This is because, by conditions (**I2-I3**),

$$dI_k^{ij}(Y_{k\delta}) = \delta_{ij}, \quad \nabla^2 I_k^i(Y_{k\delta}) = 0.$$

Note that, by (15),

$$dy_t^i = \langle E^i, H + (1/2)\nabla_{\Sigma_r}\Sigma_r\rangle dt + d\beta_t^i.$$

Here, the drift coefficient appearing before dt is uniformly bounded and β is a standard Brownian motion in $\mathbb{R}^d$.

Let, (this is the notation of (30) in Proposition 3),

$$\Delta y_k^i = y^i_{(k+1)\delta} - y^i_{k\delta}, \quad \Delta Y_k^i = I^i(Y_{k\delta}, Y_{(k+1)\delta}).$$

Then, by Itô isometry,

$$\mathbb{E}|\Delta y_k^i - \Delta Y_k^i|^2 \le C \max_j \int_{k\delta}^{(k+1)\delta} \mathbb{E}|dI_k^{ij}(Y_t) - dI_k^{ij}(Y_{k\delta})|^2 dt + \int_{k\delta}^{(k+1)\delta} \mathbb{E}|\nabla^2 I_k^i(Y_t) - \nabla^2 I_k^i(Y_{k\delta})|^2 dt,$$

where C is some positive constant (which does note depend on k).

By condition (**I1**) and the fact that the manifold $\mathcal{M}$ is compact, dI_k^{ij} and $\nabla^2 I_k^i$ are bounded and continuous. Therefore, the expectations under the integral in each term tend to zero as $\delta \downarrow 0$.

This proves that

$$\mathbb{E}|\Delta y_k^i - \Delta Y_k^i|^2 = o(\delta), \qquad (95)$$

by an extension of this result, it is straightforward to establish Proposition 2. This is now done.

As in the proposition, formula (29), let

$$R_\delta = \sum_{k\delta<t}\langle G_{k\delta}, \Delta Y_k\rangle = \sum_{k\delta<t} G_{k\delta}^i \Delta Y_k^i,$$

where $G^i_{k\delta} = \langle E^i, Gk\delta \rangle$.

By (17) and the definition of classical Itô integral, the integral $R = \int_0^t \langle G_s, dY \rangle$ is the limit in the square mean of

$$r_\delta = \sum_{k\delta<t} G^i_{k\delta} \Delta y^i_k.$$

Note as before

$$G^i_{k\delta}\Delta Y^i_k = \int_{k\delta}^{(k+1)\delta} G^i_{k\delta} dI^{ij}_k(Y_t) dy^j_t + (1/2) \int_{k\delta}^{(k+1)\delta} G^i_{k\delta} \nabla^2 I^i_k(Y_t) dt,$$

and

$$G^i_{k\delta}\Delta y^i_k = \int_{k\delta}^{(k+1)\delta} G^i_{k\delta} dI^{ij}_k(Y_{k\delta}) dy^j_t + (1/2) \int_{k\delta}^{(k+1)\delta} G^i_{k\delta} \nabla^2 I^i_k(Y_{k\delta}) dt.$$

By summing over k and using Itô isometry and the fact that $\|G_t\|$ is bounded

$$\mathbb{E}|R_\delta - r_\delta|^2 \leq C \sup_{s\leq t} \|G\|^2 \times$$
$$\sum_{k\delta<t} \max_j \int_{k\delta}^{(k+1)\delta} \mathbb{E}|dI^{ij}_k(Y_t) - dI^{ij}_k(Y_{k\delta})|^2 dt$$
$$\sum_{k\delta<t} \max_j \int_{k\delta}^{(k+1)\delta} \mathbb{E}|\nabla^2 I^i_k(Y_t) - \nabla^2 I^i_k(Y_{k\delta})|^2 dt,$$

where C is some positive constant, possibly different from before. Using again condition (**I1**) and the fact that the manifold $\mathcal{M}$ is compact it is seen that

$$\lim_{\delta\downarrow 0} \mathbb{E}|R_\delta - r_\delta|^2 = 0,$$

which, from the definition of r_δ, immediately gives Proposition 2. It is enough to write

$$\mathbb{E}|R_\delta - R|^2 \leq 2\mathbb{E}|R_\delta - r_\delta|^2 + 2\mathbb{E}|r_\delta - R|^2.$$

The first term has just been proved to converge to zero. The second term converges to zero by definition of r_δ.

As already mentioned, this proof is similar to the one in [30], but makes the additional remark that the only required properties for the mapping I are conditions (**I1-I3**). Accordingly, there is no need to restrict I to being the geodesic connector mapping (28). In its above form, the proof explicitly uses compactness of $\mathcal{M}$ in order to obtain convergence in the square mean. However, it is clear that this can be replaced by a milder assumption.

REFERENCES

[1] J. H. Manton
On the role of differential geometry in signal processing,
in *International Conference on Acoustics Speech and Signal Processing*, 2005.

[2] Y. Liu, G. Li, and Z. Shi
Tracking deformable object via particle filtering on manifolds,
in *Chinese Conference on Pattern Recognition*, 2008.

[3] G. Wang, Y. Liu, and H. Shi
Covariance tracking via geometric particle filtering,
in *Intelligent Computation Technology and Automation, 2009*, 2009.

[4] Q. Rentmeesters, P. Absil, P. Van Dooren, K. Gallivan, and A. Srivastava
An efficient particle filtering technique on the Grassman manifold,
in *International Conference on Acoustics Speech and Signal Processing*, 2010.

[5] J. Kwon, M. Choi, C. Chun, and F. C. Park
Particle filtering on the Euclidean group,
in *International Conference on Robotics and Automation*, 2007.

[6] F. Tompkins and P. J. Wolfe
Bayesian filtering on the stiefel manifold,
in *Advances in Multi-Sensor Adaptive Processing*, 2007.

[7] J. Saboune and F. Charpillet
Using interval particle filtering for marker less 3d human motion capture,
in *Tools with Artificial Intelligence*, 2005.

[8] H. Snoussi and C. Richard
Monte carlo tracking on the riemannian manifold of multivariate normal distributions,
in *Digital Signal Processing Workshop and Signal Processing Education Workshop*, 2009.

[9] A. Srivastava
A nonlinear filtering method for geometric subspace tracking,
in *Sensor Array and Multichannel Signal Processing Workshop*, 2000.

[10] M. Emery
Stochastic calculus in manifolds.
Springer-Verlag (Universitext), 1989.

[11] E. P. Hsu
Stochastic analysis on manifolds.
Americal Mathematical Society (Graduate Studies in Mathematics), 2002.

[12] J. Manton
A primer on stochastic differential geometry for signal processing,
Selected Topics in Signal Processing, IEEE Journal of, vol. 7, no. 4, pp. 681–699, 2013.

[13] J. T. Lo
Signal detection for bilinear systems,
Information Sciences, vol. 9, pp. 249–378, 1975.

[14] S. Said and J. H. Manton
On filtering with observation in a manifold: Reduction to a classical filtering problem,
SIAM Journal on Control and Optimisation, vol. 51, no. 1, pp. 767–783, 2013.

[15] M. Pontier and J. Szpirglas
Filtrage non linéaire avec observations sur une variété,
Stochastics, vol. 15, no. 2, pp. 121–148, 1985.

[16] S. K. Ng and P. E. Caines
Nonlinear filtering in Riemannian manifolds,
IMA Journal of Mathematical Control and Information, vol. 2, pp. 25–36, 1985.

[17] T. E. Duncan
Some filtering results in Riemannian manifolds,
Information and Control, vol. 35, pp. 182–195, 1977.

[18] M. Pontier
Apprixmation d'un filtre avec observation sur une variété,
Stochastics, vol. 24, no. 4, pp. 285–304, 1988.

[19] P. Del Moral
Feynman-Kac formulae: Geealogical and interacting particle systems with applications.
Springer (Probability and Applications), 2004.

[20] J. Boulanger, N. Le Bihan, S. Said, and J. Manton
Monte-carlo estimation from observation on stiefel manifold,
in *International Conference on Acoustics Speech and Signal Processing*, 2014.

[21] D. Crisan and A. Doucet
A survey of convergence results on particle filtering methods for practioners,
IEEE Transactions on Signal Processing, vol. 50, no. 3, pp. 736–746, 2002.

[22] A. Bain and D. Crisan
Fundamentals of stochastic filtering.
Springer, 2009.

[23] I. Chavel
Riemannian geometry: A modern introduction.
Cambridge University Press, 2006.

[24] L. C. G. Rogers and D. Williams
Diffusions, Markov processes and Martingales, Vol. 1.
Cambridge University Press, 2000.

[25] N. Ikeda and S. Watanabe
Stochastic differential equations and diffusion processes.
North Holland, 1988.

[26] L. C. G. Rogers and D. Williams
Diffusions, Markov processes and Martingales, Vol. 2.
Cambridge University Press, 2000.

[27] P. E. Kloeden and E. Platen
Numerical solution of stochastic differential equations.
Springer (Stochastic modelling and applied probability), 1992.

[28] M. Pontier and J. Szpirglas
Convergence des approximations de McShane d'une diffusion sur une variété compacte,
Séminaire de probabilités (Strasbourg), vol. 21, pp. 534–543, 1987.

[29] K. D. ELworthy, Y. Le Jan, and X. M. Li
On the geometry of diffusion operators and stochastic flows.
Springer (Lecture notes in mathematics 1720), 1999.

[30] R. W. R. Darling
"Martingales on manifolds and geometric Itô calculus,"
Ph.D. dissertation, The University of Warwick, 1982.

[31] J. Faraut
Analysis on Lie groups: An introduction.
Cambridge University Press, 2008.

[32] V. I. Arnold
Mathematical methods of classical mechanics.
Springer (Graduate Texts in Mathematics, 2010.

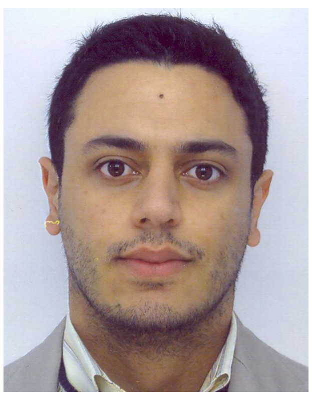
[33] S. Helgasson
Differential geometry, Lie groups and symmetric spaces.
American Mathematical Society (Graduate Studies in Mathematics), 2001.

[34] A. Estrade
Exponentielle stochastique et intègrale multiplicative discontinue,
Annales de l'Institut Henri Poincaré (B), probabilités et statistiques, vol. 28, no. 1, pp. 107–129, 1992.

[35] M. Hakim-Dowek and D. Lépingle
L'exponentielle stochastique des groupes de lie,
Séminaire de probabilités (Strasbourg), vol. 20, pp. 352–374, 1986.

[36] A. H. Jazwinski
Stochastic processes and filtering theory.
Academic Press, 1970.

[37] H. Kunita
Some extensions of itô's formula,
Séminaire de probabilités (Strasbourg), vol. 15, pp. 118–141, 1981.



Salem Said is a research scientist with CNRS since 2014, at IMS laboratory in Bordeaux, France. He currently holds an IdEx Bordeaux Chair of Installation. Previously, he held a research fellowship at the University of Melbourne school of engineering, under the supervision of Pr. Jonathan Manton. His current research focuses on information geometry and adaptive algorithms. His research also includes the areas of physical modeling and partial differential equations.

Jonathan Manton 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. Prof Manton is a Fellow of IEEE and a Fellow of the Australian Mathematical Society.

He received his Bachelor of Science (mathematics) and Bachelor of Engineering (electrical) degrees in 1995 and his 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.

Prof Manton's traditional research interests range from pure mathematics (e.g. commutative algebra, algebraic geometry, differential geometry) to engineering (e.g. signal processing, wireless communications, systems theory). More recently, he has become interested in systems biology and systems neuroscience.