ON FILTERING WITH OBSERVATION IN A MANIFOLD: REDUCTION TO A CLASSICAL FILTERING PROBLEM

SALEM SAID, JONATHAN H. MANTON *

Abstract. The current paper deals with filtering problems where the observation process, conditioned on the unknown signal, is an elliptic diffusion in a differentiable manifold. Precisely, the observation model is given by a stochastic differential equation in the underlying manifold. The main new idea is to use a Le Jan-Watanabe connection, instead of a usual Levi-Civita connection, in performing the operation of antidevelopment (both connections can be constructed from a stochastic differential equation as in the observation model). The following results are obtained. First, it is shown that antidevelopment reduces the original filtering problem to a classical one, *i.e.* an additive white noise model. Second, a new form of the Zakai and filtering equations is derived which has a structure very similar to that of classical equations. Currently, there are few well understood general numerical solution methods for filtering problems with observation in a manifold. Reduction to classical filtering problems seems desirable as it allows proven existing numerical methods for classical problems to be applied immediately. In view of practical implementation of this approach, the paper also gives an explicit pathwise construction of the antidevelopment.

Key words. stochastic filtering, stochastic differential geometry, Le Jan-Watanabe connection, elliptic diffusion, pathwise construction

AMS subject classifications. 60G35, 58J65, 60H35

1. Introduction. Stochastic filtering problems where the observation has its values in a differentiable manifold constitute an important extension of the classical additive white noise setting. They were studied in the pioneering work of Duncan [1], which was expanded by Caines [2], and also by Pontier and Szpirglas [3, 4]. While these two contributions consider different observation models, their overall methodologies are comparable. Precisely, they combine the machinery of stochastic differential geometry with standard theoretical tools, based on innovation or on change of probability. The current paper studies the same model as in [3, 4]. It proceeds along similar lines, but builds on more recent development in the geometry of diffusions.

While this may be obscured by differences in terminology, the essential operation from stochastic differential geometry in all the papers just mentioned is stochastic antidevelopment. Loosely speaking, antidevelopment sets up an equivalence between processes in a differentiable manifold and processes in a vector space of the same dimension [6]. In general, this requires a choice of affine connection. This has been taken to be a Levi-Civita connection in [1] and [3, 4]. In [2], it was an unspecified metric connection. The idea here is that a significant improvement can be made, in studying the filtering problem of [3, 4], when a Le Jan-Watanabe connection is used which is constructed in a specific way from the observation model. This is obtained by adapting a result in [5].

Using the Le Jan-Watanabe connection leads to a situation where antidevelopment yields a process depending on the unknown signal through an additive white noise model. In other words, the problem with observation in a manifold is reduced to an equivalent classical filtering problem. This suggests a two step solution method where the antidevelopment is first performed to obtain an equivalent classical problem and this is then solved using an existing classical numerical method. This seems a desirable approach, in view of the fact that numerical solution of classical filtering

---

*Department of Electrical and Electronic Engineering, The University of Melbourne, (`ssaid@unimelb.edu.au, j.manton@ieee.org`).

problems already constitutes a well developed subject [7, 8]. At least at present, this is far from being the case for filtering with observation in a manifold.

The main difference among [1, 2] and [3, 4] is the following. In the former, the observation model is given by a horizontal stochastic differential equation in the frame bundle of the underlying manifold. In the latter, the stochastic differential equation is directly on the manifold itself. In the current paper, the latter model is pursued since it is related to a great number of engineering applications; see Subsection 1.1. The stochastic differential equation in the frame bundle contains in advance (as part of the model) the dependence of the antidevelopment on the unknown signal. On the contrary, for the observation model considered here, the required dependence is uncovered by using an adequate connection.

This paper only considers continuous time observation. In fact, for purposes such as filtering, it is easier to consider continuous time than discrete time processes in manifolds; roughly speaking, there is no suitable definition for discrete time martingales in manifolds [9]. Still, a useful approach is to discretise the continuous time observation in order to obtain an approximate discrete observation problem. This was used by Solo to derive unnormalised filtering equations [10]. The rate of convergence, when a similar discretisation is applied to the model of [1, 2], was established in [11].

The following Subsection 1.1 motivates the approach taken in this paper through a special case due to Lo [15]. Subsection 1.2 describes the notation used and the main results stated in the following. The aim of the current paper is to provide systematic foundation for the two step solution method proposed above. Its implementation, numerical aspects and various applications will be the focus of future work. Here, it may be interesting to point out that reduction to a classical problem can also be used to solve theoretical questions; *e.g.* existence and uniqueness of solutions, existence of a filtering density with respect to some reference measure. Indeed [7], many such questions are solved in the classical case and the required conditions can be easily translated into conditions for the problem with observation in a manifold.

1.1. Motivation, preliminary example. Many engineering applications involve continuous time observations of the state of orientation or of configuration of a noisy physical system. Often, when such observations are possible, it is well known they are accurately described using stochastic differential equation in classical matrix manifolds [12, 13]. Local characteristics of such diffusions correspond to dynamical variables and may be controlled using inputs. As such, the problem of their estimation from observations will arise in practice. The current paper was motivated by the generalisation of the treatment in the following example.

Assume given observations of the attitude of a rigid body. Rigid body dynamics reduces to that of the angular velocity, which is controlled by torques. This reduction of dynamics is related to the existence of an underlying Lie group structure. In the engineering literature, techniques which take this group structure into account and rigorously employ stochastic calculus were developed by Willsky and Lo [14, 15]. Lo's paper [15] uses a formulation involving a general matrix Lie group. This leads to a unified treatment of several applications; *e.g.* frequency modulation and power conversion, in addition to the problem of attitude control.

Here is a "derivation" for the above example. The angular velocity $u \in \mathbb{R}^3$ of a rigid body is subject to Euler's equations [16]. If the three principal moments of inertia are equal to a common value J, this has a simple form

$$Ju = x + w \qquad \dot{x} = Q \tag{1.1}$$

where $Q \in \mathbb{R}^3$ is the applied torque. When the practical aim is to monitor the effect of unknown torques (*e.g.* actuator failures), the problem arises of tracking x. Noise is modeled by a stationary Gaussian process w with short correlation time.

Instead of $u \in \mathbb{R}^3$, consider the cross product matrix $\hat{u}$ given by $\hat{u}z = u \times z$ for $z \in \mathbb{R}^3$. The attitude of the rigid body is given by a proper orthogonal matrix Y, governed by left invariant kinematics

$$\dot{Y} = Y\hat{u} \tag{1.2}$$

When w has negligible correlation time, it may be considered as a white noise process $\sigma\dot{B}$, for some $\sigma > 0$ and a Brownian motion $B \in \mathbb{R}^3$. Equation (1.2) becomes a stochastic differential equation (the circle denotes a Stratonovich differential)

$$dY_t = Y_t \circ dy_t \qquad dy_t = (1/J)\left\{\hat{x}_t dt + \sigma d\hat{B}_t\right\} \tag{1.3}$$

where the hat notation is the same as for $\hat{u}$. The model (1.3) is essentially the one studied in [15]. The main idea used therein is that there is an equivalence between the paths (or trajectories) of y and those of Y. If a path of y is known, then solving a linear stochastic differential equation gives a corresponding path of Y. In the current example, one has access to an observed path of Y. Similarly, a corresponding path of y is given by a stochastic integral

$$y_t = \int_0^t Y_s^{-1} \circ dY_s \tag{1.4}$$

It follows from (1.3) that the process Y is a diffusion with values in the special orthogonal group $SO(3)$. Its local characteristics of drift and local diffusion are contained in the process y. This process has its values in $\mathfrak{so}(3)$, the Lie algebra of $SO(3)$.

The relation among the observation Y and the unknown signal x falls outside the framework of classical filtering problems. This is not the case for the relation among y and x, to which one could apply the Kalman-Bucy filter. Thus, it is seen that (1.4) reduces a problem with observation in a matrix manifold to a classical filtering problem. In fact, these two problems are equivalent, due to the relation between Y and y.

While Lo realised (1.4) could be applied for the paticular model (1.3), no similar reduction has been obtained for more general models; or even very similar ones where the observation does not have its values in a Lie group but in a homogeneous space. The current paper extends (1.4) to the general model of [3, 4], which is recalled in the following subsection.

1.2. Notation and main results. The approach reviewed in the previous subsection can be extended to the case where the observation is an elliptic diffusion in a differentiable manifold. To present this a clear way, here is some basic notation.

Let M be a C^3 manifold of dimension d and $(\Omega, \mathcal{A}, \mathbb{P})$ a complete probability space. On Ω, it is assumed two independent processes x and B are defined. The process x takes its values in some Polish space $(S, \mathcal{S})$, while B is a standard Brownian motion in $\mathbb{R}^d$. On M, it is assumed C^2 vector fields $(\Sigma_r; r = 1, \ldots, m)$ are given. With very little loss of generality, it is assumed M has a countable atlas. This is the case as soon as M is connected and paracompact. The augmented natural filtration of x is denoted $\mathcal{X}$. The sensor function is an application $H : \mathbb{R}_+ \times \Omega \times M \to TM$. This is a random time dependent vector field $(t, \omega, p) \mapsto H_t(\omega, p)$. For $p \in M$, it is required the

process $(H_t(p); t \geq 0)$ in T_pM given by $H_t(p)(\omega) = H_t(\omega, p)$ is $\mathcal{X}$-adapted; a typical example is when $H_t(\omega, p) = H_t(x_t(\omega), p)$. For $\omega \in \Omega$, it is required the application $(t, p) \mapsto H_t(\omega, p)$ is jointly C^1. The observation Y is assumed nonexplosive (*i.e.* for $p \in M$, if $Y_0 = p$ then Y_t is defined with values in M for $t \geq 0$) and satisfying

$$dY_t = H_t(Y_t)dt + \Sigma_r(Y_t) \circ dB_t^r \qquad (1.5)$$

Here and in all the following, summation convention is understood. A further condition will be imposed on the Σ_r to ensure Y conditioned on x is an elliptic diffusion; see Subsection 2.2. The augmented natural filtration of Y is denoted $\mathcal{Y}$. The filtering problem consists in finding the conditional distribution π_t of x_t given $\mathcal{Y}_t$ for all $t \geq 0$. As already stated, the model (1.5) is the same as in [3, 4].

The main results are Theorems 2.1 and 4.1 below, in addition to Theorem 3.1. Theorem 2.1 gives a construction generalising (1.4) to the model (1.5). That is, it defines a process y which has values in $\mathbb{R}^d$, depends on x in terms of an additive white noise model and is equivalent to Y (in the same sense as considered above). This is based on an adaptation of a result in [5]. While the construction in (1.4) follows by applying the chain rule to (1.3), that of Theorem 2.1 uses the machinery of stochastic differential geometry. Necessary background for this is provided in Subsection 2.1.

Theorem 3.1 is an application of Theorem 2.1. It requires the usual assumption that x is a time invariant Markov process; this is stated precisely in Section 3. Theorem 2.1 reduces the model (1.5) to an additive white noise model, for which filtering equations are readily available. This allows a new form of the filtering equations of [3, 4] to be recovered, by writing Y in terms of y.

Finally, Theorem 4.1 gives a practical implementation of Theorem 2.1, in the form of a pathwise construction of y. A pathwise construction is a function which maps the paths of Y into those of y. In the special case of (1.3) and (1.4), this was realised in [21]. The construction of Theorem 4.1 extends the same general technique, which was given in [20]. In a filtering problem, only an individual path of the observation Y is available. Pathwise constructions are then helpful as they lead to almost sure convergence and are robust to changes in the underlying model; see discussion in [17].

From an applied perspective, the construction of Theorem 4.1 is based on an Euler scheme but with event based sampling rather than periodic sampling. In general, an Euler scheme with periodic sampling is only convergent in probability.

The paper closes with Sections 5 and 6. Section 5 considers a particular example of Theorem 2.1. This example is meant to capture the geometric nature of the construction of (1.4). Section 6 gives some general remarks on the validity, application and further extension of the current approach.

2. Reduction to a classical problem.

2.1. Stochastic antidevelopment. As explained above, Theorem 2.1 uses the machinery of stochastic differential geometry. Precisely, it applies a transformation known as stochastic antidevelopment. The current subsection provides a succinct presentation of this concept. Prerequisites from differential geometry correspond to the first chapter of [23]. For an elegant and comprehensive introduction to stochastic differential geometry, see [22]. Stochastic differential geometry deals with semimartingales in manifolds. In order to study them, it uses affine connections. In differential geometry, affine connections distinguish certain differentiable curves as geodesics. Similarly, in stochastic differential geometry, they distinguish certain processes as being martingales. All processes in the following are pathwise continuous.

Knowing the process Y is of the form (1.5), it is natural to work with a connection which reflects the geometry of the vector fields Σ_r. This is introduced in the following Subsection. For now, let ∇ be any affine connection on M. The first definitions are of Stratonovich and Itô integrals along Y. Let $\mathcal{F} = (\mathcal{F}_t; t \geq 0)$ 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^*M, which is above Y. This means $\theta_t \in T^*_{Y_t} M$ 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

$$(\theta_t, \circ dY) = (\theta_t, H)dt + (\theta_t, \Sigma_r) \circ dB_t^r \tag{2.1}$$

$$(\theta_t, dY) = \{(\theta_t, H) + (1/2)(\theta_t, \nabla_{\Sigma_r}\Sigma_r)\} \, dt + (\theta_t, \Sigma_r)dB_t^r \tag{2.2}$$

The Stratonovich differential does not involve the chosen connection. On the other hand, the connection appears explicitly in the Itô differential. Now, 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, just like a usual Itô differential.

The case of interest will be when a Riemannian metric $\langle \cdot, \cdot \rangle$ is introduced on M. Given a process E which is $\mathcal{F}$-adapted with values in TM and above Y; (similar meaning to above). A corresponding process θ in T^*M is then given by $(\theta_t, \cdot) = \langle E_t, \cdot \rangle$. The resulting differentials of (2.1) and (2.2) are written $\langle E_t, \circ dY \rangle$ and $\langle E_t, dY \rangle$. A process such as E is called a vector field along Y.

In order to formulate the concept 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

$$\nabla_{\circ dY} E_t = 0 \tag{2.3}$$

where the left hand side is the stochastic convariant derivative. In order to give this a precise meaning, consider first the case where G is a C^1 vector field on M and $E = G(Y_t)$. Assuming the usual properties of an affine connection, (1.5) would give

$$\nabla_{\circ dY} G(Y_t) = \nabla_H G(Y_t)dt + \nabla_{\Sigma_r} G(Y_t) \circ dB_t^r \tag{2.4}$$

This is extended to a general vector field E along Y by the following transformation. For a C^2 function f on M, let $\nabla^2 f$ be the Hessian of f with respect to ∇. By definition, $\nabla_H G f = \nabla^2 f(H, G) - HGf$ and similarly for $\nabla_{\Sigma_r} G$. Now, $(\theta_t, \cdot) = \nabla^2 f(\cdot, E_t)$ is a process in T^*M as in (2.1). Thus, (2.4) can be generalised by writing

$$\nabla_{\circ dY} E_t f = dE_t f - \nabla^2 f(\circ dY, E_t) \tag{2.5}$$

which is taken as the definition of the stochastic covariant derivative. In integral notation, (2.5) defines a vector field $\int_0^t \nabla_{\circ dY} E_s$ along Y. For any C^2 function f on M, this verifies

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

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

$$d\langle E_t, K_t \rangle = \langle \nabla_{\circ dY} E_t, K_t \rangle + \langle E_t, \nabla_{\circ dY} K_t \rangle \tag{2.6}$$

It will be shown below that this identity indeed holds. In conclusion, 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 (2.3) and $\langle E_0^i, E_0^j \rangle = \delta_{ij}$. In this case, it follows from (2.6) that $\langle E_t^i, E_t^j \rangle = \delta_{ij}$ for $t \geq 0$. Given a parallel orthonormal frame (E^i), the process y is defined by

$$y_t^i = \int_0^t \langle E_s^i, \circ dY \rangle \tag{2.7}$$

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; for this, the reader could refer to [22].

An interesting property for the following is that a Stratonovich integral along Y can also be written as an integral along y. For instance, in the notation of (2.1), it is immediate from the chain rule

$$(\theta_t, \circ dY) = \theta_i(t) \circ dy_t^i \qquad \text{where} \quad \theta_i(t) = (\theta_t, E_t^i) \tag{2.8}$$

Indeed, $(\theta_t, \cdot) = \theta_i(t)\langle E_t^i, \cdot \rangle$.

To close this subsection, the proof of (2.6) is given. This uses for the first time the assumption that M has a countable atlas; say $\{(U_\kappa, z_\kappa); \kappa \in \mathbb{N}\}$. The shortest way of applying this assumption follows [22].

Fix a chart $(U_{\kappa_0}, z_{\kappa_0}) \equiv (U, z)$. Assume S, T are predictable stopping times with $S \leq T$ and $Y \in U$ on the stochastic interval $[S, T] \cap \{S < T\}$. Let $\rho(t) = t \vee S \wedge T$ and $\tilde{Y}$ the process $\tilde{Y}_t = Y_{\rho(t)}$.

Reducing to the probability subspace $\{S < T\} \subset \Omega$, the process $\tilde{Y}$ has its values in U and is adapted to $\mathcal{F}_\rho$. Moreover, in local coordinates $z_t = z(\tilde{Y}_t)$,

$$\langle E_{t\wedge T}, K_{t\wedge T} \rangle - \langle E_{t\wedge S}, K_{t\wedge S} \rangle = g_{ij}(z_t)e_t^i k_t^j - g_{ij}(z_0)e_0^i k_0^j \tag{2.9}$$

where $E_{\rho(t)} = e_t^i \partial_i$ and $K_{\rho(t)} = k_t^i \partial_i$, with $\partial_i = \partial/\partial z^i$, and g_{ij} is the metric tensor in the coordinates z. Applying the product rule, the left hand side of (2.9) reads

$$\int_0^t \partial_l g_{ij}(z_s)e_s^i k_s^j \circ dz_s^l + \int_0^t g_{ij}(z_s)k_s^j \circ de_s^i + \int_0^t g_{ij}(z_s)e_s^i \circ dk_s^j$$

If the connection ∇ is represented by Christoffel symbols Γ_{ij}^l, then it is possible to replace from the following identity

$$\partial_l g_{ij} = \Gamma_{li}^m g_{mj} + \Gamma_{lj}^m g_{mi}$$

The left hand side of (2.9) finally becomes

$$\int_0^t g_{ij}(z_s)k_s^j \left\{ \circ de_s^i + \Gamma_{lm}^i(z_s)e_s^m \circ dz_s^l \right\} + \int_0^t g_{ij}(z_s)e_s^i \left\{ \circ dk_s^j + \Gamma_{lm}^j(z_s)k_s^m \circ dz_s^l \right\}$$

On the other hand, (2.5) can be applied to the functions $f \equiv z^i$. This gives,

$$\nabla_{\circ dY} \tilde{E}_t = \left\{ \circ de_t^i + \Gamma_{lm}^i(z_t)e_t^m \circ dz_t^l \right\} \partial_i \qquad (2.10)$$

where $\tilde{E}$ is the process $\tilde{E}_t = E_{\rho(t)}$. By identification and time change, (2.9) finally reads

$$\langle E_{t \wedge T}, K_{t \wedge T} \rangle - \langle E_{t \wedge S}, K_{t \wedge S} \rangle = \int_{t \wedge S}^{t \wedge T} \langle \nabla_{\circ dY} E_s, K_s \rangle + \int_{t \wedge S}^{t \wedge T} \langle E_s, \nabla_{\circ dY} K_s \rangle$$

Under the assumption of countable atlas, this calculation can be exploited using the following result. There exist predictable stopping times $(T_n; n \in \mathbb{N})$ where $T_0 = 0$ with $T_n \uparrow \infty$ and each pair $T_n \leq T_{n+1}$ being of the form $S \leq T$ above, for some chart $\left(U_{\kappa(n)}, z_{\kappa(n)} \right)$. While this is useful in proving results such as (2.6), it should be kept in mind there is no explicit construction of the stopping times T_n.

2.2. The Le Jan-Watanabe connection. As discussed in Subsection 1.2, the aim of Theorem 2.1 is to reduce the filtering problem of (1.5) to a classical one. This amounts to replacing the process Y with values in M by a process y with values in $\mathbb{R}^d$, given as in (2.7). In order for y to have the required properties, a suitable choice of the affine connection ∇ must be made. In fact, ∇ will be the so called Le Jan-Watanabe connection corresponding to the vector fields Σ_r.

For $p \in M$, let Σ_p be the subspace of $T_p M$ spanned by the $\Sigma_r(p)$. If, independently of p, this is $\Sigma_p = T_p M$, then Y is said to be an elliptic diffusion (conditioned on x). In [3, 4], this assumption is used as follows. If Y is elliptic, elementary linear algebra implies there exists a unique Riemannian metric $\langle \cdot, \cdot \rangle$ such that,

$$\langle E, K \rangle = \langle E, \Sigma_r(p) \rangle \langle K, \Sigma_r(p) \rangle \qquad E, K \in T_p M \qquad (2.11)$$

Then, (2.7) is evaluated using the Levi-Civita connection of $\langle \cdot, \cdot \rangle$. This will *not* be the case in the following. In fact, choosing the Le Jan-Watanabe connection over the Levi-Civita connection leads to the simplification described in Theorem 2.1.

The Le Jan-Watanabe connection is defined based on [5]. Let E be any C^1 vector field on M. It follows from (2.11), this can be written $E = E^r \Sigma_r$ where $E^r = \langle E, \Sigma_r \rangle$. For $K \in T_p M$, let

$$\nabla_K E = (KE^r)\Sigma_r(p) \qquad (2.12)$$

It is readily checked this defines an affine connection ∇ compatible with $\langle \cdot, \cdot \rangle$. Precisely, for C^1 vector fields E, G on M,

$$K\langle E, G \rangle = (KE^r)G^r + E^r(KG^r) = \langle \nabla_K E, G \rangle + \langle E, \nabla_K G \rangle \qquad (2.13)$$

where (2.11) and (2.12) have been used.

The connection ∇ is known as the Le Jan-Watanabe connection. With regard to the proof of Theorem 2.1, the only necessary property of ∇ is

$$\nabla_{\Sigma_r} \Sigma_r = 0 \qquad (2.14)$$

Recall here the summation convention implies summation over r. To obtain (2.14), replace (2.11) in (2.12). Since Σ_r is a derivation,

$$\nabla_{\Sigma_r}\Sigma_r = \Sigma_r\langle\Sigma_r,\Sigma_v\rangle\Sigma_v = \Sigma_r\langle\Sigma_r,\Sigma_w\rangle\langle\Sigma_v,\Sigma_w\rangle\Sigma_v + \langle\Sigma_r,\Sigma_w\rangle\Sigma_r\langle\Sigma_v,\Sigma_w\rangle\Sigma_v$$

A simplification of the right hand side gives

$$\nabla_{\Sigma_r}\Sigma_r = \Sigma_r\langle\Sigma_r,\Sigma_w\rangle\Sigma_w + \Sigma_w\langle\Sigma_w,\Sigma_v\rangle\Sigma_v = \nabla_{\Sigma_r}\Sigma_r + \nabla_{\Sigma_w}\Sigma_w$$

which immediately gives (2.14).

THEOREM 2.1. *Let (E^i) be a parallel orthonormal frame and y given by (2.7), where the connection ∇ is defined by (2.12). 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*

$$dy_t = h_t dt + d\beta_t \qquad h_t^i = \langle E_t^i, H\rangle \qquad (2.15)$$

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 [3]. To avoid repetition, the reader is referred to Theorem IV.3 on page 135 of [3].

In order to obtain (2.15) define first

$$\beta_t^i = \int_0^t \langle E_s^i, \Sigma_r\rangle dB_s^r \qquad (2.16)$$

Clearly, β is an $\mathcal{F}$-local martingale. Moreover, (2.11) 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.

From (2.1) and (2.7),

$$dy_t = h_t dt + \int_0^t \langle E_s^i, \Sigma_r\rangle \circ dB_s^r$$

Given (2.16), it is clear (2.15) follows by showing the Stratonovich integral can be replaced by an Itô integral. Now, (2.6) and (2.4) imply

$$d\langle E_t^i, \Sigma_r\rangle = \langle E_t^i, \nabla_H\Sigma_r(Y_t)\rangle dt + \langle E_t^i, \nabla_{\Sigma_r}\Sigma_r(Y_t)\rangle \circ dB^r$$

where the fact that E^i verifies (2.3) has been used. It is finally possible to write

$$\langle E_s^i, \Sigma_r\rangle \circ dB_s^r = \langle E_s^i, \Sigma_r\rangle dB_s^r + \frac{1}{2}\langle E_t^i, \nabla_{\Sigma_r}\Sigma_r(Y_t)\rangle dt$$

but then (2.14) implies the last term is zero.   □

As already stated in Subsection 1.2, Theorem 2.1 is an adaptation of a result in [5]. In fact, this reference studies the more general case where Y is a semielliptic diffusion rather than an elliptic one. The process y then has a form somewhat different from the one in (2.15). In the current paper, obtaining (2.15) is in itself considered an important result, since this expression realises two objectives. First, it describes

a classical filtering problem. Indeed, dy depends on x through h with independent additive white noise $d\beta$. Second, along with equations (1.5) and (2.7), it generalises the setting of [15] as described in Subsection 1.1. These two aspects will be developed in Sections 3 and 5, respectively.

In anticipation of Section 5 note the following difference between (1.4) and (2.7). In the first case, y is computed directly from Y. That is, no other process is explicitly involved. On the other hand, computing y in (2.7) first requires finding the parallel orthonormal frame (E^i).

3. Recovering the filtering equations. Theorem 2.1 leads to a useful simplification of the filtering problem (1.5), described in Subsection 1.2. The statement that $\bar{\mathcal{Y}}_t = \mathcal{Y}_t$ implies the conditional distribution π_t can be computed from $\bar{\mathcal{Y}}_t$ instead of $\mathcal{Y}_t$. However, $\bar{\mathcal{Y}}$ being the natural filtration of y, it is clear π_t can be found by a slight modification of classical filtering equations.

A simple approach to deriving the filtering equations satisfied by π_t offers itself. This consists in finding the classical filtering equations giving π_t as a functional of y and then expressing Y in terms of y. This is precisely the subject of Theorem 3.1, presented in the current subsection.

The following will draw on standard techniques in stochastic filtering described in [7]. As usual, π_t is found from its application to test functions, *i.e.* bounded continuous functions $\varphi : S \to \mathbb{R}$. Indeed, π is determined up to modification by

$$\pi_t(\varphi) = \mathbb{E}\left[\varphi(x_t)|\mathcal{Y}_t\right] \tag{3.1}$$

where it is assumed π is a $\mathcal{Y}$-adapted process with values in the space of probability measures on $\mathcal{S}$; this latter space is given its weak topology. A standard assumption in deriving the filtering equations is that the process x is a time invariant Markov process. It is assumed x has generator A defined on $D(A)$, a dense subspace of the space of test functions, containing the constant function 1. For $\varphi \in D(A)$,

$$d\varphi(x_t) = A\varphi(x_t)dt + dM_t^\varphi \tag{3.2}$$

where M^φ is an $\mathcal{X}$-martingale. The distribution of x_0 is denoted μ. Recall x has a *càdlàg* modification [25].

It seems necessary to extend the notation of (3.1) beyond the class of test functions. For an $\mathcal{F}$-adapted process η with $\mathbb{E}|\eta_t| < \infty$, the expression $\pi_t(\eta)$ will stand for the conditional expectation $\mathbb{E}[\eta_t|\mathcal{Y}_t]$. Note that η can be real or vector valued.

THEOREM 3.1. *Assume $Y_0 = p$ for some $p \in M$ and $\|H\|$ is bounded, where $\|H\|^2 = \langle H, H \rangle$. Assume also x verifies (3.2). The process π of (3.1) has a càdlàg modification satisfying $\pi_0 = \mu$ and solving the Kushner-Stratonovich equation*

$$d\pi_t(\varphi) = \pi_t(A\varphi)dt + \langle \pi_t(\varphi H) - \pi_t(\varphi)\pi_t(H), dI_t \rangle \tag{3.3}$$

for $\varphi \in D(A)$. Here $\langle \cdot, dI_t \rangle = \langle \cdot, dY_t - \pi_t(H)dt \rangle$.

Proof. The proof will apply the change of probability method to (2.15). This yields the Zakai equation as a byproduct. Application of the change of probability method to (2.15) is a standard procedure, see [7]. Thus, the focus will be on showing necessary technical conditions hold and on obtaining the equation (3.3) from the classical Kushner-Stratonovich equation, which involves y.

Following [25], it can be shown Y is $\mathcal{F}$-adapted. By Theorem 2.1, the same is then true for y. Recall also, β is an $\mathcal{F}$-Brownian motion. Since $\|H\|$ is bounded, the

Cauchy-Schwarz inequality and (2.15) imply h is bounded. It follows Z defined below is a square integrable martingale, († stand for the transpose),

$$Z_t = \exp\left(-\int_0^t h_s^\dagger d\beta_s - \frac{1}{2}\int_0^t h_s^\dagger h_s ds\right) \tag{3.4}$$

Using the fact that Y is continuous it can be seen $H(Y)$ and thus h are continuous $\mathcal{F}$-adapted. This ensures the stochastic integral under the exponential is well defined.

Given Z is a square integrable martingale, it becomes possible to follow the same steps as in [7]. Consider the probability measures $\tilde{\mathbb{P}}_t$ on $\mathcal{F}_t$ given by $d\tilde{\mathbb{P}}_t = Z_t d\mathbb{P}$. These are compatible and together define a probability measure $\tilde{\mathbb{P}}$ on $\cup_{t\geq 0}\mathcal{F}_t$. Then, under $\tilde{\mathbb{P}}$, y is an $\mathcal{F}$-Brownian motion and so independent from x. Moreover, the process π is given by the Kallianpur-Striebel formula,

$$\pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(1)} \qquad \rho_t(\varphi) = \tilde{\mathbb{E}}\left[\varphi(x_t)\tilde{Z}_t | \bar{\mathcal{Y}}\right] \tag{3.5}$$

where $\tilde{\mathbb{E}}$ denotes expectation under $\tilde{\mathbb{P}}$ and

$$\tilde{Z}_t = \exp\left(\int_0^t h_s^\dagger dy_s - \frac{1}{2}\int_0^t h_s^\dagger h_s ds\right) \tag{3.6}$$

That $\pi_0 = \mu$ follows from the fact that $Y_0 = p$ is deterministic. Existence of a *càdlàg* version of π is a result of the fact that x has a *càdlàg* modification.

Clearly, the filtration $\mathcal{F}$ is right continuous. Applying the product rule to $\varphi(x)\tilde{Z}$ and using (3.2) and independence of x and y, it is possible to obtain the Zakai equation

$$d\rho_t(\varphi) = \rho_t(A\varphi)dt + \rho_t(\varphi h^\dagger)dy_t \tag{3.7}$$

and the Kushner-Stratonovich equation

$$d\pi_t(\varphi) = \pi_t(A\varphi)dt + \left[\pi_t(\varphi h^\dagger) - \pi_t(\varphi)\pi_t(h^\dagger)\right]di_t \tag{3.8}$$

where $di_t = dy_t - \pi_t(h)dt$.

Now, the main claim is that (3.8) is equivalent to (3.3). Both equations have the same term $\pi_t(A\varphi)dt$ on the right hand side. Thus, it is enough to compare the "innovation" terms. Starting with (3.8), note from (2.15)

$$\pi_t(\varphi h^\dagger)dy_t = \langle \pi_t(\varphi H), E^i\rangle\langle E_t^i, \circ dY\rangle$$

Applying (2.15) again, along with (2.2) and (2.14), shows this can also be written

$$\pi_t(\varphi h^\dagger)dy_t = \langle \pi_t(\varphi H), E^i\rangle\langle E_t^i, dY\rangle = \langle \pi_t(\varphi H), dY\rangle$$

In a similar way,

$$\pi_t(\varphi h^\dagger)\pi_t(h)dt = \langle \pi_t(\varphi H), E^i\rangle\langle \pi_t(H), E^i\rangle dt = \langle \pi_t(\varphi H), \pi_t(H)\rangle dt$$

For the remaining terms, the same calculation can be repeated but with $\pi_t(\varphi)$ replacing φ. The proof is then complete. $\qquad\square$

Here are some comments on Theorem 3.1. First, the hypothesis of bounded $\|H\|$ is clearly too strong. It can be replaced by any hypothesis guarantying that Z defined in (3.4) is a martingale (not necessarily square integrable). Of course, the point here

is more the geometric formulation of (3.3) than the technical conditions required from stochastic calculus.

Second, the transformation of (3.8) to (3.3) can also be applied to (3.7). This gives the Zakai equation

$$d\rho_t(\varphi) = \rho_t(A\varphi)dt + \langle \rho_t(\varphi H), dY \rangle \tag{3.9}$$

which has the advantage of being linear. Note also, since $A1 = 0$ from (3.2),

$$\rho_t(1) = \exp\left(\int_0^t \langle \pi_s(H), dY \rangle - \frac{1}{2} \int_0^t \|\pi_s(H)\|^2 ds \right) \tag{3.10}$$

Finally, it is important to note equations (3.3) and (3.9) have a simpler form than the equivalent ones in [3, 4]. This is because using the Le Jan-Watanabe connection leads to a more adequate expression for the Itô integral along Y. For example, in [3], the innovation dI_t is written $dI_t = (H_t - \pi_t(H))dt + \Sigma_r(Y_t)dB_t^r$. If the Levi-Civita connection is used, this cannot be expressed as an Itô integral along Y. Due to the same difficulty, the Zakai equation given in [4] involves two additional terms and ends up being a Stratonovich equation.

4. Pathwise construction of y. Theorem 2.1 suggests the filtering problem (1.5) can be tackled in two steps. First, computing the process y from (2.7). Second, since y satisfies (2.15), applying a classical method of solution; in other words, using y as the input of a classical filtering algorithm. In the current section, a practical implementation of the first step is described.

In a realistic filtering situation, available observation takes the form on an individual path of the process Y. Formally, this is the function $\hat{Y}(t) = Y_t(\omega)$ for $t \geq 0$, where $\omega \in \Omega$ is fixed. Of course, $\hat{Y}$ can only be sampled at discrete intervals. Rather than the usual requirement of periodic sampling (*i.e.* samples obtained at regular intervals), the implementation given below uses event based sampling. This is a widespread technique, generally considered more adequate for many of the applications hinted at in Subsection 1.1; see [26].

Calculations will be carried out using a countable atlas $\{(U_\kappa, z_\kappa); \kappa \in \mathbb{N}\}$. This is required to have the property that, for any differentiable function f on M and any $\kappa \in \mathbb{N}$, the local representative $f \circ z_\kappa^{-1}$ is bounded. Recall the assumption that M has countable atlas. As is well known in differential geometry [24], this implies the existence of a locally finite compatible atlas $\{(V_\kappa, z_\kappa); \kappa \in \mathbb{N}\}$ where $U_\kappa = z_\kappa^{-1}(D_1)$ form an open cover of M; here D_1 is the open unit ball in $\mathbb{R}^d$. Thus, one could restrict attention to U_κ constructed in this way.

A pathwise construction of y is usually given by defining a function which maps the paths of Y into those of y. To describe this more precisely, let $\Omega_1 = C(\mathbb{R}_+, M)$ and $\Omega_2 = C(\mathbb{R}_+, \mathbb{R}^d)$ be the spaces of continuous functions from $\mathbb{R}_+$ to M and to $\mathbb{R}^d$. Both are considered with the topology of local uniform convergence.

Without loss of generality, it is assumed functions $\hat{Y}$ defined as above are continuous for all $\omega \in \Omega$. The process Y can then be written as a measurable function $Y : \Omega \to \Omega_1$. That is, $Y(\omega) = \hat{Y}$. Similarly, it is assumed $y : \Omega \to \Omega_2$. A pathwise construction is then given by a function $\mathcal{L} : \Omega_1 \to \Omega_2$ such that $\mathcal{L} \circ Y$ is a modification of y. Moreover, $\mathcal{L}$ is required to be measurable.

In order to localise the paths of Y to coordinate charts, consider the stopping times $(\tau_n; n \in \mathbb{N})$ defined as follows. Let $\tau_0 = 0$. For $\kappa \in \mathbb{N}$, let $\tau_1^\kappa = \inf\{t \geq 0, Y_t \notin U_\kappa\}$. Define $\tau_1 = \sup_\kappa \tau_1^\kappa$. Assume τ_n has been defined, then τ_{n+1} is obtained in the same way. Let $\tau_{n+1}^\kappa = \inf\{t \geq \tau_n, Y_t \notin U_\kappa\}$ and define $\tau_{n+1} = \sup_\kappa \tau_{n+1}^\kappa$.

It is easily seen $(\tau_n; n \in \mathbb{N})$ is a strictly increasing sequence of predictable stopping times (with respect to $\mathcal{Y}$) and $\tau_n \uparrow \infty$. For $n \in \mathbb{N}$ and $\omega \in \Omega$, there exists $\kappa_n \equiv \kappa(n, \omega)$ such that $Y_t(\omega) \in \bar{U}_{\kappa_n}$ for $\tau_n(\omega) \leq t \leq \tau_{n+1}(\omega)$; where $\bar{U}_{\kappa_n}$ is the closure of U_{κ_n}. It is possible to choose κ_n such that it is $\mathcal{Y}_{\tau_{n+1}}$-measurable.

THEOREM 4.1. *Let (E^i) be a parallel orthonormal frame and y given by (2.7). There exists a measurable $\mathcal{L} : \Omega_1 \to \Omega_2$ such that $\mathcal{L} \circ Y$ is a modification of y.*

Proof. $\mathcal{L}$ can be constructed recursively. For $n \in \mathbb{N}$, assume given a measurable $\mathcal{L}_n : \Omega_1 \to \Omega_2$ such that $\mathcal{L}_n \circ Y$ is a modification of y^{τ_n}; that is, of the process y stopped at τ_n. The aim is to construct $\mathcal{L}_{n+1}$ with a similar property for $y^{\tau_{n+1}}$. Let $\sigma(t) = (\tau_n + t) \wedge \tau_{n+1}$ and $\tilde{Y}$ and $\tilde{E}^i$ the processes $\tilde{Y}_t = Y_{\sigma(t)}$ and $\tilde{E}^i_t = E^i_{\sigma(t)}$. Also, let $\tilde{B}$ be the Brownian motion in $\mathbb{R}^m$ where $\tilde{B}_t = B_{\tau_n+t} - B_{\tau_n}$. Consider the process $\tilde{y}$ defined by $\tilde{y}_t = y_{\sigma(t)} - y_{\tau_n}$. Note here τ_n is an $\mathcal{F}$-stopping time, since it is a $\mathcal{Y}$-stopping time. The processes just defined are $\mathcal{F}_\sigma$-adapted. Fix an arbitrary $\kappa \in \mathbb{N}$ and reduce to the probability subspace $\{\kappa_n = \kappa\}$. Then, $\tilde{Y}$ has its values in $\bar{U}_\kappa$. In local coordinates $z_t = z_\kappa(\tilde{Y}_t)$, let $\tilde{E}^i_t = e^j_i(t)\partial_j$ where $\partial_j = \partial/\partial z^j_\kappa$. Then, z and e^j_i verify the following stochastic differential equations

$$de^j_i(t) = -\Gamma^j_{lm}(z_t)e^m_i(t) \circ dz^l_t \tag{4.1}$$

$$dz^l_t = \left[h^l(z_t) + (1/2)\sigma^m_r(z_t)\partial_m\sigma^l_r(z_t)\right]dt + \sigma^l_r(z_t)d\tilde{B}^r_t \tag{4.2}$$

Here, h^l and σ^l_r are local representatives of H and Σ_r; Γ^j_{lm} are Christoffel symbols. By construction of U_κ, these are assumed to be functions defined on $\mathbb{R}^d$ which are differentiable and bounded along with their derivatives. The same assumption is made in the following for other local expressions. The process $\tilde{y}$ can be written in matrix notation, (when passing to matrix notation, superscripts number matrix rows and subscripts number matrix columns)

$$\tilde{y}_t = \int_0^t v_t \circ dz_t \tag{4.3}$$

Here, the matrix v_t has elements $v^i_l(t)$ given by $v^i_l(t)e^l_j(t) = \delta_{ij}$. Indeed, this follows from the definition of $\tilde{y}$ and orthonormality of $(\tilde{E}^i)$. Applying the product rule to (4.1), it follows from the definition of v_t that

$$dv_t = v_t d\bar{\Gamma}_t \tag{4.4}$$

Note that this equation has been written in Itô form. Here, $d\bar{\Gamma}_t$ has matrix elements $\Gamma^k_{lj}(z_t)dz^l_t + R^k_j(z_t)dt$ where $2R^k_j = g^{ml}\partial_m\Gamma^k_{lj} + g^{lp}\Gamma^k_{pm}\Gamma^m_{lj}$. In order to obtain $\mathcal{L}_{n+1}$ it is enough to construct $\ell_{n+1} : \Omega_2 \to \Omega_2$ measurable such that $\ell_{n+1} \circ z$ is a modification of $\tilde{y}$; note again the assumption $z, \tilde{y} : \Omega \to \Omega_2$. Indeed, $\mathcal{L}_{n+1}$ can then be obtained from ℓ_{n+1} by composing with z_κ and adding to $\mathcal{L}_n$.

Since (4.4) is linear, it is enough to consider the case $v_0 = I_d$, the $d \times d$ identity matrix. For lighter notation, let $\ell \equiv \ell_{n+1}$. This is defined following [20]. Precisely, ℓ is obtained as a limit of continuous functions $(\ell^k; k \in \mathbb{N})$. For $k \in \mathbb{N}$ and $\hat{z} \in \Omega_2$ let $a^k_0(\hat{z}) = 0$ and define a^k_j for $j \geq 1$ by $a^k_{j+1}(\hat{z}) = \inf\{t > a^k_j(\hat{z}), |\hat{z}_t - \hat{z}(a^k_j(\hat{z}))| \geq 2^{-k}\}$. Here, $|\cdot|$ is the Euclidean norm in $\mathbb{R}^d$. The image $\hat{y} = \ell^k(\hat{z})$ of $\hat{z}$ is given by the following formula,

$$\hat{y}(t) = \sum_{j \geq 0} \hat{v}(a^k_j(\hat{z}))\Delta^k_j\hat{z}(t) + \sum_{j \geq 0} \hat{v}(a^k_j(\hat{z}))q(a^k_j(\hat{z}))\Delta^n_i(t) \tag{4.5}$$

where $q \in \Omega_2$ has coordinates $q^k = g^{jl}(\hat{z})\Gamma_{lj}^k(\hat{z})$ and $\hat{v}$ is the matrix valued function

$$\hat{v}(t) = \prod_{j \geq 0} \left[I_d + \Delta_j^k \hat{\Gamma}(t) \right] \tag{4.6}$$

Here, $\Delta_j^k \hat{\Gamma}(t)$ has matrix elements $\Gamma_{kl}^j(\hat{z}(a_j^j(\hat{z}))\Delta_j^k \hat{z}^k(t)$. In the above formulas,

$$\Delta_j^k \hat{z}(t) = \hat{z}(t \wedge a_{j+1}^k(\hat{z})) - \hat{z}(t \wedge a_j^k(\hat{z})) \qquad \Delta_i^n(t) = t \wedge a_{j+1}^k(\hat{z}) - t \wedge a_j^k(\hat{z})$$

It is lengthy but straightforward that ℓ^k defined by (4.5) and (4.6) is continuous. Now, $\ell(\hat{z})$ is defined to be the limit in Ω_2 of $\ell^k(\hat{z})$ where it exists and a zero function otherwise. By construction, ℓ is measurable. It follows from Theorem 1 on page 126 of [20] that $\ell \circ z$ is a modification of $\tilde{y}$. For this, note restriction to a probability subspace preserves local characteristics of semimartingales. $\qquad \square$

Admittedly, the proof of Theorem 4.1 has a somewhat terse form. This should not obscure the fact that its application in a practical implementation is straightforward. Given a path $\hat{Y}$, the prescription is to apply the numerical scheme given by (4.5) and (4.6) in each chart U_κ visited by $\hat{Y}$. Adding the resulting contributions $\hat{y}$ together yields a pathwise local uniform approximation of y. The use of event based sampling is expressed in the proof through the functions a_j^k. Indeed, it is enough to sample the path $\hat{Y}$ when its local representative $\hat{z}$ crosses the levels $j2^{-k}$, $j \in \mathbb{N}$.

5. Example of a parallelisable manifold. Based on Theorem 2.1, it is possible to claim that (2.7) is a generalisation of (1.4). Indeed, (2.7) applies in a far more general context than (1.4) and realises the same objective of reduction to an additive white noise model. However, this claim is not entirely satisfactory. If (2.7) is a generalisation of (1.4), then it should be possible to apply (2.7) directly in the context of (1.3) and obtain (1.4).

In the current section, it is shown this is true. It turns out that, from a geometric point of view, (1.4) corresponds to the application of (2.7) to an elliptic diffusion in a parallelisable manifold.

To begin, note that (1.3) can be written in the form (1.5). Returning to the notation of Subsection 1.1, it is well known $\hat{u} = u^i \sigma_i$ where σ_i is the matrix of infinitesimal counterclockwise rotation around the ith coordinate axis. To each σ_i there corresponds a left invariant vector field on $SO(3)$. This is noted Σ_i, with $\Sigma_i(R) = R\sigma_i$ for $R \in SO(3)$. With this notation, (1.3) becomes

$$dY_t = H_t dt + \Sigma_i(Y_t) \circ dB_t^i \qquad H_t = x_t^i \Sigma_i(Y_t) \tag{5.1}$$

Recall that $(\sigma_i; i = 1, 2, 3)$ is a basis of $\mathfrak{so}(3)$. It follows that Y is an elliptic diffusion in $SO(3)$; that is, (5.1) is indeed a special case of (1.5). This is since $(\Sigma_i(R); i = 1, 2, 3)$ is a basis of $T_R SO(3)$ for each $R \in SO(3)$. The application of (2.7) to (5.1) only depends on this last fact. Thus, it is interesting to study the following general situation.

In the setting of (1.5), assume the manifold M is parallelisable. Precisely, that $(\Sigma_i(p); i = 1, \ldots, m)$ is a basis of $T_p M$ for each $p \in M$; thus $m = d$. Now, (2.11) and (2.12) simplify to the following relations

$$\langle \Sigma_i, \Sigma_j \rangle = \delta_{ij} \qquad \nabla_{\Sigma_i} \Sigma_j = 0 \tag{5.2}$$

The latter identity is a kind of special case of (2.14). Let $(E^i; i = 1, \ldots, m)$ be a parallel orthonormal frame. It follows from (2.6) and (2.4) that

$$d\langle E_t^i, \Sigma_j(Y_t) \rangle = \langle E_t^i, \nabla_{\circ dY} \Sigma_j(Y_t) \rangle = 0$$

For the second equality, note from (5.2) that $\nabla_K \Sigma_j(p) = 0$ for $p \in M$ and $K \in T_p M$. Thus, there exists an orthogonal matrix a such that

$$E_t^i = a_j^i \Sigma_j(Y_t) \tag{5.3}$$

In this case, (2.7) reads

$$y_t^i = \int_0^t \langle a_j^i \Sigma_j(Y_s), \circ dY \rangle \tag{5.4}$$

This formula should be compared to the remark made after the proof of Theorem 2.1, regarding the difference between (1.4) and (2.7). In (5.4), the (E^i) do not appear explicitly. Thus y can be computed directly from Y, just like in (1.4). Moreover, it follows from (2.1) and (5.2) that

$$dy_t^i = a_j^i \left\{ h_t^j dt + dB_t^j \right\} \qquad h_t^j = \langle \Sigma_j(Y_t), H \rangle \tag{5.5}$$

which is a special case of (2.15).

Now, it will be shown that (5.4) applied to (5.1) yields (1.4). For simplicity, put $a_j^i = \delta_{ij}$. This corresponds to a choice of initial condition. The Riemannian metric $\langle \cdot, \cdot \rangle$ will be computed as

$$\langle E, K \rangle = -\frac{1}{2} \mathrm{tr} \left(R^{-1} E R^{-1} K \right) \qquad E, K \in T_R SO(3) \tag{5.6}$$

here tr denotes the trace of a matrix. This is compatible with (5.2) since $\mathrm{tr}(\sigma_i \sigma_j) = -2\delta_{ij}$, by a direct calculation. It follows from (5.4) and (5.6) that

$$y_t^i = \int_0^t \langle \Sigma_i(Y_s), \circ dY \rangle = \left\langle \sigma_i, \int_0^t Y_s^{-1} \circ dY_s \right\rangle$$

This immediately implies (1.4) where $y_t = y_t^i \sigma_i$.

The study of (1.3), or equivalently (5.1), by considering the case of a parallelisable manifold M is somewhat unusual. More often, these equations are studied as a special case of left invariant diffusions defined on Lie groups; see for example page 27 of [28]. If M is taken to be a Lie group, then the connection given in (5.2), (the Le Jan-Watanabe connection corresponding to $(\Sigma_r; i = 1, \ldots, m)$), is known as the Cartan-Schouten $(-)$-connection. In this case, the antidevelopment y has been introduced under the name of Lie group stochastic logarithm, in [29, 30].

6. Remarks on modeling and application. The existing mathematical literature concerning filtering with observation in a manifold relies in a fundamental way on stochastic differential geometry. On the other hand, applying the machinery of stochastic differential geometry depends on a choice of affine connection. Thus, from a mathematical point of view, the main idea in the current paper is to use the Le Jan-Watanabe connection instead of the Levi-Civita connection when considering the model (1.5). In effect, this is the key ingredient in both Theorems 2.1 and 3.1.

As noted in Subsection 1.1, the engineering literature had already "discovered" the Le Jan-Watanabe connection in the special case of processes in matrix Lie groups. A rigorous form of this statement was given in Section 5 and related to the general case of abstract Lie groups.

Theorems 2.1 and 3.1 show that the simple and meaningful results arising in the matrix Lie group case extend to arbitrary manifolds, provided the Le Jan-Watanabe

connection is used. In principle, this means existing engineering applications which consider processes in matrix Lie group can be extended to a far more general context. However, in order to proceed in a concrete way, it is necessary to distinguish the engineering applications in which elliptic diffusions in manifolds will constitute useful mathematical models. This is an open question, requiring some kind of theory as to the physical (or biological, *etc*) origin of these diffusions. Still, it is possible to give a more specific answer.

The differentiable manifolds usually of interest to engineering are those related in some direct way to matrix Lie groups. The usual examples are Grassmann or Stiefel manifolds of various dimensions. These support usual mathematical models in telecommunications, array signal processing and path planning for mechanical systems. The reason why these manifolds arise is the existence of extensive symmetries in many engineering dynamical systems. While diffusions in Grassmann or Stiefel manifolds are closely related to diffusions in matrix Lie groups, they do not admit an easy reduction as the one described in Subsection 1.1. This observation should lead to concrete examples of applications which stand to benefit from Theorem 2.1.

In view of such applications, some general remarks are possible. The chosen mathematical approach to an engineering application should take into account, among other things, performance requirements and the availability of data. In the current paper, the proposed implementation of Theorem 2.1 is given by Theorem 4.1. This at least has the property of being a pathwise construction. However, even in a purely discrete time situation, Theorem 2.1 can be used to motivate pertinent implementations.

It might be considered undesirable that Theorem 4.1 uses local coordinate charts rather than directly assuming given an embedding in Euclidean space. Now, the use of local coordinate charts is completely general. Also, it should be noted that even when a manifold is naturally embedded in Euclidean space, there is no guarantee that the embedding gives a simple description of parallel transport with respect to the Le Jan-Watanabe connection. However, this is the central ingredient in Theorem 2.1. For either Grassmann or Stiefel manifolds, there are several available choices of atlases with the properties necessary for Theorem 4.1 and using them presents no special difficulty. In general, the use of local coordinate charts offers the advantage that numerical errors do not lead to points "falling off" the manifold.

A reasonable generalisation of the model (1.3) is to the case where the resulting diffusion may be semielliptic rather than elliptic. Indeed, this already arises quite often for processes in matrix Lie groups; for which it can be dealt with using the same approach as in Subsection 1.1.

Under some additional requirements, the reduction of semielliptic diffusions in a way similar to Theorem 2.1 can also be realised using the Le Jan-Watanabe connection. However, the resulting antidevelopment is not anymore given by an additive white noise model. This is because, roughly speaking, noise only appears in a subspace of the tangent space. A natural question would be to find the filtering equations giving the corresponding generalisation of Theorem 3.1. This could be the basis for further extension of the current approach.

The reduction of a filtering problem with observation in a manifold to an equivalent classical filtering problem may also be used to answer theoretical question. Indeed, the classical setting offers well known results concerning issues such as robustness, existence of smooth densities and approximation. It should be possible to transform all of these into similar results for the original problem with observation in a manifold.

REFERENCES

[1] T. E. Duncan, *Some filtering results in Riemann manifolds*, Information and Control, 35 (1977), pp. 182–195.

[2] S. K. Ng and P. E. Caines, *Nonlinear filtering in Riemannian manifolds*, IMA J. Math. Control Inf., 2 (1985), pp. 25–36.

[3] M. Pontier and J. Szpirglas, *Filtering non linéaire avec observation sur une variété*, Stochastics, 15 (1985), pp. 121–148.

[4] M. Pontier and J. Szpirglas, *Filtering on manifolds*, Lecture Notes in Control and Information Sciences, 91 (1987), pp. 147–160.

[5] K. D. Elworthy and Y. Le Jan and X. Li, *On the geometry of diffusion operators and stochastic flows*, Springer-Verlag. Berlin-Heidelberg, 1999.

[6] E. P. Hsu, *Stochastic analysis on manifolds*, American Mathematical Society. 2002.

[7] A. Bain and D. Crisan, *Fundamentals of stochastic filtering*, Springer Science. 2010.

[8] D. Crisan and B. Rozovskii (editors), *The Oxford handbook of nonlinear filtering*, Oxford University Press, 2011.

[9] M. Emery and G. Mokobodzki, *Sur le barycentre d'une probabilité dans une varieté*, Séminaire de Probabilités (Strasbourg), 25 (1991), pp. 220–233.

[10] V. Solo, *On nonlinear state estimation in a Riemannian manifold*, 48th IEEE Conference on Decision and Control, (2009), pp. 8500–8505.

[11] M. Pontier, *Approximation d'un filtre avec observation sur une varieté compacte*, Stochastics 24 (1988), pp. 285–304.

[12] G. S. Chirikjian, *Stochastic models, information theory and Lie groups, Volume 1*, Birkhäuser. Boston, 2009.

[13] G. S. Chirikjian, *Stochastic models, information theory and Lie groups, Volume 2*, Birkhäuser. Boston, 2011.

[14] A. S. Willsky, *Estimation and detection of signals in multiplicative noise*, IEEE Trans. Inf Theory., 21 (1975), pp. 472–474.

[15] J. T. H. Lo, *Signal detection for bilinear systems*, Information Sciences, 9 (1975), pp. 249–278.

[16] V. Jurdjevic, *Geometric control theory*, Cambridge University Press. New York, 1997.

[17] R. L. Karandikar, *On pathwise stochastic integration*, Stoch. Proc. Appl. 57 (1995), pp. 11–18.

[18] D. Talay, *Resolution trajectorielle et analyse numerique des equations differentielles stochastiques*, Stochastics 9 (1983), pp. 275–306.

[19] M. H. A. Davis, *A pathwise solution of the equations of nonlinear filtering*, Teor. Veroyatnost. i Primenen., 27 (1982), pp. 160–167.

[20] R. L. Karandikar, *Pathwise solution of stochastic differential equations*, Sankhayā Ser. A, 43 (1981), pp. 121–132.

[21] R. L. Karandikar, *A.s. approximation results for multiplicative stochastic integrals*, Séminaire de Probabilités (Strasbourg), 16 (1982), pp. 384–391.

[22] M. Emery, *Stochastic calculus in manifolds*, Springer-Verlag. Berlin-Heidelberg, 1989.

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

[24] J. E. Marsden and T. Ratiu and R. Abraham, *Manifolds, tensor analysis and applications*, Springer-Verlag. New York, 2011.

[25] H. Kunita, *Stochastic flows and stochastic differential equations*, Cambridge University Press. 1990.

[26] K. J. Åström and B. Bernhardsson, *Comparison of Riemann and Lebesgue sampling for first order stochastic systems*, in Proceedings of the 41st IEEE Conference on Decision and Control. Vol. 2, pp. 2011–2016.

[27] N. Ikeda and Sh. Watanabe, *Stochastic differential equations and diffusion processes*, North-Holland Publishing Company. 1981.

[28] M. Liao, *Lévy processes in Lie groups*, Cambridge University Press. 2004.

[29] M. Hakim-Dowek and D. Lépingle, *L'exponentielle stochastique des groupes de Lie*, Séminaire de Probabilités (Strasbourg), 20 (1986), pp. 352–374.

[30] A. Estrade, *Exponentielle stochastique et intégrale multiplicative discontinues*, Séminaire de Probabilités (Strasbourg), 28 (1992), pp. 107–129.