King's Research Portal

DOI:

[10.1007/978-3-319-47058-0_5](https://doi.org/10.1007/978-3-319-47058-0_5)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Armstrong, J., & Brigo, D. (2016). Extrinsic projection of Ito SDEs on submanifolds with applications to non-linear filtering. In F. Crichtley, & K. Dodson (Eds.), *Computational information geometry for image and signal processing* (pp. 101-120). (Signals and Communication Technology). Springer. https://doi.org/10.1007/978-3-319-47058-0_5

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Extrinsic projection of Itô SDEs on submanifolds with applications to non-linear filtering

John Armstrong¹ and Damiano Brigo²

¹ King's College London, Strand, London, WC2R 2LS. john.1.armstrong@kcl.ac.uk

² Imperial College, London, SW7 2AZ. damiano.brigo@imperial.ac.uk

Abstract. We define the notion of the extrinsic Itô projection of a stochastic differential equation (SDE) on a submanifold. This allows one to systematically develop low dimensional approximations to high dimensional SDEs in a differential geometric setting. We consider the example of approximating the non-linear filtering problem with a Gaussian distribution and show how the Itô projection leads to improved approximations in the Gaussian family. We briefly discuss the approximations for more general families of distribution. We perform a numerical comparison of our projection filters with the classical Extended Kalman Filter to demonstrate the efficacy of the approach.

Keywords: stochastic differential equations, manifolds, SDEs on manifolds, stochastic differential geometry, Itô calculus, Stratonovich calculus, Itô projection, Stratonovich projection.

AMS classification codes: 58J65, 60H10, 60J60

1 Introduction

In this paper we consider two notions of projecting a stochastic differential equation (SDE) onto a manifold M . We will call the two approaches Stratonovich projection and the extrinsic Itô projection.

The purpose of these projection methods is to transform an infinite dimensional SDE to a finite dimensional SDE which can then be solved numerically. We will benchmark the performance of these two competing projection techniques using a non-linear filtering problem. We will also compare the performance of our approach to comparable established approaches to non-linear filtering.

To explain the idea, let us first consider projecting an ordinary differential equation (ODE) onto a manifold $M \subseteq \mathbb{R}^r$. An ODE can be thought of as defining a vector field in $\mathbb{R}^r$. At every point $x \in M$ we can use the Euclidean metric to project the vector at x onto the tangent space $T_x M$. In this way one obtains a vector field on M which can be thought of as a new ODE on M that approximates the full ODE in $\mathbb{R}^r$.

We now wish to consider projecting stochastic differential equations onto a manifold. One possible answer has been proposed previously which we shall call

the Stratonovich projection. The Stratonovich projection is obtained by simply applying the projection operator to the coefficients of the SDE written in Fisk-Stratonovich-McShane [9, 18, 15] (Stratonovich from now on) calculus form. No optimality result has been derived for the Stratonovich projection, it has simply been derived heuristically from the deterministic case. Nevertheless, it appears to be a good approximation in practice and it has been used to find good quality numerical solutions to the non-linear filtering problem (See [8], [7], [3], [4]).

It is obvious to anyone with experience of stochastic differential equations on manifolds that simply applying the projection operator to the coefficients of the SDE written in Itô form will not work. This is because solutions to the projected equation don't stay on the manifold. Nevertheless we will be able to obtain a modification of this idea, which we will call the extrinsic Itô projection, which does give a well defined SDE on the manifold. We will show elsewhere how this extrinsic Itô projection can be derived from an optimality argument and so this new projection is in some sense an optimal approximation of the original SDE on the manifold. The extrinsic Itô projection is described in Section 2. We prove directly that it is a well-defined stochastic differential equation. For the benefit of the reader, we include a brief review of stochastic differential equations on manifolds in Subsection 2.1.

Having defined the extrinsic Itô projection, we can apply it to find approximate solutions to difficult stochastic differential equations. In particular we will apply it to the non-linear filtering problem. This application is discussed in Section 3. We will derive general projection formulae for the non-linear filtering problem. We will then apply this to the problem of approximating a non-linear filter using a Gaussian distribution. A reader who is unfamiliar with non-linear filtering will want to consult Subsection 3.1 for a brief review.

Gaussian approximations to non-linear filters are widely used in practice [5]. In particular the Extended Kalman Filter is a popular approximation technique. Other Gaussian approximations exist such as Assumed Density Filters and filters derived from the Stratonovich projection. Our theory indicates that all these classical techniques can be improved upon by using the extrinsic Itô projection (at least over small time intervals). We confirm this with a numerical example.

The utility of the projection method is by no means restricted to the filtering problem nor to such simple approximations as Gaussian filters. Our previous work shows how the Stratonovich projection can be used to generate far more sophisticated filters and it is clear that the idea of projection should be widely applicable in the study of SDEs and ODEs. Nevertheless by focussing on Gaussian filters we can examine in detail the idea that there may be many useful ways of approximating an SDE on a submanifold and examine in detail the relative performance of the extrinsic Itô projection. The point we wish to emphasize is that the extrinsic Itô projection is able to tell us something new even about the well-worn topic of approximating the non-linear filtering problem using Gaussian distributions.

Note that the development of the extrinsic Itô projection does not invalidate previous work using the Stratonovich projection, it merely indicates that

alternative approximations are possible. In a future paper we will consider in what sense the extrinsic Itô projection is an optimal approximation over a small time horizon. As we will show in that paper, the notion of optimality for a finite dimensional approximation to an SDE is far more subtle than the comparable notion for ODEs. This theoretical analysis will explain why the extrinsic Itô projection gives the excellent results demonstrated numerically in this paper, but will also provide an explanation for why there are occasions when the Stratonovich projection is still a superior approach.

2 Projecting stochastic differential equations

2.1 Itô SDEs on manifolds

It is well known that one can write SDEs on manifolds in Stratonovich form. However, in our experience there seems to be some confusion about whether one can, or should, write SDEs on manifolds in Itô form. Itô himself in [11] defined the notion of an SDE on a manifold using Itô calculus. Nevertheless we believe it may be useful to the non-expert if we explicitly define an Itô SDE on a manifold and explain the motivation behind the definition.

Given a n -dimensional manifold M , we can write down stochastic differential equations in a neighbourhood U of a point x by choosing a chart $\phi : U \rightarrow \mathbb{R}^n$ and then writing the stochastic differential equation in local coordinates. The equation written in local coordinates will depend upon the choice of chart ϕ . Thus the data for a stochastic differential equation locally consists of:

1. A vector valued Brownian motion W_t (the theory can also be extended to continuous semi-martingale integrators, but we will use Brownian motion for simplicity)
2. A chart ϕ
3. The coefficient functions a, b of a stochastic differential equation written in local coordinates:

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t.$$

To define a stochastic differential equation over the entire manifold we will need local data of this form for a complete atlas of charts. Where charts overlap, we will need some compatibility conditions on these local SDEs. The “correct” compatibility condition should be chosen so that the solutions of the SDE in one chart ϕ are mapped to the solutions of the SDE in another chart, Φ by the transition function $\tau = \Phi \circ \phi^{-1}$. Since this requirement is expressed in terms of the *solutions* to a stochastic differential equation it is a mathematically complex requirement. We would prefer to write the requirement in terms of the much simpler data of the *coefficients* of our stochastic differential equation. We can informally calculate the correct compatibility condition on the coefficients using Itô’s lemma.

Let (W_t, ϕ, a, b) be the data for the SDE in one chart, and (W_t, Φ, A, B) be the data in another chart. We will suppose that W_t takes values in $\mathbb{R}^m$ and will

write W^α for the components of W_t . Similarly, X^i , a^i , b_α^i are the components of the vectors X , a and the tensor b . We have chosen to label indices such that Roman indices run from 1 through to n (the dimension of our manifold) and Greek indices run from 1 through to m (the dimension of the process W_t).

We can now write out the SDE (3) in full detail in local coordinates as:

$$dX_t^i = a^i(X_t, t)dt + \sum_{\alpha} b_{\alpha}^i(X_t, t)dW_t^{\alpha}.$$

We will write τ^i for the components of τ and will use coordinates x^i for $\mathbb{R}^n$. With this notation in place we can apply Itô's lemma to write out an SDE for $\tau(X_t)$ as follows:

$$\begin{aligned} d\tau^i(X_t) = & \left(\sum_j \frac{\partial \tau^i}{\partial x^j} a^j(X_t, t) + \sum_{j,k,\alpha,\beta} \frac{1}{2} \frac{\partial^2 \tau^i}{\partial x^j \partial x^k} b_{\alpha}^j(X_t, t) b_{\beta}^k(X_t, t) g^{\alpha\beta} \right) dt \\ & + \left(\sum_{j,\alpha} \frac{\partial \tau^i}{\partial x^j} b_{\alpha}^j(X_t, t) \right) dW_t^{\alpha}. \end{aligned}$$

Here $g^{\alpha\beta} = [W^{\alpha}, W^{\beta}]_t$ denotes the quadratic covariation of W^{α} and W^{β} . We use the letter g because this term defines a metric on $\mathbb{R}^m$. Our local coordinate notation is very cumbersome, so we will adopt various conventions taken from differential geometry:

- (i) The Einstein summation convention: if an index appears as both an upper index and lower index, one should sum over that index.
- (ii) We write ∂_i as an abbreviation for $\frac{\partial}{\partial x^i}$.
- (iii) We drop parameters to coefficient functions when it is clear from the context what the parameter values should be. For example, one might abbreviate $a(X_t, t)$ to simply a . Similarly one might write X instead of X_t .

With these conventions in place we can rewrite our SDE for $\tau(X)$ as:

$$d\tau^i(X) = \left(a^j \partial_j \tau^i + \frac{1}{2} b_{\alpha}^j b_{\beta}^k g^{\alpha\beta} \partial_j \partial_k \tau^i \right) dt + b_{\alpha}^j \partial_j \tau^i dW_t^{\alpha}.$$

Note that all the terms on the right are evaluated at X . With all these preliminaries we can now formally define what we mean by a Stochastic differential equation on a manifold M .

Definition 1. *An Itô SDE on a manifold consists of the initial conditions together with an equivalence class of quadruples (W_t, ϕ, a, b) under the equivalence relation $\sim$ defined by*

$$(W_t, \phi, a, b) \sim (V_t, \Phi, A, B) \quad \text{if} \quad \begin{cases} W_t = V_t \\ A^j = a^j \partial_j \tau^i + \frac{1}{2} b_{\alpha}^j b_{\beta}^k g^{\alpha\beta} \partial_j \partial_k \tau^i \\ B^j = b_{\alpha}^j \partial_j \tau^i \end{cases}$$

for the transition function $\tau = \Phi \circ \phi^{-1}$.

The first condition could be written in more general terms, but for simplicity we assume pathwise equality between the two Brownian motions.

Rather less formally, one might say that an Itô SDE is an SDE whose coefficients obey Itô's lemma when one changes coordinates.

The definition we have chosen for an Itô SDE is exactly analogous to the common definition of a vector field on a manifold as a set of coordinate functions that transform in a particular way when one changes coordinates.

As is well known, when writing SDEs on $\mathbb{R}^n$ one can choose to write the equation in either Itô or Stratonovich form. We could attempt to define SDEs on manifolds using Stratonovich equations. In this case the local data would be a pair $(W_t, \phi, \bar{a}, b)$ where W_t and ϕ are as before but now $\bar{a}, b$ are the coefficients of the Stratonovich equation:

$$dX_t = \bar{a} dt + b \circ dW_t.$$

Rather than use Itô's lemma in our informal derivation, we would now use the chain rule. The end result is the following definition:

Definition 2. *A Stratonovich SDE on a manifold consists of the initial conditions together with an equivalence class of quadruples $(W_t, \phi, \bar{a}, b)$ under the equivalence relation $\sim$ defined by*

$$(W_t, \phi, \bar{a}, b) \sim (V_t, \Phi, \bar{A}, B) \quad \text{if} \quad \begin{cases} W_t = V_t \\ \bar{A}^j = \bar{a}^j \partial_j \tau^i \\ B^j = b_\alpha^j \partial_j \tau^i \end{cases} . \quad (1)$$

As we know, if the coefficients of the SDEs are smooth enough, Stratonovich SDEs and Itô SDEs on $\mathbb{R}^n$ are essentially equivalent, i.e. an Itô SDE can be transformed to an SDE in Stratonovich form which has the same solutions and vice versa. One sees immediately that Itô SDEs and Stratonovich SDEs on manifolds are essentially equivalent in precisely the same sense.

Since the chain rule is rather simpler than Itô's lemma, the definition of a Stratonovich SDE is rather simpler than that of an Itô SDE. In addition, we can easily replace the complicated index notation with coordinate free notation. We will write $T_x X$ to denote the tangent space at a point x of a manifold X . We will write $f_* : T_x X \rightarrow T_{f(x)} Y$ to denote the differential of a smooth map between manifolds $f : X \rightarrow Y$. We can then rewrite Eq. (1) more elegantly as:

$$\begin{cases} W_t = V_t, \\ \bar{A} = \tau_*(\bar{a}), \\ \bar{B} = \tau_*(b). \end{cases}$$

Thus the transformation rule for the coefficients in this case is precisely the same as the transformation rule for the coefficients of a vector field. This allows one to devise alternative definitions for stochastic differential equations in terms of vector fields without needing to mention the less attractive details about equivalence relations.

2.2 Projecting SDEs

Let M be a submanifold of $\mathbb{R}^r$ with chart $\psi : U \rightarrow \mathbb{R}^n$ for some open neighbourhood in M and inverse $\phi = \psi^{-1}$.

Given an SDE defined on $\mathbb{R}^r$, we would like to approximate solutions in $\mathbb{R}^r$ with solutions to an SDE defined on M .

Definition 3. Let W_t be an $\mathbb{R}^m$ valued Brownian motion. Given a Stratonovich SDE on $\mathbb{R}^r$

$$dX = \bar{a} dt + b_\alpha \circ dW_t^\alpha$$

and a chart $\psi : U \rightarrow \mathbb{R}^n$ for some neighbourhood in $N \subseteq M$ we define the Stratonovich projection of the SDE to be:

$$d\tilde{X} = \bar{A} dt + B_\alpha \circ dW_t^\alpha$$

where:

$$\begin{aligned}\phi &= \psi^{-1} \\ \bar{A}(\tilde{X}, t) &= (\psi_*) \Pi_{\phi(\tilde{X})}(\bar{a}(\phi(\tilde{X}), t)) \\ B_\alpha(\tilde{X}, t) &= (\psi_*) \Pi_{\phi(\tilde{X})}(b_\alpha(\phi(\tilde{X}), t))\end{aligned}$$

where Π is the projection of $\mathbb{R}^r$ onto $\phi_*(\mathbb{R}^n)$ defined by the Euclidean metric.

Because we know that the projection of vector fields can be defined similarly, and because we know that the coefficients of Stratonovich SDEs transform like vector fields, we see that the definition above defines a Stratonovich SDE on M . Indeed, if one is willing to accept that projection of vector fields onto a submanifold is well-defined, then one could define the projection of a Stratonovich SDE as the projection of the coefficient functions.

For an Itô SDE one cannot simply apply projection to the coefficient functions because the coefficients of an Itô SDE on a manifold do not transform like vector fields.

The *Stratonovich projection* of an Itô SDE is trivially defined by the recipe:

- (i) Rewrite the Itô SDE as a Stratonovich SDE
- (ii) Apply the Stratonovich projection as defined above
- (iii) Rewrite the resulting Stratonovich SDE as an Itô SDE.

We see that the use of Stratonovich calculus is just limited to the projection procedure. After and before that, we wish to use Itô calculus because of the good probabilistic properties of the Itô integral. However, we can try and avoid transiting through Stratonovich calculus and define the *extrinsic Itô projection* as follows:

Definition 4. Let W_t be an $\mathbb{R}^m$ valued Brownian motion. Given an Itô SDE on $\mathbb{R}^r$

$$dX = a dt + b_\alpha dW_t^\alpha$$

and a chart $\psi : U \rightarrow \mathbb{R}^n$ for some neighbourhood in $N \subseteq M$ we define the extrinsic Itô projection of the SDE to be:

$$dY = A dt + B_\alpha dW_t^\alpha$$

where:

$$\begin{aligned} \phi &:= \psi^{-1} \\ B_\alpha(Y_t, t) &:= (\psi_*)_{\phi(Y_t)} \Pi_{\phi(Y_t)} b_\alpha(\phi(Y_t), t) \\ A(Y_t, t) &:= (\psi_*)_{\phi(Y_t)} \Pi_{\phi(Y_t)} \left(a(\phi(Y_t), t) - \frac{1}{2} (\nabla_{B_\alpha(\phi(Y_t), t)} \phi_*) B_\beta(\phi(Y_t), t) g^{\alpha\beta} \right) \end{aligned}$$

where ∇ is the gradient operator defined on $\mathbb{R}^n$.

We will discuss the motivation for this definition in detail in a future paper. For now we will simply remark that it can be derived by searching for the optimal approximation over a small time horizon in the metric defined on $\mathbb{R}^n$. We call it the extrinsic Itô projection because the optimality is defined via the use of the metric on the extrinsic space $\mathbb{R}^n$ rather than using the Riemannian metric of the sub-manifold. As we will show in subsequent papers [2], there is an alternative notion of the intrinsic Itô projection which one may consider. Defining the intrinsic Itô projection is best done using the differential geometric language of 2-jets. Introducing this machinery now would take us too far afield, which is why we have given only this brief motivation for the definition. We will show in our future work that the notion of “optimality” is far more subtle for SDEs than for ODEs. For example, there are occasions where the Stratonovich projection actually out-performs the extrinsic Itô projection.

The Stratonovich projection is manifestly well-defined. We must work harder for the extrinsic Itô projection.

Theorem 1. *The extrinsic Itô projection defines an SDE.*

Proof. Let M be a submanifold of $\mathbb{R}^r$ and let $x_1 : U \rightarrow \mathbb{R}^k$ and $x_2 : U \rightarrow \mathbb{R}^k$ be two charts for some open set U containing a point X_0 . Let $\tau = x_1 \circ x_2^{-1}$ be the transition function between the charts.

Write ϕ_i for the inverse of x_i . Write $(P_i)_{x_i} = (x_i)_* \Pi_{\phi_i(x_i)}$ for the projection map associated with the chart x_i . Note that $P_1 = \tau_* P_2$. If the SDE in $\mathbb{R}^n$ has Itô coefficients a and b then the Itô projected SDE w.r.t the coordinates x_i is:

$$\begin{aligned} dY &= \left[P_i a(\phi_i(Y), t) - \frac{1}{2} P_i ((\nabla_{P_i b(\phi_i(Y))} (\phi_i)_*) (P_i b(\phi_i(Y)))) \right] dt \\ &\quad + (P_i b(\phi_i(Y))) dW, \\ Y_0 &= x_i(X_0). \end{aligned} \tag{2}$$

What we want to show is that equation (2) for x_2 transformed using τ gives the equation (2) for x_1 . With this in mind we transform equation (2) for x_2 using τ

to obtain an equation for $Z = \tau(Y)$:

$$\begin{aligned} dZ &= \left[\tau_* P_2 a(\phi_1(Z), t) - \frac{1}{2} \tau_* P_2 ((\nabla_{\tau_* P_2 b(\phi_1(Z), t)} (\phi_2)_*) P_2 b(\phi_1(Z), t)) \right. \\ &\quad \left. + \frac{1}{2} (\nabla_{\tau_* P_2 b(\phi_1(Z), t)} \tau_*) (P_2 b(\phi_1(Z), t)) \right] dt \\ &\quad + \tau_* P_2 b(\phi_1(Z), t) dW, \\ Z_0 &= x_1(X_0). \end{aligned} \tag{3}$$

We now simplify this using the following identities:

$$\begin{aligned} P_1 &= \tau_* P_2, \\ \phi_2 &= \phi_1 \circ \tau, \\ (\phi_2)_* &= (\phi_1)_* \circ \tau_*. \end{aligned}$$

This last identity is the chain rule. So by the product rule we have:

$$\nabla_Y (\phi_2)_* = (\nabla_Y (\phi_1)_*) \circ \tau_* + (\phi_1)_* \circ (\nabla_Y \tau_*) \tag{4}$$

for vectors Y . This allows us to rewrite (3) as follows:

$$\begin{aligned} dZ &= [(P_1 a(x)) \\ &\quad - \frac{1}{2} P_1 ((\nabla_{P_1 b(\phi_1(Z), t)} \phi_*^1) (\tau_* P_2 b(\phi_1(Z), t))) \\ &\quad - \frac{1}{2} P_1 (\phi_*^1 ((\nabla_{P_1 b(\phi_1(Z), t)} \tau_*) (P_2 b(\phi_1(Z), t)))) \\ &\quad + \frac{1}{2} (\nabla_{P_1 b(\phi_1(Z), t)} \tau_*) (P_2 b(\phi_1(Z), t))] dt \\ &\quad + P_1 b(\phi_1(Z), t) dW, \\ Z_0 &= x_1(X_0). \end{aligned}$$

We can now use the fact that $P_1(\phi_*^1)$ is the identity and again use the identity $P_1 = \tau_* P_2$ to simplify this. Two unwanted terms cancel leaving us with equation:

$$\begin{aligned} dZ &= [(P_1 a(\phi_1(Z), t)) - \frac{1}{2} P_1 ((\nabla_{P_1 b(\phi_1(Z), t)} \phi_*^1) (\tau_* P_1 b(\phi_1(Z), t)))] dt \\ &\quad + P_1 b(\phi_1(Z), t) dW, \\ Z_0 &= x_1(X_0). \end{aligned}$$

This is (2) for x_1 as claimed.

We also want to show that the extrinsic Itô projection is distinct from the Stratonovich projection. We will show this in the next sections by explicitly computing examples. Moreover we will demonstrate numerically that the extrinsic Itô projection gives superior results to the Stratonovich projection when applied to non-linear filtering.

3 Application of the Projection to Non-linear Filtering

3.1 The Kushner Stratonovich equation

We suppose that the state $X_t \in \mathbb{R}^n$ of a system evolves according to the equation:

$$dX_t = f(X_t, t) dt + \sigma(X_t, t) dW_t$$

where f and σ are smooth $\mathbb{R}^n$ valued functions and W_t is a Brownian motion.

We suppose that an associated process, the observation process, $Y_t \in \mathbb{R}^d$ evolves according to the equation:

$$dY_t = b(X_t, t) dt + dV_t$$

where b is a smooth $\mathbb{R}^d$ valued function and V_t is a Brownian motion independent of W_t . Note that the filtering problem is often formulated with an additional constant in terms of the observation noise. For simplicity we have assumed that the system is scaled so that this can be omitted.

The filtering problem is to compute the conditional distribution of X_t given a prior distribution for X_0 and the values of Y for all times up to and including t .

Subject to various bounds on the growth of the coefficients of this equation, the assumption that the distribution has a density p_t and suitable bounds on the growth of p_t one can show that p_t satisfies the Kushner–Stratonovich equation:

$$dp = \mathcal{L}^* p dt + p[b - E_p(b)]^T [dY - E_p(b)dt] \quad (5)$$

where E_p denotes the expectation with respect to the density p and the forward diffusion operator $\mathcal{L}_t^*$ is defined by:

$$\mathcal{L}_t^* \phi = - \sum_{i=1}^n \frac{\partial}{\partial x_i} [f_t^i \phi] + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} [a^{ij} \phi] \quad (6)$$

where $a = \sigma \sigma^T$.

In the event that the coefficient functions f and b are all linear and σ is a deterministic function of time one can show that so long as the prior distribution for X is Gaussian, or deterministic, the density p will be Gaussian at all subsequent times. This allows one to reduce the infinite dimensional equation (5) to a finite dimensional stochastic differential equation for the mean and covariance matrix of this normal distribution. This finite dimensional problem is known as the Kalman filter.

For more general coefficient functions, however, equation (5) cannot be reduced to a finite dimensional problem [10]. Instead one might seek approximate solutions of (5) that belong to some given statistical family of densities. This is a very general setup and includes, for example, approximating the density using piecewise linear functions to derive a finite difference approximation or approximating the density with Hermite polynomials to derive a spectral method. Other examples include exponential families (considered in [7, 8]) and mixture families (considered in [3, 4]).

Our projection theory tells us how one can find good approximations on a given statistical family with respect to a given metric on the space of distributions. We illustrate this by writing down the extrinsic Itô projection of (5) for the L^2 and Hellinger metrics onto a general manifold³.

We will then examine some numerical results regarding the very specific case of seeking approximate solutions using Gaussian distributions. The idea of approximating the solution to the filtering problem using a Gaussian distribution has been considered by numerous authors who have derived variously, the extended Kalman filter [16], assumed density filters [13] and Stratonovich projection filters [8]. We will be able to derive extrinsic Itô projection filters which outperform all these other filters (assuming performance is measured over small time intervals using the appropriate Hilbert space metric).

3.2 Itô projections

The extrinsic Itô projection filter in the L^2 direct metric Let us suppose that the density p lies in L^2 and so we can use the L^2 norm to measure the accuracy of an approximate solution to equation (5). For a discussion on conditions under which a unnormalized version of p is in L^2 (Zakai Equation) see for example [1].

We wish to consider an m -dimensional family of distributions p parameterized by m real valued parameters $\theta^1, \theta^2, \dots, \theta^m$. For example we will consider the 2 dimensional Gaussian family:

$$p(x) = \frac{1}{(\theta^2)\sqrt{2\pi}} \exp\left(-\frac{(x - (\theta^1))^2}{2(\theta^2)^2}\right). \quad (7)$$

Note that we have chosen to follow differential geometry convention and use upper indices for the coordinate functions θ^i so we have been careful to distinguish powers from indices using brackets.

More formally, an m -dimensional family is given by a smooth embedding $\phi : \mathbb{R}^m \rightarrow L^2(\mathbb{R}^n)$. The tangent vectors $\phi_* \frac{\partial}{\partial \theta^i} \in L^2(\mathbb{R}^n)$ are simply the partial derivatives

$$\frac{\partial p}{\partial \theta^i}.$$

Let us write:

$$g_{ij} = \int_{\mathbb{R}} \frac{\partial p}{\partial \theta^i} \frac{\partial p}{\partial \theta^j} dx.$$

³ Note that it is also possible to consider projecting the Zakai equation. However, as explained in [4], one expects that projecting the Kushner–Stratonovich will lead to smaller error terms

This defines the induced metric tensor on the manifold $\phi(\mathbb{R}^m)$. We will write g^{ij} for the inverse of the matrix g_{ij} . The projection operator $\Pi_{\phi(\theta)}$ is then given by

$$\begin{aligned}\Pi_{\phi(\theta)}(v) &= \sum_{i,j=1}^m g^{ij} \left\langle v, \phi_* \frac{\partial}{\partial \theta^i} \right\rangle_{L^2} \phi_* \frac{\partial}{\partial \theta^j} \\ &= \sum_{i,j=1}^m g^{ij} \left(\int_{\mathbb{R}^n} v(x) \frac{\partial p}{\partial \theta^i} dx \right) \phi_* \frac{\partial}{\partial \theta^j}.\end{aligned}$$

Thus

$$\phi_*^{-1} \Pi_{\phi(\theta)}(v) = \sum_{i,j=1}^m g^{ij} \left(\int_{\mathbb{R}^n} v(x) \frac{\partial p}{\partial \theta^i} dx \right) \frac{\partial}{\partial \theta^j}.$$

We can now write down the extrinsic Itô projection of (5) with respect to the L^2 metric. It is:

$$d\theta^i = A^i dt + B^i dY_t$$

where:

$$B^i = \sum_{j=1}^m g^{ij} \left(\int_{\mathbb{R}} (p(b - E_{p(\theta)}(b)))^T \frac{\partial p}{\partial \theta^j} dx \right)$$

and

$$A^i = \sum_{j=1}^m g^{ij} \left(\int_{\mathbb{R}^n} \left(\mathcal{L}^* p - p(b - E_{p(\theta)}(b))^T E_{p(\theta)}(b) - \frac{1}{2} \sum_{k=1}^m \frac{\partial^2 p}{\partial \theta^j \partial \theta^k} B^k \right) \frac{\partial p}{\partial \theta^j} dx \right).$$

Example 1. Consider as a test case the 1-dimensional problem with $f(x, t) = 0$, $\sigma(x, t) = 1$ and $b(x, t) = x + \epsilon x^3$ for some small constant ϵ . This problem is a perturbation of a linear filter so one might expect that a Gaussian approximation will perform reasonably well at least for small times. Thus we will use the 2 dimensional manifold of Gaussian distributions given in equation (7).

We first calculate the metric tensor g_{ij} which is diagonal in this case:

$$g_{ij} = \frac{1}{4\sqrt{\pi}(\theta^2)^3} \begin{pmatrix} 1 & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$

This is easily inverted to compute g^{ij} . We compute the expectation $E_p(b)$:

$$E_p(b) = \frac{\epsilon (\sqrt{2\pi}(\theta^1)^3(\theta^2) + 3\sqrt{2\pi}(\theta^1)(\theta^2)^3)}{\sqrt{2\pi}(\theta^2)} + (\theta^1).$$

One can now see that computing the projection equation will simply involve integrating a number of terms of the form a polynomial in x times a Gaussian.

The end result is:

$$\begin{aligned} d\theta^1 &= \left(-\frac{1}{4}\theta^1 (\theta^2)^2 \left(3\epsilon^2 \left(4(\theta^1)^4 - 4(\theta^2)^2 (\theta^1)^2 - 3(\theta^2)^4 \right) + 16\epsilon (\theta^1)^2 + 4 \right) \right) dt \\ &\quad + \left(\frac{1}{2} (\theta^2)^2 \left(3\epsilon \left(2(\theta^1)^2 + (\theta^2)^2 \right) + 2 \right) \right) dY_t, \\ d\theta^2 &= \left(-\frac{9\epsilon^2 (\theta^2)^8 + (\theta^2)^4 \left(60\epsilon^2 (\theta^1)^4 + 48\epsilon (\theta^1)^2 + 4 \right) + 6\epsilon (\theta^2)^6 \left(9\epsilon (\theta^1)^2 + 2 \right) - 4}{8\theta^2} \right) dt \\ &\quad + \left(3\epsilon\theta^1 (\theta^2)^3 \right) dY_t. \end{aligned}$$

The extrinsic Itô projection filter in the Hellinger metric The Hellinger metric is a metric on probability measures. In the case of two probability density functions $p(x)$ and $q(x)$ on $\mathbb{R}^n$, that now need only be in L^1 , the Hellinger distance is given by the square root of:

$$\frac{1}{2} \int (\sqrt{p(x)} - \sqrt{q(x)})^2 dx.$$

In other words, up to the constant factor of $\frac{1}{2}$ the Hellinger metric corresponds to the L^2 norm on the square root of the density function rather than on the density itself (as in the previous subsection). The Hellinger metric has the important advantage of making the metric independent of the particular background density that is used to express measures as densities. The L^2 direct distance introduced earlier does not satisfy this background independence.

Now, to compute the extrinsic Itô projection with respect to the Hellinger metric we first want to write down an Itô equation for the evolution on $\sqrt{p}$.

Applying Itô's lemma to equation (5) we formally obtain:

$$\begin{aligned} d\sqrt{p} &= \left(\frac{\mathcal{L}^* p - p(b - E_p(b))^T E_p(b)}{2\sqrt{p}} - \frac{p^2(b - E_p(b))^T (b - E_p(b))}{8p\sqrt{p}} \right) dt \\ &\quad + \left(\frac{p(b - E_p(b))^T}{2\sqrt{p}} \right) dY_t. \\ &= \left(\frac{\mathcal{L}^* p}{2\sqrt{p}} - \frac{1}{8}\sqrt{p}(b - E_p(b))^T (b + 3E_p(b)) \right) dt \\ &\quad + \left(\frac{1}{2}\sqrt{p}(b - E_p(b))^T \right) dY_t. \end{aligned}$$

A family of distributions now corresponds to an embedding ϕ from $\mathbb{R}^m$ to $L^2(\mathbb{R}^n)$ but now $p = \phi(\theta)^2$. The tangent space is spanned by the vectors:

$$\phi_* \frac{\partial}{\partial \theta^i} = \frac{\partial \sqrt{p}}{\partial \theta^i}.$$

We define a metric on the tangent space by:

$$h_{ij} = \int_{\mathbb{R}^n} \frac{\partial \sqrt{p}}{\partial \theta^i} \frac{\partial \sqrt{p}}{\partial \theta^j} dx.$$

We write h^{ij} for the inverse matrix of h_{ij} . The projection operator with respect to the Hellinger metric is:

$$\Pi_{\phi(\theta)}(v) = \sum_{i,j=1}^m h^{ij} \left(\int_{\mathbb{R}^n} v(x) \frac{\partial \sqrt{p}}{\partial \theta^i} dx \right) \phi_* \frac{\partial}{\partial \theta^j}.$$

We can now write down the extrinsic Itô projection of (5) with respect to the Hellinger metric. It is:

$$d\theta^i = A^i dt + B^i dY_t$$

where:

$$B^i = \sum_{j=1}^m h^{ij} \left(\int_{\mathbb{R}} \frac{1}{2} \sqrt{p}(b - E_{p(\theta)}(b))^T \frac{\partial \sqrt{p}}{\partial \theta^j} dx. \right)$$

and

$$A^i = \sum_{j=1}^m h^{ij} \left(\int_{\mathbb{R}^n} \left(\frac{\mathcal{L}^* p}{2\sqrt{p}} - \frac{1}{8} \sqrt{p}(b - E_{p(\theta)}(b))^T (b + 3E_{p(\theta)}(b)) \right. \right. \\ \left. \left. - \frac{1}{2} \sum_{k=1}^m \frac{\partial^2 \sqrt{p}}{\partial \theta^j \partial \theta^k} B^k \right) \frac{\partial \sqrt{p}}{\partial \theta^j} dx. \right).$$

Example 2. We may repeat example 1 but projecting using the Hellinger metric. We first calculate the metric tensor h_{ij} which is diagonal also in this case:

$$h_{ij} = \frac{1}{4\theta_2^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

This is easily inverted to compute h^{ij} . We obtain the following SDEs:

$$d\theta^1 = \left(-\theta^1 (\theta^2)^2 \left(3\epsilon^2 \left((\theta^1)^4 + 4(\theta^2)^2 (\theta^1)^2 + 6(\theta^2)^4 \right) + \epsilon \left(4(\theta^1)^2 + 6(\theta^2)^2 + 1 \right) \right) dt \right. \\ \left. + \left((\theta^2)^2 \left(3\epsilon \left((\theta^1)^2 + (\theta^2)^2 \right) + 1 \right) \right) dY_t, \right. \\ d\theta^2 = \left(-\frac{27\epsilon^2 (\theta^2)^8 + (\theta^2)^4 \left(15\epsilon^2 (\theta^1)^4 + 12\epsilon (\theta^1)^2 + 1 \right) + 9\epsilon (\theta^2)^6 \left(6\epsilon (\theta^1)^2 + 1 \right) - 1}{2\theta^2} \right) dt \\ \left. + \left(3\epsilon \theta^1 (\theta^2)^3 \right) dY_t. \right.$$

Note that this satisfies the important check that when ϵ is zero, this reduces to the Kalman filter.

3.3 Other Gaussian Approximate Filters

We will show in a subsequent paper that the filters above are in some sense optimal with respect to the relevant Hilbert space metric. Nevertheless many other Gaussian approximate filters have been proposed in the past. We will briefly review a number of different Gaussian approximate filters that can be found in the literature and calculate the relevant stochastic differential equations for our example 1. We will then compare the performance of these filters numerically.

The Stratonovich projection filter Instead of using the extrinsic Itô projection, one can use the Stratonovich projection.

Example 3. General formulae for performing the Stratonovich L^2 projection are given in [3]. In the specific case of example 1 the resulting Itô SDEs are:

$$\begin{aligned} d\theta^1 &= \left(-\frac{1}{4}\theta^1(\theta^2)^2 \left(3\epsilon^2 \left(4(\theta^1)^4 - 4(\theta^2)^2(\theta^1)^2 - 3(\theta^2)^4 \right) + 16\epsilon(\theta^1)^2 + 4 \right) \right) dt \\ &\quad + \left(\frac{1}{2}(\theta^2)^2 \left(3\epsilon \left(2(\theta^1)^2 + (\theta^2)^2 \right) + 2 \right) \right) dY_t, \\ d\theta^2 &= \left(-\frac{47\epsilon^2(\theta^2)^8 + (\theta^2)^4(60\epsilon^2(\theta^1)^4 + 48\epsilon(\theta^1)^2 + 4) + 2\epsilon(\theta^2)^6(33\epsilon(\theta^1)^2 + 8) - 4}{8\theta^2} \right) dt \\ &\quad + \left(3\epsilon\theta^1(\theta^2)^3 \right) dY_t. \end{aligned}$$

Example 4. General formulae for performing the Stratonovich Hellinger projection are given in [7]. In the specific case of example 1 the resulting SDEs are:

$$\begin{aligned} d\theta^1 &= \left(-\theta^1(\theta^2)^2 \left(3\epsilon^2 \left((\theta^1)^4 + 4(\theta^2)^2(\theta^1)^2 + 6(\theta^2)^4 \right) + \epsilon \left(4(\theta^1)^2 + 6(\theta^2)^2 + 1 \right) \right) \right) dt \\ &\quad + \left((\theta^2)^2 \left(3\epsilon \left((\theta^1)^2 + (\theta^2)^2 \right) + 1 \right) \right) dY_t, \\ d\theta^2 &= \left(-\frac{36\epsilon^2(\theta^2)^8 + (\theta^2)^4(15\epsilon^2(\theta^1)^4 + 12\epsilon(\theta^1)^2 + 1) + 9\epsilon(\theta^2)^6(6\epsilon(\theta^1)^2 + 1) - 1}{2\theta^2} \right) dt \\ &\quad + \left(3\epsilon\theta^1(\theta^2)^3 \right) dY_t. \end{aligned}$$

The Extended Kalman Filter The Extended Kalman Filter (EKF) is a heuristically derived method of finding approximate solutions to the filtering problem based on the idea of linearising the problem and then using the solution to the linear problem. In particular one assumes that the solution can be well approximated by a Gaussian distribution. For the EKF see [12, 1]. A definition and heuristic derivation is given in [5] (which is based, in turn, on the derivation given in [16]).

The EKF can be shown to work well on condition that the initial position of the signal is approximated well, the non-linearities of f are small, b is injective and the observation noise is small [17]. Moreover, the EKF is widely used in practice, see [5] for references to applications.

Example 5. For the example problem $b(x) = x + \epsilon x^3$ the EKF is:

$$\begin{aligned} d\theta^1 &= \left((\theta^2)^2 \left(-\left(3\epsilon(\theta^1)^2 + 1 \right) \right) \left(\theta^1 + \epsilon(\theta^1)^3 \right) \right) dt \\ &\quad + \left((\theta^2)^2 \left(3\epsilon(\theta^1)^2 + 1 \right) \right) dY_t, \\ d\theta^2 &= \left(\frac{1 - (\theta^2)^4 \left(3\epsilon(\theta^1)^2 + 1 \right)^2}{2\theta^2} \right) dt. \end{aligned}$$

Assumed density filters Assumed density filters (ADFs) provide a finite dimensional method of finding approximate solutions to the filtering problem. They have been considered in, for example, [13], [14] and [8].

The general setup is to consider a statistical family $\pi(\cdot, \eta)$ of probability measures parameterized by some coordinates $\eta = (\eta^1, \dots, \eta^m)$. This parameterization is not arbitrary. It must be chosen in such a way that, for elements of the statistical family, the values of η correspond to the expectations of some twice differentiable scalar functions $\{c^1, \dots, c^m\}$ defined on $\mathbb{R}^n$.

$$\eta^i = E_{\pi(\cdot, \eta)}(c^i) =: E_{\eta_t}(c^i)$$

where for brevity we are using the abbreviation E_{η_t} for $E_{\pi(\cdot, \eta)}$.

For example one might take the statistical family of normal distributions parameterized by its first and second moments η_1 and η_2 , so $c^1(x) = x$, $c^2(x) = (x)^2$.

Given a statistical family parameterized in this way, we define the Itô ADF to be:

$$d\eta_t^i = E_{\eta_t}(\mathcal{L}_t c^i) dt + (E_{\eta_t}(b_t c^i) - E_{\eta_t}(b_t) \eta_t^i)^T (dY_t - E_{\eta_t}(b_t) dt).$$

This is motivated by the fact that under the conditions used to derive equation (5), we have that the c_i -moments of π_t , the true solution to the filtering problem, satisfy the Itô equation:

$$d\pi_t(c_i) = \pi_t(\mathcal{L}_t c^i) dt - \frac{1}{2}(\pi_t(b_t^2 c^i) - \pi_t(b_t)^2 \pi_t(c^i))(dY - \pi_t(b_t) dt).$$

Thus if it were true that the true density was a member of our chosen statistical family then the Itô ADF would certainly be satisfied. One just hopes that the Itô ADF will continue to give a reasonable approximation even though we know that the true density isn't a member of the chosen statistical family.

With a similar motivation we define the Stratonovich ADF to be:

$$\begin{aligned} d\eta_t^i &= E_{\eta_t}(\mathcal{L}_t c^i) dt - \frac{1}{2}(E_{\eta_t}(|b_t|^2 c^i) - E_{\eta_t}(|b_t|^2) \eta_t^i) dt \\ &\quad + (E_{\eta_t}(b_t c^i) - E_{\eta_t}(b_t) \eta_t^i)^T \circ dY_t. \end{aligned}$$

If it were true that the density was a member of our statistical family then the Itô ADF and the Stratonovich ADF would be equivalent equations. Since we only expect to be able to approximate the true density with our statistical family, we must expect that the Itô ADF and Stratonovich ADF are in fact inequivalent equations. Intuitively, we can say that the local moment matching approximation on which the ADF heuristics are based and the Ito-Stratonovich transformation do not commute.

The justification just given for ADFs is far from convincing. We are relying on little other than hope that these equations will give good approximations. However, it was shown in [8] that in fact for exponential families, the Stratonovich projection filter in the Hellinger metric coincides with the Stratonovich ADF, and in [6] that for the Gaussian case, this filter approaches the optimal filter under small observation noise.

Example 6. If we calculate the Itô assumed density filter corresponding to example 1 and the family of normal distributions, and then change coordinates to θ^1 and θ^2 as used in the previous examples, we obtain the SDEs:

$$\begin{aligned} d\theta^1 &= \left(-\theta^1 (\theta^2)^2 \left(3\epsilon^2 \left((\theta^1)^4 + 4(\theta^2)^2 (\theta^1)^2 + 3(\theta^2)^4 \right) + \epsilon \left(4(\theta^1)^2 + 6(\theta^2)^2 \right) + 1 \right) \right. \\ &\quad \left. + \left((\theta^2)^2 \left(3\epsilon \left((\theta^1)^2 + (\theta^2)^2 \right) + 1 \right) \right) \right) dY_t, \\ d\theta^2 &= \left(-\frac{9\epsilon^2 (\theta^2)^8 + (\theta^2)^4 \left(15\epsilon^2 (\theta^1)^4 + 12\epsilon (\theta^1)^2 + 1 \right) + 3\epsilon (\theta^2)^6 \left(15\epsilon (\theta^1)^2 + 2 \right) - 1}{2\theta^2} \right) dt \\ &\quad + \left(3\epsilon \theta^1 (\theta^2)^3 \right) dY_t. \end{aligned}$$

Example 7. The family of normal distributions is an exponential family, therefore the Stratonovich assumed density filter is equivalent to the Stratonovich projection filter in the Hellinger metric.

3.4 Numerical Results

We simulated the example problem $b(x) = x + \epsilon x^3$ for all of the above approximate filters with $\epsilon = 0.05$. We also computed an “exact” solution using a finite difference method with a fine grid. We define the L^2 residual to be the L^2 distance between the approximate solution and the “exact” solution. We define the Hellinger residual similarly.

In Figure 1 we see the L^2 residuals for the various methods. As predicted by our theory the extrinsic Itô projection in the L^2 metric results in the lowest residuals. The Stratonovich projection in the L^2 metric comes a close second. The projection methods based on optimizing the Hellinger metric perform the worst.

For the Hellinger metric, we have plotted the ratio of the Hellinger residual for each method to the residual of the Itô Hellinger projection method. This is because the residuals themselves are too difficult to distinguish visually. The result is shown in figure 2. Over short time periods the expected value of this ratio should be greater than 1 for all the competing methods. This is born out by the numerical experiments. Note that we do not have a theory over which method will perform better in the longer term and so the fact that this relative residual eventually drops below 1 for both the Itô ADF and the Stratonovich projection does not contradict our theoretical results. Having said that, this behaviour does appear to be fairly consistent for our simple example problem. We see therefore that our “greedy” approach of finding the best residual in the short term will not necessarily lead to the best long term result.

3.5 Conclusion

We have defined a new way to approximate a high dimensional SDE with a lower dimensional SDE on a submanifold. This approximation is based on a new notion of projection, the extrinsic Itô projection method. We show that this

Fig. 1. L^2 residuals for each approximation method

projection leads to an SDE on the submanifold and briefly discuss its optimality property compared to the classic Stratonovich projection used previously in similar contexts.

We then apply this Itô projection to nonlinear filtering. We project the infinite dimensional stochastic PDE of the optimal filter on a finite dimensional Gaussian family. Our explicit calculations show that the extrinsic Itô projection gives rise to new filters for both the L^2 and Hellinger metrics, and shows in particular that the Itô projection is different from the Stratonovich projection.

Numerical results show that our extrinsic Itô projection filters often outperform existing filters over small time horizons. The difference between the extrinsic Itô projection and the Stratonovich projection approaches is small in practice and will be small whenever the extrinsic Itô projection provides a good approximation. Thus the Stratonovich projection approach can be justified in practice and arguably has the merit of being slightly simpler to calculate.

Importantly, as we will demonstrate in a subsequent paper, unlike the heuristic arguments used to justify existing Gaussian filters, we can show that our filters are in some sense “optimal” for the given Hilbert space metric.

Fig. 2. Hellinger residuals for each approximation method divided by residual for Itô Hellinger projection

4 Acknowledgments

We would like to thank the organizers of the workshop on computational information geometry for image and signal processing at ICMS in Edinburgh, 21-25 September 2015, and the participants for the lively discussion and feedback.

References

1. N. U. Ahmed. *Linear and Nonlinear Filtering for Scientists and Engineers*. World Scientific, Singapore, 1998.
2. J. Armstrong and D. Brigo. Coordinate free stochastic differential geometry with jets. *Working paper, forthcoming on arxiv.org*.
3. John Armstrong and Damiano Brigo. Stochastic filtering via L2 projection on mixture manifolds with computer algorithms and numerical examples. *arXiv preprint arXiv:1303.6236*, 2013.
4. John Armstrong and Damiano Brigo. Nonlinear filtering via stochastic PDE projection on mixture manifolds in L^2 direct metric. *Mathematics of Control, Signals and Systems*, 28(1):1–33, 2016.
5. Alan Bain and Dan Crisan. *Fundamentals of stochastic filtering*, volume 3. Springer, 2009.
6. Damiano Brigo. On the nice behaviour of the gaussian projection filter with small observation noise. *Syst. Control Lett.*, 26(5):363–370, 1995.

7. Damiano Brigo, Bernard Hanzon, François Le Gland, et al. Approximate nonlinear filtering by projection on exponential manifolds of densities. *Bernoulli*, 5(3):495–534, 1999.
8. Damiano Brigo, Bernard Hanzon, and François LeGland. A differential geometric approach to nonlinear filtering: the projection filter. *Automatic Control, IEEE Transactions on*, 43(2):247–252, 1998.
9. D. Fisk. Quasi-martingales and stochastic integrals. *PhD Dissertation, Michigan State University, Department of Statistics*, 1963.
10. Michiel Hazewinkel, SI Marcus, and HJ Sussmann. Nonexistence of finite-dimensional filters for conditional statistics of the cubic sensor problem. *Systems & control letters*, 3(6):331–340, 1983.
11. K. Itô. Stochastic differential equations in a differentiable manifold. *Nagoya Math. J.*, (1):35–47, 1950.
12. A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
13. Harold J Kushner. Approximations to optimal nonlinear filters. *Automatic Control, IEEE Transactions on*, 12(5):546–556, 1967.
14. Peter S Maybeck. *Stochastic models, estimation, and control*, volume 3. Academic press, 1982.
15. E.J. McShane. *Stochastic Calculus and Stochastic Models (2nd Ed.)*. Academic Press, New York, 1974.
16. Étienne Pardoux. Filtrage non linéaire et équations aux dérivées partielles stochastiques associées. In *Ecole d’Eté de Probabilités de Saint-Flour XIX 1989*, pages 68–163. Springer, 1991.
17. Jean Picard. Efficiency of the extended kalman filter for nonlinear systems with small noise. *SIAM Journal on Applied Mathematics*, 51(3):843–885, 1991.
18. R. L. Stratonovich. A new representation for stochastic integrals and equations. *SIAM Journal on Control*, 4(2):362–371, 1966.