STOCHASTIC DIFFERENTIAL EQUATIONS WITH APPLICATION TO MANIFOLDS AND NONLINEAR FILTERING

by

Rajesh Rugunanan

A thesis submitted in fulfillment of
the requirements for the degree of

Master of Science

University of the Witwatersrand - Johannesburg

Department of Statistics and Actuarial Science

2005

Approved by____

(Supervisor: Professor F. Beichelt)

TABLE OF CONTENTS

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

____

(Signature of candidate)

________________ day of ________________ 20 ________________

STOCHASTIC DIFFERENTIAL EQUATIONS WITH APPLICATION TO MANIFOLDS AND NONLINEAR FILTERING

by

Rajesh Rugunanan

Abstract

University of the Witwatersrand - Johannesburg

Supervisor:

Professor F. Beichelt

Department of Statistics and Actuarial Science

This thesis follows a direction of research that deals with the theoretical foundations of stochastic differential equations on manifolds and a geometric analysis of the fundamental equations in nonlinear filtering theory. We examine the importance of modern differential geometry in developing an invariant theory of stochastic processes on manifolds, which allow us to extend current filtering techniques to an important class of manifolds. Furthermore, these tools provide us with greater insight to the infinite-dimensional nonlinear filtering problem. In particular, we apply our geometric analysis to the so called unnormalized conditional density approach expounded by M. Zakai. We exploit the geometric setting to study the geometric and algebraic properties of the Zakai equation, which is a linear stochastic partial differential equation. In particular, we investigate the use of Lie algebras and group invariance techniques for dimension analysis and for the reduction of the Zakai equation. Finally, we utilize simulation to demonstrate the superiority of the Zakai equation over the extended Kalman filter for a passive radar tracking problem.

DEDICATION

I dedicate this thesis to my beloved parents who have being a constant source of inspiration and motivation throughout my life. And to my beautiful wife, for being my long awaited companion who has chosen to journey through life by my side.

ACKNOWLEDGMENTS

I would like to extend my sincere gratitude to my supervisor, Professor Frank Beichelt, for supporting my research and for making my graduate studies an enjoyable and memorable experience. Lastly, I want to thank the National Research Foundation (NRF) for their financial assistance.

ACRONYMS

EKF: Extended Kalman Filter

PDE: Partial Differential Equation

SPDE: Stochastic Partial Differential Equation

UCD: Unnormalized Conditional Density

LIST OF FIGURES

LIST OF TABLES

Part I

Introduction

Chapter 1

INTRODUCTION TO IMPORTANT THEMES

The objective of this chapter is to provide an introduction to the themes concerned with the study of stochastic differential equations on manifolds. In mathematics, this area of specialization is called *stochastic differential geometry*. We also focus on the important application of this theory and differential geometry in general, to the problem of nonlinear filtering. This is a subject that has gained considerable interest and is sometimes referred to as *geometric filtering theory*. The synergy between these fields of study is understood through the introduction to deterministic and stochastic dynamical systems and by illustrating their relation to filtering theory.

1.1 A Classical Approach

1.1.1 Deterministic Dynamical Systems

In mathematical systems theory, the plant of a deterministic dynamical system may be defined by a map F_d, which has an input space $\mathcal{U}$ as its domain and an output space $\mathcal{Y}$ as its range, that is $F_d : \mathcal{U} \to \mathcal{Y}$, refer to Fig. 1.1.

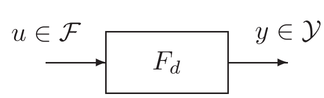

Figure 1.1: Deterministic Dynamical Systems

Traditionally, these systems have been modelled by differential equations in Euclidean spaces. However, a significant advancement was made when the great mathematician Sophus Lie created a revolutionary discipline in mathematics called *Lie Group Theory*. The motivation behind Lie's work was to find a structured approached for the advance treatment of differential equations. Lie groups have an algebraic structure and are also subsets of space, which implies that they have a geometry. Some of these have properties similar to Euclidean space, making it possible to do analysis on them.

1.1.2 Stochastic Dynamical Systems

We now introduce the idea of a stochastic dynamical system, Fig. 1.2; compare with Davis et al [15] page 4.

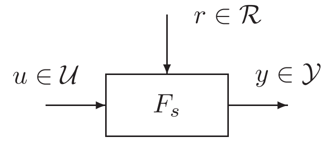

Figure 1.2: Stochastic Dynamical Systems

In this case, the plant may be described by the map $F_s : \mathcal{U} \times \mathcal{R} \to \mathcal{Y}$, where $\mathcal{R}$ denotes an uncertainty space. The map F_s is assumed to be nonanticipative, this means that $\forall r \in \mathcal{R}$ the output $y \in \mathcal{Y}$ is independent of the input $u \in \mathcal{U}$. With regard to the uncertainty variable, we restrict ourselves to the case where there is only a dependence on a stochastic phenomena, hence, $\mathcal{R}$ may be identified with the outcome space of a general random variable. From a modelling and simulation perspective, especially in filtering theory, it is convenient to represent stochastic dynamical systems using differential equations perturbed with white noise terms. This allows one to preserve the important notion of the state of the system, thus maintaining some parallelism with deterministic dynamical systems. Markov theory, Maybeck [52], fits perfectly in this framework. The rigorous mathematical treatment of such processes requires Ito stochastic calculus.

1.1.3 Filtering Theory

Consider the stochastic dynamical system described above. In addition to the observation y, which can be measured, there is another output z, which takes on values in the space $\mathcal{Z}$ and that represents the signal to be estimated, refer to Fig. 1.3.

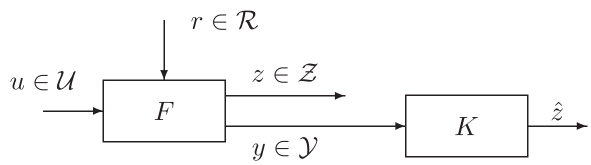

Figure 1.3: Nonlinear Filtering

Our plant can now be modelled by the map $F : \mathcal{U} \times \mathcal{R} \to \mathcal{Y} \times Z$. The central theme in filtering theory is to construct a nonanticipating map $K : \mathcal{Y} \to \mathcal{Z}$ such that $\hat{z} = Ky$ is a least squares estimate of z, Davis et al [15] page 9. In principal, the observations y are used to update a state estimate $\hat{x}$. The random vector is then determined as a function of the estimated state.

1.2 A Contemporary Approach

1.2.1 Deterministic Dynamical Systems

In differential geometry, one considers abstract spaces called manifolds, which resemble Euclidean space locally but have a very general global structure. An important feature of manifolds is that they are equipped with a measure of distances. Recent advances in differential geometry include an elegant geometric interpretation of Lie groups and their associated algebras. These play a central role in fibre bundle theory, which represents the state of art of differential geometry. A contemporary approach to dynamical systems involves the geometric interpretation of the evolution of differential equations on manifolds and the study of corresponding groups. The geometric approach allows one to describe both quantitative and qualitative *global* properties of state trajectories for such systems. Thus group theory and differential geometry provide a natural framework for the study of deterministic dynamical systems, Arnold [3].

1.2.2 Stochastic Dynamical Systems

Stochastic differential geometry deals with the systematic extension of stochastic calculus to differentiable manifolds. The earliest work that formally dealt with the study of diffusion processes on a manifold was initiated by Ito [28] who identified certain peculiarities concerning their construction. Ito findings revealed that one cannot construct a global process by piecing together solutions of local equations without further formal assumptions. Furthermore, Ito differential equations do not behave as ordinary differential equations and therefore cannot be treated as a vector fields. This is because under coordinate transformations, stochastic equation differentials are transformed under Ito's formula, which are different from the conventional rules of differential calculus, however, stochastic differential equations in the Stratonovich form do not have this problem. A general consensus shared by probabilists and mathematicians is that a successful formulation of the Ito and Stratonovich calculi on manifolds must incorporate the concepts associated with the geometry of fibre bundles, Belopolskaya et al [5] and Hsu [25]. Furthermore, advance topics in Lie groups and Lie algebras have proven to be indispensable tools to classify the dimension of stochastic differential equations and to understand their symmetry and reduction to simpler equations. The application of these techniques to nonlinear filtering has profound implications.

1.2.3 Filtering Theory

The geometric theory of filtering starts with the so called Zakai equation. We now embark on a short exposition based on Davis et al [15], page 10, which will illustrate the rich interplay between the various disciplines described earlier. Consider the following stochastic dynamical system:

$$dx_t = f(x_t)dt + g(x_t)dw_t, \; x_{t_0} = x_0 \tag{1.1}$$

$$dy_t = h(x_t)dt + dv_t, \; y_{t_0} = 0 \tag{1.2}$$

$$z_t = k(x_t) \tag{1.3}$$

where

1. $x_t \in \mathbb{R}^n$ represents the state of the system
2. $y_t \in \mathbb{R}^p$ represents the observations of the system and
3. $z_t \in \mathbb{R}^q$ represents a function of the state which we wish to estimate
4. (w_t, v_t) represents the n-dimensional state and the p-dimensional measurement noise respectively; we assume that they are mutually independent Wiener processes and they are independent of x_0
5. $f(x_t) \in \mathbb{R}^n$ represents the dynamic coupling vector
6. $g(x_t) \in \mathbb{R}^{n \times n}$ represents the process noise coupling matrix
7. $h(x_t) \in \mathbb{R}^p$ represents the measurement coupling vector

We assume that f, g and h are continuous functions.

From probability theory, it is well known that the conditional expectation $\hat{z}_t = E[z_t \mid y_\tau, \; t_0 \leq \tau \leq t]$ is the best estimate for z_t in a least squares sense, that is $E[\|z_t - \hat{z}_t\|^2]$ is minimized. If we assume that the stochastic processes associated with Equ.(1.1) possesses the Markov property then, the conditional distribution $\pi_t = P(x_t \mid y_\tau, \; t_0 \leq \tau \leq t)$ can be used to estimate the state of the system. It can be shown that there exists an equation

$$d\pi_t = A_1(\pi_t)dt + B_1(\pi_t)dy_t \tag{1.4}$$

which can be solved for π_t to determine the estimate

$$\hat{z}_t = \int_X k(x)\pi_t dx \tag{1.5}$$

with π_{t_0} being the initial distribution of x_0.

In the chapters that follow, we show that the normalized conditional distribution π_t is related to another distribution p_t in the following manner

$$\pi_t = \frac{p_t}{\int_X p_t dx} \tag{1.6}$$

We appropriately call p_t the normalized conditional distribution of π_t.

It turns out that p_t is a solution of a bilinear stochastic partial differential equation

$$dp_t = A_2 p_t dt + h p_t dy_t \tag{1.7}$$

which we call the Zakai equation in honour of Dr. Moshe Zakai for his pioneering contribution to nonlinear filtering theory and the application and theory of stochastic processes, Zakai [69]. Duncan [16] and Mortensen [57] have also been independently credited for their work on Equ.(1.7), which is sometimes called the Duncan-Mortensen-Zakai equation. The greatest advantage that the Zakai equation, Equ.(1.7), has in filtering theory is that it is considerably easier to work with than the update equation for π_t, Equ.(1.4). The estimate $\hat{z}_t$ can then be determined by the following equation:

$$\hat{z}_t = \frac{\int_X k(x)p_t dx}{\int_X p_t dx} \tag{1.8}$$

Some of the more recent developments in stochastic filtering include:

1. the discovery of finite dimensional nonlinear filters

2. the introduction of Lie algebraic and differential geometric methods

3. group representation methods

4. the rigorous formulation of the theory of partial stochastic differential equations and

5. the development of robust or pathwise solutions of the filtering problem

The first two fields of study are related and together with the third constitute the bulk of contemporary research. To briefly illustrate their importance, consider the following discussion.

Recall from an earlier definition that the filter can be associated with a non-anticipative function that maps the observations y_t to estimate $\hat{z}_t$. The update equations for π_t and p_t are realizations of this map. However, the difficulty associated with these maps is that they are in general infinite dimensional. More precisely, they map $\mathbb{R}^n$ to $\mathbb{R}$, that is an infinite dimensional object. Brockett utilized geometric tools to study the bilinear structure of the Zakai equation to determine conditions for the existence of finite dimensional filters, which admits a finite dimensional realization; this implies the existence of a differential map and an output equation

$$\dot{m}_t = v(m, \dot{y}) \tag{1.9}$$
$$\hat{z} = w(m) \tag{1.10}$$

on a finite dimensional manifold $\mathcal{M}$. These maps can be associated with the same input/output map for π_t and p_t. Therefore, the concept of dimension provides a deeper understanding of the underlying geometry of the filter. In the case of p_t the resulting filter will have the following form:

$$\dot{m}_t = \alpha(m) + \beta(m)\dot{y} \tag{1.11}$$
$$\hat{z} = \gamma(m) \tag{1.12}$$

where $\alpha(m)$ and $\beta(m)$ are vector fields on the manifold $\mathcal{M}$. The above system of equations, Equ.(1.11) and Equ.(1.12), is similar to the Kalman-Bucy filter. The main tool in this analysis is the Lie algebra of the operators associated with the Zakai equation, which has a very deep geometric structure.

From our discussion, it is clearly apparent that a geometric theory for stochastic dynamical systems and filtering theory has an important place in mathematical analysis.

1.3 Outline of Thesis

The invariant theory of stochastic equations on manifolds is *not* a precursor to the geometric theory of nonlinear filtering. However, it is important to develop a coalescent theory concerning their construction on manifolds before we attempt to understand their geometric and algebraic properties. Our application to nonlinear filtering theory is motivated by the fact that it is a subject of immense sophistication that can benefit from new developments in infinite-dimensional analysis. Furthermore, the practical importance of filtering theory in a multitude of diverse disciplines cannot be overemphasized.

Chapters 2-5 constitute part II of this thesis, which deals with the theoretical foundations of modern differential geometry, stochastic calculus and nonlinear filtering theory.

In chapter 2 we provide a contemporary introduction to the geometry of fibre bundles, which forms the foundation for much of the pertinent conceptual framework of our research. In particular, we examine the importance of Lie groups and their associated algebras in constructing connections on principal fibre bundles. The connection allows us to define horizontal lifts, parallel transport and covariant differentiation of tangent vectors, which constitute the basic machinery for doing calculus on a manifold. The chapter concludes with an introduction to the frame bundle as a principal fibre bundle and some concepts in Riemannian differential geometry, which allow us to describe stochastic processes on manifolds. Another advantage gained from the fibre bundle approach is that the associated Lie group theory will enable us to introduce tools that are essential for the advance treatment of the nonlinear filtering problem.

Chapter 3 focuses on the calculus of semimartingales, which feature prominently in nonlinear filtering theory and stochastic differential geometry. We introduce both Ito and Stratonovich stochastic equations and Ito's lemma for semimartingales. The chapter concludes with a summary of Markov and diffusion processes.

Chapter 4 specializes on the fundamental equations of the nonlinear filtering problem. We examine both the innovations approach as well as the unnormalized conditional density approach to illustrate the distinct advantages the latter technique has over the former.

In chapter 5, we introduce the Kalman-Bucy filter as a special case of nonlinear filtering problem. This is done primarily to make geometric nonlinear filtering accessible to practitioners of Kalman filtering theory so that a deeper appreciation can be gained for the techniques described in this thesis.

Part III of this thesis comprises of chapters 6-7, which highlights the synergy between the various disciplines introduced in part I.

In chapter 6 we take a detailed look at semimartingales on an arbitrary manifold. Using the connection on a frame bundle, we explain how such processes can be associated with a semimartingale in Euclidean space. The chapter concludes with some applications to nonlinear filtering.

In chapter 7 we describe a geometric analysis of the Zakai equation. In particular, we will show that the Lie algebra of operators associated with the Zakai equation and its invariance under an appropriate group of transformations exhibits the structure of the filtering problem. We investigate symmetries that simplify the resolution of the Zakai equation via group invariant solutions.

In chapters 8 and 9 of part IV, we examine numerical techniques to solve the Zakai equation and consider an application.

In chapter 8 we introduce numerical algorithms to solve the Zakai equation. The first algorithm is based on a combination of semigroup techniques and gauge transformations. The second algorithm relies entirely on the method of finite differences, which have been successfully applied to partial differential equations. We also introduce the discrete extended Kalman filter, which provides a useful benchmark for studies undertaken in nonlinear filtering.

In chapter 9 we use simulation to demonstrate the superiority of the Zakai equation against the extended Kalman filter for a nonlinear passive radar tracking problem.

Lastly, we bring our thesis to conclusion in chapter 10 of part V.

Part II

A Review of Fibre Bundles, Stochastic Calculus and Nonlinear Filtering Theory

Chapter 2

THE GEOMETRY OF FIBRE BUNDLES

In this chapter we provide an introduction to the necessary differential geometry and Lie group theory that is required to understand the geometry of fibre bundles. We place great emphasis on fibre bundle theory to give a complete description of the connection and its relation to horizontal lifts, parallel transport and covariant differentiation. As we will discover in later chapters, these concepts form the basic tools of stochastic differential geometry. For further details, Nakahara [58], Isham [27], Gockeler et al [20] and Lee [48] provide an excellent introduction to the subject.

2.1 Modern Differential Geometry

Differential geometry is essentially a generalization of the mathematics of curves and surfaces to arbitrary objects, which we call manifolds.

2.1.1 Manifolds and Submanifolds

Definition 2.1.1 (pages 81 & 171-172 Nakahara [58]). *We say that $\mathcal{M}$ is an m-dimensional differentiable **manifold** if:*

1. *$\mathcal{M}$ is a **topological space**. This means that there exists a **topology** on $\mathcal{M}$ which is defined by the set*

$$\mathcal{T} = \{U_i \subset \mathcal{M} \mid \textit{where } U_i \textit{ are open sets } \forall i \in I\} \tag{2.1}$$

$\mathcal{T}$ satisfies the following requirements:

(a) $\emptyset, \mathcal{M} \in \mathcal{T}$

(b) $\cup_{j\in J} U_j \in \mathcal{T}$, *where* $J \subseteq I$

(c) $\cap_{k\in K} U_j \in \mathcal{T}$, *where* $K \subseteq I$

2. *$\mathcal{M}$ is equipped with an **atlas** $\mathcal{A} = \{(U_\alpha, \phi_\alpha) \mid \alpha \in I\}$, which is a family of pairs (U, ϕ) called **charts**. The family $\{U_\alpha \mid \alpha \in I\}$ of open sets of $\mathcal{A}$ covers $\mathcal{M}$ in the following sense:*

$$\mathcal{M} = \cup_{\alpha\in I} U_\alpha$$

*The functions $\phi_\alpha : U_\alpha \to \mathbb{R}^m$ of $\mathcal{A}$ are called **co-ordinate functions**. They belong to a class of continuous and injective functions that map from one space to another, we call these functions **homeomorphisms**. Furthermore, given U_α and U_β such that $U_\alpha \cap U_\beta \neq \emptyset$, the map*

$$\phi_\beta \circ \phi_\alpha^{-1} \mid_{\phi_\alpha(U_\alpha \cap U_\beta)}: \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$$

is in infinitely differentiable, that is $\phi_\beta \circ \phi_\alpha^{-1} \in C^\infty \; \forall \alpha, \beta \in I$

Given the homeomorphism $f : \mathcal{M} \to \mathcal{N}$, we say that $\mathcal{M}$ is ***diffeomorphic*** to $\mathcal{N}$ if f and its bijection f^{-1} are C^∞ functions.

Example 2.1.1 (page 173 Nakahara [58]). *The most trivial example of a manifold is the Euclidean space $\mathbb{R}^n$, which has a single chart that covers the whole space and the co-ordinate function ϕ can be the identity map.*

Example 2.1.2 (page 174 Nakahara [58]). *We consider a special case of a 1-dimensional manifold $\mathcal{M}$ given by the unit circle $S^1 = \{(x^1, x^2) \in \mathbb{R}^2 \mid (x^1)^2 + (x^2)^2 = 1\}$. We need to define two charts; let us define the first chart as*

$$\phi_1^{-1} : (0, 2\pi) \to S^1, \; \theta \mapsto \phi_1^{-1}(\theta) := (\cos\theta, \sin\theta) \tag{2.2}$$

whose image is $Im(\phi_1^{-1}) = S^1 - \{1, 0\}$. Similarly, define the second chart as

$$\phi_2^{-1} : (-\pi, \pi) \to S^1,\ \theta \mapsto \phi_2^{-1}(\theta) := (\cos\theta, \sin\theta) \tag{2.3}$$

whose image is $Im(\phi_2^{-1}) = S^1 - \{-1, 0\}$. *The functions* ϕ_1^{-1} *and* ϕ_2^{-1} *are invertible and all maps* ϕ_1, ϕ_2, ϕ_1^{-1} *and* ϕ_2^{-1} *are continuous. This implies that* ϕ_1 *and* ϕ_2 *are homeomorphisms. Furthermore, the compositions* $\phi_1 \circ \phi_2^{-1}$ *and* $\phi_2 \circ \phi_1^{-1}$ *are smooth.*

In our thesis we will encounter smooth subsets of a manifold that are endowed with the structure of a manifold that has additional properties; we call these submanifolds.

Definition 2.1.2 (page 64 Isham [27]). *Let* $\mathcal{M}$ *be an* m*-dimensional manifold with atlas* $\mathcal{A}$. *Consider the subset* $\tilde{\mathcal{M}} \subset \mathcal{M}$ *which is equipped with the subset topology of* $\mathcal{M}$. *The set* $\tilde{\mathcal{M}}$ *is said to be an* n*-dimensional* ***submanifold*** *of* $\mathcal{M}$ *if* $\forall p \in \tilde{\mathcal{M}}$ $\exists (U_p, \phi_p) \in \mathcal{A}$ *such that*

1. $p \in U_p \subset \mathcal{M}$

2. $\phi_p : U_p \to \mathbb{R}^m$ *satisfies*

$$\phi_p(U_p \cap \tilde{\mathcal{M}}) = \phi_p(U_p) \cap \mathbb{R}^n \text{ where } 0 < n \leq m \tag{2.4}$$

The above definition can be explained as follows. The submanifold $\tilde{\mathcal{M}}$ of $\mathcal{M}$ comprises of a set of points of $\mathcal{M}$ which have the following property: in an open neigbouhood U_p about some point $p \in \mathcal{M}$ there exists a coordinate system of $\mathcal{M}$ where the points of $\tilde{\mathcal{M}}$ in U_p are characterized by $x^1 = \ldots = x^{m-n} = 0$.

2.1.2 Lie Group Theory

Lie group theory began with the study of the symmetries of systems of ordinary differential equations. The reason why Lie groups are so important is because they are equipped with a natural differentiable structure, hence, the tangent space at any point on the Lie group is a well defined geometric object. What is remarkable is that the tangent space at the identity element of the group has a well defined Lie algebraic structure. It turns out that many properties of the group are reflected on its Lie algebra. The geometric interpretation of Lie group theory is central to the study of fibre bundles and the theory of connections.

Definition 2.1.3 (page 116 Gockeler et al [20]). *A **Lie group** is a smooth manifold $\mathcal{G}$ that is also a group, with the following properties:*

1. *group multiplication $M : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$, which is defined as $(g, h) \mapsto M(g, h) := gh$ is continuous*

2. *group inversion $I : \mathcal{G} \to \mathcal{G}$, where $g \mapsto I(g) := g^{-1}$is continuous.*

Example 2.1.3 (page 207 Nakahara [58]). *Let S^1 be the unit circle in the complex plane*

$$S^1 = \{e^{i\theta} \mid \theta \in \mathbb{R} \ (mod\ 2\pi)\ \} \tag{2.5}$$

The group operations defined by $e^{i\theta}e^{i\phi} = e^{i(\theta+\phi)}$ and $(e^{i\theta})^{-1} = e^{-i\theta}$ are differentiable. Thus S^1 is a Lie group.

Lie Group Actions and Representations

In fibre bundle theory, the focus of our attention is not the abstract structure of Lie groups but on the action of a Lie group on a manifold and how it affects it.

Let $\mathcal{M}$ be a manifold and $\mathcal{G}$ a group. The ***representation*** ρ of $\mathcal{G}$ on $\mathcal{M}$ is a Lie group ***homomorphism***, that is it preserves the algebraic structure on $\mathcal{G}$

$$\rho : \mathcal{G} \xrightarrow{\simeq} Diff(\mathcal{M}),\ g \mapsto \rho_g \tag{2.6}$$

where $Diff(\mathcal{M}) = \{f \in C^{\infty}(\mathcal{M}) \mid f : \mathcal{M} \to \mathcal{M}\}$ denotes the set of all diffeomorphisms on $\mathcal{M}$. We now in a position to describe two fundamental group actions on a manifold.

Given a topological group $\mathcal{G}$ and a manifold $\mathcal{M}$. A ***left action*** of $\mathcal{G}$ on $\mathcal{M}$ is a differentiable map

$$L : \mathcal{G} \times \mathcal{M} \to \mathcal{M},\ (g, m) \mapsto L(g, m) := \rho_g m \tag{2.7}$$

The map $\rho_g : \mathcal{M} \to \mathcal{M},\ m \mapsto \rho_g m := gm$ is a diffeomorphism of $\mathcal{M}$ $\forall g \in \mathcal{G}$ where

1. $\rho_e = id_{\mathcal{M}}$

2. the homomorphism property is satisfied: $\rho_g \circ \rho_h = \rho_{gh}\ \forall g, h \in \mathcal{G}$

Note that ρ_g is an ***isomorphism*** because it is homomorphic and bijective.

Given a topological group $\mathcal{G}$ and a manifold $\mathcal{M}$. A ***right action*** of $\mathcal{G}$ on $\mathcal{M}$ is a differentiable map

$$R : \mathcal{M} \times \mathcal{G} \to \mathcal{M},\ (m, g) \mapsto R(m, g) := \delta_g m \tag{2.8}$$

The map $\delta_g : \mathcal{M} \to \mathcal{M},\ m \mapsto \delta_g m := mg$ is a diffeomorphism of $\mathcal{M}$ $\forall g \in \mathcal{G}$ where

1. $\delta_e = id_{\mathcal{M}}$

2. the antihomomorphism property is satisfied: $\delta_g \circ \delta_h = \delta_{gh}\ \forall g, h \in \mathcal{G}$

Given a left action $g \mapsto \rho_g$ we can define, page 178 Isham [27], a right action as

$$\delta_g := \rho_{g^{-1}} = (\rho_g)^{-1} \tag{2.9}$$

Note that we define $\rho_{g^{-1}} m = mg$.

In this section we introduce a map whose image gives rise to an equivalence relation on a differentiable manifold. This leads to a very powerful geometric interpretation, which we will use to introduce principal fibre bundles.

Given a representation ρ of $\mathcal{G}$ on $\mathcal{M}$, for any $x \in \mathcal{M}$ define the map

$$bit_x : \mathcal{G} \to \mathcal{M},\ g \mapsto bit_x(g) := \rho_g x \tag{2.10}$$

with $bit_x e = x$.

We define the ***orbit*** of x by the image of bit_x

$$\mathcal{O}_x = bit_x(\mathcal{G}) = \{\rho_g x \in \mathcal{M} \mid g \in \mathcal{G}\} \subset \mathcal{M} \tag{2.11}$$

The orbit of x consists of all points in $\mathcal{M}$ that can be reached from x by group transformation. $x, y \in \mathcal{M}$ are in the same orbit with respect to $\mathcal{G} \iff \exists g \in \mathcal{G}$ such that $y = \rho_{g^{-1}} x$. Therefore, $\mathcal{O}_x$ defines an equivalence relation on x. Thus, the manifold $\mathcal{M}$ can be decomposed into a disjoint union of orbits

$$\mathcal{M} = \cup_{x \in \mathcal{M}} \mathcal{O}_x \tag{2.12}$$

The set of all orbits defines a quotient space, that is, an orbit space

$$\mathcal{M}/\mathcal{G} = \{\mathcal{O}_x \mid x \in \mathcal{M}\} \tag{2.13}$$

The quotient map $\bar{\pi} : \mathcal{M} \to \mathcal{M}/\mathcal{G}$ is continuous and is defined by $x \mapsto \bar{\pi}(x) := \mathcal{O}_x$. The notion of the orbit together with some of the special properties of $\mathcal{G}$ form the basis of principal fibre bundles. We now introduce two important subgroups of $\mathcal{G}$ that allow us to understand the properties of representations more completely.

The ***isotropy group*** or stability group is a subgroup of $\mathcal{G}$

$$I(x) = \{g \in \mathcal{G} \mid \rho_g x = x\} \subset \mathcal{G} \tag{2.14}$$

that identifies all the elements of $\mathcal{G}$ that fix $x \in \mathcal{M}$.

The ***kernel*** of a $\mathcal{G}$-action is the subgroup of $\mathcal{G}$ defined by

$$K = \cap_{x \in \mathcal{M}} I(x) \subset \mathcal{G} \tag{2.15}$$

The kernel represents that component of the group that is not involved in the group action, it is a normal subgroup of $\mathcal{G}$.

A representation is said to be:

1. ***transitive*** if there is only one group orbit $\mathcal{O}_x$, in this case, $\mathcal{M}$ is said to be homogeneous under the given action

2. ***effective*** or faithful if $K = \{e\}$, this implies that the homomorphism $g \mapsto \rho_g$ is injective

3. ***free*** if the only element of $\mathcal{G}$ that fixes $x \in \mathcal{M}$ is the identity element $e \in \mathcal{G}$, that is $I(x) = \{e\}$. This implies that $\mathcal{O}_x \simeq \mathcal{G}$, note that if the representation is free, it also implies that the action is effective

We now restrict our attention to the important case of the action of Lie groups on itself.

1. The simplest example of such representations arise with the left action and right action on a Lie group $\mathcal{G} \times \mathcal{G} \to \mathcal{G}$. Let $(g, h) \in \mathcal{G} \times \mathcal{G}$, the above mentioned actions give rise to the following respective maps $(g, h) \mapsto L(g, h) := L_g h$ and $(g, h) \mapsto R(g, h) := R_g h$.

   We define the ***left translation*** and ***right translation*** of $g \in \mathcal{G}$, denoted by L_g and R_g respectively, by the following maps, $L_g : \mathcal{G} \to \mathcal{G}$ and $R_g : \mathcal{G} \to \mathcal{G}$, where $h \mapsto L_g(h) = gh$ and $h \mapsto R_g(h) = hg$ respectively. L_g and R_g are diffeomorphisms with inverses given by $(L_g)^{-1} = L_{g^{-1}}$ and $(R_g)^{-1} = R_{g^{-1}}$.

   These translations are very important because they endow the Lie group with a well defined differentiable structure given a single chart at the identity element

$g \in \mathcal{G}$, page 3 Schmid [62]. The notion of left translation is important to give a more technical description of Lie algebras.

2. The left and right translation allows a Lie group to act on itself through ***inner automorphisms*** $Ad_g = L_g \circ R_{g^{-1}} : \mathcal{G} \to \mathcal{G}$ where $(h) \mapsto Ad_g(h) := ghg^{-1}$.

2.1.3 Fibre Bundle Theory

Theoretical Background

Definition 2.1.4 (page 133 Gockeler et al [20]). *A differentiable **fibre bundle** $(\mathcal{B}, \mathcal{M}, \mathcal{F}, \pi)$, Fig. 2.1, is defined by:*

1. *a differentiable manifold $\mathcal{B}$ called the **bundle space***

2. *a differentiable manifold $\mathcal{M}$ called the **base space***

3. *a differentiable manifold $\mathcal{F}$ called the **standard fibre***

4. *a diffeomorphism $\Psi : \mathcal{B} \to \mathcal{F}$*

5. *a Lie group $\mathcal{G}$ or **structure group** which is represented effectively on $\mathcal{F}$. Also, $\forall g \in \mathcal{G}$ we have the diffeomorphic representation $\rho_g : \mathcal{F} \to \mathcal{F}$*

6. *a smooth map $\pi : \mathcal{B} \to \mathcal{M}$ called the **projection** which satisfies the following properties:*

   (a) π is surjective and differentiable

   (b) $\forall x \in \mathcal{M}$, $\mathcal{B}_x = \pi^{-1}(x)$ is an embedded submanifold of $\mathcal{B}$ which is diffeomorphic to $\mathcal{F}$, $\mathcal{B}_x$ is the fibre over x and represents the orbits of the $\mathcal{G}$-action

   (c) π defines a local product structure on $\mathcal{B}$, this means that there exists an open covering $\{\mathcal{U}_\alpha\}$ of $\mathcal{M}$ where $\forall x \in \mathcal{M}$ $\exists \mathcal{U}_\alpha \subset \mathcal{M}$ such that the restriction

$$\psi_\alpha \mid_{\pi^{-1}(\mathcal{U}_\alpha)} : \pi^{-1}(\mathcal{U}_\alpha) \to \mathcal{U}_\alpha \times \mathcal{F} \tag{2.16}$$

is diffeomorphic. The diffeomorphism ψ_α respects the composition $\pi_1 \circ \psi_\alpha = \pi$ where $\pi_1 : \mathcal{U}_\alpha \times \mathcal{F} \to \mathcal{U}_\alpha$ is the projection onto the first factor.

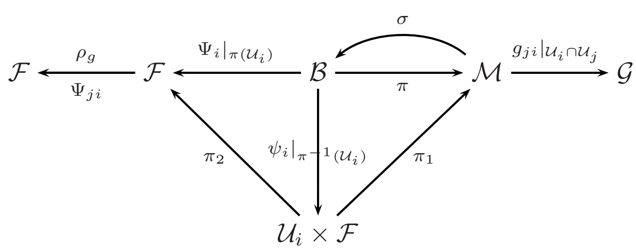

Figure 2.1: Fibre Bundles

The dimensions of the manifolds involved in the above definition satisfies the following rule

$$\dim\mathcal{B} = \dim\mathcal{M} + \dim\mathcal{F} \tag{2.17}$$

A bundle that is isomorphic to the product bundle $\mathcal{M} \times \mathcal{F}$ is called a ***trivial bundle***. Since π defines a local structure on $\mathcal{B}$, we define the local trivialization as follows:

$$\psi_\alpha \mid_{\pi^{-1}(\mathcal{U}_\alpha)} : \pi^{-1}(\mathcal{U}_\alpha) \to \mathcal{U}_\alpha \times \mathcal{F},\ p \mapsto \psi_\alpha(p) := (\pi(p), \Psi_\alpha(p)) \tag{2.18}$$

The map Ψ_α is also a diffeomorphism, which arises from the restriction $\Psi_\alpha = \Psi \mid_{\pi^{-1}(\mathcal{U}_\alpha)}$. Ψ_α can further be restricted to the fibre $\mathcal{B}_x = \pi^{-1}(x)$ over $x \in \mathcal{U}_\alpha$ giving rise to the map $\Psi_{\alpha,x} = \Psi_\alpha \mid_{\pi^{-1}(x)}$.

$(\mathcal{B}, \mathcal{M}, \mathcal{F}, \pi)$ may be identified with an n-dimensional topological bundle over $\mathcal{M}$ having bundle atlas

$$\mathcal{B} = \{(\pi^{-1}(\mathcal{U}_\alpha), \psi_\alpha) \mid \alpha \in I\} \tag{2.19}$$

where $\mathcal{U}_\alpha$ forms a covering for $\mathcal{M}$

$$\mathcal{M} = \cup_{\alpha \in I} \mathcal{U}_\alpha \tag{2.20}$$

and $\forall \alpha \neq \beta \ \exists g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathcal{G}$ such that

$$\psi_\beta \circ \psi_\alpha^{-1} \mid_{(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)\times\mathcal{F}}: (\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \times \mathcal{F} \to (\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \times \mathcal{F},\ (x, v) \mapsto (x, (g_{\beta\alpha}(x))(v)) \tag{2.21}$$

The elements

$$\{g_{\beta\alpha} \mid g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathcal{G},\ \alpha, \beta \in I\} \tag{2.22}$$

are called ***transition functions***, where $\rho(g_{\beta\alpha}(x)) = \Psi_{\beta,\alpha}$. These functions are required to be smooth and they satisfy the following conditions:

1. the cocyclic condition $g_{\beta\alpha}(x) g_{\alpha\gamma}(x) = g_{\beta\gamma}(x)$ and

2. $g_{\beta\alpha}(x) = ((g_{\alpha\beta})(x))^{-1}$

The transition functions indicate how the spaces $\mathcal{U}_\alpha \times \mathcal{F}$ and $\mathcal{U}_\beta \times \mathcal{F}$ are glued together. For $x \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta$ the map $\Psi_{\beta,\alpha} = \Psi_{\beta,x} \circ \Psi_{\alpha,x}{}^{-1}$ corresponds to an element $g_{\beta\alpha}(x)$ of the structure group $\mathcal{G}$ via the representation ρ.

Sections

A ***section*** of a fibre bundle is a map

$$\sigma : \mathcal{M} \to \mathcal{B},\ \pi \circ \sigma = id_{\mathcal{M}} \tag{2.23}$$

Vector Bundles and Principal Fibre Bundles

A very important example of a fibre bundle is a ***vector bundle***, which we denote by (E, M, π). In this case, the fibres are vector spaces, the structure group is a subgroup of $GL(\mathcal{F})$ and the diffeomorphisms $\Psi_{\alpha,x}$ are vector space isomorphisms. The set of all smooth sections $\Gamma^\infty(\mathcal{M},\mathcal{E}) = \{\sigma \in C^\infty \mid \sigma : \mathcal{M} \to \mathcal{E}\}$ on $\mathcal{E}$ has a vector space structure. Vector bundles are very important in the study of differential geometry because they are generalizations of the following bundles:

1. ***tangent bundle*** $(T\mathcal{M},\mathcal{M},\pi)$, where $T\mathcal{M} = \cup_{p\in\mathcal{M}} T_p\mathcal{M}$.

   We call the smooth section $X : \mathcal{M} \to T\mathcal{M}$ a ***vector field***, the set of all vector fields is denoted by $\mathfrak{X}(\mathcal{M}) = \Gamma^\infty(\mathcal{M}, T\mathcal{M})$. The vector field $X \in \mathfrak{X}(\mathcal{M})$ maps $p \in \mathcal{M}$ to a ***tangent vector*** $X_p : C^\infty(\mathcal{M}) \to \mathbb{R}$ such that $f \mapsto X_p(p) := (Xf)(p)$ where $Xf \in C^\infty(\mathcal{M})$. The fibre $\pi^{-1}(p) = T_p\mathcal{M}$ is called the ***tangent space*** of $\mathcal{M}$ at p and is defined as the set of all ***tangent vectors*** X_p at $p \in \mathcal{M}$

   $$T_p\mathcal{M} = \{X_p \mid X_p : C^\infty(\mathcal{M}) \to \mathbb{R}; p \in \mathcal{M}\} \tag{2.24}$$

   $T_p\mathcal{M}$ has the structure of a real vector space.

2. ***cotangent bundle*** $(T^*\mathcal{M},\mathcal{M},\pi)$, where $T^*\mathcal{M} = \cup_{p\in\mathcal{M}} T_p^*\mathcal{M}$.

   The smooth section $\omega : \mathcal{M} \to T^*\mathcal{M}$ a ***covector field***, the set of all such fields is denoted by $\mathfrak{X}^*(\mathcal{M}) = \Gamma^\infty(\mathcal{M}, T^*\mathcal{M})$. We call $T_p^*\mathcal{M}$ the ***cotangent space*** at p and define it to be the dual space of $T_p\mathcal{M}$, that is $T_p^*\mathcal{M} = (T_p\mathcal{M})^*$. The elements of $T_p^*\mathcal{M}$ are called ***covectors*** at $p \in \mathcal{M}$ and are denoted by ω_p.

3. ***tensor bundle*** $(T_s^r(\mathcal{M}),\mathcal{M},\pi)$, where $T_s^r(\mathcal{M}) = \cup_{p\in\mathcal{M}} T_s^r(T_p\mathcal{M})$.

   $T_s^r(T_p\mathcal{M})$ is a tensor space over a tangent space $T_p\mathcal{M}$ that we define by

   $$T_s^r(T_p\mathcal{M}) = T^r(T_p\mathcal{M}) \otimes T_s(T_p\mathcal{M}) = \underbrace{T_p^*\mathcal{M} \otimes \ldots \otimes T_p^*\mathcal{M}}_{r\ copies} \otimes \underbrace{T_p\mathcal{M} \otimes \ldots \otimes T_p\mathcal{M}}_{s\ copies} \tag{2.25}$$

A ***tensor field*** of type (r, s) on a manifold $\mathcal{M}$ is a C^∞ section of a tensor bundle $T^r_s(\mathcal{M})$. We denote by $\mathcal{T}^r_s(\mathcal{M}) = \Gamma^\infty(\mathcal{M}, T^r_s(\mathcal{M}))$ the set of all C^∞ sections on $\mathcal{M}$.

4. ***bundle of exterior*** k***-forms*** $(\Lambda^k(\mathcal{M}), \mathcal{M}, \pi)$, where $\Lambda^k(\mathcal{M}) = \cup_{p\in\mathcal{M}}\Lambda^k(T_p\mathcal{M})$

   We denote by $\Lambda^k(T_p\mathcal{M})$ the vector space of all alternating covariant $(0, k)$-tensors on $T_p\mathcal{M}$. An element of $\Lambda^k(T_p\mathcal{M})$ is called a k-form. The set of all continuous sections of exterior k-forms on $\mathcal{M}$ are denoted by $\Omega^k(\mathcal{M}) = \Gamma^\infty(\mathcal{M}, \Lambda^k(\mathcal{M}))$.

Remark 2.1.1. *1. Let $\phi : \mathcal{M} \to \mathcal{N}$ be a map between two manifolds. At $p \in \mathcal{M}$ we define the* ***push forward*** *of ϕ by the map $\phi_* : T_p\mathcal{M} \to T_{\phi_{(p)}}\mathcal{N}$, Fig. 2.2. The map ϕ_* is also called the differential of ϕ at $p \in \mathcal{M}$ and is sometimes denoted by $T_p\phi$. The* ***pull back*** *is defined by the map $\phi^* : T^*_{\phi(p)}\mathcal{N} \to T^*_p\mathcal{M}$, Fig. 2.2.*

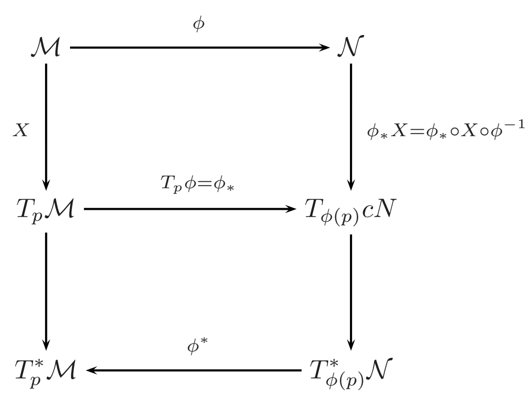

Figure 2.2: Push Forwards and Pull Backs

2. This section is based on the well known ***Tensor Characterization Lemma****, page 265 Lee [48]. It will allow us to reconcile the abstract notion of tensor fields with operators encountered in classical Riemannian geometry. The multilinear map $\tau : \mathfrak{X}^*(\mathcal{M})^r \times \mathfrak{X}(\mathcal{M})^s \to C^\infty(\mathcal{M})$ is induced by a smooth tensor field of degree (r, s) if and only if it is multilinear over $C^\infty(\mathcal{M})$. Likewise, the multilinear map $\tau' : \mathfrak{X}^*(\mathcal{M})^r \times \mathfrak{X}(\mathcal{M})^s \to \mathfrak{X}(\mathcal{M})$ is induced by the tensor field of degree*

$(r+1,s)$ if and only if it is multilinear over $C^\infty(\mathcal{M})$. In both cases, the induced maps τ and τ' can be associated with the respective tensor fields.

$(\mathcal{P},\mathcal{M},\pi)$ is a $\mathcal{G}$-bundle if $\mathcal{P}$ is a right action $\mathcal{G}$-space and if there is a bundle isomorphism $(\mathcal{P},\mathcal{M},\pi)\xrightarrow{\cong}(\mathcal{P},\mathcal{P}/\mathcal{G},\bar{\pi})$, where $\bar{\pi}$ represents the projection map and $\mathcal{P}/\mathcal{G}$ is the orbit space, page 220-221 Isham [27]. The fibre of the $\mathcal{G}$-bundle are the orbits of the $\mathcal{G}$-action on $\mathcal{P}$. If $\mathcal{G}$ acts freely on $\mathcal{P}$, then $(\mathcal{P},\mathcal{M},\pi)$ is called a ***principal $\mathcal{G}$-bundle*** and $\mathcal{G}$ is called the structure group of the bundle. The freedom of the $\mathcal{G}$-action implies that all the orbits of the group action are homomorphic to the group, that is $\mathcal{O}_x\xrightarrow{\simeq}\mathcal{G}\ \forall x\in\mathcal{P}$.

Associated Bundles

In general, for a Lie group $\mathcal{G}$ and a principal $\mathcal{G}$-bundle $(\mathcal{P},\mathcal{M},\pi)$ we consider a manifold $\mathcal{E}$ with a left $\mathcal{G}$-action $L:\mathcal{G}\times\mathcal{E}\to\mathcal{E}$ and associate to this a fibre bundle $(\mathcal{P}_\mathcal{E},\mathcal{M},\pi_\mathcal{E})$ with fibre $\mathcal{E}$. To do this, we define the total space $\mathcal{P}_\mathcal{E}$ as the orbit space

$$\mathcal{P}_\mathcal{E}=(\mathcal{P}\times\mathcal{E})/\mathcal{G}=\mathcal{P}\times_\mathcal{G}\mathcal{E} \tag{2.26}$$

given that the $\mathcal{G}$-action on $\mathcal{P}\times\mathcal{E}$ is defined by

$$(p,x)\cdot g=(pg,g^{-1}x) \tag{2.27}$$

where $p\in\mathcal{P},x\in\mathcal{E}$ and $g\in\mathcal{G}$. $\mathcal{P}_\mathcal{E}$ is therefore a quotient space for the equivalence relation $\sim$, where $(p,x)\sim(q,y)$ if and only if $\exists g\in\mathcal{G}$ such that $q=pg$ and $y=gx$. The projection map $\pi_\mathcal{E}:\mathcal{P}_\mathcal{E}\to\mathcal{M}$ is induced by the composition $\mathcal{P}\times\mathcal{E}\xrightarrow{proj}\mathcal{P}\xrightarrow{\pi}\mathcal{M}$.

The significance of associated bundles lies in the following important fact. Given a principal fibre bundle $(\mathcal{P},\mathcal{M},\pi)$ with structure group $\mathcal{G}$, there exists a vector bundle $(\mathcal{E},\mathcal{M},\pi)$ associated with $(\mathcal{P},\mathcal{M},\pi)$ by the natural action of $GL(n;\mathbb{R})$ on $\mathbb{R}^n$. We will briefly explain how this arises.

Let $\rho : \mathcal{G} \to GL(n;\mathbb{R})$, $g \mapsto \rho_g$, be a representation of $\mathcal{G}$ on $\mathbb{R}^n$. On the product manifold $\mathcal{P} \times \mathbb{R}^n$ we define the right action of $\mathcal{G}$ by $\mathcal{G} \times (\mathcal{P} \times \mathbb{R}^n) \to \mathcal{P} \times \mathbb{R}^n$, where $(g,(p,\xi)) \mapsto (R_g p, \rho_{g^{-1}}\xi)$. Let $\mathcal{E} = \mathcal{P} \times_{\mathcal{G}} \mathbb{R}^n$ be the quotient space under this action and let $p\xi \in \mathcal{E}$ represent the equivalence class of $(p,\xi) \in \mathcal{P} \times \mathbb{R}^n$. The projection $(p,\xi) \in \mathcal{P} \times \mathbb{R}^n \mapsto \pi(p) \in \mathcal{M}$ induces the projection $\pi_{\mathcal{E}} : \mathcal{E} \to \mathcal{M}$.

Hence, we have shown that one can always associate a vector bundle $(\mathcal{E}, \mathcal{M}, \pi)$ to the principal bundle $(\mathcal{P}, \mathcal{M}, \pi)$. We will exploit this result to show that a connection on a principal fibre bundle $(\mathcal{P}, \mathcal{M}, \pi)$ may be used to define a covariant derivative on an associated vector bundle $(\mathcal{E}, \mathcal{M}, \pi)$.

Lie Algebras

It turns out that all C^∞ vector fields on a neighbourhood U of a manifold $\mathcal{M}$ form a Lie algebra, page 47 Schutz [63]. However, it is of greater interest to restrict our attention to a smaller subset of these vector fields that are related to the invariance properties of the manifold and its associated invariance groups, which are usually Lie groups. The study of the action of Lie groups on manifolds allows us to make this identification possible.

A very important feature of a Lie group is the existence of an associated Lie algebra, which is intimately related to the tangent space $T_e\mathcal{G}$ at the identity element of the Lie group. The left and right translations of the Lie group allow us to study the tangent space structure of the group manifold.

Let $\mathcal{M}$ be a smooth manifold. For two vector fields $X, Y \in \mathfrak{X}(\mathcal{M})$ we define the ***Lie bracket*** $[X,Y]_p$ of X and Y at $p \in \mathcal{M}$ by

$$[X,Y]_p(f) = X_p(Y(f)) - Y_p(X(f)) \tag{2.28}$$

where $f \in C^\infty(\mathcal{M})$.

The following definition allows us to identify the tangent space at the identity element of a Lie group with its Lie algebra. The following lemma is required before we proceed with the definition.

Lemma 2.1.1. *Let $\phi : \mathcal{M} \to \mathcal{N}$ be a smooth bijective map between two manifolds. If $X, Y \in \mathfrak{X}(\mathcal{M})$ are two vector fields on $\mathcal{M}$, then:*

1. $T\phi(X) \in \mathfrak{X}(\mathcal{N})$

2. *the map $T\phi : \mathfrak{X}(\mathcal{M}) \xrightarrow{\simeq} \mathfrak{X}(\mathcal{N})$ is a* ***Lie algebra homomorphism*** *in the sense that $[T\phi(X), T\phi(Y)] = T\phi([X, Y])$.*

Definition 2.1.5 (pages 3-4 Schmid [62]). *Consider the space $\mathfrak{X}(\mathcal{G})$ of all vector fields on $\mathcal{G}$, which is infinite dimensional with Lie bracket $[X, Y] = XY - YX \ \forall X, Y \in \mathfrak{X}(\mathcal{G})$. A vector field $X \in \mathfrak{X}(\mathcal{G})$ is called* ***left invariant*** *if and only if*

$$T_h L_g(X(h)) = X(gh) \ \forall g \in \mathcal{G} \tag{2.29}$$

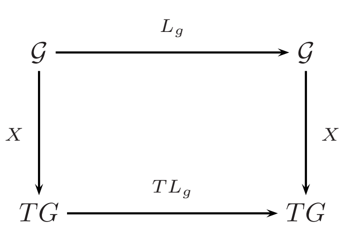

Figure 2.3: Left Invariant Maps

That is, X it is invariant under all left translations. $T_h L_g$ is a linear mapping which implies that a linear combination of invariant vector fields is invariant. If $X, Y \in \mathfrak{X}(\mathcal{G})$ are left invariant then

$$T_h L_g[X(h), Y(h)] = [T_h L_g X(h), T_h L_g Y(h)] = [X(gh), Y(gh)] \tag{2.30}$$

that is, $[X(gh), Y(gh)]$ is also left invariant. Hence, the space of left invariant vector fields $\mathfrak{X}_L(\mathcal{G})$ is a Lie subalgebra of $\mathfrak{X}(\mathcal{G})$. The ***Lie algebra*** *$\mathfrak{g}$ of a Lie group $\mathcal{G}$ is defined as $\mathfrak{g} = \mathfrak{X}_L(\mathcal{G})$.*

Let $\xi \in T_e\mathcal{G}$, then $X_\xi(g) = T_eL_g(\xi) \in T_g\mathcal{G}$ defines a left invariant vector field and $X_\xi(e) = \xi$. This defines an isomorphism $T_eG \xrightarrow{\cong} \mathfrak{X}_L(\mathcal{G})$, by Lem. 2.1.1. *Define the Lie bracket for any $\xi, \eta \in T_e\mathcal{G}$ by*

$$[\xi, \eta] = [X_\xi, X_\eta](e) \tag{2.31}$$

This bracket satisfies conditions for a Lie algebra, that is, the Lie bracket $[\cdot, \cdot]$ is:

1. *bilinear:* $[\alpha\xi_1 + \beta\xi_2] = \alpha[\xi_1, \xi_2] + \beta[\xi_1, \xi_2]\ \forall\alpha, \beta \in \mathbb{R}$
2. *skew symmetric:* $[\xi, \eta] = -[\eta, \xi]$
3. *Jacobi identity:* $[[\xi, \eta], \zeta] + [[\eta, \zeta], \xi] + [[\zeta, \xi], \eta] = 0$

We can therefore identify $\mathfrak{g} \cong T_e\mathcal{G}$ as a Lie algebra. Therefore, the Lie algebra $\mathfrak{g}$ of a Lie group $\mathcal{G}$ is defined as the space of left invariant vector fields $\mathfrak{X}_L(\mathcal{G})$ on $\mathcal{G}$ that is isomorphic to the tangent space $T_e\mathcal{G}$ at the identity element e, $\mathfrak{g} = \mathfrak{X}_L(\mathcal{G}) \xrightarrow{\cong} T_e\mathcal{G}$.

The following result is a consequence of the above definition. Consider any point g on the group manifold $\mathcal{G}$, the tangent space at this point $T_g\mathcal{G}$ can be canonically identified with the Lie algebra $\mathfrak{g}$ of $\mathcal{G}$. To elaborate, consider the differential map of the left translation $L_g : \mathcal{G} \to \mathcal{G}$ at the tangent space of $\mathcal{G}$ at e to the tangent space of $\mathcal{G}$ at g

$$T_eL_g : T_e\mathcal{G} \to T_g\mathcal{G},\ X(e) \mapsto T_eL_gX(e) := X(g) \tag{2.32}$$

If $dim(\mathcal{G}) = d$ and $X_1(e), \ldots, X_d(e)$ is a basis for $T_e\mathcal{G}$, the corresponding vector fields $X_1, \ldots, X_d$ are also a basis for $\mathfrak{g}$, Gockeler et al [20]. It follows that $T_eL_g : \mathfrak{g} \to T_g\mathcal{G}$. This result has far reaching implications in the following sense. Given a smooth manifold $\mathcal{M}$ and the map $\phi : \mathcal{G} \to \mathcal{M}$, one has $T_g\phi : T_g\mathcal{G} \to T_{\phi(g)}\mathcal{M}$.

With the identification $T_eL_g : \mathfrak{g} \to T_g\mathcal{G}$, we get the following composition of maps, Fig. 2.4,

$$T_g\phi \circ T_e L_g : \mathfrak{g} \to T_{\phi(g)}\mathcal{M} \tag{2.33}$$

which will play an important role in defining the connection on a principal fibre bundle.

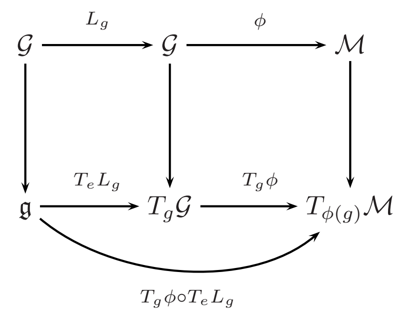

Figure 2.4: Composition map from $\mathfrak{g}$ to $T_{\phi(g)}\mathcal{M}$

We will now define the exponential map of a Lie group, which maybe used to construct a connection on a principal fibre bundle.

Definition 2.1.6 (page 3 Schmid [62]). *The* ***exponential map*** *is defined as follows:*

$$\exp : T_e\mathcal{G} \cong \mathfrak{g} \to \mathcal{G},\ \xi \mapsto \exp(\xi) := \phi_\xi(t) \mid_{t=1} \tag{2.34}$$

where $\phi_\xi : \mathbb{R} \to \mathcal{G}$ *is a one-parameter subgroup of* $\mathcal{G}$ *which describes the flow* $\phi_\xi(t)$ *of the left invariant vector field associated with* $\xi \in T_e\mathcal{G}$ *through* $e \in \mathcal{G}$*, that is,* $\phi_\xi(t) \mid_{t=0}= e$ *and* $\dot{\phi}_\xi(t) = X_\xi(\phi(t)) \in \mathfrak{X}_L(\mathcal{G})$*.*

Finally, we need to examine how a Lie group acts on itself.

Definition 2.1.7. *We define the* ***adjoint representation of a Lie group*** $\mathcal{G}$ *by*

$$Ad : \mathcal{G} \to GL(\mathfrak{g}),\ g \mapsto Ad(g) := Ad_g = L_g \circ R_{g^{-1}} \tag{2.35}$$

where $Ad_g : \mathcal{G} \to \mathcal{G}$. The differential map $TAd : \mathfrak{g} \to gl(\mathfrak{g})$, which we denote by ad, is called the ***adjoint representation of a Lie algebra*** possesses the following properties:

1. $ad(X)Y = [X, Y]$

2. $Ad(\exp X) = \exp(adX)$

Connections on Principal Bundles

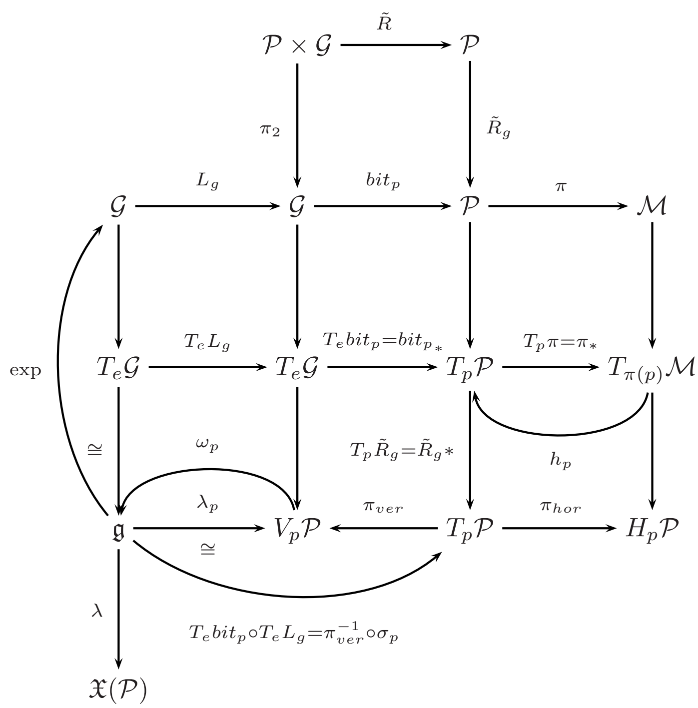

Figure 2.5: Connections on Principal Fibre Bundles

The following discussion on connections is based on pages 5-6 Kunzinger et al [41]. Consider a principal fibre bundle $(\mathcal{P}, \mathcal{M}, \pi)$, refer to Fig. 2.5.

Definition 2.1.8. *A tangent vector $X_p \in T_p\mathcal{P}$ is called **vertical** if $T_p\pi(X_p) = 0$, we denote them by X_p^{ver}. The set of all vertical tangent vectors at $p \in \mathcal{P}$ is denoted by the set $V_p\mathcal{P}$ which is defined as follows:*

$$V_p\mathcal{P} = \{X_p^{ver} \in T_p\mathcal{P} \mid T_p\pi(X_p^{ver}) = \pi_* X_p^{ver} = 0\} = Ker(T_p\pi) \tag{2.36}$$

These vectors are tangent to the fibre $\mathcal{P}_x = \pi^{-1}(x)$, which is a submanifold of $\mathcal{P}$. Hence, the vertical space $V_p\mathcal{P}$ at $p \in \mathcal{P}$ is the tangent space $T_p\mathcal{P}_x$ to the fibre $\mathcal{P}_x$ at p.

We now show that this space is isomorphic to the Lie algebra of the structure group $\mathcal{G}$, that is $V_p\mathcal{P}\xrightarrow{\cong}\mathfrak{g}$.

Consider the embedding whose image yields an orbit in $\mathcal{P}$

$$bit_p : \mathcal{G} \to \mathcal{P},\ g \mapsto bit_p(g) \tag{2.37}$$

The structure group $\mathcal{G}$ of $(\mathcal{P}, \mathcal{M}, \pi)$ has a well defined free action on $\mathcal{P}$, namely $\tilde{R} : \mathcal{P} \times \mathcal{G} \to \mathcal{P}$ where $(p, g) \mapsto \tilde{R}(p, g) = \tilde{R}_g(p) := pg$, allows us to make the identification $bit_p(g) = \tilde{R}_g(p) = pg$. The associated differential of bit_p induces an injective map

$$bit_{p*} = T_e bit_p : T_e\mathcal{G} \to T_p\mathcal{P} \tag{2.38}$$

and the quotient space by the image of $T_e bit_p$ is mapped isomorphically onto $T_{\pi(p)}\mathcal{M}$ by $\pi_* = T_p\pi$ of π. From the previous section, Equ.(2.33), we identified a map $T_e bit_p \circ T_e L_g$ that maps left invariant vector fields of $\mathfrak{g}$ to tangent vectors of $T_p\mathcal{P}$. This natural identification leads to the idea that it is possible to construct a map that maps from $\mathfrak{g}$ to $V_p\mathcal{P}$.

Let $A \in \mathfrak{g}$, then $\phi_A(t) = \exp(tA)$ is a 1-parameter subgroup of $\mathcal{G}$ generated by $T_e L_g A$. Then by the right action

$$\tilde{R}_{\exp(tA)} p = p \exp(tA) \tag{2.39}$$

a curve through p is defined in $\mathcal{P}$. Since the group action moves points along a fibre, we have

$$\pi(p) = \pi(\tilde{R}_{\exp(tA)} p) = x \tag{2.40}$$

Hence, this curve lies in $\mathcal{G}_x$. Now define a vector $A^* \in T_p\mathcal{P}$ by

$$A^* f(p) = \frac{d}{dt} f(p \exp(tA)) \mid_{t=0} \tag{2.41}$$

where $f : \mathcal{P} \to \mathbb{R}$ is an arbitrary function. The vector A^* is tangent to the fibre, therefore $A^* \in V_p\mathcal{P}$ is a vertical vector field. A^* is called the ***fundamental vector field*** generated by A. There is a vector space isomorphism $\lambda : \mathfrak{g} \xrightarrow{\cong} \mathfrak{X}(\mathcal{P})$ that maps the vector field $A \in \mathfrak{g}$ to the fundamental vector field $A^* \in \mathfrak{X}(\mathcal{P})$, $A \mapsto \lambda(A) := A^*$. The above map is a Lie algebra homomorphism since $\lambda([A, B]) = [\lambda(A), \lambda(B)]$.
Note that λ_p can be identified with $T_e bit_p(X(e))$. We now prove that this map gives rise vectors that are tangent to the fibres of $\mathcal{P}$. Since $\forall g \in \mathcal{G}$

$$\pi \circ bit_p(g) = \pi(pg) = \pi(\tilde{R}_g p) = \pi(p) \in \mathcal{M} \tag{2.42}$$

we conclude that $\pi \circ bit_p$ is constant on $\mathcal{G}$, consequently,

$$\begin{aligned} T_p\pi \circ \lambda_p &= T_p\pi \circ T_e bit_p &(2.43)\\ &= T_e(\pi \circ bit_p) &(2.44)\\ &= 0 &(2.45) \end{aligned}$$

We see that the Lie algebra homomorphism has the important property of identifying left-invariant vector fields on $\mathcal{G}$, which are Lie algebra elements with fundamental vector fields on $\mathcal{P}$. The restriction $\lambda_p : \mathfrak{g} \xrightarrow{\cong} V_p\mathcal{P}$ is a linear isomorphism. The right action of $\mathcal{G}$ on $\mathcal{P}$ also induces a corresponding right action on the fundamental vector fields which satisfies the condition

$$\tilde{R}_{g^*}(\lambda(A)) = \lambda(Ad(g^{-1})A) \tag{2.46}$$

where Ad denotes the adjoint representation $Ad : \mathcal{G} \to L(\mathfrak{g}, \mathfrak{g})$ of $\mathcal{G}$ on $\mathfrak{g}$.

Definition 2.1.9 (page 6 Kunzinger et al [41]). *A **connection form** $\omega \in \Omega^1(\mathcal{P}, \mathfrak{g})$ is a smooth 1-form on $\mathcal{P}$ with values in the Lie algebra $\mathfrak{g}$ which satisfies the following conditions*

1. $\omega_p \circ T_e bit_p = id_{\mathfrak{g}}$

2. $\forall X_p^{ver} \in V_p\mathcal{P}$, $\omega_p(X_p^{ver}) = \lambda_p^{-1}(X_p^{ver})$

3. $\forall X_p^{ver} \in T_{\tilde{R}_g p}\mathcal{P}$, $\omega_{\tilde{R}_g p}(X_p^{ver}) = Ad(g^{-1})\omega_p(\tilde{R}_{g*}^{-1} X_p^{ver})$

Definition 2.1.10 (page 6 Kunzinger et al [41]). *Given a connection 1-form ω we call a tangent vector $X_p \in T_p\mathcal{P}$ **horizontal** if it is annihilated by ω, we denote such tangent vectors by X_p^{hor}. The set of all horizontal tangent vectors at $p \in \mathcal{P}$ is denoted by $H_p\mathcal{P}$ and is defined by*

$$H_p\mathcal{P} = \{X_p^{hor} \in T_p\mathcal{P} \mid \omega_p(X_p^{hor}) = 0\} = Ker(\omega_p) \tag{2.47}$$

Kunzinger et al [41] claims that since $\omega_p \circ \lambda_p = id$, the null space of $\omega_p : T_p\mathcal{P} \to \mathfrak{g}$ is an m-dimensional vector space transverse to the fibre. This leads to a natural and unique decomposition of the vector X_p into vertical and horizontal components such that $X_p = X_p^{ver} + X_p^{hor}$, where $X_p^{ver} \in V_p\mathcal{P}$ and $X_p^{hor} \in H_p\mathcal{P}$. Consequently, it is possible to define projection maps $\pi_{ver} : T_p\mathcal{P} \to V_p\mathcal{P}$ and $\pi_{hor} : T_p\mathcal{P} \to H_p\mathcal{P}$ that map vectors on $T_p\mathcal{P}$ to their respective vertical and horizontal components on $V_p\mathcal{P}$ and $H_p\mathcal{P}$. The condition Equ.(2.46) and (i) above ensures that condition (ii) holds for vertical vectors. Condition (ii) leads to the requirement that $\tilde{R}_{g^*}$ takes horizontal vectors to horizontal vectors. This leads to an alternate definition for connections.

Definition 2.1.11 (page 6 Kunzinger et al [41]). *A **connection** on a principal bundle $(\mathcal{P}, \mathcal{M}, \pi)$ is a family of subspaces $H_p\mathcal{P} \subset T_p\mathcal{P}$ called horizontal tangent spaces, such that:*

1. *$H_p\mathcal{P}$ is complementary to $V_p\mathcal{P}$ in $T_p\mathcal{P}$, that is, $T_p\mathcal{P} \simeq V_p\mathcal{P} \oplus H_p\mathcal{P}$*

2. *$H_p\mathcal{P}$ depends smoothly on $\mathcal{P}$, that is, $H_p\mathcal{P}$ is locally spanned by smooth vector fields on $\mathcal{P}$*

3. $T_p\tilde{R}_g(H_p\mathcal{P}) = H_{\tilde{R}_g p}\mathcal{P} = H_{pg}\mathcal{P}\ \forall g \in \mathcal{G}$

The following definition is equivalent to the above. It gives rise to the notion of the horizontal lift of the connection, which plays a central role in stochastic differential geometry.

Definition 2.1.12 (page 8 Morrison [56]). *A horizontal lift of the connection is a smooth choice* $\forall p \in \mathcal{P}$ *of a linear map* $h_p : T_{\pi(p)}\mathcal{M} \to T_p\mathcal{P}$ *so that:*

$$T_p\pi \circ h_p = id_{T_{\pi(p)}\mathcal{M}} \tag{2.48}$$

$$h_{pg} = T_p\tilde{R}_g(h_p) \tag{2.49}$$

Horizontal Lifts of Curves and Parallel Translation

Given a connection form ω in $(\mathcal{P}, \mathcal{M}, \pi)$ we show that it is possible to lift a curve γ on the base manifold $\mathcal{M}$ to a horizontal curve $\bar{\gamma}$ on the bundle $\mathcal{P}$.

Definition 2.1.13 (page 381 Nakahara [58]). *Let* $(\mathcal{P}, \mathcal{M}, \pi)$ *be a* $\mathcal{G}$ *bundle and let* $\gamma : [0,1] \to \mathcal{M}$ *be a curve on* $\mathcal{M}$*. A curve* $\bar{\gamma} : [0,1] \to \mathcal{P}$ *is said to be the* ***horizontal lift of a curve*** γ *if:*

1. $\pi \circ \bar{\gamma} = \gamma$ *and*

2. *the tangent vector to* $\bar{\gamma}(t)$ *always belongs to* $H_{\bar{\gamma}(t)}\mathcal{P}$ *or equivalently,* $\omega(\dot{\bar{\gamma}}(t)) = 0$

Before we define parallel transport, we need to establish the uniqueness of horizontal lifts.

Theorem 2.1.1 (page 381 Nakahara [58]). *Let* $\gamma : [0,1] \to \mathcal{M}$ *be a curve in* $\mathcal{M}$ *and let* $u_0 \in \mathcal{P}_{\gamma(0)} = \pi^{-1}(\gamma(0))$*, then* $\exists \bar{\gamma}(t)$ *in* $\mathcal{P}$ *such that* $\bar{\gamma}(0) = u_0$*.*

Let $\gamma : [0,1] \to \mathcal{M}$ be a curve. Consider the point $u_0 \in \mathcal{P}_{\gamma(0)} = \pi^{-1}(\gamma(0))$. Then, there exists a unique horizontal lift $\bar{\gamma}(t)$ of $\gamma(t)$ through u_0 and hence a unique point $u_1 = \bar{\gamma}(1) \in \mathcal{P}_{\gamma(1)} = \pi^{-1}(\gamma(1))$. The point u_1 is called the ***parallel transport*** of u_0 along the curve γ. This defines the map $\Gamma(\bar{\gamma}) : \mathcal{P}_{\gamma(0)} = \pi^{-1}(\gamma(0)) \to \mathcal{P}_{\gamma(1)} = \pi^{-1}(\gamma(1))$ such that $u_0 \mapsto u_1$.

Covariant Derivatives

Previously, we have shown that associated with every principal fibre bundle $(\mathcal{P}, \mathcal{M}, \pi)$ is a vector bundle $(\mathcal{E}, \mathcal{M}, \pi)$. We will now demonstrate how a principal fibre bundle $(\mathcal{P}, \mathcal{M}, \pi)$ together with its connection may be used to define a covariant derivative on an associated vector bundle $(\mathcal{E}, \mathcal{M}, \pi)$. In this section, we follow Kunzinger et al [41] pages 14-15 closely.

Consider the points x_0 and x_1 that lie in the neigbourhood U of $\mathcal{M}$. Let $\Phi : \pi^{-1}(U) \to U \times \mathcal{G}$ and $\Psi : \pi_{\mathcal{E}}^{-1}(U) \to U \times \mathbb{R}^n$ be the local trivializations on U. Now consider a curve $\gamma : [0,1] \to \mathcal{M}$ that connects $x_0, x_1 \in U$ in the following manner: $\gamma(0) = x_0$ and $\gamma(1) = x_1$. Let $\bar{\gamma}$ denote the horizontal lift of γ to $\mathcal{P}$. Then we may use γ to define the parallel transport

$$\Gamma(\bar{\gamma}) : \pi_{\mathcal{E}}^{-1}(x_0) \xrightarrow{\cong} \pi_{\mathcal{E}}^{-1}(x_1),\ V_0 \mapsto \Gamma(\bar{\gamma})(V_0) = V_1 \tag{2.50}$$

by requiring that $\xi_1 = \rho(g_1 g_0^{-1})\xi_0$ such that $\Psi(V_i) = (x_i, \xi_i)$ and $\Phi(\bar{\gamma}(i)) = (x_i, g_i)$ for $i = 0, 1$. Note that V_1 is only dependent on the connection on $\mathcal{P}$ and the curve γ. We use this concept to define the covariant derivative of a vector field $V(x) \in \Gamma^\infty(\mathcal{M}, \mathcal{E})$ in the direction of the tangent to the curve at the point $\gamma(0)$ as

$$\nabla_{\dot{\gamma}(0)} V(\gamma(0)) = \lim_{h \to 0} \frac{1}{h} [\Gamma(\bar{\gamma})(V(\gamma(h))) - V(\gamma(0))] \tag{2.51}$$

$\nabla_{\dot{\gamma}(0)} V(\gamma(0))$ only depends on the direction of the tangent $X = \dot{\gamma}(0)$ and not the curve γ. So we use the above definition to define the quantity $\nabla_X V$ at $\gamma(0)$. By applying this definition, we may define the covariant of the field $V \in \Gamma^\infty(\mathcal{M}, \mathcal{E})$ with respect to the vector field $X \in \mathfrak{X}(\mathcal{M})$ by the map

$$\nabla : \mathfrak{X}(\mathcal{M}) \times \Gamma^\infty(\mathcal{M}, \mathcal{E}) \to \Gamma^\infty(\mathcal{M}, \mathcal{E}),\ (X, V) \mapsto \nabla(X, V) := \nabla_X V \tag{2.52}$$

The covariant derivative defined this way satisfies the following properties:

1. $\nabla_X V$ is $\mathbb{R}$-linear in V

2. $\nabla_X V$ is $C^\infty(\mathcal{M})$-linear in X

3. $\nabla_X(\lambda V) = \lambda \nabla_X V + X(\lambda) \cdot V$, $\forall \lambda \in C^\infty(\mathcal{M})$

Geometry of the Frame Bundles

Let $\mathcal{M}$ be an n-dimensional manifold and let $p \in \mathcal{M}$. Let $T_x\mathcal{M}$ denote the tangent space at $x \in \mathcal{M}$. The $\mathbb{R}$-linear isomorphism is called a ***frame*** at x

$$u : \mathbb{R}^n \xrightarrow{\cong} T_x\mathcal{M}, \; e \mapsto ue \tag{2.53}$$

The frame has the property of mapping the unit vectors $e \in \mathbb{R}^n$ to tangent vectors $ue \in T_x\mathcal{M}$, which form a basis for $T_x\mathcal{M}$. We define the collection of all frames at $x \in \mathcal{M}$ by the set

$$\mathcal{F}(\mathcal{M})_x = \{u \mid u : \mathbb{R}^n \xrightarrow{\cong} T_x\mathcal{M} \text{ where } x \in \mathcal{M}\} \tag{2.54}$$

We define the action of the general linear group $GL(n;\mathbb{R})$ on $\mathcal{F}(\mathcal{M})_x$ by $u \mapsto ug$, where ug is interpreted as a composition

$$\mathbb{R}^n \xrightarrow{g} \mathbb{R}^n \xrightarrow{u} T_x\mathcal{M} \tag{2.55}$$

The ***frame bundle***, which is defined by the union of all frames

$$\mathcal{F}(\mathcal{M}) = \cup_{x \in \mathcal{M}} \mathcal{F}(\mathcal{M})_x \tag{2.56}$$

is a $(n + n^2)$-dimension manifold. The projection map for this bundle

$$\pi : \mathcal{F}(\mathcal{M}) \to \mathcal{M} \tag{2.57}$$

identifies the point $x \in \mathcal{M}$ associated with the tangent space $T_x\mathcal{M}$ of a particular frame $U : \mathbb{R}^n \to T_x\mathcal{M}$. $GL(n;\mathbb{R})$ can be taken as the standard fibre for this bundle because each linear frame determines an element of $GL(n;\mathbb{R})$ and each element of

$GL(n;\mathbb{R})$ determines a linear frame. Therefore, the frame bundle $\mathcal{F}(\mathcal{M})$ of an n-dimensional manifold $\mathcal{M}$ is a principal bundle with structure group $GL(n;\mathbb{R})$, page 37 Hsu [25], with the tangent bundle $T\mathcal{M}$ being an associated bundle of $\mathcal{F}(\mathcal{M})$ over $\mathcal{M}$

$$T\mathcal{M} = \mathcal{F}(\mathcal{M}) \times_{GL(n;\mathbb{R})} \mathbb{R}^n,\ (u,e) \mapsto ue \tag{2.58}$$

The connection ∇ on $\mathcal{F}(\mathcal{M})$ gives rise to the following decomposition of $T_u\mathcal{F}(\mathcal{M})$

$$T_u\mathcal{F}(\mathcal{M}) = V_u\mathcal{F}(\mathcal{M}) \oplus H_u\mathcal{F}(\mathcal{M}) \tag{2.59}$$

where $V_u\mathcal{F}(\mathcal{M})$ and $H_u\mathcal{F}(\mathcal{M})$ denote the vertical and horizontal spaces of $T_u\mathcal{F}(\mathcal{M})$. Let $e \in \mathbb{R}^n$, the vector field H_e on $\mathcal{F}(\mathcal{M})$ defined at $u \in \mathcal{F}(\mathcal{M})$ by

$$H_e(u) = (ue)^{hor} \tag{2.60}$$

is the horizontal lift of $ue \in T_{\pi u}\mathcal{M}$. If e_i are the coordinate unit vectors of $\mathbb{R}^n$, where $i = 1, \dots, n$, then $H_i = H_{e_i}$ are the ***fundamental horizontal vector fields*** of $\mathcal{F}(\mathcal{M})$ that span $H_u\mathcal{F}(\mathcal{M})$ at each $u \in \mathcal{F}(\mathcal{M})$. The action of $GL(n;\mathbb{R})$ preserves the fundamental horizontal fields

$$T_uR_gH_e(u) = H_{ge}(gu),\ u \in \mathcal{F}(\mathcal{M}) \tag{2.61}$$

where $T_uR_g : T_u\mathcal{F}(\mathcal{M}) \to T_{ug}\mathcal{F}(\mathcal{M})$.

Example 2.1.4 (pages 41-42 Schutz [63]). *Consider the* 1*-dimensional manifold* $\mathcal{M} = S^1$*. Without having to work through all the detail, it can be shown that the frame bundle* $\mathcal{F}(\mathcal{M})$ *for* S^1 *has the same structure as the tangent bundle* $T\mathcal{M} = TS^1$*, which is isomorphic to* $S^1 \times \mathbb{R}^1$*. The structure group for* $\mathcal{F}(\mathcal{M})$ *is* $\mathcal{G} = \mathbb{R} - \{0\}$*, which is a group under multiplication. The bundle* $\mathcal{F}(\mathcal{M})$ *has a fibre* $\mathcal{F}$ *that comprises of the set of all bases for* TS^1*. Since* $\mathcal{F}(\mathcal{M})$ *is a principal fibre bundle, the frame bundle for* S^1 *has fibres homeomorphic to its structure group* $\mathcal{G}$*. This implies that* $\mathcal{F}$ *may be identified with* $\mathbb{R} - \{0\}$*.*

Gauge Transformation

Heuristically, a gauge can be thought of as a degree of freedom within a mathematical theory that has no physical observable effect. A gauge transformation acts on this degree of freedom. Sometimes, gauge transformations are used to simplify a complex system without modifying any of its physical observable properties. We will exploit this concept to obtain a simplified solution to the Zakai equation. In this section we describe the essence of gauge theory in context to fibre bundles, page 459 Mitter et al [53] and page 32 Svetlichny [66].

Consider a principal $\mathcal{G}$-bundle over $\mathcal{M}$ $(\mathcal{P}, \mathcal{M}, \pi)$. A ***gauge-transformation*** is a bundle isomorphism

$$\phi : \mathcal{P} \xrightarrow{\cong} \mathcal{P}, \; p \mapsto \phi(p) \tag{2.62}$$

that commutes with the right action. The free right action of $\mathcal{G}$ on $\mathcal{P}$ implies that there exist a map

$$\gamma : \mathcal{P} \to \mathcal{G}, \; p \mapsto \gamma(p) \tag{2.63}$$

such that

$$\phi(p) = p \cdot \gamma(p) \tag{2.64}$$

To commute with the right action, it is necessary and sufficient that

$$(p \cdot g) \cdot \gamma(p \cdot g) = (p \cdot \gamma(p)) \cdot g \tag{2.65}$$

which implies that

$$\gamma(p \cdot g) = g^{-1}\gamma(p)g \tag{2.66}$$

The ***gauge-group*** of $\mathcal{P}$, which we denote by $\mathcal{G}(\mathcal{P})$, is the set of all gauge-transformations

$$\mathcal{G}(\mathcal{P}) = \{\gamma : \mathcal{P} \to \mathcal{G} \mid \gamma(p \cdot g) = g^{-1}\gamma(p)g\} \tag{2.67}$$

Given that $\mathcal{M}$ is a manifold and $\mathcal{G}$ is a Lie group, $\mathcal{G}(\mathcal{P})$ is an infinite-dimensional Lie group.

We can study the Lie algebra of the gauge group $\mathcal{G}(\mathcal{P})$ by considering the one-parameter family $\gamma_t(p)$ of gauge-transformations defined by

$$\gamma_t(p) := \exp^{t\theta(p)} \tag{2.68}$$

where $\theta(p) \in \mathfrak{g}$. Then $\forall t$, Equ.(2.66) is equivalent to

$$\theta(p \cdot g) = Ad_{g^{-1}}\theta(p) \tag{2.69}$$

The maps $\theta : \mathcal{P} \to \mathfrak{g}$ constitute the Lie algebra of $\mathcal{G}(\mathcal{P})$ and are called ***infinitesimal gauge-transformations***.

2.2 Riemannian Differential Geometry

2.2.1 Riemannian Metric

Let $\mathcal{M}$ be a smooth manifold. A ***Riemannian metric*** $g = (g_{ij})$ on $\mathcal{M}$ is a smooth $(0, 2)$ tensor field $g : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to C^\infty(\mathcal{M})$ such that $\forall p \in \mathcal{M}$ the restriction $g_p = g \mid_{T_p\mathcal{M}\otimes T_p\mathcal{M}}: T_p\mathcal{M} \otimes T_p\mathcal{M} \to \mathbb{R}$ has the following properties, page 281 Ikeda et al [26]:

1. g is symmetric, that is, $g_p(X, Y) = g_p(Y, X)$

2. g is positive definite, that is, $g_p(X, X) \geq 0$ $\forall p$ and $v \in \mathbb{R}^d$, $v \neq 0$

3. g defines an inner product on each tangent space $T_p\mathcal{M}$ by

   $$\langle \omega, Y \rangle = g(X_\omega, Y)$$

   where $\omega \in T_p^*\mathcal{M}$, $X_\omega \in T_p\mathcal{M}$ and $\langle \cdot, \cdot \rangle : T_p^*\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}$

We adopt the following notation when basis tangent vectors are considered

$$g_{ij} = g(\partial_i, \partial_j), \text{ where } \partial_i = \tfrac{\partial}{\partial x_i} \text{ and } \partial_j = \tfrac{\partial}{\partial x_j} \tag{2.70}$$

The pair $(\mathcal{M}, g)$ is called a ***Riemannian manifold***. Geometric properties of $(\mathcal{M}, g)$ which depend on the metric g are called ***intrinsic*** (or ***metric***) properties.

Consider the n-dimensional manifold $\mathcal{M} = \mathbb{R}^n$. The Euclidean metric of $\mathbb{R}^n$ is given by $\delta = diag(1, \ldots, 1)$. Then $(\mathcal{M}, g)$=$(\mathbb{R}^n, \delta)$ is a Riemannian manifold.

The Induced Metric

Let $\mathcal{M}$ be an m-dimensional submanifold of an n-dimensional Riemmanian manifold $\mathcal{N}$ with metric $g_{\mathcal{N}}$. Consider the map $f : \mathcal{M} \to \mathcal{N}$ which induces the submanifold structure of $\mathcal{M}$. The pullback f^* induces the metric $g_{\mathcal{M}} = f^* g_{\mathcal{N}}$ on $\mathcal{M}$, which has components

$$g_{\mathcal{M}ij}(x) = g_{\mathcal{M}kl}(f(x)) \frac{\partial f^k}{\partial x^i} \frac{\partial f^l}{\partial x^j} \tag{2.71}$$

The induced metric plays a very import role in the calculus of manifolds, especially when we regard such manifolds as submanifolds that are embedded in higher dimensional Euclidean spaces. This enables us to take the metric properties of curves and surfaces in $\mathbb{R}^n$ and study them in an intrinsic manner on the respective submanifolds.

Example 2.2.1. *Consider the* 1*-dimensional manifold* $\mathcal{M} = S^1$*. We may regard* S^1 *as a submanifold of the Riemmanian manifold* $(\mathbb{R}^2, \delta)$*. The embedding function is given by*

$$f : S^1 \to \mathbb{R}^2,\ (r, \theta) \mapsto f(r, \theta) := (r \cos\theta, r \sin\theta) \tag{2.72}$$

Then the induced metric on S^1 *is given by*

$$
\begin{aligned}
g_{ij}dx^i \otimes dx^j &= \delta_{kl}\frac{\partial f^k}{\partial x^i}\frac{\partial f^l}{\partial x^j} \\
&= dr \otimes dr + r^2 d\theta \otimes d\theta
\end{aligned}
\tag{2.73}
$$

2.2.2 The Connection

Let $(\mathcal{E}, \mathcal{M}, \pi)$ be a smooth vector bundle over $\mathcal{M}$. Let

$$
\nabla : \mathfrak{X}(\mathcal{M}) \times \mathcal{T}_1^0(\mathcal{M}) \to \mathcal{T}_1^0(\mathcal{M}),\ (X, Y) \mapsto \nabla(X, Y) := \nabla_X Y \tag{2.74}
$$

be the connection on $(\mathcal{E}, \mathcal{M}, \pi)$.

The operator ∇_X is called covariant differentiation with respect to X. The components of the connection ∇ are defined in terms of the set of functions $\{\Gamma^i_{jk}(x)\}$ such that in local coordinates

$$
\nabla_{\partial_i}\partial_j = \Gamma^k_{ij}(x)\frac{\partial}{\partial x_k},\ \partial_i = \frac{\partial}{\partial x_i} \tag{2.75}
$$

We say that the connection is symmetric if $\Gamma^k_{ij} = \Gamma^k_{ji}$. In local coordinates we may express the covariant derivative of Y with respect to X as, page 278 Ikeda et al [26],

$$
\nabla_X Y = \left[X^i(x)\frac{\partial}{\partial x_i}Y^k(x) + \Gamma^k_{ij}(x)X^i(x)Y^i(x)\right]\frac{\partial}{\partial x_k} \tag{2.76}
$$

where

$$
X = X^i(x)\tfrac{\partial}{\partial x_i} \text{ and } Y = Y^i(x)\tfrac{\partial}{\partial x_i} \tag{2.77}
$$

Consider a smooth curve $c : I = [t_0, t_1] \to \mathcal{M}$. Let $X(t) \in T_{c(t)}\mathcal{M}\ \forall t \in I$. We say that $X(t)$ is parallel along c, with respect to ∇, if

$$
\frac{d}{dt}X^i(t) + \Gamma^i_{kj}(c(t))X^j(t)\frac{d}{dt}c(t) = 0,\ \forall t \in I \tag{2.78}
$$

For $t_0, t_1 \in I$, where $t_0 \leq t_1$, $X(t_1)$ is uniquely determined from $X(t_0)$ by parallel displacement along $c(t)$.

2.2.3 The Levi-Civita Connection

Consider the tangent bundle $(T\mathcal{M}, \mathcal{M}, \pi)$, if g is a Riemannian metric on $\mathcal{M}$, we define the ***torsion*** $T : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$ of ∇ by

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] \tag{2.79}$$

where $[\cdot, \cdot]$ is the Lie bracket on $\mathfrak{X}(\mathcal{M})$.

Definition 2.2.1. *There is a unique connection ∇ on $(T\mathcal{M}, \mathcal{M}, \pi)$ called the **Levi-Civita connection** which is compatible with g in the following sense:*

$$\nabla_Z g(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla_Z Y), \ \forall X, Y, Z \in \mathfrak{X}(\mathcal{M}) \tag{2.80}$$

and which is torsion free $T(X, Y) = 0 \ \forall X, Y \in \mathfrak{X}(\mathcal{M})$.

The Levi-Civita connection ∇ is an intrinsic object since, by definition, it is determined by the metric g.

This is in essence the fundamental theorem of Riemannian geometry. On $(T\mathcal{M}, \mathcal{M}, \pi)$, the connection ∇ can be thought of as a rule for differentiating a vector field $Y \in \mathfrak{X}(\mathcal{M})$ in the direction of another $X \in \mathfrak{X}(\mathcal{M})$. If g is a Riemannian metric on $\mathcal{M}$, the Levi-Civita connection is the only way of doing this in a metric and torsion-free manner. The corresponding Christoffel symbols for the Levi-Civita connection are given by, page 282 Ikeda et al [26]:

$$\Gamma^k_{ij} = \frac{1}{2}\left[\frac{\partial}{\partial x_i} g_{mj} + \frac{\partial}{\partial x_j} g_{im} - \frac{\partial}{\partial x_m} g_{ij}\right] g^{km} \tag{2.81}$$

Example 2.2.2 (page 248 262 Nakahara [58]). *The induced metric on $\mathcal{M} = S^1$ is given by Equ.(2.73). The non vanishing components of the Levi-Civita connection coefficients are:*

$$\Gamma^{\theta}_{r\theta} = 1/r \tag{2.82}$$
$$\Gamma^{\theta}_{\theta r} = 1/r \tag{2.83}$$
$$\Gamma^{r}_{\theta\theta} = -r \tag{2.84}$$

Orthonormal Frame Bundle

Consider a Riemannian manifold $\mathcal{M}$ and metric g. We now restrict ourselves to a smaller set of frames called ***orthonormal frames***. Let $\mathcal{O}(\mathcal{M})$ denote the ***orthonormal frame bundle***. In keeping with the general theory of frame bundles, the element $u \in \mathcal{O}(\mathcal{M})$ is a Euclidean isometry $u : \mathbb{R}^n \to T_x\mathcal{M}$ which we call the frame at $x \in \mathcal{M}$. In this case, the action group is the orthogonal group $O(n)$, which is a subset of the general linear group $GL(n;\mathbb{R})$. Therefore, $\mathcal{O}(\mathcal{M})$ is a principal fibre bundle with structure group $O(n)$. The Riemannian connection on $\mathcal{M}$ splits the tangent spaces $T_u\mathcal{O}(\mathcal{M})$ into vertical $V_u\mathcal{O}(\mathcal{M})$ and horizontal $H_u\mathcal{O}(\mathcal{M})$ subspaces, where

$$H_u\mathcal{O}(\mathcal{M}) = \left\{ X = a^i \left(\frac{\partial}{\partial x_i}\right)_x - \Gamma^i_{kl}(x) e^l_j a^k \left(\frac{\partial}{\partial e^i_j}\right) \mid (a^i) \in \mathbb{R}^d \right\} \tag{2.85}$$

which is independent of the choice of local coordinates (x^i, e^i_j), page 279 Ikeda et al [26]. Given the vector field $X \in \mathfrak{X}(\mathcal{M})$, there exists a unique $X^{hor} \in \mathfrak{X}(\mathcal{O}(\mathcal{M}))$ such that $X^{hor}_u \in H_u\mathcal{O}(\mathcal{M})$ if the horizontal lift of $\pi(X^{hor}_u) = X_{\pi u}$ $\forall u \in \mathcal{O}(\mathcal{M})$. In local coordinates we have, page 280 Ikeda et al [26],

$$X^{hor} = X^i(x)\frac{\partial}{\partial x_i} - \Gamma^q_{ij}(x) X^i(x) e^j_p \frac{\partial}{\partial e^q_p} \tag{2.86}$$

where $X = X^i(x)\frac{\partial}{\partial x_i}$. Similarly, given a smooth curve $c(t)$ on $\mathcal{M}$, we can lift $c(t)$ to $\tilde{c}(t)$ on $\mathcal{O}(\mathcal{M})$ if

1. $\frac{d\tilde{c}}{dt}(t)$ is horizontal and

2. $\pi(\tilde{c}(t)) = c(t)$

The curve $\tilde{c}(t)$ is given by $\tilde{c}(t) = (c(t), e(t))$ where the frames e are parallel transported along $c(t)$ - the basis of stochastic differential geometry if founded on this fundamental notion. In local coordinates (x^i, e^i_j), the ***fundamental horizontal vector fields*** $H_j \in \mathfrak{X}(\mathcal{O}(\mathcal{M}))$ may be expressed as, page 280 Ikeda et al [26],

$$H_j = e^i_j \frac{\partial}{\partial x^i} - \Gamma^q_{kl} e^k_j e^l_p \frac{\partial}{\partial e^q_p},\ j = 1, \ldots, n \tag{2.87}$$

Chapter 3

CALCULUS OF SEMIMARTINGALES

The mathematical tools that describe the time evolution of physical systems is differential equations. However, in most circumstances such models have to be refined by incorporating random behaviour to model environmental and/or measurement effects. The associated tools for such models are stochastic processes. Most stochastic processes can be expressed as the sum of a mean motion and a fluctuation from the mean. In this chapter we will introduce the necessary theory to understand the properties of semimartingales, which play an important role in the stochastic differential geometry and filtering theory. A basic understanding of measure theory and probability theory is assumed, however, important definitions and results will be stated to improve the readability of the thesis and to set the notation for other chapters that follow. For further details Brzezniak et al [7], Oksendal [59], Karatzas [33] and Kunita [40] should be consulted.

3.1 Concepts in Probability

3.1.1 Introduction

Probability Spaces and Random Variables

Consider a nonempty set $\Omega \neq \emptyset$. The set of all subsets of Ω is called the powerset of Ω, which we denote by 2^{Ω}. We call a nonempty set $\mathcal{F} \subseteq 2^{\Omega}$ a class. $\mathcal{F}$ is called a ***sigma-algebra***, denoted by σ-algebra , if:

1. $\Omega \in \mathcal{F}$ and

2. is closed under the operations of:

(a) countable unions: $A_i \in \mathcal{F} \Longrightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$, where $i = 1, 2, \ldots$ and

(b) complements: $A, B \in \mathcal{F} \Longrightarrow B \setminus A \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a ***measurable space***. The elements of Ω are called samples and those of $\mathcal{F}$ are called events. A map $P : \mathcal{F} \to [0, 1]$ is called a ***probability measure*** on $(\Omega, \mathcal{F})$ if:

1. $P(\emptyset) = 0$

2. $P(\Omega) = 1$

3. if $A_i, A_j \in \mathcal{F}$ are disjoint $A_i \cap A_j = \emptyset$ for $i \neq j$, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ where $i, j = 1, 2, \ldots$ and $i \neq j$

The triplet $(\Omega, \mathcal{F}, P)$ is called a ***probability space***. We assume that $(\Omega, \mathcal{F}, P)$ is ***complete*** in the sense that $B \subseteq A \in \mathcal{F}$ with $P(A) = 0$, then $B \in \mathcal{F}$ with $P(B) = 0$.

A finite collection of events $\{A_1, \ldots, A_n\}$ is called ***independent*** if $P(\cap_{l=1}^{k} A_{i_l}) = \prod_{l=1}^{k} P(A_{i_l})$ holds for any subset $\{A_{i_1}, \ldots, A_{i_k}\}$ of $\{A_1, \ldots, A_n\}$.

We assume the existence of a probability space $(\Omega, \mathcal{F}, P)$. Let Σ be some separable complete metric space. Given a family of open sets $\mathcal{U} = \{U \mid U \subset \Sigma\}$ there is a smallest σ-algebra $\mathcal{F}_{\mathcal{U}}$ containing $\mathcal{U}$

$$\mathcal{F}_{\mathcal{U}} = \cap\{\mathcal{F} \mid \mathcal{F} \text{ is a } \sigma\text{-algebra of } \Omega, \mathcal{U} \in \mathcal{F}\} \tag{3.1}$$

$\mathcal{F}_{\mathcal{U}}$ is called the σ-algebra generated by $\mathcal{U}$, which is denoted by $\sigma(\mathcal{U})$. Let $\Sigma = \mathbb{R}^n$ and let $\mathcal{U}$ denote a collection of rectangular sets of the form $U = (a, b] = \{x \in \mathbb{R}^n \mid a_i < x_i \leq b_i,\ i = 1, \ldots, n\}$, where $a, b \in \mathbb{R}^n$, then

$$\mathcal{B}(\mathbb{R}^n) = \mathcal{F}_{\mathcal{U}} \tag{3.2}$$

is called the ***Borel*** σ-algebra on $\mathbb{R}^n$, which is the smallest σ-algebra containing rectangular sets of the form $U = (a, b]$ on $\mathbb{R}^n$. The sets of $\mathcal{B}(\mathbb{R}^n)$ are called Borel sets and are denoted by B.
Let $(\Omega^i, \mathcal{F}^i)$, $i = 1, \ldots, n$ be a finite sequence of measurable spaces. Define the product σ-algebra $\mathcal{F} = \mathcal{F}^1 \otimes \ldots \otimes \mathcal{F}^n$ on $\Omega = \Omega^1 \times \ldots \times \Omega^n$ as the σ-algebra generated by products of measurable sets:

$$\mathcal{F} = \sigma(\{A \mid A = A_1 \times \ldots \times A_n \text{ where } A_i \in \mathcal{F}^i,\ i = 1, \ldots, n\}) \tag{3.3}$$

It can be shown that $\mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^n)$.

A ***random variable*** on Ω is a mapping $X : \Omega \to \Sigma$. If $X^{-1}(\mathcal{B}(\Sigma)) = \{\omega \in \Omega \mid X(\omega) \in \mathcal{B}(\Sigma)\} \in \mathcal{F}$, then the random variable X is said to be $\mathcal{F}/\mathcal{B}(\Sigma)$-measurable or just ***$\mathcal{F}$-measurable***. This is equivalent to

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{F},\ \forall B \in \mathcal{B}(\Sigma) \tag{3.4}$$

Sometimes, we denoted $X^{-1}(B)$ by $\{X \in B\}$. If $x \in \Sigma$ then,

$$X^{-1}(x) = \{\omega \in \Omega \mid X(\omega) \leq x\} \tag{3.5}$$

may be represented by $\{X \leq x\}$. The smallest σ-algebra such that X is measurable on $(\Omega, \mathcal{F})$ is denoted by $\sigma(X)$, where

$$\sigma(X) = \mathcal{F}(X) = \{X^{-1}(B) \in \mathcal{F} \mid B \in \mathcal{B}(\Sigma)\} \subseteq \mathcal{F} \tag{3.6}$$

Every random variable $X : \Omega \to \Sigma$ gives rise to a probability measure P_X defined on the σ-algebra of Borel sets $B \in \mathcal{B}(\Sigma)$. P_X is defined in terms of the composition

$$P_X(B) = P \circ X^{-1}(B) = P(\{\omega \in \Omega \mid X(\omega) \in B\}) = P(\{X \in B\}). \tag{3.7}$$

P_X is called the ***distribution*** of X. The function $F_X : \Sigma \to [0, 1]$ defined by

$$F_X(x) = P_X(x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}) = P(\{X \leq x\}) \tag{3.8}$$

is called the ***distribution function*** of X, which is non-decreasing and right continuous. If there is a Borel function $f_X : \Sigma \to \Sigma$ such that

$$P_X(B) = P \circ X^{-1}(B) = \int_B f_X(x)dx, \forall x \in \Sigma \tag{3.9}$$

then X is said to be a random variable with absolutely continuous distribution and

$$f_X(x) = \frac{d}{dx} F_X(x) \tag{3.10}$$

is called the ***density function*** of X. A random variable $X : \Omega \to \Sigma$

$$\int_\Omega |X(\omega)|^p dP(\omega) < \infty \tag{3.11}$$

is said to be ***integrable*** for $p = 1$ and p-th integrable for $p > 0$. We use the notation $L^1(\Omega, \mathcal{F}, P)$ and $L^p(\Omega, \mathcal{F}, P)$ to denote the class of all integrable and ***p-th integrable*** random variables, respectively, on $(\Omega, \mathcal{F}, P)$. Another important class is the class of all ***square integrable*** random variables $L^2(\Omega, \mathcal{F}, P)$ $(p = 2)$ on $(\Omega, \mathcal{F}, P)$. The class $L^p(\Omega, \mathcal{F}, P)$ forms a Banach space with the norm

$$\|X\|_p = \left(\int_\Omega |X(\omega)|^p dP(\omega)\right)^{1/p} \tag{3.12}$$

Let $X \in L^1(\Omega, \mathcal{F}, P)$, the ***expectation*** of X is defined as

$$E(X) = \int_\Omega X(\omega)dP(\omega) = \int_\Sigma x dP_X(x) = \int_\Sigma x f_X(x)dx \tag{3.13}$$

If the two random variables $X, Y \in L^1(\Omega, \mathcal{F}, P)$ are ***independent***, then

$$E(XY) = E(X) \cdot E(Y) \tag{3.14}$$

Conditional Expectation

Given a probability space $(\Omega, \mathcal{F}, P)$, consider a sub-σ-algebra $\mathcal{H} \subseteq \mathcal{F}$ and a random variable $X \in L^2(\Omega, \mathcal{F}, P)$. The integral of X over the set $H \in \mathcal{H}$ will be a measure on $\mathcal{H}$

$$P_X(H) = \int_H X dP \tag{3.15}$$

This measure has the property that if $P(H) = 0$, then $P_X(H) = 0$. From the Radon-Nikodyn Theorem there exists a unique random variable $\bar{X}$ such that $\bar{X} \in L^1(\Omega, \mathcal{F}, P)$ and

1. $\bar{X} \in L^1(\Omega, \mathcal{H}, P)$ is $\mathcal{H}$-measurable

2. $\int_H \bar{X} dP = \int_H X dP \ \forall H \in \mathcal{H}$

We denote $\bar{X}$ by $E[X \mid \mathcal{H}]$ and call it the conditional expectation of X with respect to $\mathcal{H}$.

The conditional probability of the event H given the σ-algebra $\mathcal{H}$ is defined by $P(A \mid \mathcal{H}) = E[\chi_A \mid \mathcal{H}]$ where χ_A is the characteristic function of the set A

$$\chi_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

3.1.2 Stochastic Processes

In this section we will introduce three basic stochastic processes, which are central to the topics discusses in this thesis. These are Brownian motion, martingales and Markov processes.

A ***stochastic process*** $\{\xi_t\}$, $t \in \mathcal{T}$ (or simply $\{\xi_t\}$ where $t \in \mathcal{T}$ is assumed), is a parameterized collection of random variables defined on $(\Omega, \mathcal{F}, P)$ with values in Σ, that is $\xi_t : \Omega \to \Sigma \ \forall t \in \mathcal{T}$. Unless explicitly stated we will assume that the interval

$\mathcal{T} = [0, \infty)$. A stochastic process may also be defined by the mapping $\xi : \mathcal{T} \times \Omega \to \Sigma$, $(t, \omega) \mapsto \xi(t, \omega)$ which can be thought of as the state of the state space Σ. The process $\{\xi_t\}$ is said to be ***measurable*** if $\xi(\cdot, \cdot)$ is $\mathcal{B}(\mathcal{T}) \otimes \mathcal{F}/\mathcal{B}(\Sigma)$-measurable. The map ξ_t is used to denote the ***random variable*** $\xi(t, \cdot)$. The function $\xi(\cdot, \omega)$ is called the ***sample function*** of ξ. If

$$\{\omega \in \Omega \mid \omega \mapsto \xi_t(\omega) = \xi(t, \omega) \leq x\} \in \mathcal{F} \tag{3.16}$$

for each $x \in \Sigma$ then the random variable ξ_t is said to be $\mathcal{F}_t$-measurable.

Filtrations

Given a measurable space $(\Omega, \mathcal{F})$ consider a monotone family of sub-σ-fields $\mathcal{F}_t \subseteq \mathcal{F}$, $t \in \mathcal{T}$:

$$\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2}, \forall t_1, t_2 \in \mathcal{T} \text{ s.t. } t_1 \leq t_2 \tag{3.17}$$

The family $\{\mathcal{F}_t\}$, $t \in \mathcal{T}$ (or simply $\{\mathcal{F}_t\}$ where $t \in \mathcal{T}$ is assumed) is also a σ-algebra and is called a ***filtration***; it contains all the information generated by a stochastic process $\{\xi_t\}$ on the interval $\mathcal{T}$. An event A is $\mathcal{F}_t$-measurable, denoted by A $\in \mathcal{F}_t$, if it is possible to decide whether A has occurred based on the observation of the trajectory $\{\xi_t\}$. Given a filtration $\{\mathcal{F}_t\}$ on a probability space $(\Omega, \mathcal{F}, P)$, define

$$\mathcal{F}_{t_+} = \cap_{s>t}\mathcal{F}_s,\ \mathcal{F}_{t_-} = \sigma(\cup_{s<t}\mathcal{F}_s), \text{ where } s, t \in \mathcal{T} \tag{3.18}$$

We say that a process $\{\mathcal{F}_t\}$ is right (left) continuous if $\mathcal{F}_t = \mathcal{F}_{t_+}$ ($\mathcal{F}_t = \mathcal{F}_{t_-}$) $\forall t \in \mathcal{T}$. We say that the filtration satisfies the ***usual condition*** if it is right continuous and $\mathcal{F}_0$ contains all the null events.

The stochastic process $\{\xi_t\}$ is ***adapted*** to the filtration $\{\mathcal{F}_t\}$ if $\forall t \in \mathcal{T} = [0, T]$, where $0 < T < \infty$ is the final time, the random variable ξ_t is a $\mathcal{F}_t/\mathcal{B}(\Sigma)$-measurable function - adaptedness means that the σ-algebra generated by the measurable random vector

ξ_t up to time T is contained in the filtration, that is $\sigma(\xi_t) \subseteq \{\mathcal{F}_t\}$. We sometimes say that $\{\xi_t\}$ is ***nonanticipative*** when it is $\{\mathcal{F}_t\}$-adapted.

Given a stochastic process $\{\xi_t\}$, $t \in \mathcal{T} = [0, T]$, the simplest choice of filtration is that generated by the process itself, that is $\mathcal{F}_T^\xi = \sigma(\{\xi_t\})$ which is the smallest σ-algebra with respect to ξ_t that is measurable $\forall t \in \mathcal{T}$.

Brownian Motion

Let $(\Omega, \mathcal{F}, P)$ be a filtered probability space. Consider an $\{\xi_t\}$-adapted $\mathbb{R}^n$-valued process $w_t = (w_t^1, \ldots, w_t^d)$ with mean $\mu(t) = E[w_t]$ and covariance $V(s,t) = E[(w_s - \mu(s))(w_t - \mu(t))^t]$. The process w_t is called ***Brownian motion*** if it has independent increments, that is, given $t_0 = s$, $t_n = T$ and $t_k < t_{k+1}$, where $0 \leq k \leq n-1$, the increments $w_{t_{k+1}} - w_{t_k}$ are independent random variables. The covariance $V(s,t)$ satisfies the following:

1. $V(s,t) = V(r,r)$ where $r = \min(s,t)$

2. $V(t) = V(t,t)$ increases with t.

A Brownian motion is called ***standard*** if $\mu(t) = 0$ and $V(t,t) = tI_{n \times n}$.

Martingales

A ***martingale*** M (***submartingale***/***supermartingale***) with respect to a filtration $\{\mathcal{F}_t\}$ is a stochastic process $M = \{M_t\}$, $t \in \mathcal{T}$, (or simply $M = \{M_t\}$ where $t \in \mathcal{T}$ is assumed) which is adapted to $\{\mathcal{F}_t\}$ such that:

1. the random variables M_t are integrable, i.e. $M_t \in L^1(\Omega, \mathcal{F}, P)$ $\forall t \in \mathcal{T}$

2. $E[M_t|\mathcal{F}_r] = M_r$ $(\geq / \leq)$ $P - a.s$ $\forall r < t$

We sometimes use the notation $M = \{M_t, \mathcal{F}_t\}$ to refer to a martingale (submartingale/supermatingale).

One of the most important examples of martingales is a Brownian motion.

Stopping times play an important role in the theory of submartingales. Let $\mathcal{T} = [0, \infty)$ or $[0, T]$. A random variable $\tau : \Omega \to \mathcal{T}$ is called a ***stopping time*** if the set of all $\omega \in \Omega$ that satisfies the inequality $\tau(\omega) \leq t$ for any $t \in \mathcal{T}$ belongs to $\mathcal{F}_t$, that is $\{\omega \mid \tau(\omega) \leq t \in \mathcal{T}\} \in \mathcal{F}_t$. If σ, τ are stopping times of the same filtration, then $\sigma \wedge \tau = \min\{\sigma, \tau\}$ is a minimum stopping time if $\{\sigma \wedge \tau \leq t\} = \{\sigma \leq t\} \cup \{\tau \leq t\} \in \mathcal{F}_t$. Consider a general stochastic process $\{\xi_t\}$ where $t \in \mathcal{T}$, a ***stopped process*** is defined as $\xi^\tau = \{\xi^\tau_t\}$, where $\xi^\tau_t = \xi_{t \wedge \tau}$ is a random variable $\omega \mapsto \xi_{t \wedge \tau(\omega)}(\omega)$. For a given stopping time, set $\mathcal{F}_\tau = \{A \in \mathcal{F} \mid A \cap \{\tau(\omega) \leq t\} \in \mathcal{F}_t, \ \forall t \in \mathcal{T}, \omega \in \Omega\}$. $\mathcal{F}_\tau$ is a sub σ-algebra of $\mathcal{F}$.

Markov Processes

This this section we provide an elementary definition that identifies a ***Markov processes*** as a stochastic process. A stochastic process $\{\xi_t\}$, with $t \in \mathcal{T}$ is a Markov process if for any $0 \leq s \leq t \leq T$ and any Borel set B of the state space Σ we have

$$P(\xi_t \in B \mid \mathcal{F}_s) = P(\xi_t \in B \mid \xi_s) \tag{3.19}$$

which implies that the probability given $\mathcal{F}_s$ is exactly the same as the probability given the information available at time s.

We will take a more detailed look at Markov processes and look at their relation to diffusion processes in the sections that follow.

3.2 Continuous Semimartingales and Stochastic Integrals

3.2.1 Introduction

The purpose of this section is to introduce stochastic integrals based on continuous semimartingales, which is used to establish Ito's formula. We will also introduce localmartingales and semimartingales which are generalizations of martingales and submartingales, Kunita [40] provides an excellent overview of the calculus of semimartingales.

Localmartingales

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space equipped with the filtration $\{\mathcal{F}_t\}$ of sub σ-fields of $\mathcal{F}$ with time interval $\mathcal{T} = [0, T]$. A continuous valued $\{\mathcal{F}_t\}$-adapted process $M = \{M_t, \mathcal{F}_t\}$ is called a ***localmartingale*** if there exists and increasing sequence of stopping times $\{\tau_n\}$ such that $P(\tau_n < T) \to 0$ as $n \to \infty$ and each stopped process $M^{\tau_n} = \{M_t^{\tau_n}\}$ is a martingale. A continuous local-submartingale and a continuous local-supermartingale are defined in a similar manner.

Let $\mathcal{L}_c$ be the linear space of all real valued continuous stochastic processes. We introduce the norm $\|\cdot\|$ by $\|M\| = E[\sup_t |M_t|^2]^{\frac{1}{2}}$ and denote by $\mathcal{L}_c^2$ the set of all elements in $\mathcal{L}_c$ with finite norms. Let $\mathcal{M}_c^2$ be the set of all continuous square integrable martingales M with $M_0 = 0$. From Doob's inequality, the norm $\|M\|$ is finite for any $M \in \mathcal{M}_c^2$. Hence, $\mathcal{M}_c^2$ is a subset of $\mathcal{L}_c^2$. We denote by $\mathcal{M}_c^{loc}$ the set of all continuous localmartingales M such that $M_0 = 0$, which is subset of $\mathcal{L}_c$. We now state the ***Doob-Meyer decomposition*** theorem, which plays an important role in stochastic analysis.

Theorem 3.2.1 (Doob-Meyer Decomposition - page 14 Hsu [24]). *Let $t \in \mathcal{T}$, consider a probability space $(\Omega, \mathcal{F}, P)$, we assume that the filtration $\{\mathcal{F}_t\}$ satisfies the usual conditions. Let $\{\xi_t\}$ be a $\mathcal{F}_t$-submartingale with continuous sample paths. Then there exists a unique continuous local $\mathcal{F}_t$-martingale $M = \{M_t, \mathcal{F}_t\}$ and a continuous $\mathcal{F}$-adapted nondecreasing process $\{A_t\}$ with $A_0 = 0$ such that*

$$\xi_t = M_t + A_t,\ t \in \mathbb{R}_+ \tag{3.20}$$

Semimartingales

Next we define semimartingales and other related stochastic processes. Let $M = \{M_t, \mathcal{F}_t\}$ be a continuous $\{\mathcal{F}_t\}$-adapted process. It is called an increasing process if for almost all ω, $M_t(\omega)$ is an increasing function of t such that $M_0(\omega) = 0$. It is called a process of bounded variation if it is written as the difference of two increasing processes. It is called a ***semimartingale*** if it is written as the sum of a localmartingale and a process of bounded variation.

Proposition 3.2.1 (page 16 Hsu [24]). *A semimartingale can be uniquely decomposed as a sum of a continuous local martingale and a continuous process of bounded variation.*

3.2.2 Quadratic Variational Processes

In this section we study the quadratic variation of continuous stochastic processes, which is a measure of path oscillation for processes that have infinite variation.

Quadratic Variations of Continuous Semimartingales

Let $\{\xi_t\}$, $t \in \mathcal{T} = [0,T]$, be a continuous stochastic process. Let $\triangle = \{0 = t_0 < \ldots < t_l = T\}$ denote the partition of the interval $[0,T]$. Let $|\triangle| = \max_k(t_{k+1} - t_k)$ where $0 \leq k \leq l-1$. The ***quadratic variation along a partition*** $\triangle$ for ξ is defined by

$$\langle \xi \rangle_t^{\triangle} = \sum_{k=0}^{l-1} (\xi_{t \wedge t_{k+1}} - \xi_{t \wedge t_k})^2 \tag{3.21}$$

Let $\triangle_1, \ldots, \triangle_n$ be a sequence of partitions such that $|\triangle_n| \to 0$. If $\forall t$ the limit

$$\lim_{|\triangle_n| \to 0} \langle \xi \rangle_t^{\triangle_n} = \langle \xi \rangle_t, \ \forall t \in \mathcal{T} \tag{3.22}$$

exists in probability and is independent of the choice of sequences $\triangle_1, \ldots, \triangle_n$ a.s., it is called the ***quadratic variation*** of ξ and is denoted by $\langle \xi \rangle_t$. The ***quadratic variation process*** is defined by $\langle \xi \rangle = \{\langle \xi \rangle_t\}$.

Let $\{\eta_t\}$, $t \in \mathcal{T} = [0,T]$, be a continuous stochastic process. Then the ***joint quadratic variation*** of ξ and η associated with the partition $\triangle = \{0 = t_0 < \ldots < t_l = T\}$ is defined by

$$\langle \xi, \eta \rangle_t^{\triangle} = \sum_{k=0}^{l-1} (\xi_{t\wedge t_{k+1}} - \xi_{t\wedge t_k})(\eta_{t\wedge t_{k+1}} - \eta_{t\wedge t_k}) \tag{3.23}$$

By definition $\langle \xi, \xi \rangle_t^{\triangle} = \langle \xi \rangle_t^{\triangle}$. The joint quadratic variation may also be defined in terms of the polarization identity

$$\langle \xi, \eta \rangle_t = \frac{\langle \xi + \eta \rangle_t - \langle \xi - \eta \rangle_t}{4} \tag{3.24}$$

Consider the semimartingales $\xi_i = M_i + A_i$, where $i = 1, 2$, then $\langle \xi_i \rangle_t = \langle M_i \rangle_t$ and $\langle \xi_1, \xi_2 \rangle_t = \langle M_1, M_2 \rangle_t$. It can be shown that $\langle M_i \rangle_t$ can be defined as the unique continuous increasing process such that $M_i^2 - \langle M_i \rangle_t$ is a continuous local martingale.

3.2.3 Stochastic Integrals and Ito's Formula

Ito's Integrals and Stratonovich Integrals

Let $\mathcal{T} = [0, T]$, consider a continuous localmartingale $M = \{M_t, \mathcal{F}_t\}$ and let f_t be a continuous $\{\mathcal{F}_t\}$-adapted process. We will embark on defining the stochastic integral of f_t with respect to the differential dM_t, we will restrict our attention to continuous localmartingales and continuous semimartingales.

Consider the partion $\triangle = \{0 = t_0 < \ldots < t_l = T\}$ of the interval $[0, T]$. Define

$$L_t^{\triangle} = \sum_{k=0}^{l-1} f_{t\wedge t_k}(M_{t\wedge t_{k+1}} - M_{t\wedge t_k}) \tag{3.25}$$

Lemma 3.2.1 (page 56 Kunita [40]). *Let $L_t^{\triangle}$ is a continuous localmartingale. Its quadratic variation is given by*

$$\langle L^{\triangle} \rangle_t = \int_0^t |f_s^{\triangle}|^2 d\langle M \rangle_s \tag{3.26}$$

where $\langle M \rangle_t$ is the quadratic variation of M_t and $f_t^{\triangle}$ is the simple process defined from setting $f_t^{\triangle} = f_{t_k}$ if $t_k \le t < t_{k+1}$.

Now let $\{\triangle_n\}$ be a sequence of partitions of $[0,T]$ such that $|\triangle_n| \to 0$. Then, it can be shown that the sequence $\{L^{\triangle_n}\}$ is a Cauchy sequence in $\mathcal{M}_c^{loc}$, this allows us apply various convergence results that leads to the limit

$$\lim_{n\to\infty}\{L^{\triangle_n}\} = L_t \tag{3.27}$$

We call this limit, Equ.(3.27), the ***Ito integral*** of f_t by dM_t, which we formally define as

$$L_t := \int_0^t f_s dM_s \tag{3.28}$$

Let ξ be a continuous semimartingale decomposed to a continuous localmartingale M and a continuous process of bounded variation A. Let $|A|_t$ be the total variation of A_s $0 \le s \le t$. It is a continuous increasing process. For an arbitrary element f of $L^2(\langle M\rangle) \cap L^1(|A|)$ we define the Ito integral by $d\xi_t$

$$\int_0^t f_s d\xi_s = \int_0^t f_s dM_s + \int_0^t f_s dA_s \tag{3.29}$$

It is a continuous martingale. Its joint quadratic variation with a continuous semimartingale η_t satisfies

$$\langle \int f d\xi, \eta\rangle_t = \int_0^t f_s \langle \xi, \eta\rangle_s = \int_0^t f_s d\langle M, N\rangle_s \tag{3.30}$$

where N_t is the localmartingale part of η.

We will define another stochastic integral by the differential $\circ d\xi_t$:

$$\int_0^t f_s \circ d\xi_s = \lim_{|\triangle|\to 0} \sum_{k=0}^{l-1} \frac{1}{2}(f_{t\wedge t_{k+1}} + f_{t\wedge t_k})(\xi_{t\wedge t_{k+1}} - \xi_{t\wedge t_k}) \tag{3.31}$$

The above limit exists in the same sense of the convergence in probability, it is called the ***Stratonovich integral*** of f_s by $\circ d\xi_s$.

$$\int_0^t f_s \circ d\xi_s = \int_0^t f_s d\xi_s + \frac{1}{2}\langle f, \xi\rangle_t \tag{3.32}$$

Ito's Formula

Theorem 3.2.2 (page 64 Kunita [40]). *Let $\xi_t = (\xi_t^1, \dots, \xi_t^n)$ be a continuous semimartingale. If $\phi(x^1, \dots, x^n) \in C^2$, then $\phi(\xi_t)$ is a continuous semimartingale and satisfies the following formula:*

$$\phi(\xi_t) - \phi(\xi_0) = \sum_{i=1}^n \int_0^t \phi_{x^i}(\xi_s) d\xi_s^i + \frac{1}{2}\sum_{i,j=1}^n \int_0^t \phi_{x^i x^j}(\xi_s) d\langle \xi^i, \xi^j\rangle_s \tag{3.33}$$

Furthermore, if $\phi \in C^3$, then we have

$$\phi(\xi_t) - \phi(\xi_0) = \sum_{i=1}^n \int_0^t \phi_{x^i}(\xi_s) \circ d\xi_s^i \tag{3.34}$$

3.3 Markov and Diffusion Processes

The purpose of this section is to study the properties of Markov processes, which feature prominently in filtering theory. In particular, we present Kolmogorov's equation for the evolution of the unconditional distribution of a stochastic process, which can be regarded as a precursor to the conditional distribution of the stochastic process given the observations. We follow the concise exposition by Davis et al [15], pages 57-62.

A stochastic process $\{\xi_t\}$, $t \in \mathcal{T} = [0, T]$, is a ***Markov Process*** if for any $0 \le s \le t \le T$ and any Borel set $B \in \mathcal{B}(\Sigma)$

$$P(\xi_t \in B \mid \mathcal{F}_s) = P(\xi_t \in B \mid \xi_s) \tag{3.35}$$

The ***transition probability function*** for a Markov process $\{\xi_t\}$ is defined as

$$P(s, x, t, B) := P(\xi_t \in B \mid \xi_s = x) \tag{3.36}$$

which satisfies the ***Chapman-Kolmogorov equation***

$$P(s,x,t,B) = \int_{\Sigma} P(u,y,t,B)P(s,x,u,dy) \tag{3.37}$$

for any $0 \leq s \leq u \leq t \leq T$. One of the important properties that Markov processes have is that all finite dimensional distributions are determined by its initial distribution and transition probability function. A Markov process $\{\xi_t\}$ is said to be ***homogeneous*** if the transition probability function is invariant with respect to a time shift in the following sense

$$P(s+u,x,t+u,B) = P(s,x,t,B) \tag{3.38}$$

$\forall\ 0 \leq s \leq t \leq T$ and $0 \leq s+u \leq t+u \leq T$. Consider a bounded measurable real-valued function f on Σ, that is $f \in \mathcal{B}(\Sigma)$, define

$$T_t f(x) = E_x[f(\xi_t)] := \int_{\Sigma} f(y)P(0,x,t,dy) \tag{3.39}$$

for a homogeneous Markov process $\{\xi_t\}$. Using the Chapman-Kolmogorov equation we can show that T_t is a ***semigroup*** of operators acting on $\mathcal{B}(s)$, that is $T_{t+s}f(x) = T_t(T_s f)(x)$ for $t, s \geq 0$. The process $\{\xi_t\}$ or the operator T_t has a ***generator***, which we denote by $\mathcal{L}$, that acts on a domain $\mathcal{D}(\mathcal{L}) \subset \mathcal{B}(\Sigma)$ and is defined by

$$\mathcal{L}\phi = \lim_{t\to 0} \frac{1}{t}(T_t\phi - \phi) \tag{3.40}$$

where, the limit is uniform in $x \in \Sigma$ and $\mathcal{D}(\mathcal{L})$ comprises of all the functions that give rise to a finite limit. It can be shown that the time derivative of the semigroup operator T_t is related to the generator $\mathcal{L}$ in the following manner

$$\frac{d}{dt}T_t\phi = \mathcal{L}T_t\phi \tag{3.41}$$

This equation is the abstract version of the so called ***backward equation*** for the process. By representing this equation in integral form and using the definition of T_t, one arrives at the ***Dynkin formula***

$$E_x[\phi(\xi_t) - \phi(x)] = E_x \int_0^t \mathcal{L}\phi(\xi_s)ds \tag{3.42}$$

From this we conclude that the process M_t^ϕ defined for $\phi \in \mathcal{D}(\mathcal{L})$

$$M_t^\phi = \phi(\xi_t) - \phi(x) - \int_0^t \mathcal{L}\phi(\xi_s)ds \tag{3.43}$$

is a martingale; this property can be used as a definition for $\mathcal{L}$. The operator $\mathcal{L}$ is known as the ***extended generator*** of $\{\xi_t\}$ because M_t^ϕ can be a martingale for certain $\phi \notin \mathcal{D}(\mathcal{L})$. In context to filtering theory, there exists an another interesting semigroup of operators associated with $\{\xi_t\}$, which propagates the initial distribution of the process with respect to time. More precisely, let $M(\Sigma)$ be a set of probability measures on Σ and denote the inner product of $\phi \in \mathcal{B}(\Sigma)$ and $\mu \in M(\Sigma)$ by

$$\langle \phi, \mu \rangle = \int_\Sigma \phi(x)\mu(dx) \tag{3.44}$$

Suppose ξ_0 has distribution $\pi \in M(\Sigma)$, the distribution for ξ_t is given by

$$P[\xi_t \in A] = E[I_A(\xi_t)] = \langle T_t I_A, \pi \rangle \tag{3.45}$$

We denote the distribution $P[\xi_t \in A]$ by $U_t\pi(A)$. This shows that U_t is adjoint to T_t since

$$\langle \phi, U_t\pi \rangle = \langle T_t\phi, \pi \rangle = E[\phi(x_t)] \tag{3.46}$$

for $\phi \in \mathcal{B}(\Sigma)$, $\pi \in M(\Sigma)$. Thus the generator of U_t is $\mathcal{L}^*$, the adjoint of $\mathcal{L}$ and $\pi_t := U_t\pi$ satisfies

$$\tfrac{d}{dt}\pi_t = \mathcal{L}^*\pi_t \text{ where } \pi_0 = \pi \tag{3.47}$$

This is the ***forward equation*** for $\{\xi_t\}$ since it gives the evolution of the distribution π_t of ξ_t. The objective of filtering theory is to obtain a similar description of the *conditional distribution* of ξ_t given $\{\eta_s \mid s \leq t\}$.

Theorem 3.3.1 (page 60 Davis et al [15]). *The solution of*

$$\xi_t = \xi_0 + \int_0^t f(\xi_s)ds + \int_0^t g(\xi_s)d\beta_s \tag{3.48}$$

where $\xi_t \in \mathbb{R}^n$ and $\beta_t \in \mathbb{R}^m$ is a vector of independent Brownian motion, is a homogeneous Markov process with infinitesimal generator

$$\mathcal{L} = \sum_{i=1}^{n} f^i(x)\frac{\partial}{\partial x^i} + \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(x)\frac{\partial^2}{\partial x^i \partial x^j} \tag{3.49}$$

where $A(x) = [a^{ij}(x)] = g(x)g^T(x)$ and f^i and x^i denote the i^{th} components of f and x respectively. In this case Ito's rule Equ.(3.33) can be written as:

$$\begin{aligned}\phi(\xi_t) - \phi(\xi_0) &= \sum_{i=1}^{n}\int_0^t \phi_{x^i}(\xi_s)d\xi_s^i + \frac{1}{2}\sum_{i,j=1}^{n}\int_0^t \phi_{x^i x^j}(\xi_s)a^{ij}(\xi_s)ds \\ \Rightarrow \phi(\xi_t) - \phi(\xi_0) &= \int_0^t \mathcal{L}\phi(\xi_s)ds + \int_0^t \nabla\phi^T(\xi_s)g(x_s)d\beta_s \end{aligned} \tag{3.50}$$

emphasizing that M_t^{ϕ} is a martingale, where ∇ in Equ.(3.50) denotes the gradient with respect to x.

The following theorem deals with the Kolmogorov forward equation.

Theorem 3.3.2 (page 60 Davis et al [15]). *Assume that the stochastic process $\{\xi_t\}$ that satisfies Equ.(3.48) has a transition density $p(s,x,t,y)$*

$$P(s,x,t,B) = \int_B p(s,x,t,y)dy \tag{3.51}$$

which is continuous and bounded in s,t and x for $t - s > \delta > 0$ $p(s,x,t,y)$. We assume that the partial derivatives

$$\frac{\partial f}{\partial x^i}, \frac{\partial A}{\partial x^i}, \frac{\partial^2 A}{\partial x^i \partial x^j}, \frac{\partial p}{\partial t}, \frac{\partial p}{\partial y}, \frac{\partial p}{\partial y^2}$$

*exist. Then for $0 < s < t$, p satisfies the **Kolmogorov forward equation***

$$\begin{aligned}\frac{\partial p}{\partial t}(s,x,t,y) &= -\sum_{i=1}^{n}\frac{\partial}{\partial y^i}(f^i(y)p(s,x,t,y)) \\ &\quad +\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2}{\partial y^i\partial y^j}(a^{ij}(y)p(s,x,t,y)) \qquad (3.52)\\ &= \mathcal{L}^*p(s,x,t,y) \qquad (3.53)\end{aligned}$$

where $\mathcal{L}^*$ is the formal adjoint of $\mathcal{L}$. Also, the initial condition is

$$\lim_{t\to s} p(s,x,t,y) = \delta(y-x) \qquad (3.54)$$

The formal adjoint of the Kolmogorov forward equation gives rise to the ***Kolmogorov backward equation***

$$\frac{\partial}{\partial s}p(s,x,t,y) + \mathcal{L}p(s,x,t,y) = 0 \qquad (3.55)$$

The backward and forward equations play an important role in stochastic optimal control and filtering theory, respectively. Consider the following stochastic differential equation, page 201 Maybeck [52]:

$$d\xi_t = f(\xi_t,t)dt + g(\xi_t,t)dw_t \qquad (3.56)$$

where w_t is Brownian motion with diffusion $Q(t)$ $\forall t \in \mathcal{T}$, the transition probability density $p(s,x,t,y)$ for ξ_t satisfies the forward equation. Given that ξ_t is Markov, the update equations for the mean and covariance for ξ_t can be determined as follows:

$$\begin{aligned}\dot{\mu}_\xi(t) &= E[f(\xi_t,t)] \qquad (3.57)\\ \dot{P}_\xi(t) &= [E[f(\xi_t,t)\xi_t^T] - E[f(\xi_t,t)]\mu_\xi^T(t)] + \\ &\quad [E[\xi_t f^t(\xi_t,t)] - \mu_\xi(t)E[f^T(\xi_t,t)]] + \\ &\quad E[g(\xi_t,t)Q(t)g^T(\xi_t,t)] \qquad (3.58)\end{aligned}$$

In the special case of linear stochastic differential equations $f(\xi_t,t) = F(t)\xi_t$, $g(\xi_t,t) = G(t)$ the update equations for the mean and covariance for ξ_t simplify to

$$\dot{\mu}_\xi(t) = F(t)\mu_\xi(t) \tag{3.59}$$

$$\dot{P}_\xi(t) = F(t)P_\xi(t) + P_\xi(t)F^T(t) + G(t)Q(t)G^T(t) \tag{3.60}$$

which appears to be a subset of the Kalman-Bucy filter, Cha. 5.

Chapter 4

NONLINEAR FILTERING

The general filtering problem concerns estimating some aspect of an unobservable stochastic process $\{x_t\}$ given the observations of a related process $\{y_t\}$. The problem of interest is to calculate the conditional distribution of $\{x_t\}$ given $\{y_s \mid 0 \leq s \leq t\}$. In the context of linear filtering theory, this problem has been solved in what is now known as the *Kalman-Bucy* filter, Kalman [32] and Kalman et al [31]. The nonlinear filtering problem is, in general, infinite-dimensional and therefore more difficult to solve. However, considerable progress was made in the early sixties in as far as describing the evolution of the conditional distribution is concerned by Bucy [8], Kushner [38], Shiryaev [64], Stratonovich [65] and Wonham [68]. In 1968, Kailath [30] utilized the so called *innovations* approach for the linear filtering. To extend this approach to the nonlinear problem, Frost [18] advocated that filtering theory be formulated on martingale theory. The first authoritative treatment based on this idea was provided by Fujisaki et al [19]. Despite these developments, the nonlinear filtering problem was purely of academic interest because of the complicated nature of the conditional distribution. In 1969 Zakai [69] made a significant break through. Using the so called *reference probability method*, Wong [67], Zakai obtained a substantially simpler set of equations that made the problem less intractable and more attractive for numerical implementation.

The mainstream approach to nonlinear filtering theory follows two schools based on:

1. the Kushner-Stratonovich or Fujisaki-Kallianpur-Kunita equation, - Fujisaki et al [19] and

2. the Zakai equation, - Zakai [69].

4.1 Introduction

The following hypotheses, page 57 Davis et al [15], are made in all nonlinear estimation processes:

H1] $\{y_t\}$ is a real valued process

H2] $\{v_t\}$ is at standard Brownian motion process

H3] $E[\int_0^T z_s^2 ds] < \infty$

H4] $\{z_t\}$ is independent of $\{v_t\}$

The general problem statement is given as follows:

Problem Statement 4.1.1 (page 817 Marcus [51] and page 55 Davis et al [15]). *The* ***state process*** $\{x_t\}$ *is a stochastic process that cannot be observed directly. The information concerning* $\{x_t\}$ *is contained in the* ***observation process*** $\{y_t\}$

$$y_t = \int_0^t z_s ds + v_t \tag{4.1}$$

where $\{z_t\}$ *is a process related to* $\{x_t\}$ *by*

$$z_t = h(x_t) \tag{4.2}$$

and $\{v_t\}$ *is a Brownian motion process. The end goal is determine the least squares estimates of functions of* x_t *given past observations* $\{y_s \mid 0 \leq s \leq t\}$*, that is the computation of the conditional expectation of the form* $E[\phi(x_t) \mid y_s, 0 \leq s \leq t]$*. For practical purposes, due to the nature of the observation process, which is typically based on discrete-time measurements of a physical process, it is highly desirable for this computation to occur in a recursive manner in terms of the statistic* $\pi = \{\pi_t\}$ *which can be updated using new observations when they become available:*

$$\pi_{t+\tau} = \alpha(t, \tau, \pi_t, \{y_s \mid 0 \le s \le t + \tau\}) \tag{4.3}$$

Then, π_t may be used to compute estimates in a memoryless manner

$$E[\phi(x_t) \mid y_s, 0 \le s \le t] = \beta(t, y_t, \pi_t) \tag{4.4}$$

One can regard π_t as a representation of the ***conditional distribution*** of $\{x_t\}$ given $\{y_s \mid 0 \le s \le t\}$. There are conditions under which π_t can be computed with a finite set of stochastic differential equations driven by $\{y_t\}$. The significance of recursiveness is that it leads to a real-time implementation of the filter described by Equ.(4.3) and Equ.(4.4). More specifically:

1. α and β can be thought of as representing that aspect of the filter that does not depend on data, it can be hardware or software related

2. π_t represents the state of the filter, which has a memory requirement

3. $\{y_s \mid t \le s \le t + \tau\}$ or dy_t can be thought of as the new information to the filter that is encoded in the measurement process, which becomes available at certain time increments

To progress with the problem through the application of certain results from stochastic calculus, we make the following assumptions:

1. $\{x_t\}$ is a semimartingale

2. $\{x_t\}$ is a Markov process or a vector diffusion process of the form Equ.(3.48)

We will see in later chapters that the structure of the recursive nonlinear filtering problem leads to the natural application of geometric techniques involving Lie algebras.

4.2 The Innovations Approach to Nonlinear Filtering

In this section we assume that our nonlinear filtering problems are well posed and satisfy the hypotheses H1-H4 outlined in the previous section. In particular, we examine the stochastic differential equations that describe the evolution of conditional statistics and of the conditional density, which is similar to the Kolmogorov forward equation for Markov processes. We follow the innovations approach of Fujisaki et al [19].

Assume that observations y_t have the form Equ.(4.1) and that the hypotheses H1 to H4 are valid. Define $\mathcal{Y}_t = \sigma\{y_s \mid 0 \leq s \leq t\}$. Now for any arbitrary process η_t we use the notation

$$\hat{\eta}_t = E[\eta_t \mid y_t] \tag{4.5}$$

Now we define the ***innovations process*** by the residual

$$\nu_t = y_t - \int_0^t \hat{z}_s ds \tag{4.6}$$

The increment $\nu_{t+h} - \nu_t$ represents new information of the process $\{z_t\}$ that is determined from the observations on the interval $[t, t+h]$, such that the residual $\nu_{t+h} - \nu_t$ is independent of $\mathcal{Y}_t$. The following properties of the innovation are important.

Lemma 4.2.1 (page 63 Davis et al [15]). *The process $\{\nu_t, \mathcal{Y}_t\}$ is a standard Brownian motion process. Furthermore, $\mathcal{Y}_s$ and $\sigma\{\nu_u - \nu_t \mid 0 \leq s \leq t < u \leq T\}$ are independent.*

It is important to note that the above lemma provides a precise result concerning the structure of the innovations process without any restriction on the distribution of z_t.

Before we proceed to the next lemma, which is related to Kailath's innovations conjecture, consider the following discussion. By definition ν_t is $\mathcal{Y}_t$-measurable and $\sigma\{\nu_s \mid 0 \leq s \leq t\} \subset \mathcal{Y}_t$. The innovations conjecture asserts that $\mathcal{Y}_t \subset \sigma\{\nu_s \mid 0 \leq s \leq t\}$, which implies that two σ-algebras are equal; this means that the observations and the innovations contain the same information. This important fact was unknown at the time that Fujisaki et al [19] was published, however, it was later shown by Allinger et al [1] that the conjecture is true when H1-H4 are respected. It is well known in stochastic analysis that all martingales of Brownian motion are stochastic integrals. The innovations conjecture asserts that any $\mathcal{Y}_t$-martingale can be written as a stochastic integral with respect to the innovations process $\{\nu_t\}$. The importance of the contribution made by Fujisaki et al [19] was to show that this representation holds whether or not the innovations conjecture is valid.

Lemma 4.2.2 (Fujisaki et al [19]). *Every square integrable martingale $\{m_t, \mathcal{Y}_t\}$ with respect to observation σ-algebras $\mathcal{Y}_t$ is sample continuous and has the representation*

$$m_t = E[m_0] + \int_0^t \eta_s d\nu_s \tag{4.7}$$

where $\int_0^T E[\eta_s^2]ds < \infty$ and $\{\eta_t\}$ is jointly measurable and adapted to $\mathcal{Y}_t$.

The lemma proves that m_t can be written as a stochastic integral with respect to the innovations process. However, it should be noted that $\{\eta_t\}$ is adapted to $\mathcal{Y}_t$ and not necessarily to $\mathcal{F}_t^\nu$

To obtain a general filtering equation, consider a real-valued $\mathcal{F}_t$-semimartingale ξ_t

$$\xi_t = \xi_0 + \int_0^t \alpha_s ds + n_t \tag{4.8}$$

where $\{n_t, \mathcal{F}_t\}$ is a martingale and derive an equation satisfied by $\hat{\xi}_t$, which has the form of Equ.(4.5).

Theorem 4.2.1 (page 64 Davis et al [15]). *Assume the $\{\xi_t\}$ and $\{y_t\}$ are given by Equ.(4.8) and Equ.(4.1) respectively and that $\langle n, w\rangle_t = 0$. Then $\{\hat{\xi}_t\}$ satisfies the stochastic differential equation*

$$\hat{\xi}_t = \hat{\xi}_0 + \int_0^t \hat{\alpha}_s ds + \int_0^t [\widehat{\xi_s z_s} - \hat{\xi}_s \hat{z}_s] d\nu_s \tag{4.9}$$

Equ.(4.9) is not recursive, however, it can be used to obtain further results for filtering of Markov processes, in particular the following theorem.

Theorem 4.2.2 (page 66 Davis et al [15]). *Assume that:*

1. *$\{x_t\}$ is a homogeneous Markov process with infinitesimal generator $\mathcal{L}$*
2. *that $\{y_t\}$ is given by Equ.(4.1) with $z_t = h(x_t)$ and*
3. *that $\{x_t\}$ and $\{v_t\}$ are independent.*

Then for any $\phi \in \mathcal{D}(\mathcal{L})$, $\langle \pi_t, \phi \rangle = \pi_t(\phi) = E[\phi(x_t) \mid \mathcal{Y}_t]$ satisfies

$$\pi_t(\phi) = \pi_0(\phi) + \int_0^t \pi_s(\mathcal{L}\phi)ds + \int_0^t [\pi_s(h\phi) - \pi_s(h)\pi_s(\phi)]d\nu_s \tag{4.10}$$

Equ.(4.10) can be regarded as a recursive infinite-dimensional stochastic differential equation for the conditional measure π_t of x_t given $\mathcal{Y}_t$ because $\{\pi_t(\phi) \mid \phi \in \mathcal{D}(\mathcal{L})\}$ determines a measured valued stochastic process π_t. Furthermore, $\pi_t(\phi)$ is the conditional statistic computed from π_t in a memoryless fashion. In general it is not possible to derive a finite dimensional recursive filter even for the conditional mean $\hat{x}_t$. However, there are special cases where finite dimensional filters are known to exist.

Assume that $\{x_t\}$ is a diffusion process of the form given by Equ.(3.48) with infinitesimal generator $\mathcal{L}$ Equ.(3.49) and that the conditional distribution of x_t given $\mathcal{Y}_t$ has a density $\tilde{p}(t,x)$. Using the differentiability hypotheses, Liptser et al [49], one can perform an integration by parts on

$$\pi_t(\phi) = \pi_0(\phi) + \int_0^t \pi_s(\mathcal{L}\phi)ds + \int_0^t [\pi_s(h\phi) - \pi_s(h)\pi_s(\phi)]d\nu_s \tag{4.11}$$

to obtain a stochastic partial differential equation

$$d\tilde{p}(t,x) = \mathcal{L}^*\tilde{p}(t,x)dt + \tilde{p}(t,x)[h(x) - \pi_t(h)]d\nu_t \tag{4.12}$$

where

$$\pi_t(h) = \int h(x)\tilde{p}(t,x)dx \tag{4.13}$$

This is a recursive equation for the computation of $\tilde{p}(t,x)$. However, it is infinite dimensional and has the complication of the integral in Equ.(4.13). The equation for $d\tilde{p}(t,x)$, Equ.(4.12), is the analog of the Kolmogorov forward equation; Equ.(4.12) reduces to Equ.(3.55) as the observation noise approaches ∞. The conditional mean cannot in general be computed with a finite dimensional recursive filter as seen by letting $\phi(x) = x$ in Equ.(4.12). Hence, $\pi_t(f)$, $\pi_t(hx)$ and $\pi_t(h)$ are necessary for the computation of $\hat{x}_t$

$$\hat{x}_t = \hat{x}_0 + \int_0^t \pi_s(f)ds + \int_0^t [\pi_s(hx) - \pi_s(h)\hat{x}_s]d\nu_s \tag{4.14}$$

One case in which this is possible is in the ***Kalman-Bucy Filter***.

4.3 The Unnormalized Equations

The conditional measure π_t satisfies Equ.(4.12), however, is often more convenient to work with a less complicated equation, which is obtained by considering an unnormalized version of π_t. These equations were originally derived by Zakai [69] and Wong [67]. We will adhere to the derivation by Davis et al [15], page 69, because it is more concise and is based on Equ.(4.12) and Ito's rule. We make the following assumptions:

1. $\{x_t\}$ is a homogeneous Markov process with infinitesimal generator $\mathcal{L}$,

2. $\{y_t\}$ is given by Equ.(4.1),

3. $z_t = h(x_t)$ and

4. $\{x_t\}$ and $\{v_t\}$ are independent.

To derive the unnormalized equations, consider the first step to define a measure P_0 on the measurable space $(\Omega, \mathcal{F})$ by

$$P_0(A) = \int_A \frac{dP_0}{dP} P(d\omega) \tag{4.15}$$

$\forall A \in \mathcal{F}$ where

$$\frac{dP_0}{dP} = \exp\left(-\int_0^T h(x_s)dy_s + \frac{1}{2}\int_0^T h^2(x_s)ds \right) \tag{4.16}$$

is the Radon-Nikodyn derivative of P_0 with respect to P. Under this new measure, Marcus [51], $\{y_t\}$ is a standard Brownian motion, $\{x_t\}$ and $\{y_t\}$ are independent and the distributions for $\{x_t\}$ remain invariant.

Lemma 4.3.1 (page 9 Wong [67]). *P_0 has the following properties*

1. *P_0 is a probability measure with $P_0(\Omega) = 1$*
2. *under P_0 $\{y_t\}$ is a standard Brownian motion*
3. *under P_0 $\{x_t\}$ and $\{y_t\}$ are independent*
4. *$\{x_t\}$ has the same distribution under P_0 as under P*
5. *P is absolutely continuous with respect to P_0 with Radon-Nikodyn derivative*

$$\frac{dP}{dP_0} = \left(\frac{dP_0}{dP}\right)^{-1} \tag{4.17}$$

$$= \exp\left(\int_0^T h(x_s)dy_s - \frac{1}{2}\int_0^T h^2(x_s)ds \right) \tag{4.18}$$

Hence, conditional statistics of x_t given $\mathcal{Y}_t$ in terms of the original measure P can be calculated in terms of conditional statistics under the measure P_0 by a Bayes formula, page 818 Marcus [51]:

$$\langle \pi_t, \phi \rangle = \pi_t(\phi) := E[\phi(x_t) \mid \mathcal{Y}_t] = \frac{E_0[\phi(x_t)\Lambda_t \mid \mathcal{Y}_t]}{E_0[\Lambda_t \mid \mathcal{Y}_t]} := \frac{\sigma_t(\phi)}{\sigma_t(1)} \tag{4.19}$$

where E_0 is the expectation with respect to P_0 and

$$\Lambda_t := \exp\left(\int_0^T h(x_s)dy_s - \frac{1}{2}\int_0^T h^2(x_s)ds\right) \tag{4.20}$$

is a martingale with respect to $\mathcal{F}_t$ and P_0 so that

$$\Lambda_t = E_0\left[\frac{dP}{dP_0} \mid \mathcal{F}_t\right] \tag{4.21}$$

We now require a recursive equation for the measure σ_t. Since $\sigma_t(\phi) = \sigma_t(1)\pi_t(\phi)$ we have the equation for $\pi_t(\phi)$, an equation for $\sigma_t(\phi)$ is derived by finding a stochastic differential equation for $\sigma_t(1) = E_0[\Lambda_t\ \mathcal{Y}_t]$ and applying Ito's rule.

Lemma 4.3.2 (page 70 Davis et al [15]). $E_0[\Lambda_t \mid \mathcal{Y}_t]$ *is given by the formula*

$$\hat{\Lambda}_t = E_0[\Lambda_t \mid \mathcal{Y}_t] = \exp\left(\int_0^T \pi_s(h)dy_s - \frac{1}{2}\int_0^T \pi_s^2(h)ds\right) \tag{4.22}$$

Theorem 4.3.1 (page 72 Davis et al [15]). *For any* $\phi \in \mathcal{D}(\mathcal{L})$, $\sigma(\phi)$ *satisfies*

$$\sigma_t(\phi) = \sigma_0(\phi) + \int_0^t \sigma_s(\mathcal{L}\phi)ds + \int_0^t \sigma_s(h\phi)dy_s \tag{4.23}$$

The Stratonovich version of Equ.(4.23) is

$$\sigma_t(\phi) = \sigma_0(\phi) + \int_0^t \sigma_s(\tilde{\mathcal{L}}\phi)ds + \int_0^t \sigma_s(h\phi) \circ dy_s \tag{4.24}$$

where

$$\tilde{\mathcal{L}}\phi(x) = \mathcal{L}\phi(x) - \frac{1}{2}h^2(x)\phi(x) \tag{4.25}$$

Since $\{\sigma_t(\phi), \phi \in \mathcal{D}(\mathcal{L})\}$ determines a measure-valued stochastic process, $\sigma_t(\phi)$ can be regarded as a recursive infinite dimensional linear stochastic differential equation for the unnormalized conditional measure σ_t of x_t given $\mathcal{Y}_t$.

If we assume that x_t is a diffusion process and that the unnormalized conditional measure has a sufficiently smooth density $p(t,x)$, that is:

$$\sigma_t(\phi) = \int \phi(x)p(t,x)dx \tag{4.26}$$

then under appropriate hypotheses one can obtain from Equ.(4.23) stochastic partial differential equations for $p(t,x)$, Kunita [39]. In particular, we assume that x_t is a diffusion of the form given by Equ.(3.48) with generator $\mathcal{L}$. Then

$$dp(t,x) = \mathcal{L}^* p(t,x)dt + p(t,x)h(x)dy_t \tag{4.27}$$

$$dp(t,x) = \left[\mathcal{L}^* - \frac{1}{2}h^2(x)\right] p(t,x)dt + p(t,x)h(x) \circ dy_t \tag{4.28}$$

where $\mathcal{L}^*$ is the formal adjoint of $\mathcal{L}$. The above equation is a bilinear stochastic partial differential equation, which we call the Zakai equation. Note that

$$\pi_t(\phi) = \frac{\int_{-\infty}^{\infty} \phi(x)p(t,x)dx}{\int_{-\infty}^{\infty} p(t,x)dx} \tag{4.29}$$

is used to determine the estimated state of the system.

Chapter 5

THE KALMAN-BUCY FILTER IN RELATION TO THE NONLINEAR FILTERING PROBLEM

In this chapter we explore the relationship between the nonlinear filtering theory and the Kalman-Bucy filter.

5.1 The Kalman-Bucy Filter

Consider the operator, page 20-21 Crisan [10],

$$A(t)\varphi(x) = \sum_{i=1}^{d} F_i(t)\frac{\partial\varphi(x)}{\partial x^i} + \sum_{i,j=1}^{d} Q_{ij}(t)\frac{\partial^2\varphi(x)}{\partial x^i\partial x^j} \tag{5.1}$$

Let $\mathcal{T} = [0,T]$, we assume that $F_i, Q_{ij} \in C(\mathcal{T})$ and that $Q_{ij} \geq 0$. Consider the case where x_t

$$x_t = x_0 + \int_0^t F(s)x_s ds + \int_0^t Q(s)dB_s^1 \tag{5.2}$$

is the solution of the martingale problem, with infinitesimal operator A. Consider an observation process y_t given by

$$y_t = x_0 + \int_0^t H(s)x_s ds + \int_0^t R(s)dB_s^2 \tag{5.3}$$

where $\{B_t^1, \mathcal{F}_t\}$ and $\{B_t^2, \mathcal{F}_t\}$, $t \in \mathcal{T}$, are standard independent Brownian motions. We assume that the initial data π_0 is given by

$$\pi_0(\psi) = \int \psi(x_0 + P_0^{\frac{1}{2}}\xi)\frac{\exp^{-\frac{1}{2}|\xi|^2}}{(2n)^{\frac{n}{2}}}d\xi \tag{5.4}$$

where x_0, P_0 are given. P_0 is symmetric and nonnegative. Therefore, π_0 is Gaussian with mean x_0 and covariance matrix P_0.
Bensoussan [6] proves that the solution of the Zakai and Kushner equations exist and are unique and proves the following theorem

Theorem 5.1.1 (page 101 Bensoussan [6]). *The unnormalized conditional density of the signal given the observation is Gaussian with mean $\hat{x}_t$ and covariance matrix P_t is given by*

$$p_t(\varphi) = \left[\int \varphi(\hat{x}_t + P_t^{\frac{1}{2}}\xi)\right] s_t \frac{\exp^{-\frac{1}{2}|\xi|^2}}{(2n)^{\frac{1}{2}}} d\xi \tag{5.5}$$

where P_t is the solution of the Riccati equation

$$\begin{aligned} \dot{P} + PH^T R^{-1} HP - Q - FP - PF^T &= 0 && (5.6) \\ P(t) &= P_0 && (5.7) \end{aligned}$$

and $\hat{x}_t$ is the Kalman filter solution of

$$\begin{aligned} d\hat{x}_t &= F\hat{x}_t dt + PH^T R^{-1}[dy_t - H\hat{x}_t dt] && (5.8) \\ \hat{x}(0) &= x_0 && (5.9) \end{aligned}$$

The process s_t is given by

$$s_t = \exp^{\left[\int_0^t (\hat{x}^T H^T) R^{-1} - \frac{1}{2}\int_0^t ((\hat{x}^T H^T) R^{-1}(\hat{x}H)) ds\right]} \tag{5.10}$$

Part III

Stochastic Differential Geometry and Geometric Nonlinear Filtering Theory

Chapter 6

STOCHASTIC CALCULUS ON MANIFOLDS

The most successful formulation of Brownian motion on a manifold is the Eells-Elworthy-Malliavin construction, which is founded on the idea that parallel transport and horizontal lifts can be applied to manifold-valued semimartingales. We have seen that when a manifold is equipped with a connection that it is possible to lift a curve to a horizontal curve on the frame bundle $\mathcal{O}(\mathcal{M})$. If $c : I \to \mathcal{M}$ is such a curve, choose the frame $u(0)$ over $c(0)$ and let $u(t)$ be the parallel transport of $u(0)$ along $c[0,t]$, which is achieved by solving an ordinary differential equation in local charts. The lift $\{u(t) \mid t \in I\}$ defines a unique smooth curve $q(t) = u(t)^{-1} \cdot c(t)$ in the Euclidean space of the same dimension. The curve q is called an anti-development of c. Therefore, there is a one-to-one correspondence between a set of smooth curves on a manifold, starting at some fixed point, and their anti-developments in Euclidean space up to the action by the general linear group. This construction can be extended to semimartingales on a manifold that is equipped with a connection by solving an appropriate horizontal stochastic differential equation on the frame bundle.

6.1 Calculus on Manifolds

Before we proceed, we need to formally define a semimartingale $\Sigma = \{\Sigma_t\}$, $t \geq 0$, on an m-dimensional manifold $\mathcal{M}$. We assume that all processes are defined on a fixed probability space $(\Omega, \mathcal{F}, P)$ and are $\{\mathcal{F}_t\}$-adapted.

Definition 6.1.1 (page 19 Hsu [25]). *A semimartingale $\Sigma = \{\Sigma_t\}$, $t \geq 0$ on $\mathcal{M}$ is an $\mathcal{M}$-valued $\{\mathcal{F}_t\}$-adapted process such that $\{f(\Sigma_t)\}$, $t \geq 0$, is a real-valued semimartingale for all smooth real-valued functions f on $\mathcal{M}$.*

From Stratonovich calculus in Euclidean spaces, we know that the Ito Formula gives rise to the following equation, compare with Equ.(3.34):

$$d\phi(\Sigma) = \phi_x(\Sigma) \circ d\Sigma \tag{6.1}$$

where $\phi \in C^3$. For simplicity, we follow Kendall's notation, Kendall [34] & Kendall [35], and rewrite the above as

$$d_S\phi(\Sigma) = \phi_x(\Sigma) d_S\Sigma \tag{6.2}$$

This shows that $d_S\Sigma$ behaves as an ordinary differential. Thus, if Σ is an $\mathcal{M}$-valued semimartingale, then its Stratonovich differential can be formally associated with a tangent vector. Let $\phi : \mathcal{M} \to \mathcal{N}$ be a map between two manifolds. Then the tangent map of ϕ is $T\phi : T\mathcal{M} \to T\mathcal{N}$. It can then be shown, Kendall [34], that the semimartingale $\phi(\Sigma)$ satisfies the following equation:

$$d_S\phi(\Sigma) = T\phi(\Sigma) d_S\Sigma \tag{6.3}$$

Equ.(6.3) is invariant of the choice of coordinate system chosen. Consider the following sets, which define the tangent spaces associated with the semimartingales Σ and $\phi(\Sigma)$, respectively,

$$T_\Sigma\mathcal{M} = \{T_x\mathcal{M} \mid x \in \Sigma\} \tag{6.4}$$

$$T_{\phi(\Sigma)}\mathcal{N} = \{T_x\mathcal{N} \mid x \in \phi(\Sigma)\} \tag{6.5}$$

The Stratonovich differentials $d_S\Sigma$ and $d_S\phi(\Sigma)$ can then be associated with the sets $T_\Sigma\mathcal{M}$ and $T_{\phi(\Sigma)}\mathcal{M}$, respectively.

6.2 Semimartingales on Manifolds - Stochastic Development and Parallel Transport

Stochastic development and parallel transport is essentially a technique that relies on geometric tools to *flatten* out a semimartingale Σ on an m-dimensional manifold

$\mathcal{M}$ by associating it with a system of Stratonovich stochastic differential equations driven by a real-valued semimartingale $\Lambda = \{\Lambda_t\}$, $t \geq 0$ that resides in a reference Euclidean space $V = \mathbb{R}^m$ of the same dimension as $\mathcal{M}$.

We start by considering the linear isometry, which relates the differential $d_S\Lambda \in V$ to the differential $d_S\Sigma \in T_\Sigma\mathcal{M}$

$$\Xi : V \to T_\Sigma\mathcal{M} \tag{6.6}$$

In order to define the Stratonovich stochastic differential equation from this map

$$\Xi d_S\Lambda = d_S\Sigma \tag{6.7}$$

we require that Ξ itself be a manifold-valued semimartingale $\Xi = \{\Xi_t\}$, $t \geq 0$ in the orthonormal frame bundle $\mathcal{O}(\mathcal{M})$. For this definition to be correct, we require a rule for the evolution of Ξ in terms of Σ or Λ. We therefore require a rule for lifting $d_S\Sigma \in T_\Sigma\mathcal{M}$ to the frame bundle to drive Ξ. This follows naturally since the orthonormal frame bundle $\pi : \mathcal{O}(\mathcal{M}) \to \mathcal{M}$ is a principal fibre bundle that is equipped with a connection, refer to Sec. 2.2.3.

Given that $\Xi \in \mathcal{O}(\mathcal{M})$, we require the projection map of the orthonormal frame bundle to map Ξ to Σ on $\mathcal{M}$

$$\pi : \mathcal{O}(\mathcal{M}) \to \mathcal{M}, \Xi \mapsto \pi(\Xi) := \Sigma \tag{6.8}$$

Associated with π is the differential

$$T\pi : T_\Xi\mathcal{O}(\mathcal{M}) \to T_\Sigma\mathcal{M}, d_S\Xi \mapsto T\pi(d_S\Xi) := d_S\Sigma \tag{6.9}$$

Now, from fibre bundle theory we have the horizontal lift of the connection on $\mathcal{O}(\mathcal{M})$, Def. 2.1.12, which for each $u \in \mathcal{O}(\mathcal{M})$, where $u : V \to T_x\mathcal{M}$ and $x \in \mathcal{M}$, connects $T_{\pi u}\mathcal{M}$ to $T_u\mathcal{O}(\mathcal{M})$

$$h_u : T_{\pi u}\mathcal{M} \to T_u\mathcal{O}(\mathcal{M}), \tag{6.10}$$

If h varies smoothly with u then

$$d_S\Xi = h_\Xi d_S\Sigma \tag{6.11}$$

is a stochastic differential equation with smooth coefficients, which has a solution that exists up to a possibly finite explosion time. We summarize the process using the commutative diagram in Fig. 6.1

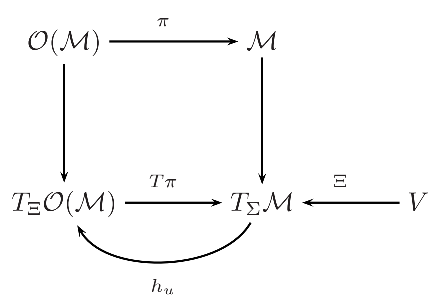

Figure 6.1: Stochastic Development and Parallel Transport

We call Λ and Ξ the ***stochastic development*** and ***stochastic parallel transport*** of Σ, respectively.
We need to ensure that the connection respects the following requirements, Kendall [34]:

1. the relation $\pi(\Xi) = \Sigma$ must be maintained otherwise the differential $\Xi d_S\Lambda$ will not be relevant for Σ; upon differentiation both sides we have the following requirement

$$(T\pi)(h_u d_S\Sigma) = d_S\Sigma \tag{6.12}$$

   this leads to the requirement

$$T\pi(h_u) = id_{T\pi u\mathcal{M}},\ \forall u \in \mathcal{O}(\mathcal{M}) \tag{6.13}$$

which is exactly the condition that is respected in Equ.(2.48) of Def. 2.1.12

2. the equivariance condition; though Ξ depends on its initial $\Xi_0 = u$, a change in the initial condition should only have the effect of subjecting the solution Ξ to a fixed isometrical transformation of V, which is a consequence of Equ.(2.49) of Def. 2.1.12

There is another restriction that we impose on the connection: it should have zero torsion. As we have seen earlier, there is one and only one such orthonormal connection for each Riemannian manifold, the Levi-Civita connection, refer to Def. 2.2.1. This brings us to the following important theorem:

Theorem 6.2.1 (Equations for Stochastic Development and Parallel Transport - page 11 Kendall [35]). *To each manifold-valued semimartingale Σ in an m-dimensional Riemannian manifold $\mathcal{M}$ we can associate a stochastic anti-development Λ in $\mathbb{R}^m$ and a stochastic parallel transport ζ in the orthonormal frame bundle $\mathcal{O}(\mathcal{M})$ such that*

$$d_S\Sigma = \Sigma d_S\Lambda \tag{6.14}$$

$$d_S\Xi = h_\Xi d_S\Sigma \tag{6.15}$$

Furthermore, Λ, Ξ are unique when the initial frame Ξ_0 is prescribed (and depending equivariantly on this choice); and on the other hand Σ can be defined by prescribing the initial condition Ξ_0 and hence Σ_0 and the driving flat semimartingale Λ.

Intuitively, the stochastic development Λ can be associated with the trajectory generated on V when V is rolled without slipping on $\mathcal{M}$ with the constraint that the point of contact on $\mathcal{M}$ is along the semimartingale Σ on $\mathcal{M}$. Also, the parallel transport Ξ can be associated with an inertial reference frame that is attached to the semimartingale Σ which encodes the orientation of V as it rolls. The importance of the connection in this construction justifies the detailed analysis of principal fibre bundles, refer to Sec. 2.1.3.

The stochastic differential equations may be used to categorize semimartingales Σ on $\mathcal{M}$ in terms of their stochastic development Λ on V. There are two process that are particularly important. When Λ is curve on V the process Σ on $\mathcal{M}$ is said to be a **Γ*-geodesics***. If Λ is Brownian motion on V then we say that Σ is a **Γ*-Brownian*** motion on $\mathcal{M}$. These processes are prefixed with the symbol Γ to emphasize the importance of the connection (Γ is a symbol that conventionally represents the coordinate representation of a connection).

6.3 Local Description of Γ-Geodesics and Γ-Brownian Motion on $\mathcal{M}$

In this section, we refer to Ikeda et al [26] to illustrate how stochastic development and parallel transport may be used to provide a local description of Γ-geodesics and Γ-brownian motion on $\mathcal{M}$.

Let $\mathcal{M}$ be an n-dimensional Riemannian manifold equipped with an affine connection $\nabla = \{\Gamma_{ij}^k\}$ compatible with the Riemannian metric g. Using the connection ∇, the manifold $\mathcal{M}$ can be rolled along a curve $\gamma(t)$ in $\mathbb{R}^n$ to generate a curve $c(t)$ on $\mathcal{M}$. The infinitesimal motion of $c(t)$ at $x \in \mathcal{M}$ is that of $\gamma(t)$ in $T_x\mathcal{M} \cong V$. By choosing an orthonormal frame, the infinitesimal motion of the frame is given by the connection.

Let $\gamma : I = [0, \infty) \to V$ be a smooth curve. Let $u = (x, e) \in \mathcal{O}(\mathcal{M})$. Define the curve $\tilde{c}(t) = (c(t), e(t))$ in $\mathcal{O}(\mathcal{M})$

$$\begin{aligned} \tfrac{dc}{dt}(t) &= \tfrac{d\gamma^\alpha}{dt}(t)e_\alpha(t),\ c(0) = x \\ \nabla_{\dot{c}(t)}e(t) &= 0,\ e(0) = e \end{aligned} \tag{6.16}$$

In local coordinates we have

$$\begin{aligned} \tfrac{dc^i}{dt}(t) &= e^i_\alpha(t)\tfrac{d\gamma^\alpha}{dt}(t)e_\alpha(t),\ c^i(0) = x^i \\ \tfrac{de^l_\alpha}{dt}(t) &= -\Gamma^i_{ml}(c(t))e^l_\alpha(t)\tfrac{dc^m}{dt}(t),\ e^i_\alpha(0) = e^i_\alpha \end{aligned} \tag{6.17}$$

where $i, \alpha = 1, \ldots, n$. The coefficients of the Levi-Civita connection Γ^i_{ml} in Equ.(6.17) have been define in Equ.(2.81). Equ.(6.16) may be written more compactly as follows:

$$\frac{d\tilde{c}}{dt}(t) = H_\alpha(\tilde{c}(t))\frac{d\gamma^\alpha}{dt}(t),\ \tilde{c}(0) = u \tag{6.18}$$

where H_α's in Equ.(6.18) are the fundamental horizontal vector fields defined in Equ.(2.87).

The curve $c(t) = \pi(\tilde{c}(t))$ in $\mathcal{M}$ is a Γ-geodesic on $\mathcal{M}$ which depends on the initial frame e at $x \in \mathcal{M}$. Let us denote the curve $c(t)$ by $c(t, u, \gamma)$, where $u = (x, e)$. It can be shown that, Ikeda et al [26] page 283,

$$c(t, ua, \gamma) = c(t, u, a\gamma),\ t \in [0, \infty),\ a \in O(n) \tag{6.19}$$

where the curve $(a\gamma)(t) = a\gamma(t)$.

The above discussion easily generalizes to Γ-Brownian motion as follows. Let $w(t) = (w^\alpha(t))$ be a canonical realization of n-dimensional Brownian motion. Let $u(t) = (u(t, u, w))$ be a solution of the stochastic differential equation

$$du(t) = H_\alpha(u(t)) \circ dw^\alpha(t),\ u(0) = u \tag{6.20}$$

In local coordinates, Equ.(6.20) is equivalent to, Ikeda et al [26] page 284,

$$\begin{aligned} d\Sigma^i(t) &= \Xi^i_\alpha(t) \circ dw^\alpha(t) \\ d\Xi^i_\alpha(t) &= -\Gamma^i_{mk}(\Sigma(t))\Xi^k_\alpha(t) \circ d\Sigma^m(t) \end{aligned} \tag{6.21}$$

where $i, \alpha = 1, \ldots, n$. The solution $u(t) = (\Sigma^i(t), \Xi^i_\alpha(t))$ lies on $\mathcal{O}(\mathcal{M})$ if $u(0) \in \mathcal{O}(\mathcal{M})$. The stochastic curve $\Sigma(t)$ on $\mathcal{M}$ is defined by $\Sigma(t) = \pi(u(t))$.

6.4 Nonlinear Filtering on Manifolds

An important application of the above theory concerns nonlinear estimation on manifolds.

The first example we consider deals with the nonlinear filtering on the unit circle S^1, which has important applications in engineering, for example FM demodulation, frequency stability and single degree of freedom gyroscope analysis. The problem was originally investigated by Lo et al [50], who proved that linear filtering techniques can be applied together with some nonlinear transformations. The first geometric analysis of the problem was initiated by James [29] who showed that differential geometry is a natural mathematical setting for solving the nonlinear estimation problem because S^1 has the properties of a manifold and Lie group, Fig. 6.2

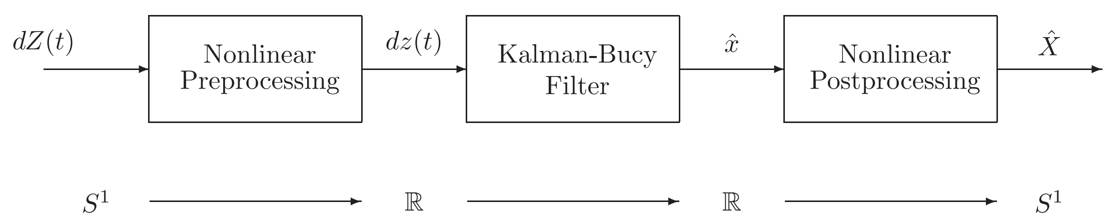

Figure 6.2: Nonlinear Filtering on Manifolds

However, the problem of generating Brownian motion on S^1 using tools from stochastic differential geometry was not completely addressed. The theory covered in this chapter together with the examples discussed in preceding chapters addresses this problem.

Other applications include more sophisticated filtering techniques like the geometrically intrinsic nonlinear recursive filter due to Darling [13], [11] & [12].

Chapter 7

GEOMETRIC NONLINEAR FILTERING THEORY

This chapter is largely based on a collection of papers by Mitter et al [55] and Lara [44], [43], [46], [45] & [47]. It is aimed at emphasizing the role of geometric tools in the analysis of the Zakai equation, which is concerned with the classification of the dimension and the reduction of the Zakai equation through symmetry.

7.1 Introduction

Consider the partially observed stochastic process given by Equ.(1.1) and Equ.(1.2). The unnormalized conditional law σ_t satisfies

$$d\sigma_t(\phi) = [M_0\phi]\sigma_t(\phi)dt + \sum_{i=1}^{p}[M_i\phi]\sigma_t(\phi) \circ dy_t^i, \ \sigma_0 = \mu_0 \tag{7.1}$$

for some $\phi \in \mathcal{D}(\mathbb{R}^n)$, which is just the Stratonovich differential equation of Equ.(4.24) extended for a p-dimensional observation process. The operators $M_0, M_1, \ldots, M_p$ are defined as follows:

$$M_0\phi = \mathcal{L}\phi + H\phi \tag{7.2}$$

$$M_i\phi = h_i\phi, \ i = 1, \ldots, p \tag{7.3}$$

where

$$\begin{aligned}\mathcal{L}\phi &= \frac{1}{2}\sum_{k=1}^{m} L_{g_k}^2\phi + L_f\phi \\ &= \frac{1}{2}\sum_{i,j=1}^{m}\sigma_{ij}(x_t)\frac{\partial^2}{\partial x^i \partial x^j}\phi + \sum_{i=1}^{m} f^i(x)\frac{\partial}{\partial x^i}\phi \qquad (7.4) \\ H\phi &= -\frac{1}{2}h^2\phi \qquad (7.5)\end{aligned}$$

The $\sigma_{ij}(x)$ is just the ij-th element of the matrix $\sigma(x_t) = g(x_t)g(x_t)^T$. In the above, we make use of the Lie derivative notation, where for any vector field $X = \sum_{i=1}^{n} X^i \frac{\partial}{\partial x^i}$ and a smooth function ϕ

$$L_X\phi(\xi) = \sum_{i=1}^{n} X^i \frac{\partial \phi}{\partial x^i} \qquad (7.6)$$

By convention, $L_X^0\phi = \phi$ and $L_X^k\phi = L_X(L_X^{k-1})\ \forall k \geq 1$. If the unnormalized conditional density $p(x,t)$ of the measure σ_t is smooth, then it satisfies the following stochastic partial differential equation:

$$dp(x,t) = M_0^* p(x,t)dt + \sum_{i=1}^{p} M_i^* p(x,t) \circ dy_t^i \qquad (7.7)$$

which is a generalization of Equ.(4.28) given a p-dimensional observation process.

In the sections that follow, we apply geometric tools to classify the dimension of the estimation algebra for the nonlinear filtering problem. We also introduce some background to the infinitesimal symmetries of a parabolic operator and show how these techniques may be applied to reduce the Zakai equation to a stochastic partial differential equation on a lower dimensional space.

7.2 Estimation Algebra

We are now ready to introduce the Lie estimation algebra associated with the Zakai equation. Let us first emphasize the importance this concept has in filtering theory.

In the linear Gaussian filtering problem, the solution of the Zakai equation evolves in the domain of unnormalized Gaussian measures. The corresponding finite number of parameters satisfy a stochastic differential equation on a finite dimensional manifold, which is driven by the observations of the physical. The extension of this idea to nonlinear filtering problems has inspired the notion of finite dimensional filters. The Lie estimation algebra associated the Zakai equation, which we denote by $\mathcal{LA}$, plays an important role in identifying whether finite dimensional filters may exist for nonlinear filtering problems.

Let us consider a simplification in the generalized Zakai equation Equ.(7.7) for the 1-dimensional state and observation equation given by Equ.(1.1) and Equ.(1.2) respectively.

$$dp(x,t) = M_0^* p(x,t)dt + M_1^* p(x,t) \circ dy_t \tag{7.8}$$

We then have

Proposition 7.2.1 (page 206 Mitter [54]). $\mathcal{LA} = \{M_0^*, M_1^*\}$ *is finite dimensional only in the case where the nonlinear state and measurement coupling functions f and h respectively, in Equ.(1.1) and Equ.(1.2), have the following form:*

1. $h = ax + \beta$
2. $f_x + f^2 = ax^2 + bx + c$, *this is called the Riccati equation*

where the Lie algebra operators are computed on $C^\infty(\mathbb{R})$.

When we seek the diffusion process x_t to be globally defined on $\mathbb{R}^1$, we require the coefficients of the Riccati equation in Pro. 7.2.1 to have the following properties:

1. $a \geq 0$ and
2. $(a, b, c) \neq 0$

If the Lie algebra $\mathcal{LA} = \{M_0^*, M_1^*\}$ is finite dimensional, then it should be possible to construct a filter by integrating the Lie algebra. However, if the Lie algebra is infinite dimensional we cannot represent the solution of the Zakai equation by means of a finite dimensional sufficient statistic which is represented by the stochastic differential equation

$$d\alpha_t = a(\alpha_t)dt + b(\alpha_t)dy_t \tag{7.9}$$

Assume that we are interested in determining the unnormalized conditional statistics

$$\hat{\phi}_t = \int_{\mathbb{R}} \phi(x)p(t,x)dx \tag{7.10}$$

$\hat{\phi}_t$ may be represented by a function of the solution to Equ.(7.9)

$$\hat{\phi}_t = c(\alpha_t) \tag{7.11}$$

Brockett proved, page 210 Mitter [54], that this occurs when there exists a homomorphism between the Lie algebra of operators $\mathcal{LA}\{M_0^*, M_1^*\}$ and the Lie algebra of vector fields $\mathcal{LA}\{a - \frac{1}{2}b_x b, b\}$.

7.3 Symmetry and Reduction

We will now introduce the theory that provides conditions and methods under which the general Zakai equation Equ.(7.7) with a solution on $\mathbb{R}^n$ can be reduced to a stochastic partial differential equation with a solution on a lower dimensional space $\mathbb{R}^m$, where $m < n$. We first introduce the group invariant techniques for the deterministic case and then extend it to the stochastic case.

7.3.1 Infinitesimal Symmetries of a Parabolic Operator

Consider a second-order time-varying linear partial differential operator on $\mathbb{R}^n$ of the form

$$A = \sum_{i,j=1}^{n} a^{ij}(x)\partial^2_{x_i x_j} + \sum_{i=1}^{n} b^i(t,x)\partial_{x_i} + c(t,x) \tag{7.12}$$

where:

1. the functions $a^{ij}(\cdot), b^i(\cdot), c(\cdot,\cdot) \in C^\infty(\mathbb{R}^n)$

2. $(a^{ij}(x))_{i,j=1,\ldots,n}$ is positive definite $\forall x \in \mathbb{R}^n$, making A elliptic

Then the associated parabolic equation given A is

$$\partial_t u - Au = 0 \tag{7.13}$$

The set of all infinitesimal symmetries of this form is denoted by the set $\mathfrak{G}_{\partial_t - A}$. If u satisfies Equ.(7.13), then the symmetry group of $\partial_t - A$ allows one to determine other solutions by taking the graph of u onto the graphs of solutions of Equ.(7.13), page 1444 Lara [42].
It turns out that the set of admissible perturbations of the operator A forms a Lie algebra of operators of order less or equal to one, we call this the perturbation algebra $\mathscr{P}_A$, which is isomorphic to the Lie algebra of infinitesimal generators of the symmetry group of the parabolic operator $\partial_t - A$. Lara [42] proves that this symmetry group is finite dimensional by relating it to the Lie group of homothetic transformations on a Riemannian manifold. Lara [42] introduces geometric tools to prove this result, furthermore, these tools are used to characterize $\mathscr{P}_A$.
The application of this theory to nonlinear filtering can be motivated if we consider the Zakai equation for a partially observed stochastic process as a perturbation of a deterministic parabolic differential equation.

Time-Dependent Case

Proposition 7.3.1 (page 128 Lara [44]). $\mathfrak{G}_{\partial_t - A}$ *is a Lie algebra that can be written as a direct sum*

$$\mathfrak{G}_{\partial_t - A} = [\mathfrak{G}_{\partial_t - A} \cap \mathfrak{U}(\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})] \oplus \mathfrak{G}^{\infty}_{\partial_t - A} \tag{7.14}$$

where:

1. $\mathfrak{U}(\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})$ *is a subalgebra of* $\mathfrak{X}(\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})$ *of vector fields* ζ *of the form*

$$\zeta = \zeta^0(t)\partial_t + \sum_{i=1}^{n} \zeta^i(t,x)\partial_{x_i} + \zeta^{n+1}(t,x)y\partial_y \tag{7.15}$$

2. $\mathfrak{G}^{\infty}_{\partial_t - A}$ *reflects the linearity of* ***Equ.(7.13)*** *and consists of vector fields of the form* $\zeta = u(t,x)\partial_y$ *with* u *a solution of* ***Equ.(7.13)****.*

The first part $[\mathfrak{G}_{\partial_t - A} \cap \mathfrak{U}(\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})]$ of the Lie algebra $\mathfrak{G}^{\infty}_{\partial_t - A}$ can be characterized as follows.

Definition 7.3.1 (page 128 Lara [44]). *Let* $A_{quad} = \sum_{i,j=1}^{n} a^{ij}(x)\partial^2_{x_i x_j}$ *be the second order part of* A *that is time independent. Let* $\mathcal{F}^A = RA_{quad} \oplus \mathfrak{X}(\mathbb{R}^n) \oplus C^{\infty}(\mathbb{R}^n)$ *denote the space of smooth linear operators* P *of order less that or equal to two on* $\mathbb{R}^n$ *of the form* $P\varphi = \alpha A_{quad}\varphi + X\varphi + m\varphi$ *for* $\varphi \in C^{\infty}(\mathbb{R}^n)$*, where* $(\alpha, X, m) \in \mathbb{R} \times \mathfrak{X}(\mathbb{R}^n) \times C^{\infty}(\mathbb{R}^n)$*. For any* $\zeta \in \mathfrak{U}(\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})$ *of the form given by* ***Equ.(7.15)*** *and* $t \in \mathbb{R}$*, let* $\hat{\zeta}_t$ *and* $\partial_t\hat{\zeta}_t \in \mathcal{F}^A$ *be given by*

$$\hat{\zeta}_t = -\zeta^0(t)A_{quad} - \sum_{i=1}^{n} \zeta^i(t,x)\partial_{x_i} + \zeta^{n+1}(t,x) \tag{7.16}$$

$$\partial_t\hat{\zeta}_t = -\partial_t\zeta^0(t)A_{quad} - \sum_{i=1}^{n} \partial_t\zeta^i(t,x)\partial_{x_i} + \partial_t\zeta^{n+1}(t,x) \tag{7.17}$$

This gives a characterization of $\mathfrak{G}_{\partial_t - A}$ *in terms of differential operators on* $\mathbb{R}^n$*.*

Theorem 7.3.1 (page 128 Lara [44]). *If* $\zeta \in \mathfrak{U}(\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})$*, then the following equivalence is satisfied:*

$$\zeta \in \mathfrak{G}_{\partial_t - A} \iff \forall t \in \mathbb{R},\ \partial_t\hat{\zeta}_t = [A, \hat{\zeta}_t] \tag{7.18}$$

Time-Independent Case

When the functions b^i and c in Equ.(7.12) do not depend on time, we have the following refinement:

Theorem 7.3.2 (page 128 Lara [44]). *There exists a finite-dimensional Lie algebra $\mathscr{P}_A$ called the perturbation algebra of the elliptic operator A, consisting of linear partial differential operators of order less that or equal to one on $\mathbb{R}^n$ such that $\forall \zeta \in \mathfrak{U}(\mathbb{R}^n), \zeta \in \mathfrak{G}_{\partial_t - A} \iff$*

$$\partial_t \hat{\zeta}_t = [A, \hat{\zeta}_t], \; \forall \tag{7.19}$$

$\iff$

$$\hat{\zeta}_t = \exp\left(t ad_A\right)(\hat{\zeta}_0) = \sum_{k=0}^{\infty} \frac{t^k}{k!} ad_A^k(\hat{\zeta}_0), \; \forall t \in \mathbb{R} \tag{7.20}$$

where $\hat{\zeta}_0 \in RA \oplus \mathscr{P}_A$.

The above result implies that the characterization of infinitesimal symmetries of Equ.(7.13) relies upon the characterization of the Lie algebra $\mathscr{P}_A$.

A Geometric Characterization of the Perturbation Algebra

Definition 7.3.2 (page 129 Lara [44]). *Let X, Y and Z be vector fields, α be a 1-form and φ a smooth function.*

1. *D is the Levi-Civita connection associated with the metric g. L_Z is the Lie derivative and A_Z the derivative $A_Z = L_Z - D_Z$.*

2. *A_Z induces a $(1,1)$-tensor field A_Z by $A_Z X = -D_X Z$ whose adjoint A_Z^* is defined by $g(A_Z^* X, Y) = g(X, A_Z Y)$*

3. *$Z^\flat$ is the one form defined by $< Z^\flat, X >= g(Z, X)$ ($\natural$ is the inverse operation of $\flat$).*

4. $\nabla_g \varphi$ *is the vector field* $(d\varphi)^\natural$*, that is* $g(\nabla_g \varphi, X) = X\varphi = L_X \varphi$

5. the divergence of Z *is the function* $div_g Z = -trace(A_Z)$ *and the Laplacian* $\triangle_g \varphi = div_g(\nabla_g \varphi)$

Definition 7.3.3 (page 129 Lara [44]). *$T \in \mathfrak{X}(\mathbb{R}^n)$ is said to be a homothetic infinitesimal transformation of $(\mathbb{R}^n, g)$ if there exists $\lambda \in \mathbb{R}$ such that $L_T g = \lambda g$. We denote by $\mathfrak{H}_g$ the space of all homothetic infinitesimal transformations of $(\mathbb{R}^n, g)$ and define the linear form η_g on $\mathfrak{H}$ by $\eta_g(T) = \lambda$ if $L_T g = \lambda g$. $\mathfrak{F}_g = Ker\eta_g$ is the space of infinitesimal isometries of $(\mathbb{R}^n, g)$.*

We have now characterized $\mathscr{P}_A$.

Theorem 7.3.3 (page 129 Lara [44]). *The smooth differential operator $X + m \in \mathfrak{X}(\mathbb{R}^n) \oplus C^\infty(\mathbb{R}^n)$ belongs to $\mathscr{P}_A \iff \exists (X_i)_{i\in\mathbb{N}}$, a sequence, in $\mathscr{H}_g(M)$ that satisfies one of the equivalent inductions:*

$$\begin{aligned} X_0 &= X \\ X_1 &= K_A X_0 + \nabla_g(m - g(X_0, B)) \\ X_{i+2} &= K_A X_{i+1} + \tfrac{1}{2}\nabla_g(L_{X_i} H_A + \eta_g(X_i) H_A) \end{aligned} \tag{7.21}$$

Here the skew-symmetric $(1,1)$ tensor field K_A and the function H_A are given by

$$K_A = A_U - A_U^*, \; H_A = div_g U + g(U,U) - 2c \tag{7.22}$$

Moreover, we have

$$ad_A^{k+1}(X + m) = \eta_g(X_k)A + X_{k+1} + m_{k+1} \tag{7.23}$$

where the sequence of functions m_k satisfy the induction

$$m_{k+1} = (\frac{1}{2}\nabla_g + U)m_k - (L_{X_k} + \eta_g(X_k))c, \; m_0 = m \tag{7.24}$$

7.3.2 Computation of Infinitesimal Symmetries

We are now in a position to extend the deterministic invariance group techniques to the stochastic case. This will allow us to identify solutions of the Zakai equation that have a certain degree of symmetry, that is, invariant under some group action. Such solutions enable us to reduce the Zakai to a stochastic partial differential equation of lower dimension.

The assumptions that we make are:

A1] The coefficients of the linear partial differential operators $M_0, M_1, \ldots, M_p$ are smooth functions.

A2] M_0 is elliptic $M_0 = \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(x)\partial^2_{x_i x_j} + \ldots$, then the symmetric matrix $(a^{ij}(x))_{i,j=1,\ldots,n}$ is positive definite $\forall x \in \mathbb{R}^n$.

Shifting from the symmetries of a stochastic PDE to symmetries of a family of PDE's

Let $(y(t))_{t\geq 0} = (y^1(t), \ldots, y^p(t))_{t\geq 0}$ be a smooth trajectory in $\mathbb{R}^p$. Consider a deterministic partial differential equation

$$\frac{\partial u}{\partial t} = M_0^* u + \sum_{k=1}^{p} y^k(t) M_k^* u \tag{7.25}$$

The primary objective now is to seek invariant solutions of Equ.(7.25) and to characterize them and examine their relation to the Zakai equation. Basically, this involves finding particular solutions of Equ.(7.25), this can be done by studying its invariances under group actions. We can adapt to the filtering case using the computations discussed by Lara [44]. Once a symmetry is identified, one can associate special solutions having these symmetries which satisfy simpler or reduced equations.

A Necessary and Sufficient Condition for Existence

We find that due to the parabolic nature of the Zakai equation there exists geometric conditions for the existence of infinitesimal symmetries, which prove the existence of a reduced equation.

Theorem 7.3.4 (page 124 Lara [44]). *Let $\hat{\zeta}_0$ be the linear partial differential operator of order less than or equal to two, given by*

$$\hat{\zeta}_0 = -\zeta^0(0)M_0^* - \sum_{i=1}^{n} \zeta^i(0,x)\partial_{x_i} + \zeta^{n+1}(0,x) \tag{7.26}$$

and let $\hat{\zeta}_0^*$ be the dual operator. Then, $\zeta \in \mathfrak{U}(\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})$ is an infinitesimal symmetry of the partial differential equation $\forall (y(t))_{t\geq 0} \iff$

$$\hat{\zeta}_0^* \in \mathbb{R}M_0 \oplus \mathscr{P}_{M_0} \qquad \text{and} \qquad \forall i = 1, \ldots, p \qquad ad_{M_i} ad_{M_0}^k(\hat{\zeta}_0^*) = 0, \qquad \forall k \in \mathbb{N} \tag{7.27}$$

Geometry of the Operator M_0 and the Existence of Infinitesimal Symmetries

The objective of this section is to show how the existence of nontrivial infinitesimal symmetries are related to the geometric properties of the parabolic operator M_0.

Lemma 7.3.1 (page 125 Lara [44]). *There exists a Riemannian metric g on $\mathbb{R}^n$ such that if ∇_g is the Laplace-Beltrami operator on the Riemannian space $V_n = (\mathbb{R}^n, g)$, then M_0 can be written in the following form*

$$M_0 = \frac{1}{2}\nabla_g + U + c \tag{7.28}$$

where U is a smooth vector field on $\mathbb{R}^n$ and c is a smooth function.

Proposition 7.3.2 (page 125 Lara [44]). *Let $M_i = h_i$, where $i = 1, \ldots, p$ be nonconstant functions. Also let $\zeta \in \mathfrak{U}(\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})$ be such that $\hat{\zeta}_0^* \in \mathscr{P}_{M_0}$ and let*

$(X_k)_{k\in\mathbb{N}}$ in $\mathscr{H}_g(M)$ be the associated sequence given by **Equ.(7.21)**. *Two cases are possible:*

1. *$\forall k \in \mathbb{N}$ such that $X_k \notin \mathscr{I}_g(M)$ $\eta_g(X_k) \neq 0$. Then ζ is an infinitesimal symmetry of* **Equ.(7.25)** *$\forall (y(t))_{t\geq 0} \iff$*

$$L_{X_i} h_i = 0, \qquad \forall i = 1, \ldots, p \tag{7.29}$$

2. *$\exists k \in \mathbb{N}$ such that $X_k \notin \mathscr{I}_g(M)(\eta_g(X_k) \neq 0)$. Then ζ is not an infinitesimal symmetry of* **Equ.(7.25)** *$\forall (y(t))_{t\geq 0}$*

Pro. 7.3.2 proves that the existence of infinitesimal symmetries is related to the geometry of the infinitesimal generator of the signal process.

Part IV

Numerical Solution of the Zakai Equation with Applications

Chapter 8

NUMERICAL TECHNIQUES

We have explored the importance of geometric tools in the analysis and reduction of nonlinear filtering problems. However, to take advantage of these in a practical situation involves solving the Zakai Equation, which is a complex problem because it is infinite dimensional.

8.1 Solution of the Nonlinear Filtering Problem

As stated previously in chapter 4, the nonlinear filter concerns computing the conditional law π_t of the state x_t given past observations $\mathcal{Y}_t$

$$\pi_t(\phi) = E[\phi(x_t) \mid \mathcal{Y}_t] \tag{8.1}$$

for a bounded continuous function $\phi : S \to \mathbb{R}^n$ and a positive finite measure μ on S

$$\langle \mu, \phi \rangle = \int_S \phi(x) d\mu(x) \tag{8.2}$$

The conditional law is a solution of the Kushner-Stratonovich equation, which is a nonlinear stochastic partial differential equation. We have shown that the conditional law can be obtained from an unnormalized positive measure σ_t

$$\mu_t(\phi) = \frac{\sigma_t(\phi)}{\sigma_t(1)} \tag{8.3}$$

where σ_t represents the solution of the Zakai equation, which is a linear stochastic partial differential equation. The Zakai equation was an important development because it is more amenable to analysis than the Kushner-Stratonovich equation. However, both the Kushner-Stratonovich and Zakai equation are classed as infinite dimensional stochastic differential equations. This implies that their solution through numerical techniques are not computable because the full solution requires updating an infinite dimensional vector. Therefore the solution of the nonlinear filtering problem must be approximated. This raises two important problems:

1. the form of the finite-dimensional measure μ we need to employ, and

2. an algorithm for updating the finite dimensional description as the observations become available

8.1.1 Approximation of the Finite-Dimensional Probability Measure

Generally, there are four techniques that may be used to give a reasonable finite approximation of the probability measure μ:

1. finite sampling

2. exponential family

3. truncated expansion

4. finite grid

We will discuss the last technique because it is more relevant to our problem, details regarding the other techniques can be obtained from Darling [14].

Finite Grid

This approach emanates from numerical techniques used to solve partial differential equations. This method sets up a grid space which assigns the probability $\mu[s]$ to each grid element s. If the grid has k elements for each dimension d, then k^d values have to be pre-allocated. The computational effort may be simplified by neglecting those grid elements for which $\mu[s] < \delta$, for some δ, and boosting other values so that they still sum to 1.

8.1.2 Algorithms for the Approximate Solution of the Nonlinear Filtering Problem

Darling [14] cites a multitude of algorithms that have been investigated in recent years:

1. classical Monte Carlo particle method
2. integration of measured-valued stochastic differential equation
3. space discretization of Markov chains
4. genetic resampling particle method
5. Wiener chaos expansion
6. minimum variance branching method
7. projection on finite-dimensional manifolds of densities
8. the extended Kalman filter

From the above, the extended Kalman filter is without question the most widely applied algorithm in the scientific and engineering community, Grewal et al [21]. For this reason, we discuss the algorithm in detail and use it to benchmark two algorithms that we investigated, namely:

1. solution by gauge transformation and semigroup techniques and
2. finite difference approximation

Gauge Transformation and Semigroup Techniques

In this section we describe a numerical technique for determining the unnormalized conditional density from the Zakai equation. Consider the matrix formulation of the Zakai equation given in the Ito form

$$
\begin{aligned}
dp(x,t) &= \mathcal{L}^* p(x,t)dt + h^T(x)p(x,t)dy_t && (8.4)\\
p(x,t) &= p_0(x) && (8.5)\\
\mathcal{L}^* p(x,t) &= \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x^i \partial x^j}[\sigma_{ij}(x)p(x,t)] - \sum_{i=1}^{n} \frac{\partial}{\partial x^i}[f_i(x)p(x,t)] && (8.6)\\
\sigma(x) &= \frac{1}{2} g(x) g^T(x) && (8.7)
\end{aligned}
$$

where $(x,t) = \Omega = \mathbb{R}^n \times [0,T]$ and $\mathcal{L}^*$ is just the formal adjoint of the diffusion operator $\mathcal{L}$. The entity $\sigma_{ij}(x)$ is the (i,j) element of the matrix $\sigma(x)$. We assume that the initial value for the unnormalized conditional density is $p_0(x)$. To ensure the Zakai equations is solvable, we make the following assumptions:

A1] $\mathcal{L}^*$ is uniformly elliptic in the sense that for some $\lambda > 0$

$$z^T \sigma(x) x \geq \lambda z^T z \qquad (8.8)$$

A2] the functions $f(x)$, $\sigma(x)$ and $h(x)$ together with $\frac{\partial}{\partial x_i} f_i(x)$, $\frac{\partial}{\partial x_i}\sigma_{ij}(x)$, $\frac{\partial}{\partial x_j}\sigma_{ij}(x)$, $\frac{\partial}{\partial x_i} h_k(x)$, $\frac{\partial^2}{\partial x_i \partial x_j} h_k(x)$ and $\frac{\partial^2}{\partial x_i \partial x_j}\sigma_{ij}(x)$ for $i,j = 1,\ldots,n$ and $k = 1,\ldots,p$ are uniformly bounded and Lipschitz continuous.

We employ a Gauge transformation to simplify the analysis and to prove the necessary convergence results. Once the unnormalized conditional density is transformed, it is approximated and transformed back to the original and thus yielding an approximation to the unnormalized conditional density.

We proceed by factoring $\mathcal{L}^*\ p(x,t)$, Equ.(8.6), into the following form:

$$\mathcal{L}^*p(x,t) = A^*p(x,t) + C(x) \tag{8.9}$$

where A^* is an operator that contains all the derivatives with respect to $p(x,t)$ and $C(x)$ is a term that has a multiplicative relationship to $p(x,t)$. Using Equ.(8.9) we can express Equ.(8.4) in the following form:

$$dp(x,t) = A^*p(x,t)dt + C(x)p(x,t)dt + h^T(x)p(x,t)dy_t \tag{8.10}$$

Equ.(8.10) makes numerical implementation easier. Defining

$$\phi(x, y_t, t) = h^T(x)dy_t + \left[C(x) - \frac{1}{2}\|h(x)\|^2\right]t \tag{8.11}$$

we use a Gauge transformation, Baras et al [4],

$$r(x,t) = \exp^{-\phi(x,y_t,t)} p(x,t) \tag{8.12}$$

which gives

$$p(x,t) = \exp^{\phi(x,y_t,t)} r(x,t) \tag{8.13}$$

This solution of the unnormalized conditional density must satisfy the Zakai equation, substituting $p(x,t)$, Equ.(8.13), in Equ.(8.10), we obtain

$$dr(x,t) = \exp^{-\phi(x,y_t,t)} \left(A^* \left[\exp^{\phi(x,y_t,t)} r(x,t)\right]\right) dt \tag{8.14}$$

which is a classical parabolic partial differential equation. Using semigroup theory, Baras et al [4], we are able to show that the solution to $dr(\mathbf{x},t)$, Equ.(8.14), is given by

$$r(x,t) = \exp^{-\phi(x,y_t,t)} \exp^{A^*t} \exp^{\phi(x,y_t,t)} p_0(x) \tag{8.15}$$

Thus, the unnormalized conditional density has a solution given by

$$p(x,t) = \exp^{A^*t} \exp^{h^T(x)dy_t - \frac{1}{2}\|h(x)\|^2 t + C(x)t} p_0(x) \tag{8.16}$$

Standard numerical techniques can be applied with respect to temporal and spatial discretization to obtain a solution for $p(x,t)$, for multidimensional problems the algorithms developed by Rodriguez [61] are recommended.

For our purpose consider a 1-dimensional Zakai equation

$$dp(x,t) = \mathcal{L}^* p(x,t)dt + h(x)p(x,t)dy_t \tag{8.17}$$

where

$$\begin{aligned}\mathcal{L}^* p(x,t) &= \bar{a}(x)\frac{\partial^2 p(x,t)}{\partial x^2} + \bar{b}(x)\frac{\partial p(x,t)}{\partial x} + C(x)p(x,t) \\ &= \left(\bar{a}(x)\frac{\partial^2}{\partial x^2} + \bar{b}(x)\frac{\partial}{\partial x}\right) p(x,t) + C(x)p(x,t) \\ &= A^* p(x,t) + C(x)p(x,t)\end{aligned} \tag{8.18}$$

The solution for the Zakai equation is

$$p(x,t) = \exp^{A^* t} \varphi(x,t) \tag{8.19}$$

where

$$\varphi(x,t) = \exp^{h(x)dy_t - \frac{1}{2}h^2(x)t + C(x)t} p_0(x) \tag{8.20}$$

The unnormalized density may be approximated by the Taylor series approximation for the exponential $\exp^{A^* t}$

$$\begin{aligned}p(x,t) &\approx [1 + A^* t]\varphi(x,t) \\ &= \varphi(x,t) + \left(\bar{a}(x)\frac{\partial^2 \varphi(x,t)}{\partial x^2} + \bar{b}(x)\frac{\partial \varphi(x,t)}{\partial x}\right) t\end{aligned} \tag{8.21}$$

We can now introduce the finite difference technique, Ames [2], to approximate the spatial derivatives of a general function θ at (x,t)

$$\frac{\partial\theta(x,t)}{\partial t} \approx \frac{\theta(x,t+\triangle t)-\theta(x,t)}{\triangle t} \quad (8.22)$$

$$\frac{\partial\theta(x,t)}{\partial x} \approx \frac{\theta(x+\triangle x,t)-\theta(x-\triangle x,t)}{2\triangle x} \quad (8.23)$$

$$\frac{\partial^2\theta(x,t)}{\partial x^2} \approx \frac{\theta(x+\triangle x,t)-2\theta(x,t)+\theta(x-\triangle x,t)}{\triangle x^2} \quad (8.24)$$

where $(\triangle x, \triangle t)$ represent the spatial and temporal perturbations about (x,t). Substituting Equ.(8.23) and Equ.(8.24) into Equ.(8.21) we obtain the approximate solution for $p(x,t)$

$$\begin{aligned} u(x,t) &= \varphi(x,t) + \left(\bar{a}(x)\frac{\varphi(x+\triangle x,t)-2\varphi(x,t)+\varphi(x-\triangle x,t)}{\triangle x^2}\right)t + \\ &\quad \left(\bar{b}(x)\frac{\varphi(x+\triangle x,t)-\varphi(x-\triangle x,t)}{2\triangle x}\right)t \end{aligned} \quad (8.25)$$

The above equation involves points separated by $\triangle x$ in space and $\triangle t$ in time, which can be associated with a uniform mesh with spatial and temporal discretization. If the spatial component has a range L, then it can be divided into N equally spaced discretized intervals of length $\triangle x = L/N$. Each point in the spatial range can be determined by

$$x_i = i\triangle x,\ 0 \leq i \leq N \quad (8.26)$$

Similarly, points in the temporal range are given by

$$t_j = j\triangle t,\ 0 \leq j \leq M \quad (8.27)$$

The exact solution of the unnormalized density $p(x,t)$ at (x_i, t_j) can be approximated by $u(x_i, t_j)$. We introduce the following notation

$$u(x_i, t_j) = u_{(i,j)} \quad (8.28)$$

$$\varphi(x_i, t_j) = \varphi_{(i,j)} \quad (8.29)$$

The discretized version of $\varphi(x,t)$ can be expressed as follows:

$$\varphi(x_i, t_j) = \exp^{h(x_i)\triangle y_j - \frac{1}{2}h^2(x_i)\triangle t + C(x_i)\triangle t} u(x_i, t_{j-1}) \tag{8.30}$$

The recursive update equation for $u_{(i,j+1)}$ is given by

$$u_{(i,j+1)} = \bar{c}_1 \varphi_{(i+1,j)} + \bar{c}_0 \varphi_{(i,j)} + \bar{c}_{-1} \varphi_{(i-1,j)} \tag{8.31}$$

where:

$$\begin{aligned} r &= \tfrac{\triangle t}{\triangle x^2} \\ \bar{c}_1 &= \bar{a}_i r + \tfrac{\bar{b}_i r}{2} \triangle x \\ \bar{c}_0 &= 1 - 2\bar{a}_i r \\ \bar{c}_{-1} &= \bar{a}_i r - \tfrac{\bar{b}_i r}{2} \triangle x \end{aligned} \tag{8.32}$$

Finite Difference Approximation

Kloeden [36], page 10, suggests that numerical methods for parabolic stochastic partial differential equations can be constructed by applying a scheme with constant time step to the n-dimensional Ito stochastic differential equation by finite difference approximations for the spatial derivatives.

To illustrate this technique, consider a 1-dimensional stochastic partial differential equation, which can be rewritten in the following form

$$\frac{\partial p(x,t)}{\partial t} = \tilde{a}(x,t)\frac{\partial^2 p(x,t)}{\partial x^2} + \tilde{b}(x,t)\frac{\partial p(x,t)}{\partial x} + \tilde{c}(x,t)p(x,t) + \tilde{d}(x,t) \tag{8.33}$$

The coefficient $\tilde{c}(x,t)$ is usually a function of a white noise processes, which is associated with the Brownian motion of the stochastic partial differential equation. Now we can apply Equ.(8.22), Equ.(8.23) and Equ.(8.24) to approximate the temporal and spatial derivatives in Equ.(8.33). The solution of the stochastic differential equation at any point (x,t) can be approximated by

$$\begin{aligned}
&\frac{u(x,t+\triangle t)-u(x,t)}{\triangle t} = \\
&\quad\left(\tilde{a}(x,t)\frac{u(x+\triangle x,t)-2u(x,t)+u(x-\triangle x,t)}{\triangle x^2}\right)u(t,x)+ \\
&\quad\left(\tilde{b}(x,t)\frac{u(x+\triangle x,t)-u(x-\triangle x,t)}{2\triangle x}\right)u(t,x)+ \\
&\quad\tilde{c}(x,t)u(t,x)+\tilde{d}(x,t)
\end{aligned} \tag{8.34}$$

Equ.(8.34) may be written as follows

$$\begin{aligned}
&\frac{u_{(i,j+1)}-u_{(i,j)}}{\triangle t} = \\
&\quad\tilde{a}_{(i,j)}\left(\frac{u_{(i+1,j)}-2u_{(i,j)}+u_{(i-1,j)}}{\triangle x^2}\right)+ \\
&\quad\tilde{b}_{(i,j)}\left(\frac{u_{(i+1,j)}-u_{(i-1,j)}}{2\triangle x}\right)+\tilde{c}_{(i,j)}u_{(i,j)}+\tilde{d}_{(i,j)}
\end{aligned} \tag{8.35}$$

Solving for $u_{(i,j+1)}$ in Equ.(8.35) yields:

$$u_{(i,j+1)} = \tilde{c}_1 u_{(i+1,j)} + \tilde{c}_0 u_{(i,j)} + \tilde{c}_{-1} u_{(i-1,j)} + \tilde{d}_{(i,j)}\triangle t \tag{8.36}$$

where:

$$\begin{aligned}
r &= \tfrac{\triangle t}{\triangle x^2} \\
\tilde{c}_1 &= \tilde{a}_{(i,j)}r + \tfrac{\tilde{b}_{(i,j)}r}{2}\triangle x \\
\tilde{c}_0 &= 1-2\tilde{a}_{(i,j)}r+\tilde{c}_{(i,j)}r\triangle x^2 \\
\tilde{c}_{-1} &= \tilde{a}_{(i,j)}r - \tfrac{\tilde{b}_{(i,j)}r}{2}\triangle x
\end{aligned} \tag{8.37}$$

The system of equations described in Equ.(8.32) and Equ.(8.37) contain a dimensionless quantity r, which we call the ***stability parameter***. When applying these techniques, the following stability criterion must be respected to ensure numerical stability, page 487 Haberman [23],

$$r = \frac{\triangle t}{\triangle x^2} \leq \frac{1}{4} \tag{8.38}$$

In cases where one wishes to optimize on simulation time, the Crank-Nicholson scheme, page 496 Haberman [23], is recommended; it ensures numerical stability $\forall r$.

The Extended Kalman Filter

In chapter 4, we derived the Kalman-Bucy filter as a special case of the nonlinear filtering problem. However, there are many problems in science and engineering where nonlinearities occur in the dynamic and measurement models. In this section we focus on a special class of problems that have linear dynamic and nonlinear observation models; the nonlinear tracking problem that we will encounter in next chapter falls into this class.

$$
\begin{aligned}
dynamics: \quad x_t &= x_0 + \int_0^t F x_s ds + \int_0^t G u(s) ds + \int_0^t \sqrt{Q} dB_s^1 \qquad (8.39) \\
measurements: \quad y_t &= y_0 + \int_0^t h(x_s, s) x_s ds + \int_0^t \sqrt{R} dB_s^2 \qquad (8.40)
\end{aligned}
$$

In filtering theory it is sometimes more convenient to rewrite the above system of equations as differential equations containing white noise terms

$$
\begin{aligned}
dynamics: \quad \dot{x}_t &= F x_t + G u(t) + w_t \qquad (8.41) \\
measurements: \quad z_t &= h(x_t, t) + v_t \qquad (8.42)
\end{aligned}
$$

where:

1. $x_t \in \mathbb{R}^n$ represents the ***state*** of the stochastic dynamical system
2. $z_t = \frac{dy_t}{dt} \in \mathbb{R}^l$ represents the ***measurement*** vector
3. $u(t) \in \mathbb{R}^r$ represents a ***deterministic input*** vector
4. $F \in \mathbb{R}^{n \times n}$ represents the time-timevariant***dynamic coupling*** matrix
5. $G \in \mathbb{R}^{n \times r}$ represents the time-invariant ***input coupling*** matrix

6. $h(x_t, t) \in \mathbb{R}^l$ represents the nonlinear ***measurement drift*** function

7. $w_t = \sqrt{Q}\frac{dB_s^1}{dt} \in \mathbb{R}^r$ represents the ***process noise*** process which is $w_t \sim \mathcal{N}(0, Q)$

8. $v_t = \sqrt{R}\frac{dB_s^2}{dt} \in \mathbb{R}^l$ represents the ***measurement noise*** process which is $v_t \sim \mathcal{N}(0, R)$

We assume that w_t and v_t are independent white noise processes. The function $h(x_t, t)$ is assumed to be nonlinear. The extended Kalman filter for this particular problem is similar to the Kalman-Bucy filter. However, the measurement residual in the state update equation is calculated using the observation and the nonlinear function $h(x_t, t)$ using current state estimate

$$\dot{\hat{x}}_t = F\hat{x}_t + PH^T R^{-1}[z_t - h(\hat{x}_t, t)] \tag{8.43}$$

The linear version of $h(x_t, t)$

$$H(t) \approx \frac{\partial h(x_t, t)}{\partial x_t}|_{x_t = \hat{x}_t} \tag{8.44}$$

is used to propagate the covariance and the state estimates. The equations for the Extended Kalman filter in continuous-time can be easily discretized for implementation on a digital computer, Grewal et al [21].
The state transition matrix $\Phi(t)$ and its discrete counterpart is defined by the following set of equations

$$\Phi(t) = \exp^{Ft} \approx I + Ft \tag{8.45}$$

$$\Phi_k = \Phi(\triangle t) \tag{8.46}$$

The pertinent equations are summarized below, compare with Grewal et al [21] and Zarchan et al [70]:

1. **Nonlinear dynamic model**:

$$x_k = \Phi_{k-1} x_{k-1} + w_{k-1}, \quad w_k \sim \mathcal{N}(0, Q_k) \tag{8.47}$$

where

$$Q_k = \int_0^{\triangle t} \Phi(\tau) Q \Phi^T(\tau) d\tau \tag{8.48}$$

2. **Nonlinear measurement model**:

$$z_k = h_k(x_k) + v_k, \quad v_k \sim \mathcal{N}(0, R_k) \tag{8.49}$$

3. **Nonlinear implementation equations**:

   (a) *predicted state*:

$$\hat{x}_k^- = \Phi_{k-1} \hat{x}_{k-1}^+ + G_{k-1} u_{k-1} \tag{8.50}$$

   where

$$G_k = \int_0^{\triangle t} \Phi(\tau) G d\tau \tag{8.51}$$

   (b) *computing predicted measurement*:

$$\hat{z}_k = h_k(\hat{x}_k^-) \tag{8.52}$$

4. **Linear approximation equations**:

$$H_k \approx \frac{\partial h_k}{\partial x} \mid_{x=\hat{x}_{k-1}^-} \tag{8.53}$$

(a) *conditioning the predicted estimate on the measurement*:

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(z_k - \hat{z}_k) \tag{8.54}$$

where

$$\hat{z}_k = h_k(\hat{x}_k^-) \tag{8.55}$$

(b) *computing the priori covariance matrix*:

$$P_k^- = \Phi_{k-1} P_{k-1}^+ \Phi_{k-1}^T + Q_{k-1} \tag{8.56}$$

(c) *computing the Kalman gain*:

$$K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1} \tag{8.57}$$

(d) *computing the posteriori covariance matrix*:

$$P_k^+ = [I - K_k H_k] P_k^- \tag{8.58}$$

The overall architecture of a typical software implementation of a Kalman filter is shown in Fig. 8.1, which clearly depicts the relevant process and measurement models and the estimator.

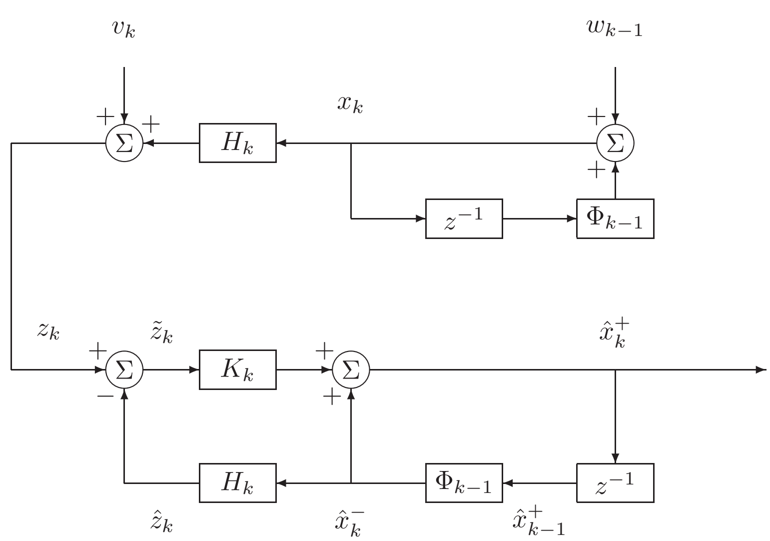

Figure 8.1: Discrete Kalman Filter

Chapter 9

NONLINEAR TRACKING PROBLEM

In this chapter, we focus on the application of the techniques described in this thesis to solving problems in science and engineering. We have chosen the classical passive radar tracking problem, which is a common benchmark problem in the literature: Rao et al [60], Kouritzin [37] and Challa et al [9]. A complete geometric analysis of the problem has been carried out by Lara [44], however, there are:

1. no numerical studies have been conducted with regard to solving the Zakai equation and

2. furthermore, there are no performance related studies with standard nonlinear filter schemes like the extended Kalman filter

9.1 Problem Statement

The tracking problem comprises of a radar and an aircraft. We assume that the aircraft is travelling toward the radar at a velocity $V(t)$. The objective of the tracking problem is to determine the coordinates of the aircraft based on the measurements taken by the radar, which tend to be noisy due to sensor imperfections. To simplify the problem, we assume that the aircraft is confined to a two dimensional plane. Furthermore, we assume that the velocity of the aircraft is constant, that is $V(t) = V$. If we assume perfect measurements, then the coordinates of the aircraft may be determined as the solution of the following system of differential equations

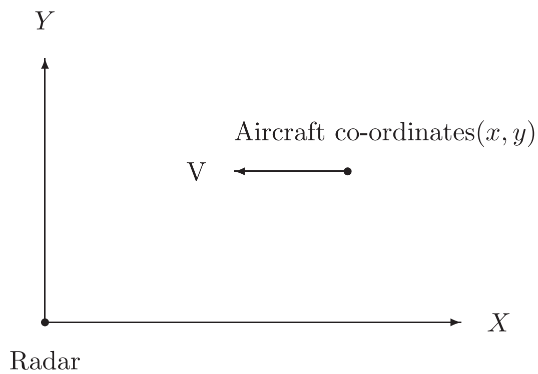

Figure 9.1: The Nonlinear Radar Tracking Problem

$$\dot{x} = -V \tag{9.1}$$

$$\dot{y} = 0 \tag{9.2}$$

To apply filtering theory to this problem, we need to consider the state of the aircraft $x_a = (x, y)$ as the solution to following system of differential equations

$$dx_t = -V dt + \sqrt{\epsilon} dw_t \tag{9.3}$$

$$dy_t = \sqrt{\epsilon} dv_t \tag{9.4}$$

where ϵ is associated with the intensity of the noise process of the state. The observation process is modelled as

$$dz_t = h(x_a)dt + dw_t \tag{9.5}$$

We assume that u_t, v_t and w_t are independent Brownian motions.

9.2 The Zakai Equation

The Stratonovich and Ito versions of the Zakai equation for the conditional density p_t of the state given the observations $\{z(s) \mid s \leq t\}$ are given below

$$Stratonovich: \quad dp_t = \left(\frac{\epsilon}{2}\frac{\partial^2}{\partial x^2} + \frac{\epsilon}{2}\frac{\partial^2}{\partial y^2} + V\frac{\partial}{\partial x} - \frac{1}{2}h^2\right)p_t dt + hp_t \circ dz_t \quad (9.6)$$

$$Ito: \quad dp_t = \left(\frac{\epsilon}{2}\frac{\partial^2}{\partial x^2} + \frac{\epsilon}{2}\frac{\partial^2}{\partial y^2} + V\frac{\partial}{\partial x}\right)p_t dt + hp_t dz_t \quad (9.7)$$

This is a two-dimensional stochastic partial differential equation. It turns out the Zakai equation is infinite dimensional.

9.3 Estimation Algebra

The nonlinear filtering problem for this application is clearly infinite dimensional since the measurement coupling function in Equ.(9.5) is clearly not linear with respect to the state, refer to Pro. 7.2.1.

9.4 Computation of Infinitesimal Symmetries

In this section, we will explore the existence of infinitesimal symmetries associated with the Zakai equation Equ.(9.6). If we can demonstrate the existence of such symmetries, then the Zakai equation can be simplified. We follow Lara [44], page 126, closely.

From Equ.(9.6) we can identify the operators M_0 Equ.(7.2) and M_1 Equ.(7.3) as follows:

$$M_0 = \mathcal{L} + H \quad (9.8)$$

$$M_1 = h \quad (9.9)$$

where $\mathcal{L}$ Equ.(7.4) and H Equ.(7.5) are given by:

$$\mathcal{L} = \frac{\epsilon}{2}\frac{\partial^2}{\partial x^2} + \frac{\epsilon}{2}\frac{\partial^2}{\partial y^2} + V\frac{\partial}{\partial x} \quad (9.10)$$

$$H = \frac{1}{2}h^2 \quad (9.11)$$

Let $g_{\mathbb{R}^2}$, $\nabla_{\mathbb{R}^2}$ and $\Delta_{\mathbb{R}^2}$ represent the metric, gradient and Laplacian, respectively, on $\mathbb{R}^2$. Expressing the operator M_0 in terms of these metrics yields the following equation:

$$M_0 = \frac{\epsilon}{2}\Delta_{\mathbb{R}^2} - V\frac{\partial}{\partial x} - \frac{1}{2}h^2 \tag{9.12}$$

The metric is therefore just

$$g = \frac{1}{\epsilon}g_{\mathbb{R}^2} \tag{9.13}$$

From Equ.(7.28) we identify the following components of Equ.(9.12)

$$U = -V\frac{\partial}{\partial x} \tag{9.14}$$

$$c = -\frac{1}{2}h^2 \tag{9.15}$$

Since

$$U = \nabla_{\mathbb{R}^2}(-Vx) = \frac{1}{\epsilon}\nabla_g(-Vx) \tag{9.16}$$

we can show by The. 7.3.3 that

$$K_{M_0} = 0 \tag{9.17}$$

Also, with reference to The. 7.3.3, since

$$div_g U = 0 \tag{9.18}$$

and

$$g(U,U) = \frac{1}{\epsilon}V^2 \tag{9.19}$$

we can show that

$$H_{M_0} = div_g U + \frac{1}{2}g(U,U) - 2c \tag{9.20}$$

is a function of r only if h is a function of r. Lara [43], page 1595, proves that

$$\zeta = -\frac{\partial}{\partial\theta} - \frac{Vr}{\epsilon}\sin\theta u\frac{\partial}{\partial u} \tag{9.21}$$

is an infinitesimal symmetry. Then, by The. 7.3.4 we consider

$$\hat{\zeta}_0^* = X + m = -\frac{\partial}{\partial\theta} - \frac{Vr}{\epsilon}\sin\theta \tag{9.22}$$

where:

1. $X = -\frac{\partial}{\partial\theta}$ is an isometry of the Riemannian manifold $\mathcal{M} = \mathbb{R}^2$ with metric g since it is an isometry of $\mathcal{M} = \mathbb{R}^2$ with metric $g_{\mathbb{R}^2}$

2. $m = -\frac{Vr}{\epsilon}\sin\theta$ is such that $X_1 = 0$ in induction by The. 7.3.3:

$$\begin{aligned} X_0 &= X && (9.23)\\ X_1 &= \nabla_g(2m - g(X,U)) && (9.24)\\ X_{i+2} &= \frac{1}{2}\nabla_g(L_{X_i}H_{M_0} + \eta_g(X_i)H_{M_0}) && (9.25) \end{aligned}$$

Since

$$L_{X_0}H_{M_0} = -\frac{\partial H_{M_0}}{\partial\theta} = 0 \tag{9.26}$$

we have $X_2 = 0$, then by induction $X_i = 0\ \forall i \geq 3$, therefore, $\hat{\zeta}_0^* \in \mathscr{P}_{M_0}$. Note that $X_0 = -\frac{\partial}{\partial\theta}$ and $X_i = 0\ \forall i \geq 1$. Hence, Equ.(7.29) is satisfied because h depends on r

$$L_{X_0}h = \frac{\partial h(r)}{\partial\theta} = 0 \tag{9.27}$$

9.5 Reduction of the Zakai Equation

For the purpose of simplifying the Zakai equation let us define the drift term in the observation as follows:

$$h(x_a) = -V \frac{\partial (H \circ \theta)}{\partial x}(x_a) \tag{9.28}$$

where

$$\theta(x_a) = \tan^{-1}\left(\frac{y_t}{x_t}\right) \tag{9.29}$$

Lara [44] claims that there is no smooth function H for which h is linear. Thus Kalman filtering techniques do not apply. However, the Zakai equation may be simplified by shaping the observations with

$$H(\theta) = \ln \tan(\theta/2) \tag{9.30}$$

Substituting Equ.(9.30) into Equ.(9.28) we obtain a simplified expression for the drift term

$$\begin{aligned} h(x_a) &= -V \frac{\partial}{\partial x}\left[\ln(\tan(\theta/2))\right] \\ &= V \frac{\sin(\theta)}{r} \frac{\partial}{\partial \theta}\left[\ln(\tan(\theta/2))\right] \\ &= V \frac{\sin(\theta)}{r} \frac{1}{\tan(\theta/2)} \frac{\partial}{\partial \theta}\left[\tan(\theta/2)\right] \\ &= V \frac{\sin(\theta)}{2r} \frac{\cos(\theta/2)}{\sin(\theta/2)} \frac{1}{\cos^2(\theta/2)} \\ &= \frac{V}{r} \end{aligned} \tag{9.31}$$

where $r = \sqrt{x^2+y^2}$. When the drift function can be expressed in this form, Equ.(9.31), it is possible to reduce the Zakai equation to a one dimensional stochastic partial differential equation. However, the resulting Zakai equation remains infinite dimensional.

Proposition 9.5.1 (page 122 Lara [44]). *When h depends only on r, there exists a particular solution of the Zakai equation of the form*

$$p_t(x,y) = \exp^{\left(\frac{Vx}{\epsilon}\right)} q_t\left(\sqrt{x^2+y^2}\right) = \exp^{\left(\frac{Vr}{\epsilon}\cos\theta\right)} q_t(r) \tag{9.32}$$

where the function q_t satisfies the following stochastic partial differential equation

$$Stratonovich: \quad dq_t = \left(\frac{\epsilon}{2}\frac{\partial^2 q_t}{\partial r^2} + \frac{\epsilon}{2r}\frac{\partial^2 q_t}{\partial r} + \frac{V^2}{2\epsilon}q_t - \frac{h^2}{2}q_t\right)dt + hq_t \circ dz_t \quad (9.33)$$

$$Ito: \quad dq_t = \left(\frac{\epsilon}{2}\frac{\partial^2 q_t}{\partial r^2} + \frac{\epsilon}{2r}\frac{\partial^2 q_t}{\partial r} + \frac{V^2}{2\epsilon}q_t\right)dt + hq_t dz_t \quad (9.34)$$

The invariant solution only exists when the initial condition of the state x_a has the same invariance property

$$p_0(r,\theta) = \exp^{\left(\frac{Vr}{\epsilon}\cos\theta\right)} q_0(r) \quad (9.35)$$

This fact is explained by showing the existence of infinitesimal symmetries. It is important to note that while Pro. 9.5.1 asserts that there exists a simplified solution it is by no means suitable for numerical implementation because the exponential function is dependent on geometric range, which can be very large. However, instead of pursuing this topic further, it perhaps more important to illustrate the distinct advantages the Zakai equation has over conventional nonlinear filtering schemes. We do this by simplifying the tracking problem further.

9.6 Approximation to the Nonlinear Filtering Problem

Without loss of generality, we assume that the altitude or y component of the aircrafts position is known at initialization y_0 and remains constant for all time, hence, $y = y_0$. This assumption reduces the Zakai equation from a 2-dimensional to a 1-dimensional stochastic partial differential equation

$$dp_t = \left(\frac{\epsilon}{2}\frac{\partial^2}{\partial x^2} + V\frac{\partial}{\partial x}\right)p_t dt + hp_t dz_t \quad (9.36)$$

where

$$h = \frac{V}{\sqrt{x^2 + y_0^2}} \quad (9.37)$$

We emphasize that the simplification that arises from our assumption is purely for numerical simplicity.

9.6.1 Discretization of the Zakai Equation

The coefficients associated with Equ.(8.32) are

$$\begin{aligned} \bar{a}(x) &= \tfrac{\epsilon}{2} \\ \bar{b}(x) &= V \\ \bar{c}(x) &= 0 \end{aligned} \tag{9.38}$$

Source Code Listing

Refer to Appendix A and B.

9.6.2 Finite Difference Approximation of the Zakai Equation

The following identification can be made with Equ.(8.37)

$$\begin{aligned} \tilde{a}(x,t) &= \tfrac{\epsilon}{2} \\ \tilde{b}(x,t) &= V \\ \tilde{c}(x,t) &= h\tfrac{dz_t}{dt} \\ \tilde{d}(x,t) &= 0 \end{aligned} \tag{9.39}$$

Source Code Listing

Refer to Appendix A and C.

9.6.3 Solution by Extended Kalman Filtering

Consider the nonlinear tracking problem written out in differential equation form

$$\dot{x}_t = -V + w_t \tag{9.40}$$

$$z_t = h(x_a) + v_t \tag{9.41}$$

In this section we proceed with identifying the parameters for the discrete extended Kalman filter.

Description	**Continuous-Time**	**Discrete-Time**
state	x_t	x_k
measurement	$z_t = \frac{dy_t}{dt}$	z_k
input	$u(t) = -V$	u_k
dynamic coupling	$F = 0$	
input coupling	$G = 1$	$G_k = \triangle t$
measurement drift	$h(x_t, t) = \tan^{-1}(\frac{y_0}{x_t})$	h_k
linearized drift	$H(x_t, t) = -(\frac{y_0}{r^2})$	H_k
state transition matrix	$\Phi(t) = 0$	$\Phi_k = 1$
process noise	$w_t \sim \mathcal{N}(0, Q)$	$w_k \sim \mathcal{N}(0, Q_k)$
process noise covariance	$Q = \epsilon$	$Q_k = Q\triangle t$
measurement noise	$v_t \sim \mathcal{N}(0, R)$	$v_k \sim \mathcal{N}(0, R_k)$
measurement noise covariance	R	$R_k = \sigma_\theta^2$

Table 9.1: Extended Kalman Filter for the Nonlinear Tracking Problem

Source Code Listing

Refer to Appendix D.

9.7 Simulation Results

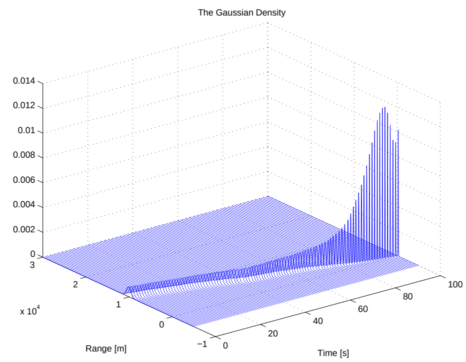

Figure 9.2: Gaussian Density

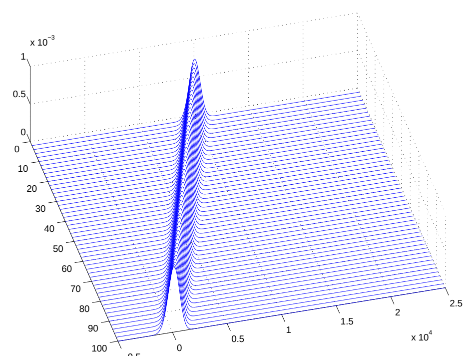

Figure 9.3: UCD Using Semigroup Techniques

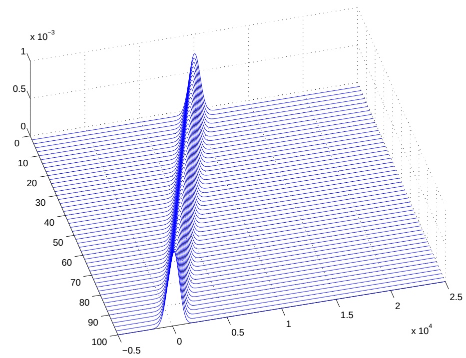

Figure 9.4: UCD Using Finite Differences

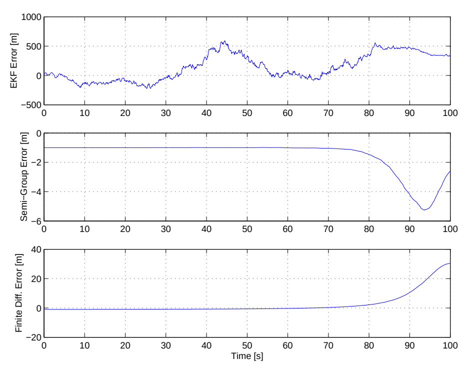

Figure 9.5: Error in Estimation

9.7.1 Discussion

In the case of the extended Kalman filter, the simulation was setup with a large covariance matrix P_0^- to reflect the uncertainty associated with the initial state x_0^-. By determining these two parameters, we fix the initial gaussian density of the filter. Ordinarily, one would have to determine the equivalent initial unnormalized conditional density from the initial gaussian density to solve the Zakai equation. However, as Elliot et al [17] have proven, page 944 The. 5.1, the two densities differ from each other only by a constant. Since we are essentially interested in determining the state estimate from Equ.(4.29), we may use the same density as an initial condition for determining the solution of the Zakai equation.

From Fig. 9.2 we observe that the gaussian density evolves in a very erratic manner compared with the densities determined from the semigroup and finite difference techniques, Fig. 9.3 and Fig. 9.4 respectively. Also note that for the unnormalized conditional density, the two independent numerical techniques provide near identical solutions, Fig. 9.3 and Fig. 9.4 respectively.

On close examination we observe that the state estimate as determined from the solution of the Zakai equation is by far more accurate, subplot 2 & 3 of Fig. 9.5, than the solution of the extended Kalman filter, subplot 1. This phenomena may be explained as follows. In the case of the extended Kalman filter the Kalman gain tends to be very large when the uncertainty associated with the state is large. This has the effect of amplifying the measurement residual in the state update equation x_k^+, Equ.(8.54). Consequently, this manifests as a very noisy estimate of the state. Note that the semigroup method, subplot 2 of Fig. 9.5, tends to be more accurate than the finite difference method. The observed deviation between the two methods is probably due to the numerical problems that arise as the range tends to zero.

Our simulation studies demonstrate the accuracy of the Zakai equation over the extended Kalman filter. There are a many areas in science and engineering that can benefit from these new techniques. However, one problem of concern is to develop algorithms that can run in real-time on embedded microprocessors. This continues to be a field of active research and some interesting developments have been made in this regard, Guo [22].

Part V

Conclusion

Chapter 10

CONCLUSION

In conclusion, it is perhaps worthwhile to reflect on the contribution of this thesis. As any specialist of stochastic analysis, differential geometry and nonlinear filtering theory would acknowledge, there have been a number of independent developments in these fields in recent years. Much of our research reflects these developments, in fact, there is no new theoretical research presented in this thesis. However, the unification of these ideas as well as their application is regarded as our greatest contribution. We have shown how state of the art geometric tools maybe used to analyze the Zakai equation and how they may be applied to study the existence of symmetries, which provides greater insight to the latent physical and geometric properties of a system. Furthermore, the existence of these symmetries provide us with a deeper understanding of group invariant solutions, which maybe used to simplify the Zakai equation; this will undoubtedly have wide spread application once numerical techniques for solving stochastic partial differential equations have reached greater maturity. Also, using simulation we have demonstrated the optimality of the Zakai equation over the extended Kalman filter for the passive radar tracking problem - to the best of our knowledge, such benchmark studies have not been done before for the tracking problem using the numerical techniques discussed in this thesis. This supports the call for continual research and justifies the application of new tools founded on differential geometry. Lastly, we would like to add that this research has presented us with an opportunity to appreciate the applicability of abstract geometric theories to physical systems. We sincerely hope that this presentation has conveyed our insights and enthusiasm for the subject.

Part VI

Appendix

APPENDIX A: INITIALIZATION FILE

```
% Configuration Control Preamble--------------------------------------------%
% File Name         : ZakaiNumIni.m                                         %
% Purpose           : m-file for the initialization of numerical algorithms %
%                     for the solution of the Zakai equation by finite      %
%                     difference and semi group methods for the nonlinear   %
%                     target tracking problem.                              %
% Program Language  : Matlab m-file                                         %
% Version           : 1.0                                                   %
% Programmers       : R. Rugunanan                                          %
% Date Written      : 04 February 2005                                      %
% Updates           : Reasons                             Date         By   %
%---------------------------------------------------------------------------%


close all;
clear all;

load IniData;
DenXVec = DenData(:,1);
DenVec  = DenData(:,2);

% Use a fixed random state
randn('state',777);

% PDE
k       = 1;                 % PDE constant
s_test  = 0.25;
epsilon = 400;
V       = 100;
X0      = 10100;             % Initial x-comp to target position  [m]
y0      = 500;
```

```
% Spatial range
Xf = 25000;
Xi = -5000;
L  = Xf-Xi;
dx = 50;           % Increment in x
N  = L/dx;         % Total number of intervals in x
X  = [Xi:dx:Xf];   % X interval
P  = max(size(X)); % P = N+1

% Time range
dt = 1/100;        % Increment in t_l
s  = k*dt/dx^2;    % Stability parameter

Ti = 0;
Tf = 100;
M  = (Tf-Ti)/dt;   % Total number of intervals in t_l
T  = [Ti:dt:Tf];   % t_l interval
Q  = max(size(T)); % Q = M+1

A = zeros([P,Q]);
U = zeros([P,Q]);

% set initial condition and bundary cinditions
% u_j_m ... j = 1,...,k-1 and m = 0

for Ik=1:P       % scan space
     A(Ik,1) =  interp1(DenXVec, DenVec, X(Ik),'linear');
end

U=A;



%Dynamic model
I       = 1;                 % Identity matrix
sigpro  = sqrt(epsilon/dt);
sigob   = 0.025;       % Standard deviation on observation
```

```
% Discrete model
X_ln = X0;
w_ln = 0;



% Sample control input
U_ln = -V;

% Setup process noise matrix
Q_l  = epsilon*dt;


% State transition matrix
Phi_ln =  I;

% Set up control matrix
G_ln   = dt;
```

APPENDIX B: GAUGE TRANSFORMATION AND SEMIGROUP APPROACH APPLIED TO THE ZAKAI EQUATION

```
% Configuration Control Preamble---------------------------------------------%
% File Name         : ZakaiNumSG.m                                           %
% Purpose           : m-file to implement the numerical solution of the      %
%                     Zakai equation based on semi group methods             %
%                     for the nonlinear target tracking problem.             %
% Program Language  : Matlab m-file                                          %
% Version           : 1.0                                                    %
% Programmers       : R. Rugunanan                                           %
% Date Written      : 04 February 2005                                       %
% Updates           : Reasons                            Date          By    %
%----------------------------------------------------------------------------%



% Call initialization file
ZakaiNumInit;

% Solve PDE
% u_j_m ... j = 1,...,k-1 and m = 1,...,M

for Il=1:Q-1          % scan time



        % Generate noises
        w_l    = sqrt(Q_l)*randn; %sigpro*randn;
        v_ln   = sigob*randn;

        % Discrete dynamics systems model
        X_l       = Phi_ln*X_ln + G_ln*U_ln + w_ln;
```

```
for Ik=1:P-2     % scan space

    U_kn1_l    =  A(Ik,   Il);
    U_k_l      =  A(Ik+1, Il);
    U_kp1_l    =  A(Ik+2, Il);

    t_l   = Il*dt;
    x_kn1 = X(Ik);
    x_k   = X(Ik+1);
    x_kp1 = X(Ik+2);

    % Generate measurement
    range_l = sqrt(X_l^2+y0^2);
    u_l     = V/range_l + v_ln;

    h_kn1 = V/sqrt(x_kn1^2+y0^2);
    h_k   = V/sqrt(x_k^2+y0^2);
    h_kp1 = V/sqrt(x_kp1^2+y0^2);


    a = epsilon/2;
    b = V;
    c = 0;
    d = 0;

    cp1 = a*s + b*s*dx/2;
    c0  = 1 - 2*a*s;
    cn1 = a*s - b*s*dx/2;

    P_kn1_l = exp(h_kn1*u_l*dt - 0.5*h_kn1^2*dt)*U_kn1_l;
    P_k_l   = exp(h_k  *u_l*dt - 0.5*h_k^2  *dt)*U_k_l;
    P_kp1_l = exp(h_kp1*u_l*dt - 0.5*h_kp1^2*dt)*U_kp1_l;
```

```
         U_k_lp1 = cp1*P_kp1_l + c0*P_k_l + cn1*P_kn1_l;


         A(Ik+1,Il+1)   =  U_k_lp1;


     end
     X_ln=X_l;
     w_ln=w_l;


        intn = approx_int(dx,X.*A(:,Il)');
        intd = approx_int(dx,A(:,Il)');
        estimate = intn(end)/intd(end);
        DataZSG(Il,:) = [t_l x_k estimate X_l u_l];

     if((mod(t_l,2)<dt/2)==1)
        t_l
        figure(1);hold on;grid on;plot3(t_l*ones(size(X)),X,A(:,Il));view(75,70);
        figure(2);hold on;grid on;plot3(t_l*ones(size(X)),X,A(:,Il));view(75,70);
        %figure(2);hold on;grid on;plot(t_l,estimate,'r.');
     end

 end



save ZakaiNumSGSim DataZSG
hgsave(1,'ZakaiDensity');



figure(2);
FileName = ['C:\TEX\MSc\Thesis\ZakaiDenSG'];
string = ['print  -depsc ' FileName];
eval(string);
```

APPENDIX C: FINITE DIFFERENCE APPROACH APPLIED TO THE ZAKAI EQUATION

```
% Configuration Control Preamble--------------------------------------------%
% File Name         : ZakaiNumFD.m                                          %
% Purpose           : m-file to implement the numerical solution of the     %
%                     Zakai equation based on finite difference methods     %
%                     for the nonlinear target tracking problem.            %
% Program Language  : Matlab m-file                                         %
% Version           : 1.0                                                   %
% Programmers       : R. Rugunanan                                          %
% Date Written      : 04 February 2005                                      %
% Updates           : Reasons                            Date         By    %
%---------------------------------------------------------------------------%




% Call initialization file
ZakaiNumInit;



% Solve PDE
% u_j_m ... j = 1,...,k-1 and m = 1,...,M

for Il=2:Q             % scan time



        % Generate noises
        w_l    = sqrt(Q_l)*randn; %sigpro*randn;
        v_ln   = sigob*randn;

        % Discrete dynamics systems model
        X_l       = Phi_ln*X_ln + G_ln*U_ln + w_ln;
```

```
for Ik=1:P-2     % scan space

    U_kn1_ln1  =  A(Ik,  Il-1);
    U_k_ln1    =  A(Ik+1,Il-1);
    U_kp1_ln1  =  A(Ik+2,Il-1);

    t_l = Il*dt;
    x_k = X(Ik);



    % Generate measurement
    range_l = sqrt(X_l^2+y0^2);
    u_l     = V/range_l + v_ln;



    h = V/sqrt(x_k^2+y0^2);

    a = epsilon/2;
    b = V;
    c = h*u_l;
    d = 0;

    cp1 = a*s + b*s*dx/2;
    c0  = 1 - 2*a*s + c*s*dx^2;
    cn1 = a*s - b*s*dx/2;

    U_k_lp1  = cp1*U_kp1_ln1 + c0*U_k_ln1 + cn1*U_kn1_ln1 + d*dt;

    A(Ik+1,Il)   =  U_k_lp1;

end
X_ln=X_l;
w_ln=w_l;
```

```
        intn = approx_int(dx,X.*A(:,Il)');
        intd = approx_int(dx,A(:,Il)');
        estimate = intn(end)/intd(end);
        DataZFD(Il-1,:) = [t_l x_k estimate X_l u_l];

     if((mod(t_l,2)<dt/2)==1)
        t_l
        figure(1);hold on;grid on;plot3(t_l*ones(size(X)),X,A(:,Il));view(75,70);
        figure(2);hold on;grid on;plot3(t_l*ones(size(X)),X,A(:,Il));view(75,70);
        %figure(2);hold on;grid on;plot(t_l,estimate,'r.');
     end
 end


save ZakaiNumFDSim DataZFD
hgsave(1,'ZakaiDensity');


figure(2);
FileName = ['C:\TEX\MSc\Thesis\ZakaiDenFD'];
string = ['print  -depsc ' FileName];
eval(string);
```

APPENDIX D: SOLUTION BY EXTENDED KALMAN FILTER

```
% Configuration Control Preamble--------------------------------------------%
% File Name         : EKFTarSim1DNL.m                                       %
% Purpose           : m-file to implement an EKF for the target tracking    %
%                     problem.                                              %
% Program Language  : Matlab m-file                                         %
% Version           : 1.0                                                   %
% Programmers       : R. Rugunanan                                          %
% Date Written      : 04 February 2005                                      %
% Updates           : Reasons                             Date          By  %
%---------------------------------------------------------------------------%

% Close all figures and clear workspace
close all
clear all

% Use a fixed random state
randn('state',777);

ShapeObservations = 1;       % 1 -> Use observation shaping for nonlinear drift
                             % 0 -> Use bearing model for nonlinear drift
% Constants
r2d=180/pi;                  % Convert radians to degrees
d2r=1/r2d;                   % Convert degrees to radians

% Time interval
% Continuous time
Ti = 0;                      % Initial time                          [s]
Tf = 100;                    % Final time                            [s]
dt = 0.01;                   % Time delt for continuous states       [s]
T  = [Ti:dt:Tf];             % Time interval                         [s]
```

```
M  = (Tf-Ti)/dt;            % Number of samples                      [s]
N  = M+1;                   %

% Sample time
Ts  = 0.1;                  % Time step of EKF                       [s]

%Target model
V   = 100;                  % Target velocity magnitude              [m/s]

%True model
%Initial postion
X0  = 10100;                % Initial x-comp to target position      [m]
y0  = 500;                  % Initial y-comp to target position      [m]



%Estimates
X_err = 500;               % Initial position error for covariance   [m]

%Initial postion
hX0   = X0;                 % Assume initial state estimate is known exactly



%Dynamic model
I       = 1;                % Identity matrix
epsilon = 400;              % Process noise coupling matrix          []
sigpro  = sqrt(epsilon/dt); % Spectral intensity on process noise    []
sigob   = 0.025;            % Standard deviation on observation      [rad]

% Counters for simulation
nk       = 0
SamCount = 0;
j        = 1;



for i=1:N
```

```
% Initialization
if(i==1)

   % True model
   X    =  X0;
   U    = -V;
   F    =  0;
   G    =  1;

   % Discrete model
   X_kn = X0;
   w_kn = 0;

   % Estimate
   hX_kn_p  = hX0;

   % Setup process noise matrix
   Q_k  = epsilon*Ts;
   Q_kn = Q_k;



   % Setup measurement noise
   R_k      = sigob^2;

   % Setup covariance matrix
   P_kn_p      = X_err^2;

   % State transition matrix
   Phi_kn =  I;

   % Set up control matrix
   G_kn   = Ts;



   % Sample control input
   U_kn = U;
```

```
   % Save initial density to boot Zakai Solution
   GaussDenData0=PlotGaussDen(0,hX_kn_p,P_kn_p);

end



 % Continuous-part of simulation
 % Evolution of Time
 t = T(i);

 % Determine outputs based on available state info
 % Ideal observations
 the   = atan2(y0,X);
 range = sqrt(X^2+y0^2);

 if(t<Ts)
  range0 = range;
  the0    = the;
 end

 Z1   = V/range;
 Z2   = the;



 if(ShapeObservations)
  Z        =  Z1;
 else
  Z        =  Z2;
 end






 % Discrete component of simulation
 if(SamCount==Ts/dt | i==1)
```

```
        % Generate noises
        w_k    = sqrt(Q_k)*randn; %sigpro*randn;
        v_kn   = sigob*randn;

        % Discrete dynamics systems model
        X_k       = Phi_kn*X_kn + G_kn*U_kn + w_kn;

        X_kn = X_k;
        w_kn = w_k;


        % Generate measurement
        the_k   = atan2(y0,X_k);
        range_k = sqrt(X_k^2+y0^2);

        Z1_k    = V/range_k + v_kn;
        Z2_k    = the_k + v_kn;



        %Kalaman Filter equations
%        P_k_n    = Phi_kn*P_kn_p*Phi_kn'+Q_kn;
%        K_k      = P_k_n*H_k'*inv(H_k*P_k_n*H_k'+R_k);
%        P_k_p    = (I-K_k*H_k)*P_k_n;


        % Projected state
         hX_k_n     = Phi_kn*hX_kn_p + G_kn*U_kn;

        % Measurement matrix
         hthe_k_n   = atan2(y0,hX_k_n);
         hrange_k_n = sqrt(hX_k_n^2+y0^2);
```

```
 H1_k       =  [-V*hX_k_n/(hrange_k_n^3)];
 H2_k       =  [-y0/hrange_k_n^2];

 hZ1_k      =  [V/hrange_k_n];
 hZ2_k      =  [hthe_k_n];



if(ShapeObservations)
 Z_k        =  Z1_k;
 H_k        =  H1_k;
 hZ_k       =  hZ1_k;
else
 Z_k        =  Z2_k;
 H_k        =  H2_k;
 hZ_k       =  hZ2_k;
end



% Residual
 Zr     = Z_k-hZ_k;


% Program for numerical stability
% 1] compute P_k_n using P_kn_p, Phi_kn and Q_kn
 H_k_T                                = H_k';
 Phi_kn_T                             = Phi_kn';
 Phi_kn_mult_P_kn_p                   = Phi_kn*P_kn_p;
 Phi_kn_mult_P_kn_p_mult_Phi_kn_T     = Phi_kn_mult_P_kn_p*Phi_kn_T;
 P_k_n                                = Phi_kn_mult_P_kn_p_mult_Phi_kn_T+Q_kn;

% 2] Compute K_k using P_k_n, H_k and R_k
 H_k_mult_P_k_n                       = H_k*P_k_n;
 H_k_mult_P_k_n_mult_H_k_T            = H_k_mult_P_k_n*H_k_T;
 H_k_mult_P_k_n_mult_H_k_T_plus_R_k   = H_k_mult_P_k_n_mult_H_k_T+R_k;
 inv_H_k_mult_P_k_n_mult_H_k_T_plus_R_k = inv(H_k_mult_P_k_n_mult_H_k_T_plus_R_k);
```

```
 P_k_n_mult_H_k_T                       = P_k_n*H_k_T;
 K_k                                    =
 P_k_n_mult_H_k_T*inv_H_k_mult_P_k_n_mult_H_k_T_plus_R_k;




% 3] Compute P_k_p using K_k, P_k_n
 K_k_mult_H_k                           = K_k*H_k;
 I_min_K_k_mult_H_k                     = I-K_k_mult_H_k;
 P_k_p                                  = I_min_K_k_mult_H_k*P_k_n;

% 4] Compute hX_k_p using K-k and Z_k data
 hX_k_p                                 = hX_k_n + K_k*Zr;


% Model delays for next cycle
 P_kn_p  = P_k_p;
 hX_kn_p = hX_k_p;




% Post processing data

% Determine Gaussian density
 if(mod(t,1)==0)
  GaussDenData   = PlotGaussDen(t,hX_k_p,P_kn_p);
 end

%Output for post processing
 X_err    = X - hX_k_p;
 SP(1)    = sqrt(P_k_p(1,1));


 TVec(j,:)     =  t;
 XVec(j,:)     =  X';
```

```
        XkVec(j,:)    =  X_k';
        XerrVec(j,:)  =  X_err';
        hXkpVec(j,:)  =  hX_k_p';

        ZVec(j,:)     =  Z';
        ZkVec(j,:)    =  Z_k';
        Z1Vec(j,:)    =  Z1';
        Z2Vec(j,:)    =  Z2';
        Z1kVec(j,:)   =  Z1_k';
        Z2kVec(j,:)   =  Z2_k';

        KkVec(j,:)    =  K_k';

        ZrVec(j,:)    =  Zr';



        SPVec(j,:)    =  SP;


    SamCount = 0;
    j=j+1;
    end




       % State space model
        Xd = F*X + G*U;

       % Euler integration
        X  = X + Xd*dt;




SamCount=SamCount+1;
end
```

```
% Saving data
DenData = [GaussDenData0(:,2),GaussDenData0(:,3)];
save IniData DenData range0 the0 TVec ZkVec XkVec
save EKFSim XVec TVec hXkpVec

% Plotting data
figure
plot(TVec,XerrVec(:,1),TVec,SPVec(:,1),TVec,-SPVec(:,1)),grid on; hold on; zoom on;
ylabel('Error in Estimate of x-component [m]')
xlabel('Time [s]')

figure
plot(TVec,KkVec(:,1)),grid on; hold on; zoom on;
ylabel('Kalman Gain []')
xlabel('Time [s]')

figure
plot(TVec,ZrVec(:,1)),grid on; hold on; zoom on;
ylabel('Measurement Residual []')
xlabel('Time [s]')



figure
plot(TVec,XVec, 'r', TVec, XkVec, 'b', TVec, hXkpVec, 'g');grid on; hold on; zoom on;
legend('True','Discrete','Estimated');
ylabel('Downrange [m]')
xlabel('Time [s]')

figure
plot(TVec,XVec-XkVec,'b',TVec,XerrVec(:,1),'r');grid on; hold on; zoom on;
ylabel('Noise on Position [m]')
xlabel('Time [s]')

figure
```

```
subplot(2,1,1);plot(TVec,Z1Vec(:,1),'b',TVec,Z2Vec(:,1),'r'),grid on; hold on; zoom on;
legend('shaping function','true bearing');
ylabel('Measurements [rad]')
subplot(2,1,2);plot(TVec,Z1kVec(:,1),'b',TVec,Z2kVec(:,1),'r'),grid on; hold on; zoom on;
legend('shaping function','true bearing');
ylabel('Measurements [rad]')
xlabel('Time [s]')




FigureList = sort(get(0,'children'));

for(i=1:max(size(FigureList)))
   CurrentFigure = FigureList(i);
   figure(CurrentFigure);
   FileName = ['C:\TEX\MSc\Thesis\EKF_' num2str(CurrentFigure)];
   string = ['print  -depsc ' FileName];
   eval(string);
end
```

APPENDIX E: SOFTWARE USED IN THIS THESIS

The purpose of this appendix is to provide an short overview of the software that used in preparing this thesis.

LATEX2E: The wordprocessor that was selected was LaTeX2e because of its simplicity and ease of use when dealing with relatively complicated mathematical formulas. We highly recommend the thesis template based on the University of Washington thesis class by Jim Fox; LaTeX2e version 1995 from the latex209 thesis.sty style (1990-1).

PSTRICKS: All commutative diagrams were completed using PsTricks.

MATLAB: All the software programs and simulation results were developed in MatLab (Version 6.1.0.450 (R12.1)), which is a trademark of Mathworks.

LaTeX2e and PsTricks are freeware products available for download on the internet.

BIBLIOGRAPHY

[1] Alinger, D., and Mitter, S. New Results on the Innovations Problem for Nonlinear Filtering. *Stochastics 4* (1981), 339–348.

[2] Ames, W. *Numerical Methods for Partial Differential Equations*. Academic Press, 1977.

[3] Arnold, V. *Ordinary Differential Equations*. MIT Press, 1978.

[4] Baras, J., and LaVigna, A. Real Time Sequential Detection for Diffusion Signals. Preprint: Number: TR 1986-27, 1986. http://techreports.isr.umd.edu/.

[5] Belopolskaya, Y. I., and Dalecky, Y. L. Ito's Equations and Differential Geometry. *Russian Math Surveys Ref. 37:3* (1982), 95–142.

[6] Bensoussan, A. *Stochastic Control of Partially Observable Systems*. Cambridge University Press, 1992.

[7] Brzeźniak, Z., and Zastawniak, T. *Basic Stochastic Processes*. Springer-Verlag, 1999.

[8] Bucy, R. Nonlinear Filtering. *IEEE Trans. Automatic Control AC-10* (1965), 198.

[9] Challer, C., and Bar-Shalom, Y. Nonlinear Filter Design Using Fokker-Plank-Kolmogorov Probability Density Evolutions. *IEEE Transactions on Aerospace and Electronic Systems 36*, 1 (January 2000), 309–314.

[10] Crisan, D. *The Problem of NonLinear Filtering.* PhD thesis, The University of Edinburgh, Department of Mathematics, 1996.

[11] Darling, R. Geometrically Intrinsic Nonlinear Recursive Filters I: Algorithms. Preprint, 1997. `http://www.stat.berkeley.edu/tech-reports/`.

[12] Darling, R. Geometrically Intrinsic Nonlinear Recursive Filters II: Foundations. Preprint, 1998. `http://www.stat.berkeley.edu/tech-reports/`.

[13] Darling, R. Intrinsic Location Paramter of a Diffusion Process. *Electronic Journal of Probability 7*, 3 (2002), 1–23.

[14] Darling, R. Nonlinear Filtering Online Survey. Preprint, 2004. `http://members.verizon.net/~vze543kh/rwrdarling/FILTERING/`.

[15] Davis, M., and Marcus, S. An Introduction to Nonlinear Filtering. In *Stochastic Systems: The Mathematics of Filtering and Identification and Applications* (1981), M. Hazewinkel and J. Willems, Eds., vol. 78 of *C*, Procedings of the NATO Advance Study Institute held at Les Arcs, Savoie, France, June 22 - July 5, 1980, D. Reidel, pp. 53–75.

[16] Duncan, T. *Probability Densities for Diffusion Processes with Application to Nonlinear Filtering Theory and Diffusion Theory.* PhD thesis, Stanford University, Stanford, CA, 1967.

[17] Elliot, R., and Krishnamurthy, V. New Finite-Dimensional Filters for Parameter Estimation of Discrete-Time Linear Gaussian Models. *IEEE Trans. On Automatic Control 44*, 5 (May 1999), 938–951.

[18] Frost, P., and Kailath, T. An Innovations Approach to Least Squares Estimation - Part III: Linear Filtering in Additive White Noise. *IEEE Trans. Automatic Control AC-16* (1971), 217–226.

[19] Fujisaki, M., Kallianpur, G., and Kunita, H. Stochastic Differential Equations for the Nonlinear Filtering Problem. *Osaka J. Math. 1* (1972), 19–40.

[20] Gckeler, M., and Schcker, T. *Differential Geometry, Gauge Theories, and Gravity*. Cambridge University Press, 1989.

[21] Grewal, M., and Andrews, A. *Kalman Filtering : Theory and Practice Using MATLAB*. Wiley-Interscience 2 edition, 2001.

[22] Guo, Z. Software Implementation of the Yau Filtering System. Master's thesis, National Cheng Kung University - Taiwan, Department of Electrical Engineering, 2004.

[23] Haberman, R. *Elementary Applied Partial Differential Equations*. Prentice Hall, 1983.

[24] Hsu, E. Martingale Theory (Lecture Notes - Chapter 1). Preprint, 2000. `http://www.math.northwestern.edu/~elton/`.

[25] Hsu, E. *Stochastic Analysis on Manifolds*. American Mathematical Society, 2002.

[26] Ikeda, N., and Watanebe, S. *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam, 1981.

[27] Isham, C. *Modern Differential Geometry for Physicists*. World Scientific Publishing Company, 1999.

[28] Ito, K. On Stochastic Differential Equations on a Differentiable Manifold. *Nagoya Math. J. 1* (1950), 35–47.

[29] James, M. Introduction to Manifolds, Lie Groups, and Estimation on the Circle. Preprint: TR 1987-194, 1987. `http://techreports.isr.umd.edu/`.

[30] Kailath, T. An Innovations Approach to Least Squares Estimation - Part I: Linear Filtering in Additive White Noise. *IEEE Trans. Automatic Control AC-13* (1968), 646–655.

[31] Kalman, R. E., and Bucy, R. New Results in Linear Filtering and Prediction Theory. *J. Basic Engr. ASME Series D 83* (1961), 95–108.

[32] Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng. ASME 82* (1960), 33–46.

[33] Karatzas, I., and Shreve, S. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 1997.

[34] Kendall, W. Stochastic Differential Geometry: An Introduction. *Acta Applicande Mathematicae 9* (1987), 29–60.

[35] Kendall, W. From Stochastic Parallel Transport to Harmonic Maps. Preprint, 1998. `citeseer.ist.psu.edu/kendall98from.html`.

[36] Kloeden, P. A Brief Overview of Numerical Methods for Stochastic Differential Equations. Preprint, 2001. `http://www.maths.uq.edu.au/~pmd/milano.ps`,.

[37] Kouritzin, M. On Exact Filters for Continuous Signals with Discrete Observations. Preprint: 1409, 1996. `http://www.ima.umn.edu/preprints/May96/May96.html`.

[38] Kshner, H. On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes. *SIAM J. control 2* (1964), 106–119.

[39] Kunita, H. *Lecture Notes in Mathematics - Stochastic Differential Equations - Connections with Nonlinear Filtering*, vol. 972. Springer Verlag, 1982.

[40] Kunita, H. *Lecture Notes on Stochastic Flows and Applications*. Springer Verlag, 1986.

[41] Kunzinger, M. D., Steinbauer, R., and Vickers, J. Generalized Connections and Curvature. Preprint: Math. Proc. Cambridge Philos. Soc (to appear), September 2004. `http://arxiv.org/`.

[42] Lara, M. D. A Note of the Symmetry Group and Perturbation Algebra of a Parabolic Differenctial Equation. *Journal of Mathamatical Physics 32*, 6 (1991), 1444–1449.

[43] Lara, M. D. Geometric and Symmetric Properties of a Nondegenerate Diffusion Process. *The Annuals of Probability 23*, 4 (1995), 1557–1604.

[44] Lara, M. D. Reduction of the Zakai Equation by Invariance Group Techniques. *Stochastic Prosesses and their Applications 73* (1998), 119–130.

[45] Lara, M. D. Finite-Dimensional Filters. Part II: Invariance Group Techniques. *SIAM J. Control Optim. 35*, 3 (May 1997), p 1002–1029.

[46] Lara, M. D. Finite-Dimensional Filters. Part I: The Wei-Norman Technique. *SIAM J. Control Optim. 35*, 3 (May 1997), 980–1001.

[47] Lara, M. D. Entangling and Disentangling Noise and Dynamics. Preprint, May 2001. `citeseer.ist.psu.edu/isard98condensation.html`.

[48] Lee, J. *Introduction to Smooth Manifolds*. Springer-Verlag, 2002.

[49] Liptser, R., and Shiryaev, A. *Statistics of Random Processes I*. New York, 1972.

[50] Lo, J., and Willsky, A. Estimation for Rotational Processes with One Degree of Freedom - Parts I,II & III. *IEEE Trans. Aut. Control AC-20*, 1 (Feb. 1975).

[51] Marcus, S. Algebraic and Geomtric Methods in Nonlinear Filtering. *SAIM J. Control and Optimization 22*, 6 (1984), 817–844.

[52] Maybeck, P. *Stochastic Models, Estimation and Control: Volume 1*. Mathematics in Science and Engineering, 1994.

[53] Mitter, P., and Viallet, C. On the Bundle of Connections and the Gauge Orbit Manifold in Yang-Mills Theory. *Communications in Mathematical Physics 79* (1981), 457–472.

[54] Mitter, S. On The Analogy Between Mathematical Problems Of Non-Linear Filtering and Quantum Physics. *Ricerche Di Automatica 10* (1979), 163–216.

[55] Mitter, S. *Lecture Notes in Mathematics - Lectures on Nonlinear Filtering and Stochastic Control*, vol. 972. Springer Verlag, 1982.

[56] Morrison, S. Connections on Principal Fibre Bundles. Preprint, 2000. `http://math.berkeley.edu/~scott/wiki/Mathematics/`.

[57] Mortensen, R. *Optimal Control of Continuous-Time Stochastic Systems*. PhD thesis, Uniersity of California, Berkeley, 1966.

[58] Nakahara, M. *Geometry, Topology and Physics*. Institute of Physics, 2003.

[59] Oksendal, B. *Stochastic Differential Equations : An Introduction with Applications*. Springer-Verlag, 2003.

[60] Rao, C., Rozovsky, B., and Tartakovsky, A. Domain Pursuit Method for Tracking Ballistic Targets. Preprint, 1998. `http://www.usc.edu/dept/LAS/CAMS/usr/facmemb/tartakov/`.

[61] Rodriguez, S. Wald Sequential Detection with Non-Guassian Pulsed Radar Data Using the Zakai Equation. Preprint Number: TR 1990-10, 1990. `http://techreports.isr.umd.edu/`.

[62] Schmid, R. Infinite Dimensional Lie Groups with Applications to Mathematical Physics. *Geometry and Symmetry in Physics 1* (2004), 1–67.

[63] Schutz, B. *Geometrical Methods of Mathematical Physics*. Cambridge University Press, 1988.

[64] Shiryaev, A. Some New Results in the Theory of Controlled Stochastic Processes. In *Trans. 4th Prague Conference on Information Theory* (1967), Czech Academy of Sciences, Prague.

[65] Stratonovich, R. *Conditional Markov Processes and Their Application to the Theory of Optimal Control*. New York: Elsevier, 1968.

[66] Svetlichny, G. Preparation for Gauge Theory. Preprint, 1999. `http://arxiv.org/abs/math-ph/9902027/`.

[67] Wong, E. *Stochastic Processes in Information and Dynamical Systems*. McGraw-Hill, 1971.

[68] Wonham, W. Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering. *SIAM J. Control 2* (1965), 347–369.

[69] Zakai, M. On The Optimal Filtering of Diffusion Processes. *Z. Wahrsch. Verw. Geb. 11* (1969), 230–243.

[70] Zarchan, P., and Musoff, H. *Fundamentals of Kalman Filtering: A Practical Approach*. AIAA (American Institute of Aeronautics & Ast), 2001.

INDEX

VITA

Rajesh Rugunanan was born on the 13^{th} August 1972 in Johannesburg. In 1994, he obtained his BSc. degree from Rhodes University with majors in Applied Mathematics (with distinction), Physics with Electronics and Mathematical Statistics (with distinction). In 1995 he obtained his BSc.Hons. (with distinction) in Electronics with the department of Physics and Electronics at Rhodes. He is currently employed by the Denel Aerospace Systems group as a Senior Engineer with a special interest in optimal control theory and estimation.