Lévy Processes and Filtering Theory

Stefan Blackwood

Thesis presented for examination for the degree of Doctor of Philosophy

School of Mathematics and Statistics
The University of Sheffield

June 2014

Abstract

Stochastic filtering theory is the estimation of a continuous random system given a sequence of partial noisy observations, and is of use in many different financial and scientific areas. The main aim of this thesis is to explore the use of Lévy processes in both linear and non-linear stochastic filtering theory.

In the existing literature, for the linear case the use of square integrable Lévy processes as driving noise is well known. We extend this by dropping the assumption of square integrability for the Lévy process driving the stochastic differential equation of the observations. We then explore a numerical example of infinite variance alpha-stable observations of a mean reverting Brownian motion with a Gaussian starting value, by comparing our derived filter with that of two others.

The rest of the thesis is dedicated to the non-linear case. The scenario we look at is a system driven by a Brownian motion and observations driven by an independent Brownian motion and a generalised jump processes. The result of our efforts is the famous Zakai equation which we solve using the change of measure approach. We also include conditions under which the change of measure is a martingale. Next, via computing a normalising constant, we derive the Kushner-Stratonovich equation.

Finally we prove the uniqueness of solution to the Zakai equation, which in turn leads to the uniqueness of solution to the Kusher-Stratonovich equation.

I am unable to find the words to thank my supervisor David Applebaum enough. His time, energy, patience and sense of humour are truly unbounded.

As eternity
is reckoned
there's a lifetime
in a second.
Piet Hein

Introduction

Suppose you have a random system which is not directly observable, instead you have a sequence of partial observations. Using the information gathered from these observations, what can we infer about the underlying system? Using stochastic models to make these deductions is known as stochastic filtering. The primary objective of this thesis is to explore stochastic filtering theory using Lévy processes as driving noise.

We begin by introducing the relevant background theory and some properties of square integrable Lévy processes which will be required in the subsequent chapters. Once this has been completed we move onto the main subject of this thesis, that of stochastic filtering theory. We begin by looking at the linear case, by deriving the famous Kalman-Bucy filter in finite dimensions for a square integrable Lévy processes with given finite variances. This is carried out by following the usual innovations process methodology.

We then extend the Kalman-Bucy filter by dropping the square integrability assumption on the observation noise. The main tools here will be Itó's lemma, approximation by bounded jumps, and limiting arguments. The key result here is that as we pass to the limit, the noise term in the Kalman-Bucy filter disappears. We conclude this chapter by looking at numerical simulation, we take our result for the infinite variance Kalman-Bucy filter and compare it with two other filters at their point of overlap with our filter. We show that in the case of infinite variance α-stable observations of a mean reverting Gaussian system, the calculation of estimates in the absence of noise gives a more accurate representation of the system.

The next chapter looks at the non-linear filtering case. We start by deriving the cornerstone of this field - the Zakai equation, this time using a Brownian motion and generalised jump process in the observations and standard Brownian motion in the system. The Zakai equation is solved using the standard change of measure approach; sufficient conditions for the change of measure to be a martingale are also proved. We complete this chapter by deriving the Kusher-Stratonovich equation, this involves computing a normalising constant to turn the measure valued Zakai equation into the probability measure valued process of the Kushner-Stratonovich equation. In order to find this normalising constant, we require a technical stopping time argument. We also present the beginnings of a alternative method which would require a further assumption.

Chapter 4 deals with the uniqueness of solution of the Zakai equation, which in turn leads to uniqueness of the Kusher-Stratonovich equation. This is done by transforming the solution to a stochastic differential equation in a Hilbert space and then using some estimates based on Hilbert space theory.

Contents

Notation

$\mathbb{R}^d$ is a d-dimensional Euclidean space, where $d \in \mathbb{N}$. The elements of $\mathbb{R}^d$ are vectors $x = (x_1, x_2, \ldots, x_d)$ with each $x_i \in \mathbb{R}$ for $1 \leq i \leq d$. The inner product in $\mathbb{R}^d$ is denoted by (x, y) where $x, y \in \mathbb{R}^d$ so that,

$$(x, y) = \sum_{i=1}^{d} x_i y_i,$$

or as $x^T y$ where x^T is the transpose of the vector x.

We then have the Euclidean norm $|x| = (x, x)^{1/2} = \left(\sum_{i=1}^{d} x_i^2\right)^{1/2}$.

The determinant of a square matrix A is written as $\det(A)$ and its trace is $\mathrm{tr}(A)$. The identity matrix will always be denoted I. The space of $n \times m$ real valued matrices is denoted $\mathcal{M}_{n,m}(\mathbb{R})$. For any matrix A, the transpose of A will be written as A^T.

We will sometimes write $\mathbb{R}^+ = [0, \infty)$, and $\mathbb{R}_0^d = \mathbb{R}^d - \{0\}$.

For $1 \leq n \leq \infty$ we write $C_c^n(\mathbb{R}^d)$ to denote the space of all n times differentiable functions of compact support from $\mathbb{R}^d$ to $\mathbb{R}$, all of whose derivatives are continuous. The jth partial derivative of $f \in C^1(\mathbb{R}^d)$ at $x \in \mathbb{R}^d$ will sometimes be written $(\partial_j f)(x)$.

Let $\mathcal{B}(S)$ be the Borel σ-algebra of a Borel set $S \subseteq \mathbb{R}^d$. Elements of $\mathcal{B}(S)$ are called Borel sets, and any measure on $(S, \mathcal{B}(S))$ is called a Borel measure. Let $(S, \mathcal{S}, \mu)$ be an arbitrary measure space, then will write μ_A to denote the restriction of the measure μ to the set $A \in \mathcal{B}(S)$. The Lebesgue measure on $\mathbb{R}$ is written as $\mathcal{L}eb$.

We will make use of Landau notation, according to which $(o(n), n \in \mathbb{N})$ is any real valued sequence for which $\lim_{n\to\infty}(o(n)/n) = 0$ and $(O(n), n \in \mathbb{N})$ is any non negative sequence for which $\limsup_{n\to\infty}(O(n)/n) < \infty$.

Given two random variables X and Y we will write $\mathrm{Cov}(X, Y)$ meaning the covariance of X and Y, and $\mathrm{Var}(X)$ meaning the variance of X. We will also write $X \stackrel{d}{=} Y$ if X is equal to Y in distribution.

For $1 \leq p < \infty$ let $L^p(S, \mathcal{S}, \mu; \mathbb{R}^d)$ be the Banach space of all equivalence classes of mappings $S \to \mathbb{R}^d$ which coincide a.e with respect to μ such that $||f||_p < \infty$ where $|| \cdot ||_p$ denotes the norm

$$||f||_p = \left[\int_S |f(x)|^p \mu(dx)\right]^{1/p}$$

In particular when $p = 2$ we have a Hilbert space with respect to the inner product

$$\langle f, g\rangle = \int_S (f(x), g(x))\, \nu(dx).$$

The open ball of radius r centered at x in $\mathbb{R}^d$ is denoted $B_r(x) = \left\{x \in \mathbb{R}^d; |y - x| < r\right\}$ and we write $\hat{B} = B_1(0)$. if f is a mapping between two sets A and B, we denote its range as $\mathrm{Ran}(f) = \{y \in B; y = f(x) \text{ for some } x \in A\}$.

1 Review of Stochastic Calculus for Lévy Processes

In this chapter we give a brief introduction to the background and required probability tools that will be used throughout this thesis.

1.0.1 Filtrations and Adapted Processes

Let $(\Omega, \mathcal{F}, P)$ be a probability space where $\mathcal{F}$ is a σ-algebra of a given set Ω. A collection of sub σ-algebras of $\mathcal{F}$, denoted $(\mathcal{F}_t, t \geq 0)$ is termed a filtration if we have,

$$\mathcal{F}_s \subseteq \mathcal{F}_t \quad \text{for all} \;\; s \leq t.$$

A probability space is said to be *filtered* if it is equipped with a filtration, as above. During this thesis we will be working with a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t, t \geq 0), P)$, which is a probability space satisfying the usual hypotheses listed below, and unless stated otherwise all random variables are defined on this space.

Definition 1.0.1. (Usual Hypotheses) *Given a probability space* $(\Omega, \mathcal{F}, P)$ *a filtration* $\{\mathcal{F}_t\}_{t \geq 0}$ *is said to satisfy the* usual hypotheses *if,*

1 $\mathcal{F}_0$ *contains all the* P*-null sets of* $\mathcal{F}$.

2 $\mathcal{F}_t = \bigcap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}$; *that is the filtration* $\mathcal{F}$ *is right continuous.*

In the following work, when there is mention of measurable mappings taking values in a subset of $\mathbb{R}^d$ we will assume that $\mathbb{R}^d$ is equipped with the Borel σ-algebra, $\mathcal{B}(\mathbb{R}^d)$, i.e the smallest σ-algebra of subsets of $\mathbb{R}^d$ containing all open sets.

Let $X = (X(t), t \geq 0)$ be a stochastic process defined on $(\Omega, \mathcal{F}, P)$, then we say X is adapted to the filtration $(\mathcal{F}_t, t \geq 0)$ if $X(t)$ is $\mathcal{F}_t$-measurable for each $t \geq 0$. Given an $\mathcal{F}_t$ adapted process X which satisfies $\mathbb{E}\left[|X(t)|\right] < \infty$ for each $t \geq 0$, then if for $0 \leq s < t < \infty$,

$$\mathbb{E}\left[X(t)|\mathcal{F}_s\right] = X(s) \;\; \text{a.s.},$$

we call X a martingale. For more on martingales see [52].

1.0.2 Characteristic Functions

We define the indicator function $1_A(x)$ which equals 1 when $x \in A$ and equals 0 when $x \notin A$. Let μ_1 and μ_2 be two probability measures on $\mathbb{R}^d$, then we define the convolution of these two measures as

$$(\mu_1 * \mu_2)(A) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_A(x+y)\mu_1(dx)\mu_2(dy) \tag{1}$$

for each $A \in \mathcal{B}(\mathbb{R}^d)$.

Given a random variable X on $(\Omega, \mathcal{F}, P)$ taking values on $\mathbb{R}^d$ with probability law p_X we define its *characteristic function* $\phi_X : \mathbb{R}^d \to \mathbb{C}$ by

$$\phi_X(u) = \mathbb{E}[e^{i(u,X)}] = \int_{\mathbb{R}^d} e^{i(u,x)} p_X(dx),$$

for each $u \in \mathbb{R}^d$. The characteristic function is therefore similar to a moment generating function but with the real part replaced with iu, this has the advantage that it always exists since $x \to e^{i(u,x)}$ is bounded. Note the following fundamental properties of the characteristic function.

(i) If μ_1 and μ_2 have characteristic functions $\phi_1(u)$ and $\phi_2(u)$ then $\mu_1 * \mu_2$ has characteristic function $\phi_1(u)\phi_2(u)$.

(ii) The characteristic function uniquely determines the distribution.

See [7, p.342] for details. More generally, if μ is a probability measure on $\mathbb{R}^d$ then its characteristic function is defined by $\phi_\mu(u) = \int_{\mathbb{R}^d} e^{i(u,y)}\mu(dy)$ for $u \in \mathbb{R}^d$.

1.1 Lévy Processes

Before defining a Lévy process we begin by contrasting the similarities and differences of two well known processes, Brownian motion and Poisson processes, which are covered below.

We say a process has *independent increments* if $\forall\ n \in \mathbb{N}, 0 \leq t_1 < t_2 < \cdots < t_{n+1} < \infty$ the random variables $(X(t_{j+1}) - X(t_j), 1 \leq j \leq n)$ are independent. A process has stationary increments if each $X(t_{j+1}) - X(t_j) \stackrel{d}{=} X(t_{j+1} - t_j) - X(0)$.

It was Norbert Wiener [56] in 1923 who defined Brownian motion in a mathematical sense and showed it to exist.

Definition 1.1.1. (Brownian Motion)

A real valued process $B = (B(t), t \geq 0)$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is said to be a Brownian motion *if the following hold.*

(i) The paths of B are P-(a.s) continuous.

(ii) $B(0) = 0$ *a.s.*

(iii) B *has stationary and independent increments.*

(iv) *For each* $t > 0$*,* $B(t)$ *is equal in distribution to a normal random variable with mean zero and variance* t*.*

Definition 1.1.2. (Poisson Process)
A non-negative integer valued process $N = (N(t), t \geq 0)$ *defined on a probability space* $(\Omega, \mathcal{F}, P)$ *is said to be a* Poisson process *with intensity* $\lambda > 0$ *if the following hold.*

(i) *The paths of* N *are* P*-(a.s) right continuous with left limits.*

(ii) $N(0) = 0$ *a.s.*

(iii) N *has stationary and independent increments.*

(iv) *For each* $t > 0$*,* $N(t)$ *is equal in distribution to a Poisson random variable with parameter* λt*.*

On first look, these two processes can appear to be rather different. B has continuous paths [7, p.501], [31, p.387], whereas N does not, N is non-decreasing whereas B is not. The paths of N are almost surely of finite variation over finite time intervals [32, p.228] whereas the paths of B are almost surely of infinite variation over finite time intervals [38], [3, p.112]. On closer inspection however, we see that they do have some common properties. Both processes begin from the origin, have stationary and independent increments and have sample paths that are right continuous with left limits. We can then use these properties to define a more general class of stochastic processes, which we call Lévy processes.

Definition 1.1.3. (Lévy Process)
A stochastic process $X = (X(t), t \geq 0)$ *defined on a probability space* (Ω, F, P) *is a Lévy process if it has the following properties.*

$(L1)$ $X(0) = (0)$ $(a.s)$,

$(L2)$ X *has stationary and independent increments*,

$(L3)$ X *is* stochastically continuous, *i.e* $\forall\ a > 0, s \geq 0$

$$\lim_{t \to s} P(|X(t) - X(s)| > a) = 0.$$

For other aspects of Lévy processes, including subordinators, potential theory and Markov properties see [6], and for fluctuation theory see [38] and [19]. In Chapter 2 we will need additive processes which are a little more general than Lévy processes so we introduce them briefly here.

Definition 1.1.4. (Additive Processes)
A stochastic process $(X(t), t \geq 0)$ *on* $\mathbb{R}^d$ *is an additive process if the following conditions are satisfied.*

(1) $X_0 = 0$ *(a.s)*

(2) X *has independent increments*

(3) X *is stochastically continuous.*

Before moving on, we give a more precise mathematical formulation of sample paths that are right continuous with left limits.

Definition 1.1.5. (Càdlàg Functions)
A function $f : [0, T] \to \mathbb{R}^d$ *is said to be càdlàg if it is right continuous with left limits, i.e for each* $t \in [0, T]$ *the limits,*

$$f(t-) = \lim_{s \to t, s < t} f(s) \qquad\qquad f(t+) = \lim_{s \to t, s > t} f(s)$$

exist and are finite and $f(t) = f(t+)$.

Obviously every continuous function is càdlàg, but càdlàg functions can have discontinuities. If t is a point at which a discontinuity occurs then we denote the jump of a function f at time t by,

$$\Delta f(t) = f(t) - f(t-). \tag{2}$$

A càdlàg function can have at most a countable number of jump discontinuities, i.e $\{t \in [0, T] : f(t) \neq f(t-)\}$ is countable; see [21] for proof. Further to this, every Lévy process has a càdlàg modification that is itself a Lévy process see [3, Theorem 2.1.8], and we will always use this modification.

Remark 1.1.6. *When dealing with stochastic processes we should always read càdlàg as (a.s)-càdlàg, i.e if* X *is a càdlàg Lévy process then* $\exists\ \Omega_0 \in \mathcal{F}$ *where* $P(\Omega_0) = 1$ *such that* $t \to X(t)(\omega)$ *is càdlàg* $\forall\ \omega \in \Omega_0$.

For more information on càdlàg functions see [3, p.139], [15, p.37] [27, p.34] [8, p.119].

1.1.1 The Lévy-Khintchine Formula

Definition 1.1.7. (Infinite Divisibility)

Let $\mathcal{M}_1(\mathbb{R}^d)$ be the set of all Borel probability measures on $\mathbb{R}^d$ then $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ is infinitely divisible if for each $n \in \mathbb{N}$ there exists another measure $\mu^{1/n} \in \mathcal{M}_1(\mathbb{R}^d)$ such that $\mu = \left(\mu^{1/n}\right)^{*n}$, where μ^{*n} is the n-fold convolution of μ.

We also introduce infinite divisibility for random variables. Let X be a random variable taking values in $\mathbb{R}^d$ with law μ_X. We say X is *infinitely divisible* if, $\forall\ n \in \mathbb{N}$ there exist i.i.d random variables $Y_1^{(n)}, \dots, Y_n^{(n)}$ such that

$$X \stackrel{d}{=} Y_1^{(n)} + \cdots + Y_n^{(n)}.$$

Note that if X is infinitely divisible then its law μ_X is infinitely divisible in the sense of Definition 1.1.7.

To present the Lévy-Khintchine formula, firstly we need the concept of a Lévy measure.

Definition 1.1.8. (Lévy Measure)

Let ν be a Borel measure defined on $\mathbb{R}^d - \{0\}$. We say that it is a Lévy measure if

$$\int_{\mathbb{R}^d - \{0\}} (|x|^2 \wedge 1)\nu(dx) < \infty. \tag{3}$$

This is just a streamlined way of stating that on $\mathbb{R}^d - \{0\}$,

$$\int_{|x|<1} |x|^2 \nu(dx) < \infty \quad \text{and} \quad \nu(|x| \geq 1) < \infty.$$

It is worth noting that every Lévy measure on $\mathbb{R}^d - \{0\}$ is σ-finite, see [3, p.29].

Theorem 1.1.9. (Lévy-Khintchine Formula)

A probability measure μ on $\mathbb{R}^d$ is infinitely divisible if and only if there exists a vector $b \in \mathbb{R}^d$ a non-negative definite symmetric $d \times d$ matrix A and a Lévy measure ν on $\mathbb{R}^d - \{0\}$ such that for all $u \in \mathbb{R}^d$

$$\phi_\mu(u) = \exp\left\{ i(b,u) - \frac{1}{2}(u, Au) + \int_{\mathbb{R}^d - \{0\}} [e^{i(u,y)} - 1 - i(u,y)1_{\hat{B}}(y)]\nu(dy) \right\}, \tag{4}$$

where $\hat{B}$ is the open unit ball without zero, i.e $\{x : |x| < 1\} - \{0\}$.

$\square$

For full proof see [54, p.40], [43], [40], [24].

We say that (b, A, ν) are the characteristics of the Lévy-Khintchine formula; for example, in the Gaussian case b is the mean, A is the covariance and $\nu = 0$. Also for notational convenience we denote $\phi_\mu(u) = e^{\eta(u)}$.

We can see that Gaussian random variables are infinitely divisible with each $Y_j^{(n)} \sim N(b/n, (1/n)A)$ for all $n \in \mathbb{N}$ and each $1 \leq j \leq n$ and have characteristic function

$$\phi_X(u) = \exp[i(b, u) - \frac{1}{2}(u, Au)], \tag{5}$$

see [3, Example 1.2.8] for details.

If $(Z(n), n \in \mathbb{N})$ is a sequence of i.i.d $\mathbb{R}^d$-valued random variables, then we define a compound Poisson random variable, $X = Z(1) + \cdots + Z(N)$ where N is an independent Poisson process with intensity λ. If X is a compound Poisson process we write $X \sim \pi(\lambda, \mu_Z)$ and it is infinitely divisible with each $Y_j^{(n)} \sim \pi(\lambda/n, \mu_Z)$ and characteristic function

$$\phi_X(u) = \exp[\lambda(\phi_Z(u) - 1)]. \tag{6}$$

See [3, Proposition 1.2.11] for details.

We can see that for each $t \geq 0$ a Lévy process $X(t)$ is infinitely divisible, see [3, p.43], [38, p.4]. What is not quite so clear is whether given an infinitely divisible distribution we can construct a Lévy process X, such that $X(1)$ has this distribution.

Theorem 1.1.10. (Lévy-Khintchine Formula for a Lévy Process) *Let $b \in \mathbb{R}^d$, A be a non-negative definite symmetric $d \times d$ matrix, and ν a Lévy measure. From this triple define for each $u \in \mathbb{R}$,*

$$\eta(u) = i(b, u) - \frac{1}{2}(u, Au) + \int_{\mathbb{R}-\{0\}} [e^{i(u,y)} - 1 - i(u, y)1_{\hat{B}}(y)]\nu(dy). \tag{7}$$

Then there exists a Lévy process $X = (X(t), t \geq 0)$ such that

$$\mathbb{E}[e^{i(u,X(t))}] = e^{t\eta(u)} \tag{8}$$

where (b, A, ν) are the characteristics of $X(1)$. $\square$

See [3, p.127] for proof.

We will require the following theorem, proved here, in Theorem 3.3.3.

Theorem 1.1.11. *Given a probability space $(\Omega, \mathcal{F}, P)$, and filtration $(\mathcal{F}_t, t \geq 0)$ then if $(X(t), t \geq 0)$ is a càdlàg adapted stochastically continuous process such that for all $0 \leq s \leq t < \infty$, $u \in \mathbb{R}^d$ and $X(0) = 0$ a.s where*

$$\mathbb{E}\left[e^{i(u,X(t)-X(s))}|\mathcal{F}_s\right] = e^{(t-s)\eta(u)} \tag{9}$$

then X is a Lévy process.

Proof. Firstly, let $s = 0$, by (9), $\mathbb{E}\left[e^{i(u,X(t))}\right] = \mathbb{E}\left[e^{i(u,X(t))}|\mathcal{F}_0\right] = e^{t\eta(u)}$.

We need to show that X has stationary and independent increments. Starting with stationarity, from (9)

$$\begin{aligned}\mathbb{E}\left[e^{i(u,X(t)-X(s))}\right] &= \mathbb{E}\left[\mathbb{E}\left[e^{i(u,X(t)-X(s))}|\mathcal{F}_s\right]\right] \\ &= \mathbb{E}\left[e^{(t-s)\eta(u)}\right] \\ &= \mathbb{E}\left[e^{i(u,X(t-s))}\right]. \end{aligned} \tag{10}$$

Now since the characteristic function uniquely determines the law of X

$$X(t) - X(s) \stackrel{d}{=} X(t-s).$$

For independent increments: for $r < s < t$, $u_1, u_2 \in \mathbb{R}^d$,

$$\begin{aligned}\mathbb{E}\left[e^{i(u_1,X(r))+i(u_2,X(t)-X(s))}\right] &= \mathbb{E}\left[e^{i(u_1,X(r))}\mathbb{E}\left[e^{i(u_2,X(t)-X(s))}|\mathcal{F}_s\right]\right] \\ &= \mathbb{E}\left[e^{i(u_1,X(r))}\right] e^{(t-s)\eta(u_2)} \\ &= \mathbb{E}\left[e^{i(u_1,X(r))}\right] \mathbb{E}\left[e^{i(u_2,X(t-s))}\right] \\ &= \phi_{X(r)}(u_1)\phi_{X(t-s)}(u_2) \\ &= \phi_{X(r)}(u_1)\phi_{X(t)-X(s)}(u_2), \text{ by (10).}\end{aligned}$$

So by Kac's theorem (see [3, p.18]) we have independent increments. □

1.1.2 Stable Processes

We now introduce stable processes, as they will form the basis of the numerical simulations which will be carried out at the end of chapter 2.

Definition 1.1.12. (Stable Random Variables)
A random variable Y is said to have a stable distribution if for all $n \geq 1$ we have the following distributional equality,

$$Y_1 + \cdots + Y_n \stackrel{d}{=} a_n Y + m_n \tag{11}$$

where $Y_1, \ldots, Y_n$ are independent copies of Y, $a_n > 0$ and $m_n \in \mathbb{R}$. The name stable *comes from the above mentioned stability under addition property.*

By subtracting m_n/n from each term on the left hand side of (11) we can see that this implies that any stable random variable is also infinitely divisible. It can be shown, see [53, Corollary 2.1.3], that the only choice of

a_n in (11) is of the form $n^{1/\alpha}$, $0 < \alpha \leq 2$. This constant is termed the index of stability.

Consider the sum,

$$S_n = \frac{1}{\sigma_n}(Y_1 + \cdots + Y_n - c_n) \tag{12}$$

where $(c_n, n \in \mathbb{N})$ is a sequence in $\mathbb{R}$, $(\sigma_n, n \in \mathbb{N})$ is a sequence in $\mathbb{R}^+$. By the usual central limit theorem, if each $c_n = nb$ and $\sigma_n = \sqrt{n}\sigma$ then S_n converges in distribution to Y which is normally distributed with mean b and variance σ^2. It can be shown that a random variable is stable if it arises as a limit of (12), and so is therefore a generalisation of the normal CLT. It is not difficult to see that (12) and (11) are equivalent, see [9], [22]. The Gaussian case has finite variance and index of $\alpha = 2$. It can be shown that $\mathbb{E}[Y^2] < \infty$ if and only if $\alpha = 2$ and also that $\mathbb{E}[|Y|] < \infty$ if and only if $1 < \alpha \leq 2$, see [15, p.95].

The characteristics in the Lévy-Khintchine formula are given by the result below.

Theorem 1.1.13. (Characteristics for a Stable Random Variable)
If X is a stable real-valued random variable, then its characteristics must take one of the two following forms:

(1) When $\alpha = 2, \nu = 0$ so $X \sim N(b, A)$;

(2) When $\alpha \neq 2, A = 0$ and

$$\nu(dx) = \frac{c^+}{x^{1+\alpha}} 1_{(x>0)}(x)dx + \frac{c^-}{|x|^{1+\alpha}} 1_{(x<0)}(x)dx$$

where $c^+ \geq 0, c^- \geq 0$ and $c_1 + c_2 > 0$.

□

A proof can be found in [54, p.80].

1.2 The Lévy Itô Decomposition

For the rest of this report we will require the following assumptions as they are required for the proof of the Lévy Itô decomposition, and we refer the reader to [3, p.90] for further details.

A1 Every Lévy process $X = \big(X(t), t \geq 0\big)$ is $\mathcal{F}_t$-adapted and has càdlàg sample paths,

A2 Sharpen (L2) to $X(t) - X(s)$ is independent of $\mathcal{F}_s$ for all $0 \leq s < t < \infty$.

1.2.1 The Jumps of Lévy Processes

Recall from (2) the jumps of a càdlàg function were defined as $\Delta f(t) = f(t) - f(t-)$. We now introduce the jump of a Lévy process $(\Delta X(t), t \geq 0)$ by

$$\Delta X(t) = X(t) - X(t-) \quad \text{for each } t \geq 0.$$

From the above definition, it is clear that ΔX is an adapted process, however it is not a Lévy process as it can be shown to fail the independent increments property, see [3, Example 2.3.1].

Definition 1.2.1. (Counting jumps)
Let $0 \leq t < \infty$ and $A \in \mathcal{B}(\mathbb{R}^d - \{0\})$, then we define

$$N(t, A)(\omega) = \#\{0 \leq s \leq t | \Delta X(s)(\omega) \in A\} = \sum_{0 \leq s \leq t} 1_A \left(\Delta X(s)(\omega) \right), \quad (13)$$

if $\omega \in \Omega_0$ and by convention $N(t, A)(\omega) = 0$ if $\omega \in \Omega_0^c$, the complement of Ω_0. Recall our definition of Ω_0 from Remark 1.1.6.

We say A is bounded below if 0 is not in $\bar{A}$, the closure of A. Given this condition then $N(t, A) < \infty$ for all $t \geq 0$; see [3, p.101] for a proof. If A is not bounded below, i.e we allow 0 in the closure of A then $N(t, A)$ may no longer be finite for all $t \geq 0$ as infinite quantities of small jumps may accumulate to give infinite mass.

Let S be a subset of $\mathbb{R}^d - \{0\}$ and N an integer valued random random measure on $U = \mathbb{R}^+ \times S$. For a definition of random measures see [3, p.102],[26] and [49, Appendix F].

Definition 1.2.2. *N is a Poisson random measure if,*

(1) For each $t \geq 0$ and $A \in \mathcal{B}(S)$ bounded below, $N(t, A)$ has a Poisson distribution.

(2) For each $t_1, \ldots, t_n \in \mathbb{R}^+$ and each disjoint family $A_1, \ldots, A_n \in \mathcal{B}(S)$ bounded below, the random variables $\big(N(t_i, A_i), 1 \leq i \leq n\big)$ are independent.

The intensity measure of N is defined as $\mu(A) = \mathbb{E}[N(1, A)]$ for all A bounded below and $\mu(A) < \infty$.

We now define the compensated Poisson random measure for A bounded below as

$$\tilde{N}(t, A) = N(t, A) - t\mu(A). \quad (14)$$

1.2.2 Poisson Integration

Define the Poisson integral of a Borel measurable function f from $\mathbb{R}^d \to \mathbb{R}^d$ as a random finite sum by

$$\int_A f(x)N(t,dx) = \sum_{x\in A} f(x)N(t,\{x\}) = \sum_{0\le u\le t} f(\Delta X(u))1_A(\Delta X(u)) \quad (15)$$

where A is bounded below, $t > 0$. The last equality is due to the fact that $N(t,\{x\}) \neq 0$ if and only if $\Delta X(u) = x$ for at least one value of u between $[0,t]$.

Theorem 1.2.3. *Let A be bounded below. Then:*

(1) $\int_A f(x)N(t,dx)$ *has a compound Poisson distribution such that for each* $t \geq 0, u \in \mathbb{R}^d$,

$$\mathbb{E}\left(\exp\left[i\left(u, \int_A f(x)N(t,dx)\right)\right]\right) = \exp\left(t\int_{\mathbb{R}^d}(e^{i(u,x)}-1)\mu_{f,A}(dx)\right)$$

where $\mu_{f,A}(B) = \mu(A\cap f^{-1}(B))$ *for each* $B \in \mathcal{B}(\mathbb{R}^d)$.

(2)

$$\mathbb{E}\left[\int_A f(x)N(t,dx)\right] = t\int_A f(x)\mu(dx)$$

for $f \in L^1(A,\mu_A)$.

(3)

$$Var\left[\left|\int_A f(x)N(t,dx)\right|\right) = t\int_A |f(x)|^2\mu(dx).$$

for $f \in L^2(A,\mu_A)$.

□

For proof see [3, Theorem 2.3.7], [54, Proposition 19.5], also see [30].

We can now define the compensated Poisson integral for $f \in L^1(A,\mu_A), t \geq 0$ and A bounded below by,

$$\int_A f(x)\tilde{N}(t,dx) = \int_A f(x)N(t,dx) - t\int_A f(x)\mu(dx). \quad (16)$$

A result that we will use frequently later is that the compensated Poisson integral is a martingale, see [38, p.46] for proof. From here on we write,

$$\int_{|x|<1}\tilde{N}(t,dx) = \lim_{n\to\infty}\int_{\epsilon_n<|x|<1}\tilde{N}(t,dx)$$

where $(\epsilon_n, n \in \mathbb{N})$ is any sequence monotonically decreasing to zero, and the limit is taken in the L^2 sense. For existence of the limit see [3, Theorem 2.4.11].

1.2.3 The Lévy-Itô Decomposition

Let A be a non-negative definite symmetric $d \times d$ matrix, with square root σ such that σ is a $d \times m$ matrix and $A = \sigma\sigma^T$. Let B be a standard Brownian motion in $\mathbb{R}^m$, then we write

$$B_A(t) = \sigma B(t).$$

The Lévy-Itô decomposition was established by Lévy in [40] and [41] and proved by Itô in [24]. For a fuller account see [54, Ch.4], [3, Ch.2], [51, p.327].

Theorem 1.2.4. (The Lévy-Itô Decomposition)
If X is a Lévy process, then there exists;

(i) $b \in \mathbb{R}^d$,

(ii) A Brownian motion B_A with covariance matrix A ,

(iii) An independent Poisson random measure N on $\mathbb{R}^+ \times (\mathbb{R}^d - \{0\})$,

such that, for each $t \geq 0$;

$$X(t) = bt + B_A(t) + \int_{|x|<1} x\tilde{N}(t, dx) + \int_{|x|\geq 1} xN(t, dx) \tag{17}$$

where $b = \mathbb{E}\Big(X(1) - \int_{|x|\geq 1} xN(1, dx)\Big)$.

□

Remark 1.2.5. *The choice of integrating above and below 1, is purely for convenience. But the random processes are very different, the integral below one is an L^2 limit of a compensated sum of jumps, summed over $\epsilon_n < |x| < 1$. The integral greater than one describes the "large" jumps and is a compound Poisson process.*

1.3 Square Integrable Lévy Processes

In this section we develop square integrable Lévy processes which will be used extensively in chapter 2. We will then define orthogonal increment processes and show how they relate back to square integrable Lévy processes. We finish the section by calculating the variance for a mean zero square integrable Lévy processes.

Firstly we note the following useful result.

Theorem 1.3.1. *If X is a Lévy process and $n \in \mathbb{N}$, $\mathbb{E}[|X(t)|^n] < \infty$ for all $t > 0$ if and only if $\int_{|x|\geq 1} |x|^n \nu(dx) < \infty$.*

See [3, p.132] for proof. □

If $\mathbb{E}[|X(t)|] < \infty$ for all $t > 0$ then we may define,

$$\int_{|x|\geq 1} x\tilde{N}(t, dx) = \int_{|x|\geq 1} xN(t, dx) - t\int_{|x|\geq 1} \nu(dx)$$

and we may also compensate over the whole range of jumps by defining,

$$\int_{R^d-\{0\}} x\tilde{N}(t, dx) := \int_{|x|<1} x\tilde{N}(t, dx) + \int_{|x|\geq 1} x\tilde{N}(t, dx).$$

Lemma 1.3.2. *The Lévy-Itó decomposition for a square integrable Lévy process is of the form*

$$X(t) = b't + B_A(t) + \int_{\mathbb{R}^d-\{0\}} x\tilde{N}(t, dx), \tag{18}$$

where $b' = b + \int_{|x|\geq 1} x\nu(dx)$.

Proof. By Theorem 1.3.1 for all $t \geq 0$ and using [3, Theorem 2.4.7],

$$\begin{aligned} \mathbb{E}(|X(t)|^2) < \infty \iff & \mathbb{E}\left(\left|\int_{|x|\geq 1} xN(t, dx)\right|^2\right) < \infty \\ \iff & \int_{|x|\geq 1} |x|^2\nu(dx) < \infty. \end{aligned} \tag{19}$$

Also note that by using the Cauchy-Schwarz inequality,

$$\begin{aligned}\left|\int_{|x|\geq 1} x\nu(dx)\right|^2 &= \left|\int_{|x|\geq 1} x\cdot 1\nu(dx)\right|^2 \\ &\leq \int_{|x|\geq 1} |x|^2\nu(dx)\nu\{|x|\geq 1\} < \infty,\end{aligned}$$

and so b' is finite. Therefore from Theorem 1.2.4,

$$\begin{aligned}X(t) &= bt + B_A(t) + \int_{|x|<1} x\tilde{N}(t,dx) + \int_{|x|\geq 1} xN(t,dx) \\ &= bt + B_A(t) + \int_{|x|<1} x\tilde{N}(t,dx) + \int_{|x|\geq 1} x\tilde{N}(t,dx) + t\int_{|x|\geq 1} x\nu(dx) \\ &= b't + B_A(t) + \int_{\mathbb{R}^d-\{0\}} x\tilde{N}(t,dx).\end{aligned}$$

□

Next we look at the properties of these square integrable Lévy processes.

Theorem 1.3.3. *A Lévy process is a martingale if and only if it is of the form*

$$X(t) = B_A(t) + \int_{\mathbb{R}^d-\{0\}} x\tilde{N}(t,dx). \tag{20}$$

Proof. The form (20) is a martingale as it consists of a sum of a Brownian motion, widely known to be a martingale see [27, p.271], and a compensated Poisson integral proven to be a martingale in [38, p.46]. Conversely, if a Lévy process is a martingale then by Theorem 1.3.1 for each $t \geq 0$, $\mathbb{E}\big(|X(t)|\big) < \infty \iff \int_{|x|\geq 1} |x|\nu(dx) < \infty$ thus by a similar argument to the proof of Theorem 1.3.2 we can write $X(t)$ as in (18). By definition martingales have constant expectation $\forall\ t \geq 0$, and $\mathbb{E}\big(X(t)\big)$ is constant if and only if $b' = 0$. □

This now allows us to show the following.

Corollary 1.3.4. *A Lévy process X has mean zero if and only if it is a martingale.*

Proof. Suppose X is a martingale, then X is of the form (20), and it is known that $\big(B_A(t), t \geq 0\big)$ and $\big(\int_{\mathbb{R}^d-\{0\}} x\tilde{N}(t,dx), t \geq 0\big)$ are centred martingales, i.e each has mean zero, hence $\mathbb{E}[X(t)] = 0$. Conversely, if X is mean zero, following the same argument as Lemma 1.3.2, X can be written in the form (18) and $b' = 0$. Then from Theorem 1.3.3 it is a martingale. □

1.4 Orthogonal Increments

We want to consider a general square integrable stochastic process $(X(t), t \geq 0)$ taking values in $\mathbb{R}^d$.

Lemma 1.4.1. *If X is a square integrable stochastic process then for any pair of non-overlapping intervals $(s,t),(s',t')$*

$$\left|\mathbb{E}\left[(X(t)-X(s))\left(X(t')-X(s')\right)^T\right]\right|^2 < \infty.$$

Proof. The result follows from a straightforward application of the Cauchy-Schwarz inequality. $\square$

Definition 1.4.2. *A square integrable stochastic process is said to have* orthogonal increments *if for any non overlapping intervals $(s,t),(s',t')$*

$$\mathbb{E}\left[(X(t)-X(s))\left(X(t')-X(s')\right)^T\right] = 0. \tag{21}$$

Theorem 1.4.3. *A square-integrable martingale has orthogonal increments.*

Proof. For $0 \leq s' \leq t' \leq s \leq t \leq \infty$

$$\begin{aligned}
&\mathbb{E}\left[(X(t)-X(s))\left(X(t')-X(s')\right)^T\right] \\
&\qquad = \mathbb{E}\left[\mathbb{E}\left[(X(t)-X(s))\left(X(t')-X(s')\right)^T\right]\Big|\mathcal{F}_s\right] \\
&\qquad = \mathbb{E}\left[\mathbb{E}\left[\big(X(t)-X(s)\big|\mathcal{F}_s\big)\left(X(t')-X(s')\right)^T\right]\right] \\
&\qquad = 0,
\end{aligned}$$

where in the last step we used the martingale property that $\mathbb{E}\left[X(t)|\mathcal{F}_s\right] = X(s)$. $\square$

Corollary 1.4.4. *A mean zero square integrable Lévy process has orthogonal increments.*

Proof. This is a direct consequence of Theorem 1.4.3 and Corollary 1.3.4. $\square$

1.4.1 Computation of Variance

The variance of a one dimensional zero mean square integrable Lévy process is $\mathrm{Var}[X(t)] = \mathbb{E}[X(t)^2]$. Furthermore using (20),

$$\begin{aligned}\mathbb{E}[X(t)^2] &= \mathbb{E}[B(t)^2] + 2\mathbb{E}[B(t)]\mathbb{E}\left[\int_{\mathbb{R}-\{0\}} x\tilde{N}(t,dx)\right] \\ &\quad + \mathbb{E}\left[\left(\int_{\mathbb{R}-\{0\}} x\tilde{N}(t,dx)\right)^2\right] \\ &= t\left(\sigma^2 + \int_{\mathbb{R}-\{0\}} |x|^2\nu(dx)\right) \quad \text{using [3, p.109].} \qquad (22)\end{aligned}$$

Similarly in d dimensions we have for $1 \leq i, j \leq d$ and recalling the characteristics of a Lévy process (b, A, ν)

$$\begin{aligned}\mathrm{Cov}\left[X_i(t), X_j(t)\right] &= \mathbb{E}\left[X_i(t)X_j(t)^T\right] \\ &= A_{i,j}t + t\left(\int_{R^d-\{0\}} x_i x_j \nu(dx)\right). \qquad (23)\end{aligned}$$

We will denote $A_{i,j}t + t\left(\int_{R^d-\{0\}} x_i x_j \nu(dx)\right) = \rho_{i,j}$ for future convenience.

1.5 Stochastic Integration

This section introduces some key elements of stochastic calculus for Lévy processes. The main topics here will be the construction of stochastic integrals based on Lévy processes, Itô's formula for a Lévy process, Itô's product formula and some basic properties of the space $\mathcal{H}_2(t)$.

We refer the reader to [3, Section 2.2.1] for the definition of the predictable σ-algebra. To be able to deal with jumps we need a slight generalisation of this:

Fix $E \in \mathcal{B}(\mathbb{R}^d)$ and $0 < T < \infty$ and let $\mathcal{P}$ denote the smallest σ-algebra with respect to which the mappings $F : [0,T] \times E \times \Omega \to \mathbb{R}$ satisfying the following two conditions are measurable

(1) for $t \in [0,T]$ the mapping $(x, \omega) \to F(t,x,\omega)$ is $\mathcal{B}(E) \otimes \mathcal{F}_t$ measurable

(2) for each $x \in E$, $\omega \in \Omega$ the mapping $t \to F(t,x,\omega)$ is left continuous.

In order to ensure existence of the Lévy stochastic integral we must first define two spaces, firstly $\mathcal{P}_2(T,E)$ the space of all predictable mappings $F : [0,T] \times E \times \Omega \to \mathbb{R}^d$ which satisfy

$$P\left[\int_0^T \int_E |F(t,x)|^2 \nu(dy) dt < \infty\right] = 1$$

and secondly, $\mathcal{P}_2(T)$ the space of all predictable mappings $F : [0,T] \times \Omega \to \mathbb{R}^d$ satisfying,

$$P\left[\int_0^T |F(t)|^2 < \infty\right] = 1.$$

1.5.1 Lévy Type Stochastic Integrals

Take $E = \hat{B} - \{0\}$. We define an $\mathbb{R}^d$ valued stochastic process $Y = (Y(t), t \geq 0)$ to be a Lévy type stochastic integral if it is of the following form,

$$\begin{aligned} Y(t) = Y_0 &+ \int_0^t G(s)ds + \int_0^t F(s)dB(s) \\ &+ \int_0^t \int_{|x|<1} H(s,x)\tilde{N}(ds,dx) \\ &+ \int_0^t \int_{|x|\geq 1} K(s,x)N(ds,dx), \end{aligned} \tag{24}$$

for $1 \le i \le d$, $1 \le j \le m$, $t \ge 0$ and $|G^i|^{\frac{1}{2}}, F^i_j \in \mathcal{P}_2(T)$, $H^i \in \mathcal{P}_2(T, E)$ and K^i predictable. The construction of these integrals can be found in [3, p.229-232].

1.5.2 Itô's Formula

Itô's formula for a Lévy type stochastic differential equation as introduced below is both an extension of the well known Brownian case, and a special case of the more general semimartingale framework detailed in [50, p.271]. So we simply state, Itô's lemma for Lévy processes and refer the reader to [3, p.251-p.256] for further details.

From now on we will impose the following assumption of local boundedness on the small jumps

$$P\left(\sup_{0\le s\le t}\sup_{0<|x|<1}|H(s,x)| < \infty\right) = 1,$$

this is required in the the proof of Itô's lemma for Lévy processes see [3, p.252] for details.

In order to introduce Itô's lemma and product formula we firstly need to define the quadratic variation process of a Lévy type process $[Y, Y] = ([Y, Y](t), t \ge 0)$ for the (i, j)th entry $(1 \le i, j \le d)$ of $(Y(t), t \ge 0)$ of the form (24).

$$\begin{aligned}[Y^i, Y^j](t) = &\sum_{k=1}^{m}\int_0^T F^i_k(s)F^j_k(s)ds + \int_0^t\int_{|x|<1} H^i(s,x)H^j(s,x)N(ds,dx) \\ &+ \int_0^t\int_{|x|\ge 1} K^i(s,x)K^j(s,x)N(ds,dx)\end{aligned}$$

see [3, p.257] for details. In the following we will use Y_c to denote the *continuous part* of Y i.e.

$$Y^i_c(t) = \int_0^t G^i(s)ds + \sum_{j=1}^{m}\int_0^t F^i_j(s)dB_j(s),$$

for each $t \ge 0$, $1 \le i \le d$.

Theorem 1.5.1. (Itô's lemma for Lévy Processes)
If Y is a Lévy-type stochastic integral of the form (24), for each $f \in C^2(\mathbb{R}^d), t \ge 0$, with probability 1 we have;

$$
\begin{aligned}
f\big(Y(t)\big) - f\big(Y(0)\big) &= \int_0^t \partial_i f\big(Y(s-)\big) dY_c^i(s) \\
&+ \frac{1}{2}\int_0^t \partial_i\partial_j f\big(Y(s-)\big) d[Y_c^i, Y_c^i](s) \\
&+ \int_0^t \int_{|x|\geq 1} \Big[f\big(Y(s-)+K(s,x)\big) - f\big(Y(s-)\big)\Big] N(ds,dx) \\
&+ \int_0^t \int_{|x|<1} \Big[f\big(Y(s-)+H(s,x)\big) - f\big(Y(s-)\big)\Big] \tilde{N}(ds,dx) \\
&+ \int_0^t \int_{|x|<1} \Big[f\big(Y(s-)+H(s,x)\big) - f\big(Y(s-)\big) \\
&- H^i(s,x)\partial_i f\big(Y(s-)\big)\Big] \nu(dx)ds \qquad (25)
\end{aligned}
$$

□

Theorem 1.5.2. (Itô's Product Formula)

If Y_1 and Y_2 are real valued Lévy stochastic integrals of the form (24) then with probability 1 we have,

$$
\begin{aligned}
Y_1(t)Y_2(t) = Y_1(0)Y_2(0) &+ \int_0^t Y_1(s-)dY_2(s) \\
&+ \int_0^t Y_2(s-)dY_1(s) + [Y_1, Y_2](t). \qquad (26)
\end{aligned}
$$

□

For proof see [3, Theorem 4.4.13].

1.5.3 The Space $\mathcal{H}_2(T)$

We require the following norms, firstly the operator norm $||F||_{OP}$ for $F \in \mathcal{M}_{n,m}(\mathbb{R})$ defined by,

$$||F||_{OP} = \sup\left\{||Fv||_{\mathbb{R}^n} : v \in \mathbb{R}^m, ||v||_{\mathbb{R}^m} = 1\right\},$$

where $||\cdot||_{\mathbb{R}^n}$ and $||\cdot||_{\mathbb{R}^m}$ are the Euclidean norms in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Secondly the Hilbert Schmidt norm $||F||_{HS}$ defined by,

$$||F||_{HS}^2 = \mathrm{tr}(FF^T).$$

The following useful inequality is well known when $m = n$. We include the proof for the more general case.

Proposition 1.5.3. *For $A \in \mathcal{M}_{n,m}(\mathbb{R})$ and $B \in \mathcal{M}_{m,r}(\mathbb{R})$,*

$$||AB||_{HS} \leq ||A||_{OP}||B||_{HS} \tag{27}$$

Proof. Choose an orthonormal basis $\{e_i; 1 \leq i \leq m\}$ of $\mathbb{R}^m$. Using the property that the trace is invariant under cyclic permutations (see [46, Proposition 4.36] for proof) we have,

$$\begin{aligned} ||AB||_{HS}^2 &= \text{tr}(ABB^T A^T) \\ &= \text{tr}(A^T ABB^T) \\ &= \sum_{i=1}^{n} ||A^T ABB^T e_i|| \\ &\leq ||A^T A||_{OP}||B||_{HS}^2 \end{aligned}$$

and the result follows. □

Using a similar argument we can show that,

$$||AB||_{HS} \leq ||A||_{HS}||B||_{OP}. \tag{28}$$

Let F be a measurable mapping from $[0, T]$ to $\mathcal{M}_{n,m}(\mathbb{R})$. Let $\mathcal{H}_2(T)$ be the space of all such mappings which satisfy,

$$\int_0^T \left[||F(s)\rho^{1/2}||_{HS}^2\right] ds = \int_0^T \text{tr}\left(F(s)\rho F(s)^T\right) ds < \infty,$$

for some non-negative definite symmetric matrix $\rho \in \mathcal{M}_{m,m}(\mathbb{R})$. Then $\mathcal{H}_2(T)$ is a Hilbert space with inner product,

$$\int_0^t \left\langle F(s)\rho^{1/2}, G(s)\rho^{1/2}\right\rangle_{HS} ds = \int_0^T \text{tr}\left(F(s)\rho G(s)^T\right) ds,$$

for $F, G \in \mathcal{H}_2(T)$.

Lemma 1.5.4. *Let F be a left continuous mapping from $[0, T] \to \mathcal{M}_{n,m}(\mathbb{R})$ and $X = (X(t), t \geq 0)$ be an $\mathbb{R}^m$ valued square integrable mean zero Lévy process with covariance $\rho = A + \int_{\mathbb{R}^m - \{0\}} xx^T \nu(dx)$ as in (23). Then,*

$$\mathbb{E}\left[\left|\int_0^T F(t)dX(t)\right|^2\right] = \int_0^T ||F(s)\rho^{1/2}||_{HS}^2 ds.$$

Proof.

$$\mathbb{E}\left[\left|\int_0^T F(t)dX(t)\right|^2\right] = \mathbb{E}\left[\left(\int_0^T F(t)dX(t)\right)^T \int_0^T F(t)dX(t)\right]$$
$$= \sum_{i,j,k} \mathbb{E}\left[\int_0^T dX_j(t)F_{j,i}(t) \int_0^T F_{i,k}(t)dX_k(t)\right] \quad (29)$$

and so by Itô's isometry (see [3, p.271]),

$$\mathbb{E}\left[\left|\int_0^T F(t)dX(t)\right|^2\right] = \sum_{i,j,k} \mathbb{E}\left[\int_0^T F_{i,k}(t)dX_k(t) \int_0^T dX_j(t)F_{j,i}(t)\right]$$
$$= \sum_{i,j,k} \left[\int_0^T F_{i,k}(t)\rho_{k,j}F_{j,i}(t)\right]$$
$$= \int_0^T \left[\mathrm{tr}\left(F(t)\rho F(t)^T\right)\right] dt$$
$$= \int_0^T \left[||F(s)\rho^{1/2}||^2_{HS}\right] ds$$

□

We can see that the above is a version of Itô's isometry between the spaces $L^2(\Omega, \mathcal{F}, P; \mathbb{R}^m)$ and $\mathcal{H}_2(T)$. For the extension to the case where $\mathbb{R}^m$ is generalised to a separable Hilbert space see [2].

1.5.4 Existence and Uniqueness

We finish this chapter by looking at systems of Lévy stochastic differential equations, focusing on the following form;

$$dY^i(t) = b^i(Y(t-))dt + \sum_{j=1}^n \sigma^i_j(Y(t-))dB_j(t) + \int_{|x|<1} F^i(Y(t-),x)\tilde{N}(dt,dx)$$
$$+ \int_{|x|\geq 1} G^i(Y(t-),x)N(dt,dx), \quad (30)$$

with initial condition $Y(0) = Y_0$ (a.s.), where Y_0 is a given $\mathbb{R}^d$ valued $\mathcal{F}_0$ measurable random variable. The mappings $b^i : \mathbb{R}^d \to \mathbb{R}$, $\sigma^i_j : \mathbb{R}^d \to \mathbb{R}$, $F^i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $G^i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ are all assumed to be measurable. The solution to (30) when it exists, is an $\mathbb{R}^d$-valued adapted stochastic process

$(Y(t), t \geq 0)$. The existence and uniqueness of the solution to (30) is described by the next theorem. We will require the following conditions which make use of the matrix seminorm on $d \times d$ matrices

$$||a|| = \sum_{i=1}^{d} |a_i^i|.$$

(C1) **Lipschitz Condition** There exists $K_1 > 0$ such that for all $y_1, y_2 \in \mathbb{R}^d$,

$$|b(y_1) - b(y_2)|^2 + ||a(y_1, y_1) - 2a(y_1, y_2) + a(y_2, y_2)|| \\ + \int_{|x|<1} |F(y_1, x) - F(y_2, x)|^2 \nu(dx) \leq K_1 |y_1 - y_2|^2, \tag{31}$$

where $a(y_1, y_2) = \sigma(y_1)\sigma(y_2)^T$ for each $y_1, y_2 \in \mathbb{R}^d$.

(C2) **Growth condition** There exists $K_2 > 0$ such that for all $y \in \mathbb{R}^d$

$$|b(y)|^2 + ||a(y, y)|| + \int_{|x|<1} |F(y, x)^2 \nu(dx) \leq K_2(1 + |y|^2) \tag{32}$$

(G) We assume that the mapping $y \to G(y, x)$ is continuous for all $|x| > 1$.

Theorem 1.5.5. *There exists a unique càdlàg adapted solution to (30).*

$\square$

See [3, Theorem 6.2.9] for the full proof.

2 The Filtering Problem

2.1 Background and Motivation

The discrete time filtering problem goes back as far as the work of Kolmogorov in 1939 and 1941 ([33] and [34]) and Krein in 1945 (see [35] and [36]). Meanwhile Wiener developed the continuous time case see [57] which discussed the optimal estimation in a dynamical system with the presence of noise. This was originally provided to defence engineers in 1942 but only declassified and published as a book in 1949. The Kalman filter was developed in the 60's, see [28] [29], and produces estimates of the true values of measurements by predicting a value, estimating the uncertainty of this predicted value, and computing a weighted average of the predicted value and the measured value. The most weight is given to the value with the least uncertainty.

The continuous time version of the Kalman filter is called the Kalman-Bucy filter. It was developed by Richard Bucy see [10] and is what we will be focussing on in this chapter. The Kalman filter and the Kalman - Bucy filter provide a computationally viable method for estimation of the filtering problem and they have had a massive impact on a wide host of applications; not exclusively consisting of radar detection, stock market analysis, aerospace engineering such as orbit determination, 3D modelling, satellite navigation and weather forecasting.

In most of the literature the Kalman-Bucy filter is derived using Brownian motion, or white noise processes (or in the case of [17] orthogonal increment processes) with unit variance. Therefore our first task will be to derive the Kalman-Bucy filter using Lévy processes of finite variance, which is a straightforward extension of existing methodology. We take a route similar to [47, Ch.6], in that we use an innovations process method to derive the filter. When using Brownian motion, we have the nice result that the innovations process is also a Brownian motion. In our case however the analogous result no longer holds true, but by looking into the properties of our innovations process we can show that it is an orthogonal increment process with càdlàg paths.

We then move on to the main focus of this chapter, we aim to answer the question: is there a Kalman-Bucy filter for a square integrable Lévy driven system with an infinite variance Lévy driven observation process? We achieve this by generalising the Kalman-Bucy filter by extending the noise in the observations process to an infinite variance Lévy process, all of which is new and original work. This is achieved by truncating the size of the jumps in the observation noise. We then take the Kalman-Bucy filter derived in the first half of this chapter and pass to the limit. This allows an L^1 linear estimate of the system. In this case the Riccati equation for the mean square error also linearises and as a result it can no longer be interpreted as a mean square error even though it is a limit of these terms.

Before we move on, we draw the readers attention to two papers. Firstly the paper by Ahn and Feldman [1]; this paper deals with a similar problem to ours, but it allows a non-linear optimal recursive filter for best measurable estimate of the system given an infinite variance observation process. In this they require that the system be driven by Gaussian noise, which they describe as a "significant limitation" in the closing paragraph of their paper. Secondly, the paper by Le Breton and Musiela [39] uses a very different approach to the one here, and they state that due to the limitations of their methodology, the filter only seems to work for α-stable Lévy processes. To summarise, the work here deals with more general noise types than those presented above, and does so in a finite dimensional framework.

2.2 Framework

For the rest of this thesis we work on the interval $[0, T]$ for some fixed $T > 0$.

We begin by formulating the general linear filtering problem within the context of Lévy processes. Suppose the state $Y(t) \in \mathbb{R}^d$ at time $t \in [0, T]$ of a system is given by the stochastic differential equation

$$dY(t) = A(t)Y(t)dt + B(t)dX_1(t), \tag{33}$$

where $X_1(t)$ is a zero mean Lévy process on $\mathbb{R}^p$, $B(t)$ is a Borel measurable, locally bounded left continuous $(d \times p)$ matrix valued function, $A(t)$ is a Borel measurable locally bounded left continuous $(d \times d)$ matrix valued function, and the initial condition Y_0 is a random variable on $\mathbb{R}^d$.

We assume that the observations $Z(t) \in \mathbb{R}^m$ of $Y(t)$ are performed continuously and are of the form

$$dZ(t) = C(t)Y(t)dt + D(t)dX_2(t), \tag{34}$$

where $X_2(t)$ is a zero mean Lévy process on $\mathbb{R}^r$ independent of $X_1(t)$. $D(t)$ is a Borel measurable, locally bounded left continuous $(m \times r)$ matrix valued function such that $D(t)D(t)^T$ is invertible, and $C(t)$ is a Borel measurable locally bounded left continuous $(m \times d)$ matrix valued function.

Let $(V(t), 0 \leq t \leq T)$ be any stochastic process taking values in $\mathbb{R}^m$ such that $\mathbb{E}\left[|V(t)|^2\right] < \infty$ for all $0 \leq t \leq T$, then

$$\begin{aligned}&\mathcal{L}(V, T) = \text{The closure in } L^2(\Omega, \mathcal{F}, P; \mathbb{R}^d) \text{ of all linear combinations}\\ &c_0 + c_1V(t_1) + \cdots + c_kV(t_k); \quad 0 \leq t_i \leq T, c_0 \in \mathbb{R}^d, c_j \in \mathcal{M}_{d,m}(\mathbb{R}).\end{aligned}$$

It is straightforward yet worth noting that $\mathcal{L}(V,T_1) \subseteq \mathcal{L}(V,T_2)$ for $T_1 \leq T_2$, as $\mathcal{L}(V,T_2)$ will contain all linear combinations that were contained in $\mathcal{L}(V,T_1)$.

Let $\mathcal{P}_{\mathcal{L}}$ denote the projection from $L^2(\Omega,\mathcal{F},P;\mathbb{R}^m)$ onto $\mathcal{L}(Z,T)$. We can then define for each $0 \leq t \leq T$, $\hat{Y}(t)$ to be the projection of $Y(t)$ onto $\mathcal{L}(Z,T)$, i.e $\hat{Y}(t) = \mathcal{P}_{\mathcal{L}}(Y(t))$. From this we can see that $Y(t) - \hat{Y}(t) = (I - \mathcal{P}_{\mathcal{L}})Y(t)$, and so $Y(t) - \hat{Y}(t) \perp \mathcal{L}(Z,t)$.

The key question for us is: given the observations $Z(s)$ satisfying (34) for $0 \leq s \leq t$, what is the best estimate $\hat{Y}(t)$ of the state $Y(t)$ of the system (33) based on these observations?

Within the following steps, we follow [47, Ch.6] inserting Lévy processes and reworking proofs as necessary. The first step in [47] shows that in a Gaussian framework linear estimates are equivalent to measurable estimates. We are unable to show this holds once Gaussianity is removed, and so we will continue with just linear estimates.

From here on, until otherwise stated, we will be working with square integrable zero mean Lévy processes. Therefore from Theorem 1.3.3 for $i = 1,2$, $X_i(t)$ is as follows;

$$X_i(t) = B_i(t) + \int_{\mathbb{R}_0} x\tilde{N}_i(t,dx) \tag{35}$$

with covariance matrix,

$$\rho_i = \mathbb{E}[X_i(1)X_i(1)^T] = A_i + \int_{|x|\neq 0} xx^T\nu_i(dx)$$

by (22). We will also need

$$\lambda_i = \mathrm{tr}(\rho_i) = \mathrm{tr}(A_i) + \int_{|x|\neq 0} |x|^2\nu_i(dx) \tag{36}$$

.

We begin by finding an explicit form for $(Y(t), 0 \leq t \leq T)$.

Lemma 2.2.1. *For all $t \geq 0$, $Y(t)$ takes the following form:*

$$Y(t) = \exp\left(\int_0^t A(u)du\right)Y_0 + \int_0^t \exp\left(\int_s^t A(u)du\right)B(s)dX_1(s). \tag{37}$$

Proof. For each $0 \leq t \leq T$ let

$$W(t) = \exp\left(-\int_0^t A(u)du\right)Y(t).$$

Therefore by (33),

$$\begin{aligned}
dW(t) &= -A(t)\exp\left(-\int_0^t A(u)du\right)Y(t)dt \\
&\quad + \exp\left(-\int_0^t A(u)du\right)\big[A(t)Y(t)dt + B(t)dX_1(t)\big] \\
&= \exp\left(-\int_0^t A(u)du\right)B(t)dX_1(t).
\end{aligned}$$

Integrating over $[0,t]$ we get

$$W(t) = W_0 + \int_0^t \exp\left(-\int_0^s A(u)du\right)B(s)dX_1(s),$$

and so

$$\begin{aligned}
Y(t) &= \exp\left(\int_0^t A(u)du\right)W(t) \\
&= \exp\left(\int_0^t A(u)du\right)Y_0 \\
&\quad + \exp\left(\int_0^t A(u)du\right)\left[\int_0^t \exp\left(-\int_0^s A(u)du\right)B(s)dX_1(s)\right] \\
&= \exp\left(\int_0^t A(u)du\right)Y_0 + \int_0^t \exp\left(\int_s^t A(u)du\right)B(s)dX_1(s).
\end{aligned}$$

□

More generally, if $0 \le r \le t$ by a similar argument we get,

$$Y(t) = \exp\left(\int_r^t A(u)du\right)Y(r) + \int_r^t \exp\left(\int_s^t A(u)du\right)B(s)dX_1(s) \quad (38)$$

The following is not necessary for the derivation of the Kalman-Bucy filter, however it is an interesting observation.

Corollary 2.2.2. *Let $Y'(t) = Y(t) - Y_0$, then $(Y'(t), t \ge 0)$ is an additive process.*

Proof. By (38), for all $0 \le t \le T$,

$$\begin{aligned}
Y'(t) &= \left[\exp\left(\int_0^t A(u)du\right) - 1\right]Y_0 + \int_0^t \exp\left(\int_s^t A(u)du\right)B(s)dX_1(s) \\
&= G(t) + \int_0^t F(u)dX_1(u)
\end{aligned}$$

where, $G(t) := \left(\exp\left(\int_0^t A(u)du\right) - 1\right) Y_0$ and $F(u) := \exp\left(\int_s^u A(r)dr\right) B(u)$.

We note that firstly $G(t)$ is deterministic, and secondly that $\int_0^t F(u)dX_1(u)$ is of the form given in [3, Lemma 4.3.12], from which it follows that $Y'(t) - Y'(s)$ is independent of $\mathcal{F}_s$. We can subsequently use [4, Lemma 2.1]to show that $\int_0^t F(s)dX(s)$ is stochastically continuous. So Y' has independent increments, and is stochastically continuous and since $Y'_0 = 0$, $(Y'(t), t \geq 0)$ is an additive process. □

2.3 The Innovation Process

Before introducing the innovation process it is necessary to establish a representation of the functions in the space $\mathcal{L}(Z,T)$. In order to derive this representation we require the following lemma which generalises [47, 6.2.11].

We require the following assumption.

$$\inf_{0 \leq t \leq T} \lambda_{\min}(t) > 0 \tag{39}$$

where $\lambda_{\min}(t)$ is the smallest eigenvalue of $D(t)\rho D(t)^T$.

Lemma 2.3.1. *Let $F \in \mathcal{H}_2(T)$, then*

$$A_4 \int_0^T ||F(t)||_{HS}^2 dt \leq \mathbb{E}\left[\left|\left|\int_0^T F(t)dZ(t)\right|\right|_{HS}^2\right] \leq A_3 \int_0^T ||F(t)||_{HS}^2 dt \tag{40}$$

for some $A_3, A_4 \geq 0$.

Proof. For all $0 \leq t \leq T$,

$$\begin{aligned}
\mathbb{E}\left[\left|\left|\int_0^T F(t)dZ(t)\right|\right|_{HS}^2\right] &= \mathbb{E}\left[\left|\left|\int_0^T F(t)C(t)Y(t)dt\right|\right|_{HS}^2\right] \\
&+ \mathbb{E}\left[\left|\left|\int_0^T F(t)D(t)dX_2(t)\right|\right|_{HS}^2\right] \\
&+ 2\mathbb{E}\left[\left\langle\int_0^T F(t)C(t)Y(t)dt, \int_0^T F(t)D(t)dX_2(t)\right\rangle_{HS}\right]
\end{aligned} \tag{41}$$

Then using the Cauchy-Schwarz inequality, Fubini's theorem and Proposition 1.5.3

$$
\begin{aligned}
\mathbb{E}\left[\left|\left|\int_0^T F(t)C(t)Y(t)dt\right|\right|_{HS}^2\right] &\leq \mathbb{E}\left[\left(\int_0^T ||F(t)C(t)Y(t)||_{HS}\, dt\right)^2\right] \\
&\leq \mathbb{E}\left[\left(\int_0^T ||F(t)||_{HS}||C(t)Y(t)||_{OP} dt\right)^2\right] \\
&\leq \int_0^T ||F(t)||_{HS}^2 dt \mathbb{E}\left[\int_0^T ||C(t)Y(t)||_{OP}^2 dt\right] \\
&\leq C_1 \int_0^T ||F(t)||_{HS}^2 dt \int_0^T \mathbb{E}\left[||Y(t)||_{HS}^2\right] dt \\
&= A_1 \int_0^T ||F(t)||_{HS}^2 dt, \qquad (42)
\end{aligned}
$$

where $C_1 = \sup_{0\leq t\leq T} ||C(t)||_{OP}^2$ and $A_1 = \sup_{0\leq t\leq T} ||C(t)||_{OP}^2 \int_0^T \mathbb{E}\left[||Y(t)||_{HS}^2\right] dt$. This can be seen to be finite by,

$$
\begin{aligned}
\int_0^T \mathbb{E}\left[||Y(t)||_{HS}^2\right] dt &= \int_0^T \mathbb{E}\left[\left|\left|\exp\left(\int_0^t A(u)du\right) Y_0 \right.\right.\right. \\
&\quad \left.\left.\left. + \int_0^t \exp\left(\int_s^t A(u)du\right) B(s)dX_1(s)\right|\right|_{HS}^2\right] dt \\
&\leq \int_0^T \mathbb{E}\left[\left|\left|\exp\left(\int_0^t A(u)du\right) Y_0\right|\right|_{HS}^2\right] dt \\
&\quad + \int_0^T \mathbb{E}\left[\left|\left|\int_0^t \exp\left(\int_s^t A(u)du\right) B(s)dX_1(s)\right|\right|_{HS}^2\right] dt \\
&= \int_0^T \mathbb{E}\left[\left|\left|\exp\left(\int_0^t A(u)du\right) Y_0\right|\right|_{HS}^2\right] dt \\
&\quad + \int_0^T \mathbb{E}\left[\left|\left|\int_0^t \exp\left(\int_s^t A(u)du\right) B(s)\rho_1 ds\right|\right|_{HS}^2\right] dt \\
&< \infty. \qquad (43)
\end{aligned}
$$

By independence and (38),

$$\begin{aligned}
&\mathbb{E}\left[\left\langle \int_0^T F(t)C(t)Y(t)dt, \int_0^T F(t)D(t)dX_2(t)\right\rangle_{HS}\right] \\
&\quad = \left\langle \int_0^T F(t)C(t)Y(t)dt, \mathbb{E}\left[\int_0^T F(t)D(t)dX_2(t)\right]\right\rangle_{HS} \\
&\quad = 0. \qquad (44)
\end{aligned}$$

By Lemma 1.5.4 and (27),

$$\begin{aligned}
\mathbb{E}\left[\left|\left|\int_0^T F(t)D(t)dX_2(t)\right|\right|_{HS}^2\right] &= \int_0^T \left|\left|F(t)D(t)\rho_2^{\frac{1}{2}}\right|\right|_{HS}^2 dt \\
&\leq \int_0^T ||F(t)||_{HS}^2 ||D(t)\rho_2^{\frac{1}{2}}||_{OP}^2 dt \\
&\leq ||\rho_2||_{OP} \sup_{0\leq t\leq T} ||D(t)||_{OP}^2 \int_0^T ||F(t)||_{HS}^2 dt \\
&= A_2 \int_0^T ||F(t)||_{HS}^2 dt \qquad (45)
\end{aligned}$$

where $A_2 = ||\rho_2||_{OP}^2 \sup_{0\leq t\leq T} ||D(t)||_{OP}^2$.

We now obtain a lower bound, by (41) and [58, Theorem 1],

$$\begin{aligned}
\mathbb{E}\left[\left|\left|\int_0^T F(t)dZ(t)\right|\right|_{HS}^2\right] &\geq \mathbb{E}\left[\left|\left|\int_0^T F(t)D(t)dX_2(t)\right|\right|_{HS}^2\right] \\
&= \int_0^T \left|\left|F(t)D(t)\rho_2^{1/2}\right|\right|_{HS}^2 dt \\
&= \int_0^T \mathrm{tr}\left(F(t)D(t)\rho_2^{1/2}\rho_2^{1/2}D(t)^T F(t)^T\right) dt \\
&= \int_0^T \mathrm{tr}\left(D(t)\rho_2 D(t)^T F(t)^T F(t)\right) dt \\
&\geq \int_0^T \lambda_{\min}(t)\mathrm{tr}\left(F(t)^T F(t)\right) dt \\
&\geq \inf_{0\leq t\leq T} \lambda_{\min}(t) \int_0^T ||F(t)||_{HS}^2 dt.
\end{aligned}$$

We now have,

$$A_4 \int_0^T ||F(t)||_{HS}^2 dt \leq \mathbb{E}\left[\left|\left|\int_0^T F(t)dZ(t)\right|\right|_{HS}^2\right] \leq A_3 \int_0^T ||F(t)||_{HS}^2 dt$$

for $A_1, A_2, A_3, A_4 > 0$ where,

- $A_3 = A_1 + A_2$
- $A_4 = \inf_{-\leq t \leq T} \lambda_{\min}(t) > 0$ by (39).

$\square$

We now establish a representation of the functions in $\mathcal{L}(Z,T)$. The following lemma is a generalisation of [47, Lemma 6.2.4].

Lemma 2.3.2. $\mathcal{L}(Z,T) = \left\{ c_0 + \int_0^T F(t)dZ(t) \middle| F \in \mathcal{H}_2(T), c_0 \in \mathbb{R}^d \right\}$.

Proof. Denote the right hand side by $\mathcal{N}(Z,T)$. Then we wish to show the following,

a) $\mathcal{N}(Z,T) \subset \mathcal{L}(Z,T)$

b) $\mathcal{N}(Z,T)$ contains all linear combinations of the form

$$c_0 + c_1 Z(t_1) + \cdots + c_k Z(t_k); \qquad 0 \leq t_i \leq T$$

c) $\mathcal{N}$ is closed in $L^2(\Omega, \mathcal{F}, P; \mathbb{R}^d)$,

from which the result follows.

a): If F is continuous, then using dyadic intervals and limits in $L^2(\Omega, \mathcal{F}, P; \mathbb{R}^d)$, $\int_0^T F(t)dZ(t)$ is of the form $c_1 Z(t_1) + \cdots + c_k Z(t_k)$.
as shown in [47, Lemma 6.2.4]. Note that if F is not continuous we can approximate using simple functions.

b): Suppose $0 = t_0 \leq t_1 < \cdots < t_k \leq t_{k+1} = T$, $c_i := c'_{i-1} - c'_i$, $c_0 = 0$ for convenience, $\Delta Z(j) := Z(t_{j+1}) - Z(t_j)$ then,

$$\begin{aligned}
\sum_{i=1}^{k} c_i Z(t_i) &= \sum_{j=0}^{k-1} c'_j \Delta Z(j) \\
&= \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} c'_j dZ(t) \\
&= \int_0^T \Big(\sum_{j=0}^{k-1} 1_{(t_j, t_{j+1}]}(t) c'_j \Big) dZ(t).
\end{aligned}$$

c): Let $\left(\int_0^T F_n(t)dZ(t), n \in \mathbb{N}\right)$ be a sequence which converges in $L^2(\Omega, \mathcal{F}, P; \mathbb{R}^d)$ to say $Q(T)$. From Lemma 2.3.1,

$$A_4 \int_0^T ||F_n(t) - F_m(t)||_{HS}^2 dt \leq \mathbb{E}\left[\left|\left|\int_0^T \big(F_n(t) - F_m(t)\big)dZ(t)\right|\right|_{HS}^2\right] \leq A_3 \int_0^T ||F_n(t) - F_m(t)||_{HS}^2 dt. \tag{46}$$

It follows that $(F_n, n \in \mathbb{N})$ is Cauchy in $\mathcal{H}_2(T)$ and so converges to say, F. By taking limits as $n \to \infty$

$$A_4 \int_0^T ||F_n(t) - F(t)||_{HS}^2 dt \leq \mathbb{E}\left[\left|\left|\int_0^T F_n(t)dZ(t) - Q(T)\right|\right|_{HS}^2\right] \leq A_3 \int_0^T ||F_n(t) - F(t)||_{HS}^2 dt.$$

We also know that,

$$A_4 \int_0^T ||F_n(t) - F(t)||_{HS}^2 dt \leq \mathbb{E}\left[\left|\left|\int_0^T \Big(F_n(t) - F(t)\Big)dZ(t)\right|\right|_{HS}^2\right] \leq A_3 \int_0^T ||F_n(t) - F(t)||_{HS}^2 dt. \tag{47}$$

Therefore from the uniqueness of limits in $L^2(\Omega, \mathcal{F}, P; \mathbb{R}^d)$

$$Q(T) = \int_0^T F(t)dZ(t),$$

therefore $\mathcal{N}(Z, T)$ is closed.

□

We now define the innovation process $\big(M(t), t \geq 0\big)$, where $M(t) \in \mathbb{R}^m$ is as follows:

$$M(t) = Z(t) - \int_0^t C(s)\hat{Y}(s)ds,$$

and so

$$dM(t) = C(t)(Y(t) - \hat{Y}(t))dt + D(t)dX_2(t). \tag{48}$$

Lemma 2.3.3. *a)* $\mathcal{L}(M, t) = \mathcal{L}(Z, t)\ \forall\ t \geq 0$

b) $\big(M(t), t \geq 0\big)$ *has orthogonal increments.*

Proof. a): $\mathcal{L}(M,t) \subset \mathcal{L}(Z,t)$ since,

$$\begin{aligned}
c_0 + &c_1 M(t_1) + \cdots + c_k M(t_k) \\
&= c_0 + c_1 \left(Z(t_1) - \int_0^{t_1} C(s)\hat{Y}(s)ds \right) \\
&\quad + \cdots + c_k \left(Z(t_k) - \int_0^{t_k} C(s)\hat{Y}(s)ds \right) \\
&= c_0 + c_1 Z(t_1) + \cdots + c_k Z(t_k) \\
&\quad - \sum_{i=1}^{k} \int_0^{t_i} C(s)\hat{Y}(s)ds \\
&\in \mathcal{L}(Z,t),
\end{aligned}$$

$\int_0^{t_i} C(s)\hat{Y}(s)ds \in \mathcal{L}(Z,t)$ because,

$$\begin{aligned}
\int_0^t C(s)\hat{Y}(s)ds &= \int_0^t C(s)\mathcal{P}_{\mathcal{L}}(Y(s))ds \\
&= \mathcal{P}_{\mathcal{L}} \left(\int_0^t C(s)Y(s)ds \right)
\end{aligned}$$

by the continuity of $\mathcal{P}_{\mathcal{L}}$.

To establish the reverse inclusion, see [47, p92-93].

b): If $s < t$ consider $K \in \mathcal{L}(Z,s)$. By Fubini's theorem,

$$\begin{aligned}
&\mathbb{E}\big[(M(t) - M(s)) \cdot K\big] \\
&= \mathbb{E}\left[\left(\int_s^t C(r)(Y(u) - \hat{Y}(u))du + \int_s^t D(u)dX_2(u) \right) \cdot K \right] \\
&= \int_s^t C(t)\mathbb{E}[(Y(u) - \hat{Y}(u)) \cdot K]du + \mathbb{E}\left[\left(\int_s^t D(u)dX_2(u) \right) \cdot K \right] \\
&= 0,
\end{aligned}$$

since $Y(u) - \hat{Y}(u) \perp \mathcal{L}(Z,u) \supset \mathcal{L}(Z,s)$ for $u \geq s$ and the expectation of the stochastic integral is zero.

$\square$

Recall that $D(t)D(t)^T$ was assumed to be invertible and bounded away from 0 on bounded intervals. We will find it convenient as in [17, p.136] to define

$$G(t) = \left(D(t)D(t)^T\right)^{-\frac{1}{2}}$$

i.e. $G(t)$ is any $m \times m$ matrix such that $G(t)^T G(t) = G(t)G(t)^T = \left(D(t)D(t)^T\right)^{-1}$.

Define the process $R(t) \in \mathbb{R}^m$ by,

$$dR(t) = G(t)dM(t). \tag{49}$$

The next lemma will be of use when finding the first two moments of $R(t)$.

Lemma 2.3.4.

$$\mathbb{E}[\hat{Y}(t)] = \mathbb{E}[Y(t)],$$

for all $0 \leq t \leq T$.

Proof. $\mathbb{E}[Y(t)] = \langle Y(t), 1\rangle$ thus,

$$\begin{aligned}\mathbb{E}[\hat{Y}(t)] &= \langle \hat{Y}(t), 1\rangle = \langle P_{\mathcal{L}}(Y(t)), 1\rangle \\ &= \langle Y(t), P_{\mathcal{L}}(1)\rangle = \langle Y(t), 1\rangle = \mathbb{E}[Y(t)].\end{aligned}$$

□

We now have a look at some of the properties of the process R. In a scenario where the system and observation processes are both driven by Brownian motion, it can be shown that R is a Brownian motion. We will show that given a Lévy setup we have an orthogonal increment process with càdlàg paths.

Lemma 2.3.5. $(R(t), 0 \leq t \leq T)$ *has the following properties,*

a) R *has càdlàg paths,*

b) R *has orthogonal increments,*

c) $\mathbb{E}[R(t)] = 0$ *and* $\mathbb{E}[R(t)R(s)^T] = \Sigma_2(s)$ *if* $s < t$ *where*

$$\Sigma_2(t) = \int_0^t G(s)D(s)\rho_2 D(s)^T G(s)^T ds$$

Proof. a): This follows since $M(t)$ has càdlàg paths, see [3, p.140].

b): This follows since $M(t)$ has orthogonal increments.

c): Using Fubini's theorem and Lemma 2.3.4

$$
\begin{aligned}
\mathbb{E}[R(t)] &= \mathbb{E}\left[\int_0^t G(s)\left(C(s)(Y(s)-\hat{Y}(s))ds + D(s)dX_2(s)\right)\right] \\
&= \int_0^t \left[G(s)C(s)\mathbb{E}(Y(s)-\hat{Y}(s))\right] ds - \mathbb{E}\left[\int_0^t G(s)D(s)dX_2(s)\right] \\
&= 0,
\end{aligned}
$$

By orthogonal increments and the fact that $\mathbb{E}[R(s)] = 0$ for all s, for $s < t$ we have,

$$
\begin{aligned}
\mathbb{E}[R(t)R(s)^T] &= \mathbb{E}[(R(t)-R(s))R(s)^T] + \mathbb{E}[R(s)R(s)^T] \\
&= \mathbb{E}[R(t)-R(s)]\mathbb{E}[R(s)^T] + \mathbb{E}[R(s)R(s)^T] \\
&= \mathbb{E}[R(s)R(s)^T]. \qquad (50)
\end{aligned}
$$

So by using Itô's product formula,

$$
\begin{aligned}
d[R(t)R(t)^T] &= R(t)dR(t)^T + dR(t)R(t)^T + dR(t)dR(t)^T \\
&= R(t)(G(t)dM(t))^T + G(t)dM(t)R(t)^T \\
&\quad + G(t)D(t)\left[\sigma_2\sigma_2^T dt + \left(\int_{\mathbb{R}_0} xx^T N(dt,dx)\right)\right] D(t)^T G(t)^T.
\end{aligned}
$$

Taking expectations and integrating gives,

$$
\begin{aligned}
\mathbb{E}[R(t)R(t)^T] &= R_0R_0^T + \mathbb{E}\left[\int_0^t G(s)D(s)\rho_2 D(s)^T G(s)^T dt\right] \\
&= R_0R_0^T + \int_0^t G(s)D(s)\rho_2 D(s)^T G(s)^T ds \\
&= \Sigma_2(t), \qquad (51)
\end{aligned}
$$

where $R_0 = 0$ is clear by definition. Therefore by (50)

$$
\mathbb{E}[R(t)R(s)^T] = \Sigma_2(s)
$$

for $s < t$.

□

We will require the following result later in this chapter.

Corollary 2.3.6. *Recall the definition of λ_2 from (36) then,*

$$
\mathbb{E}\left[R(t)^T R(t)\right] = \lambda_2 t
$$

Proof.

$$\begin{aligned}
d(R(t)^T R(t)) &= dR(t)^T R(t) + R(t)^T dR(t) + dR(t)^T dR(t) \\
&= (G(t)dM(t))^T R(t) + R(t)^T G(t)dM(t) \\
&\quad + dM(t)^T G(t)^T G(t)dM(t) \\
&= (G(t)dM(t))^T R(t) + R(t)^T G(t)dM(t) \\
&\quad + dX_2(t)^T D(t)^T G(t)^T G(t)D(t)dX_2(t) \\
&= (G(t)dM(t))^T R(t) + R(t)^T G(t)dM(t) \\
&\quad + \operatorname{tr}(A)dt + \int_{\mathbb{R}_0} x^T x N(dt, dx).
\end{aligned}$$

Integrating, using (35) and taking expectations yields,

$$\begin{aligned}
\mathbb{E}\left[R(t)^T R(t)\right] &= R_0^T R_0 + \operatorname{tr}(A)t + \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} x^T x N(dt, dx)\right] \\
&= \operatorname{tr}(A)t + t\int_{\mathbb{R}_0} |x|^2 \nu(dx) \\
&= \lambda_2 t. \qquad (52)
\end{aligned}$$

□

We now wish to replicate Lemma 2.3.3 (a) for the process R.

Lemma 2.3.7.

$$\mathcal{L}(M,t) = \mathcal{L}(R,t) \qquad (53)$$

Proof.

$$\begin{aligned}
\mathcal{L}(M,t) &= \left\{c_0 + \int_0^T F(t)dM(t) \middle| F \in \mathcal{H}_2(T), c_0 \in \mathbb{R}^d\right\} \qquad (54) \\
&= \left\{c_0 + \int_0^T F(t)G(t)^{-1}dR(t) \middle| F \in \mathcal{H}_2(T), c_0 \in \mathbb{R}^d\right\} \\
&\subseteq \left\{c_0 + \int_0^T H(t)dR(t) \middle| H \in \mathcal{H}_2(T), c_0 \in \mathbb{R}^d\right\} \\
&= \mathcal{L}(R,t).
\end{aligned}$$

We can see $H \in \mathcal{H}_2(T)$ since,

$$\begin{aligned}
\mathbb{E}\left[\left|\int_0^T F(t)G(t)^{-1}(t)dR(t)\right|^2\right] &= \int_0^t \mathbb{E}\left[\left|F(t)G(t)^{-1}\right|^2\right]\Sigma_2(t)dt \\
&\leq \sup_{0\leq t\leq T} |D(t)|^2 \int_0^T \mathbb{E}[|F(t)|^2]\Sigma_2(t)dt \\
&\leq K_T \mathbb{E}\left[\int_0^t |F(t)|dR(t)\right]^2
\end{aligned}$$

The reverse inclusion follows the same argument. □

We now find a representation of $(\hat{Y}(t), 0 \leq t \leq T)$, which is a generalisation of [47, Lemma 6.2.7], in that it is done in finite dimensions for a square integrable Lévy process with a non unit variance.

Lemma 2.3.8. *For all* $0 \leq t \leq T$*,*

$$\hat{Y}(t) = \mathbb{E}[Y(t)] + \int_0^t \frac{\partial}{\partial s}\mathbb{E}[Y(t)R(s)^T]\Sigma_2(s)^{-1}dR(s)$$

Proof. From Lemma 2.3.2 and Lemma 2.3.7 and using Proposition 4.1.1 in [17] we have

$$\hat{Y}(t) = c_0(t) + \int_0^t J(s)dR(s) \quad \text{for some } J \in \mathcal{H}_2(T), c_0(t) \in \mathbb{R}^d.$$

Now if we take expectations we find that from Lemma 2.3.4 $c_0(t) = \mathbb{E}[\hat{Y}(t)] = \mathbb{E}[Y(t)]$, and from Lemma 2.3.3 and Lemma 2.3.2

$$Y(t) - \hat{Y}(t) \perp \int_0^t F(s)dR(s) \;\text{ for all } F \in \mathcal{H}_2(T),$$

as $\int_0^t F(s)dR(s) \in \mathcal{L}(R,t) = \mathcal{L}(Z,t) \perp (I - \mathcal{P}_{\mathcal{L}})Y(t) = Y(t) - \hat{Y}(t)$. Therefore using Itô's isometry, and Lemma 2.3.5 (c)

$$\begin{aligned}
\mathbb{E}\left[Y(t)\left(\int_0^t F(s)dR(s)\right)^T\right] &= \mathbb{E}\left[\hat{Y}(t)\left(\int_0^t F(s)dR(s)\right)^T\right] \\
&= \mathbb{E}\left[\int_0^t J(s)dR(s)\left(\int_0^t F(s)dR(s)\right)^T\right] \\
&= \int_0^t J(s)\Sigma_2(s)F(s)^T ds. \qquad (55)
\end{aligned}$$

Let $F = 1_{[0,r]}E_{i,j}$ for some $r \leq t$, where $(E_{i,j}, 1 \leq i \leq d, 1 \leq j \leq m)$ is the natural basis for $\mathcal{M}_{d,m}(\mathbb{R})$ defined by

$$E_{i,j} = \begin{cases} 1 & \text{in the } (i,j)^{th} \text{ entry of the matrix} \\ 0 & \text{everywhere else} \end{cases}$$

$$\int_0^t 1_{[0,r)} E_{i,j} dR(s) = \begin{pmatrix} 0 \\ \dots \\ R_j(r) \\ \dots \\ 0 \end{pmatrix}_{(i)}$$

where $R_j(r)$ is in the i^{th} row. Therefore

$$\mathbb{E}\left[Y(t)\left(\int_0^t F(s)dR(s)\right)^T\right] = \begin{pmatrix} 0 & \dots & \mathbb{E}\left[Y_1(t)R_j(r)\right] & \dots & 0 \\ \vdots & \dots & \mathbb{E}\left[Y_2(t)R_j(r)\right] & \vdots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & \mathbb{E}\left[Y_m(t)R_j(r)\right] & \dots & 0 \end{pmatrix}_{(i)} \tag{56}$$

Now,

$$\Sigma_2(s)E_{j,i} = \begin{pmatrix} 0 & \dots & \Sigma_2(s)_{1,j} & \dots & 0 \\ \vdots & \dots & \Sigma_2(s)_{2,j} & \vdots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & \Sigma_2(s)_{m,j} & \dots & 0 \end{pmatrix}_{(i)}.$$

In words, the j^{th} column of Σ_2 is moved to the i^{th} column of the resulting matrix. So we have that,

$$J(s)\Sigma_2(s)E_{j,i} = \begin{pmatrix} 0 & \dots & \sum_{k=1}^m J_{1,k}\Sigma_2(s)_{k,j} & \dots & 0 \\ \vdots & \dots & \sum_{k=1}^m J_{2,k}\Sigma_2(s)_{k,j} & \vdots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & \sum_{k=1}^m J_{n,k}\Sigma_2(s)_{k,j} & \dots & 0 \end{pmatrix}_{(i)}$$

hence, by (55) and (56), for all $1 \leq i \leq d$, $1 \leq j \leq m$, $\mathbb{E}\left[Y_i(t)R_j(r)\right] = \int_0^r \left(J(s)\Sigma_2(s)\right)_{i,j} ds$, and it follows that,

$$\mathbb{E}\left[Y(t)R(r)^T\right] = \int_0^r J(s)\Sigma_2(s)ds.$$

We can clearly see that the matrix valued function $\mathbb{E}[Y(t)R(r)^T]$ is differentiable component-wise with respect to r hence,

$$J(r) = \frac{\partial}{\partial r}\mathbb{E}[Y(t)R(r)^T]\Sigma_2(t)^{-1}.$$

$\square$

2.4 The Differential Equation for $\hat{Y}(t)$

We now make the representation of $\hat{Y}$ from Lemma 2.3.8 more precise.

Lemma 2.4.1. *For all* $0 \leq t \leq T$,

$$\begin{aligned}\hat{Y}(t) &= \mathbb{E}[Y(t)] \\ &\quad + \int_0^t \exp\left(\int_s^t A(u)du\right) S(s)C(s)^T G(s)^T \Sigma_2(s)^{-1} dR(s),\end{aligned} \tag{57}$$

where

$$S(s) = \mathbb{E}\left[\tilde{Y}(s)\tilde{Y}(s)^T\right] = \mathbb{E}\left[\left(Y(s) - \hat{Y}(s)\right)\left(Y(s) - \hat{Y}(s)\right)^T\right] \tag{58}$$

is the mean square error.

Proof. Let $f(s,t) = \dfrac{\partial}{\partial s}\mathbb{E}[Y(t)R(s)^T]\Sigma_2^{-1}(s)$ for $s \leq t$ then Lemma 2.3.8 states that,

$$\hat{Y}(t) = \mathbb{E}[Y(t)] + \int_0^t f(s,t)dR(s). \tag{59}$$

From (48) and (49) we have,

$$R(s) = \int_0^s G(r)C(r)\big(Y(r) - \hat{Y}(r)\big)dr + \int_0^s G(r)D(r)X_2(r)$$

and so we obtain by independence of Y and X_2,

$$\mathbb{E}[Y(t)R(s)^T] = \int_0^s \mathbb{E}[Y(t)\tilde{Y}(r)^T]C(t)^T G(t)^T dr \tag{60}$$

where $\tilde{Y}(t) = Y(t) - \hat{Y}(t)$ is the error between the system and the best estimate. Using (38) for $Y(t)$, and recalling that the expectation of the stochastic integral is zero, we obtain

$$\begin{aligned}\mathbb{E}[Y(t)\tilde{Y}(r)^T] &= \mathbb{E}\Bigg[\Bigg(\exp\left(\int_r^t A(u)du\right)Y(r) \\ &\qquad + \int_r^t \exp\left(\int_s^t A(u)du\right)B(s)dX_1(s)\Bigg)\tilde{Y}(r)^T\Bigg] \\ &= \exp\left(\int_r^t A(u)du\right)\mathbb{E}[Y(r)\tilde{Y}(r)^T] \\ &= \exp\left(\int_r^t A(u)du\right)S(r).\end{aligned} \tag{61}$$

Substituting (61) back into (60) we get

$$\mathbb{E}[Y(t)R(s)^T] = \int_0^s \exp\left(\int_r^t A(u)du\right) S(r)C(r)^T G(r)^T dr,$$

and so,

$$\frac{\partial}{\partial s}\mathbb{E}[Y(t)R(s)^T] = \exp\left(\int_s^t A(u)du\right) S(s)C(s)^T G(s)^T$$

therefore

$$f(s,t) = \exp\left(\int_s^t A(u)du\right) S(s)C(s)^T G(s)^T \Sigma_2(s)^{-1}. \tag{62}$$

and the result follows. □

We now show that the mean square error $(S(t), 0 \leq t \leq T)$ satisfies a deterministic Riccati equation.

Theorem 2.4.2. $(S(t), 0 \leq t \leq T)$ *satisfies the* Riccati equation*,*

$$\begin{aligned}\frac{dS(t)}{dt} &= A(t)S(t) + S(t)A(t)^T + B(t)\rho_1 B(t)^T \\ &\quad - \left(S(t)C(t)^T G(t)^T \Sigma_2(t)^{-1} G(t)C(t)S(t)^T\right).\end{aligned} \tag{63}$$

Proof. We can see that

$$\begin{aligned}&\mathbb{E}\left[\left(Y(t) - \hat{Y}(t)\right)\left(Y(t) - \hat{Y}(t)\right)^T\right] \\ &\qquad = \mathbb{E}[Y(t)Y(t)^T] - \mathbb{E}[\hat{Y}(t)\hat{Y}(t)^T],\end{aligned}$$

and so from (59) and Itô's isometry,

$$\begin{aligned}S(t) &= \mathbb{E}[Y(t)Y(t)^T] - \mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T - \int_0^t f(s,t)\Sigma_2(s)f(s,t)^T ds \\ &= T(t) - \int_0^t f(s,t)\Sigma_2(s)f(s,t)^T ds - \mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T\end{aligned} \tag{64}$$

where $T(t) = \mathbb{E}[Y(t)Y(t)^T]$. Now from (38)

$$\begin{aligned}T(t) &= \exp\left(\int_0^t A(u)du\right)\mathbb{E}[Y_0Y_0^T]\exp\left(\int_0^t A(u)^T du\right) \\ &\quad + \int_0^t \exp\left(\int_s^t A(u)du\right) B(s)\rho_1 B(s)^T \exp\left(\int_s^t A(u)^T du\right) ds.\end{aligned}$$

Differentiating gives

$$\begin{aligned}
\frac{dT(t)}{dt} &= A(t)\exp\left(\int_0^t A(u)du\right)\mathbb{E}[Y_0Y_0^T]\exp\left(\int_0^t A(u)^T du\right) \\
&\quad + \exp\left(\int_0^t A(u)du\right)\mathbb{E}[Y_0Y_0^T]\exp\left(\int_0^t A(u)^T du\right)A(t)^T + B(t)\rho_1 B(t)^T \\
&\quad + \int_0^t A(t)\exp\left(\int_s^t A(u)du\right)B(s)\rho_1 B(s)^T\exp\left(\int_s^t A(u)^T du\right)ds \\
&\quad + \int_0^t \exp\left(\int_s^t A(u)du\right)B(s)\rho_1 B(s)^T\exp\left(\int_s^t A(u)^T du\right)A(t)^T ds \\
&= B(t)\rho_1 B(t)^T + A(t)T(t) + T(t)A(t)^T. \qquad (65)
\end{aligned}$$

We also need to differentiate $\mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T$.

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T &= \frac{d}{dt}\mathbb{E}\left[\exp\left(\int_0^t A(u)du\right)Y_0Y_0^T\exp\left(\int_0^t A(u)^T du\right)\right] \\
&= A(t)\mathbb{E}\left[\exp\left(\int_0^t A(u)du\right)Y_0Y_0^T\exp\left(\int_0^t A(u)^T du\right)\right] \\
&\quad + \mathbb{E}\left[\exp\left(\int_0^t A(u)du\right)Y_0Y_0^T\exp\left(\int_0^t A(u)^T du\right)\right]A(t)^T \\
&= A(t)\mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T + \mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T A(t)^T \qquad (66)
\end{aligned}$$

Differentiating (64) and combining it with (65) and (66) we get

$$\begin{aligned}
\frac{dS(t)}{dt} &= \frac{dT(t)}{dt} - f(t,t)\Sigma_2(t)f(t,t)^T - \int_0^t f(s,t)\Sigma_2(t)\frac{\partial}{\partial t}f(s,t)^T ds \\
&\quad - \int_0^t \frac{\partial}{\partial t}f(s,t)\Sigma_2(t)f(s,t)^T \\
&\quad - A(t)\mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T - \mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T A(t)^T \\
&= B(t)\rho_1 B(t)^T + A(t)T(t) + T(t)A(t)^T \\
&\quad - \left(S(t)C(t)^T G(t)^T \Sigma_2(t)^{-1}\right)\Sigma_2(t)\left(\Sigma_2(t)^{-1}G(t)C(t)S(t)^T\right) \\
&\quad - A(t)\int_0^t f(s,t)\Sigma_2(s)f(s,t)^T ds - \int_0^t f(s,t)\Sigma_2(s)f(s,t)^T ds A(t)^T \\
&\quad - A(t)\mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T - \mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T A(t)^T \quad \text{by (65), (62) and (66)} \\
&= B(t)\rho_1 B(t)^T + A(t)T(t) + T(t)A(t)^T \\
&\quad - \left(S(t)C(t)^T G(t)^T \Sigma_2^{-1}(t)G(t)C(t)S(t)^T\right) \\
&\quad - A(t)\int_0^t f(s,t)\Sigma_2(s)f(s,t)^T ds - \int_0^t f(s,t)\Sigma_2(s)f(s,t)^T ds A(t)^T \\
&\quad - A(t)\mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T - \mathbb{E}[Y(t)]\mathbb{E}[Y(t)]^T A(t)^T \\
&= A(t)S(t) + S(t)A(t)^T + B(t)\rho_1 B(t)^T \\
&\quad - \left(S(t)C(t)^T G(t)^T \Sigma_2^{-1}(t)G(t)C(t)S(t)^T\right).
\end{aligned}$$

□

The last part is to find the SDE for $\hat{Y}(t)$.

Theorem 2.4.3. *The solution $\hat{Y}(t) = P_{\mathcal{L}}(Y(t))$ of the finite-dimensional linear filtering problem satisfies the stochastic differential equation,*

$$\begin{aligned}
d\hat{Y}(t) &= A(t)\hat{Y}(t)dt \\
&\qquad + S(t)C(t)^T G(t)^T \Sigma_2(t)^{-1} G(t)\left[dZ(t) - C(t)\hat{Y}(t)dt\right]
\end{aligned} \tag{67}$$

with initial condition $\hat{Y}_0 = \mathbb{E}[Y_0]$ and,

$$S(0) = \mathbb{E}\left[(Y_0 - \mathbb{E}[Y_0])(Y_0 - \mathbb{E}[Y_0])^T\right].$$

Proof. From (59),

$$\hat{Y}(t) = c_0(t) + \int_0^t f(s,t)dR(s) \qquad \text{where} \quad c_0(t) = \mathbb{E}[Y(t)].$$

It follows that

$$d\hat{Y}(t) = c_0'(t)dt + f(t,t)dR(t) + \left(\int_0^t \frac{\partial}{\partial t} f(s,t)dR(s)\right) dt$$

and so by (62) and since $c_0'(t) = \dfrac{d}{dt}\mathbb{E}[Y(t)] = \dfrac{d}{dt}\mathbb{E}\left[\exp\left(\int_0^t A(u)du\right) Y_0\right] = A(t)\mathbb{E}[Y_t] = A(t)c_0(t)$, we obtain,

$$\begin{aligned}
d\hat{Y}(t) &= c_0'(t)dt + S(t)C(t)^T G(t)^T \Sigma_2(t)^{-1} dR(t) + A(t)\left(\int_0^t f(s,t)dR(s)\right) dt \\
&= c_0'(t)dt + A(t)\big(\hat{Y}(t) - c_0(t)\big)dt + S(t)C(t)^T G(t)^T \Sigma_2(t)^{-1} dR(t) \\
&= A(t)\hat{Y}(t)dt + S(t)C(t)^T G(t)^T \Sigma_2(t)^{-1} dR(t).
\end{aligned}$$

By (49) we have,

$$\begin{aligned}
dR(t) &= G(t)dM(t) \\
&= G(t)\left[C(t)\big(Y(t) - \hat{Y}(t)\big)dt + D(t)dX_2(t)\right] \qquad \text{by (48)} \\
&= G(t)\left[C(t)Y(t)dt + D(t)dX_2(t) - C(t)\hat{Y}(t)dt\right] \\
&= G(t)\left[dZ(t) - C(t)\hat{Y}(t)dt\right],
\end{aligned}$$

substituting in the above we conclude that,

$$\begin{aligned}
d\hat{Y}(t) &= A(t)\hat{Y}(t)dt \\
&\quad + S(t)C(t)^T G(t)^T \Sigma_2(t)^{-1} G(t)\left[dZ(t) - C(t)\hat{Y}(t)dt\right].
\end{aligned}$$

□

2.5 Extension of the Kalman - Bucy Filter

Given an observation process with a fixed finite variance, we know from (67) what the appropriate Kalman-Bucy filter is. However given the scenario of an observation process which is allowed to have infinite variance what is the best estimate of the system in this case? What can we say about the mean square error?

We now wish to extend the Kalman-Bucy filter derived in Theorem 2.4.3 to allow an infinite variance observation process. We do this by dropping the assumption that the process $(X_2(t), t \geq 0)$ has finite second moment for all $t \geq 0$. However we continue to insist that it has finite first moment. i.e

$$\mathbb{E}[|X_2(t)|] < \infty.$$

A necessary and sufficient condition for integrability of $X_2(t)$ for all $t \geq 0$ is,

- $\int_{|x|\geq 1} |x|\nu_2(dx) < \infty$

see [3, Theorem 2.5.2] for the proof.

In order to derive a Kalman-Bucy filter we will find it convenient to approximate X_2 by a sequence of Lévy processes $(X_2^{(n)}(t), t > 0, n \in \mathbb{N})$ where each $\mathbb{E}[|X_2^{(n)}(t)|^2] < \infty$ for all $n \in \mathbb{N}$. We do this by truncating the jumps, i.e.

$$X_2^{(n)}(t) = B_2(t) + \int_{|x|\leq n} x\tilde{N}_2(t, dx), \tag{68}$$

and we note that,

$$\mathbb{E}[X_2^{(n)}(1)X_2^{(n)}(1)^T] = \rho_2^{(n)} = A_2 + \int_{|x|\leq n} xx^T \nu_2(dx). \tag{69}$$

We now have that $\int_{|x|\geq 1} |x|^2\nu_2(dx) = \infty$ and we assume that $\int_{|x|\geq 1} x_i^2\nu_2(dx) = \infty$ for all values of i where $1 \leq i \leq r$.

We also write $\beta_{\nu_2}(n) := \max\left\{\int_{|x|\leq n} x_i^2\nu_2(dx); 1 \leq i \leq r\right\}$. In order to develop a manageable theory, we will also make the assumption that $\left|\int_{|x|\geq 1} x_i x_j \nu(dx)\right| < \infty$ for all $i \neq j$.

An example of the above set up would be when the components of X_2 are independent, centred one-dimensional Lévy processes, all of which have infinite variance. In that case, for each $i \neq j$,

$$\int_{|x|\geq 1} x_i x_j \nu(dx) = \left(\int_{|x|\geq 1} x_i \nu(dx)\right)\left(\int_{|x|\geq 1} x_j \nu(dx)\right) = 0.$$

Let $\psi^{(n)} = (\psi_{ij}^{(n)})$ be the covariance matrix corresponding to the jump part of X_2^n as in (22). In this section, we will need to consider the existence of limits of the form $\lim_{n\to\infty}(\psi^{(n)})^{-1}$. If $r = 1$ it is easy to see that this limit exists and is zero. If $r > 1$, then each element of the matrix of cofactors of $\psi^{(n)}$ is $O(\beta_\nu(n)^\alpha)$ where $1 \leq \alpha \leq r-1$, however $\det(\psi^{(n)})$ is $O(\beta_\nu(n)^r)$. It follows that $\lim_{n\to\infty}(\psi^{(n)})^{-1} = 0$ also in this case.

We will write $\gamma = \mathrm{diag}\left(\int_{|x|\leq 1} x_1^2 \nu(dx), \ldots, \int_{|x|\leq 1} x_r^2 \nu(dx)\right)$ and also $\gamma_a := a + \gamma$.

Theorem 2.5.1. *For each $t \geq 0$,*

$$\lim_{n\to\infty} \left|\left|X_2(t) - X_2^{(n)}(t)\right|\right|_1 = 0.$$

Proof.

$$\begin{aligned}
\left|\left|X_2(t) - X_2^{(n)}(t)\right|\right|_1 &= \mathbb{E}\left[\left|\int_0^t \int_{|x|\geq n} x\tilde{N}(ds, dx)\right|\right] \\
&= \mathbb{E}\left[\left|\int_0^t \int_{|x|\geq n} xN(ds, dx) - \int_0^t \int_{|x|\geq n} x\nu(dx)ds\right|\right] \\
&\leq \mathbb{E}\left[\left|\int_0^t \int_{|x|\geq n} xN(ds, dx)\right|\right] + t\int_{|x|\geq n} |x|\nu(dx).
\end{aligned}$$

Now if we define $U_A = \big(U_A(t), t \geq 0\big)$ as $U_A(t) = \int_A xN(t, dx)$ then using [3, p.231]

$$\begin{aligned}
\mathbb{E}\left[\left|\int_0^t \int_{|x|\geq n} xN(ds, dx)\right|\right] &= \mathbb{E}\left[\left|\sum_{0\leq u\leq t} \Delta U_{(|x|\geq n)}(u) 1_{(|x|\geq n)}\big(\Delta U_{(|x|\geq n)}(u)\big)\right|\right] \\
&\leq \mathbb{E}\left[\sum_{0\leq u\leq t} \left|\Delta U_{(|x|\geq n)}(u) 1_{(|x|\geq n)}\big(\Delta U_{(|x|\geq n)}(u)\big)\right|\right] \\
&= \mathbb{E}\left[\int_0^t \int_{|x|\geq n} |x| N(ds, dx)\right] \\
&= \mathbb{E}\left[\int_{|x|\geq n} |x| N(t, dx)\right] \\
&= t\int_{|x|\geq n} |x|\nu(dx).
\end{aligned}$$

We conclude that,

$$\left|\left|X_2(t) - X_2^{(n)}(t)\right|\right|_1 \leq 2t\int_{|x|\geq n} |x|\nu(dx) \to 0 \text{ as } n \to \infty.$$

$\square$

Now we need to show that each $\mathbb{E}[|X_2^{(n)}(t)|^2] < \infty$ or equivalently each $\int_{|x|\leq n} |x|^2 \nu(dx) < \infty$ see [3, Theorem 2.5.2], so that $X_2^{(n)} \in L^2(\Omega, \mathcal{F}, P; \mathbb{R}^r)$.

Lemma 2.5.2.

$$\int_{|x|\leq n} |x|^2 \nu(dx) < \infty$$

Proof.

$$\begin{aligned}
\int_{|x|\leq n} |x|^2 \nu(dx) &= \int_{|x|<1} |x|^2 \nu(dx) + \int_{1\leq|x|\leq n} |x|^2 \nu(dx) \\
&\leq \int_{|x|<1} |x|^2 \nu(dx) + n^2 \nu\{1 \leq |x| \leq n\} \\
&< \infty.
\end{aligned}$$

□

The observation process $(Z(t), t \geq 0)$ is defined as previously as the solution of the stochastic differential equation

$$dZ(t) = C(t)Y(t) + D(t)dX_2(t),$$

but note that it is now of infinite variance. We obtain a sequence of finite variance observation processes by incorporating the sequence of Lévy processes we defined in (68).

$$dZ^{(n)}(t) = C(t)Y(t)dt + D(t)dX_2^{(n)}(t). \tag{70}$$

It therefore follows directly from Lemma 2.5.2 that each
$Z^{(n)}(t) \in Ł^2(\Omega, \mathcal{F}, P; \mathbb{R}^m)$.

Theorem 2.5.3. *For each $t \geq 0$,*

$$\lim_{n\to\infty} \left|\left|Z(t) - Z^{(n)}(t)\right|\right|_1 = 0.$$

Proof. From (70) and (34),

$$\begin{aligned}
\left|\left|Z(t)-Z^{(n)}(t)\right|\right|_1 &= \mathbb{E}\left[\left|\int_0^t D(s)dX_2(s) - \int_0^t D(s)dX_2^{(n)}(s)\right|\right] \\
&= \mathbb{E}\left[\left|\int_0^t \int_{|x|\geq n} D(s)x\tilde{N}(ds,dx)\right|\right] \\
&= \mathbb{E}\left[\left|\int_0^t \int_{|x|\geq n} D(s)xN(ds,dx) - t\int_0^t \int_{|x|\geq n} D(s)x\nu(dx)ds\right|\right] \\
&\leq \mathbb{E}\left[\left|\int_0^t \int_{|x|\geq n} D(s)xN(ds,dx)\right|\right] + t\left|\int_0^t \int_{|x|\geq n} D(s)x\nu(dx)ds\right| \\
&\leq \mathbb{E}\left[\left|\int_0^t \int_{|x|\geq n} D(s)xN(ds,dx)\right|\right] \\
&\quad + t\int_0^t \int_{|x|\geq n} ||D(s)||_{OP}\cdot|x|\nu(dx)ds \\
&\leq \mathbb{E}\left[\left|\int_0^t \int_{|x|\geq n} D(s)xN(ds,dx)\right|\right] \\
&\quad + t\sup_{0\leq s\leq t}||D(s)||_{OP}\int_{|x|\geq n}|x|\nu(dx).
\end{aligned}$$

Following a similar argument to the proof of Theorem 2.5.1 we get,

$$\left|\left|Z(t)-Z^{(n)}(t)\right|\right|_1 \leq 2t\sup_{0\leq s\leq t}||D(s)||_{OP}\int_{|x|\geq n}|x|\nu(dx) \to 0 \text{ as } n\to\infty.$$

□

We now need to make sense of $\hat{Y}(t)$ for $0 \leq t \leq T$. This was previously defined using the projection from $L^2(\Omega,\mathcal{F},P;\mathbb{R}^d)$ onto $\mathcal{L}(Z,T)$. Let us define $\mathcal{L}^{(n)}(Z,t)$ to be the closure of the linear span of all vectors of the form,

$$c_0 + c_1Z^{(n)}(t_1) + \cdots + c_kZ^{(n)}(t_k), \quad 0\leq t_1,\ldots,t_k\leq t, n\in\mathbb{N}. \tag{71}$$

For all $n\in\mathbb{N}$ let $\mathcal{P}^{(n)}$ be the orthogonal projection of $L^2(\Omega,\mathcal{F},P;\mathbb{R}^d)$ onto $\mathcal{L}^{(n)}(Z,t)$, and we write $\hat{Y}^{(n)}(t) = \mathcal{P}^{(n)}(Y(t))$.

We can clearly see that for each $n\in\mathbb{N}$, $\hat{Y}^{(n)}$ satisfies a Kalman-Bucy filter derived previously, but not so clear is that the mean square error $\left(S^{(n)}(t), n\in\mathbb{N}, t>0\right)$ defined as

$$S^{(n)}(t) = \mathbb{E}\left[\left(Y(t)-\hat{Y}^{(n)}(t)\right)\left(Y(t)-\hat{Y}^{(n)}(t)\right)^T\right]$$

satisfying,

$$\frac{dS^{(n)}(t)}{ds} = A(t)S^{(n)}(t) + S^{(n)}(t)A(t)^T + B(t)\rho_1 B(t)^T$$
$$- \left(S^{(n)}(t)C(t)^T G(t)^T (\Sigma_2^{(n)}(t))^{-1} G(t)C(t)S^{(n)}(t)^T \right),$$

converges to a limit. Clearly for all $0 \leq t \leq T$ the sequence $(S^{(n)}(t), n \in \mathbb{N})$ is bounded below by 0.

We will find the following inequality useful in the sequel.

Lemma 2.5.4. *For all $n \in \mathbb{N}, 0 \leq t \leq T$,*

$$||S^{(n)}(t)||_{HS} \leq d||Y(t)||_2^2 \tag{72}$$

Proof. For $1 \leq i, j \leq d$,

$$\begin{aligned} |S^{(n)}(t)_{ij}| &\leq \mathbb{E}\left[|Y(t)_i - \hat{Y}(t)_i||Y(t)_j - \hat{Y}(t)_j|\right] \\ &\leq \mathbb{E}\left[|Y(t) - \hat{Y}(t)|_n^2\right] \\ &= ||P^{(n)\perp}Y(t)||_2^2 \\ &\leq ||Y(t)||_2^2 \end{aligned}$$

Taking the Hilbert-Schmidt norm we get,

$$\begin{aligned} ||S^{(n)}(t)||_{HS} &= \left(\sum_{i,j=1}^{d} S^{(n)}(t)_{i,j}^2 \right)^{\frac{1}{2}} \\ &\leq d||Y(t)||_2^2 \end{aligned}$$

□

We must next find an analogue of the Riccati equation. To simplify expressions we introduce the notation $Q(t) := G(t)C(t)$ for $0 \leq t \leq T$. We note that since $\lim_{n\to\infty} \varphi^{(n)} \to 0$, for each $0 \leq t \leq T, \Sigma_2^{(\infty)}(t)^{-1} = \lim_{n\to\infty} \Sigma_2^{(n)}(t)^{-1}$ exists and is the zero matrix.

Theorem 2.5.5. *For each $0 \leq t \leq T$ the sequence $(S^{(n)}(t), n \in \mathbb{N})$ converges to a matrix $S^{(\infty)}(t)$. The mapping $t \to S^{(\infty)}$ is differentiable on $[0, T]$ and $S^{(\infty)}$ is the unique solution of the Riccati equation*

$$\frac{dS^{(\infty)}(t)}{dt} = A(t)S^{(\infty)}(t) + S^{(\infty)}(t)A(t)^T + B(t)\rho_1 B(t)^T$$

Proof. Let $(\Gamma(t), 0 \leq t \leq T)$ be the unique solution of,

$$\frac{d\Gamma(t)}{dt} = A(t)\Gamma(t) + \Gamma(t)A(t)^T + B(t)\rho_1 B(t)^T, \quad \text{with } \Gamma(0) = \mathrm{Cov}(Y_0).$$

Then we wish to show that $\Gamma(t) = S^{(\infty)}$ for all $0 \leq t \leq T$. Using (63), for all $n \in \mathbb{N}$ we have,

$$\begin{aligned}\frac{d}{dt}\left(\Gamma(t) - S^{(n)}(t)\right) &= A(t)\left(\Gamma(t) - S^{(n)}(t)\right) + \left(\Gamma(t) - S^{(n)}(t)\right)A(t)^T \\ &\quad + \left(S^{(n)}(t)Q(t)^T\left(\Sigma_2^{(n)}(t)\right)^{-1}Q(t)S^{(n)}(t)^T\right).\end{aligned}$$

Integrating we get,

$$\begin{aligned}\Gamma(t) - S^{(n)}(t) &= \int_0^t A(u)\left(\Gamma(u) - S^{(n)}(u)\right)du + \int_0^t \left(\Gamma(u) - S^{(n)}(u)\right)A(u)^T du \\ &\quad + \int_0^t \left(S^{(n)}(u)Q(u)^T\left(\Sigma_2^{(n)}(u)\right)^{-1}Q(u)S^{(n)}(u)^T\right)du.\end{aligned}$$

Taking norms and using the triangle inequality and (72)

$$\begin{aligned}\left|\left|\Gamma(t) - S^{(n)}(t)\right|\right|_{HS} &= \left|\left|\int_0^t A(u)\left(\Gamma(u) - S^{(n)}(u)\right)du \right.\right.\\ &\quad + \int_0^t \left(\Gamma(u) - S^{(n)}(u)\right)A(u)^T du \\ &\quad \left.\left. + \int_0^t \left(S^{(n)}(u)Q(u)^T\left(\Sigma_2^{(n)}(u)\right)^{-1}Q(u)S^{(n)}(u)^T\right)du\right|\right|_{HS} \\ &\leq \left|\left|\int_0^t A(u)\left(\Gamma(u) - S^{(n)}(u)\right)du\right|\right|_{HS} \\ &\quad + \left|\left|\int_0^t \left(\Gamma(u) - S^{(n)}(u)\right)A(u)^T du\right|\right|_{HS} \\ &\quad + \left|\left|\int_0^t \left(S^{(n)}(u)Q(u)^T\Sigma_2^{(n)}(u)^{-1}Q(u)S^{(n)}(u)^T\right)du\right|\right|_{HS} \\ &\leq 2\int_0^t ||A(u)||_{OP}\left|\left|\Gamma(u) - S^{(n)}(u)\right|\right|_{HS}du \\ &\quad + \int_0^t \left|\left|S^{(n)}(u)\right|\right|_{OP}^2 ||Q(u)||_{OP}^2 \left|\left|\Sigma_2^{(n)}(u)^{-1}\right|\right|_{HS}du \\ &\leq 2\int_0^t ||A(u)||_{OP}\left|\left|\Gamma(u) - S^{(n)}(u)\right|\right|_{HS}du \\ &\quad + d\int_0^t ||Y(u)||_2^4 ||Q(u)||_{OP}^2 \left|\left|\Sigma_2^{(n)}(u)^{-1}\right|\right|_{HS}du.\end{aligned}$$

Now using Gronwall's inequality,

$$\left|\left|\Gamma(t) - S^{(n)}(t)\right|\right|_{HS} = d\exp\left(\int_0^t ||A(u)||_{OP}du\right)$$
$$\times \int_0^t ||Y(u)||_2^4 \, ||Q(u)||_{OP}^2 \left|\left|\Sigma_2^{(n)}(u)^{-1}\right|\right|_{HS} du$$

By assumptions on D and the sequence (ψ^n) we have that

$$\sup_{n\in N}\sup_{0\leq t\leq T}\left|\left|\Sigma_2^{(n)}(t)^{-1}\right|\right|_{HS}^2 \leq \sup_{0\leq t\leq T}\left|\left|(G(t)D(t)\gamma_a D(t)^T G(t)^T)^{-1}\right|\right|_{HS}^2 < \infty,$$

and so by dominated convergence, $\lim_{n\to\infty}||\Gamma(t) - S_n(t)||_{HS} = 0$, and the result follows. □

This is a very surprising result, as the non-linear term has vanished as we've passed to the limit and so the Riccati equation is now a simple first order differential equation with an exact solution. It can be solved in one dimension by the integrating factor method, to give,

$$S^{(\infty)}(t) = \rho_1 \exp\left(2\int_0^t A(u)du\right)\int_0^t \exp\left(-2\int_0^s A(u)du\right)B(s)^2 ds. \quad (73)$$

In the finite dimensional case if $A(t)$ is symmetric for all $0 \leq t \leq T$, this can be solved using a similar methodology.

Our next result is the desired L^1 Kalman-Bucy filter.

Theorem 2.5.6. *For each* $0 \leq t \leq T$ *the sequence* $(\hat{Y}^{(n)}(t), n \in \mathbb{N})$ *converges in* $L^1(\Omega, \mathcal{F}, P; \mathbb{R}^d)$ *to* $\hat{Y}(t)$. *The process* $\hat{Y}$ *is the solution of the following DE:*

$$d\hat{Y}(t) \quad = A(t)\hat{Y}(t)dt \quad (74)$$

Proof. Let $\Psi = (\Psi(t), 0 \leq t \leq T)$ be the unique solution of the DE

$$d\Psi(t) = A(t)\Psi(t)dt \quad (75)$$

with the initial condition of $\Psi(0) = \mu_0$ (a.s). We are required to show that $\lim_{n\to\infty}||\Psi(t) - \hat{Y}_n(t)||_1 = 0$ and the result then follows by uniqueness of limits. We use a similar argument to that in the proof of Theorem 2.5.5. Using (67) we find that

$$\begin{aligned}\Psi(t) - \hat{Y}_n(t) = &\int_0^t A(r)\left(\Psi(r) - \hat{Y}^{(n)}(r)\right)dr \\ &- \int_0^t S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}Q(r)\hat{Y}^{(n)}(r)dr \\ &- \int_0^t S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}G(r)^T dZ^{(n)}(r) \quad (76)\end{aligned}$$

Taking the L^1-norm we find that

$$\begin{aligned}
||\Psi(t) - \hat{Y}^{(n)}(t)||_1 &\leq \int_0^t ||A(r)||_{OP}||\Psi(r) - \hat{Y}^{(n)}(r)||_1 dr \\
&+ \mathbb{E}\left[\left|\int_0^t S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}G(r)dZ^{(n)}(r)\right|\right] \\
&+ \int_0^t ||S^{(n)}(r)||_{OP}||Q(r)^T||_{OP}||\Sigma_2^{(n)}(r)^{-1}||_1 \\
&\times ||Q(r)||_{OP}||\hat{Y}^{(n)}(r)||dr. \qquad (77)
\end{aligned}$$

To proceed further we need some additional estimates. First note that for all $n \in \mathbb{N}$,

$$||\hat{Y}^{(n)}(t)||_1 \leq ||\hat{Y}^{(n)}(t)||_2 = ||P^{(n)}Y(t)||_2 \leq ||\hat{Y}(t)||_2.$$

Secondly, using (34):

$$\begin{aligned}
&\mathbb{E}\left[\left|\int_0^t S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}G(r)dZ^{(n)}(r)\right|\right] \\
&\quad \leq \mathbb{E}\left[\left|\int_0^t S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}Q(r)Y(r)dr\right|\right] \\
&\quad + \mathbb{E}\left[\left|\int_0^t S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}G(r)D(r)dB_2(r)\right|\right] \\
&\quad + \mathbb{E}\left[\left|\int_0^t \int_{|y|<n} S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}G(r)D(r)\tilde{N}_2(dy,dr)\right|\right] \\
&\quad \leq \int_0^t ||S^{(n)}(r)||_{OP}||Q(r)^T||_{OP}||\Sigma_2^{(n)}(r)^{-1}||_{OP}||Q(r)||_{OP}||Y(r)||_1 dr \\
&\quad + \left(\mathbb{E}\left[\left|\int_0^t S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}G(r)D(r)dB_2(r)\right|^2\right]\right)^{\frac{1}{2}} \\
&\quad + 2\int_{\mathbb{R}_0} |y|\nu_2(dy)\int_0^t ||S^{(n)}(r)||_{OP}||Q(r)^T||_{OP} \\
&\qquad \times ||\Sigma_2^{(n)}(r)^{-1}||_{OP}||G(r)D(r)||_{OP}dr,
\end{aligned}$$

and by Itô's isometry,

$$\mathbb{E}\left[\left|\int_0^t S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}G(r)D(r)dB_2(r)\right|^2\right]$$
$$= \int_0^t ||S^{(n)}(r)Q(r)^T\Sigma_2^{(n)}(r)^{-1}G(r)D(r)A^{\frac{1}{2}}||_{HS}^2 dr$$
$$\leq \mathrm{tr}(A)\int_0^t ||(S^{(n)}(r)||_{OP}^2||Q(r)^T||_{OP}^2$$
$$\times ||\Sigma_2^{(n)}(r)^{-1}||_{OP}^2||G(r)D(r)||_{OP}^2 dr.$$

We now apply these estimates to (77) and also make use of Lemma 2.5.4 to conclude that there exist bounded measurable positive functions $t \to H_j(t)$ defined on $[0,T]$ for $j = 1,2,3,4$ so that

$$\begin{aligned}||\Psi(t)-\hat{Y}^{(n)}(t)||_1 &\\ \leq \int_0^t H_1(r)||\Psi(r)-\hat{Y}^{(n)}(r))||_1 dr &+ \int_0^t H_2(r)||\Sigma_2^{(n)}(r)^{-1}||_{OP}dr\\ + \left(\int_0^t H_3(r)||\Sigma_2^{(n)}(r)^{-1}||_{OP}dr\right)^{\frac{1}{2}} &+ \int_{\mathbb{R}_0}|y|\nu(dy)\int_0^t H_4(r)||\Sigma_2^{(n)}(r)^{-1}||_{OP}dr.\end{aligned}$$

Then by Gronwall's inequality

$$\begin{aligned}||\Psi(t)-\hat{Y}^{(n)}(t)||_1 \leq \exp\left(\int_0^t H_1(r)dr\right)&\left[\int_0^t H_2(r)||\Sigma_2^{(n)}(r)^{-1}||_{OP}dr\right.\\ &+\left(\int_0^t H_3(r)||\Sigma_2^{(n)}(r)^{-1}||_{OP}dr\right)^{\frac{1}{2}}\\ &\left.+\int_{\mathbb{R}_0}|y|\nu(dy)\int_0^t ||\Sigma_2^{(n)}(r)^{-1}||_{OP}H_4(r)dr\right],\end{aligned}$$

and the result follows by using dominated convergence (which is justified by using (2.5.4) and the fact that $\int_{\mathbb{R}}|y|\nu(dy) < \infty$). $\square$

2.6 Numerics

In this section we look at three different methods for filtering an infinite variance observation process in a linear framework. Our aim will be to investigae the value of the result in (74).

2.6.1 Infinite Variance Observations of a Mean Reverting Brownian Motion

The example we look at is a simple numerical example of filtering a mean reverting Brownian motion from infinite variance observations. We aim to compare the performance of the different filters in an ideal situation where all parameters are known. We choose the following two existing filters in this field i.e the work of [39] and [1]. The work of [39] is only applicable in the case of α-stable noise, and the work of [1] is only applicable in the case of a Gaussian system and Gaussian observations with a pure Lévy jump process. Therefore the only point of intersection for these three filters will be a Gaussian system with α stable observations. In order to proceed we will employ the use of the commonly used Euler approximation see [23, Chapter 2].

We assume the observations to occur at a fixed rate and set the last observation to occur at time T. Our comparative example is a system described by a mean reverting Brownian motion, i.e

$$Y(t) = Y_0 - \int_0^t Y(s)ds + X_1(t)$$

where in this case $A(t) = -1$ and $B(t) = 1$, $X_1(t)$ is a standard Brownian motion and Y_0 is a standard Gaussian random variable which is independent of $(X_1(t), t \geq 0)$. We take observations of this system of the following form,

$$dZ(t) = Y(t)dt + dX_2(t)$$

where $C(t) = 1$ and $D(t) = 1$, and $X_2(t)$ is an symmetric α-stable process with infinite variance, and finite expectation, so that we must take $1 < \alpha < 2$ and we fix the dispersion parameter c (see Theorem 1.1.13) to be 1.

For the filter described in (74), $S^{(\infty)}(t)$ is of the form,

$$\frac{dS^{(\infty)}(t)}{dt} = -2S^{(\infty)}(t) + 1, \text{where} \quad S_0 = 0$$

which we integrate using the integrating factor method to get,

$$S^{(\infty)}(t) = \frac{1}{2}\left[1 - \exp(-2t)\right]$$

and we compute our best estimate using (74) which yields,

$$d\hat{Y}(t) = -\hat{Y}(t)dt,$$

and so,

$$\hat{Y}(t) = Y_0 e^{-t}.$$

The filter of Le Breton and Musiela [39] takes the form,

$$d\hat{Y}(t)^* = -\hat{Y}(t)^* dt - |\gamma|^{q/p}(dZ(t) + \hat{Y}(t)^* dt)$$

where γ is given by,

$$\frac{d\gamma}{dt} = -p\gamma + 1 - (p-1)|\gamma|^q.$$

We take the values $p = 1.1$ and $q = 11$ as were taken in [1]. Finally the filter of Ahn and Feldman, [1] takes the form of the following recursion,

$$\hat{Y}^{**}(t_{j+1}) = -\lambda_j \sigma_{j+1}^2 \frac{f'_{j+1}}{f_{j+1}} \left(\Delta Z(t_{j+1}) - \gamma_j \hat{Y}^{**}(t_j) \right) + \exp\left(-\delta \hat{Y}^{**}(t_j) \right),$$

with the following coefficients, $p = \exp(\delta) - 1$, $r = p - p\exp(-\delta) - \delta$

$$\lambda_{i+1} = \frac{\sigma_{i+1}^2 + r}{p\sigma_{i+1}^2}$$

$$\sigma_{i+1}^2 = \frac{1}{2}p^2(1 - \exp(2\delta)) + 2p(\exp(-\delta) - 1) + 3\delta + p\exp(-\delta)V_i$$

$$V_{i+1} = 2p^{-2}\delta - p^{-2}\sigma_{i+1}^2(1 - p\lambda_i)^2.$$

We simulate the system and observation processes using the Euler approximation. In order to do this, we will make use of the self-similarity property of stable processes, (see [54, Proposition 13.5] for a proof that a Lévy process is self similar if and only if it is a strictly stable processes). Firstly we consider the system, this results in us being required to simulate the increment of a standard Brownian motion, which we can do by the following

$$B(t + \Delta t) - B(t) \approx N(0, \Delta t) \approx \sqrt{\Delta t} \cdot N(0, 1).$$

For the observations we need to simulate the increment of an α-stable process which can be done as follows,

$$X_2(t + \Delta t) - X_2(t) \approx X_2(\Delta t \cdot 1) \approx (\Delta t)^{\frac{1}{\alpha}} X_2(1).$$

To measure and compare the filters accuracy we used 3 different values of α, i.e $\{1.1, 1.5, 1.9\}$. We then calculated the error, this was done by taking the mean square error of the system from the predicted value. We set the inter-arrival times of the observations to be 0.01, and the expiration time to be $T = 10$.

100,000 Monte-Carlo simulations were generated to allow us to estimate the error with some accuracy, we then took a median of these values due to the skewed nature of the distribution of the results.

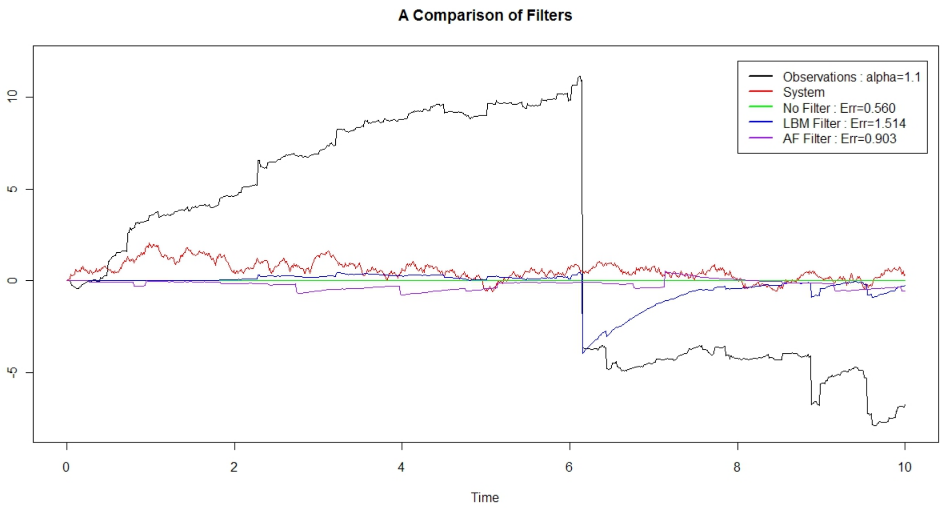
From Table 1 we note that having no filter provides a closer estimate of the underlying system (for $\alpha = \{1.1, 1.5\}$. Whilst these results are particular only to the setup proposed, it is still worth noting that there may be other examples where computing estimates in the absence of noise will give results closer than when making best estimates in the presence of infinite variance noise.

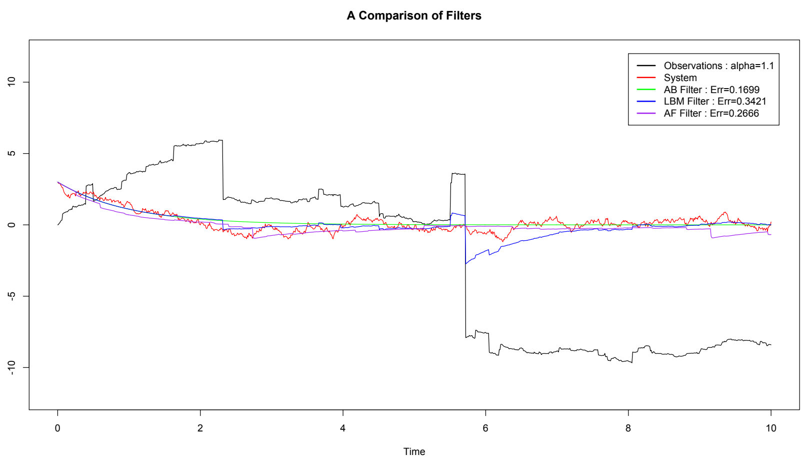
Table 1: A Comparison of Filters

α	No Filter Error	LBM Filter Error	AF Filter Error
1.1	0.6601	0.7819	0.7485
1.5	0.6593	0.6819	0.7507
1.9	0.6603	0.6446	0.7508

The following figure provides a graphical representation of one instance of the above calculations. It shows that the filter of Le Breton and Musiela is quite prone to large jumps in the observations, causing a higher error in this instance. The filter of Ahn and Feldman remains relatively close to 0, but having no filter provides the lowest error. Whilst this may seem counter-intuitive it should be noted that when $\alpha = 1.1$ the noise in the observation process is very wild, and so the observation process bears almost no resemblance to the system process.

For the code, see appendix B.





The above figures give a comparison of the three filters, incorporating two different starting values. a

3 Non-Linear Filtering

In this chapter we move away from the linear Kalman-Bucy filter and look towards the non-linear arena. The output of the filter is the distribution of the estimated process given all the available data of the observations, otherwise known as the posterior distribution. We will start by introducing the relevant background before moving on to the celebrated filtering equations of Kallianpur-Striebel, Zakai and Kushner-Stratonovich. During this chapter we mainly follow and extend the works of [5] and [45]. Whilst [5] expertly deals with the Brownian case, more general noise processes such as those with jumps are not dealt with. Meanwhile the papers [45] and [44] deal with Lévy processes but leave out many technical difficulties which will be fully dealt with in this chapter.

3.1 Introduction

We can draw parallels to the previous chapter in that we have an observation process $(Z(t), t \geq 0)$ and a signal process $(Y(t), t \geq 0)$ which is not directly observable. Similarly to the Kalman-Bucy filter derived earlier the observation process is a function Z of the signal Y and a measurement noise particular to the observations say $(X_2(t), t \geq 0)$, i.e.

$$Z(t) = f\left(Y(t), X_2(t)\right).$$

This will be defined rigorously later.

Let $(\mathcal{Z}_t, t \geq 0)$ be the σ-algebra generated by the observations Z. Then this can be interpreted as the collective information gathered from all the observations up to time t. We wish to use this information to try and answer questions such as: what is the best estimate of the signal at time t, and can we calculate the conditional probability that the signal lies inside some set G at time t, i.e. $P(Y(t) \in G|\mathcal{Z}_t)$? The first question is generally referred to in a mean square sense, by which we wish to minimise $\mathbb{E}\left[\left(Y(t) - \hat{Y}(t)\right)^2 |\mathcal{Z}_t\right]$ and therefore this relates to computing $\hat{Y}(t) = \mathbb{E}\left[Y(t)|\mathcal{Z}_t\right]$.

Readers may recall that in the previous chapter we used linear estimates, in this chapter we will be using measurable estimates, and these two estimations are generally different. To expand on this, we have the following inequality in the L^2 framework,

$$\mathbb{E}\left[(Y - \mathbb{E}\left[Y|\mathcal{F}\right]))^2\right] \leq \mathbb{E}\left[(Y - \mathcal{P}_{\mathcal{L}}(Y))^2\right]$$

for more on this, see [47, p.92] and [16, p.19].

3.2 Framework

In this section we are following the approach of [5, Chapter 2]. As mentioned in the introduction we are interested in finding the conditional distribution of the signal Y given the σ-algebra generated by the observations. We now give a more precise definition of this observation σ-algebra, which we denote $\mathcal{Z}_t$.

$$\mathcal{Z}_t = \sigma(Z_t, 0 \leq s \leq t) \vee \mathcal{N},$$

where $\mathcal{N}$ is the collection of null sets of the complete probability space $(\Omega, \mathcal{F}, P)$. It is mandatory that we incorporate these null sets into our observation sigma algebra as we will need to modify $\mathcal{Z}_t$-adapted processes. This augmentation of $\mathcal{Z}_t$ by the addition of null sets ensures that any modified process is also $\mathcal{Z}_t$ adapted. We also require the right-continuous enlargement of the observation filtration $\mathcal{Z}_t$. Therefore we choose the following enlargement, $(\mathcal{Z}_{t+}, t \geq 0)$ where $\mathcal{Z}_{t+} = \cap_{s>t} \mathcal{Z}_s$, and to save notation we will write $\mathcal{Z}_t$ instead of $\mathcal{Z}_{t+}$ from now on.

Another very important point is that the conditional distribution of the signal, otherwise known as the posterior distribution, can be thought of as a stochastic process taking values in the space of probability measures. To formalise this we let the signal process Y take values in $\mathbb{R}$. We now define $\mathcal{P}(\mathbb{R})$, the space of Borel probability measures over $\mathbb{R}$ and a $\mathcal{P}(\mathbb{R})$-valued $\mathcal{Z}_t$ adapted stochastic process $(\pi_t, t \geq 0)$ which satisfies the following,

$$\pi_t(A) = P(Y(t) \in A | \mathcal{Z}_t),$$

where A is an arbitrary set in the σ-algebra $\mathcal{B}(\mathbb{R})$ and $\pi_0 \in \mathcal{P}(\mathbb{R})$.

For the following work we now consider one particular case of a signal process, i.e. when Y is a diffusion process. We let the state space of the signal be $\mathbb{R}$ and let Y be the solution of the stochastic differential equation driven by a Brownian motion $(B^Y(t), t \geq 0)$,

$$Y(t) = Y_0 + \int_0^t b(Y(s))ds + \int_0^t \sigma(Y(s))dB^Y(s). \tag{78}$$

We assume that $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are globally Lipschitz. This global Lipschitz assumption then ensures that (78) has a unique solution using [5, Theorem B.38].

In later work we will require the martingale $(M_t^\varphi, t \geq 0)$ where

$$M_t^\varphi = \int_0^t \varphi'(Y(s))dB(s)$$

for any $\varphi \in C_c^2(\mathbb{R})$ the space of twice differentiable compactly supported continuous functions on $\mathbb{R}$. This is derived from Itô's formula from which we obtain

$$M_t^{\varphi} = \varphi(Y(t)) - \varphi(Y(0)) - \int_0^t A\varphi(Y(s))ds \quad t \geq 0. \tag{79}$$

where A is the second order differential operator,

$$A = b\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2}, \tag{80}$$

and we assume that the distribution of $Y(0)$ is π_0. (79) suggests that more general signal processes could be described as the solution of martingale problems. There is some discussion of this in [5].

The following work extends and clarifies the work of [45] where the Zakai equation is stated but not derived, nor are any conditions for existence of solutions obtained. The paper [44] offers a proof of the Zakai equation, along with some conditions for existence of its solution. The proof is somewhat lacking in its thoroughness, which will be dealt with here in a much more detailed fashion, and in the upcoming lemmas we will show that a certain martingale condition, which was imposed in [44] is not required and will be derived from our assumptions. Other work in this area includes the papers by Colaneri and Ceci [13] and [12] which have a system and observation process that employs the use of Poisson random measures with common jump times, and a proof of uniqueness of solution. This paper focuses on deriving the Kushner-Stratonovich equation using the filtered martingale problem approach. The paper [48] has a system driven by Brownian motion and a Poisson random measure with observations of either a diffusion or a Poisson process. This paper follows the change of measure approach but again misses out many details in its derivation of the Zakai equation. Finally for results regarding a discontinuous system and continuous observations see [55].

After deriving the Zakai equation we will then obtain the normalised Kushner-Stratonovich equation.

We begin by outlining the non-linear filtering problem we wish to investigate. Let $h : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ be a measurable function such that

$$P\left(\int_0^t |h(s, Y(s))|ds < \infty\right) = 1. \tag{81}$$

Suppose that the observation process is given by the solution of the stochastic differential equation,

$$dZ(t) = h(t, Y(t))dt + dB^Z(t) + \int_{\mathbb{R}_0} yN_\lambda(dt, dy) \tag{82}$$

where B^Z is a Brownian motion and N_λ is an independent integer valued random measure which has the predictable compensator

$$\lambda(t, Y(t), y)dt\nu(dy)$$

for a Lévy measure ν which we restrict to be finite, and a measurable function $\lambda : [0, T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. See [25, p.32] for more details of such random measures.

We then note that,

$$\mathbb{E}\left[N_\lambda(t, A)\right] = \mathbb{E}\left[\int_0^t \int_A \lambda(u, Y(u), y)\nu(dy)du\right].$$

We further assume that $\lambda > 0$ a.e. and finally we will need,

$$\int_0^t \int_{\mathbb{R}_0} |y|^2 \nu(dy) < \infty.$$

The need for this requirement will be made apparent later in the work.

The solution to (82) can be constructed using the interlacing procedure similar to [3, Theorem 6.2.9] outlined here.

Let $P = (P(t), t \geq 0)$ be the point process defined by

$$P(t) = \int_0^t \int_{\mathbb{R}_0} yN_\lambda(ds, dy)$$

with jumps at times $\tau_1, \tau_2, \tau_3 \dots$.

Let $A = (A(t), t \geq 0)$ be the continuous process defined by,

$$A(t) = Z_0 + \int_0^t h(s, Y(s))ds + B^Z(t) \tag{83}$$

then we construct the process $Z(t)$ as follows,

$$\begin{aligned} Z(t) &= A(t) & 0 \leq t < \tau_1, \\ Z(\tau_1) &= A(\tau_1) + \Delta P(\tau_1) & t = \tau_1 \\ Z(t) &= Z(\tau_1) + A(t) - A(\tau_1) & \tau_1 < t < \tau_2 \\ Z(\tau_2) &= Z(\tau_2-) + \Delta P(\tau_2) & t = \tau_2. \end{aligned}$$

and so on recursively.

The initial condition is assumed to be a random variable independent of $Y(t)$ and $B^Z(t)$. For notational simplicity we will henceforth write $\lambda(t,y) = \lambda(t, Y(t), y)$, but the reader should always keep it in mind that λ is also a function of a random variable. Note that the condition (81) is required to ensure the existence of the integral in (83).

So, the filtering problem is the calculation of the conditional distribution π_t of the signal Y at time t given the information gathered from observing the process Z up to time $t \geq 0$. That is for $\varphi \in \mathbb{C}_c^2(\mathbb{R})$ we require

$$\pi_t(\varphi) = \mathbb{E}\left[\varphi(Y(t))|\mathcal{Z}_t\right].$$

We solve this problem by using the change of measure approach. This method works by constructing a new measure under which the observations Z become a Lévy process and we show that the probability measure valued process π can be represented in terms of an unnormalised measure valued process ρ. We then show that ρ satisfies a linear evolution equation called the Zakai equation which we obtain via Itô's formula and approximation techniques.

3.3 A Change of Measure

In this section we find an equivalent probability measure on Ω to transform the process Z into a Lévy process. To be more precise we define Q on $(\Omega, \mathcal{F})$ via $dQ = \Omega_T dP$ with Radon-Nikodym density

$$\Omega_T = \exp\left\{ -\int_0^T h(t, Y(t)) dB^Z(t) - \frac{1}{2}\int_0^T h(t, Y(t))^2 dt \right.$$
$$\left. - \int_0^T \int_{\mathbb{R}_0} \log(\lambda(t, y)) N_\lambda(dt, dy) - \int_0^T \int_{\mathbb{R}_0} (1 - \lambda(t, y))\nu(dy) dt \right\}. \quad (84)$$

If we were to only consider Brownian motion as the signal and observation noise we could impose the classic Novikov condition [42] to ensure that $(\Omega_t, 0 \leq t \leq T)$ is a martingale, i.e. that

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t h(t, Y(t))^2 ds\right)\right] < \infty.$$

This can be quite a difficult condition to verify, and so in the course of the next few lemmas we will seek to provide conditions which ensure that Ω is a martingale.

The stochastic derivative in the following lemma will be of use in showing the change of measure (84) is a martingale and in the derivation of the Zakai equation.

Lemma 3.3.1.

$$d\Omega_t = \Omega_{t-}\left\{ -h(t, Y(t)) dB^Z(t) + \int_{\mathbb{R}_0} \left[\lambda(t, y)^{-1} - 1\right] \tilde{N}_\lambda(ds, dy) \right\} \quad (85)$$

Proof. Applying Theorem 1.5.1 to (84),

$$\begin{aligned} d\Omega_t &= \Omega_{t-}\left\{ -h(t, Y(t)) dB^Z(t) - \frac{1}{2} h(t, Y(t))^2 dt \right\} + \frac{1}{2}\Omega_{t-} h(t, Y(t))^2 dt \\ &\quad - \Omega_{t-}\int_{\mathbb{R}_0} \left[1 - \lambda(t, y)\right] \nu(dy) dt + \int_{\mathbb{R}_0} \left[\exp\left\{-\eta(t-) - \log\lambda(t, y)\right\} - \Omega_{t-}\right] N_\lambda(dt, dy) \\ &= \Omega_{t-}\left\{ -h(t, Y(t)) dB^Z(t) - \int_{\mathbb{R}_0} \left[1 - \lambda(t, y)\right] \nu(dy) dt \right\} \\ &\quad + \int_{\mathbb{R}_0} \left[\lambda(t, y)^{-1} - 1\right] N_\lambda(ds, dy) \\ &= \Omega_{t-}\left\{ -h(t, Y(t)) dB^Z(t) + \int_{\mathbb{R}_0} \left[\lambda(t, y)^{-1} - 1\right] \tilde{N}_\lambda(ds, dy) \right\} \end{aligned}$$

□

The following lemma is a generalisation of [5, Lemma 3.9], and shows us the conditions required in order for the change of measure to be a martingale.

Lemma 3.3.2. *Let h be the measurable process defined earlier, and N_λ the integer valued random measure with predictable compensator λ. If we impose the following conditions;*

$$\mathbb{E}\left[\int_0^t h(s,Y(s))^2 ds\right] < \infty \quad \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} \left(\lambda(s,y)^{-1} + \lambda(s,y)\right) \nu(dy)ds\right] < \infty \tag{86}$$

and if the processes Ω, λ and h satisfy the following for all $t \geq 0$,

$$\begin{aligned} \mathbb{E}\left[\int_0^t \Omega_{s-} h(s,Y(s))^2 ds\right] &< \infty, \\ \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} \Omega_{s-}(\lambda(s,y)-1)^2 \nu(dy)ds\right] &< \infty, \end{aligned} \tag{87}$$

then Ω is a martingale.

Proof. By Lemma 3.3.1 and using a similar result to [3, p289] we can see that Ω is a càdlàg local martingale. Using stopping times and Fatou's lemma as in [3, p.289] we can show this is a supermartingale.

Therefore to show that Ω is a martingale we need to calculate its expectation and show that this is constant, see [3, Lemma 5.2.3].

We use the following elementary result that given

$$f(x) = \frac{x}{(1+ax)}$$

then for $c \in \mathbb{R}$

$$f(cx) - f(x) = \frac{(c-1)x}{(1+ax)(1+acx)}.$$

By Theorem 1.5.1, for arbitrary $\varepsilon > 0$ and using the above result,

$$
\begin{aligned}
\frac{\Omega_t}{1+\varepsilon\Omega_t} &= \frac{1}{1+\varepsilon} - \int_0^t \frac{\Omega_{s-}}{(1+\varepsilon\Omega_{s-})^2} h(s,Y(s))dB^Z(s) \\
&- \int_0^t \frac{\varepsilon\Omega_{s-}^2 h(s,Y(s))^2}{(1+\varepsilon\Omega_{s-})^3} ds \\
&+ \int_0^t \int_{\mathbb{R}_0} \frac{\Omega_{s-}(\lambda(s,y)^{-1}-1)}{(1+\varepsilon\lambda(s,y)^{-1}\Omega_{s-})(1+\varepsilon\Omega_{s-})} \tilde{N}_\lambda(ds,dy) \\
&+ \int_0^t \int_{\mathbb{R}_0} \frac{\varepsilon(\lambda(s,y)-1)(\lambda(s,y)^{-1}-1)\Omega_{s-}^2}{(1+\varepsilon\lambda(s,y)^{-1}\Omega_{s-})(1+\varepsilon\Omega_{s-})^2} \nu(dy)ds \qquad (88)
\end{aligned}
$$

Using the square integrability condition of h:

$$
\mathbb{E}\left[\int_0^t \left(\frac{\Omega_{s-}}{(1+\varepsilon\Omega_{s-})^2}\right)^2 h(s,Y(s))^2 ds\right] \le \frac{1}{\varepsilon^2}\mathbb{E}\left[\int_0^t h(s,Y(s))^2 ds\right] < \infty.
$$

Now using condition (86) and the finiteness of ν,

$$
\begin{aligned}
\mathbb{E}&\left[\int_0^t \int_{\mathbb{R}_0} \left|\frac{\Omega_{s-}(\lambda(s,y)^{-1}-1)}{(1+\varepsilon\lambda(s,y)^{-1}\Omega_{s-})(1+\varepsilon\Omega_{s-})}\right|^2 \lambda(s,y)\nu(dy)ds\right] \\
&\le \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} \left|\frac{\Omega_{s-}(\lambda(s,y)^{-1}-1)}{\Omega_{s-}}\right|^2 \lambda(s,y)\nu(dy)ds\right] \\
&= \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (\lambda(s,y)^{-1}-1)^2 \lambda(s,y)\nu(dy)ds\right] \\
&= \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} \left(\lambda(s,y)^{-1}+\lambda(s,y)-2\right)\nu(dy)ds\right] < \infty
\end{aligned}
$$

Hence both the second and fourth terms of (88) are square integrable martingales with zero expectation, therefore if we take expectations we are left with,

$$
\begin{aligned}
\mathbb{E}\left[\frac{\Omega_t}{1+\varepsilon\Omega(t)}\right] &= \frac{1}{1+\varepsilon} - \mathbb{E}\left[\int_0^t \frac{\varepsilon\Omega_{s-}^2 h(s,Y(s))^2}{(1+\varepsilon\Omega_{s-})^3} ds\right] \\
&+ \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} \left|\frac{\varepsilon(\lambda(s,y)-1)(\lambda(s,y)^{-1}-1)\Omega_{s-}^2}{(1+\varepsilon\lambda(s,y)^{-1}\Omega_{s-})(1+\varepsilon\Omega_{s-})^2}\right| \nu(dy)ds\right].
\end{aligned}
$$

We now take limits of the above as $\varepsilon \to 0$. It is clear the first term will yield 1. For convergence of the second term to zero, the reader is directed to [5, p.53].

For the final term we use (87)

$$\mathbb{E}\left[\int_0^t\int_{\mathbb{R}_0}\left|\frac{\varepsilon(\lambda(s,y)-1)(\lambda(s,y)^{-1}-1)\Omega_{s-}^2}{(1+\varepsilon\lambda(s,y)^{-1}\Omega_{s-})(1+\varepsilon\Omega_{s-})^2}\right|\nu(dy)ds\right]$$
$$\leq \mathbb{E}\left[\int_0^t\int_{\mathbb{R}_0}\left|\frac{\varepsilon\Omega_{s-}^2(\lambda(s,y)-1)(\lambda(s,y)^{-1}-1)}{\varepsilon\lambda(s,y)^{-1}\Omega_{s-}}\right|\nu(dy)ds\right]$$
$$= \mathbb{E}\left[\int_0^t\int_{\mathbb{R}_0}\Omega_{s-}(\lambda(s,y)-1)^2\nu(dy)ds\right] < \infty$$

and using dominated convergence we see that,

$$\lim_{\varepsilon\to 0}\mathbb{E}\left[\int_0^t\int_{\mathbb{R}_0}\frac{\varepsilon(\lambda(s,y)-1)(\lambda(s,y)^{-1}-1)\Omega_{s-}^2}{(1+\varepsilon\lambda(s,y)^{-1}\Omega_{s-})(1+\varepsilon\Omega_{s-})^2}\nu(dy)ds\right] = 0.$$

Therefore, the expectation is constant and we have a martingale. □

For the rest of this thesis, we will make the assumptions (86) and (87).

Theorem 3.3.3. *$Z = (Z(t), t \geq 0)$ is a Q Lévy process.*

Proof. We will derive the Lévy-Khintchine formula for the given process.

Firstly using basic principles of martingales, for $u \in \mathbb{R}$

$$\mathbb{E}_Q\left[e^{iuZ(t)}\right] = \mathbb{E}_P\left[e^{iuZ(t)}\Omega_T\right] = \mathbb{E}_P\left[\mathbb{E}_P\left[e^{iuZ(t)}\Omega_T|\mathcal{F}_t\right]\right] = \mathbb{E}_P\left[e^{iuZ(t)}\Omega_t\right].$$

We now apply Itô's product rule to get,

$$d\left(e^{iuZ(t)}\Omega_t\right) = d(e^{iuZ(t)})\Omega_{t-} + e^{iuZ(t-)}d\Omega_t + d\left(e^{iuZ(t)}\right)d\Omega_t.$$

So using Lemma 3.3.1 and (82),

$$
\begin{aligned}
d\left(e^{iuZ(t)}\Omega_t\right) = e^{iuZ(t-)}\Omega_{t-} \Bigg\{ & iudB^Z(t) + iuh(t,Y(t))dt - \frac{1}{2}u^2dt + \\
& + \int_{\mathbb{R}_0} \left(e^{iuy} - 1\right) N_\lambda(dt,dy) \\
& - h(t,Y(t))dB^Z(t) + \int_{\mathbb{R}_0} \left(\lambda(t,y)^{-1} - 1\right) \tilde{N}_\lambda(dt,dy) \\
& - iuh(t,Y(t))dt + \int_{\mathbb{R}_0} \left(e^{iuy} - 1\right)\left(\lambda(t,y)^{-1} - 1\right) N_\lambda(dt,dy) \Bigg\} \\
= e^{iuZ(t-)}\Omega_{t-} \Bigg\{ & (iu - h(t,Y(t))dB^Z(t) - \frac{1}{2}u^2dt \\
& + \int_{\mathbb{R}_0} \left(\lambda^{-1}(t,y)(e^{iuy} - 1)\right) N_\lambda(dt,dy) \\
& + \int_{\mathbb{R}_0} \left(\lambda(t,y)^{-1} - 1\right) \tilde{N}_\lambda(dt,dy) \\
= e^{iuZ(t-)}\Omega_{t-} \Bigg\{ & (iu - h(t,Y(t))dB^Z(t) - \frac{1}{2}u^2dt \\
& + \int_{\mathbb{R}_0} \left(\lambda(t,y)^{-1}e^{iuy} - 1\right) \tilde{N}_\lambda(dt,dy) \\
& + \int_{\mathbb{R}_0} \left(e^{iuy} - 1\right) \nu(dy)dt \Bigg\}.
\end{aligned} \tag{89}
$$

Integrating yields,

$$
\begin{aligned}
e^{iuZ(t)}\Omega_t = & \int_0^t e^{iuZ(s-)}\Omega_{s-}(iu - h(s,Y(s-)))dB^Z(s) \\
& + \int_0^t \int_{\mathbb{R}_0} e^{iuZ(s-)}\Omega_{s-} \left(e^{iuy}\lambda(s,y)^{-1} - 1\right) \tilde{N}_\lambda(ds,dy) \\
& + \int_0^t \int_{\mathbb{R}_0} e^{iuZ(s-)}\Omega_{s-} \left(e^{iuy} - 1\right) \nu(dy)ds - \frac{1}{2}u^2 \int_0^t e^{iuZ(s-)}\Omega_{s-}ds.
\end{aligned} \tag{90}
$$

Now we let $\phi_u(s,t) = \mathbb{E}\left[e^{iuZ(t)}\Omega_t | \mathcal{F}_s\right]$ then,

$$
\phi_u(s,t) = e^{iuZ(s)}\Omega_s + \eta(u) \int_s^t \phi_u(s,r)dr,
$$

where $\eta(u) = -\frac{1}{2}u^2 + \int_{\mathbb{R}_0}(e^{iuy} - 1)\,\nu(dy)$. Now as $e^{iuZ(s)(\omega)}\Omega_s(\omega) \neq 0$ for almost all $\omega \in \Omega$ we can define

$$g(s,t) = \frac{\phi_u(s,t)}{e^{iuZ(s)}\Omega_s} = \mathbb{E}_P\left[e^{iu(Z(t)-Z(s))}\frac{\Omega_t}{\Omega_s}\Big|\mathcal{F}_s\right]$$

and so we have,

$$g(s,t) = 1 + \eta(u)\int_s^t g(s,r)dr.$$

Now, it follows that $t \to g(s,t)$ is differentiable for $s < t$ and $\frac{d}{dt}g(s,t) = \eta(u)g(s,t)$ with initial condition $g(s,s) = 1$. Therefore,

$$g(s,t) = e^{(t-s)\eta(u)}.$$

Then it follows from Lemma 1.1.11 that we have a Lévy process. □

As shown above, we have transformed the observation process into a pure noise process consisting of a Q Brownian motion and a pure jump Lévy process, i.e (82) becomes,

$$dZ(t) = dB^Q(t) + dL(t), \tag{91}$$

where $Z(t)$ is now a Lévy process independent of $Y(t)$ with

$$B^Q(t) = B^Z(t) + \int_0^t h(s, Y(s))ds \tag{92}$$

as the Brownian part (from Girsanov's theorem), and

$$L(t) = \int_0^t \int_{\mathbb{R}_0} yN(ds,dy) \tag{93}$$

as the jump component. Since Z is a Q -Lévy process, N is a Poisson random measure with compensator $\nu(dy)dt$ as has been defined earlier in (16). At this point we calculate the Q expectation of the observation process for each $t \geq 0$:

$$\begin{aligned}\mathbb{E}_Q\left[Z(t)\right] &= \mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} yN(ds,dy)\right]\\ &= \int_0^t \int_{\mathbb{R}_0} y\nu(dy)ds. \end{aligned}\tag{94}$$

Let $\tilde{\Omega}_t = (\tilde{\Omega}_t, t \geq 0)$ be the process defined as $\tilde{\Omega}_t = \Omega_t^{-1}$ i.e,

$$\tilde{\Omega}_t = \exp\left\{\int_0^t h(s, Y(s))dB^Z(s) + \frac{1}{2}\int_0^t h(s, Y(s))^2 ds \right.$$
$$\left. + \int_0^t \int_{\mathbb{R}_0} \log(\lambda(t, y))N_\lambda(ds, dy) + \int_0^t \int_{\mathbb{R}_0} (1 - \lambda(t, y))\,\nu(dy)ds\right\}. \quad (95)$$

When we find an expression for the unnormalised conditional distribution, we will take Q expectations and transform them back to P expectations, therefore we will find the following lemma of use.

Lemma 3.3.4.

$$d\tilde{\Omega}_t = \tilde{\Omega}_{t-}\left\{h(t, Y(t))dB^Q(t) + \int_{\mathbb{R}_0} [\lambda(t, y) - 1]\, N_\lambda(ds, dy)\right.$$
$$\left. - \int_0^t (\lambda(t, y) - 1)\nu(dy)ds\right\} \quad (96)$$

Proof. Using (95),(92) and Itô's formula,

$$d\tilde{\Omega}_t = \tilde{\Omega}_{t-}\left\{h(t, Y(t))dB^Z(t) + \frac{1}{2}h^2(t, Y(t))dt\right\} + \frac{1}{2}\tilde{\Omega}_{t-}h^2(t, Y(t))dt$$
$$+ \tilde{\Omega}_{t-}\int_{\mathbb{R}_0} [1 - \lambda(t, y)]\,\nu(dy)dt + \int_{\mathbb{R}_0}\left[\exp\left\{\eta(t-) + \log\lambda(t, y)\right\} - \tilde{\Omega}_{t-}\right] N_\lambda(ds, dy)$$
$$= \tilde{\Omega}_{t-}\left\{h(t, Y(t))dB^Q(t) + \int_{\mathbb{R}_0} [1 - \lambda(t, y)]\nu(dy)dt\right.$$
$$\left. + \int_{\mathbb{R}_0} [\lambda(t, y) - 1]\, N_\lambda(ds, dy)\right\}$$

□

It is convenient to rewrite (95) from a Q perspective as follows,

$$\tilde{\Omega}_t = \exp\left\{\int_0^t h(s, Y(s))dB_s^Q - \frac{1}{2}\int_0^t h(s, Y(s))^2 ds\right.$$
$$\left. + \int_0^t \int_{\mathbb{R}_0} \log(\lambda)N_\lambda(ds, dy) + \int_0^t \int_{\mathbb{R}_0} (1 - \lambda)\,\nu(dy)ds\right\} \quad (97)$$

The final proposition in this section is important when deriving the non linear filtering equations using the change of measure approach. It shows us that we can replace the σ-algebra $\mathcal{Z}_t$ with a fixed, non time dependent σ-algebra $\mathcal{Z}$. The intriguing interpretation of this is that it tells us the future observations will not affect our best estimate.

Proposition 3.3.5. *Let U be a Q integrable $\mathcal{F}_t$ measurable random variable. Then we have,*

$$\mathbb{E}_Q\left[U|\mathcal{Z}_t\right] = \mathbb{E}_Q\left[U|\mathcal{Z}\right]$$

where

$$\mathcal{Z} = \bigvee_{0\leq t\leq T} \mathcal{Z}_t.$$

$\square$

For proof see [5, Proposition 3.15].

3.4 The Unnormalised Conditional Distribution

We start this section by first stating the Kallianpur-Striebel formula and then go on use this to define the unnormalised conditional distribution.

Theorem 3.4.1. (Kallianpur-Striebel)
For every $\varphi \in C_c^2(\mathbb{R})$ and for fixed $t \in [0, \infty)$,

$$\pi_t(\varphi) = \frac{\mathbb{E}_Q\left[\tilde{\Omega}_t \varphi(Y(t))|\mathcal{Z}\right]}{\mathbb{E}_Q\left[\tilde{\Omega}_t|\mathcal{Z}\right]} \quad P(Q)a.s. \tag{98}$$

$\square$

For proof see [5, Proposition 3.16].
In the following work we will define $(\beta(t), 0 \leq t \leq T)$ as

$$\beta(t) = \mathbb{E}_Q\left[\tilde{\Omega}_t|\mathcal{Z}_t\right]$$

and we will identify it with a càdlàg version which always exists by the argument in [5, p.58]. This leads us nicely to the following definition of the unnormalised conditional distribution.

Definition 3.4.2. *Define the measure valued process $\rho = (\rho_t, t \geq 0)$, to be the unnormalised conditional distribution of the signal Y. It is given by*

$$\rho_t(\varphi) := \pi_t(\varphi)\beta(t),$$

for $\varphi \in B(\mathbb{R})$ and $t \geq 0$.

Lemma 3.4.3. *The process $(\rho_t, t \geq 0)$ is càdlàg and $\mathcal{Z}_t$ adapted. Furthermore, for any $t \geq 0$*

$$\rho_t(\varphi) = \mathbb{E}_Q\left[\tilde{\Omega}_t \varphi(Y(t))\big|\mathcal{Z}_t\right] \quad P(Q)a.s.$$

$\square$

For proof see [16, Lemma 3.18].

It is tempting, in the following, to write $\pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(\mathbf{1})}$ but we avoid this at this stage since it is not clear if $\rho_t(\mathbf{1}) \neq 0$ (a.s).

Corollary 3.4.4. *For $\varphi \in C_c^2(\mathbb{R})$ and $t \geq 0$,*

$$\pi_t(\varphi)\rho_t(1) = \rho_t(\varphi) \quad P(Q)a.s.$$

□

For proof see [5, Corollary 3.19]. It is of vital importance to our filtering equations to find the conditions under which we can pass conditional expectations through an integral. However before we look at this problem we first need the following lemma.

Lemma 3.4.5. *Let $X = (X(t), t \geq 0)$ be a Lévy process on a probability space $(\Omega', \mathcal{F}', P')$ adapted to a given filtration $(\mathcal{F}'_t, t \geq 0)$. Define $S_t = \{(\varepsilon^r_t, t \geq 0) | r \in L^\infty([0,t], \mathbb{R})\}$ where,*

$$\varepsilon^r_t = \exp\left(i \int_0^t r(s)dX(s) - \int_0^t \eta(r(s))ds\right) \tag{99}$$

and $\eta(\cdot)$ is the Lévy symbol corresponding to the process X. Then S_t is a total set in $L^1(\Omega, \mathcal{F}, Q)$, by this we mean if $a \in L^1(\Omega, \mathcal{F}, Q)$ and $\mathbb{E}[a\varepsilon_t] = 0$ for all $\varepsilon \in S_t$ then $a = 0$ a.s. Also, each ε^r_t in S_t satisfies an SDE of the following form,

$$\varepsilon^r_t = 1 + \int_0^t i\varepsilon^r_s r(s)dB(s) + \int_0^t \int_{\mathbb{R}_0} \left(e^{ir(s)y} - 1\right) \varepsilon^r_s \tilde{N}(ds, dy) \tag{100}$$

□

The proof of the first part is a direct generalisation of [5, B.39] and is included in the appendix A.0.6. For the representation in the second part, the reader is referred to [3, Lemma 5.3.3(ii)], where only an application of Itô's formula is required.

The following result is a generalisation of [5, Lemma 3.21].

Lemma 3.4.6. *Let $(u_t, t \geq 0)$ be an $\mathcal{F}_t$ predictable process such that for all $(t \geq 0)$ we have*

$$\mathbb{E}_Q\left[\int_0^t u_s^2 ds\right] < \infty, \tag{101}$$

then for all $t \geq 0$ for our observation process Z defined earlier, we have

$$\mathbb{E}_Q\left[\int_0^t u_s dZ(s) \Big| \mathcal{Z}\right] = \int_0^t \mathbb{E}_Q\left[u_s | \mathcal{Z}\right] dZ(s). \tag{102}$$

Proof. Every ε^r_t from the total set defined in Lemma 3.4.5 satisfies the following stochastic integral equation (100) and so we have, using (91), (94) and the conditional Fubini theorem,

$$
\begin{aligned}
\mathbb{E}_Q\left[\varepsilon_t^r \mathbb{E}_Q\left[\int_0^t u_s dZ(s)\Big|\mathcal{Z}\right]\right] &= \mathbb{E}_Q\left[\varepsilon_t^r \int_0^t u_s dZ(s)\right] \\
&= \mathbb{E}_Q\left[\int_0^t u_s dZ(s)\right] + \mathbb{E}_Q\left[\int_0^t i\varepsilon_s^r r_s u_s ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} y\left(e^{ir(s)y}-1\right)\varepsilon_s^r u_s \nu(dy)ds\right] \\
&= \mathbb{E}_Q\left[\mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} y u_s \nu(dy)ds\Big|\mathcal{Z}\right]\right] \\
&\quad + \mathbb{E}_Q\left[\mathbb{E}_Q\left[\int_0^t i\varepsilon_s^r r_s u_s ds\Big|\mathcal{Z}\right]\right] \\
&\quad + \mathbb{E}_Q\left[\mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} y\left(e^{ir(s)y}-1\right)\varepsilon_s^r u_s \nu(dy)ds\Big|\mathcal{Z}\right]\right] \\
&= \mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} y\mathbb{E}_Q\left[u_s|\mathcal{Z}\right]\nu(dy)ds\right] + \mathbb{E}\left[\int_0^t i\varepsilon_s^r r_s \mathbb{E}\left[u_s|\mathcal{Z}\right]ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} y\left(e^{ir(s)y}-1\right)\varepsilon_s^r \mathbb{E}_Q\left[u_s|\mathcal{Z}\right]\nu(dy)ds\right] \\
&= \mathbb{E}_Q\left[\varepsilon_t^r \int_0^t \mathbb{E}_Q\left[u_s|\mathcal{Z}\right]dZ(s)\right].
\end{aligned}
$$

□

We will need the following lemma for the proof of the Zakai equation.

Lemma 3.4.7. *Let N be a Poisson random measure, with compensator $\tilde{N}(ds,dx) = N(ds,dx) - \nu(dx)ds$, and B an independent standard Brownian motion defined on the probability space $(\Omega', \mathcal{F}', P')$. Let both ψ_n, ψ and ϕ_n, ϕ be predictable mappings such that $\int_0^t \int_{\mathbb{R}_0} |\psi(s,y)|^2\nu(dy)ds < \infty$ a.s. and $\int_0^t \phi(s)^2 ds < \infty$ a.s.*

a) If

$$
\lim_{n\to\infty} \int_0^t \int_{\mathbb{R}_0} |\psi_n(s,y) - \psi(s,y)|^2\nu(dy)ds = 0 \tag{103}
$$

in probability then,

$$
\lim_{n\to\infty} \sup_{0\le t\le T} \left| \int_0^t \int_{\mathbb{R}_0} (\psi_n(s,y) - \psi(s,y))\tilde{N}(ds,dy) \right| = 0
$$

in probability.

b) If

$$\lim_{n\to\infty}\int_0^t |\phi_n(s)-\phi(s)|^2 ds = 0$$

in probability then

$$\lim_{n\to\infty}\sup_{0\le t\le T}\left|\int_0^t (\phi_n(s)-\phi(s))dB(s)\right| = 0$$

in probability.

The proofs of these results follow a similar argument, and the reader is directed to [16, Proposition B.41] for part (*b*) and the appendix A.0.5 for part (*a*).

3.5 The Zakai Equation

We now need to make the following assumptions, for all $t \geq 0$,

$$Q\left[\int_0^t \rho_s \left(h(s,Y(s))^2 ds < \infty\right] = 1. \tag{104}$$

$$Q\left[\int_0^t \int_{\mathbb{R}_0} \left[\rho_t\left(\lambda^{-1}(t,y)+\lambda(t,y)\right)\right]^2 \nu(dy)ds < \infty\right] = 1 \tag{105}$$

The assumption (104) is found in [5] and will be used to deal with the Brownian terms in the upcoming proof. The condition (105) is analogous to the first but will be used to control the jump terms.

Now we are ready to derive the celebrated Zakai equation. Note that our derivation here agrees with the one found in [5, Theorem 3.24] when there is no jump part.

Theorem 3.5.1. (Zakai Equation)
The process $(\rho_t, t \geq 0)$ satisfies the following evolution equation.

$$\begin{aligned}\rho_t(\varphi) = \pi_0(\varphi) &+ \int_0^t \rho_s[A\varphi]ds \\ &+ \int_0^t \rho_s\left[(\varphi(Y(s))h(s,Y(s))\right] dB^Q(s) \\ &+ \int_0^t\int_{\mathbb{R}_0} \rho_s\left[\varphi(Y(s))(\lambda(s,y)-1)\right]\tilde{N}(ds,dy)\end{aligned} \tag{106}$$

Proof. This proof unfolds over the next few pages and is quite delicate. We adapt the technique used in [5, p.62-65], and so our first aim is to approximate $\tilde{\Omega}_t$ by,

$$\tilde{\Omega}_t^\epsilon = \frac{\tilde{\Omega}_t}{1+\epsilon\tilde{\Omega}_t}$$

where $\epsilon > 0$. The second step is to use Itô's formula, integrate and then take conditional expectations, which will require certain conditions being met in order to pass the conditional expectations through the integrals. Finally in our third step we take limits as $\varepsilon \to 0$.

Using Itô's formula we see that

$$d\left(\tilde{\Omega}_t^\epsilon\varphi(Y(t))\right) = d\tilde{\Omega}_t^\epsilon\varphi(Y(t-)) + \tilde{\Omega}_{t-}^\epsilon d\varphi(Y(t)) + d\tilde{\Omega}_t^\epsilon d\varphi(Y(t)),$$

Now using Lemma 3.3.4 and (79),

$$\begin{aligned}
d\left(\tilde{\Omega}_t^\epsilon\varphi(Y(t))\right) &= \tilde{\Omega}_{t-}^\epsilon\left[A\varphi(Y(t-))\right]dt + \tilde{\Omega}_t^\varepsilon dM_t^\varphi \\
&\quad + \varphi(Y(t-))\frac{\tilde{\Omega}_{t-}}{(1+\varepsilon\tilde{\Omega}_{t-})^2}\left\{h(t,Y(t))dB^Q(t) + \int_{\mathbb{R}_0}(1-\lambda(t,y))\nu(dy)dt\right\} \\
&\quad - \frac{\varepsilon\varphi(Y(t-))}{(1+\varepsilon\tilde{\Omega}_{t-})^3}\tilde{\Omega}_{t-}^2 h(t,Y(t))^2 dt \\
&\quad + \int_{\mathbb{R}_0}\frac{\varphi(Y(t))\tilde{\Omega}_{t-}(\lambda(t,y)-1)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{t-})(1+\varepsilon\tilde{\Omega}_{t-})}N(dy,dt) \\
&= \tilde{\Omega}_t^\epsilon\left[A\varphi(Y(t-))\right]dt + \tilde{\Omega}_t^\varepsilon dM_t^\varphi \\
&\quad + \varphi(Y(t-))\frac{\tilde{\Omega}_{t-}}{(1+\varepsilon\tilde{\Omega}_{t-})^2}\left\{h(t,Y(t))dB^Q(t) + \int_{\mathbb{R}_0}(1-\lambda(t,y))\nu(dy)dt\right\} \\
&\quad - \frac{\varepsilon\varphi(Y(t-))}{(1+\varepsilon\tilde{\Omega}_{t-})^3}\tilde{\Omega}_{t-}^2 h(t,Y(t))^2 dt \\
&\quad + \int_{\mathbb{R}_0}\frac{\varphi(Y(t-))\tilde{\Omega}_{t-}(\lambda(t,y)-1)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{t-})(1+\varepsilon\tilde{\Omega}_{t-})}\tilde{N}(dt,dy) \\
&\quad + \int_{\mathbb{R}_0}\frac{\varphi(Y(t))\tilde{\Omega}_{t-}(\lambda(t,y)-1)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{t-})(1+\varepsilon\tilde{\Omega}_{t-})}\nu(dy) \\
&=
\end{aligned} \tag{107}$$

$$
\begin{aligned}
&= \tilde{\Omega}^{\epsilon}_{t-}\left[A\varphi(Y(t-))\right]dt + \tilde{\Omega}^{\varepsilon}_{t} dM^{\varphi}_{t} \\
&\quad + \varphi(Y(t))\frac{\tilde{\Omega}_{t-}}{(1+\varepsilon\tilde{\Omega}_{t-})^2}\left\{h(t,Y(t))dB^{Q}(t)\right\} \\
&\quad - \frac{\varepsilon\varphi(Y(t-))}{(1+\varepsilon\tilde{\Omega}_{t-})^3}\tilde{\Omega}^2_{t-}h(t,Y(t))^2 dt \\
&\quad + \int_{\mathbb{R}_0}\frac{\varphi(Y(t-))\tilde{\Omega}_{t-}(\lambda(t,y)-1)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{t-})(1+\varepsilon\tilde{\Omega}_{t-})}\tilde{N}(dt,dy) \\
&\quad - \int_{\mathbb{R}_0}\frac{\varphi(Y(t-))\varepsilon(\lambda(t,y)-1)^2\tilde{\Omega}^2_{t-}}{(1+\varepsilon\tilde{\Omega}_{t-})^2(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{t-})}\nu(dy)dt
\end{aligned} \tag{108}
$$

For convenience we now label the terms in (108) from (1) to (6). We need to integrate over $[0,t]$ and pass conditional expectations through the integrals.

For the first term since $\varphi \in C^2_c(\mathbb{R})$ and $\tilde{\Omega}^{\varepsilon}_t$ is bounded then it is clear that,

$$
\mathbb{E}_Q\left[\int_0^t \mathbb{E}_Q\left[\left|\tilde{\Omega}^{\varepsilon}_s\varphi(Y(s-))\right|\mathcal{Z}\right]ds\right] < \infty
$$

and so by the conditional form of Fubini's theorem we have,

$$
\mathbb{E}_Q\left[\int_0^t \tilde{\Omega}^{\varepsilon}_s\varphi(Y(s-))ds\big|\mathcal{Z}\right] = \int_0^t \mathbb{E}_Q\left[\tilde{\Omega}^{\varepsilon}_s\varphi(Y(s-))\big|\mathcal{Z}\right]ds
$$

For the second term we note that since $\tilde{\Omega}^{\varepsilon}_t$ is bounded,

$$
\mathbb{E}_Q\left[\int_0^t \tilde{\Omega}^{\varepsilon}_s dM^{\varphi}_s\big|\mathcal{Z}\right] = 0
$$

For terms (3) and (5) we will use Lemma 3.4.6 to take conditional expectations and pass them through the integrals, but in order to do this we must first satisfy condition (101).

To this end, we require the following elementary inequality,

$$
\frac{1}{1+\alpha} \leq \min\left\{1, \frac{1}{\alpha}\right\} \quad \text{for all} \quad \alpha \geq 0. \tag{109}
$$

Now by Itô's isometry,

$$
\begin{aligned}
&\mathbb{E}_Q\left[\left(\int_0^t\int_{\mathbb{R}_0}\frac{\varphi(Y(s))\tilde{\Omega}_{s-}(\lambda(s,y)-1)}{(1+\varepsilon\lambda(s,y)\tilde{\Omega}_{s-})(1+\varepsilon\tilde{\Omega}_{s-})}\tilde{N}(ds,dy)\right)^2\right]\\
&\qquad\leq\frac{||\varphi||_\infty^2}{\varepsilon^2}\mathbb{E}_Q\left[\int_0^t\int_{\mathbb{R}_0}(\lambda(s,y)-1)^2\nu(dy)ds\right]\\
&\qquad=\frac{||\varphi||_\infty^2}{\varepsilon^2}\mathbb{E}\left[\int_0^t\int_{\mathbb{R}_0}\Omega_{s-}(\lambda(s,y)-1)^2\nu(dy)ds\right]\\
&\qquad<\infty.
\end{aligned}\tag{110}
$$

The proof of

$$
\mathbb{E}\left[\left(\int_0^t\varphi(Y(s))\frac{\tilde{\Omega}_{s-}}{(1+\varepsilon\tilde{\Omega}_{s-})^2}h(s,Y(s))dB^Q(s)\right)^2\right]<\infty
$$

follows from (87), and can be seen in [5, p.62].

For term (6) we need the result of the following in order to use the conditional form of Fubini's theorem. So, using (109), (87) and the tower property:

$$
\begin{aligned}
&\mathbb{E}_Q\left[\int_0^t\int_{\mathbb{R}_0}\mathbb{E}_Q\left[\left|\frac{\varphi(Y(s))\varepsilon(\lambda(s,y)-1)^2\tilde{\Omega}_{s-}^2}{(1+\varepsilon\tilde{\Omega}_{s-})^2(1+\varepsilon\lambda(s,y)\tilde{\Omega}_{s-})}\right|\Big|\mathcal{Z}\right]\nu(dy)ds\right]\\
&\qquad=\mathbb{E}_Q\left[\int_0^t\int_{\mathbb{R}_0}\left|\frac{\varphi(Y(s))\varepsilon(\lambda(s,y)-1)^2\tilde{\Omega}_{s-}^2}{(1+\varepsilon\tilde{\Omega}_{s-})^2(1+\varepsilon\lambda(s,y)\tilde{\Omega}_{s-})}\right|\nu(dy)ds\right]\\
&\qquad\leq\varepsilon||\varphi||_\infty\mathbb{E}_Q\left[\int_0^t\int_{\mathbb{R}_0}\frac{(\lambda(s,y)-1)^2\tilde{\Omega}_{s-}^2}{(1+\varepsilon\tilde{\Omega}_{s-})^2(1+\varepsilon\lambda(s,y)\tilde{\Omega}_{s-})}\nu(dy)ds\right]\\
&\qquad\leq\varepsilon||\varphi||_\infty\mathbb{E}_Q\left[\int_0^t\int_{\mathbb{R}_0}\frac{(\lambda(s,y)-1)^2\tilde{\Omega}_{s-}^2}{(\varepsilon\tilde{\Omega}_{s-})^2}\nu(dy)ds\right]\\
&\qquad=\frac{||\varphi||_\infty}{\varepsilon}\mathbb{E}_Q\left[\int_0^t\int_{\mathbb{R}_0}(\lambda(s,y)-1)^2\nu(dy)ds\right]\\
&\qquad=\frac{||\varphi||_\infty}{\varepsilon}\mathbb{E}\left[\int_0^t\int_{\mathbb{R}_0}\Omega_{s-}(\lambda(s,y)-1)^2\nu(dy)ds\right]<\infty
\end{aligned}
$$

Finally all that remains is to show that we can use the conditional form of Fubini's theorem on term (4). Using (109) and (87),

$$
\begin{aligned}
&\mathbb{E}_Q\left[\int_0^t \mathbb{E}_Q\left[\left|\frac{\varepsilon\varphi(Y(s-))}{(1+\varepsilon\tilde{\Omega}_{s-})^3}\tilde{\Omega}^2_{s-}h(s,Y(s))^2\right|\Big|\mathcal{Z}\right]ds\right] \\
&\quad= \mathbb{E}_Q\left[\int_0^t \left|\frac{\varepsilon\varphi(Y(s-))}{(1+\varepsilon\tilde{\Omega}_{s-})^3}\tilde{\Omega}^2_{s-}h(s,Y(s))^2\right|ds\right] \\
&\quad\le \varepsilon||\varphi||_\infty\mathbb{E}_Q\left[\int_0^t \left|\frac{\tilde{\Omega}^2_{s-}}{(1+\varepsilon\tilde{\Omega}_{s-})^3}h(s,Y(s))^2\right|ds\right] \\
&\quad\le \frac{||\varphi||_\infty}{\varepsilon}\mathbb{E}_Q\left[\int_0^t h(s,Y(s))^2 ds\right] \\
&\quad= \frac{||\varphi||_\infty}{\varepsilon}\mathbb{E}\left[\int_0^t \Omega_{s-}h(s,Y(s))^2 ds\right] < \infty.
\end{aligned}
$$

Now returning to (108), using the results just proved, integrating and applying Lemma 3.4.6

$$
\begin{aligned}
\mathbb{E}_Q\left[\tilde{\Omega}^\varepsilon_t\varphi(Y(t))|\mathcal{Z}\right] &= \frac{\pi_0}{1+\varepsilon} + \int_0^t \mathbb{E}_Q\left[\tilde{\Omega}^\varepsilon_s A\varphi(Y(s))|\mathcal{Z}\right]ds \\
&+ \int_0^t \mathbb{E}_Q\left[\varphi(Y(s))\frac{\tilde{\Omega}_{s-}}{(1+\varepsilon\tilde{\Omega}_{s-})^2}h(s,Y(s))\Big|\mathcal{Z}\right]dB(s) \\
&+ \int_0^t \mathbb{E}_Q\left[\frac{\varepsilon\varphi(Y(t))}{(1+\varepsilon\tilde{\Omega}_{t-})^3}\tilde{\Omega}^2_{t-}h(t,Y(t))^2\Big|\mathcal{Z}\right]dt \\
&+ \int_0^t\int_{\mathbb{R}_0} \mathbb{E}_Q\left[\frac{\varphi(Y(s))\tilde{\Omega}_{s-}(\lambda(t,y)-1)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{s-})(1+\varepsilon\tilde{\Omega}_{s-})}\Big|\mathcal{Z}\right]\tilde{N}(ds,dy) \\
&- \int_0^t\int_{\mathbb{R}_0} \mathbb{E}_Q\left[\frac{\varphi(Y(s))\varepsilon(\lambda(t,y)-1)^2\tilde{\Omega}^2_{s-}}{(1+\varepsilon\tilde{\Omega}_{s-})^2(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{s-})}\Big|\mathcal{Z}\right]\nu(dy)ds.
\end{aligned} \tag{111}
$$

We now wish to let ε tend to 0 so that $\lim_{\varepsilon\to 0}\tilde{\Omega}^\varepsilon_t = \tilde{\Omega}_t$ $P(Q)$ a.s. Using Lemma 3.4.3 and the conditional form of the dominated convergence theorem (see [20, p.397] for statement and proof) it is shown in [5, p.62] that

$$
\int_0^t \mathbb{E}_Q\left[\tilde{\Omega}^\varepsilon_s A\varphi(Y(s))|\mathcal{Z}\right] \to \int_0^t \rho_s\left[A\varphi(Y(s))\right]ds.
$$

The proof that

$$\lim_{\varepsilon\to 0}\int_0^t \mathbb{E}\left[\varphi(Y(s))\frac{\tilde{\Omega}_{s-}}{(1+\varepsilon\tilde{\Omega}_{s-})^2}h(s,Y(s))\Big|\mathcal{Z}\right]dB(s)$$
$$=\int_0^t \rho_s\left[\varphi(Y(s))h\left(s,Y(s)\right)\right]dB(s)$$

and

$$\lim_{\varepsilon\to 0}\int_0^t \mathbb{E}_Q\left[\frac{\varepsilon\varphi(Y(s))}{(1+\varepsilon\tilde{\Omega}_{s-})^3}\tilde{\Omega}_{s-}^2 h(s,Y(s))^2\Big|\mathcal{Z}\right]ds=0$$

can also be found in [5, p.64].

The next stage of this proof requires us to show that the last term in (111) tends to zero as $\varepsilon\to 0$.

To justify the use of dominated convergence we first compute

$$\int_0^t\int_{\mathbb{R}_0}\left|\frac{\varphi(Y(s))\varepsilon(\lambda(s,y)-1)^2\tilde{\Omega}_{s-}^2}{(1+\varepsilon\tilde{\Omega}_{s-})^2(1+\varepsilon\lambda(s,y)\tilde{\Omega}_{s-})}\right|\nu(dy)ds$$
$$<\varepsilon||\varphi||_\infty\int_0^t\int_{\mathbb{R}_0}\left|\frac{\left(\lambda(s,y)-1\right)^2\tilde{\Omega}_{s-}^2}{(\varepsilon\tilde{\Omega}_{s-})^2}\right|\nu(dy)ds$$
$$=\frac{||\varphi||_\infty}{\varepsilon}\int_0^t\int_{\mathbb{R}_0}\left|(\lambda(s,y)-1)^2\right|\nu(dy)ds.$$

The right hand side of the above can seen to be Q-integrable using (87) to obtain,

$$\mathbb{E}_Q\left[\int_0^t\int_{\mathbb{R}_0}(\lambda(t,y)-1)^2\nu(dy)ds\right]=\mathbb{E}\left[\int_0^t\int_{\mathbb{R}_0}\Omega_s(\lambda(t,y)-1)^2\nu(dy)ds\right]<\infty$$

and so we use the conditional form of dominated convergence again to show that

$$\lim_{\varepsilon\to 0}\int_0^t\int_{\mathbb{R}_0}\mathbb{E}_Q\left[\frac{\varphi(Y(s))\varepsilon(\lambda(t,y)-1)^2\tilde{\Omega}_{s-}^2}{(1+\varepsilon\tilde{\Omega}_{s-})^2(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{s-})}\Big|\mathcal{Z}\right]\nu(dy)ds$$
$$=\int_0^t\int_{\mathbb{R}_0}\mathbb{E}_Q\left[\lim_{\varepsilon\to 0}\frac{\varphi(Y(s))\varepsilon(\lambda(t,y)-1)^2\tilde{\Omega}_{s-}^2}{(1+\varepsilon\tilde{\Omega}_{s-})^2(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{s-})}\Big|\mathcal{Z}\right]\nu(dy)ds$$
$$=0. \tag{112}$$

Our last step in this proof is to show that as $\varepsilon \to 0$

$$\int_0^t \int_{\mathbb{R}_0} \mathbb{E}_Q \left[\frac{\varphi(Y(s))\tilde{\Omega}_{s-}(\lambda(t,y)-1)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{s-})(1+\varepsilon\tilde{\Omega}_{s-})} \Big| \mathcal{Z} \right] \tilde{N}(ds,dy)$$
$$\to \int_0^t \int_{\mathbb{R}_0} \rho_s \left[\varphi(Y(s))(\lambda(t,y)-1) \right] \tilde{N}(ds,dy). \qquad (113)$$

We aim to show this convergence using Lemma 3.4.7. First, we consider the local martingale

$$t \to \int_0^t \int_{\mathbb{R}_0} \mathbb{E}_Q \left[\frac{\varphi(Y(s))\tilde{\Omega}_{s-}(\lambda(t,y)-1)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{s-})(1+\varepsilon\tilde{\Omega}_{s-})} \Big| \mathcal{Z} \right] \tilde{N}(ds,dy) \qquad (114)$$

and we show that this is square integrable. By Itô's isometry, the conditional Jensen inequality, Fubini's theorem and (87),

$$\begin{aligned}
&\mathbb{E}_Q \left[\left(\int_0^t \int_{\mathbb{R}_0} \mathbb{E}_Q \left[\frac{\varphi(Y(s))\tilde{\Omega}_{s-}(\lambda(s,y)-1)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{s-})(1+\varepsilon\tilde{\Omega}_{s-})} \Big| \mathcal{Z} \right] \tilde{N}(ds,dy) \right)^2 \right] \\
&\quad = \mathbb{E}_Q \left[\int_0^t \int_{\mathbb{R}_0} \left(\mathbb{E}_Q \left[\frac{\varphi(Y(s))\tilde{\Omega}_{s-}(\lambda(s,y)-1)}{(1+\varepsilon\lambda(s,y)\tilde{\Omega}_{s-})(1+\varepsilon\tilde{\Omega}_{s-})} \Big| \mathcal{Z} \right] \right)^2 \nu(dy)ds \right] \\
&\quad \leq \mathbb{E}_Q \left[\int_0^t \int_{\mathbb{R}_0} \mathbb{E}_Q \left[\left(\frac{\varphi(Y(s))\tilde{\Omega}_{s-}(\lambda(s,y)-1)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_{s-})(1+\varepsilon\tilde{\Omega}_{s-})} \right)^2 \Big| \mathcal{Z} \right] \nu(dy)ds \right] \\
&\quad \leq \frac{||\varphi||_\infty^2}{\varepsilon^2} \int_0^t \int_{\mathbb{R}_0} \mathbb{E}_Q \left[(\lambda(s,y)-1)^2 \right] \nu(dy)ds \\
&\quad = \frac{||\varphi||_\infty^2}{\varepsilon^2} \int_0^t \int_{\mathbb{R}_0} \mathbb{E} \left[\Omega_s(\lambda(s,y)-1)^2 \right] \nu(dy)ds \\
&\quad < \infty
\end{aligned}$$

and so the process in (114) is a martingale. We postulate that the limit process as $\varepsilon \to 0$ is the well defined local martingale,

$$\begin{aligned}
t \to &\int_0^t \int_{\mathbb{R}_0} \mathbb{E}_Q \left[\tilde{\Omega}_s \varphi(Y(s)(\lambda(t,y)-1) | \mathcal{Z} \right] \tilde{N}(ds,dy) \\
&= \int_0^t \int_{\mathbb{R}_0} \rho_s \left[\varphi(Y(s)(\lambda(s,y)-1)) \right] \tilde{N}(ds,dy)
\end{aligned}$$

First observe that the difference of the above with (114) is also a local martingale taking the form,

$$t \to \int_0^t \int_{\mathbb{R}_0} \mathbb{E}_Q \left[\frac{\varphi(Y(s))(1-\lambda(s,y))\tilde{\Omega}_s^2(1+\lambda(s,y)+\varepsilon\lambda(s,y)\tilde{\Omega}_s)\varepsilon}{(1+\varepsilon\lambda(s,y)\tilde{\Omega}_s)(1+\varepsilon\tilde{\Omega}_s)} \Big| \mathcal{Z} \right] \tilde{N}(dy,ds). \tag{115}$$

We now use Lemma 3.4.7 to prove that the integral above converges to 0 Q-a.s.

We compute,

$$\begin{aligned}
&\left| \frac{\varphi(Y(t))(\lambda(t,y)-1)\tilde{\Omega}_s^2(1+\lambda(t,y)+\varepsilon\lambda(t,y)\tilde{\Omega}_s)\varepsilon}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_s)(1+\varepsilon\tilde{\Omega}_s)} \right| \\
&\qquad \leq ||\varphi||_\infty \left| \frac{\varepsilon(\lambda(t,y)-1)\tilde{\Omega}_s^2}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_s)(1+\varepsilon\tilde{\Omega}_s)} + \frac{\varepsilon(\lambda(t,y)-1)\tilde{\Omega}_s^2(1+\varepsilon\tilde{\Omega}_s)\lambda(t,y)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_s)(1+\varepsilon\tilde{\Omega}_s)} \right| \\
&\qquad \leq ||\varphi||_\infty \left(\left| \frac{\varepsilon(\lambda(t,y)-1)\tilde{\Omega}_s^2}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_s)(1+\varepsilon\tilde{\Omega}_s)} \right| + \left| \frac{\varepsilon(\lambda(t,y)-1)\tilde{\Omega}_s^2\lambda(t,y)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_s)} \right| \right) \\
&\qquad \leq ||\varphi||_\infty \left(\left| \frac{\varepsilon(\lambda(t,y)-1)\tilde{\Omega}_s^2}{\varepsilon\lambda(t,y)\tilde{\Omega}_s} \right| + \left| \frac{\varepsilon(\lambda(t,y)-1)\tilde{\Omega}_s^2\lambda(t,y)}{\varepsilon\lambda(t,y)\tilde{\Omega}_s} \right| \right) \\
&\qquad = ||\varphi||_\infty \left(\left| (1-\lambda(t,y)^{-1})\tilde{\Omega}_s \right| + \left| (\lambda(t,y)-1)\tilde{\Omega}_s \right| \right) \\
&\qquad \leq ||\varphi||_\infty \tilde{\Omega}_s \left(\lambda^{-1}(t,y) + \lambda(t,y) + 2 \right).
\end{aligned} \tag{116}$$

We now show the right-hand side of (116) is Q integrable a.e. First observe that

$$\mathbb{E}_Q \left[\tilde{\Omega}_s \left(\lambda^{-1}(t,y) + \lambda(t,y) + 2 \right) \right] = 2 + \mathbb{E}\left[\lambda^{-1}(t,y) + \lambda(t,y) \right].$$

By Fubini's theorem and (86)

$$\mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (\lambda(s,y) + \lambda^{-1}(s,y))\nu(dy)ds \right] = \int_0^t \int_{\mathbb{R}_0} \mathbb{E}\left[(\lambda(s,y) + \lambda^{-1}(s,y)) \right] \nu(dy)ds$$
$$< \infty$$

and this implies that $\mathbb{E}\left[(\lambda(t,y) - \lambda^{-1}(s,y))\right] < \infty$ a.e ($\mathcal{L}eb \times \nu$).

Using the dominated convergence theorem once more we can conclude that for almost every $t \geq 0$ and $y \in \mathbb{R}$,

$$\lim_{\varepsilon\to 0}\mathbb{E}_Q\left[\frac{\varphi(Y(t))(\lambda(t,y)-1)\tilde{\Omega}_t^2(1+\lambda(t,y)+\varepsilon\lambda(t,y)\tilde{\Omega}_t)\varepsilon}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_t)(1+\varepsilon\tilde{\Omega}_t)}\right] = 0 \quad Q\text{-a.s.} \tag{117}$$

Using (116) we have,

$$\begin{aligned}\left|\mathbb{E}_Q\left[\frac{\varphi(Y(s))(\lambda(t,y)-1)\tilde{\Omega}_s^2(1+\lambda(t,y)+\varepsilon\lambda(t,y)\tilde{\Omega}_s)\varepsilon}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_s)(1+\varepsilon\tilde{\Omega}_s)}\Big|\mathcal{Z}\right]\right| \\ \leq ||\varphi||_\infty\left|\mathbb{E}_Q\left[\tilde{\Omega}_s\left(\lambda^{-1}(t,y)+\lambda(t,y)+2\right)|\mathcal{Z}\right]\right| \\ = ||\varphi||_\infty\left|\rho_t\left(\lambda^{-1}(t,y)+\lambda(t,y)+2\right)\right|\end{aligned}$$

and therefore using (105) it follows that Q-a.s,

$$\begin{aligned}\int_0^t\int_{\mathbb{R}_0}\left(\mathbb{E}_Q\left[\varepsilon\frac{\varphi(Y(s))(\lambda(t,y)-1)\tilde{\Omega}_s^2(1+\lambda(t,y)+\varepsilon\lambda(t,y)\tilde{\Omega}_s)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_s)(1+\varepsilon\tilde{\Omega}_s)}\Big|\mathcal{Z}\right]\right)^2\nu(dy)ds \\ \leq ||\varphi||_\infty^2\int_0^t\int_{\mathbb{R}_0}\left[\rho_t\left(\lambda^{-1}(t,y)+\lambda(t,y)+2\right)\right]^2\nu(dy)ds \\ < \infty.\end{aligned}$$

Now using dominated convergence, we can show that,

$$\begin{aligned}&\lim_{\varepsilon\to 0}\int_0^t\int_{\mathbb{R}_0}\left(\mathbb{E}_Q\left[\varepsilon\,\frac{\varphi(Y(s))(\lambda(t,y)-1)\tilde{\Omega}_s^2(1+\lambda(t,y)+\varepsilon\lambda(t,y)\tilde{\Omega}_s)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_s)(1+\varepsilon\tilde{\Omega}_s)}\Big|\mathcal{Z}\right]\right)^2\nu(dy)ds \\ =&\int_{\mathbb{R}_0}\left(\mathbb{E}_Q\left[\lim_{\varepsilon\to 0}\int_0^t\varepsilon\,\frac{\varphi(Y(s))(\lambda(t,y)-1)\tilde{\Omega}_s^2(1+\lambda(t,y)+\varepsilon\lambda(t,y)\tilde{\Omega}_s)}{(1+\varepsilon\lambda(t,y)\tilde{\Omega}_s)(1+\varepsilon\tilde{\Omega}_s)}\Big|\mathcal{Z}\right]\right)^2\nu(dy)ds \\ =& 0 \quad Q-\text{a.s.}\end{aligned} \tag{118}$$

We have now satisfied condition (103) and so as a consequence of the above we can now use Lemma 3.4.7 to establish the convergence proposed in (113).

□

3.6 The Kushner Stratonovich Equation

In the previous section we derived the form for the unnormalised conditional distribution. We now wish to extend this by finding an equation for the normalised version π.

In this section we will require the following assumptions. They can however be shown to be implied by the stronger conditions (86) and (87) stated earlier.

$$P\left(\int_0^t \pi_s\left(h(s,Y(s))\right)^2 < \infty\right) = 1 \tag{119}$$

$$P\left(\int_0^t \int_{\mathbb{R}_0} \pi_s(\lambda(s,y)-1)\nu(dy)ds < \infty\right) = 1 \tag{120}$$

Before moving on to the derivation of the normalised conditional distribution, we see from Corollary 3.4.4 that in order to find $\pi_t(\varphi)$, we will be required to divide by $\rho_t(\mathbf{1})$. We already know from (3.4.3) that $\rho_t(\mathbf{1}) = \mathbb{E}_Q\left[\tilde{\Omega}_t|\mathcal{Z}\right]$. Since $\tilde{\Omega}_t > 0$ for all t, it also follows that $\mathbb{E}_Q\left[\tilde{\Omega}_t|\mathcal{Z}\right] > 0$ a.s. and so, $\rho_t(\mathbf{1}) > 0$ a.s for all $t > 0$. However this result is not enough, and in fact we need to show the following,

$$P\left(\rho_t(1) > 0, \forall t > 0\right) = 1$$

We give thanks to Aleksandar Mijatović for his guidance on the proof below.

Lemma 3.6.1. *Given any càdlàg martingale $M = (M_t, t \geq 0)$ such that $M_t > 0$ a.s for all $t > 0$ then,*

$$P\left(M_t > 0, \forall t > 0\right) = 1$$

Proof. Let $\tau = \inf\{t > 0; M_t = 0\}$ and assume τ is bounded then $P(M_\tau = 0) = 1$. We define the set,

$$A_s = \tau^{-1}\left([0,s]\right), s > 0$$

and define the stopped σ-algebra,

$$\mathcal{Z}_\tau = \left\{A \in \mathcal{F}; A \cap (\tau \leq t) \in \mathcal{Z}_t, \forall t \geq 0\right\}.$$

Then if we let $b > \tau$ we note that

$$M(\tau) = 0 \text{ (a.s)} \implies \mathbb{E}\left[M_\tau 1_{A_s}\right] = 0 \implies \mathbb{E}\left[1_{A_s}\mathbb{E}\left[M_b|\mathcal{Z}_\tau\right]\right] = 0$$

by the optional stopping theorem, and since $A_s \in \mathcal{Z}_\tau$. Therefore with probability 1

$$1_{A_s}\mathbb{E}\left[M_b|\mathcal{Z}_\tau\right] = 0.$$

It follows that,

$$\lim_{s\to\infty} \mathbb{E}\left[1_{A_s}M_b|\mathcal{Z}_\tau\right] = 0$$

and we note that $1_{A_s}M_b \uparrow 1_{A_\infty}M_b$, where $A_\infty = \tau^{-1}[0,\infty) = \{\tau < \infty\}$. Using the conditional version of the monotone convergence theorem,

$$\mathbb{E}\left[1_{A_\infty}M_b|\mathcal{Z}_\tau\right] = 0$$

and finally taking expectations of the above we have,

$$\mathbb{E}\left[1_{\{\tau<\infty\}}M_b\right] = 0$$

which in turn implies that $1_{\{\tau<\infty\}}M_b = 0$ a.s., i.e. that $M_b(\omega) = 0$ for all ω in a set of positive measure. Now assume that τ is unbounded. Since

$$\{\tau < \infty\} = \bigcup_{n\in\mathbb{N}} \{n-1 \le \tau < n\}$$

if $P(\tau < \infty) > 0$, then at least one of the sets $A_n = \{n-1 \le \tau < n\}$ must have positive measure. Let A_n be such a set. Then if $b > n$ the same argument as above for bounded stopping times shows that

$$1_{A_n}M(b) = 0,$$

on a set of positive measure, and we have our desired contradiction.

□

Using the above lemma it is now easy to see that $\rho_t(1) > 0$ for all t with probability 1.

Lemma 3.6.2. *Given assumptions (86),(87), (105) and (104) then the process $t \to \rho_t(\mathbf{1})$ has the following representation,*

$$\begin{aligned}\rho_t(1) = \exp\Bigg\{ &\int_0^t \pi_s(h(s,Y(s)))dB^Q(s) \\ &- \int_0^t \frac{1}{2}\pi_s^2(h(s,Y(s)))ds \\ &+ \int_0^t \int_{\mathbb{R}_0} \log(\pi_s(\lambda)N(ds,dy) \\ &- \int_0^t \int_{\mathbb{R}_0} \pi_s(\lambda(s,y)-1)\nu(dy)ds \Bigg\}\end{aligned} \qquad (121)$$

Proof. Since h is not necessarily bounded it is not trivial to see that h is integrable with respect to π_t, i.e that $\pi_t(h)$ exists. However conditions (86) and (87) ensures that it is defined $\mathcal{L}eb \otimes P$-a.s.

Using the Zakai equation and the identity $A1 = 0$ it follows that $\rho_t(1)$ satisfies the following identity,

$$\begin{aligned}\rho_t(1) = 1 &+ \int_0^t \rho_s[h(s, Y(s))]dB^Q(s) \\ &+ \int_0^t \int_{\mathbb{R}_0} \rho_s[(\lambda(s, y) - 1)]\tilde{N}(dy, ds)\end{aligned}$$

If we now apply Corollary 3.4.4 we can rewrite the above as,

$$\begin{aligned}\rho_t(1) = 1 &+ \int_0^t \rho_s(1)\pi_s(h(s, Y(s)))dB^Q(s) \\ &+ \int_0^t \int_{\mathbb{R}_0} \rho_s(1)\pi_s(\lambda(s, y) - 1)\tilde{N}(dy, ds)\end{aligned} \tag{122}$$

we now apply Itô's formula to $\log(\rho_t(1))$.

$$\begin{aligned}d\log(\rho_t(1)) = &\left(\frac{1}{\rho_t(1)}\right)\rho_t(1)\pi_t(h)dB^Q(t) \\ &- \frac{1}{2}\frac{1}{\rho_t(1)^2}\left(\pi_t(h)\rho_t(1)\right)^2 dt \\ &+ \int_{\mathbb{R}_0} [\log\left(\rho_t(1) + \rho_t(1)\pi_t(\lambda(t, y) - 1)\right) \\ &\quad - \log\left(\rho_t(1)\right)]\,\tilde{N}(dy, dt) \\ &+ \int_{\mathbb{R}_0} \Bigg[\log\left(\rho_t(1) + \rho_t(1)\pi_t(\lambda(t, y) - 1)\right) \\ &\quad - \log\left(\rho_t(1)\right) - \frac{\rho_t(1)\pi_t(\lambda(t, y) - 1)}{\rho_t(1)}\Bigg]\nu(dy)dt \\ = &\,\pi_t(h)dB^Q(t) - \frac{1}{2}\pi_t(h)^2 dt \\ &+ \int_{\mathbb{R}_0} \log(\pi_t(h))N(dy, dt) - \int_{\mathbb{R}_0} \pi_t(\lambda(t, y) - 1)\nu(dy)ds\end{aligned} \tag{123}$$

Now integrating and taking exponentials gives the required result. $\square$

In the work of [5, p.67] the authors do not use Lemma 3.6.1, instead an approximation argument similar to the one in the Zakai equation is used. This is needed because the author cannot use Itô's formula directly since $x \to \log(x)$ is not defined at $x = 0$, and if it was not for the previous lemma we would not know that $\rho_t(1) > 0$ for all $t \geq 0$ a.s. It is however never negative which allows the following approximation. This works well when the observation noise is Gaussian, however when we add jump noise the result does not follow so easily. We have decided to include partial workings here to warn others against following Alice down the rabbit-hole [11].

During the rest of the proof, we will take h to mean, $(h(t, Y(t)), t \geq 0)$. By using Itô's formula applied to (122) we get,

$$
\begin{aligned}
d\left(\log\sqrt{\varepsilon + \rho_t(1)^2}\right) = &\frac{\rho_t(1)^2}{(\rho_t(1)^2 + \varepsilon)}\pi_s(h)dB^Q(t) \\
&+ \frac{1}{2}\frac{\varepsilon - \rho_t(1)^2\pi_t(h)^2}{(\rho_t(1)^2 + \varepsilon)^2}dt \\
&+ \int_{\mathbb{R}_0} \frac{1}{2}\log\left(\frac{\varepsilon + [\rho_t(1)(1 + \pi_t(\lambda(s,y) - 1))]^2}{\varepsilon + \rho_t(1)^2}\right)\tilde{N}(dy, dt) \\
&+ \int_{\mathbb{R}_0} \left[\frac{1}{2}\log\left(\frac{\varepsilon + [\rho_t(1)(1 + \pi_t(\lambda(s,y) - 1))]^2}{\varepsilon + \rho_t(1)^2}\right)\right. \\
&\left. -\pi_t(\lambda(s,y) - 1)\frac{\rho_t(1)^2}{(\rho_t(1)^2 + \varepsilon)}\right]\nu(dy)dt.
\end{aligned}
\tag{124}
$$

Noting that

$$
\begin{aligned}
\rho_t(1)(1 + \pi_t(\lambda(s,y) - 1)) &= \rho_t(1) + \rho_t(\lambda(t,y) - 1) \\
&= \rho_t(\lambda(t,y)).
\end{aligned}
$$

we can simplify the above to,

$$
\begin{aligned}
d\left(\log\sqrt{\varepsilon+\rho_t(1)^2}\right) &= \frac{\rho_t(1)^2}{(\rho_t(1)^2+\varepsilon)}\pi_s(h)dB^Q(t)+\frac{1}{2}\frac{\varepsilon-\rho_t(1)^2\pi_t(h)^2}{(\rho_t(1)^2+\varepsilon)^2}dt \\
&+\int_{\mathbb{R}_0}\frac{1}{2}\log\left(\frac{\varepsilon+\rho_t(\lambda(s,y))^2}{\varepsilon+\rho_t(\mathbf{1})^2}\right)\tilde{N}(dy,dt) \\
&+\int_{\mathbb{R}_0}\left[\frac{1}{2}\log\left(\frac{\varepsilon+\rho_t(\lambda(s,y))^2}{\varepsilon+\rho_t(1)^2}\right)\right. \\
&\left.-\pi_t(\lambda(s,y)-1)\frac{\rho_t(1)^2}{(\rho_t(1)^2+\varepsilon)}\right]\nu(dy)dt
\end{aligned}
\tag{125}
$$

The first and second terms above are expertly dealt with in [5], however in order to control the jump terms this author suggests an assumption such as the following would be required,

$$
\int_0^t\int_{\mathbb{R}_0}\left[\max\left\{\frac{1+\rho_s(\lambda(s,y))^2}{\rho_s(1)^2},\frac{1+\rho_s(1)^2}{\rho_s(\lambda(s,y))^2}\right\}\right]\nu(dy)ds<\infty.
$$

We now continue from Lemma 3.6.2 to derive the normalised conditional distribution.

Theorem 3.6.3. **(The Kushner-Stratonovich Equation)** *If conditions (86), (87), (119) and (120) hold, then the conditional distribution of the signal π_t satisfies the following evolution equation called the Kushner-Stratonovich equation.*

$$
\begin{aligned}
\pi_t(\varphi) &= \pi_0(\varphi)+\int_0^t\pi_s(A\varphi)ds \\
&+\int_0^t\left[\pi_s(\varphi h)-\pi_s(h)\pi_s(\varphi)\right]\left[dB^Q(s)-\pi_s(h)ds\right] \\
&+\int_0^t\int_{\mathbb{R}_0}\left[\pi_s(\lambda(s,y)^{-1})\left(\pi_s(\varphi(\lambda(s,y))-\pi_s(\varphi)+1\right)-1\right]N(dt,dy) \\
&+\int_0^t\int_{\mathbb{R}_0}\left[\pi_t(\lambda(t,y)-1)-\pi_t(\varphi(\lambda(t,y)-1))\right]\nu(dy)dt
\end{aligned}
$$

for any $\varphi\in C_c^2(\mathbb{R})$.

Proof. From Lemma 3.6.2 we obtain,

$$\frac{1}{\rho_t(1)} = \exp\left\{-\int_0^t \pi_s(h)dB^Q(s) + \frac{1}{2}\int_0^t \pi_s(h)^2 ds \right.$$
$$\left. - \int_0^t \int_{\mathbb{R}_0} \log(\pi_t(\lambda(s,y)))N(dt,dy) + \int_0^t \int_{\mathbb{R}_0} \pi_t(\lambda(s,y)-1)\nu(dy)dx\right\}$$

and so,

$$d\left(\frac{1}{\rho_t(1)}\right) = \frac{1}{\rho_t(\mathbf{1})}\left[-\pi_t(h)dB^Q(t) + \pi_t^2(h)dt + \int_{\mathbb{R}_0} \pi_t(\lambda(s,y)-1)\nu(dy)dt \right.$$
$$\left. + \int_{\mathbb{R}_0} \left(\pi_t(\lambda(s,y))^{-1} - 1\right) N(dt,dy)\right]$$

Now using the Zakai equation (106) for $\rho_t(\varphi)$ and the Kallianpur-Striebel formula (98), we obtain by Itô's product formula,

$$\pi_t(\varphi) = \rho_t(\varphi)\cdot\frac{1}{\rho_t(1)}$$
$$d\left(\pi_t(\varphi)\right) = d\left(\rho_t(\varphi)\right)\frac{1}{\rho_{t-}(1)} + \rho_{t-}(\varphi)d\left(\frac{1}{\rho_t(1)}\right) + d\left(\rho_t(\varphi)\right)d\left(\frac{1}{\rho_t(1)}\right)$$

Breaking these parts up for easier digestion,

1)

$$d\left(\rho_t(\varphi)\right)\frac{1}{\rho_t(1)}$$
$$= \left(\rho_t(A\varphi)dt + \rho_t(\varphi h)dB^Q(t) + \int_{\mathbb{R}_0} \rho_t(\varphi(\lambda(t,y)-1))\tilde{N}(dt,dy)\right)\frac{1}{\rho_t(1)}$$
$$= \pi_t(A\varphi)dt + \pi_t(\varphi h)dB^Q(t) + \int_{\mathbb{R}_0} \pi_t(\varphi(\lambda(s,y)-1))\tilde{N}(ds,dy)$$

2)

$$\begin{aligned}
&\rho_t(\varphi) d\left(\frac{1}{\rho_t(1)}\right)\\
&\quad= \frac{\rho_t(\varphi)}{\rho_t(1)}\left[-\pi_t(h)dB^Q(t) + \pi_t^2(h)dt + \int_{\mathbb{R}_0} \pi_t(\lambda(t,y)-1)\nu(dy)dt\right.\\
&\qquad\left. + \int_{\mathbb{R}_0} \left(\pi_t(\lambda(t,y))^{-1} - 1\right) N(dt,dy)\right]\\
&\quad= \pi_t(\varphi)\left[-\pi_t(h)dB^Q(t) + \pi_t^2(h)dt + \int_{\mathbb{R}_0} \pi_t(\lambda(t,y)-1)\nu(dy)dt\right.\\
&\qquad\left. + \int_{\mathbb{R}_0} \left(\pi_t(\lambda(t,y))^{-1} - 1\right) N(dt,dy)\right]
\end{aligned}$$

3)

$$\begin{aligned}
&d\left(\rho_t(\varphi)\right) d\left(\frac{1}{\rho_t(1)}\right)\\
&\quad= \left(\rho_t(A\varphi)dt + \rho_t(\varphi h)dB^Q(t) + \int_{\mathbb{R}_0} \rho_t(\varphi(\lambda(s,y)-1))\tilde{N}(dt,dy)\right)\\
&\qquad\times \frac{1}{\rho(1)}\left[-\pi_t(h)dB^Q(t) + \pi_t^2(h)dt + \int_{\mathbb{R}_0} \pi_t(\lambda(s,y)-1)\nu(dy)dt\right.\\
&\qquad\qquad\left. + \int_{\mathbb{R}_0} \left(\pi_t(\lambda(s,y))^{-1} - 1\right) N(dt,dy)\right]\\
&\quad= -\pi_t(\varphi h)\pi_t(h)dt + \int_{\mathbb{R}_0} \pi_t(\varphi(\lambda(s,y)-1)) \times \left(\pi_t(\lambda(s,y))^{-1} - 1\right) N(dt,dy)
\end{aligned}$$

Now combining these back together we get,

$$
\begin{aligned}
d\left(\pi_t(\varphi)\right) &= \pi_t(A\varphi)dt + \pi_t(\varphi h)dB^Q(t) + \int_{\mathbb{R}_0} \pi_t(\varphi(\lambda(t,y)-1))\tilde{N}(dt,dy) \\
&\quad + \pi_t(\varphi)\Bigg[-\pi_t(h)dB^Q(t) + \pi_t^2(h)dt + \int_{\mathbb{R}_0} \pi_t(\lambda(t,y)-1)\nu(dy)dt \\
&\quad + \int_{\mathbb{R}_0} \left(\pi_t(\lambda(t,y))^{-1} - 1\right) N(dt,dy)\Bigg] \\
&\quad - \pi_t(\varphi h)\pi_t(h)dt + \int_{\mathbb{R}_0} \pi_t(\varphi(\lambda(s,y)-1))(\pi_t(\lambda(s,y))^{-1}-1)N(dt,dy) \\
&= \pi_t(A\varphi)dt + \pi_t(\varphi)\pi_t^2(h)dt - \pi_t(\varphi h)\pi_t(h)dt \\
&\quad + \pi_t(\varphi h)dB^Q(t) - \pi_t(\varphi)\pi_t(h)dB^Q(t) \\
&\quad + \int_{\mathbb{R}_0} \big\{\big[\pi_t(\varphi(\lambda(t,y)-1)) + \big(\pi_t(\lambda(t,y))^{-1}-1\big)\big] \\
&\quad + \pi_t(\varphi(\lambda(s,y)-1))(\pi_t(\lambda(s,y))^{-1}-1)\big\} N(dt,dy) \\
&\quad + \int_{\mathbb{R}_0} \pi_t(\lambda(t,y)-1) - \pi_t(\varphi(\lambda(t,y)-1))\nu(dy)dt
\end{aligned}
$$

Finally integrating and simplifying gives,

$$
\begin{aligned}
\pi_t(\varphi) &= \pi_0(\varphi) + \int_0^t \pi_s(A\varphi)ds \\
&\quad + \int_0^t \left[\pi_s(\varphi h) - \pi_s(h)\pi_s(\varphi)\right]\left[dB^Q(s) - \pi_s(h)ds\right] \\
&\quad + \int_0^t \int_{\mathbb{R}_0} \left[\pi_s(\lambda(s,y)^{-1})\left(\pi_s(\varphi(\lambda(s,y)) - \pi_s(\varphi) + 1\right) - 1\right] N(dt,dy) \\
&\quad + \int_0^t \int_{\mathbb{R}_0} \left[\pi_t(\lambda(t,y)-1) - \pi_t(\varphi(\lambda(t,y)-1))\right]\nu(dy)dt
\end{aligned}
$$

$\square$

4 Uniqueness of the Solution to the Zakai Equation

In this chapter we prove the uniqueness of the solution to the Zakai equation (106) by following a similar methodology to [59, Chapter 6] which in turn is based on [37]. We achieve this by transforming the solution to that of a stochastic differential equation in a Hilbert space and subsequently using estimates based on Hilbert-space theory.

To establish this transformation, we will need some well known results about Hilbert spaces and we refer the reader to [18] for details.

We also need to introduce the space of finite signed measures on $\mathbb{R}$ which is denoted by $\mathcal{M}_S(\mathbb{R})$. Given any $v \in \mathcal{M}_S(\mathbb{R})$, the Jordan decomposition [14, Corollary 4.15] states that there exists two positive measures v^+ and v^- (at least one of which is finite) so that $v = v^+ - v^-$. We now define the variation of the measure v as the positive measure $|v|$ as $|v| = v^+ + v^-$.

We also define the *total variation* norm by $||v|| = |v|(X)$, and by [14, Proposition 4.1.7], $\mathcal{M}_S(\mathbb{R})$ is complete under the total variation norm.

4.1 Transformation to a Hilbert Space

We start this section by recalling that for all $\varphi \in C_b(\mathbb{R}), t \geq 0$,

$$\mathbb{E}\left[\varphi(Y(t))|\mathcal{Z}_t\right] = \int_{\mathbb{R}} \varphi(x)\pi_t(dx)$$

where π_t is the conditional probability measure. By (98), $\mathbb{E}_Q\left[\Omega_t\varphi(Y(t))|\mathcal{Z}_t\right]$ is the measure valued solution to the Zakai equation. Therefore it can be written as $\int_{\mathbb{R}} \varphi(x)V_t(dx)$ for some conditional measure V_t and some $\varphi \in \mathbb{C}_c^2(\mathbb{R})$. We can therefore rewrite the Kallianpur-Striebel formula (98) as,

$$\mathbb{E}\left[\varphi(Y(t))|\mathcal{Z}_t\right] = \int_{\mathbb{R}} \varphi(x)\pi_t(dx) = \frac{\int_{\mathbb{R}} \varphi(x)V_t(dx)}{V_t(\mathbb{R})}$$

Now, from the Zakai equation we have,

$$
\begin{aligned}
\rho_t(\varphi) &= \int_{\mathbb{R}} \varphi(x) V_t(dx) \\
&= \int_{\mathbb{R}} \varphi(x)\pi_0(dx) + \int_0^t \int_{\mathbb{R}} A\varphi(x) V_s(dx) ds \\
&+ \int_0^t \int_{\mathbb{R}} \varphi(x) h(s,x) V_s(dx) dB(s) \\
&+ \int_0^t \int_{\mathbb{R}_0} \int_{\mathbb{R}} \varphi(x)(\lambda(s,y)-1) V_s(dx) \tilde{N}(ds,dy). \qquad (126)
\end{aligned}
$$

Recall that $L^2(\mathbb{R})$ is the Hilbert space of square integrable functions on $\mathbb{R}$. To obtain relevant estimates to study uniqueness we will need to transform an $\mathcal{M}_S(\mathbb{R})$-valued process to an $L^2(\mathbb{R})$-valued process.

For the rest of this chapter we will also require the following assumptions.

Assumption 4.1.1. *The mapping h defined earlier is bounded and uniformly Lipschitz, i.e. for all $x_1, x_2 \in \mathbb{R}, t \geq 0$ there exists $K > 0$ such that*

$$|h(t,x_1) - h(t,x_2)| \leq K|x_1 - x_2|$$

and the predictable compensator λ satisfies

$$\int_{R_0} |\lambda(t,x_1,y) - \lambda(t,x_2,y)| \nu(dy) \leq K|x_1 - x_2|.$$

For $\delta \geq 0$ we define the operator $T_\delta : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ by,

$$T_\delta \phi(x) = \int_{\mathbb{R}} G_\delta(x-y)\phi(y) dy \quad \forall \phi \in L^2(\mathbb{R}),$$

where G_δ is the heat kernel given by,

$$G_\delta(x) = (2\pi\delta)^{-\frac{1}{2}} \exp\left(-\frac{|x|^2}{2\delta}\right) \text{ for all } x \in \mathbb{R}.$$

For any $\gamma \in \mathcal{M}_S(\mathbb{R})$ and $\delta > 0$, let,

$$(T_\delta\gamma)(x) = \int_{\mathbb{R}} G_\delta(x-y)\gamma(dy).$$

Lemma 4.1.2. *For all $\delta > 0$ $T_\delta : \mathcal{M}_S(\mathbb{R}) \to C_b(\mathbb{R})$.*

Proof. For boundedness we note that,

$$\begin{aligned}\sup_{x\in\mathbb{R}}|T_\delta(\gamma)(x)| &= \sup_{x\in\mathbb{R}}\left|\int_{\mathbb{R}} G_\delta(x-y)\gamma(dy)\right| \\ &\le (2\pi\delta)^{-\frac{1}{2}}||\gamma|| \\ &< \infty.\end{aligned}$$

Let $x \in \mathbb{R}$ and $(x_n, n \in \mathbb{N})$ be a sequence such that $x_n \to x$ as $n \to \infty$. Let $\gamma \in \mathcal{M}_S(\mathbb{R})$ then,

$$T_\delta(\gamma)(x) - T_\delta(\gamma)(x_n) = \int_{\mathbb{R}} \left(G_\delta(x-y) - G_\delta(x_n-y)\right)\gamma(dy).$$

We can use dominated convergence to show that,

$$\begin{aligned}&\lim_{x_n\to x}\int_{\mathbb{R}} \left(G_\delta(x-y) - G_\delta(x_n-y)\right)\gamma(dy) \\ &= \int_{\mathbb{R}}\left(G_\delta(x-y) - \lim_{x_n\to x} G_\delta(x_n-y)\right)\gamma(dy) \\ &= 0\end{aligned}$$

□

The following is standard in semigroup theory, for a proof see [59, Lemma 6.6].

Lemma 4.1.3. *The operators $(T_t, t \ge 0)$ form a contraction semigroup on $L^2(\mathbb{R})$ i.e. for all $t, s \ge 0$ and $\phi \in L^2(\mathbb{R})$,*

$$T_{t+s} = T_t T_s \quad \textit{and} \quad ||T_t\phi|| \le ||\phi||.$$

The following results will also be required.

Lemma 4.1.4. *If $\gamma \in \mathcal{M}_S(\mathbb{R})$ and $\delta > 0$ then;*

1) $T_\delta\gamma \in L^2(\mathbb{R})$,

2) $||\ T_{2\delta}|\gamma|\ || \le ||\ T_\delta|\gamma|\ ||$.

See [59, Lemma 6.7] for proof.

The following result is key, since it shows how we can transfer the measure valued solution to an SDE in a Hilbert space.

Lemma 4.1.5. *For any $t \ge 0$, $\gamma \in \mathcal{M}_S(\mathbb{R})$ and $f \in L^2(\mathbb{R})$ we have,* *1)*

$$\langle T_\delta \gamma, f\rangle = \int_{\mathbb{R}} T_\delta f(x)\gamma(dx).$$

2) If f is such that $\mathcal{D}f \in L^2(\mathbb{R})$ then,

$$\mathcal{D}T_\delta f = T_\delta \mathcal{D} f$$

where $\mathcal{D}f = \frac{df}{dx}$.

For proof see [59, Lemma 6.8].

If we replace φ with $T_\delta\varphi$ and write $Z_s^\delta = T_\delta V_s$ where V_s is an $\mathcal{M}_S(\mathbb{R})$ valued solution to the Zakai equation, using Lemma 4.1.5 (1) we obtain from (126):

$$\begin{aligned}\langle Z_t^\delta, \varphi\rangle &= \int_{\mathbb{R}} T_\delta\varphi(x)V_t(dx)\\ &= \int_{\mathbb{R}} T_\delta\varphi(x)\pi_0(dx) + \int_0^t\int_{\mathbb{R}} AT_\delta\varphi(x)V_s(dx)ds\\ &\quad + \int_0^t\int_{\mathbb{R}} T_\delta\varphi(x)h(s,Y(s))V_s(dx)dB(s)\\ &\quad + \int_0^t\int_{\mathbb{R}_0}\int_{\mathbb{R}} T_\delta\varphi(x)(\lambda(s,y)-1)V_s(dx)\tilde{N}(ds,dy). \end{aligned} \tag{127}$$

Recall from (80) that A is a diffusion operator, then using Lemma 4.1.5 (2) we obtain,

$$\begin{aligned}\int_{\mathbb{R}} AT_\delta\varphi(x)V_t(dx) &= \int_{\mathbb{R}}\left(\frac{1}{2}\sigma^2\mathcal{D}^2(T_\delta\varphi(x)) + b\mathcal{D}(T_\delta\varphi(x))\right)V_t(dx)\\ &= \frac{1}{2}\int_{\mathbb{R}}\sigma^2 T_\delta\mathcal{D}^2\varphi(x)V_t(dx) - \int_{\mathbb{R}} bT_\delta\mathcal{D}\varphi(x)V_t(dx).\end{aligned}$$

Now using Lemma 4.1.5 (1) we get,

$$\begin{aligned}\int_{\mathbb{R}} AT_\delta\varphi(x)V_t(dx) &= \frac{1}{2}\left\langle T_\delta(\sigma^2V_t), \mathcal{D}^2\varphi\right\rangle - \langle T_\delta(bV_t), \mathcal{D}\varphi\rangle\\ &= \frac{1}{2}\left\langle \mathcal{D}^2T_\delta(\sigma^2V_t), \varphi\right\rangle + \langle \mathcal{D}T_\delta(bV_t), \varphi\rangle.\end{aligned}$$

Similarly we can show that,

$$\int_{\mathbb{R}} T_\delta\varphi(x)h(t,x)V_t(dx) = \langle T_\delta(h(t,Y(t))V_t),\varphi\rangle$$

and for all $y \in \mathbb{R}$,

$$\int_{\mathbb{R}} T_\delta\varphi(x)(\lambda(t,y)-1)V_t(dx) = \langle T_\delta((\lambda(t,y)-1)V_t),\varphi\rangle$$

Inserting these back into the the Zakai equation we get,

$$\begin{aligned}
\langle Z_t^\delta,\varphi\rangle &= \langle Z_0^\delta,\varphi\rangle + \frac{1}{2}\int_0^t \left\langle \mathcal{D}^2 T_\delta(\sigma^2 V_s),\varphi\right\rangle ds \\
&\quad + \int_0^t \langle \mathcal{D}T_\delta(bV_s),\varphi\rangle\, ds + \int_0^t \langle T_\delta(h(s,Y(s))V_s),\varphi\rangle\, dB^Q(s) \\
&\quad + \int_0^t \int_{\mathbb{R}_0} \langle T_\delta((\lambda(s,y)-1)V_s),\varphi\rangle\, \tilde{N}(ds,dy)
\end{aligned}$$

We now seek the stochastic differential of $\langle Z_t^\delta,\varphi\rangle^2$. By Itô's product formula we have,

$$d\langle Z_t^\delta,\varphi\rangle^2 = 2\langle Z_{t-}^\delta,\varphi\rangle d\langle Z_t^\delta,\varphi\rangle + d\langle Z_t^\delta,\varphi\rangle d\langle Z_t^\delta,\varphi\rangle.$$

Now using Itô's formula we get,

$$\begin{aligned}
d\langle Z_t^\delta,\varphi\rangle &= \frac{1}{2}\langle \mathcal{D}^2 T_\delta(\sigma^2 V_t),\varphi\rangle dt + \langle \mathcal{D}T_\delta(bV_t),\varphi\rangle dt \\
&\quad + \langle T_\delta(h(t,Y(t))V_t,\varphi\rangle dB^Q(t) \\
&\quad + \int_{\mathbb{R}_0} \big(\langle Z_{t-}^\delta + T_\delta(\lambda(t,y)-1)V_t,\varphi\rangle - \langle Z_{t-}^\delta,\varphi\rangle\big)\, \tilde{N}(dt,dy) \\
&\quad + \int_{\mathbb{R}_0} \Big(\langle Z_{t-}^\delta + T_\delta(\lambda(t,y)-1)V_t,\varphi\rangle - \langle Z_{t-}^\delta,\varphi\rangle \\
&\quad - \langle T_\delta(\lambda(t,y)-1)V_t,\varphi\rangle\Big)\nu(dy,dt), \qquad (128)
\end{aligned}$$

and the Itô correction term is,

$$\begin{aligned}
d\langle Z_t^\delta,\varphi\rangle d\langle Z_t^\delta,\varphi\rangle &= \langle T_\delta(h(t,Y(t)))V_t,\varphi\rangle^2 dt \\
&\quad + \int_{\mathbb{R}_0} \langle T_\delta(\lambda(t,y)-1)V_t,\varphi\rangle^2 N(dt,dy).
\end{aligned}$$

Combining the above and integrating gives,

$$
\begin{aligned}
\left\langle Z_t^\delta,\varphi\right\rangle^2 = \langle Z_0^\delta,\varphi\rangle^2 &+ \int_0^t \langle Z_s^\delta,\varphi\rangle_0 \left\langle \mathcal{D}^2 T_\delta(\sigma^2 V_s),\varphi\right\rangle ds \\
&+ \int_0^t 2\langle Z_s^\delta,\varphi\rangle\langle \mathcal{D}T_\delta(bV_s),\varphi\rangle ds \\
&+ \int_0^t 2\langle Z_s^\delta,\varphi\rangle\langle T_\delta(h(s,Y(s))V_s),\varphi\rangle dB^Q(s) \\
&+ \int_0^t \langle T_\delta(h(s,Y(s))V_s),\varphi\rangle^2 ds \\
&+ \int_0^t \int_{\mathbb{R}_0} 2\langle Z_{s-}^\delta,\varphi\rangle \left\langle T_\delta\left((\lambda(s,y)-1)V_s\right),\varphi\right\rangle \tilde{N}(ds,dy) \\
&+ \int_0^t \int_{\mathbb{R}_0} \left\langle T_\delta\left((\lambda(s,y)-1)V_s\right),\varphi\right\rangle^2 N(ds,dy). \qquad (129)
\end{aligned}
$$

Summing over φ in a complete orthonormal system of $L^2(\mathbb{R})$ and using Parseval's formula we get,

$$
\begin{aligned}
||Z_t^\delta||^2 = ||Z_0^\delta||^2 &+ \int_0^t \left\langle Z_s^\delta, \mathcal{D}^2 T_\delta(\sigma^2 V_s)\right\rangle ds \\
&+ \int_0^t 2\langle Z_s^\delta, \mathcal{D}T_\delta(bV_s)\rangle ds \\
&+ \int_0^t 2\left\langle Z_s^\delta, T_\delta(h(s,Y(s))V_s)\right\rangle dB(s) \\
&+ \int_0^t ||T_\delta(h(s,Y(s))V_s)||^2 ds \\
&+ \int_0^t \int_{\mathbb{R}_0} 2\left\langle Z_{s-}^\delta, T_\delta((\lambda(s,y)-1)V_s)\right\rangle \tilde{N}(ds,dy) \\
&+ \int_0^t \int_{\mathbb{R}_0} ||T_\delta\left((\lambda(s,y)-1)V_s\right)||^2 N(ds,dy).
\end{aligned}
$$

Taking expectations we obtain,

$$
\begin{aligned}
\mathbb{E}_Q\left[||Z_t^\delta||^2\right] &= \mathbb{E}_Q\left[||Z_0^\delta||^2\right] + \mathbb{E}\left[\int_0^t \left\langle Z_s^\delta, \mathcal{D}^2 T_\delta(\sigma^2 V_s)\right\rangle ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t 2\langle Z_s^\delta, \mathcal{D} T_\delta(bV_s)\rangle ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t 2\left\langle Z_s^\delta, T_\delta(h(s,Y(s))V_s)\right\rangle dB^Q(s)\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t ||T_\delta(h(s,Y(s))V_s)||^2\, ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} 2\left\langle Z_s^\delta, T_\delta((\lambda(s,y)-1)V_s)\right\rangle \tilde{N}(ds,dy)\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} ||T_\delta\left((\lambda(s,y)-1)V_s\right)||^2 N(ds,dy)\right] \\
&= \mathbb{E}_Q\left[||Z_0^\delta||^2\right] + \mathbb{E}\left[\int_0^t \left\langle Z_s^\delta, \mathcal{D}^2 T_\delta(\sigma^2 V_s)\right\rangle ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t 2\langle Z_s^\delta, \mathcal{D} T_\delta(bV_s)\rangle ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t ||T_\delta(h(s,Y(s))V_s)||^2\, ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} ||T_\delta\left((\lambda(s,y)-1)V_s\right)||^2 \nu(dy)ds\right] \qquad (130)
\end{aligned}
$$

To pass the expectations through the integrals requires a dominated convergence argument that will be dealt with once we have introduced the necessary estimates in the next section.

4.2 Useful Inequalities

Let $f_i : \mathbb{R} \to \mathbb{R}, i = 1, 2$ be bounded Lipschitz continuous functions such that there exists $K > 0$ so that,

$$|f_i(x) - f_i(y)| \leq K|x - y| \quad \forall x, y \in \mathbb{R} \quad \text{and} \quad |f_i(x)| \leq K \ \forall x \in \mathbb{R}.$$

Lemma 4.2.1. *Suppose that $g \in L^2(\mathbb{R})$ is such that $\mathcal{D}g \in L^2(\mathbb{R})$. Then,*

$$|\langle g, f_1 \mathcal{D} g\rangle| \leq \frac{1}{2} K ||g||^2.$$

For proof see [59, Lemma 6.9].

Note that for $\zeta \in \mathcal{M}_S(\mathbb{R}), f_1\zeta$ is an $\mathbb{R}$ - valued signed measure.

Lemma 4.2.2. *$T_\delta(f_1\cdot) : \mathcal{M}_S(\mathbb{R}) \to L^2(\mathbb{R})$ for all $\delta > 0$.*

Proof. For all $\zeta \in \mathcal{M}_S(\mathbb{R})$,

$$\begin{aligned}
\int_{\mathbb{R}} |T_\delta(f_1\zeta)|^2\, dx &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} G_\delta(x-y) f_1(y)\zeta(dy) \right|^2 dx \\
&\leq \int_{\mathbb{R}} \left(\int_R G_\delta(x-y)|\zeta|(dy) \right) \\
&\quad \times \left(\int_{\mathbb{R}} G_\delta(x-y)|f_1(y)|^2|\zeta|dy \right) dx \\
&\leq (2\pi\delta)^{-1/2}||\zeta||\ ||f_1||_\infty^2 \int_{\mathbb{R}} \int_{\mathbb{R}} G_\delta(x-y)dx|\zeta|(dy) \\
&\leq (2\pi\delta)^{1/2}||f_1||_\infty^2||\zeta||^2 < \infty.
\end{aligned}$$

□

Lemma 4.2.3. *There exists K_1 and K_2 dependent on K such that for any $\zeta \in \mathcal{M}_G(\mathbb{R})$,*

$$||T_\delta(f_1\zeta)|| \leq ||T_\delta(|f_1| \cdot |\zeta|)|| \leq K||T_\delta(|\zeta|)|| \tag{131}$$

$$|\langle T_\delta(f_2\zeta), \mathcal{D}T_\delta(f_1\zeta)\rangle| \leq K_2||T_\delta(|\zeta|)||^2 \tag{132}$$

see [59, Lemma 6.10] for proof.

Lemma 4.2.4. *There exists a constant K_1 such that for any $\zeta \in \mathcal{M}_S(\mathbb{R})$ we have,*

$$\langle T_\delta\zeta, \mathcal{D}^2(c^2\zeta)\rangle + ||\mathcal{D}T_\delta(c\zeta)||^2 \leq K_1||T_\delta(|\zeta|)||^2.$$

see [59, Lemma 6.11] for proof.

Carrying on from (130), and using the above estimates

$$\begin{aligned}
\mathbb{E}_Q\left[||Z_t^\delta||^2\right] &= \mathbb{E}_Q\left[||Z_0^\delta||^2\right] + \mathbb{E}_Q\left[\int_0^t \left\langle Z_s^\delta, \mathcal{D}^2 T_\delta(\sigma^2 V_s)\right\rangle ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t 2\langle Z_s^\delta, \mathcal{D} T_\delta(bV_s)\rangle ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t ||T_\delta(h(s,Y(s))V_s)||^2\, ds\right] \\
&\quad + \mathbb{E}_Q\left[\int_0^t \int_{\mathbb{R}_0} ||T_\delta\left((\lambda(s,y)-1)V_s\right)||^2 \nu(dy) ds\right] \\
&\leq ||Z_0^\delta||_0^2 + K\mathbb{E}_Q\left[\int_0^t ||T_\delta(|V_s|)||^2 ds\right] \\
&\leq ||Z_0^\delta||_0^2 + K\mathbb{E}_Q\left[\int_0^t ||\ |V_s|\ ||^2 ds\right] < \infty \qquad (133)
\end{aligned}$$

By dominated convergence,

$$\begin{aligned}
\mathbb{E}_Q\left[||Z_t^\delta||^2\right] &= \mathbb{E}_Q\left[||Z_0^\delta||^2\right] + \int_0^t \mathbb{E}_Q\left[\left\langle Z_s^\delta, \mathcal{D}^2 T_\delta(\sigma^2 V_s)\right\rangle\right] ds \\
&\quad + \int_0^t 2\mathbb{E}_Q\left[\langle Z_s^\delta, \mathcal{D} T_\delta(bV_s)\rangle\right] ds \\
&\quad + \int_0^t \mathbb{E}_Q\left[||T_\delta(h(s,Y(s))V_s)||^2\right] ds \\
&\quad + \int_0^t \int_{\mathbb{R}_0} \mathbb{E}_Q\left[||T_\delta\left((\lambda(s,y)-1)V_s\right)||^2\right] \nu(dy) ds \qquad (134)
\end{aligned}$$

4.3 Uniqueness for the Zakai Equation

Theorem 4.3.1. *If V is a $\mathcal{M}_S(\mathbb{R})$ valued solution to the Zakai equation and $Z^\delta = T_\delta V$ then,*

$$\mathbb{E}_Q\left[||Z_t^\delta||^2\right] \leq ||Z_0^\delta||^2 + K_1 \int_0^t \mathbb{E}_Q\left[||T_\delta(|V_s|)||^2\right] ds$$

where K_1 is a constant.

Proof. Looking at the estimate (134) the first four terms are dealt with as in [59, Theorem 6.12], so all that is required is to note that the last term is bounded by a constant multiplied by $||T_\delta(|V_s|)||^2$ using (4.2.4). □

Let $\mathcal{M}_{\mathcal{B}}(\mathbb{R})$ be the space of finite Borel measures on $\mathbb{R}$.

The following corollary requires some explanation in that it states that given a measure valued process $(V_t, 0 \le t \le T)$, such that $V_t \in \mathcal{M}_{\mathcal{B}}(\mathbb{R})$ for each t, we can find $\tilde{V}_t \in L^2(\mathbb{R})$ such that for arbitrary $f \in L^2(\mathbb{R})$,

$$\langle \tilde{V}_t, f \rangle = \int_{\mathbb{R}} f(x) V_t(dx) \ \text{ a.s..} \tag{135}$$

From now on, when we write $V_t \in L^2(\mathbb{R})$ it should be understood in the sense of (135), i.e. we identify the measure V_t with the class of functions $\tilde{V}_t$. As a consequence of Theorem 4.3.1 we have the following corollary.

Corollary 4.3.2. *If V is a solution to the Zakai equation taking values in the space of Borel measures $\mathcal{M}_{\mathcal{B}}(\mathbb{R})$, and $V_0 \in L^2(\mathbb{R})$, then $V_t \in L^2(\mathbb{R})$ a.s. and $\mathbb{E}_Q[||V_t||^2] < \infty$ for all $t \ge 0$.*

The following theorem is a generalisation of [59, Theorem 6.14]

Theorem 4.3.3. *Suppose that $V_0 \in L^2(\mathbb{R})^+$. Then the Zakai equation has at most one $\mathcal{M}_B(\mathbb{R})$ valued solution.*

Proof. Let V_t^1 and V_t^2 be two measure valued solutions with the same initial value V_0. By Corollary 4.3.2, V_t^1 and $V_t^2 \in L^2(\mathbb{R})$ a.s. Let $V_t = V_t^1 - V_t^2$, then $V_t \in L^2(\mathbb{R})$ a.s. and

$$\mathbb{E}_Q\left[||T_\delta V_t||^2\right] \le \int_0^t \mathbb{E}_Q\left[||T_\delta(|V_s|)||^2\right] ds.$$

To apply dominated convergence we require the following, using Lemma 4.1.3,

$$\begin{aligned}\int_0^t \mathbb{E}_Q\left[||T_\delta(|V_s|)||^2\right] ds &\le \int_0^t \mathbb{E}_Q\left[||V_s||^2\right] ds \\ &< \infty.\end{aligned}$$

Now as we let $\delta \to 0$,

$$\mathbb{E}_Q\left[||V_t||^2\right] \le K_1 \int_0^t \mathbb{E}_Q\left[|| \ |V_s| \ ||^2\right] = K_1 \int_0^t \mathbb{E}_Q\left[||V_s||^2\right] ds.$$

Then, via an application of Gronwall's inequality we see that V_t must be zero a.s. □

Given that we have now shown that the Zakai equation has at most one solution, we can use this to show that the Kushner Stratonovich equation has a unique solution using the following theorem.

The following is an adaptation of [5, Theorem 4.19].

Corollary 4.3.4. *The Kushner-Stratonovich equation has at most one probability measure valued solution.*

Proof. Let π^1 and π^2 be two distinct probability measure valued solutions to the Kushner Stratonovich equation in Theorem 3.6.3. From the proof of Theorem 3.6.3, using Itô's product formula, we have for each π_i for $i = 1, 2$, $\varphi \in C_c^2(\mathbb{R})$,

$$\pi^i(\varphi)\rho^i(1) = \rho^i(\varphi).$$

We know from Theorem 4.3.3 that the Zakai equation has a unique solution such that ρ^1 and ρ^2 coincide. Therefore,

$$\rho_t^1(1) = \rho_t^2(1) \quad \text{(a.s.)}$$

for all $t \geq 0$. Hence for all $t \geq 0$,

$$\pi_t^1(\varphi) = \frac{\rho_t^1(\varphi)}{\rho_t^1(1)} = \frac{\rho_t^2(\varphi)}{\rho_t^2(1)} = \pi_t^2(\varphi) \quad \text{(a.s.)}.$$

□

A Proof of Lemma 3.4.7

Lemma A.0.5. *Let N be a Poisson random measure defined on $\mathbb{R}^+ \times \mathbb{R}_0$ with compensator $\tilde{N}$ and intensity measure ν where ν is a Lévy measure and let ψ_n and ψ be predictable mappings such that $\int_0^t \int_{\mathbb{R}_0} |\psi(s,y)|^2 \nu(dy)ds < \infty$ a.s and*

$$\lim_{n\to\infty} \int_0^t \int_{\mathbb{R}_0} |\psi_n(s,y) - \psi(s,y)|^2 \nu(dy)ds = 0$$

in probability then

$$\lim_{n\to\infty} \sup_{0\le t\le T} \left| \int_0^t \int_{\mathbb{R}_0} (\psi_n(s,y) - \psi(s,y)) \tilde{N}(ds,dx) \right| = 0$$

in probability.

Proof. Let $\Psi_n(s,y) = \psi_n(s,y) - \psi(s,y)$. For some arbitrary t, ε, η and $\delta > 0$ we define

$$\tau_\eta = \inf\left\{ t : \int_0^t \int_{\mathbb{R}_0} |\Psi_n(s,y)|^2 \nu(dy)ds > \eta \right\}$$

where,

$$\Psi_n^{(\eta)}(s,y) = \Psi_n(s,y) 1_{[0,\tau_\eta]}(s).$$

Then,

$$\begin{aligned}
P&\left(\sup_{0\le s\le t} \left| \int_0^s \int_{\mathbb{R}_0} \Psi_n(s,y)\tilde{N}(dy,ds) \right| \ge \varepsilon \right) \\
&\le P\left(\tau_\eta < t : \sup_{0\le s\le t} \left| \int_0^s \int_{\mathbb{R}_0} \Psi_n(s,y)\tilde{N}(dy,ds) \right| \ge \varepsilon \right) \\
&+ P\left(\tau_\eta \ge t : \sup_{0\le s\le t} \left| \int_0^s \int_{\mathbb{R}_0} \Psi_n(s,y)\tilde{N}(dy,ds) \right| \ge \varepsilon \right) \\
&\le P(\tau_\eta < t) + P\left(\sup_{0\le s\le t} \left| \int_0^s \int_{\mathbb{R}_0} \Psi_n^{(\eta)}(s,y)\tilde{N}(dy,ds) \right| \ge \varepsilon \right) \\
&\le P\left(\int_0^t \int_{\mathbb{R}_0} |\Psi_n(s,y)|^2 \nu(dy)ds > \eta \right) \\
&+ P\left(\sup_{0\le s\le t} \left| \int_0^s \int_{\mathbb{R}_0} \Psi_n^{(\eta)}(s,y)\tilde{N}(dy,ds) \right| \ge \varepsilon \right)
\end{aligned}$$

Now using Chebychev's inequality and Doob's martingale inequality we can show the second term on the right hand side above is bounded.

$$
\begin{aligned}
P&\left(\sup_{0\le s\le t}\left|\int_0^s\int_{\mathbb{R}_0}\Psi_n^{(\eta)}(s,y)\tilde{N}(dy,ds)\right|\ge\varepsilon\right)\\
&\le\frac{1}{\varepsilon^2}\mathbb{E}\left[\left(\sup_{0\le s\le t}\left|\int_0^t\int_{\mathbb{R}_0}\Psi_n^{(\eta)}(s,y)\tilde{N}(dy,ds)\right|\right)^2\right]\\
&\le\frac{4}{\varepsilon^2}\mathbb{E}\left[\left(\int_0^t\int_{\mathbb{R}_0}\Psi_n^{(\eta)}(s,y)\tilde{N}(dy,ds)\right)^2\right]\\
&=\frac{4}{\varepsilon^2}\mathbb{E}\left[\int_0^t\int_{\mathbb{R}_0}\left(\Psi_n^{(\eta)}(s,y)\right)^2\nu(dy)ds\right]\\
&\le\frac{4\eta}{\varepsilon^2}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
P&\left(\sup_{0\le t\le T}\left|\int_0^t\int_{\mathbb{R}_0}\Psi_n(s,y)\tilde{N}(dy,ds)\right|\ge\varepsilon\right)\\
&\qquad\le P\left(\int_0^t\int_{\mathbb{R}_0}|\Psi_n(s,y)|^2\nu(dy)ds>\eta\right)+\frac{4\eta}{\varepsilon^2}
\end{aligned}
\tag{136}
$$

Now, given some $\delta>0$ we choose $\eta<\delta\varepsilon^2/8$ so that $4\eta/\varepsilon^2<\delta/2$. From the condition proposed at the beginning of the lemma, we can see that there exists $N(\eta)$ such that for $n\ge N(\eta)$ the first term is bounded by $\delta/2$ and so the whole right side is bounded by δ and we are done.

□

The following is very closely based on [5, B.39] but whereas the result therein used only Brownian motion, we have a general Lévy process.

Lemma A.0.6. *Let $X=(X(t),t\ge 0)$ be a Lévy process on a probability space $(\Omega',\mathcal{F}',Q')$ adapted to a given filtration $(\mathcal{F}'_t,t\ge 0)$. Define $S_t=\{(\varepsilon_t^r,t\ge 0)|r\in L^\infty([0,t],\mathbb{R})\}$ where,*

$$
\varepsilon_t^r=\exp\left(i\int_0^t r(s)dX(s)-\int_0^t\eta(r(s))ds\right)
\tag{137}
$$

and $\eta(\cdot)$ is the Lévy symbol corresponding to the process X. Then S_t is a total set in $L^1(\Omega',\mathcal{F}',Q')$, by this we mean if $a\in L^1(\Omega',\mathcal{F}',Q')$ and $\mathbb{E}[a\varepsilon_t]=0$ for all $\varepsilon\in S_t$ then $a=0$ a.s.

Proof. Define a set

$$S'(t) = \left\{ \varepsilon_t = \exp\left(i \int_0^t r(s) dX(s) \right), r \in L^\infty([0,t], R) \right\}.$$

Fix a in $L^1(\Omega', \mathcal{F}', Q')$ such that $\mathbb{E}_Q[a\varepsilon_t] = 0$ for all $\varepsilon_t \in S'(t)$. This can then be seen to be equivalent to the statement that $\mathbb{E}_Q[a\varepsilon_t] = 0$ for all $\varepsilon_t \in S(t)$, and so we will assume this. Take $t_1, t_2, \ldots, t_p \in (0,t)$ with $t_1 < t_2 < \cdots < t_p$, then given $l_1, l_2, \ldots, l_n \in \mathbb{R}$ define

$$\mu_p = l_p, \;\; \mu_{p-1} = l_p + l_{p-1}, \ldots, \;\; \mu_1 = l_p + \cdots + l_1.$$

We set $t_0 = 0$ and define,

$$r_t = \begin{cases} \mu_h & \text{for } t \in (t_{h-1}, t_h), h = 1, \ldots, p; \\ 0 & \text{for } t \in (t_p, T). \end{cases}$$

Then since, $X(t_0) = X(0) = 0$,

$$\sum_{h=1}^p l_h X(t_h) = \sum_{h=1}^p \mu_h (X(t_h) - X(t_{h-1})) = \int_0^t r_s dX(s).$$

Therefore, for $a \in L^1(\Omega', \mathcal{F}', Q')$

$$\mathbb{E}_Q\left[a \exp\left(i \sum_{h=1}^p l_h X(t_h) \right) \right] = \mathbb{E}_Q\left[a \exp\left(i \int_0^t r_s dX(s) \right) \right] = 0,$$

the second equality follows from the assumption $\mathbb{E}[a\varepsilon_t] = 0$ for all $\varepsilon \in S'(t)$. Therefore by linearity,

$$\mathbb{E}_Q\left[a \sum_{k=1}^K c_k \exp\left(i \sum_{h=1}^p l_{h,k} X(t_h) \right) \right] = 0.$$

Let $F(x_1, \ldots, x_p)$ be a continuous bounded complex valued function defined on $\mathbb{R}^p$. By the Weierstrass approximation theorem there exists a uniformily bounded sequence of functions of the form

$$P^{(n)}(x_1, \ldots, x_p) = \sum_{k=1}^{K^n} c_k^{(n)} \exp\left(i \sum_{h=1}^p l_{h,k}^{(n)} x_h \right)$$

such that

$$\lim_{n\to\infty} P^{(n)}(x_1, \ldots, x_p) = F(x_1, \ldots, x_p).$$

So we have $\mathbb{E}_Q\left[aF(X(t_1), \ldots, X(t_p) \right] = 0$ for every continuous function F or by a further approximation a bounded measurable function. Since $t_1, t_2 \ldots t_p$ are arbitrary we obtain $\mathbb{E}_Q[ab] = 0$ for any bounded measurable function b. Also we have that $\mathbb{E}_Q[a^2 \wedge m] = 0$ for arbitrary m, hence $a = 0$ Q-a.s. $\square$

B Code Used in Numerics

B.1 Code Used to Create Graphics

```
library(fBasics)
library(stabledist)
N<-1000
T<-10
x<- 0
y0<-rnorm(1,0,1)
theta1<-c(-1,1)
theta2<-c(1,1)
Dt<- T/N

###SYSTEM###
Y<- numeric(N+1)
Y[1] <-y0
B<- rnorm(N, 0, 1)
for(i in 1:N) {
Y[i+1] <- Y[i] + (theta1[1]*Y[i])*Dt + theta1[2]*sqrt(Dt)*B[i]
}
Y<- ts(Y,start=0, deltat=T/N)

###OBSERVATIONS###
alpha<-1.9
Z<- numeric(N+1)
Z[1] <- k
A<- rstable(N, alpha, 0)
for(i in 1:N) {
Z[i+1] <- Z[i] + (theta2[1]*Y[i])*Dt + theta2[2]*(Dt^(1/alpha))*A[i]
}
Z<- ts(Z,start=0, deltat=T/N)

###MSE###
s<-numeric(N+1)
s[1]<-0.5
for(i in 1:N) {
   s[i] <- 0.5 - 0.5*exp(-2*i*Dt)
}
s<- ts(s,start=0, deltat=T/N)
```

```
###MY FILTER###
k<-Z[1]
Yhat<- numeric(N+1)
Yhat[1] <- y0
for(i in 1:N) {
Yhat[i+1] <- Yhat[i] + theta1[1]*Yhat[i]*Dt}
Yhat<- ts(Yhat,start=0, deltat=T/N)

###MSE M&B FILTER ###
gamma<-numeric(N+1)
gamma[1]=Dt
for(i in 1:N){
gamma[i+1] <-  gamma[i]+ theta1[1]*1.1*gamma[i]*Dt
+(theta1[2]^1.1)*Dt
 - 0.1*(abs((theta2[1]/theta2[2])*gamma[i])^11)*Dt}
gamma<- ts(gamma,start=0, deltat=T/N)

###M&B FILTER###
Yhoot<-numeric(N+1)
Yhoot[1]<-y0
for(i in 1:N) {
Yhoot[i+1] <-   Yhoot[i] +theta1[1]*Yhoot[i]*Dt +sign(theta2[1])
*abs(theta2[1]*gamma[i])^10 *((theta2[1]*Y[i])*Dt
 + theta2[2]*(Dt^(1/alpha))*A[i] - theta2[1]*Yhoot[i]*Dt)
}
Yhoot<- ts(Yhoot,start=0, deltat=T/N)

###AH FILTER###
beta<-exp(-Dt)
p<- exp(Dt)-1
r<- p - p*exp(-Dt)-2*Dt

sigma<-numeric(N+1)
sigma[1]<-0.019801
V<-numeric(N+1)
V[1] <- 0
lambda<-numeric(N+1)
lambda[1]<- -0.4974749
for(i in 1:N) {
sigma[i+1]<- 0.5*(p^2)*(1-exp(-2*Dt))+2*p*(exp(-Dt)-1)
+2*Dt+((p*beta)^2)*V[i]
```

```
lambda[i+1] <- (sigma[i+1]+ r)/(p*sigma[i+1])
V[i+1]<- (p)^(-2)*2*Dt - (p)^(-2)*sigma[i+1]*(1-p*lambda[i])^2
}
sigma<- ts(sigma,start=0, deltat=T/N)
lambda<- ts(lambda,start=0, deltat=T/N)
V<-ts(V,start=0, deltat=T/N)

h<-numeric(N+1)
for(i in 1:N){
h[i]<- (0.01*A[i]+B[i])/sqrt(abs(sigma[i]))}

f<-numeric(N+1)
for(i in 1:N){
f[i+1]<- density(h, from=h[i], to=h[i], n=1)$y
}
g<-numeric(N+1)
for(i in 1:N){
g[i+1]<- (density(h, from=h[i], to=h[i]+0.01, n=1)$y
- density(h, from=h[i], to=h[i], n=1)$y)/0.01
}

zero<-which(f==0)
g[zero]<-0
f[zero]<-10

x<-0
YAH<-numeric(N+1)
YAH[1]<-y0
for(i in 1:N){
YAH[i+1]<- -lambda[i]*(abs(sigma[i+1])^0.5)*(g[i+1]/f[i+1])
+ beta*YAH[i]}
YAH<- ts(YAH,start=0, deltat=T/N)

###ERROR ESTIMATION###

error<-(Yhat-Y)^2
error1<-(Yhoot-Y)^2

error2<-(YAH-Y)^2

print((mean(error)))
```

```
print((mean(error1)))
print((mean(error2)))

###PLOTS###
plot(Z,type="l",ylim=c(-12,12),ylab="")
lines(Y,col="red")
lines(Yhat,col="green")
lines(Yhoot,col="blue")
lines(YAH,col="purple")
```

References

[1] H. Ahn and R.E. Feldman. Optimal filtering of a Gaussian signal in the presence of Lévy noise. *SIAM J. Appl. Math.*, 60(1):359–369, November 1999.

[2] D Applebaum. Martingale-valued measures, Ornstein-Uhlenbeck processes with jumps and operator self-decomposability in Hilbert space. *In Memoriam Paul-André Meyer, Séminaire de Probabilités*, XXXIX Springer LNM 1874:171–196, 2006.

[3] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[4] D. Applebaum. On the infinitesimal generators of Ornstein Uhlenbeck processes with jumps in Hilbert space. *Potential Analysis*, 26:79–100, 2007.

[5] A. Bain and D. Crişan. *Fundamentals of Stochastic Filtering*. Applications of Mathematics. Springer, 2008.

[6] J. Bertoin. *Lévy Processes*. Cambridge Tracts in Mathematics. Cambridge University Press, 1998.

[7] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. J. Wiley & Sons, 1995.

[8] V.S. Borkar. *Probability Theory: an Advanced Course*. Universitext (1979). Springer, 1995.

[9] L. Breiman. *Probability*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1968.

[10] R.S. Bucy and P.D. Joseph. *Filtering for Stochastic Processes with Applications to Guidance*. AMS Chelsea Publishing Series. AMS Chelsea Pub., 1987.

[11] L. Carroll. *Alice's Adventures in Wonderland*. Mac Millan, 1869.

[12] C. Ceci and K. Colaneri. The Zakai equation of nonlinear filtering for jump-diffusion observation: existence and uniqueness. *ArXiv e-prints*, October 2012.

[13] K. Colaneri and C. Ceci. Nonlinear filtering for jump diffusion processes *Adv Appl. Prob.*, 44:678–701, 2012.

[14] D.L. Cohn. *Measure Theory.* Birkhäuser Boston, 1994.

[15] R. Cont and P. Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall/CRC financial mathematics series. Chapman & Hall/CRC, 2004.

[16] D. Crisan and B. Rozovskii. *The Oxford Handbook of Nonlinear Filtering.* Oxford Handbooks. Oxford University Press, 2011.

[17] M.H.A. Davis. *Linear Estimation and Stochastic Control.* Chapman and Hall Mathematics Series. Chapman and Hall, 1977.

[18] L. Debnath and P. Mikusiński. *Introduction to Hilbert Spaces with Applications.* Elsevier Academic Press, 2005.

[19] R.A. Doney and J. Picard. Fluctuation Theory for Lévy Processes in *Ecole d'Eté de Probabilités de Saint-Flour XXXV-2005.* Lecture Notes in Mathematics 1897, Springer-Verlag, 2007.

[20] J.L. Doob. *Classical Potential Theory and Its Probabilistic Counterpart.* Classics in Mathematics. Springer, 2001.

[21] B. Fristedt and L.F. Gray. *A Modern Approach to Probability Theory.* Probability and its Applications. Birkhäuser, 1997.

[22] B.V. Gnedenko and A.N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley Series in Statistics. Addison-Wesley, 1968.

[23] S.M. Iacus. *Simulation and Inference for Stochastic Differential Equations: With R Examples.* Springer Series in Statistics, 1. Springer-Verlag New York, 2008.

[24] K. Itô. On stochastic processes I (infinitely divisible laws of probability). *Jap. J. Math*, 18:261–301, 1942.

[25] J. Jacod and A.N. Shiryaev. *Limit theorems for stochastic processes.* Grundlehren der mathematischen Wissenschaften. Springer-Verlag, 1987.

[26] O. Kallenberg. *Random Measures.* Akademie-Verlag, 1986.

[27] O. Kallenberg. *Foundations of Modern Probability*. Probability and its Applications (New York). Springer-Verlag, second edition, 2002.

[28] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME Journal of Basic Engineering*, (82 (Series D)):35–45, 1960.

[29] R. E. Kalman and R. S. Bucy. New Results in Linear Filtering and Prediction Theory. *Transactions of the ASME. Series D, Journal of Basic Engineering*, 83:95–107, 1961.

[30] J. F. C. Kingman. Completely random measures. *Pacific J. Math*, 21(1):59–78, 1967.

[31] J.F.C. Kingman and S.J. Taylor. *Introduction to Measure and Probability*. Cambridge University Press, 1966.

[32] F.C. Klebaner. *Introduction to Stochastic Calculus with Applications*. Imperial College Press, 2005.

[33] A. N. Kolmogorov. Sur l'interpolation et extrapolation des suites stationnaires. *C. R. Acad. Sci.*, 208:2043, 1939.

[34] A. N. Kolmogorov. Interpolation and Extrapolation. *Bulletin de l-académie des sciences de U.S.S.R., Ser. Math.*, 5:3–14, 1941.

[35] M. Krein. On a generalization of some investigations of G. Szegö, V. Smirnoff and A. Kolmogoroff. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 46:91–94, 1945.

[36] M. Krein. On a problem of extrapolation of A. N. Kolmogoroff. *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, 46:306–309, 1945.

[37] T. G. Kurtz and Jie Xiong. Particle representations for a class of nonlinear spdes. *Stochastic Processes and their Applications*, 83(1):103 – 126, 1999.

[38] A. E. Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications (Universitext)*. Springer, 2006.

[39] A. Le Breton and Marek Musiela. A generalization of the Kalman filter to models with infinite variance. *Stochastic Process. Appl.*, 47(1):75–94, 1993.

[40] P. Lévy. Sur les intégrales dont les éléments sont des variables aléatoires indépendentes. *Ann. R. Scuola Norm. Super. Pisa*, 3:337–366, 1934.

[41] P. Lévy. *Théorie de L'Addition des Variables Aléatoires.* Gauthier-Vilars, Paris, 1937.

[42] R.S. Liptser and A.N. Shiryaev. *Statistics of Random Processes: I. General Theory.* Applications of Mathematics. Springer, 2001.

[43] E. Lukacs. *Characteristic Functions.* Griffin Books of Cognate Interest. Hafner Pub. Co., 1970.

[44] V. Mandrekar and T. Meyer-Brandis. A Bayes formula for non-linear filtering with Gaussian and Cox noise. *Probability and Statistics*, 2011.

[45] T. Meyer-Brandis and F. Proske. Explicit solution of a non-linear filtering problem for Lévy processes with application to finance. *Applied Mathematics & Optimization*, 50:119–134, 2004. 10.1007/s00245-004-0798-6.

[46] V Moretti. *Spectral Theory and Quantum Mechanics: With an Introduction to the Algebraic Formulation.* 2013.

[47] B.Øksendal. *Stochastic Differential Equations: An Introduction with Applications (Universitext).* Springer, 6th edition, 2003.

[48] D. R. Poklukar. Nonlinear filtering for jump-diffusions. *Journal of Computational and Applied Mathematics*, 197(2):558 – 567, 2006.

[49] D. Pollard. *A User's Guide to Measure Theoretic Probability.* Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge University Press, 2002.

[50] P.E. Protter. *Stochastic Integration and Differential Equations.* Applications of Mathematics. Springer, 2004.

[51] M.M. Rao. *Real and Stochastic Analysis: New Perspectives.* Trends in Mathematics. Birkhäuser, 2004.

[52] J.S. Rosenthal. *A First Look at Rigorous Probability Theory.* World Scientific, 2006.

[53] G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Stochastic modeling. Chapman & Hall, 1994.

[54] K. I. Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

[55] S.S.Sritharan S.Popa. Nonlinear filtering of Itô-Lévy stochastic differential equations with continuous obervations. *Commun. Stoch. Anal.*, 3(313):30, 2009.

[56] N. Wiener. Differential space. *Journal of Mathematical Physics*, 2:131–174, 1923.

[57] N.Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series. With Engineering Applications.* The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass, 1949.

[58] Sheng-De Wang, Te-Son Kuo and Chen-Fa Hsu *Trace bounds on the solution of the algebraic matrix Riccati and Lyapunov equation.* IEEE Transactions on Automatic Control, 31(7):654-656, 1986.

[59] J. Xiong. *An Introduction to Stochastic Filtering Theory.* Oxford Graduate Texts in Mathematics. Oxford University Press, 2008.