On Galerkin Approximations for the Zakai Equation with Diffusive and Point Process Observations

An der Fakultät für Mathematik und Informatik
der Universität Leipzig
angenommene

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM
(Dr.rer.nat.)

im Fachgebiet

Mathematik.

vorgelegt

von Master of Science Ling Xu
geboren am 22.12.1977 in Hebei

Die Annahme der Dissertation haben empfohlen:

1. Professor Dr. Rüdiger Frey (Universität Leipzig)
2. Professor Wolfgang Runggaldier(University of Padova)

Die Verleihung des akademischen Grades erfolgte mit Bestehen der Verteidigung am 09.02.2011 mit dem Gesamtprädikat magna cum laude.

Abstract

We are interested in a nonlinear filtering problem motivated by an information-based approach for modelling the dynamic evolution of a portfolio of credit risky securities. We solve this problem by 'change of measure method' and show the existence of the density of the unnormalized conditional distribution which is a solution to the Zakai equation. Zakai equation is a linear SPDE which, in general, cannot be solved analytically. We apply Galerkin method to solve it numerically and show the convergence of Galerkin approximation in mean square. Lastly, we design an adaptive Galerkin filter with a basis of Hermite polynomials and we present numerical examples to illustrate the effectiveness of the proposed method. The work is closely related to the paper Frey and Schmidt (2010).

Acknowledgements

I am very grateful to Prof. Frey and Prof. Schmidt. They gave me the suggestions about my research and discussed with me the problems that happened during the accomplishment of this thesis. Without their steady support, competent help and criticism, it would have been impossible to finish this dissertation.
My thanks are also going to Frau Götz and Frau Fricke as well as other members of the group, for their enthusiastic help.
Furthermore, I thank the Max Planck Institute for Mathematics in the Sciences, and DFG (Deutschen Forschungsgemeinschaft) for their financial support to my stay in University Leipzig.
I want to say thanks to my husband, Chuangye Wang for his never-ending support.

Contents

Chapter 1

Introduction

1.1 Preliminary work

Financial motivation: Credit risk model under incomplete information

The credit markets have developed at a tremendous speed in recent years and the demand for credit derivatives, such as credit default swaps (CDSs) and collateralized debt obligations (CDOs), are growing rapidly. Consequently, recent research has highlighted the study of the credit risk model. A good model is one that captures the dynamic evolution of credit spreads and the dependent structure of default in a realistic way. Also, from a computation point of view, it should be tractable and parsimonious.

Existing credit risk models can be divided into two classes: structural models and reduced-form models. In structural models, default occurs when an assets value falls below a threshold, generally representing liabilities. Structural credit risk models are discussed in, for instance, Black and Scholes (1973), Merton (1974), Black and Cox (1976). In reduced-form models, one models directly the law of the default time where the precise mechanism leading to default is not specified. In practice, reduced-form models are usually preferred for tractability reasons. Reduced-form credit risk models are discussed in, for instance, Jarrow and Turnbull (1995), Lando (1998), Duffie and Singleton (1999), Blanchet-Scalliet and Jeanblanc (2004).

In most of the credit risk models, the distribution of default times depends on a state variable process X and in practice, X can not be fully observed by investors. This, in turn, leads to a nonlinear filtering problem in a natural way. Structural credit risk model under incomplete information are discussed in, for instance, Kusuoka (1999), Duffie and Lando (2001), Jarrow and Protter (2004), Coculescu, Geman, and Jeanblanc (2008), and Frey and Schmidt (2009). Reduced-form credit risk models under incomplete information are discussed in, for instance, Frey and Schmidt (2010), Frey and Runggaldier (2010), Schönbucher (2004), Collin-dufresne, Goldstein, and Helwege (2003), Giesecke (2004) and McNeil, Frey, and Embrechts (2005), page 448-462.

The general theories of nonlinear filtering are well-developed, see for instance Bain and Crisan (2009), and are increasingly being used in financial mathematics, see for instance Frey (2000), Frey and Runggaldier (2010), Frey and Schmidt (2009). Filtering techniques come into play when the factors cannot be observed directly. A short introduction for nonlinear filtering is presented in Section 2.1.

Model and notation

Frey and Schmidt (2010) study reduce-form portfolio credit risk models under incomplete information. They consider models where the default intensities of the firms under consideration depend on an unobservable stochastic processes X. The information for investors, the so-called market information, contains only the default history of firms and noisy price observations of traded credit derivatives.

Specifically, they work on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and on a finite time interval $[0, T]$. Here $\mathbb{P}$ is a risk-neutral measure. $\mathbb{F} = \{\mathcal{F}_t\}_{0\leq t\leq T}$ is the full-information filtration and all processes will be $\mathbb{F}$-adapted. In association with a generic process ξ, define for each $t \geq 0$ a sub-σ-field of $\mathcal{F}_t$, denoted $\mathcal{F}_t^\xi$, by

$$\mathcal{F}_t^\xi = \sigma\{\xi_s, s \in [0, t]\}. \tag{1.1.1}$$

Consider defaultable securities issued by a firm where the random time τ denotes the default time of the firm. $Y_t = \mathbf{1}_{\{\tau\leq t\}}$ is the corresponding default indicator. The default intensity is assumed to depend on some state process X, which is modelled as a d-dimensional finite state Markov chain.

Assume τ is a doubly stochastic random time with $\big(\mathbb{P}, \mathbb{F}\big)$-default intensity $\lambda_t = \lambda(X_t)$, i.e., there is a function $\lambda : \mathbb{R} \to (0, \infty)$, such that $N_t^* := Y_t - \int_0^{t\wedge\tau} \lambda(X_s)ds$ is an $\mathbb{F}$-martingale.

The informational advantage of informed market participants is modelled via observations of a process Z. Formally, the market filtration is given by $\mathbb{F}^{Z,Y} := \{\mathcal{F}_t^{Z,Y}\}_{0\leq t\leq T}$, where process Z, which is l-dimensional, solves the SDE

$$dZ_t = h(X_t)dt + dB_t, \quad t \in [0, T].$$

Here B is a l-dimensional standard Brownian motion independent of X and Y. And $h(\cdot)$ is a function from $\mathbb{R}^d$ to $\mathbb{R}^l$.

Consider a liquidly traded credit derivatives with maturity T and $\mathcal{F}_T^Y$-measurable payoff P. In order to simplify the computation, we assume the full-information value of the securities given by $\mathbb{E}(P|\mathcal{F}_t) =: \tilde{P}_t(X_t, Y_t)$. The market price of the security, which is determined by informed market-participants with market information, is defined as, by iterated conditional expectations,

$$P_t := \mathbb{E}(P|\mathcal{F}_t^{Z,Y}) = \mathbb{E}\Big[\mathbb{E}(P|\mathcal{F}_t)\Big|\mathcal{F}_t^{Z,Y}\Big] = \mathbb{E}\Big[\tilde{P}_t(X_t, Y_t)\Big|\mathcal{F}_t^{Z,Y}\Big].$$

The objective of financial mathematics is to derive the dynamics of the market price. In order to computer the price P_t, one need to obtain the conditional measure of X_t given $\mathcal{F}_t^{Z,Y}$, given by π_t. This leads to a nonlinear filtering problem in a natural way.

1.2 Overview of the thesis

Motivated by the credit risk model studied by Frey and Schmidt (2010), we extend the model with two viewpoints, realistic and mathematical. Realistically, the state process X is modelled as a diffusion process with generator $\mathcal{L}$, which is a second order differential operator with domain $D(\mathcal{L})$. Mathematically, the jump observation Y is modelled as a doubly stochastic Poisson process with stochastic intensity $\lambda(X_t)$, then τ can be viewed as the first jumping time of Y.

Given the past observation of Z and Y, the objective of this thesis is to determine the conditional measure of X_t. We solve this problem by 'change of measure method' and we show the existence of the density of the unnormalized conditional distribution which is a solution of the so-called Zakai equation. The Zakai equation is a linear stochastic partial differential equation which cannot typically be solved analytically. We will apply Galerkin method to solve it numerically, and will show the convergence of Galerkin approximation in mean square. To conclude, we design an adaptive Galerkin filter with a basis of Hermite polynomials and present numerical examples to illustrate the effectiveness of the proposed method. Following are specifics about the topics of this thesis.

The unnormalized filtering equations

We deduce the evolution equation for π using the change of measure method: A new measure $\mathbb{P}^0$ is constructed under which Z becomes a Brownian motion, Y becomes a Poisson process with intensity 1, and the process X, Z, Y are independent. Then, π has a representation in terms of an associated unnormalized version ρ. This ρ is then shown to satisfy the so-called Zakai equation, for any $f \in D(\mathcal{L})$,

$$\begin{aligned}\rho_t(f) =& \mathbb{E}[f(X_0)|\mathcal{F}_0^{Z,Y}] + \int_0^t \rho_s(\mathcal{L}f)ds + \int_0^t \rho_s(fh^\top)dZ_s \\ &+ \int_0^t \rho_{s-}\Big(f(\lambda-1)\Big)d(Y_s - s), \ \ \mathbb{P}^0 - a.s., \ \ \forall t \in [0,T].\end{aligned}$$

For detail, see Theorem 2.9. This leads to the evolution equation for π by an application of Itô's formula.

The unnormalized conditional density

The following question is interesting to answer: Does there exist a density of the conditional distribution of X_t given accumulated observation? In Chapter 5, we prove, under fairly mild conditions, the unnormalized conditional distribution ρ_t has a square integrable density q_t with respect to Lebesgue measure, which is a weak solution of a linear stochastic partial differential equation, known as the Zakai Equation. For the detail, see Theorem 5.3.

Numerical approach

Our objective is to seek a numerical method that is implementable and provides accurate solutions to the Zakai Equation.

One approach is to approximate the diffusion model of X by a finite-state Markov chain and to write down the Zakai equation for it, which yields a stochastic ODE, see Frey and Runggaldier (2010), Frey and Schmidt (2010). In this work, a different approach is proposed, namely the Zakai equation is directly approximated by means of the classical Galerkin method for solving deterministic PDEs, see Ahmed and Radaideh (1997). The solution of the Zakai equation is first approximated by a finite combination of orthogonal series. Then, it is approximated by the solution of a family of finite dimensional stochastic ordinary differential equations, which can be solved numerically or analytically. This work consists of two main parts, theoretical and numerical.

Theoretical part In Chapter 6, we prove the convergence of the Galerkin approximation of the Zakai equation in mean square sense, with usual assumptions. This is done by using a general continuity result for the solution of a mild stochastic linear differential equation on a Hilbert space with respect to the semigroup. With bounded and square integrable function φ, we have

$$\sup_{t\in[0,T]} \mathbb{E}^0 \Big| \int_{\mathbb{R}^d} \varphi(x) q_t^{(n)}(x)dx - \int_{\mathbb{R}} \varphi(x) q_t(x)dx \Big|^2 \to 0, \text{ as } n \to \infty,$$

where $\mathbb{E}^0$ is the expectation w.r.t. $\mathbb{P}^0$, n is the number of basis functions used in the Galerkin filter.

Numerical part Concerned with the Galerkin approximation convergence rate, in Chapter 7, we design an adaptive Galerkin filter with a basis of Hermite polynomials and present numerical examples to illustrate the effectiveness of the proposed method. In simulation study, we compare the proposed method with particle methods and show that the Galerkin approximation converges well. In Chapter 7, we present the Galerkin approximation strategy for solutions of the Zakai equation (5.2.1) by solving a sequence of finite dimensional stochastic differential equations. The solution of the Zakai equation can be constructed by the Galerkin method using any suitable set of basis functions from Hilbert space. It is possible to choose a complete set of basis functions, like Gaussian series and Hermite functions. However, in most nonlinear filtering problems, particularly when the observation noise is small, the conditional density is well-localized in a small region of the state space and generally cannot be predicted in advance. To overcome this difficulty, we design an adaptive Galerkin filter with Hermite polynomials. Finally, we present examples and the corresponding simulation results.

Structure of the thesis Listed below is a brief summary of the remaining chapters.

In Chapter 2: We introduce the nonlinear filtering model which we study throughout this thesis and deduce the corresponding filtering equation by ‘change of measure method’.

In Chapter 3: We present an overview of the main computational methods currently available for solving the filtering problem. Three classes of numerical method are presented: the finite dimensional filter based on approximating the conditional density by a linear combination of Gaussian functions, finite state Markov chain, and the particle filter.

In Chapter 4: Since the Zakai equation is a linear SPDE, we survey some existing results on linear SPDEs. Two approaches presented are the semigroup and the variational.

In Chapter 5: We show the existence of the unnormalized conditional density which is the solution of a Zakai equation.

In Chapter 6: We solve the Zakai equation numerically by Galerkin approximation, and we show the convergence of Galerkin approximation in mean square.

In Chapter 7: We design an adaptive Galerkin filter with Hermite polynomials and present examples of the corresponding simulation results to illustrate the effectiveness of the proposed method.

Chapter 2

Filtering model and the Zakai equation

In this chapter, we mainly introduce the filtering model which will be studied throughout the thesis. Furthermore we deduce the corresponding filtering equation by a 'change of measure method'.

Presented in Section 2.1 is a short review of a nonlinear filtering problem. In Section 2.2, we introduce the nonlinear filtering model studied in this thesis and include the state process which we are interested but, unable to observe directly. Also included are the observation processes which are the partial observations of the state, and the objective of the nonlinear filtering problem. The objective is to obtain the conditional distribution of the partially observed processes recursively. In Section 2.3, we introduce the innovation process and discuss their martingale processes, which are useful for the numerical study introduced in Chapter 7. In Section 2.4, by 'change of measure method', we derive the Zakai equation, which describes the evolution of an unnormalized version of the conditional distribution.

2.1 Stochastic filtering

This section is devoted to a short introduction on nonlinear filtering. For details, we refer to the book Bain and Crisan (2009).

We begin Section 2.1.1 with a general nonlinear filtering problem. In Section 2.1.2, we pursue the Markov case and present the filtering equation obtained by two approaches. Section 2.1.3 is devoted to finite dimensional filters. Finally, in Section 2.1.4, we introduce the nonlinear filtering problem for jump-diffusion case.

2.1.1 A general introduction

The objective, in this section, is to present a short introduction to a general nonlinear filtering problem. It concerns the following: Denote by $\mathcal{T} \subseteq \mathbb{R}^+$ a set of time points (usually $\mathcal{T} = \mathbb{R}^+$). We are interested in a signal or state process $X = \{X_t\}_{t\in\mathcal{T}}$ which can not be observed directly. Instead, the so called observation process $Z = \{Z_t\}_{t\in\mathcal{T}}$, a noisy nonlinear observation of X is obtained.

The objective of a filtering problem is to determine the conditional distribution of the state X_t given $\mathcal{F}_t^Z = \sigma(Z_s, 0 \leq s \leq t)$, which we denote by π_t. Recall, by Equation (1.1.1), $\mathcal{F}_t^Z$ denotes the past observation of Z until time t. Bain and Crisan (2009), Corollary 2.26, page 31, show that under some assumptions, the conditional distribution of the signal can be viewed as a stochastic process with values in the space of probability measures.

Additionally, from a computational point of view, it is desired that π be obtained recursively. This means that π_{t+s} can be built up from π_t and the new observation rather than the whole observation history. This allows for quick updating of the filter and avoids serious data storage issues.

Typically, π_t is infinite dimensional. However, sometimes π_t can be taken as a finite-dimensional object. This will be introduced in Section 2.1.3.

2.1.2 The filtering equations

To derive detailed results we introduce some notations as follows. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space together with a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ which satisfies the usual conditions.

Assume that X is a $\mathbb{R}^d$-valued Markov process with state space $S \subset \mathbb{R}^d$. For example, X can be a diffusion process or a finite state Markov chain, for details, see Bain and Crisan (2009), page 49. Denote by $\mathcal{L}$ be the generator of the Markov process: for $x \in S$,

$$\mathcal{L}f(x) = \lim_{t\downarrow 0} \frac{\mathbb{E}[f(X_t)|X_0 = x] - f(x)}{t},$$

the limit being uniform in $x \in S$ and $D(\mathcal{L})$ denotes the set of all bounded, measurable, real-valued functions $f : S \to \mathbb{R}$ such that this limit exists. The generator gives the expected rate of change of the process $\{f(X_t)\}_{t\in[0,T]}$.

Z, is the noisy nonlinear observation of X,

$$Z_t = \int_0^t h(X_s)ds + B_t, \quad 0 \leq t \leq T, \tag{2.1.1}$$

where B is a standard l-dimensional Brownian motion which is independent of X. $h : \mathbb{R}^d \to \mathbb{R}^l$ is a measurable function such that,

$$\mathbb{P}(\int_0^T \|h(X_s)\|ds < \infty) = 1, \tag{2.1.2}$$

where the Euclidean norm $\|\cdot\|$ is defined in the usual fashion for vectors. This ensures the Riemann integral in Equation (2.1.1) is well defined.

As we introduced before, the objective is to recursively derive $\pi_t(f) = \mathbb{E}[f(X_t)|\mathcal{F}_t^Z]$, which denotes the conditional of X_t given the past observation. The problem can be solved by two approaches, one is the innovation approach, the other is the 'change of measure method'.

In the innovation approach, if Equation (2.1.2) is satisfied, Bain and Crisan (2009), Theorem 3.30, page 68, show that, with additional assumption for h such that the stochastic integral in the following equation is well defined, $\forall f \in D(\mathcal{L})$, $t \in [0, T]$,

$$\pi_t(f) = \pi_0(f) + \int_0^t \pi_s(\mathcal{L}f)ds + \int_0^t \Big[\pi_s(hf) - \pi_s(h)\pi_s(f)\Big]\Big(dZ_s - \pi_s(h)ds\Big). \tag{2.1.3}$$

This is called the Kushner-Stratonovich equation. The second term describes the evolution of the distribution of X_t. The third describes the evolution of the conditional distribution of X_t given accumulation of observations. Equation (2.1.3) is a nonlinear stochastic equation. It is not only an infinite dimensional, but it has a complicated structure due to the presence of the term $\pi_s(h)$. In general, it is not useful for computation.

By the change of measure approach and with the same assumptions as Equation (2.1.3), one obtains an unnormalized version of π, denoted by ρ which satisfies the following linear equation. See Theorem 3.24, Bain and Crisan (2009), page 62,

$$\rho_t(f) = \rho_0(f) + \int_0^t \rho_s(\mathcal{L}f)ds + \int_0^t \rho_s(hf)dZ_s, \quad t \in [0,T],$$

which has a much simpler structure than Equation (2.1.3). This equation is called Zakai equation. And π can be obtained from ρ after normalizing

$$\pi_t(f) = \frac{\rho_t(f)}{\rho_t(1)}, \quad t \in [0,T].$$

2.1.3 Finite-dimensional filters

Recall that π_t denotes the conditional distribution of X_t given the past observation. In general, π_t can not be determined by finite number of parameters. But, in some special cases, π will be determined by a finite dimensional stochastic differential equations driven by observations. The aim of this section is to introduce some special filters for which the corresponding π is finite-dimensional.

The Kalman-Bucy filter

In this section, we introduce the very special filtering problem where the signal is Gaussian and the observation function is linear. The corresponding theory is called Kalman-Bucy filter. For this case, the conditional distribution of X_t given $\mathcal{F}_t^Z$ is a normall distribution. Hence it is determined by the conditional mean and the conditional variance. Therefore, in this case, the filter is 2-dimensional. Finally, we give the evolution equations of the two parameters.

Here, to simplify, we assume that the coefficients are 1-dimensional. We consider the following linear model, for $t \in [0,T]$,

$$\begin{cases} dX_t = (\tilde{b}_t X_t + \tilde{b}_t^0)dt + \tilde{\sigma}_t dV_t, \\ dZ_t = (\tilde{h}_t X_t + \tilde{h}_t^0)dt + dB_t, \end{cases} \tag{2.1.4}$$

where X_0 is normal distributed with mean $\mu_0 \in \mathbb{R}$ and variance $r_0^2 \in \mathbb{R}^+$, V and B are independent 1-dimensional Brownian motions, $\{\tilde{b}_t\}$, $\{\tilde{b}_t^0\}$, $\{\tilde{h}_t\}$ and $\{\tilde{h}_t^0\}$ are deterministic $\mathbb{R}$-valued processes, $\{\tilde{\sigma}_t\}$ is a deterministic $\mathbb{R}^+$-valued process, and $Z_0 = 0$. For this model, the conditional distribution of the state X_t given the past observation of Z is Gaussian, applying Lemma 6.12, Bain and Crisan (2009), page 149. A normal distribution is determined by its mean and variance. Therefore, the conditional distribution is uniquely determined by its mean, defined by $\hat{X}_t := \mathbb{E}[X_t|\mathcal{F}_t^Z]$, and variance, defined by $P_t := \mathbb{E}\Big[(X_t - \hat{X}_t)^2\Big|\mathcal{F}_t^Z\Big]$. By Proposition 6.14, Bain and Crisan (2009), page 152, the process $\big\{(\hat{X}_t, P_t)\big\}_{0\le t\le T}$ is the unique solution to the following equations, for $t \in [0,T]$,

$$\begin{cases} d\hat{X}_t = (\tilde{b}_t\hat{X}_t + \tilde{b}^0_t)dt + \tilde{h}_t P_t\Big[dZ_t - (\tilde{h}_t\hat{X}_t + \tilde{h}^0_t)dt\Big], \\ \frac{d}{dt}P_t = \tilde{\sigma}^2_t + 2\tilde{b}_t P_t - \tilde{h}^2_t P^2_t, \end{cases} \tag{2.1.5}$$

with $\hat{X}_0 = \mathbb{E}(X_0) = \mu_0$ and $P_0 = \mathbb{E}[(X_0 - \hat{X}_0)^2] = r_0^2$. The details can be found in Bain and Crisan (2009), page 148-154.

Finite state Markov chain

Next we introduce a filter, for which the corresponding conditional distribution is a finite-dimensional stochastic process, since the state space is finite dimensional. We will show that, in this case, (2.1.3) gives rise to a finite-dimensional filter.

We consider Model (2.1.1) and, for simplicity, we assume that the coefficients are 1-dimensional. Moreover, we specify to the case that X be a finite-state Markov-chain with state space $S = \{1,\dots,n\}$, where $n \in \mathbb{N}$. The generator associated to process X is a Matrix $Q = (q_{ij})_{1\le i,j\le n}$, which denotes the transition intensities of X. Let

$$p_{t,i} := \mathbb{P}(X_t = i|\mathcal{F}^Z_t) \quad \text{and} \quad \mathbf{p}_t := (p_{t,1},\dots,p_{t,n})^\top \tag{2.1.6}$$

be the vector process representing the vector of conditional probabilities. Let vector $\mathbf{h} := \Big(h(1), h(2),\dots,h(n)\Big)^\top$, and I_n be the identity matrix of size n. Davis and Marcus (1981), Example 1, page 66, and Bain and Crisan (2009), Remark 3.26, page 65, show that the conditional probabilities solve the following n-dimensional stochastic differential equation,

$$\mathbf{p}_t = \mathbf{p}_0 + \int_0^t Q^\top \mathbf{p}_s ds + \int_0^t (B - \mathbf{h}^\top \mathbf{p}_s I_n)\mathbf{p}_s\Big(dZ_s - \mathbf{h}^\top \mathbf{p}_s ds\Big), \quad t \in [0,T], \tag{2.1.7}$$

where $B := \operatorname{diag}\Big(h(1),\dots,h(n)\Big)$ is a diagonal matrix.

Equation (2.1.7) is a recursive equation for the computation of $\mathbf{p}_t$. Moreover Equation (2.1.7) is an n-dimensional SDE system for the vector $\mathbf{p}_t$ of conditional probabilities, and hence a finite-dimensional filter.

2.1.4 Filtering from point processes observations

The nonlinear filtering problem for jump-diffusion is of great interest. The fundamental to point process filtering was presented in Brémaud (1972) and Brémaud (1981). Frey and Runggaldier (2010) deal with a general case and provide recursive updating rules for the filter. Shown below is a simple example with 1-dimensional coefficients:

Let $n \in \mathbb{N}$, and we assume that X can be a finite-state Markov-chain with state space $S = \{1,\dots,n\}$. The generator associated to the process X is a Matrix $Q = (q_{ij})_{1\le i,j\le n}$. Here, observation Z is a doubly stochastic Poisson process with intensity $\lambda(X_t)$, where $\lambda : \mathbb{R} \to \mathbb{R}^+$. Let $\mathbf{p}$, defined by Equation (2.1.6), be the vector process representing the vector of conditional probabilities. Let vector $\bar{\lambda} := \Big(\lambda(1), \lambda(2),\dots,\lambda(n)\Big)^\top$. Frey and Runggaldier (2010) show that these conditional probabilities solve the following n-dimensional stochastic differential equation, for $t \in [0, T]$,

$$\mathbf{p}_t = \mathbf{p}_0 + \int_0^t Q^\top \mathbf{p}_s ds + \int_0^t \frac{1}{\lambda^\top \mathbf{p}_{s-}} (B^\lambda - \bar{\lambda}^\top \mathbf{p}_{s-} I_n) \mathbf{p}_{s-} \Big(dZ_s - \lambda^\top \mathbf{p}_s ds \Big), \tag{2.1.8}$$

where $B^\lambda := \operatorname{diag}\Big(\lambda(1), \dots, \lambda(n)\Big)$ is a diagonal matrix.

2.2 The filtering model

The objective of this section is to introduce the nonlinear filtering problem in which we are interested. We have a state process X which can not be observed. The information of X is obtained by the observation processes Z and Y, where Z is a nonlinear continuous observation, with some noise, and Y is a doubly stochastic Poisson process with the stochastic intensity which is a nonlinear function of X. We are interested in the conditional expectation of a function of X_t given the past observation.

All stochastic processes will be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and on a finite time interval $[0, T]$. $\mathbb{F} = \{\mathcal{F}_t, t \in [0, T]\}$, which satisfies the usual conditions, is the full-information filtration. All processes considered are by assumption $\mathbb{F}$-adapted. We consider the following filtering problem throughout this thesis.

2.2.1 Unobserved state process

This section is devoted to an introduction of the state process which is a diffusion process in our case. To be precise, let $X = \{X_t, 0 \le t \le T\}$ be the unobserved d-dimensional state. The state process is a stochastic process which can not be observed directly. We assume X is the solution of a d-dimensional stochastic differential equation driven by a standard m-dimensional Brownian motion $V = \{V_t, t \in [0, T]\}$,

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dV_s, \quad 0 \le t \le T. \tag{2.2.1}$$

Here, X_0 has finite second moment. Let $p_0 \in L^2(\mathbb{R}^d)$ be the density of the law of X_0, then $p_0(x) \ge 0$ a.e. and $\int_{\mathbb{R}^d} p_0(x) dx = 1$. We assume that $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times m}$ satisfy the following conditions: there exist a positive constant K, such that for all $x, y \in \mathbb{R}^d$, we have

$$\|b(x) - b(y)\| \le K\|x - y\|, \qquad \|\sigma(x) - \sigma(y)\| \le K\|x - y\|. \tag{2.2.2}$$

$$\|b(x)\| \le K(1 + \|x\|), \qquad \|\sigma(x)\| \le K(1 + \|\sigma(x)\|). \tag{2.2.3}$$

Here the Euclidean norm $\|\cdot\|$ is defined in the usual fashion for vectors, and extended to $d \times m$ matrices by considering then as $d \times m$-dimensional vectors : $\|\sigma\| = \sqrt{\sum_{i=1}^d \sum_{j=1}^m (\sigma_{ij})^2}$. Under the globally Lipschitz condition, the SDE has a unique solution. For the existence and uniqueness of the solution for Equation (2.2.1), see Øksendal (1980), Theorem 5.5, page 48, or Bain and Crisan (2009), page 7.

Let $L^\infty(\mathbb{R}^d)$ be the space of bounded measurable real-valued functions on $\mathbb{R}^d$. The generator $\mathcal{L} : D(\mathcal{L}) \to L^\infty(\mathbb{R}^d)$ associated to the process X is the second order differential operator

$$\mathcal{L} = \sum_{i=1}^d b_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}, \tag{2.2.4}$$

where $a(x) = [a_{ij}(x)] := \sigma(x)\sigma^\top(x)$, and b_i and x_i denote the ith component of b and x, respectively. $D(\mathcal{L})$ consists of all function $f \in L^\infty(\mathbb{R}^d)$, such that $\mathcal{L}f \in L^\infty(\mathbb{R}^d)$ and

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)ds, \quad t \in [0,T], \tag{2.2.5}$$

is an $\{\mathcal{F}_t\}$-adapted martingale.

2.2.2 Observations

X is partially observed, that is, information concerning X is obtained from the observation processes Z, which is continuous, and Y, which is a pure jump process. Z is a l-dimensional noisy nonlinear observation of the state X_t,

$$Z_t = \int_0^t h(X_s)ds + B_t, \quad 0 \le t \le T. \tag{2.2.6}$$

Here $B = \{B_t, t \in [0,T]\}$ is a l-dimensional standard Brownian motion independent of X and Y. We have the following:

Assumption 2.1. *We assume that $h : \mathbb{R}^d \to \mathbb{R}^l$ is a measurable function such that*

$$\mathbb{P}\Big(\int_0^T \|h(X_s)\|ds < \infty\Big) = 1. \tag{2.2.7}$$

The Equation (2.2.7) ensures that the Riemann integral in Equation (2.2.6) exists $\mathbb{P} - a.s.$

Assumption 2.2. *We assume that $\lambda : \mathbb{R}^d \to [\varpi_1, \varpi_2]$ is a continuous function, where $0 < \varpi_1 < \varpi_2$ are constants.*

We further assume that Y is a doubly stochastic Poisson process with the stochastic intensity $\{\lambda(X_t)\}_{t\in[0,T]}$ and $Y_0 = 0$. Then the process

$$Y_t - \int_0^t \lambda(X_s)ds, \quad t \in [0,T],$$

is a $(\mathbb{P}, \mathbb{F})$-martingale. Denote the jumping times of Y by the sequence $\{\tau_n\}_{n\ge1}$, then $\tau_n = \inf\{t \ge 0 | Y_t \ge n\}$.

2.2.3 The objective

Recall that $\mathcal{F}_t^{Z,Y}$ is the σ-algebra generated by $\{Z_u, 0 \le u \le t\} \bigcup \{Y_u, 0 \le u \le t\}$. Then the observation filtration is given by $\mathbb{F}^{Z,Y} := \{\mathcal{F}_t^{Z,Y}\}_{0\le t\le T}$. We are interested in, for all $f \in L^\infty(\mathbb{R}^d)$,

$$\mathbb{E}\Big[f(X_t)\Big|\mathcal{F}_t^{Z,Y}\Big], \quad t \in [0,T]. \tag{2.2.8}$$

The subject of the mathematical theory of filtering is finding suitable ways of computing this conditional expectation recursively, either exactly or approximately. Equivalently, the principal aim of solving a filtering problem is to determine the conditional distribution of the state X_t given the observation history.

To conclude, we study the following nonlinear model throughout the thesis, for $t \in [0,T]$,

$$\begin{cases} X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dV_s, \\ Z_t = \int_0^t h(X_s)ds + B_t, \\ Y \text{ is a doubly stochastic Poisson process with intensity } \lambda(X_t), \end{cases} \tag{2.2.9}$$

where X is the state process, Z and Y are observations. The objective is to determine the conditional distribution of the state X_t given the observation history.

2.3 The innovations processes

In this section, we introduce the innovations processes and discuss their martingale properties. These theoretical properties will be used in our computation strategy, see Section 7.1.2. In Chapter 7, we will use the Galerkin method to approximate the unnormalized conditional density. When the numerical result is inconsistent with its theoretical properties, it is necessary to increase the number of basis functions used in the approximation.

For a generic process η, denote $\widehat{\eta}_t := \mathbb{E}[\eta_t|\mathcal{F}_t^{Z,Y}]$. Now we introduce the innovations processes as follows: for $t \in [0,T]$,

$$M_t = Y_t - \int_0^t \widehat{\lambda}_s ds, \tag{2.3.1}$$

$$\mu_t = Z_t - \int_0^t \widehat{h}_s ds. \tag{2.3.2}$$

In what follows, we show M and μ are martingales.

Lemma 2.1. *Under Assumptions 2.2, $\{M_t\}_{t\in[0,T]}$ is a $(\mathbb{P}, \mathbb{F}^{Z,Y})$-martingale.*

Proof. It follows from Equation (2.3.1) that for $t \in [0,T]$,

$$\begin{aligned} \mathbb{E}|M_t| =& \mathbb{E}\Big|Y_t - \int_0^t \lambda(X_s)ds + \int_0^t \lambda(X_s)ds - \int_0^t \widehat{\lambda}_s ds\Big| \\ \leq& \mathbb{E}\Big|Y_t - \int_0^t \lambda(X_s)ds\Big| + 2\|\lambda\|_\infty T < \infty. \end{aligned}$$

The last inequality follows as $Y_t - \int_0^t \lambda(X_s)ds$ is a martingale and λ is bounded by Assumption 2.2. For $0 \leq s < t \leq T$,

$$\begin{aligned} \mathbb{E}[M_t|\mathcal{F}_s^{Z,Y}] =& \mathbb{E}\Big[Y_t - Y_s - \int_s^t \widehat{\lambda}_u du\Big|\mathcal{F}_s^{Z,Y}\Big] + M_s \\ =& \mathbb{E}\Big[Y_t - Y_s - \int_s^t \lambda(X_u)du\Big|\mathcal{F}_s^{Z,Y}\Big] + \mathbb{E}\Big[\int_s^t (\lambda(X_u) - \widehat{\lambda}_u)du\Big|\mathcal{F}_s^{Z,Y}\Big] + M_s. \end{aligned}$$

On the one hand,

$$\mathbb{E}\Big[Y_t - Y_s - \int_s^t \lambda(X_u)du\Big|\mathcal{F}_s^{Z,Y}\Big] = \mathbb{E}\Big[\mathbb{E}\Big(Y_t - Y_s - \int_s^t \lambda(X_u)du\Big|\mathcal{F}_s\Big)\Big|\mathcal{F}_s^{Z,Y}\Big] = 0.$$

On the other hand,

$$\begin{aligned}\mathbb{E}\Big[\int_s^t (\lambda(X_u) - \widehat{\lambda}_u)du\Big|\mathcal{F}_s^{Z,Y}\Big] =& \int_s^t \mathbb{E}\Big(\lambda(X_u) - \widehat{\lambda}_u\Big|\mathcal{F}_s^{Z,Y}\Big)du \\ =& \int_s^t \mathbb{E}\Big(\lambda(X_u) - \mathbb{E}(\lambda(X_u)|\mathcal{F}_u^{Z,Y})\Big|\mathcal{F}_s^{Z,Y}\Big)du \\ =& 0,\end{aligned}$$

where the first equality follows from Fubini's Theorem. So

$$\mathbb{E}[M_t|\mathcal{F}_s^{Z,Y}] = M_s,$$

and M is a $(\mathbb{P}, \mathbb{F}^{Z,Y})$-martingale. ☐

Lemma 2.2. *Under Assumption 2.1, $\{\mu_t\}_{t\in[0,T]}$ is a $(\mathbb{P}, \mathbb{F}^{Z,Y})$-Brownian motion.*

Proof. It follows from Equation (2.3.2) that, for $0 \le s < t \le T$,

$$\begin{aligned}\mathbb{E}[\mu_t|\mathcal{F}_s^{Z,Y}] &= \mathbb{E}\Big[Z_t - Z_s - \int_s^t \widehat{h}_u du\Big|\mathcal{F}_s^{Z,Y}\Big] + \mu_s \\ &= \mathbb{E}\Big[B_t - B_s + \int_s^t \Big(h(X_u) - \widehat{h}_u\Big)du\Big|\mathcal{F}_s^{Z,Y}\Big] + \mu_s \\ &= \int_s^t \mathbb{E}\Big[h(X_u) - \mathbb{E}\Big(h(X_u)|\mathcal{F}_u^{Z,Y}\Big)\Big|\mathcal{F}_s^{Z,Y}\Big]du + \mu_s \\ &= \mu_s,\end{aligned}$$

where the fourth equality follows from the independent increments of the Brownian motion B, Fubini's Theorem and iterated conditional expectation. Therefore μ is a $(\mathbb{P}, \mathbb{F}^{Z,Y})$-martingale.

On the other hand, it follows from Equation (2.3.2) that

$$\langle\mu\rangle_t = \langle Z\rangle_t = t.$$

Since μ is an $\mathbb{F}^{Z,Y}$-martingale starting from zero at time zero with continuous paths and with quadratic variation equal to t at each time t, μ is an $\mathbb{F}^{Z,Y}$-Brownian motion, by Shreve (2004), Theorem 4.6.4, page 168. ☐

2.4 The Zakai equations

In this section, we proceed to estimate the state X based on the available information. Generally, the conditional distribution can be computed by two methods. The first method is to solve Kushner-Stratonovich equation, which is a nonlinear stochastic partial differential equation. The second method is to solve Zakai equation, which is a linear stochastic partial differential equation describing the dynamics of the unnormalized distribution. The objective of this section is to deduce the Zakai equation of the nonlinear filtering problem.

The section is organized as follows. In Section 2.4.1, a new measure is constructed under which Z becomes a Brownian motion, Y becomes a standard Poisson process. In Section 2.4.2, we show that the conditional expectation has a representation in terms of an associated unnormalized version ρ and ρ satisfies a linear evolution equation, the so-called Zakai equation.

2.4.1 A new measure

Before deriving the Zakai Equation, we introduce a new measure where Z becomes a standard Brownian motion and Y becomes a Poisson process with intensity 1. Furthermore, under the new measure, they are independent. Prior to that, we define stochastic process $\tilde{\Lambda}$. We show it is a martingale and consequently, a new measure is constructed based on $\tilde{\Lambda}$.

Define stochastic process $\tilde{\Lambda}$ by[1]

$$\begin{aligned}\tilde{\Lambda}_t :=& \Big(\prod_{\tau_n \le t} \frac{1}{\lambda(X_{\tau_n-})}\Big) \exp\Big(-\int_0^t [h(X_s)]^\top dB_s \\ & - \frac{1}{2}\int_0^t \|h(X_s)\|^2 ds + \int_0^t (\lambda(X_s)-1)ds\Big), \quad t \in [0,T]. \end{aligned} \tag{2.4.1}$$

In what follows, we will show that $\tilde{\Lambda}$ is a $(\mathbb{P},\mathbb{F})$-martingale under some assumptions. The necessary and sufficient conditions for absolute continuity of measures have been studies for diffusion type by Liptser and Shiryaev (1974a), page 257-291, for point processes, by Liptser and Shiryaev (1974b), page 336-345, Brémaud (1981), Theorem T4, page 168, Theorem T11, page 242.

Now we introduce conditions under which $\tilde{\Lambda}$ is a martingale. For general case, the classical condition is Novikov's condition, see Protter (2005), Theorem 41, page 140. Notice, for this special case, B is independent of X. We have an alternative condition provided as follows.

Assumption 2.3. *We assume that $h : \mathbb{R}^d \to \mathbb{R}^l$ is a measurable function such that*

$$\mathbb{E}\Big[\int_0^T \|h(X_s)\|^2 ds\Big] < \infty. \tag{2.4.2}$$

Lemma 2.3. *Suppose that Assumptions 2.3 and 2.2 are fulfilled, then*

$$\mathbb{E}(\tilde{\Lambda}_t) = 1 \text{ for } t \in [0,T].$$

Proof. For sake of simplicity, define

$$\begin{aligned}\tilde{\Lambda}_{1,t} :=& \exp\Big[-\int_0^t [h(X_s)]^\top dB_s - \frac{1}{2}\int_0^t \|h(X_s)\|^2 ds\Big], \\ \tilde{\Lambda}_{2,t} :=& \Big\{\prod_{\tau_n \le t} \frac{1}{\lambda(X_{\tau_n-})}\Big\} \exp\Big[\int_0^t (\lambda(X_s)-1)ds\Big].\end{aligned}$$

With Assumption 2.3, apply Liptser and Shiryaev (1974a), Example 4, page 234, or Note 3, page 278, we get

$$\mathbb{E}(\tilde{\Lambda}_{1,t}|\mathcal{F}_t^X) = 1. \tag{2.4.3}$$

With Assumption 2.2, Brémaud (1981), Theorem T11, page 242, shows

$$\mathbb{E}(\tilde{\Lambda}_{2,t}|\mathcal{F}_t^X) = 1. \tag{2.4.4}$$

[1] The product $\prod_{\tau_n \le t}$ is taken to be 1 if $\tau_1 > t$.

Combining Equation (2.4.3) and (2.4.4), we get

$$\begin{aligned}\mathbb{E}(\tilde{\Lambda}_t) =& \mathbb{E}\Big[\mathbb{E}(\tilde{\Lambda}_t|\mathcal{F}_t^X)\Big]\\ =& \mathbb{E}\Big[\mathbb{E}(\tilde{\Lambda}_{1,t}\tilde{\Lambda}_{2,t}|\mathcal{F}_t^X)\Big]\\ =& \mathbb{E}\Big[\mathbb{E}(\tilde{\Lambda}_{1,t}|\mathcal{F}_t^X)\mathbb{E}(\tilde{\Lambda}_{2,t}|\mathcal{F}_t^X)\Big] = 1.\end{aligned}$$

The desired result is obtained.

For reader's convenience we present the proof of Equation (2.4.3) and (2.4.4) here. Note that B is independent of X, by Assumption (2.4.2), applying Shreve (2004), Theorem 4.4.9, given $\mathcal{F}_t^X$, $-\int_0^t [h(X_s)]^\top dB_s$ is normally distributed with mean zero and variance $\int_0^t \|h(X_s)\|^2 ds$. So we obtain Equation (2.4.3). Furthermore, notice

$$\mathbb{E}\Big(\tilde{\Lambda}_{2,t}\Big|\mathcal{F}_t^X\Big) = \exp\Big[\int_0^t (\lambda(X_s)-1)ds\Big]\mathbb{E}\Big(\prod_{\tau_n\le t}\frac{1}{\lambda(X_{\tau_n -})}\Big|\mathcal{F}_t^X\Big).$$

Given $\mathcal{F}_t^X$ and $\tau_j \le t < \tau_{j+1}$, the variables $\tau_1 < \tau_2 \ldots < \tau_j$ are distributed like j order statistics from a sample of independent random variable with density $\lambda(X_s)/(\int_0^t \lambda(X_s)ds)$, $s \in [0,t]$. So

$$\begin{aligned}\mathbb{E}\Big(\prod_{\tau_n\le t}\frac{1}{\lambda(X_{\tau_n -})}\Big|\mathcal{F}_t^X\Big) =& \mathbb{E}\Big(\mathbf{1}_{\{\tau_1>t\}}|\mathcal{F}_t^X\Big) + \sum_{j=1}^{\infty}\mathbb{E}\Big(\mathbf{1}_{\tau_j\le t\le\tau_{j+1}}\prod_{n=1}^{j}\frac{1}{\lambda(X_{\tau_n -})}\Big|\mathcal{F}_t^X\Big)\\ =& \exp\Big[-\int_0^t \lambda(X_s)ds\Big]\sum_{j=0}^{\infty}\frac{(\int_0^t\lambda(X_s)ds)^j}{j!}\cdot\frac{t^j}{(\int_0^t\lambda(X_s)ds)^j}\\ =& \exp\Big[-\int_0^t \lambda(X_s)ds\Big]\exp(t).\end{aligned}$$

Equation (2.4.4) is obtained. □

It is now time to show that $\tilde{\Lambda}$ is a martingale.

Proposition 2.4. *Suppose that Assumptions 2.3 and 2.2 are fulfilled, then process* $\{\tilde{\Lambda}_t\}_{0\le t\le T}$*, defined by Equation* (2.4.1)*, is a nonnegative* $(\mathbb{P},\mathbb{F})$*-martingale.*

Proof. First we show $\{\tilde{\Lambda}_t\}_{0\le t\le T}$ is a $(\mathbb{P},\mathbb{F})$-nonnegative-local martingale and a $(\mathbb{P},\mathbb{F})$-supermartingale. By Itô's formula, process $\tilde{\Lambda}$ satisfies the equation

$$\tilde{\Lambda}_t = 1 - \int_0^t \tilde{\Lambda}_{s-}[h(X_s)]^\top dB_s - \int_0^t \tilde{\Lambda}_{s-}\frac{\lambda(X_{s-})-1}{\lambda(X_{s-})}d\Big(Y_s - \int_0^s \lambda(X_u)du\Big).$$

Define

$$\tilde{S}_n = \begin{cases} \inf\Big\{t \le T|\tilde{\Lambda}_{t-}\ge n & \\ \qquad \text{or } \int_0^t \|h(X_s)\|^2 ds \ge n \text{ or } \int_0^t |\lambda(X_s)-1|ds \ge n\Big\}, & \text{if } \{...\}\ne\emptyset,\\ T, & \text{otherwise.}\end{cases}$$

Applying Brémaud (1981), (II, T8) and Shreve (2004), Theorem 11.4.5, we obtain $\tilde{\Lambda}_{t\wedge\tilde{S}_n}$ is a $(\mathbb{P},\mathbb{F})$-martingale. Now λ is bounded, therefore, Y has only finitely many jumps in $[0,T]$. Notice that $\tilde{\Lambda}_{t-}$ is a left-continuous process, we have $\sup_{t\in[0,T]}|\tilde{\Lambda}_{t-}| < \infty$, a.s.. Moreover we have $\int_0^T \|h(X_s)\|^2 ds < \infty$, $\mathbb{P}-a.s.$, λ is bounded, $\mathbb{P}-a.s.$. Therefore $\tilde{\Lambda}_t$ is a $(\mathbb{P},\mathbb{F})$-nonnegative local martingale. It is nonnegative, by Lemma 2.3

$$\mathbb{E}(|\tilde{\Lambda}_t|) = \mathbb{E}(\tilde{\Lambda}_t) = 1 < \infty.$$

It is integrable and therefore, by Brémaud (1981), (I, E8), page 8, it is a $(\mathbb{P},\mathbb{F})$-supermartingale. Furthermore, application of Lemma 2.3, Brémaud (1981) I, E6, page 7, shows that $\{\tilde{\Lambda}_t\}_{0\le t\le T}$ is a $(\mathbb{P},\mathbb{F})$-martingale. □

Define the new measure $\mathbb{P}^0$ on the measureable space $(\Omega,\mathcal{F})$ by

$$\mathbb{P}^0(A) = \int_A \tilde{\Lambda}_T(\omega)\mathbb{P}(d\omega),$$

for all $A\in\mathcal{F}$. Denote by $\mathbb{E}^0$ the expectation w.r.t. $\mathbb{P}^0$.

Proposition 2.5. *If Assumption 2.3 and 2.2 hold, then,*

1 $\mathbb{P}^0$ is a probability measure.

2 The law of the process X under $\mathbb{P}^0$ is the same as its law under $\mathbb{P}$.

2 Under $\mathbb{P}^0$, Z is a standard Brownian motion.

4 Under $\mathbb{P}^0$, Y is a standard Poisson process with intensity 1.

5 Under $\mathbb{P}^0$, the processes X, Z and Y are independent.

Proof. Results follow from Proposition 2.4 and Girsanov's theorem for semimartingales, see for instance, Jacod and Shiryaev (2003), Theorem 3.24, page 172. □

Lemma 2.6. *Suppose that Assumption 2.3 and 2.2 are fulfilled. Define*

$$\Lambda_t := \tilde{\Lambda}_t^{-1}, \quad t\in[0,T], \tag{2.4.5}$$

where $\tilde{\Lambda}$ is defined by Equation (2.4.1)*. Then Λ is a $(\mathbb{P},\mathbb{F})$-martingale. $\mathbb{E}^0(\Lambda_t) = 1$ and $\Lambda_t = \mathbb{E}^0(\frac{d\mathbb{P}}{d\mathbb{P}^0}|\mathcal{F}_t)$ for all $t\ge 0$.*

Proof. First we have

$$\mathbb{E}^0(\Lambda_t) = \mathbb{E}(\tilde{\Lambda}_t\Lambda_t) = 1. \tag{2.4.6}$$

By Brémaud (1981), (I, E6), page 7, in order to show Λ_t is a martingale with respect to $\mathcal{F}_t$ and $\mathbb{P}^0$, it is suffices to show it is a $(\mathbb{P},\mathbb{F})$-supermartingale. This can be obtained similarly as the proof of Proposition 2.4. To be precisely, by the definition of Λ,

$$\begin{aligned}\Lambda_t :=& \Big\{\prod_{\tau_n\le t}\lambda(X_{\tau_n -})\Big\}\exp\Big(\int_0^t [h(X_s)]^\top dB_s + \frac{1}{2}\int_0^t \|h(X_s)\|^2 ds - \int_0^t (\lambda(X_s)-1)ds\Big)\\ =& \Big\{\prod_{\tau_n\le t}\lambda(X_{\tau_n -})\Big\}\exp\Big(-\int_0^t [h(X_s)]^\top dY_s - \frac{1}{2}\int_0^t \|h(X_s)\|^2 ds - \int_0^t (\lambda(X_s)-1)ds\Big).\end{aligned}$$

Applying Itô's formula

$$\Lambda_t = 1 + \int_0^t \Lambda_{s-}\Big\{[h(X_s)]^\top dZ_s + (\lambda(X_{s-}) - 1)d(Y_s - s)\Big\}.$$

Define

$$S_n = \begin{cases} \inf\Big\{t \le T|\Lambda_{t-} \ge n \text{ or } \int_0^t \|h(X_s)\|^2 ds \ge n \text{ or } \int_0^t |\lambda(X_s) - 1|ds \ge n\Big\}, & \text{if } \{...\} \ne \emptyset, \\ T, & \text{otherwise.} \end{cases}$$

Applying Brémaud (1981), (II, T8) and Shreve (2004), Theorem 11.4.5, we obtain $\tilde{\Lambda}_{t\wedge S_n}$ is a $(\mathbb{P}, \mathbb{F})$-martingale. Now, Y is a Poisson processes with intensity 1, therefore, Y has only finitely many jumps in $[0, T]$. Notice that Λ_{t-} is a left-continuous process, we have $\sup_{t\in[0,T]} |\Lambda_{t-}| < \infty$, a.s.. Moreover we have $\int_0^T \|h(X_s)\|^2 ds < \infty$, $\mathbb{P} - a.s.$, λ is bounded, $\mathbb{P} - a.s.$. Therefore, Λ_t is a $(\mathbb{P}, \mathbb{F})$-nonnegative local martingale. It is nonnegative, by Equation (2.4.6),

$$\mathbb{E}(|\tilde{\Lambda}_t|) = \mathbb{E}(\tilde{\Lambda}_t) = 1 < \infty.$$

It is integrable and therefore, By Brémaud (1981), (I, E8), page 8, a $(\mathbb{P}, \mathbb{F})$-supermartingale.

□

2.4.2 The unnormalized filtering equations

Followed the idea from Bain and Crisan (2009), Proposition 3.15, page 56, we have the following result.

Proposition 2.7. *Let U be an integrable $\mathcal{F}_t$-measurable random variable. Then*

$$\mathbb{E}^0[U|\mathcal{F}_t^{Z,Y}] = \mathbb{E}^0[U|\mathcal{F}_T^{Z,Y}].$$

Proof. Let us denote by $\tilde{\mathcal{F}}_t^{Z,Y} = \sigma(Z_{t+u} - Z_t, Y_{t+u} - Y_t; 0 \le u \le T - t)$, then $\mathcal{F}_T^{Z,Y} = \sigma(\mathcal{F}_t^{Z,Y}, \tilde{\mathcal{F}}_t^{Z,Y})$. Under $\mathbb{P}^0$, $\tilde{\mathcal{F}}_t^{Z,Y} \subset \mathcal{F}_T^{Z,Y}$ is independent of $\mathcal{F}_t$ because Z is an $\mathcal{F}_t$-adapted Brownian motion and Y is Poisson process with intensity 1. Noting that U is $\mathcal{F}_t$-adapted, using the properties of the conditional expectation, we get

$$\mathbb{E}^0[U|\mathcal{F}_t^{Z,Y}] = \mathbb{E}^0\Big[U\Big|\sigma(\mathcal{F}_t^{Z,Y}, \tilde{\mathcal{F}}_t^{Z,Y})\Big] = \mathbb{E}^0[U|\mathcal{F}_T^{Z,Y}]. \tag{2.4.7}$$

□

In the following proposition, known as Kallianpur-Striebel formula, see Kallianpur and Striebel (1968), we show that the distribution of X_t given $\mathcal{F}_t^{Z,Y}$ under the original measure $\mathbb{P}$ can be calculated in terms of conditional expectations under the new measure $\mathbb{P}^0$. In other word, it is suffice to compute the numerator on the right hand side of Equation (2.4.8).

Proposition 2.8 (Kallianpur-Striebel)**.** *Suppose that Assumptions 2.3 and 2.2 are fulfilled. For any $f \in L^\infty(\mathbb{R}^d)$, $t \in [0, T]$,*

$$\mathbb{E}\Big(f(X_t)\Big|\mathcal{F}_t^{Z,Y}\Big) = \frac{\mathbb{E}^0(f(X_t)\Lambda_t|\mathcal{F}_T^{Z,Y})}{\mathbb{E}^0(\Lambda_t|\mathcal{F}_T^{Z,Y})} =: \frac{\rho_t(f)}{\rho_t(1)}. \tag{2.4.8}$$

Proof. See the proof of Bain and Crisan (2009), Proposition 6.1. □

In the following, we further assume that

$$\mathbb{P}^0\Big[\int_0^T [\rho_s(\|h\|)]^2 ds < \infty\Big] = 1, \tag{2.4.9}$$

$$\mathbb{P}^0\Big[\int_0^T [\rho_s(1)]^2 ds < \infty\Big] = 1. \tag{2.4.10}$$

Under condition (2.4.9), the stochastic integral $\int_0^t \rho_s(fh^\top)dZ_s$ is a $(\mathbb{P}^0, \mathbb{F})$-local martingale for any bounded measurable function f. Moreover, the stochastic integral $\int_0^t \rho_{s-}\big(f(\lambda-1)\big)d(Y_s - s)$ is a $(\mathbb{P}^0, \mathbb{F})$-martingale for any bounded measurable function f in view of bounded of λ, by integration theorem. For details see Brémaud (1981), Theorem T8, page 27.

Now, we are ready to present the main result of this section. In the following, we show $\{\rho_t\}_{t\in[0,T]}$ satisfies the following linear SPDE.

Theorem 2.9. *Suppose that Assumptions 2.3 and 2.2 are fulfilled. Further more, if conditions* (2.4.9) *and* (2.4.10) *are satisfied then the processes ρ satisfies*

$$\begin{aligned}\rho_t(f) =& \pi_0(f) + \int_0^t \rho_s(\mathcal{L}f)ds + \int_0^t \rho_s(fh^\top)dZ_s \\ &+ \int_0^t \rho_{s-}\Big(f(\lambda-1)\Big)d(Y_s - s), \ \ \mathbb{P}^0 - a.s., \ \ \forall t \in [0,T],\end{aligned} \tag{2.4.11}$$

for any $f \in D(\mathcal{L})$.

The approach leading to the dynamics evolution of for the unnormalized conditional measure has been developed in doctoral Duncan (1967), Mortensen (1966) and the important paper of Zakai (1969). The linear SPDE (2.4.11) is therefore known as the Duncan-Mortensen-Zakai equation, or simply, Zakai equation.

Proof of Theorem 2.9. We first approximate Λ_t with Λ_t^ε given by $\Lambda_t^\varepsilon = \frac{\Lambda_t}{1+\varepsilon\Lambda_t}$, $\varepsilon > 0$. The definition of Λ_t implies that

$$d\Lambda_t = \Lambda_{t-}\Big[h(X_t)^\top dZ_t + (\lambda(X_{t-}) - 1)d(Y_t - t)\Big]. \tag{2.4.12}$$

The Itô formula for jump process, see for instance Shreve (2004) Theorem 11.5.1, shows

$$\begin{aligned}\Lambda_t^\varepsilon =& \Lambda_0^\varepsilon + \int_0^t \frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2} h(X_s)^\top dZ_s + \int_0^t -\frac{\varepsilon\Lambda_s^2}{(1+\varepsilon\Lambda_s)^3}\|h(X_s)\|^2 ds \\ &- \int_0^t \frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}(\lambda(X_s) - 1)ds + \sum_{\tau_n \le t} \triangle \frac{\Lambda_{\tau_n}}{1+\varepsilon\Lambda_{\tau_n}}.\end{aligned} \tag{2.4.13}$$

The product rule for semimartingales, together with Equation (2.4.13) and (2.2.5), implies that

$$\begin{aligned}\Lambda_t^\varepsilon f(X_t) =& \Lambda_0^\varepsilon f(X_0) + \int_0^t \Lambda_{s-}^\varepsilon df(X_s) + \int_0^t f(X_{s-}) d\Lambda_s^\varepsilon + [\Lambda^\varepsilon, f(X)]_t \\ =& \Lambda_0^\varepsilon f(X_0) + \int_0^t \Lambda_s^\varepsilon (\mathcal{L}f)(X_s) ds + \int_0^t \Lambda_s^\varepsilon dM_s^f \\ &+ \int_0^t f(X_s) \frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2} h(X_s)^\top dZ_s \\ &+ \int_0^t f(X_s)(-\frac{\varepsilon\Lambda_s^2}{(1+\varepsilon\Lambda_s)^3}\|h(X_s)\|^2) ds \\ &- \int_0^t f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}(\lambda(X_s)-1) ds \\ &+ \int_0^t \frac{f(X_{s-})\Lambda_{s-}(\lambda(X_{s-})-1)}{(1+\varepsilon\Lambda_{s-}\lambda(X_{s-}))(1+\varepsilon\Lambda_{s-})} dY_s.\end{aligned}$$

Taking conditional expectations on both sides, we get

$$\begin{aligned}\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\left[\Lambda_t^\varepsilon f(X_t)\right] =& \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\left[\Lambda_0^\varepsilon f(X_0)\right] \\ &+ \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\int_0^t \Lambda_s^\varepsilon(\mathcal{L}f)(X_s) ds\Big] + \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\int_0^t \Lambda_s^\varepsilon dM_s^f\Big] \\ &+ \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\int_0^t f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2} h(X_s)^\top dZ_s\Big] \\ &+ \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\int_0^t f(X_s)(-\frac{\varepsilon\Lambda_s^2}{(1+\varepsilon\Lambda_s)^3}\|h(X_s)\|^2) ds\Big] \\ &- \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\int_0^t f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}(\lambda(X_s)-1) ds\Big] \\ &+ \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\int_0^t \frac{f(X_{s-})\Lambda_{s-}(\lambda(X_{s-})-1)}{(1+\varepsilon\Lambda_{s-}\lambda(X_{s-}))(1+\varepsilon\Lambda_{s-})} dY_s\Big]. \\ :=& E_1 + E_2 + E_3 + E_4 + E_5 + E_6 + E_7,\end{aligned} \tag{2.4.14}$$

correspondingly. Compare to the Equation (2.4.11), it remains to show that, $\mathbb{P}^0 - a.s.$, as $\varepsilon \to 0$,

$$\begin{aligned}&\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\left[\Lambda_t^\varepsilon f(X_t)\right] \to \rho_t(f), \quad E_1 \to \pi_0(f), \quad E_2 \to \int_0^t \rho_s(\mathcal{L}f) ds, \quad E_3 = 0, \\ &E_4 \to \int_0^t \rho_s(fh^\top) dZ_s, \quad E_5 = 0, \quad E_6 \to \int_0^t \rho_s(\lambda - 1) ds, \quad E_7 \to \int_0^t \rho_{s-}\Big(f(\lambda-1)\Big) dY_s.\end{aligned}$$

In the following, we show this step by step. First, by the pointwise convergence of $\Lambda_t^\varepsilon \to \Lambda_t$,

$$\lim_{\varepsilon\to 0} \Lambda_t^\varepsilon f(X_t) = \Lambda_t f(X_t).$$

We have that

$$\mathbb{E}^0|\Lambda_t^\varepsilon f(X_t)| \le \|f\|_\infty \mathbb{E}^0(\Lambda_t) = \|f\|_\infty \mathbb{E}(\tilde\Lambda_t\Lambda_t) = \|f\|_\infty < \infty,$$

as $\tilde\Lambda_t\Lambda_t = 1$. Then, the dominated convergence theorem gives that

$$\lim_{\varepsilon\to 0} \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}[\Lambda_t^\varepsilon f(X_t)] = \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}[\Lambda_t f(X_t)] = \rho_t(f), \;\; \mathbb{P}^0 - a.s. \tag{2.4.15}$$

In a similar way,

$$\lim_{\varepsilon\to 0} E_1 = \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}[\Lambda_0 f(X_0)] = \pi_0(f), \ \ \mathbb{P}^0 - a.s. \tag{2.4.16}$$

Now we consider E_2. Note that,

$$\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big|\int_0^t \Lambda_s^\varepsilon(\mathcal{L}f)(X_s)ds\Big| = \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big|\int_0^t \frac{\varepsilon\Lambda_s}{1+\varepsilon\Lambda_s}\frac{1}{\varepsilon}(\mathcal{L}f)(X_s)ds\Big| \le \frac{1}{\varepsilon}\|\mathcal{L}f\|_\infty T < \infty.$$

By Fubini's theorem, we rewrite E_2 as

$$E_2 = \int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\Lambda_s^\varepsilon(\mathcal{L}f)(X_s)\Big]ds. \tag{2.4.17}$$

Moreover,

$$\lim_{\varepsilon\to 0}\Lambda_s^\varepsilon(\mathcal{L}f)(X_s) = \Lambda_s(\mathcal{L}f)(X_s).$$

We have that

$$\begin{aligned}
\mathbb{E}^0\Big(\int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}|\Lambda_s^\varepsilon(\mathcal{L}f)(X_s)|ds\Big) \le& \mathbb{E}^0\Big(\int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}(\Lambda_s\|\mathcal{L}f\|_\infty)ds\Big) \\
=& \|\mathcal{L}f\|_\infty\int_0^t \mathbb{E}^0(\Lambda_s)ds \\
=& \|\mathcal{L}f\|_\infty\int_0^t \mathbb{E}(\tilde{\Lambda}_s\Lambda_s)ds = \|\mathcal{L}f\|_\infty t < \infty,
\end{aligned}$$

such that the dominated convergence theorem implies the desired result for E_2:

$$\lim_{\varepsilon\to 0} E_2 = \int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\Lambda_s(\mathcal{L}f)(X_s)\Big]ds = \int_0^t \rho_s(\mathcal{L}f)ds, \ \ \mathbb{P}^0 - a.s. \tag{2.4.18}$$

Since Λ_t^ε is bounded, apply Protter (2005), Corollary 3, page 73 and Bain and Crisan (2009), Lemma 3.21, page 59,

$$E_3 = 0. \tag{2.4.19}$$

Before considering E_4, we first have the following square integrability,

$$\begin{aligned}
\mathbb{E}^0\Big[\int_0^t f^2(X_s)\Big(\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}\Big)^2\|h(X_s)\|^2ds\Big] =& \frac{1}{\varepsilon}\mathbb{E}^0\Big[\int_0^t f^2(X_s)\frac{\varepsilon\Lambda_s}{(1+\varepsilon\Lambda_s)^4}\Lambda_s\|h(X_s)\|^2ds\Big] \\
\le& \frac{\|f\|_\infty^2}{\varepsilon}\mathbb{E}^0\Big[\int_0^t \Lambda_s\|h(X_s)\|^2ds\Big] \\
=& \frac{\|f\|_\infty^2}{\varepsilon}\int_0^t \mathbb{E}\Big[\tilde{\Lambda}_s\Lambda_s\|h(X_s)\|^2\Big]ds \\
=& \frac{\|f\|_\infty^2}{\varepsilon}\int_0^t \mathbb{E}\Big[\|h(X_s)\|^2\Big]ds < \infty.
\end{aligned}$$

The last inequality follows from Equation (2.4.2). According to Bain and Crisan (2009), Lemma 6.6, we change the order of conditional expectation and stochastic integral, and we rewrite E_4 equivalently as

$$E_4 = \int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}h(X_s)^\top\Big]dZ_s.$$

Now consider the following process

$$t \mapsto \int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\left[f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}h(X_s)^\top\right]dZ_s. \tag{2.4.20}$$

We now show this is a martingale by using Jensen's inequality, Fubini's Theorem and Equation (2.4.2),

$$\begin{aligned}
&\mathbb{E}^0\Big\{\int_0^t\Big(\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}h(X_s)\Big]\Big)^2 ds\Big\}\\
\leq&\mathbb{E}^0\Big\{\int_0^t\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\Big(f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}h(X_s)\Big)^2\Big]ds\Big\}\\
\leq&\mathbb{E}^0\Big\{\int_0^t\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[f^2(X_s)\frac{1}{\varepsilon}\frac{\varepsilon\Lambda_s}{(1+\varepsilon\Lambda_s)^4}\Lambda_s\|h(X_s)\|^2\Big]ds\Big\}\\
\leq&\frac{\|f\|_\infty^2}{\varepsilon}\mathbb{E}^0\Big[\int_0^t\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}[\Lambda_s\|h(X_s)\|^2]ds\Big]\\
=&\frac{\|f\|_\infty^2}{\varepsilon}\int_0^t\mathbb{E}^0\Big[\Lambda_s\|h(X_s)\|^2\Big]ds\\
=&\frac{\|f\|_\infty^2}{\varepsilon}\int_0^t\mathbb{E}\Big[\|h(X_s)\|^2\Big]ds\leq\infty.
\end{aligned}$$

As Z is a standard Brownian motion under measure $\mathbb{P}^0$, the stochastic integral defined in Equation (2.4.20) is a $(\mathbb{P}^0, \mathbb{F})$-martingale by Shreve (2004), Theorem 4.3.1, page 134. Moreover, from the condition (2.4.9), the postulated limit process as $\varepsilon \to 0$,

$$\int_0^t \rho_s(fh^\top)dZ_s \tag{2.4.21}$$

is a local martingale. Thus, the difference of (2.4.20) and (2.4.21) is a local martingale:

$$\int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[\frac{\varepsilon\Lambda_s^2(2+\varepsilon\Lambda_s)}{(1+\varepsilon\Lambda_s)^2}f(X_s)h(X_s)^\top\Big]dZ_s. \tag{2.4.22}$$

Define

$$\xi_s := \frac{\varepsilon\Lambda_s^2(2+\varepsilon\Lambda_s)}{(1+\varepsilon\Lambda_s)^2}f(X_s)[h(X_s)]^\top.$$

Since the pointwise limit

$$\lim_{\varepsilon\to 0}\xi_s = \mathbf{0}$$

and, taking into account Equation (2.4.9), we get the following dominating,

$$\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\|\xi_s\| \leq \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}(2\|f\|_\infty\Lambda_s\|h(X_s)\|) = 2\|f\|_\infty\rho_s(\|h\|) < \infty,\ \ \mathbb{P}^0 - a.s.,$$

for almost every $s \in [0,T]$, dominated convergence theorem shows that for almost every $s \in [0,T]$,

$$\lim_{\varepsilon\to 0}\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\xi_s = \mathbf{0},\ \ \mathbb{P}^0 - a.s.$$

Furthermore, by Equation (2.4.9)

$$\int_0^t \Big[\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}(\xi_s)\Big]^2 ds \leq 4\|f\|_\infty^2 \int_0^t [\rho_s|h|]^2 ds < \infty,\ \ \mathbb{P}^0 - a.s..$$

The value of the integral $\int_0^t \lim_{\varepsilon\to 0}[\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}(\xi_s)]^2 ds$ is not modified with the change of the function $\lim_{\varepsilon\to 0}[\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}(\xi_s)]^2$ on the set of Lebesgue zero measure. The dominated convergence theorem implies

$$\lim_{\varepsilon\to 0}\int_0^t \Big[\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}(\xi_s)\Big]^2 ds = 0,\ \ \mathbb{P}^0 - a.s..$$

By central limit theorem for stochastic integrals, see Revuz and Yor (1999), page 152, the integral in Equation (2.4.22) convergence to $\mathbf{0}$, $\mathbb{P}^0 - a.s.$ Then, we get the desired result for E_4,

$$\lim_{\varepsilon\to 0} E_4 = \int_0^t \rho_s(fh^\top)dZ_s,\ \ \mathbb{P}^0 - a.s.. \tag{2.4.23}$$

The desired result of E_5

$$\lim_{\varepsilon\to 0} E_5 = 0,\ \ \mathbb{P}^0 - a.s. \tag{2.4.24}$$

is obtain as a consequence of dominated convergence theorem by the pointwise limit,

$$\lim_{\varepsilon\to 0} f(X_s)\Big(-\frac{\varepsilon\Lambda_s^2}{(1+\varepsilon\Lambda_s)^3}\|h(X_s)\|^2\Big) = 0$$

and the dominating,

$$\begin{aligned}\mathbb{E}^0\Big[\int_0^t |f(X_s)(-\frac{\varepsilon\Lambda_s^2}{(1+\varepsilon\Lambda_s)^3}\|h(X_s)\|^2)|ds\Big] \leq & \|f\|_\infty \mathbb{E}^0\Big[\int_0^t \Lambda_s\|h(X_s)\|^2 ds\Big] \\ = & \|f\|_\infty \int_0^t \mathbb{E}\Big[\|h(X_s)\|^2\Big] ds < \infty.\end{aligned}$$

Now we consider E_6. Since, λ is bounded, we get the following boundedness,

$$\begin{aligned}\mathbb{E}^0\Big[\int_0^t |f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}(\lambda(X_s)-1)|ds\Big] \leq & \|f\|_\infty \mathbb{E}^0\Big[\int_0^t \Lambda_s|\lambda(X_s)-1|ds\Big] \\ = & \|f\|_\infty \int_0^t \mathbb{E}\Big(|\lambda - 1|\Big) ds \\ < & \infty.\end{aligned}$$

Guaranteed by Fubini's theorem, we change the order of conditional expectation and integral, and rewrite E_6 equivalently as

$$E_6 = \int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big[f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}(\lambda(X_s)-1)\Big] ds.$$

Since the pointwise limit

$$\lim_{\varepsilon\to 0} f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}(\lambda(X_s)-1) = f(X_s)\Lambda_s(\lambda(X_s)-1),$$

and, in view of bounded, the dominating

$$\mathbb{E}^0\Big[\int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}|f(X_s)\frac{\Lambda_s}{(1+\varepsilon\Lambda_s)^2}(\lambda(X_s)-1)|ds\Big] = \|f\|_\infty \int_0^t \mathbb{E}(|\lambda-1|)ds < \infty,$$

dominated convergence theorem shows the desired result for E_6,

$$\lim_{\varepsilon\to 0} E_6 = \int_0^t \rho_s(\lambda-1)ds, \;\; \mathbb{P}^0 - a.s.. \tag{2.4.25}$$

It remains to study E_7. For sake of simplicity, define

$$H_t^\varepsilon := \frac{f(X_{s-})\Lambda_{s-}(\lambda(X_{s-})-1)}{(1+\varepsilon\Lambda_{s-}\lambda(X_{s-}))(1+\varepsilon\Lambda_{s-})},$$

then E_7 can be rewritten as,

$$E_7 = \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big(\int_0^t H_s^\varepsilon dY_s\Big).$$

Notice the boundedness

$$|H_t^\varepsilon| \leq \frac{\|f\|_\infty}{\varepsilon^2} < \infty,$$

change the order of stochastic integral and conditional expectation, which is guaranteed by stochastic Fubini's theorem, see Protter (2005), Theorem 64, page 207, E_7 can be rewritten as

$$E_7 = \int_0^t \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}(H_s^\varepsilon)dY_s.$$

Combining the pointwise limit

$$\lim_{\varepsilon\to 0} H_t^\varepsilon = f(X_{t-})\Lambda_{t-}\Big(\lambda(X_{t-})-1\Big), \;\; \mathbb{P}^0 - a.s,$$

and the dominating

$$\mathbb{E}^0(\mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}|H_t^\varepsilon|) < \mathbb{E}^0\Big(\|f\|_\infty\|\lambda-1\|_\infty \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}(\Lambda_{t-})\Big) = \|f\|_\infty\|\lambda-1\|_\infty < \infty,$$

stochastic dominated convergence theorem, see Protter (2005), Theorem 32, page 174, implies that for almost every $s \in [0,T]$,

$$\lim_{\varepsilon\to 0} \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big(H_t^\varepsilon\Big) = \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}\Big(f(X_{t-})\Lambda_{t-}(\lambda(X_{t-})-1)\Big).$$

On the other hand

$$\begin{aligned}\rho_{t-}\Big(f(\lambda-1)\Big) &= \lim_{s\uparrow t}\rho_s\Big(f(\lambda-1)\Big)\\ &= \lim_{s\uparrow t}\mathbb{E}^0\Big(f(X_s)(\lambda(X_s)-1)\Lambda_s|\mathcal{F}^{Z,Y}\Big)\\ &= \mathbb{E}^0\Big(f(X_{t-}(\lambda(X_{s-})-1)\Lambda_{t-})|\mathcal{F}^{Z,Y}\Big),\end{aligned}$$

where the last equality follows from Liptser and Shiryaev (1974a), Theorem 1.6, page 17. Hence we get

$$\lim_{\varepsilon\to 0} \mathbb{E}^0_{\mathcal{F}_T^{Z,Y}}(H_t^\varepsilon) = \rho_{t-}\Big(f(\lambda-1)\Big).$$

By Equation (2.4.10), together with boundedness of f and λ, dominated convergence theorem shows that

$$\lim_{\varepsilon\to 0} E_7 = \int_0^t \rho_{s-}\Big(f(\lambda-1)\Big)dY_s, \ \ \mathbb{P}^0 - a.s.. \tag{2.4.26}$$

Summing up, the Zakai equation is obtained. □

Chapter 3

Numerical methods

We are interested in the conditional distribution of the state process X given the past observation. In general, this is an infinite dimensional problem, i.e. we cannot obtain the conditional distribution by a finite set of parameters. Therefore, it is normal to look for a finite dimensional approximation which can be used in practical applications. There are various numerical methods for this, some of which we list below. For a more comprehensive listing of these methods , we refer to Budhiraja, Chen, and Lee (2007), Bain and Crisan (2009), page 191-217.

To begin, we have the spectral approach for SPDEs which is based on the Cameron-Martin version of the Wiener Chaos expansion. The main advantage of the spectral approach, as compared to most other nonlinear filtering algorithms, is that the time consuming computations, including solving partial differential equations and evaluation of integrals, are performed in advance. The real-time part is relatively simple, even when the dimension of the state process is large. For further details, see Lototsky (2006). Unfortunately, the spectral approach only works for SPDEs driven by white noise and is not useful for Model (2.2.9) with jump.

Next, we have the extended Kalman filter which is a linearized approximation of the original problem. The EKF does not always perform well and, in fact, performs poorly if the nonlinearities are strong, see Bain and Crisan (2009), page 196. It will give a good estimate only when the coefficients are 'slightly' nonlinear, see Bain and Crisan (2009), Theorem 8.5, page 195. The accuracy of results obtained using EKF is neither verifiable, nor reliable. It only works for SPDEs driven by white noise and is not useful for Model (2.2.9) with additional jump observations.

In Section 2.1.3, we show the filter for a finite-state Markov chain is finite dimensional. Therefore, it is logical to consider approximations of the state process X by a sequence X^n which are finite-state Markov chains, see for instance Frey and Runggaldier (2010).

Similar to the Markov chain approximation method, the particle methods are a finite dimensional approximation of unnormalized measure ρ by discretisation of the state variable, see for instance Carpenter, Clifford, and Fearnhead (1999), Crisan, Moral, and Lyons (1999). However, particle methods are more flexible and easier to implement. The basic idea is to approximate the conditional expectation by Monto Carlo methods.

In Chapter 5, we will introduce an additional numerical method. We will show that the unnormalized conditional density is the solution of a partial differential equation, although a stochastic one. Therefore, we can apply classical PDE methods, such as the Galerkin method (see for instance Germani and Piccioni (1984)), to the stochastic PDEs and obtain a density approximation. A detailed introduction of Galerkin method follows in Chapters 6 and 7.

Here, in Chapter 3, we summarize the existing results concerning the finite-dimensional approximation of Model (2.2.9), as introduced in Section 2.2. Throughout this Chapter, we will assume d, the dimension of the state process X, is 1. Generalization to the multi-dimensional case is easily obtained. We present three classical filters beginning in Section 3.1 with an introduction to a finite dimensional filter based on the conditional distributions are Gaussian functions. In Section 3.2, we introduce the filter obtained by the finite-state Markov chain approximation and, in Section 3.3, we introduce particle filter.

3.1 A finite dimensional filter

Since Model (2.2.9) has additional jump observations, we can not directly solve the corresponding nonlinear filtering problem using the Kalman-Bucy filter or the extended Kalman filter(EKF). In this section, we provide a finite dimensional approximation of the nonlinear filtering problem w.r.t. Model (2.2.9), where λ is specified as a exponential function, see Fontana and Runggaldier (2010). We will show, after approximation, the conditional distribution is a linear combination of Gaussian functions and consequently, it can be determined by a set of finite parameters.

In Section 3.1.1, we introduce two special cases for which the conditional distribution is Gaussian or a linear combination of Gaussian functions. In Section 3.1.2, we introduce a numerical method for Model (2.2.9), motivated by these two special cases.

3.1.1 Two special cases

In Section 2.1.3, we gave an introduction of the classical Kalman-Bucy filter where the signal is Gaussian and the observation is linear. The conditional distribution is also Gaussian and consequently, the filter is determined by two parameters: the conditional mean and conditional variance.

Now, we will study the nonlinear filtering problem w.r.t. Model (2.2.9) which is with additional jumps observations. We introduce two special cases where for Case 1, the Gaussianity is preserved between jumps, and for Case 2, the Gaussianity is preserved at times of jump.

Case1

In this case, we consider linear model and include additional jump observation. We show, for this case, the Gaussianity is preserved until the time of the first jump.

Consider Model (2.1.4), with additional jump observation, the intensity is a quadratic function of the state process X. The reason we assume quadratic instead of linearity is to preserve the positivity of the intensity. More precisely, we consider the following model, for $t \in [0, T]$,

$$\begin{cases} dX_t = (\tilde{b}_t X_t + \tilde{b}_t^0)dt + \tilde{\sigma}_t dV_t, \\ dZ_t = (\tilde{h}_t X_t + \tilde{h}_t^0)dt + dB_t, \\ Y_t \text{ is a doubly stochastic Poisson process with intensity } \tilde{\lambda}_t^2 X_t^2 + \tilde{\lambda}_t^1 X_t + \tilde{\lambda}_t^0, \\ \qquad\qquad\qquad \text{and jumping times } \tau_1,\ \tau_2,\ \ldots. \end{cases} \tag{3.1.1}$$

Here, the assumptions of X and Z are as same as assumptions in Section 2.1.3. Moreover, we assume that the coefficients $\tilde{\lambda}_t^2$, $\tilde{\lambda}_t^1$ and $\tilde{\lambda}_t^0$ are deterministic $\mathbb{R}$-valued processes, and for $t \in [0,T]$,

$$\tilde{\lambda}_t^2 > 0, \quad \text{and} \quad \tilde{\lambda}_t^0 - \frac{(\tilde{\lambda}_t^1)^2}{4\tilde{\lambda}_t^2} \geq 0 \tag{3.1.2}$$

to guarantee the positivity of the jump intensity.

We now study the conditional distribution of X_t given past observations of Z and Y before τ_1, the time of the first jump. We have the following conjecture: Before τ_1, that is , for $0 \leq t < \tau_1$, the conditional distribution of X_t given the past observations of Z and Y is Gaussian with mean $\hat{X}_t$ and variance P_t defined as follows:

$$\hat{X}_t := \mathbb{E}[X_t | \mathcal{F}_t^{Z,Y}], \quad P_t := \mathbb{E}\Big[(X_t - \hat{X}_t)^2 \Big| \mathcal{F}_t^{Z,Y}\Big]. \tag{3.1.3}$$

And the process $\big\{(\hat{X}_t, P_t)\big\}_{t\in[0,\tau_1)}$ is the unique solution to the following equations, for $0 \leq t < \tau_1$,

$$\begin{cases} d\hat{X}_t = (\tilde{b}_t\hat{X}_t + \tilde{b}_t^0)dt + \tilde{h}_t P_t\Big[dZ_t - (\tilde{h}_t\hat{X}_t + \tilde{h}_t^0)dt\Big] - \Big[2\tilde{\lambda}_t^2 P_t\hat{X}_t + \tilde{\lambda}_t^1 P_t\Big]dt, \\ \frac{d}{dt}P_t = \tilde{\sigma}_t^2 + 2\tilde{b}_t P_t - \tilde{h}_t^2 P_t^2 - 2\tilde{\lambda}_t^2 P_t^2, \end{cases}$$

with $\hat{X}_0 = \mathbb{E}(X_0)$ and $P_0 = \mathbb{E}\Big[(X_0 - \hat{X}_0)^2\Big]$.

Remark 3.1. *The proof of the Gaussianity preserving becomes too involved to be covered here. We only present the idea of the proof. To simplify, we consider the simple case of Model* (3.1.1)*. That is for $t \in [0,T]$,*

$$\begin{cases} dX_t = \tilde{b}X_t dt + \tilde{\sigma}dV_t, \\ d\,Z_t = \tilde{h}X_t dt + dB_t, \\ Y_t \text{ is a doubly stochastic Poisson process with intensity } \tilde{\lambda}^2 X_t^2, \\ \qquad\qquad \text{and jumping times } \tau_1,\ \tau_2,\ \ldots. \end{cases} \tag{3.1.4}$$

where $\tilde{b}, \tilde{h} \in \mathbb{R}$, $\tilde{\sigma} \geq 0$, $\tilde{\lambda}^2 > 0$ are constants. The other assumptions are as same as Model (3.1.1)*. Now we assume $\big\{(\hat{X}_t, P_t)\big\}_{t\in[0,\tau_1)}$ is the solution of the following SDEs*

$$\begin{cases} d\hat{X}_t = \tilde{b}\hat{X}_t dt + \tilde{h}P_t\Big[dZ_t - \tilde{h}\hat{X}_t dt\Big] - \Big[2\tilde{\lambda}^2 P_t\hat{X}_t\Big]dt, \quad t \in [0,\tau_1) \\ \frac{d}{dt}P_t = \tilde{\sigma}^2 + 2\tilde{b}P_t - \tilde{h}^2 P_t^2 - 2\tilde{\lambda}^2 P_t^2, \end{cases}$$

with $\hat{X}_0 = \mathbb{E}(X_0)$ and $P_0 = \mathbb{E}[(X_0 - \hat{X}_0)^2]$. We then show in the following that, before τ_1, the conditional distribution of X_t given $\mathcal{F}_t^{Z,Y}$ is Gaussian with conditional mean $\hat{X}_t$ and conditional variance P_t. Let $\{q_t\}_{t\in[0,\tau_1)}$ be the density of the unnormalized conditional distribution of X, it can be shown it is the solution of the following SPDE

$$\begin{cases} dq_t = \mathcal{L}^* q_t dt + \tilde{h}x q_t dZ_t - (\tilde{\lambda}^2 x^2 - 1)q_t dt, \quad t \in [0,\tau_1) \\ q_0 = p_0, \end{cases} \tag{3.1.5}$$

where $\mathcal{L}$ is the generator of X, $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$ and

$$\mathcal{L}^* f = -\frac{\partial}{\partial x}(\tilde{b}xf) + \frac{1}{2}(\tilde{\sigma})^2\frac{\partial^2}{\partial x^2}f, \quad \forall f \in D(\mathcal{L}^*).$$

Recall, p_0 is the density of the law of X_0. Notice, X_0 is a normal distributed random variable with mean $\hat{X}_0$ and variance P_0, then

$$p_0(x) = \frac{1}{\sqrt{2\pi P_0}} e^{-\frac{(x-\hat{X}_0)^2}{2P_0}}, \quad x \in \mathbb{R}.$$

Define

$$u_t(x) := C_t \frac{1}{\sqrt{2\pi P_t}} e^{-\frac{(x-\hat{X}_t)^2}{2P_t}}, \quad t \in [0, \tau_1),$$

with $\{C_t\}_{t\in[0,\tau_1)}$ is the solution of the following SDE

$$\begin{cases} dC_t = C_t\Big[\tilde{h}\hat{X}_t dZ_t - (\tilde{\lambda}^2 P_t + \lambda(\hat{X}_t)^2 - 1)dt\Big], \quad x \in \mathbb{R}, \quad t \in [0, \tau_1), \\ C_0 = 1. \end{cases}$$

Notice, for $t \in [0, \tau_1)$, u_t are Gaussian functions, therefore, the desired result is obtained if we can show u is the unique solution of Equation (3.1.5)*. In fact, on the one hand, notice, by Itô formula,*

$$\begin{aligned} dC_t \frac{1}{\sqrt{2\pi P_t}} =& C_t \frac{1}{\sqrt{2\pi P_t}} \Big[\tilde{h}\hat{X}_t dZ_t - (\tilde{\lambda}^2 P_t + \tilde{\lambda}^2(\hat{X}_t)^2 - 1)dt\Big] + C_t \frac{1}{\sqrt{2\pi P_t}}[-\frac{1}{2P_t}]dP_t \\ =& C_t \frac{1}{\sqrt{2\pi P_t}} \Big[\tilde{h}\hat{X}_t dZ_t - (\tilde{\lambda}^2 P_t + \tilde{\lambda}^2(\hat{X}_t)^2 - 1)dt + [-\frac{1}{2P_t}]dP_t\Big], \end{aligned}$$

$$d[-\frac{(x-\hat{X}_t)^2}{2P_t}] = \frac{x-\hat{X}_t}{P_t} d\hat{X}_t - \frac{(\tilde{h})^2 P_t^2}{2P_t} dt + \frac{(x-\hat{X}_t)^2}{2P_t P_t} dP_t,$$

$$\begin{aligned} de^{-\frac{(x-\hat{X}_t)^2}{2P_t}} =& e^{-\frac{(x-\hat{X}_t)^2}{2P_t}} d[-\frac{(x-\hat{X}_t)^2}{2P_t}] + \frac{1}{2} e^{-\frac{(x-\hat{X}_t)^2}{2P_t}} [\frac{x-\hat{X}_t}{P_t}\tilde{h}P_t]^2 dt \\ =& e^{-\frac{(x-\hat{X}_t)^2}{2P_t}} \Big\{\frac{x-\hat{X}_t}{P_t} d\hat{X}_t - \frac{(\tilde{h})^2 P_t^2}{2P_t} dt + \frac{(x-\hat{X}_t)^2}{2P_t P_t} dP_t + \frac{1}{2}[(x-\hat{X}_t)\tilde{h}]^2 dt\Big\}, \end{aligned}$$

then

$$\begin{aligned} du_t =& d\Big[C_t \frac{1}{\sqrt{2\pi P_t}} \cdot e^{-\frac{(x-\hat{X}_t)^2}{2P_t}}\Big] \\ =& (dC_t \frac{1}{\sqrt{2\pi P_t}}) e^{-\frac{(x-\hat{X}_t)^2}{2P_t}} + C_t \frac{1}{\sqrt{2\pi P_t}} de^{-\frac{(x-\hat{X}_t)^2}{2P_t}} + u_t[\tilde{h}\hat{X}_t][-\frac{x-\hat{X}_t}{P_t}]\tilde{b}P_t dt \\ =& u_t\Big\{\tilde{h}\hat{X}_t dZ_t - (\tilde{\lambda}^2 P_t + \tilde{\lambda}^2(\hat{X}_t)^2 - 1)dt + [-\frac{1}{2P_t}]dP_t\} \\ &+ u_t\Big\{\frac{x-\hat{X}_t}{P_t} d\hat{X}_t - \frac{(\tilde{h})^2 P_t^2}{2P_t} dt + \frac{(x-\hat{X}_t)^2}{2P_t P_t} dP_t + \frac{1}{2}[(x-\hat{X}_t)\tilde{h}]^2 dt\Big\} \\ &+ u_t[\tilde{h}\hat{X}_t][\frac{x-\hat{X}_t}{P_t}]\tilde{h}P_t dt \\ =& u_t\Big\{\tilde{h}\hat{X}_t dZ_t - (\tilde{\lambda}^2 P_t + \tilde{\lambda}^2(\hat{X}_t)^2 - 1)dt + [-\frac{1}{2P_t}][\tilde{\sigma}^2 + 2\tilde{b}P_t - \tilde{h}^2 P_t^2 - 2\tilde{\lambda}^2 P_t^2]dt\} \\ &+ u_t\Big\{\frac{x-\hat{X}_t}{P_t}\Big(\tilde{b}\hat{X}_t dt + \tilde{h}P_t\Big[dZ_t - \tilde{h}\hat{X}_t dt\Big] - \Big[2\tilde{\lambda}^2 P_t \hat{X}_t\Big]dt\Big) - \frac{(\tilde{h})^2 P_t^2}{2P_t} dt \\ &+ \frac{(x-\hat{X}_t)^2}{2P_t P_t}[\tilde{\sigma}^2 + 2\tilde{b}P_t - \tilde{h}^2 P_t^2 - 2\tilde{\lambda}^2 P_t^2]dt + \frac{1}{2}[(x-\hat{X}_t)\tilde{h}]^2 dt\Big\} \\ &+ u_t[\tilde{h}\hat{X}_t(x-\hat{X}_t)\tilde{h}]dt. \end{aligned}$$

It can be rewritten as

$$du_t = u_t\Big\{-\tilde{b}+\tilde{b}x\frac{x-\hat{X}_t}{P_t}+\frac{1}{2}(\tilde{\sigma})^2\Big[-\frac{1}{P_t}+[\frac{x-\hat{X}_t}{P_t}]^2\Big]\Big\}dt+u_t\Big\{\tilde{h}xdZ_t-(\tilde{\lambda}^2x^2-1)dt\Big\}. \tag{3.1.6}$$

On the other hand, notice

$$\mathcal{L}^*u_t=-\frac{\partial}{\partial x}(\tilde{b}xu_t)+\frac{1}{2}(\tilde{\sigma})^2\frac{\partial^2}{\partial x^2}u_t=-\tilde{b}u_t-\tilde{b}x\frac{\partial u_t}{\partial x}+\frac{1}{2}(\tilde{\sigma})^2\frac{\partial^2}{\partial x^2}u_t,$$

with

$$\frac{\partial u_t}{\partial x}=-\frac{x-\hat{X}_t}{P_t}u_t,\qquad \frac{\partial^2 u_t}{\partial x^2}=-\frac{1}{P_t}u_t+[\frac{x-\hat{X}_t}{P_t}]^2u_t.$$

Then

$$\mathcal{L}^*u_t=u_t\Big\{-\tilde{b}+\tilde{b}x\frac{x-\hat{X}_t}{P_t}+\frac{1}{2}(\tilde{\sigma})^2\Big[-\frac{1}{P_t}+[\frac{x-\hat{X}_t}{P_t}]^2\Big]\Big\}.$$

We get

$$\begin{aligned}&\mathcal{L}^*u_tdt+\tilde{h}xu_tdZ_t-(\tilde{\lambda}^2x^2-1)u_tdt \\ =&u_t\Big\{-\tilde{b}+\tilde{b}x\frac{x-\hat{X}_t}{P_t}+\frac{1}{2}(\tilde{\sigma})^2\Big[-\frac{1}{P_t}+[\frac{x-\hat{X}_t}{P_t}]^2\Big]\Big\}dt+u_t\Big\{\tilde{h}xdZ_t-(\tilde{\lambda}^2x^2-1)dt\Big\}.\end{aligned} \tag{3.1.7}$$

Combining Equation (3.1.6) *and* (3.1.7), *u satisfies*

$$du_t=\mathcal{L}^*u_tdt+\tilde{h}xu_tdZ_t-(\tilde{\lambda}^2x^2-1)u_tdt.$$

That is $\{u_t\}_{t\in[0,\tau_1)}$ is the solution Equation (3.1.5).

Now we generalize this result to the following model, for $t\in[0,T]$,

$$\begin{cases} dX_t=[\bar{b}(t,Z,Y)X_t+\bar{b}^0(t,Z,Y)]dt+\bar{\sigma}(t,Z,Y)dV_t,\\ dZ_t=[(\bar{h}(t,Z,Y)X_t+\bar{h}^0(t,Z,Y)]dt+dB_t,\\ Y_t \text{ is a doubly stochastic Poisson process}\\ \qquad\qquad \text{with intensity } \bar{\lambda}^2(t,Z,Y)X_t^2+\bar{\lambda}^1(t,Z,Y)X_t+\bar{\lambda}^0(t,Z,Y),\\ \qquad\qquad \text{and jumping times } \tau_1,\ \tau_2,\ \ldots. \end{cases}$$

where $\bar{b}$, $\bar{b}^0$, $\bar{\sigma}$, $\bar{h}$, $\bar{h}^0$, $\bar{\lambda}^2$, $\bar{\lambda}^1$ and $\bar{\lambda}^0$ are $\{\mathcal{F}_t^{Z,Y}\}$-adapted, and again, we assume

$$\bar{\lambda}_t^2>0,\quad \text{and}\quad \bar{\lambda}_t^0-\frac{(\bar{\lambda}_t^1)^2}{4\bar{\lambda}_t^2}\geq 0 \tag{3.1.8}$$

to guarantee the positivity of the intensity.

As above, before the time of first jump of Y, that is , for $0\leq t<\tau_1$, the conditional distribution of X given the past observation of Z and Y is Gaussian with mean $\hat{X}_t$ and variance P_t. The process $\big\{(\hat{X}_t,P_t)\big\}_{t\in[0,\tau_1)}$ is the unique solution to the following equations, for $0\leq t<\tau_1$,

$$\begin{cases} d\hat{X}_t=\Big[\bar{b}(t,Z,Y)\hat{X}_t+\bar{b}^0(t,Z,Y)\Big]dt\\ \qquad +\bar{h}(t,Z,Y)P_t\Big[dZ_t-\Big(\bar{h}(t,Z,Y)\hat{X}_t+\bar{h}^0(t,Z,Y)\Big)dt\Big]\\ \qquad -2\bar{\lambda}^2(t,Z,Y)\hat{X}_tP_tdt-\bar{\lambda}^1(t,Z,Y)P_tdt,\\ \frac{d}{dt}P_t=\bar{\sigma}^2(t,Z,Y)+2\bar{b}(t,Z,Y)P_t-\Big[\bar{h}(t,Z,Y)\Big]^2P_t^2-2\Big[\bar{\lambda}^2(t,Z,Y)\Big]^2P_t^2, \end{cases} \tag{3.1.9}$$

with $\hat{X}_0 = \mathbb{E}(X_0)$ and $P_0 = \mathbb{E}[(X_0 - \hat{X}_0)^2]$.

It can be shown that Equation (3.1.9) still holds between jumps, that is for $\tau_n \le t < \tau_{n+1}$, if the conditional distribution of the state X at time τ_n given the past observation of Z and Y is Gaussian.

Case2

Consider Model (2.2.9) with the jump intensity λ specified as follows

$$\lambda(x) = c_0 + c_1 e^{c_2 x}, \quad x \in \mathbb{R}, \tag{3.1.10}$$

where $c_2 \in \mathbb{R}$ and $c_0, c_1 \in \mathbb{R}^+$. In what follows, we will show that, λ defined by Equation (3.1.10) guarantees of preserving Gaussianity at times of jump.

By Frey and Runggaldier (2010), Theorem 6.4, the update of the conditional distribution at the time of jump is as following, for $t = \tau_1, \tau_2, \ldots$,

$$p_t(x) = \frac{\lambda(x) p_{t-}(x)}{\int_{\mathbb{R}} \lambda(x) p_{t-}(x) dx}, \quad x \in \mathbb{R}.$$

Then, motivated by Fontana and Runggaldier (2010), Proposition 8, we have the following Gaussianity preserving result.

Theorem 3.1. *If $p_{t-} \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$, then p_t is a linear combination of 2 Gaussian distributions $N(\mu, \sigma^2)$ and $N(\mu + \sigma^2 c_2, \sigma^2)$ with corresponding weights w_1 and w_2, where*

$$w_1 := \frac{c_0}{c_0 + c_1 \exp(\frac{(\mu+\sigma^2 c_2)^2 - \mu^2}{2\sigma^2})}, \quad w_2 := \frac{c_1 \exp(\frac{(\mu+\sigma^2 c_2)^2 - \mu^2}{2\sigma^2})}{c_0 + c_1 \exp(\frac{(\mu+\sigma^2 c_2)^2 - \mu^2}{2\sigma^2})}. \tag{3.1.11}$$

3.1.2 A finite dimensional filter

The objective of this section is to introduce a finite dimensional approximation of the nonlinear filtering problem, motivated by Case 1 and Case 2, w.r.t. Model (2.2.9) with default intensity λ specified as some exponential functions.

Consider Model (2.2.9) with the jump intensity λ specified as (3.1.10). With this assumption the Gaussianity is preserved at times of jump by Case 2, Theorem 3.1. Between jumps, notice that Model (2.2.9) is a nonlinear model. In order to apply the result of Case 1, we linearize of the state model and the continuous observation Z, and we take the quadratic approximation of jump intensity. Consequently, the Gaussianity is preserved between jumps. To sum up, after approximation, the conditional distribution is a linear combination of Gaussian functions and consequently can be obtained by finite number of parameters. Finally, we give the algorithm of the filter for Model (2.2.9).

Filtering between jumps

We have shown in Section 3.1.1, for Model (3.1.1) the conditional distribution is Gaussian before the time of the first jump. We now applying this result to numerically solve the nonlinear filtering problem w.r.t. Model (2.2.9). One key point is to approximate Model (2.2.9) to obtain Model (3.1.1), applying Taylor expansion. Assume the prior estimate of x is $\bar{x}$ then we approximate σ, the diffusion coefficient in the state equation, by the 0-th order Taylor expansion,

$$\sigma(x) \approx \sigma(\bar{x}).$$

Next, we linearize b, h using the 1-th order Taylor expansion,

$$\begin{aligned} b(x) \approx & b(\bar{x}) + b'(\bar{x})(x - \bar{x}), \\ h(x) \approx & h(\bar{x}) + h'(\bar{x})(x - \bar{x}). \end{aligned}$$

Finally, we approximate λ by a quadratic function, using the 2-th order Taylor expansion,

$$\begin{aligned} \lambda(x) \approx & \lambda(\bar{x}) + \lambda'(\bar{x})(x - \bar{x}) + \frac{1}{2}\lambda''(\bar{x})(x - \bar{x})^2 \\ = & (c_0 + c_1 e^{c_2 \bar{x}}) + c_1 c_2 e^{c_2 \bar{x}}(x - \bar{x}) + \frac{1}{2} c_1 (c_2)^2 e^{c_2 \bar{x}} (x - \bar{x})^2 \\ > & 0, \end{aligned} \tag{3.1.12}$$

where the equality follows from the Equation (3.1.10), the definition of λ, and the inequality follows from $c_1(c_2)^2 e^{c_2 \bar{x}} > 0$ and

$$(c_0 + c_1 e^{c_2 \bar{x}}) - \frac{[c_1 c_2 e^{c_2 \bar{x}}]^2}{4 \cdot \frac{1}{2} c_1 (c_2)^2 e^{c_2 \bar{x}}} = (c_0 + c_1 e^{c_2 \bar{x}}) - \frac{1}{2} c_1 e^{c_2 \bar{x}} = c_0 + \frac{1}{2} c_1 e^{c_2 \bar{x}} > 0.$$

In order to apply this idea to Model (2.2.9), we should have an prior estimate $\bar{x}$ of the state X first. For example, the prior estimate $\bar{x}$ can be the solution of the ordinary differential equation

$$d\bar{x}_t = b(\bar{x}_t)dt \quad \text{with} \quad \bar{x}_0 = \int_{\mathbb{R}} x p_0(x) dx.$$

Then, by approximation of the nonlinear coefficients with Taylor expansion near $\bar{x}$, we have the following approximation of Model (2.2.9), for $0 \le t < \tau_1$,

$$\begin{cases} dX_t \approx [b'(\bar{x}_t)(X_t - \bar{x}_t) + b(\bar{x}_t)]dt + \sigma(\bar{x}_t)dV_t, \\ dZ_t \approx [h'(\bar{x}_t)(X_t - \bar{x}_t) + h(\bar{x}_t)]dt + dB_t, \\ Y \text{ is approximated by a doubly stochastic Poisson process} \\ \qquad \text{with intensity } \frac{1}{2}\lambda''(\bar{x}_t)(X_t - \bar{x}_t)^2 + \lambda'(\bar{x}_t)(X_t - \bar{x}_t) + \lambda(\bar{x}_t) > 0. \end{cases}$$

The positivity of intensity is guaranteed by Equation (3.1.12). The conditional distribution of X_t is approximated by a normal distribution with mean $\hat{X}_t$ and variance P_t which satisfy, by Equation (3.1.9), for $0 \le t < \tau_1$,

$$\begin{cases} d\hat{X}_t = \Big[b'(\bar{x}_t)\hat{X}_t + b(\bar{x}_t) - b'(\bar{x}_t)\bar{x}\Big]dt \\ \qquad\qquad + h'(\bar{x}_t)P_t\Big[dZ_t - (h'(\bar{x}_t)\hat{X}_t + h(\bar{x}_t) - h'(\bar{x}_t)\bar{x}_t)dt\Big] \\ \qquad\qquad - \lambda''(\bar{x}_t)\hat{X}_t P_t dt - \lambda'(\bar{x}_t)P_t dt + \lambda''(\bar{x}_t)\bar{x}_t P_t dt, \\ \frac{d}{dt}P_t = \sigma^2(\bar{x}_t) + 2b'(\bar{x}_t)P_t - \Big[h'(\bar{x}_t)\Big]^2 P_t^2 - 2\Big[\frac{1}{2}\lambda''(\bar{x}_t)\Big]^2 P_t^2. \end{cases}$$

Similarly, we can take $\hat{X}$ as the prior estimator. The filter between jumps is then the solution of the following system, for $0 \le t < \tau_1$,

$$\begin{cases} d\hat{X}_t = b(\hat{X}_t)dt + h'(\hat{X}_t)P_t\Big[dZ_t - h(\hat{X}_t)dt\Big] - \lambda'(\hat{X}_t)P_t dt, \\ \frac{d}{dt}P_t = \sigma^2(\hat{X}_t) + 2b'(\hat{X}_t)P_t - \Big[h'(\hat{X}_t)\Big]^2 P_t^2 - \frac{1}{2}\Big[\lambda''(\hat{X}_t)\Big]^2 P_t^2. \end{cases} \tag{3.1.13}$$

Filtering at a jump time

Apply Theorem 3.1, λ defined by Equation (3.1.10) guarantees of preserving Gaussianity at the time of jump.

To sum up, after approximation, at time t, the conditional density is a linear combination of Gaussian functions with conditional mean, conditional variance, defined by $\hat{X}_t^{(k)}$, $P_t^{(k)}$ and corresponding weights, defined by $\varrho_t^{(k)}$, which is the coefficient of the linear combination. Here, $k = 1, 2, \ldots$. Hence, the conditional density can be determined, if one can determined $\hat{X}_t^{(k)}$, $P_t^{(k)}$ and $\varrho_t^{(k)}$. Suppose that X_0 is normal distributed with mean μ_0 and variance r_0^2, then, the algorithm of the finite dimensional filter for Model (2.2.9) with jump intensity specifies as (3.1.10) is as follows.

Algorithm

i Set $\tau_0 = 0$, $\hat{X}_0^{(1)} = \mu_0$, $P_0^{(1)} = r_0^2$ and $\varrho_0^{(1)} = 1$. For $t = 0$, the filtering distribution is a Gaussian distribution $N(\hat{X}_0^{(1)}, P_0^{(1)})$.

ii For $n = 1, 2, \ldots$,

(a) For $t \in [\tau_{n-1}, \tau_n)$, the filtering distribution is linear combinations of 2^{n-1} Gaussian distributions $N(\hat{X}_t^{(k)}, P_t^{(k)})$ with corresponding weights $\varrho_t^{(k)}$, $k = 1, 2, \ldots, 2^{n-1}$. Here $\hat{X}_t^{(k)}, P_t^{(k)}$ are solutions of Equation (3.1.13) with starting points $\hat{X}_{\tau_{n-1}}^{(k)}, P_{\tau_{n-1}}^{(k)}$. And $\varrho_t^{(k)} = \varrho_{\tau_{n-1}}^{(k)}$.

(b) For $t = \tau_n$, the filtering distribution is linear combination of 2^n Gaussian distributions $N(\hat{X}_t^{(k)}, P_t^{(k)})$ with corresponding weights $\varrho_{\tau_n}^{(k)}$, $k = 1, 2, \ldots, 2^n$.
Here, recalling w_1 and w_2 are defined by Equation (3.1.11), for $k = 1, 2, \ldots, 2^{n-1}$,

$$\hat{X}_{\tau_n}^{(k)} = \hat{X}_{\tau_n-}^{(k)}, \ P_{\tau_n}^{(k)} = P_{\tau_n-}^{(k)}, \ \varrho_{\tau_n}^{(k)} = \varrho_{\tau_n-}^{(k)} w_1.$$

For $k = 2^{n-1} + 1, 2^{n-1} + 2, \ldots, 2^n$,

$$\hat{X}_{\tau_n}^{(k)} = \hat{X}_{\tau_n-}^{(k-2^{n-1})} + c_2 P_{\tau_n-}^{(k-2^{n-1})}, \ P_{\tau_n}^{(k)} = P_{\tau_n-}^{(k-2^{n-1})}, \ \varrho_{\tau_n}^{(k)} = \varrho_{\tau_n-}^{(k-2^{n-1})} w_2.$$

3.2 The finite-state Markov chain approximation

This section is devoted to the study of the nonlinear filtering problem w.r.t. Model (2.2.9) by finite-state Markov chain approximation.

The idea, in the finite-state Markov chain approximation, is to replace the state process X by a simpler process which approximates X well and is such, that, the corresponding expectations are easier to compute. As seen, in Section 2.1.3, the filter corresponding to a finite-state Markov chain is finite dimensional. Consider a sequence of finite-state Markov chain X^n, with state space $S^n = \{s_1^n, s_2^n, \ldots, s_n^n\}$ and generator matrix $Q^n = (\kappa_{ij}^n)_{n \times n}$, see for instance Frey and Runggaldier (2010), Frey and Schmidt (2009). Here, we assume S^n is a fixed equidistance grid. In practice, however, one uses the information from the filtering results to dynamically move the grid in a suitable manner, see for instance, Cai, Gland, and Zhang (1995). We refer to Dupuis and Kushner (2001) for details on how to construct the approximating Markov chain. Define the conditional probabilities of X^n by $\mathbf{p}_t^n = (p_t^{n1}, p_t^{n2}, \ldots, p_t^{nn})^\top$, where p_t^{ni} is the conditional probability of X_t^n in state s_i^n provided past observations of Z and Y. Then, $\mathbf{p}^n$ is the solution of a n-dimensional stochastic ODE and can be actually be computed recursively.

In Section 3.2.1, we approximate the state process X by a finite state Markov chain X^n. Then, in Section 3.2.2, we compute the corresponding filter, w.r.t. approximated state process X^n, which is the solution of a finite dimensional ordinary equation. Finally, in Section 3.2.3, we introduce the computation strategy for the corresponding ordinary equation.

3.2.1 Approximating Markov chain

The objective of this section, is to approximate the state process X, which is a diffusion process given by Equation (2.2.1), by a finite state Markov chain X^n. The key point is the finite dimensional approximation of $\mathcal{L}$ which is the generator of X. $\mathcal{L}$ is a second order differential operator defined by Equation (2.2.4). Examples for its finite dimensional approximation can be seen in Dupuis and Kushner (2001) and Frey and Schmidt (2009). Finally, we show that functioned by $\mathcal{L}$ is approximated by multiplication, by a matrix.

We first fix an equidistant grid $S^n = \{s_1^n, s_2^n, \ldots, s_n^n\}$ with distance $h^n := s_i^n - s_{i-1}^n$. Then, we construct a matrix $Q^n = (\kappa_{ij}^n)_{n\times n} \in \mathbb{R}^{n\times n}$ by approximation the derivatives in $\mathcal{L}$. We will show that

$$\kappa_{ij}^n \geq 0,\ i \neq j, \tag{3.2.1}$$

$$\kappa_{ii}^n \leq 0,\ i = 1, 2, \ldots, n, \tag{3.2.2}$$

$$\sum_{j=1}^{n} \kappa_{ij}^n = 0,\ i = 1, 2, \ldots, n, \tag{3.2.3}$$

which are the sufficient properties for the generator matrix of a Markov chain, see Bain and Crisan (2009), Exercise 3.6, page 51. Therefore, Q^n can be view as the generator matrix of a finite-state Markov chain X^n with state space S^n. Finally, X is approximated by the finite-state Markov chains X^n which has generator matrix Q^n.

So, the key step is to obtain the generator matrix Q^n. Here, Q^n is determined by finite difference method to approximate the partial differential operator $\mathcal{L}$, see for instance Dupuis and Kushner (2001) and Cai, Gland, and Zhang (1995). The finite difference method is a numerical method, for differential equation, by approximation of derivatives with finite differences. To be precise, Q^n are obtained as follows.

Using the finite difference method, the first order derivatives are approximated as, $\forall x \in \mathbb{R}$, for a twice differential function $f : \mathbb{R} \to \mathbb{R}$,

$$\frac{d}{dx} f(x) \approx \begin{cases} \frac{f(x+h^n)-f(x)}{h^n}, & \text{if } b(x) \geq 0, \\ \frac{f(x)-f(x-h^n)}{h^n}, & \text{else.} \end{cases} \tag{3.2.4}$$

Here, difference approximations for $f'(x)$ are in order to to guarantee Equation (3.2.1), (3.2.2) and (3.2.3). The second order derivatives are approximated as, $\forall x \in \mathbb{R}$,

$$\frac{d^2}{dx^2} f(x) \approx \frac{f(x+h^n) - 2f(x) + f(x-h^n)}{(h^n)^2}. \tag{3.2.5}$$

With (3.2.4) and (3.2.5), we can give the approximation of $\mathcal{L}$, which is defined by Equation by (2.2.4), for $i = 2, 3, \ldots, n-1$,

$$\begin{aligned}
\mathcal{L}f(s_i) =& \frac{\sigma^2(s_i)}{2}\frac{d^2}{dx^2}f(s_i) + b(s_i)\frac{d}{dx}f(s_i) \\
\approx& \frac{\sigma^2(s_i)}{2}\frac{f(s_i+h^n) - 2f(s_i) + f(s_i - h^n)}{(h^n)^2} \\
&+ f(s_i)(-\frac{|b(s_i)|}{h^n}) + f(s_i+h)\frac{|b(s_i)|}{h^n}\mathbf{1}_{b(s_i)>0} + f(s_i - h)\frac{|b(s_i)|}{h^n}\mathbf{1}_{b(s_i)<0} \\
=& \frac{\sigma^2(s_i)}{2}\frac{f(s_{i+1}) - 2f(s_i) + f(s_{i-1})}{(h^n)^2} \\
&+ f(s_i)(-\frac{|b(s_i)|}{h^n}) + f(s_{i+1})\frac{|b(s_i)|}{h^n}\mathbf{1}_{b(s_i)>0} + f(s_{i-1})\frac{|b(s_i)|}{h^n}\mathbf{1}_{b(s_i)<0} \\
=& f(s_i)\Big[-\frac{\sigma^2(s_i)}{(h^n)^2} - \frac{|b(s_i)|}{h^n}\Big] + f(s_{i+1})\Big[\frac{\sigma^2(s_i)}{2(h^n)^2} + \frac{|b(s_i)|}{h^n}\mathbf{1}_{b(s_i)>0}\Big] \\
&+ f(s_{i-1})\Big[\frac{\sigma^2(s_i)}{2(h^n)^2} + \frac{|b(s_i)|}{h^n}\mathbf{1}_{b(s_i)<0}\Big].
\end{aligned}$$

Consequently, Q^n is defined as below, denoting, for $i = 2, 3, \ldots, n-1$,

$$\begin{aligned}
\kappa^n_{ii} &:= -\frac{\sigma^2(s_i)}{(h^n)^2} - \frac{|b(s_i)|}{h^n}, \\
\kappa^n_{i,i+1} &:= \frac{\sigma^2(s_i)}{2(h^n)^2} + \frac{|b(s_i)|}{h^n}\mathbf{1}_{b(s_i)>0}, \\
\kappa^n_{i,i-1} &:= \frac{\sigma^2(s_i)}{2(h^n)^2} + \frac{|b(s_i)|}{h^n}\mathbf{1}_{b(s_i)<0}, \\
\kappa^n_{ij} &:= 0,\ j \neq i-1, i, i+1.
\end{aligned}$$

Secondly, κ^n_{ij}, $i = 1, n$, $j = 1, 2, \ldots, n$, are defined according to Equation (3.2.1), (3.2.2), (3.2.3) and other conditions. Finally, the other coefficients of Q^n are defined as 0.

With the definition of Q^n, we see it satisfies Equation (3.2.1), (3.2.2) and (3.2.3). Now, we have the following approximation,

$$\mathcal{L}f(s_i) \approx (Q^n\mathbf{f})_i,$$

with $\mathbf{f} := \Big(f(s^n_1), f(s^n_2), \ldots, f(s^n_n)\Big)^\top$. Using Jacod and Shiryaev (2003), Thoerem 4.21, page 558, if can be shown that X^n obtained as above weakly convergence to X, as $n \to \infty$. The sufficient conditions for weak convergence of X^n to X are also discussed by Frey and Runggaldier (2010).

3.2.2 Filter

The objective of this section is to derive the corresponding filter of X^n which is the finite state Markov chain approximation for X. Define the corresponding stochastic process Λ^n by, for $t \in [0, T]$,

$$\Lambda^n_t := \Big\{\prod_{\tau_n \le t} \lambda(X^n_{\tau_n-})\Big\} \exp\Big(\int_0^t [h(X^n_s)]^\top dB_s + \frac{1}{2}\int_0^t \|h(X^n_s)\|^2 ds - \int_0^t (\lambda(X^n_s) - 1)ds\Big). \quad (3.2.6)$$

Define the new measure $\mathbb{P}^{0,n}$ on the measureable space $(\Omega, \mathcal{F})$ by

$$\mathbb{P}^{0,n}(A) = \int_A (\Lambda_T^n)^{-1}(\omega)\mathbb{P}(d\omega)$$

for all $A \in \mathcal{F}$. Denote by $\mathbb{E}^{0,n}$ the expectation w.r.t. $\mathbb{P}^{0,n}$.

Then, in this case the filter can be represented by the n-dimensional process $\mathbf{q}_t^n = (q_t^{n1}, q_t^{n2}, \ldots, q_t^{nn})^\top$, $t \in [0,T]$, with $q_t^{ni} := \mathbb{E}^{0,n}(\Lambda_t^n \mathbf{1}_{\{X_t^n = s_i^n\}} | \mathcal{F}_t^{Z,Y})$, for $i = 1, 2, \ldots, n$. Consequently, the conditional expectation can be approximated as follows

$$\mathbb{E}^{0,n}\Big[f(X_t)\Lambda_t \Big| \mathcal{F}_t^{Z,Y}\Big] \approx \mathbb{E}^{0,n}\Big[f(X_t^n)\Lambda_t^n \Big| \mathcal{F}_t^{Z,Y}\Big] = \sum_{i=1}^n f(s_i^n) q_t^{ni}, \quad t \in [0,T],$$

recalling that Λ is defined by Equation (2.4.5). And by Kallianpur-Striebel formula (2.4.8),

$$\begin{aligned}\mathbb{E}\Big[f(X_t)\Big|\mathcal{F}_t^{Z,Y}\Big] \approx \mathbb{E}\Big[f(X_t^n)\Big|\mathcal{F}_t^{Z,Y}\Big] &= \frac{\mathbb{E}^{0,n}[f(X_t^n)\Lambda_t^n | \mathcal{F}_t^{Z,Y}]}{\mathbb{E}^{0,n}[\Lambda_t^n | \mathcal{F}_t^{Z,Y}]} = \frac{\sum_{i=1}^n f(s_i^n) q_t^{ni}}{\sum_{j=1}^n q_t^{nj}} \\ &= \sum_{i=1}^n f(s_i^n) \cdot \frac{q_t^{ni}}{\sum_{j=1}^n q_t^{nj}} = \sum_{i=1}^n f(s_i^n) p_t^{ni},\end{aligned}$$

where, by normalizing,

$$p_t^{ni} = \frac{q_t^{ni}}{\sum_{j=1}^n q_t^{nj}}, \; i = 1, 2, \ldots, n, \quad t \in [0,T].$$

Therefore the conditional probabilities are now given by

$$\Big\{(s_1^n, p_t^{n1}), (s_2^n, p_t^{n2}), \ldots, (s_n^n, p_t^{nn})\Big\}, \quad t \in [0,T],$$

with position s_i^n and corresponding conditional probabilities p_t^{ni}.

It can be shown that, similar to Equation (2.1.7) and (2.1.8), the recursive representation of $\mathbf{q}^n$ is as follows, for $i = 1, 2, \ldots, n$, $t \in [0,T]$,

$$q_t^{ni} = q_0^{ni} + \int_0^t \Big(\sum_{j=1}^n \kappa_{ji}^n q_s^{nj}\Big) ds + \int_0^t h(s_i^n) q_s^{ni} dZ_s + \int_0^t \Big[\lambda(s_i^n) - 1\Big] q_{s-}^{ni} d(Y_s - s).$$

And the matrix representation is

$$\mathbf{q}_t^n = \mathbf{q}_0^n + \int_0^t (Q^n)^\top \mathbf{q}_s^n ds + \int_0^t B^h \mathbf{q}_s^n dZ_s + \int_0^t B^\lambda \mathbf{q}_{s-}^n d(Y_s - s), \quad t \in [0,T], \tag{3.2.7}$$

where $B^h := \operatorname{diag}\Big(h(s_1^n), h(s_2^n), \ldots, h(s_n^n)\Big)$ and $B^\lambda := \operatorname{diag}\Big(\lambda(s_1^n) - 1, \ldots, \lambda(s_n^n) - 1\Big)$. Notice, this filter is much faster than the Galerkin filter, since the coefficient matrices B^h and B^λ are diagonal. The Galerkin filter is further introduced in Chapter 7.

The convergence was discussed by, for instance, Frey and Runggaldier (2010) and Frey and Schmidt (2009). Frey and Runggaldier (2010) show, under suitable assumptions, the approximating filters convergence in probability. Frey and Schmidt (2009) show the weak convergence of the filter, that is, for all bounded and continuous function, the conditional expectation of $f(X_t^n)$ given past observation converges to the conditional expectation of $f(X_t)$ given past observation, as $n \to \infty$.

3.2.3 Numerical solution

This section introduces the numerical method for Equation (3.2.7).

Equation (3.2.7) can be solved numerically with splitting-up method which we will introduce in Section 7.1.2 in detail. Let $0 = t_0 < t_1 < \cdots < t_k < \cdots < t_L = T$ be a uniform partition of the interval $[0,T]$ with time step $\Delta = t_k - t_{k-1} = \frac{T}{L}$. Assume that $\{Z_{t_k}\}$, $\{Y_{t_k}\}$, $k = 0,1,\ldots,L$ is the sampled trajectories of the observation processes Z and Y at discrete times. $\mathbf{q}_k^{(n,\Delta)} = \left(q_{k,1}^{(n,\Delta)}, q_{k,2}^{(n,\Delta)}, \ldots, q_{k,n}^{(n,\Delta)}\right)^\top$, the approximation of the unnormalized conditional distribution $\mathbf{q}_{t_k}^n$ at discrete times $(t_k, k = 0,1,\ldots,L)$ is obtained as follows, first $\mathbf{q}_0^{(n,\Delta)}$ is obtained, by recalling that p_0 is the density of the law of X_0,

$$q_{0,i}^{(n,\Delta)} = \int_{s_i^n - \frac{\Delta}{2}}^{s_i^n + \frac{\Delta}{2}} p_0(x)dx, \quad i = 1,2,\ldots,n.$$

For $k = 1,2,\ldots,L$,

1) $\bar{\mathbf{q}}_k^{(n,\Delta)} = \left(\bar{q}_{k,1}^{(n,\Delta)}, \bar{q}_{k,2}^{(n,\Delta)}, \ldots, \bar{q}_{k,n}^{(n,\Delta)}\right)^\top := \exp\left[(Q^n - B^\lambda)\Delta\right]\mathbf{q}_{k-1}^{(n,\Delta)}$,

2) $\tilde{\mathbf{q}}_k^{(n,\Delta)} = \left(\tilde{q}_{k,1}^{(n,\Delta)}, \tilde{q}_{k,2}^{(n,\Delta)}, \ldots, \tilde{q}_{k,n}^{(n,\Delta)}\right)^\top$, where for $i = 1,\ldots,n$,

$$\tilde{q}_{k,i}^{(n,\Delta)} = \exp\left[h(s_i^n)(Z_{t_k} - Z_{t_{k-1}}) - \frac{h(s_i^n)^2\Delta}{2}\right]\bar{q}_{k,i}^{(n,\Delta)},$$

3) $\mathbf{q}_k^{(n,\Delta)} = \left(q_{k,1}^{(n,\Delta)}, q_{k,2}^{(n,\Delta)}, \ldots, q_{k,n}^{(n,\Delta)}\right)^\top$, where for $i = 1,\ldots,n$,

$$q_{k,i}^{(n,\Delta)} = \lambda(s_i^n)^{(Y_{t_k} - Y_{t_{k-1}})}\tilde{q}_{k,i}^{(n,\Delta)}.$$

In the computation, the most complicated part is the matrix exponential in step 1). In general, the computation of the matrix exponential is difficult if n is large. But notice, it does not depend on the observations, it only depends on the model. Therefore $\exp\left[(Q^n - B^\lambda)\Delta\right]$ can be computed before hand. This is a computational advantage of finite-state Markov chain approximation.

3.3 Particle methods

As same as the Markov chain approximation method and Euler-Maruyama, particles methods are finite dimensional approximation of unnormalized measure ρ by discretisation of the state variable. For $n \in \mathbb{N}$, The approximation of the conditional measure, for $0 \le t \le T$, is given by the discrete probabilities $\left\{(x_t^1, p_t^1), (x_t^2, p_t^2) \ldots, (x_t^n, p_t^n)\right\}$ with position x_t^i and corresponding conditional probabilities p_t^i, $i = 1,2,\ldots,n$. Here n represents the number of particles that are used to approximate of the measure. But particle methods need not to fix a grid of state and are very flexible.

Again, this section is devoted to derive the particle filter for Model (2.2.9). The basic idea of particle filters is to approximate conditional expectations $\mathbb{E}^0\left[f(X_t)\Lambda_t \middle| \mathcal{F}_t^{Z,Y}\right]$ by Monte Carlo methods, where Λ is defined by Equation (2.4.5). By Proposition 2.5, under the new measure $\mathbb{P}^0$, the state process X, the continuous observation Z and the jump observation Y are independent and X has the same law as its law under $\mathbb{P}$. Now we can apply Monte Carlo methods. And the basic steps of the algorithm are as follows: Assume that trajectories of observations Z and Y are given.

Algorithm 3.1. *i First, according to SDE* (2.2.1), *independently generate n paths of X denoted by x_t^i, $i = 1 \ldots, n$, $0 \le t \le T$.*

ii Obtain corresponding paths of Λ, denoted by a_t^i. a_t^i is generated by the following equation, for $i = 1, 2, \ldots n$,

$$a_t^i = 1 + \int_0^t a_s^i h(x_s^i)^\top dZ_s + \int_0^t a_{s-}^i \Big[\lambda(x_s^i) - 1\Big] d(Y_s - s), \quad t \in [0, T].$$

iii Then, the conditional expectation can be approximated as a weighted average

$$\mathbb{E}^0\Big[f(X_t)\Lambda_t \Big| \mathcal{F}_t^{Z,Y}\Big] \approx \frac{1}{n}\sum_{i=1}^n f(x_t^i) a_t^i, \quad t \in [0, T]. \tag{3.3.1}$$

And consequently, apply Kallianpur-Striebel formula (2.4.8),

$$\begin{aligned} \mathbb{E}\Big[f(X_t)\Big|\mathcal{F}_t^{Z,Y}\Big] &= \frac{\mathbb{E}^0[f(X_t)\Lambda_t|\mathcal{F}_t^{Z,Y}]}{\mathbb{E}^0[\Lambda_t|\mathcal{F}_t^{Z,Y}]} \approx \frac{\frac{1}{n}\sum_{i=1}^n f(x_t^i)a_t^i}{\frac{1}{n}\sum_{i=1}^n a_t^i} \\ &= \sum_{i=1}^n f(x_t^i) \cdot \frac{a_t^i}{\sum_{j=1}^n a_t^j} = \sum_{i=1}^n f(x_t^i) p_t^i, \end{aligned}$$

where after normalizing,

$$p_t^i := \frac{a_t^i}{\sum_{j=1}^n a_t^j}, \ i = 1, 2, \ldots, n.$$

Finally, the approximation of the conditional measure is now given as

$$\Big\{(x_t^1, p_t^1), (x_t^2, p_t^2) \ldots, (x_t^n, p_t^n)\Big\}, \quad t \in [0, T],$$

with position x_t^i and corresponding conditional probabilities p_t^i, $i = 1, 2, \ldots, n$.

For the convergence and error of the approximation with only continuous observation see Bain and Crisan (2009), page 209-290. Similarly, it can be shown, the convergence is true for the case with additional point process observations.

Let $\{\rho_t^n\}_{t\in[0,T]}$ be the sequences of measure-valued processes

$$\rho_t^n := \frac{1}{n}\sum_{j=1}^n a_t^j \delta_{x_t^j}, \quad t \in [0, T].$$

Here, δ is the Dirac measure. We then have the following convergence result of ρ^n.

Lemma 3.2. *For any $f \in D(\mathcal{L})$, we get*

$$\mathbb{E}^0\Big[\Big(\rho_t^n(f) - \rho_t(f)\Big)^2\Big|\mathcal{F}_t^{Z,Y}\Big] = \frac{C_f(t)}{n},$$

where

$$C_f := \mathbb{E}^0\Big[\Big(f(X_t)\Lambda_t - \rho_t(f)\Big)^2\Big|\mathcal{F}_t^{Z,Y}\Big].$$

Particle methods are very flexible and easy to implement. It is unlike the Markov chain approximation method for which one need to fix a grid to approximation the distribution. But it suffers from severe degeneracy, especially in high dimensions. After a few steps , the majority of the weights are close to zero and all the weights tend to concentrate on a very few particles. This reduces the effective of Monte Carlo methods.

There are remedies for this. One example is the branching particle filter. At the precise time, each existing particle will die or give birth to a random number of offspring proportional to the weight. The number of particles in the system will remain constant at n. In this way, particles that stay on the right trajectories (representing by heavy weights) are explored more thoroughly while particles with unlikely trajectories/positions (representing by little weights) are not carried forward uselessly. For details and corresponding convergence results, see Bain and Crisan (2009), Budhiraja, Chen, and Lee (2007).

Chapter 4

Linear stochastic PDEs

We are interested in the solution of the Zakai equation, which is a linear parabolic PDE. Therefore, we need some existence and uniqueness results for this class of SPDE. Some solutions are available, including the variational approach, the semigroup approach, and 'method of moving frame'. Using these methods, similar results have been obtained on the same type of equations. The 'method of moving frame' is an approach, based on a time dependent coordinate transform, which reduces a wide class of SPDEs to a class of simpler SDE problems, see for instance Filipovic, Tappe, and Teichmann (2010). The main purpose of this chapter is to present a short review of existence results of linear SPDEs by semigroup approach and variational approach.

To simplify notation, let $\mathcal{B}$ be a real Banach space, and $\mathcal{H}$ be a real separable Hilbert space with norm $\|\cdot\|_{\mathcal{H}}$ and scalar product $(\cdot,\cdot)$.

This chapter is based on Peszat and Zabczyk (2007), Prato and Zabczyk (1992), Hairer (2009) and Pardoux (1979b). We refer also to Gawarecki and Mandrekar (2001), L. Gawarecki (1999), Kallianpur and Xiong (1995), Knoche (2004), Knoche (2005) and reference therein, for other interesting results of SPDEs.

In Section 4.1, we study linear SPDEs by the tool of semigroup theory. In Section 4.2, we introduce the variational approach to linear parabolic SPDEs.

4.1 Semigroup approach

In this section, we give an introduction to semigroup theory of linear SPDEs. The details can be found in the books Peszat and Zabczyk (2007), Prato and Zabczyk (1992), and Hairer (2009).

In Section 4.1.1, we mainly introduce semigroup theory and in Section 4.1.2, we introduce the result of linear SPDEs by semigroup approach.

4.1.1 Semigroup

First, we define semigroups and generators. We then show, if a linear operator is the generator of a semigroup with some regularly properties, then the corresponding linear PDEs have a so-called weak solution obtained by the semigroup. Finally, we show the characterizations for generators of semigroups.

We are interested with the solution of following linear equation in a Banach space $\mathcal{B}$,

$$\begin{cases} u_t' = \mathrm{A}_0 u_t, \ t \in [0,T], \\ u_0 = w \in \mathcal{B}. \end{cases} \tag{4.1.1}$$

Here $u_t' := \lim_{\delta\to 0} \frac{u_{t+\delta}-u_t}{\delta}$, and the limit is taken in the norm of $\mathcal{B}$. A_0 is a linear operator, in general unbounded, with domain $D(\mathrm{A}_0) \subset \mathcal{B}$, that is $\mathrm{A}_0 \in L\Big(D(\mathrm{A}_0);\mathcal{B}\Big)$.

A strong solution of Equation (4.1.1) is a function $u \in C\big([0,T];\mathcal{B}\big) \cap C^1\big((0,T);\mathcal{B}\big)$ such that $u(t) \in D(\mathrm{A}_0)$ and fulfils Equation (4.1.1).

If a unique solution of Equation (4.1.1) exists, define operators $S(t) : D(\mathrm{A}_0) \to \mathcal{B}$ as follows

$$S(t)w = u(t,w), \ \forall w \in D(\mathrm{A}_0), \ t \in [0,T].$$

Then, $S(t)$ maps the starting point $u_0 = w$ onto the solution $u(t)$ at time t. The uniqueness of the solution implies $\{S(t), t \in [0,T]\}$ is family of linear operators, which means $S(t) \in L\big(D(\mathrm{A}_0);\mathcal{B}\big)$. In other word, for $t \in [0,T]$, $w_1, w_2 \in D(\mathrm{A}_0)$

$$S(t)(c_1w_1 + c_2w_2) = c_1S(t)w_1 + c_2S(t)w_2,$$

where c_1 and c_2 are constants. And we have the so-called semigroup properties of semigroup, for $s, t, s+t \in [0,T]$,

$$S(t+s)w = S(t)S(s)w, \ w \in D(\mathrm{A}_0),$$

with $S(0) = I$.

According to the different regularity properties, one has the following definition of semigroup, also see Hairer (2009), Definition 4.1, page 28.

Definition 4.1. A semigroup on $\mathcal{B}$ is a set $\{S(t), t \in [0,T]\}$ of linear bounded operators on $\mathcal{B}$ which satisfy

$$S(t+s) = S(t)S(s), S(0) = I, \ 0 \le s,t \le s+t \le T.$$

A semigroup is furthermore called

- a C_0-semigroup if $S(\cdot)u_0 \in C([0,T];\mathcal{B})$, $\forall u_0 \in \mathcal{B}$.
- an analytic semigroup if there exists $\theta_0 > 0$ such that $\{S(t), t \in [0,T]\}$ have an analytic extension $\{S(t), t \in \mathcal{S}_{\theta_0} \cup \{0\}\}$, where $\mathcal{S}_{\theta_0} := \{z \in \mathbb{C} : |\arg(z) < \theta|\}$ is a sector in $\mathbb{C}$, and the extension satisfies
  - $S(t+s) = S(t)S(s)$, $t, s, t+s \in S(\theta_0)$.
  - $S(e^{i\theta}\cdot)u_0 \in C([0,T];\mathcal{B})$, $\forall u_0 \in \mathcal{B}$, $\forall|\theta| < \theta_0$.

The properties of a semigroup are determined by its generator, which is defined as follows, see Hairer (2009), Definition 4.6, page 29, or Prato and Zabczyk (1992), page 380.

Definition 4.2. The generator A of a C_0-semigroup $S(\cdot)$ is a linear operator defined as follows

$$\mathrm{A}x = \lim_{t\to 0^+} \frac{S(t)x - x}{t}, \ \forall x \in D(\mathrm{A}),$$

where

$$D(\mathrm{A}) := \{x \in \mathcal{B}; \lim_{t\to 0^+} \frac{S(t)x - x}{t} \text{ exists}\},$$

and the limit is taken in the sense of strong convergence.

From the definition, we see that A is an extension of A_0. We recall some definitions from functional analysis, see for instance, Showalter (1977), page 19. Recall that the domain $D(\mathrm{A}^*)$ of the adjoint A^* of an unbounded operator $\mathrm{A} : D(\mathrm{A}) \to \mathcal{B}$ is defined as the set of all elements $\phi \in \mathcal{B}'$ such that there exists an elements $\mathrm{A}^*\phi \in \mathcal{B}'$ with property that $(\mathrm{A}^*\phi)(x) = \phi(\mathrm{A}x)$ for every $x \in D(\mathrm{A})$. Here $\mathcal{B}'$ is the dual space of $\mathcal{B}$.

Let the bilinear form $\langle\cdot,\cdot\rangle : \mathcal{B}' \times \mathcal{B} \to \mathbb{R}$ be defined as

$$\langle \phi, x\rangle = \phi(x), \;\; \phi \in \mathcal{B}', \;\; x \in \mathcal{B},$$

forms a duality pairing. $\langle \phi, x\rangle$ denotes the application of $\phi \in \mathcal{B}'$ to $x \in \mathcal{B}$.

Then, we have the following result:

Theorem 4.1. *Let* A *be the generator of a C_0-semigroup. Then, $\forall x \in D(\mathrm{A})$ and $t \in (0,T]$,*

$$S(t)x \in D(\mathrm{A}) \text{ and } \frac{d}{dt}S(t)x = \mathrm{A}S(t)x = S(t)\mathrm{A}x.$$

Furthermore, for $t \in [0,T]$, $u_t := S(t)x$, $x \in \mathcal{B}$, satisfies, $\forall \phi \in D(\mathrm{A}^*)$, where A^* is the adjoint operator of A,

$$\langle \phi, u_t\rangle = \langle \phi, x\rangle + \int_0^t \langle \mathrm{A}^*\phi, u_s\rangle ds, \; \forall x \in \mathcal{B}. \tag{4.1.2}$$

For the proof, we refer to Hairer (2009), Proposition 4.7, page 30. Theorem 4.1 shows that if A is the generator of a C_0-semigroup $S(\cdot)$, then the function $t \mapsto S(t)x$, $x \in D(\mathrm{A})$, is a solution to the equation

$$u_t' = \mathrm{A}u_t, \; u_0 = x.$$

Equation (4.1.2) means the function $t \mapsto S(t)x$, $x \in \mathcal{B}$, is the solution of the above equation in a weak sense.

Moreover, if A is the generator of an analytic semigroup $S(\cdot)$, then the semigroup map $\mathcal{B}$ into the domain of any arbitrarily high power of A, see, Hairer (2009), Proposition 4.37, page 41.

Consequently, a question arises when A is the generator of a semigroup. For this, we have characterisation of the generators of C_0-semigroup, which is the so-called Hille-Yosida theorem, see Hairer (2009), page 31, and characterisation of the generators analytic semigroup, see Hairer (2009), Theorem 4.22, page 36. Furthermore, we are interested in an important subclass of generators of analytic semigroups which are related to parabolic equations. Notice, the Zakai equation is a parabolic one.

We first give the definition of coercivity assumption.

Definition 4.3. Let $\mathcal{V}$ and $\mathcal{H}$ be Hilbert spaces(with norms $\|\cdot\|_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{V}}$) such that $\mathcal{V}$ is densely embedded in $\mathcal{H}$. A bilinear operator $a : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ is said to satisfy the coercivity assumption, if there exist $c > 0$ and $\beta \geq 0$ such that

$$-a(v,v) \geq c\|v\|_{\mathcal{V}}^2 - \beta\|v\|_{\mathcal{H}}^2, \; \forall v \in \mathcal{V}.$$

The corresponding characterisations are as follows

Theorem 4.2. *Let $\mathcal{V}$ and $\mathcal{H}$ be Hilbert space(with norms $\|\cdot\|_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{V}}$) such that $\mathcal{V}$ is densely embedded in $\mathcal{H}$. Let a bilinear operator $a : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ satisfies the coercivity assumption with the corresponding parameter $c > 0$ and $\beta \geq 0$. Define*

$$D(\mathrm{A}) = \Big\{ u \in \mathcal{V} : |a(u, v)| \leq K_u \|v\|_{\mathcal{H}},\ v \in \mathcal{V} \Big\},$$

where K_u depends on u, and let $\mathrm{A} \in L(D(\mathrm{A}); \mathcal{H})$ be given by

$$a(u, v) = (\mathrm{A}u, v),\ u \in D(\mathrm{A}),\ v \in \mathcal{V}.$$

Then A generates an analytic semigroup $S(\cdot)$ on $\mathcal{H}$ such that $\|S(t)\| \leq e^{\beta t}$, $t \in [0, T]$.

For the proof, the reader is referred to Prato and Zabczyk (1992), Proposition A.10, page 389.

Remark 4.1. *In Theorem 4.2, the definition of $\mathrm{A} \in L(D(\mathrm{A}); \mathcal{H})$ makes sense. More precisely, $\forall u \in D(\mathrm{A})$, define a mapping $F_u : \mathcal{V} \to \mathbb{R}$, by*

$$F_u(v) := a(u, v), \quad \forall v \in \mathcal{V}.$$

Since $\mathcal{V}$ is densely embedded in $\mathcal{H}$, F_u can continuously extended to $\mathcal{H}$. Then by the definition of $D(\mathrm{A})$, F_u is a bounded linear operator on $\mathcal{H}$. Riesz representation theorem implies, there exists an element $w \in \mathcal{H}$, such that, for all $v \in \mathcal{H}$,

$$F_u(v) = (w, v).$$

Now define $\mathrm{A} \in L(D(\mathrm{A}); \mathcal{H})$ by, for all $u \in D(\mathrm{A})$,

$$\mathrm{A}u := w.$$

Then we get

$$a(u, v) = F_u(v) = (w, v) = (\mathrm{A}u, v).$$

4.1.2 Linear SPEDs

In this section we introduce the result of linear SPDEs driven by Lévy noise by semigroup approach. First, we give the definition of strong, weak, mild solutions of linear SPDEs. We then show, with proper assumptions, weak solution and mild solutions coincide. Therefore, it is suffice to study mild solutions which are easier to treat. At last, we show the uniqueness and existence results of mild solutions.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq T}, \mathbb{P},)$ be a stochastic basis with the usual assumption. In this section,we shall consider a linear stochastic PDE of the following form, let $\mathcal{H}$ be a real separable Hilbert space,

$$u_t = u_0 + \int_0^t \mathrm{A}u_s ds + \int_0^t Bu_s dM_s, \text{ with } u_0 = w \in \mathcal{H}, \tag{4.1.3}$$

where A, with domain $D(\mathrm{A})$, is the generator of a C_0-semigroup $S(\cdot)$ on a Hilbert space $\mathcal{H}$. For $u \in \mathcal{H}$, $Bu := (B_1 u, B_2 u, \dots, B_d u)$, where, for $i = 1, 2, \dots, d$, $B_i : \mathcal{H} \to \mathcal{H}$ are bounded linear operators. Since we are interested in the solution of the Zakai equation, we specify M as $M = (M^1, M^2, \ldots, M^d)^\top$ is a $\mathbb{R}^d$-valude square integrable martingale. It satisfies that each M^i is a one-dimensional Brownian motion or a one-dimensional compensated Poisson process with intensity 1. And, if $i \neq j$, then the processes M^i and M^j are independent. For more general assumption of M, see Peszat and Zabczyk (2007), page 122.

A logical question is what we mean by the solution to Equation (4.1.3). In the following, we define the strong, weak and mild solutions w.r.t. Equation (4.1.3).

Definition 4.4. An $\mathcal{H}$-valued $\{\mathcal{F}_t\}$-predictable process u is said to be a strong solution to Equation (4.1.3) if u takes values in $D(\mathrm{A})$, $\sup_{s\in[0,T]} \mathbb{E}\{\|u_s\|^2_{\mathcal{H}} + \|\mathrm{A}u_s\|_{\mathcal{H}}\} < \infty$ and for $t \in [0, T]$,

$$u_t = u_0 + \int_0^t \mathrm{A}u_s ds + \int_0^t Bu_s dM_s, \text{ with } u_0 = w \in \mathcal{H}$$

holds almost surely.

Here $\int_0^t Bu_s dM_s$ is a stochastic integral, where the integrand Bu_s take values in Hilbert space $\mathcal{H}$. Since B is a bounded linear operator and $\sup_{s\in[0,T]} \mathbb{E}\{\|u_s\|^2_{\mathcal{H}} + \|\mathrm{A}u_s\|_{\mathcal{H}}\} < \infty$, this stochastic integral is well-defined.

In general, this solution concept is too restrictive. Solutions are usually defined in the mild or weak sense.

Definition 4.5. An $\mathcal{H}$-valued $\{\mathcal{F}_t\}$-predictable process u is said to be a weak solution, if $\sup_{s\in[0,T]} \mathbb{E}\|u_s\|^2_{\mathcal{H}} < \infty$ and, for $t \in [0, T]$,

$$(v, u_t) = (v, u_0) + \int_0^t (\mathrm{A}^* v, u_s) ds + \int_0^t (B^* v, u_s) dM_s, \text{ with } u_0 = w \in \mathcal{H}$$

holds almost surely for every $v \in D(\mathrm{A}^*)$.

Definition 4.6. An $\mathcal{H}$-valued $\{\mathcal{F}_t\}$-predictable process u is said to be a mild solution if $\sup_{s\in[0,T]} \mathbb{E}\|u_s\|^2_{\mathcal{H}} < \infty$ and for $t \in [0, T]$,

$$u_t = S(t)u_0 + \int_0^t S(t-s)B(u_s) dM_s, \text{ with } u_0 = w \in \mathcal{H}.$$

Here $S(t-s)$ operator on $B(u_s)$ is obtained by applying $S(t-s)$ componentwise.

Theorem 4.3. *u is a mild solution if and only if u is a weak solution.*

Proof. Since A is the generator a C_0-semigroup and B is a bounded linear operator which satisfies the so-called Lipschitz-type conditions, see Peszat and Zabczyk (2007), page 142, the desired result is obtained by Theorem9.15, Peszat and Zabczyk (2007), page 151. □

Theorem 4.3 shows that weak and mild solutions coincide. Therefore it is suffice to study the existence property of mild solution which is easier to treat. And we have the following existence result for mild solutions.

Theorem 4.4. *Assume that w is an $\mathcal{F}_0$-measurable, $\mathcal{H}$-valued, square integrable random variable, then Equation* (4.1.3) *has a unique mild solution and the solution has a càdlàg modification.*

Proof. Theorem 9.29, Peszat and Zabczyk (2007), page 164, implies the desired result. ☐

In the classical theory of stochastic differential equation, one look for càdlàg solution of form

$$du_t = \mathrm{A}u_t dt + Bu_{t-}dM_t,\ u_0 = w.$$

The following theorem shows that the càdlàg solution and the predictable solution are equivalent. For more details, see Peszat and Zabczyk (2007), page 145-148.

Theorem 4.5. *Let u be a càdlàg solution to the equation*

$$du_t = \mathrm{A}u_t dt + Bu_{t-}dM_t,\ u_0 = w.$$

Then $\bar{u}_t := u_{t-}$, $t \geq 0$, is equivalent to u and is a predictable solution to

$$du_t = \mathrm{A}u_t dt + Bu_t dM_t,\ u_0 = w.$$

Proof. The desired result is obtained from Proposition 9.10, Peszat and Zabczyk (2007), page 148. ☐

4.2 The variational approach to linear parabolic SPDEs

The variational approach was introduced by Lions (1961) to solve deterministic PDEs, and it was developed by Pardoux for SPDEs.

In Section 4.2.1, we present the general setting in which we will give the main existence and uniqueness result for linear parabolic SPDE's using variational approach. In Section 4.2.2, we present the main existence result and uniqueness result of SPDEs with Gaussian noise.

4.2.1 General setting

As we introduced before, A is the differential operator associated to the state process X. It is a linear operator, possibly unbounded. For example, define the operator $\mathcal{K} : C^1(\mathbb{R}) \to C(\mathbb{R})$ which acts by taking the derivative. So, for $f \in C^1(\mathbb{R})$, $\mathcal{K}f = f'$. Take $f = x^n$, $x \in [0, 1]$. Then, $\mathcal{K}f = f' = nx^{n-1}$, and $\|f'\|_\infty = n \to \infty$, as $n \to \infty$. So this operator is not bounded.

From now on, $\bar{\mathrm{A}}$ will denote an extension of the unbounded operator A from the previous section. That is, instead of considering

$$\mathrm{A} : D(\mathrm{A}) \to L^\infty(\mathbb{R}^d),$$

we shall consider

$$\bar{\mathrm{A}} : \mathcal{V} \to \mathcal{V}',$$

where

$$D(\mathrm{A}) \subset \mathcal{V} \subset \mathcal{H} \subset \mathcal{V}' \text{ and } \bar{\mathrm{A}}|_{D(\mathrm{A})} = \mathrm{A}.$$

More precisely, the framework is as follows. Before that, we have the following definition.

- For any separable Hilbert space $\mathcal{V}$, denote by $\|\cdot\|_{\mathcal{V}}$ the norm in $\mathcal{V}$ and by $(\cdot,\cdot)_{\mathcal{V}}$ or simply $(\cdot,\cdot)$ its scalar product.

- $\mathcal{V}$ together with its dual $\mathcal{V}'$ and the bilinear form $\langle\cdot,\cdot\rangle : \mathcal{V}' \times \mathcal{V} \to \mathbb{R}$ defined as

$$\langle u, v\rangle = u(v), \quad u \in \mathcal{V}', \quad v \in \mathcal{V}$$

  forms a duality pairing. $\langle u, v\rangle$ denotes the application of $u \in \mathcal{V}'$ to $v \in \mathcal{V}$.

- We denote by $\|\cdot\|_{\mathcal{V}'}$ the norm in $\mathcal{V}'$, defined by

$$\|u\|_{\mathcal{V}'} = \sup_{v\in\mathcal{V}, \|v\|_{\mathcal{V}}\le 1} \langle u, v\rangle.$$

We consider a triple $(\mathcal{V}, \mathcal{H}, \mathcal{V}')$ of Hilbert space which satisfies the following assumption.

Assumption 4.1. *$\mathcal{V} \subset \mathcal{H}$ is a Hilbert space, which is dense in $\mathcal{H}$, with continuous injection. We identify $\mathcal{H}$ with its dual $\mathcal{H}'$, and consider $\mathcal{H}'$ as a subspace of the dual $\mathcal{V}'$ of $\mathcal{V}$, again with continuous injection. Moreover, we have*

$$\begin{aligned}
&\langle u, v\rangle = (u, v), \ \forall v \in \mathcal{V}, \ \forall u \in \mathcal{H} \subset \mathcal{V}', \\
&\|u\|_{\mathcal{H}} \le \|u\|_{\mathcal{V}}, \ \ \forall u \in \mathcal{V}.
\end{aligned}$$

For a triple $(\mathcal{V}, \mathcal{H}, \mathcal{V}')$ which satisfies Assumption 4.1, we have

$$\mathcal{V} \subset \mathcal{H} \simeq \mathcal{H}' \subset \mathcal{V}',$$

and we have the following property:

Lemma 4.6. *Given a triple $(\mathcal{V}, \mathcal{H}, \mathcal{V}')$ which satisfies Assumption 4.1, for $u \in V$,*

$$\|u\|_{\mathcal{V}'} \le \|u\|_{\mathcal{H}} \le \|u\|_{\mathcal{V}}.$$

Proof. By definition of $\|\cdot\|$ and Assumption 4.1,

$$\begin{aligned}
\|u\|_{\mathcal{V}'} &= \sup_{v\in\mathcal{V}, \|v\|_{\mathcal{V}}\le 1} \langle u, v\rangle = \sup_{v\in\mathcal{V}, \|v\|_{\mathcal{V}}\le 1} (u, v) \\
&\le \sup_{v\in\mathcal{V}, \|v\|_{\mathcal{V}}\le 1} \|u\|_{\mathcal{H}} \cdot \|v\|_{\mathcal{H}} \le \sup_{v\in\mathcal{V}, \|v\|_{\mathcal{V}}\le 1} \|u\|_{\mathcal{H}} \cdot \|v\|_{\mathcal{V}} \le \|u\|_{\mathcal{H}}.
\end{aligned}$$

□

Example 4.1. *The Sobolev spaces $\left(H^1(\mathbb{R}^d), L^2(\mathbb{R}^d), H^{-1}(\mathbb{R}^d)\right)$ form a triple which satisfies Assumption 4.1. And we have*

$$H^1(\mathbb{R}^d) \subset L^2(\mathbb{R}^d) \subset H^{-1}(\mathbb{R}^d).$$

More generally, the Sobolev spaces $\left(H^{k+j}(\mathbb{R}^d), H^k(\mathbb{R}^d), H^{k-j}(\mathbb{R}^d)\right)$, $k, j \in \mathbb{N}$ and $j > 0$ form a such triple. For an introduction of Sobolev space, see Section A.1.1.

4.2.2 Basic results for linear parabolic SPDEs with Gaussian noise

The objective of the section is to present the existence and uniqueness results for SPDEs with Gaussian noise. In Chapter 5, we will show the existence and uniqueness results for SPDEs with Gaussian and Poisson noise.

Existence and uniqueness can be established when the coercivity assumption is satisfied. The coercivity assumption of $\bar{\mathrm{A}}$ is crucial. This is explained by Pardoux (1979b), page 144 and page 164. Now we give the definition of the coercivity assumption of an operator based on Definition 4.3.

Definition 4.7. Given a triple $(\mathcal{V},\mathcal{H},\mathcal{V}')$ which satisfies Assumption 4.1, an operator $R \in L(\mathcal{V},\mathcal{V}')$, is said to satisfy the coercivity assumption, if the bilinear operator $a : \mathcal{V}\times\mathcal{V} \to \mathbb{R}$ defined by

$$a(u,v) := \langle Ru, v\rangle$$

satisfies the coercivity assumption. That is there exists $c > 0$ and $\beta \geq 0$ such that, $\forall v \in \mathcal{V}$,

$$\beta\|v\|^2_{\mathcal{H}} - \langle Rv, v\rangle \geq c\|v\|^2_{\mathcal{V}}. \tag{4.2.1}$$

Let $\mathcal{M}^2(0,T;\mathcal{V})$ denote the space of $\mathcal{V}$-valued stochastic process with some regular properties. And let $\mathcal{C}^2(0,T;\mathcal{H})$ denote the spaces of $\mathcal{H}$-valued continuous stochastic processes with some regular properties. For details of processes spaces $\mathcal{M}^2(0,T;\mathcal{V})$ and $\mathcal{C}^2(0,T;\mathcal{H})$, see Section A.1.2. Let W be , for sake of simplicity, a one-dimensional standard Brownian motion. The following theorem are proved in Pardoux (1979b).

Theorem 4.7. *Given a triple $(\mathcal{V},\mathcal{H},\mathcal{V}')$ which satisfies Assumption 4.1, and linear bounded operators $\bar{\mathrm{A}} : \mathcal{V} \to \mathcal{V}'$, $B : \mathcal{H} \to \mathcal{H}$. If $\bar{\mathrm{A}}$ is coercive, then there exists a unique solution of the following equation, $\forall v \in \mathcal{V}$,*

$$\begin{cases} u \in \mathcal{M}^2(0,T;\mathcal{V}), \\ (u_t, v) = (u_0, v) + \int_0^t \langle \bar{\mathrm{A}}u_s, v\rangle ds + \int_0^t (Bu_s, v) dW_s, \\ u_0 = w \in \mathcal{H}, \end{cases} \tag{4.2.2}$$

Moreover, the solution is a process of the space $\mathcal{C}^2(0,T;\mathcal{H})$, and satisfies

$$\|u_t\|^2_{\mathcal{H}} = 2\int_0^t \langle \bar{\mathrm{A}}u_s u_s\rangle ds + \|w\|^2_{\mathcal{H}} + 2\int_0^t (Bu_s, u_s)dW_s + \int_0^t \|Bu_s\|^2_{\mathcal{H}} ds, \ a.s..$$

For a proof, see Pardoux (1979b), Theorem 1.3, page 137.

Remark 4.2. *Equivalently, one can rewrite Equation* (4.2.2) *as follows*

$$u_t = u_0 + \int_0^t \bar{\mathrm{A}}u_s ds + \int_0^t Bu_s dW_s,$$

which can be consider as an equation in the space $\mathcal{V}'$.

Chapter 5

Unnormalized conditional density

The question we consider in this chapter is whether the conditional distribution of X_t, given the observation history, has a density with respect to a reference measure. We prove that, under fairly mild conditions, the unnormalized conditional distribution ρ_t has a square integrable density with respect to Lebesgue measure, which is a weak solution of a stochastic partial differential equation.

There are various approaches to answer this question, giving the similar results. Pardoux (1979b), Pardoux (1979a) and Germani and Piccioni (1984), using adjoint equation, studied the nonlinear filtering problem where the state process X is a Markov diffusion process. They proved results in the case of an observation corrupted by a Wiener noise. Pardoux (1979a) proved a result in the case that the observation is a marked point process (for instance a poisson process), whose predictable projection (the stochastic intensity in the case of a point process) is a given function of the signal X. Bain and Crisan (2009) studied the continuous version of nonlinear filtering problem where the state process X is a Markov diffusion process and partially observed. Bain and Crisan (2009) have shown there exists a square integrable density of the unnormalized conditional distribution and studied the differentiability of the density.

The approach presented here is that adopted by Pardoux (1979b) and Pardoux (1979a). We generalized their results to the nonlinear filtering associated with Model (2.2.9), where the observation processes have both Wiener and Poisson noise.

This chapter is organized as follows.

The unnormalized conditional density is the solution of a SPED, the so called Zakai equation. In Section 5.1, we assume some regularity for the coefficients of the nonlinear filtering problem w.r.t. Model (2.2.9), which guarantee the existence and uniqueness of the solution of the Zakai equation and its regularity. Its regularity helps us to show the solution is nothing more than the unnormalized conditional density.

In Section 5.2, we give the main result of this chapter. That is, the Zakai equation has a unique solution in some spaces and the solution is the unnormalized conditional density.

In Section 5.3, we show the ideas to obtain the main result. It is to generalize the classical Feynman-Kac formula for second-order parabolic deterministic PDEs. The key tools are the adjoint equations.

Finally, in Section 5.4, we give the proof of our main result.

5.1 Assumptions

We study the nonlinear filtering problem w.r.t. Model (2.2.9). In Section 5.2, we will show the existence of the unnormalized conditional density and it is the solution of a Zakai equation. Therefore, we need some regularity of the coefficients such that assumptions similar to Theorem 4.7 are satisfied.

We are looking for solutions of the Zakai equation, which is a SPDE, in certain functional spaces such as Sobolev spaces. To guarantee the existence of the solution, the spaces should satisfy Assumption 4.1. Going forward, let H be $L^2(\mathbb{R}^d)$, V be $H^1(\mathbb{R}^d)$, and V' be $H^{-1}(\mathbb{R}^d)$, then the triple (V, H, V') satisfies Assumption 4.1, which is mentioned in Example 4.1.

In the following, we need more regularity of b, σ, h and λ. We keep the assumptions from Section 2.2, and we add the following assumptions throughout this chapter.

Assumption 5.1. *i) $b, \sigma,$ and h are bounded on $\mathbb{R}^d$.*

ii) b and σ are continuous with bounded derivatives. Moreover σ has bounded second order derivative.

iii) Set $a(x) = \sigma(x)\sigma(x)^\top$. There exists $\alpha > 0$, such that $z^\top a(x) z \geq \alpha z^\top z$, $\forall x, z \in \mathbb{R}^d$.

iiii) Assumption 2.2 holds.

Such hypotheses guarantee that:

Proposition 5.1. *If Assumption 5.1 holds, then,*

1) The system (2.2.1) *has a unique strong solution.*

2) An operator $\mathcal{A} : V \to V'$, defined by

$$\langle \mathcal{A}u, v\rangle := -\frac{1}{2}\sum_{i,j=1}^{d}\int_{\mathbb{R}^d} a_{ij}(x)\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j}dx + \sum_{i=1}^{d}\int_{\mathbb{R}^d}\bar{a}_i\frac{\partial u}{\partial x_i}v dx, \tag{5.1.1}$$

where

$$\bar{a}_i := b_i - \frac{1}{2}\sum_{j=1}^{d}\frac{\partial a_{ij}(x)}{\partial x_j},$$

is a bounded linear operator.

3) $\mathcal{A}$ satisfies the coercivity assumption, defined by Definition 4.7.

4) The restriction of $\mathcal{A}$ to $D(\mathcal{A})$ defined as

$$D(\mathcal{A}) := \{u \in V,\ |\langle \mathcal{A}u, v\rangle| \leq K_u \|v\|_H,\ v \in V\}, \tag{5.1.2}$$

where K_u depends on u, generates an analytic semigroup $\{G_t, t \in [0,T]\}$, such that

$$\|G_t\| \leq e^{\beta t}, \tag{5.1.3}$$

where β is the coefficient in the coercivity assumption, see Lemma 5.2.

5) Multiplication by h, λ defines a bounded self-adjoint operator on H.

Remark 5.1. *$\mathcal{A}$ is an extension of $\mathcal{L}$, which is defined by Equation* (2.2.4)*. It is obtained by converting $\mathcal{L}$ into its variational representation, or weak representation. To be precise, if the partial derivatives of u and v, which are compact supported, exist in the conventional sense and is continuous up to order 2, $\mathcal{L}$ can be rewritten in its divergence form from Equation* (2.2.4)*, by integration by parts, assuming $d = 1$ for simplicity,*

$$\begin{aligned}
(\mathcal{L}u, v) =& \int_{\mathbb{R}} \Big[b(x)\frac{du}{dx} + \frac{1}{2}a(x)\frac{d^2u}{dx^2}\Big] v dx \\
=& \int_{\mathbb{R}} b(x)u'v dx + \frac{1}{2}\int_{\mathbb{R}} [a(x)v]u'' dx \\
=& \int_{\mathbb{R}} b(x)u'v dx + \frac{1}{2}\Big\{(av)u'\Big|_{-\infty}^{\infty} - \int_{\mathbb{R}} [a(x)v]'u' dx\Big\} \\
=& \int_{\mathbb{R}} b(x)u'v dx - \frac{1}{2}\Big\{\int_{\mathbb{R}} [a'(x)v + av']u' dx\Big\} \\
=& -\frac{1}{2}\int_{\mathbb{R}} au'v' dx + \int_{\mathbb{R}} \Big[b(x) - \frac{1}{2}a'(x)\Big]u'v dx \\
=& \langle \mathcal{A}u, v\rangle .
\end{aligned}$$

Proposition 5.1 shows $\mathcal{A}$ is linear and continuous from V into V'.

Proof of Proposition 5.1. In order to simplify the notations, we suppose that $d = 1$. The more general case is handled in exactly the same way.

1) This property follows from Assumption ii) and the boundedness of b and σ. By Assumption ii), there exist $x \le \theta \le y$, $x \le \xi \le y$, such that

$$\begin{aligned}
\|b(x) - b(y)\| &= \|b'(\theta)(x - y)\| \le \|b'\|_\infty \|x - y\|, \\
\|\sigma(x) - \sigma(y)\| &= \|\sigma'(\xi)(x - y)\| \le \|\sigma'\|_\infty \|x - y\|.
\end{aligned}$$

Therefore Equation (2.2.2) holds. By Assumption i)

$$\begin{aligned}
\|b(x)\| &\le \|b\|_\infty \le \|b\|_\infty(1 + \|x\|), \\
\|\sigma(x)\| &\le \|\sigma\|_\infty \le \|\sigma\|_\infty(1 + \|x\|).
\end{aligned}$$

Therefore Equation (2.2.3) holds.

2) This property follows from Assumption ii) and the boundedness of b and σ. $\forall u, v \in V$, by the

definition of $\mathcal{A}$,

$$\begin{aligned}
\Big|\langle \mathcal{A}u, v\rangle\Big| =& \Big| -\frac{1}{2}\int_{\mathbb{R}} a(x)u'(x)v'(x)dx + \int_{\mathbb{R}} \bar{a}(x)u'(x)v(x)dx\Big| \\
=& \Big|\int_{\mathbb{R}} u'(-\frac{1}{2}av' + \bar{a}v)dx\Big| \\
=& \Big|(u', -\frac{1}{2}av' + \bar{a}v)\Big| \\
\le& \|u'\|_H \cdot \| -\frac{1}{2}av' + \bar{a}v\|_H \\
\le& \|u'\|_H \Big(\|\frac{1}{2}av'\|_H + \|\bar{a}v\|_H\Big) \\
\le& \|u'\|_H(\|av'\|_H + \|\bar{a}v\|_H) \qquad (5.1.4) \\
\le& \|u\|_V\|v\|_V(\|a\|_\infty + \|\bar{a}\|_\infty) \\
\le& c\|u\|_V\|v\|_V,
\end{aligned}$$

where c is a constant only depends on a, b and a'. The last inequality follows from a, b and a' are bounded. Therefore, by the definition of $\|\cdot\|_{V'}$

$$\|\mathcal{A}u\|_{V'} = \sup_{v\in V, \|v\|_V \le 1} |\langle \mathcal{A}u, v\rangle| \le c\|u\|_V,$$

that is $\mathcal{A}$ is a bounded linear operator.

3) See Lemma 5.2.

4) This property is obtain by Theorem 4.2 or Prato and Zabczyk (1992), Proposition A.10, page 389.

5) This property follows from the boundedness of h and λ. Boundedness follows from

$$\begin{aligned}
\|hu\|_H \le& \|h\|_\infty\|u\|_H, \\
\|\lambda u\|_H \le& \|\lambda\|_\infty\|u\|_H.
\end{aligned}$$

Self-adjoint follows from

$$\begin{aligned}
(hu, v) &= \int_{\mathbb{R}^d} h(x)u(x)\cdot v(x)dx = \int_{\mathbb{R}^d} u(x)\cdot h(x)v(x)dx = (u, hv), \\
(\lambda u, v) &= \int_{\mathbb{R}^d} \lambda(x)u(x)\cdot v(x)dx = \int_{\mathbb{R}^d} u(x)\cdot \lambda(x)v(x)dx = (u, \lambda v).
\end{aligned}$$

□

Lemma 5.2. *Assumption 5.1 implies that $\mathcal{A}$ satisfies the coercivity assumption, that is, there exists $c > 0$ and $\beta \ge 0$ such that, $\forall v \in V$,*

$$\beta\|v\|_H^2 - \langle \mathcal{A}v, v\rangle \ge c\|v\|_V^2. \qquad (5.1.5)$$

Proof. It is suffice to consider $v \in V$ with compact support. Definition of $\mathcal{A}$ implies,

$$
\begin{aligned}
-\langle \mathcal{A}v, v\rangle =&\frac{1}{2}\sum_{i,j=1}^{d}\int_{\mathbb{R}^d} a_{ij}(x)\frac{\partial v}{\partial x_i}\frac{\partial v}{\partial x_j}dx - \sum_{i=1}^{d}\int_{\mathbb{R}^d}\bar{a}_i\frac{\partial v}{\partial x_i}vdx\\
\geq&\frac{1}{2}\alpha\int_{\mathbb{R}^d}\Big[\sum_{i=1}^{d}\Big(\frac{\partial v}{\partial x_i}\Big)^2\Big]dx - \frac{1}{2}\sum_{i=1}^{d}\int_{\mathbb{R}^d}\bar{a}_i\frac{\partial}{\partial x_i}(v^2)dx\\
=&\frac{1}{2}\alpha\int_{\mathbb{R}^d}\Big[\sum_{i=1}^{d}\Big(\frac{\partial v}{\partial x_i}\Big)^2+v^2\Big]dx - \frac{1}{2}\alpha\int_{\mathbb{R}^d}v^2dx\\
&-\frac{1}{2}\sum_{i=1}^{d}\Big[\bar{a}_iv^2\Big|_{-\infty}^{+\infty} - \int_{\mathbb{R}^d}\frac{\partial \bar{a}_i}{\partial x_i}v^2dx\Big]\\
=&\frac{1}{2}\alpha\|v\|_V^2 - \frac{1}{2}\alpha\|v\|_H^2 + \frac{1}{2}\sum_{i=1}^{d}\Big[\int_{\mathbb{R}^d}\frac{\partial \bar{a}_i}{\partial x_i}v^2dx\Big]\\
\geq&\frac{1}{2}\alpha\|v\|_V^2 - \frac{1}{2}\alpha\|v\|_H^2 - \frac{d}{2}\|\nabla\bar{a}\|_\infty\int_{\mathbb{R}^d}v^2dx\\
=&\frac{1}{2}\alpha\|v\|_V^2 - \Big[\frac{1}{2}\alpha + \frac{d}{2}\|\nabla\bar{a}\|_\infty\Big]\|v\|_H^2,
\end{aligned}
$$

where the second equality follows from iii) in Assumption 5.1. Taking

$$
c := \frac{1}{2}\alpha, \quad \beta := \frac{1}{2}\alpha + \frac{d}{2}\|\nabla\bar{a}\|_\infty, \tag{5.1.6}
$$

we obtain the desired result. □

It is easy to see that β can be any real number which is greater or equal to $\frac{1}{2}\alpha + \frac{d}{2}\|\nabla\bar{a}\|_\infty$.

5.2 Main result

Let $\mathcal{M}^2(0,T;\mathcal{V})$ denote the space of $\mathcal{V}$-valued stochastic process with some regular properties. And let $\mathcal{S}^2(0,T;\mathcal{H})$ denote the spaces of $\mathcal{H}$-valued càdlàg stochastic processes with some regular properties. For details of processes spaces $\mathcal{M}^2(0,T;\mathcal{V})$ and $\mathcal{S}^2(0,T;\mathcal{H})$, see Section A.1.2. Now, the main result of the chapter.

Theorem 5.3. *If Assumption 5.1 holds, then, $\forall v \in V$, Zakai equation*

$$
\begin{cases}
q \in \mathcal{M}^2(0,T;V),\\
(v,q_t) = (v,q_0) + \int_0^t\langle \mathcal{A}v, q_s\rangle ds + \int_0^t(h^\top v, q_s)dZ_s + \int_0^t\Big((\lambda-1)v, q_{s-}\Big)d(Y_s - s),\\
q_0 = p_0 \in H,
\end{cases} \tag{5.2.1}
$$

has a unique solution which satisfies moreover $q \in \mathcal{S}^2(0,T;H)$ and

$$
(v,q_t) = \rho_t(v), \;\; v \in L^\infty(\mathbb{R}^d). \tag{5.2.2}
$$

Here the solution $q(\omega) = \{q_t(\omega)\}_{t\in[0,T]}$ is a process with values in Hilbert space V with norm $\|\cdot\|_V$. By Theorem 5.3, Zakai equation (5.2.1) has a unique solution q and, applying Equation

(5.2.2) and recalling that ρ is the unnormalized conditional measure defined by Equation (2.4.8), the solution is the conditional density. So q has the meaning of an 'unnormalized conditional density'.

Here, by Proposition 2.5, Z is a standard Brownian motion and Y is a Poisson process with intensity 1 under measure $\mathbb{P}^0$. And, as argued before, $\mathcal{A}: V \to V'$ is a bounded linear operator, multiplication by h, $\lambda - 1$ defines a bounded linear self-adjoint operator of H, and the coercivity assumption, by Lemma 5.2, is satisfied. The equation is linear parabolic stochastic partial equation. We are looking for solution $q = \{q_t, t \in [0,T]\}$ of Equation (5.2.1) in certain functional spaces such as Sobolev spaces.

Remark 5.2. *We look for the solution of the Zakai equation in the stochastic processes space $\mathcal{M}^2(0,T;V)$. Together with the boundedness condition of the coefficients, the integrals appearing in Equation (5.2.1) are well defined, for $0 \le t \le T$. First we have*

$$\int_0^t \langle \mathcal{A}v, q_s \rangle ds, \quad t \in [0,T],$$

is well defined because the integrand is measurable and, noting that $\mathcal{A}$ is a bounded linear operator from V to V',

$$\begin{aligned}\mathbb{E}^0 \int_0^t \Big|\langle \mathcal{A}v, q_s \rangle\Big| ds \le & \mathbb{E}^0\Big[c\int_0^t \|v\|_V \|q_s\|_V ds\Big] \\ \le & c\|v\|_V \mathbb{E}^0\Big\{\int_0^t \|q_s\|_V ds\Big\} \\ \le & c\|v\|_V \mathbb{E}^0\Big\{\int_0^t \Big(\|q_s\|_V^2 + 1\Big) ds\Big\} < \infty,\end{aligned}$$

the last inequality follows from $q \in \mathcal{M}^2$.

Then, we have

$$\int_0^t (h^\top v, q_s) dZ_s, \quad t \in [0,T],$$

is well defined because the integrand is $\mathbb{F}^{Z,Y}$-adapted and, by Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}^0 \int_0^t \Big\|(h^\top v, q_s)\Big\|^2 ds \le & \mathbb{E}^0\Big[\|h\|_\infty \int_0^t \|q_s\|_H^2 \|v\|_H^2 ds\Big] \\ \le & \|h\|_\infty \|v\|_V^2 \mathbb{E}^0 \int_0^t \|q_s\|_V^2 ds < \infty.\end{aligned}$$

Finally, we have

$$\int_0^t \Big((\lambda-1)v, q_{s-}\Big) d(Y_s - s), \quad t \in [0,T],$$

is well defined because the integrand is predictable and, by Brémaud (1981), Theorem T8, page 27,

$$\begin{aligned}\mathbb{E}^0 \int_0^t \Big|\Big(q_{s-}, (\lambda-1)v\Big)\Big| ds \le & \mathbb{E}^0\Big[\|\lambda - 1\|_\infty \cdot \|v\|_H \int_0^t \|q_{s-}\|_H ds\Big] \\ \le & \|\lambda - 1\|_\infty \cdot \|v\|_V \mathbb{E}^0\Big[\int_0^t \|q_{s-}\|_V ds\Big] \\ \le & \|\lambda - 1\|_\infty \cdot \|v\|_V \mathbb{E}^0\Big[\int_0^t \Big(\|q(s-)\|_V^2 + 1\Big) ds\Big] < \infty.\end{aligned}$$

Remark 5.3. *By Theorem 5.3,* $q \in \mathcal{M}^2(0,T;V) \cap S^2(0,T;H)$. $q \in \mathcal{M}^2(0,T;V)$ *means that* q *is* V*-valued stochastic process, which guarantees* $\langle \mathcal{A}v, q_s\rangle$ *makes sense, and the integrals in Equation* (5.2.1) *are well defined.* $q \in S^2(0,T;H)$ *means each trajectory of* q *is* H*-valued and càdlàg.*

Let $\mathcal{A}^* \in \mathcal{L}(V,V')$ be the adjoint operator of $\mathcal{A}$, then $\langle \mathcal{A}^*v, u\rangle = \langle \mathcal{A}u, v\rangle$, $\forall u, v \in V$. More precisely, $\mathcal{A}^*$ is defined as follows. First, define a continuous linear transformation $J : V \to V''$ by

$$J(v)f = f(v), \ \ \forall v \in V, \ \ f \in V'.$$

Then, define $\mathcal{A}^* : V \to V'$ by

$$A^*(v) = J(v) \circ \mathcal{A}, \ \ \forall v \in V.$$

Finally, we have, $\forall u, v \in V$,

$$\langle \mathcal{A}^*v, u\rangle = \mathcal{A}^*v(u) = (J(v) \circ \mathcal{A})(u) = J(v)(\mathcal{A}u) = \mathcal{A}u(v) = \langle \mathcal{A}u, v\rangle,$$

where the first equality follows from the definition of duality pairing, the second equality follows from the definition of A^*, the fourth equality follows from the definition of J, and the last equality follows from the definition of duality pairing. Therefore, $\mathcal{A}^*$ is a bounded linear operator from V to V', that is $\mathcal{A}^* \in \mathcal{L}(V,V')$.

Now, Equation (5.2.1) can be rewritten as:

$$\begin{cases} q \in \mathcal{M}^2(0,T;V), \\ (q_t, v) = (q_0, v) + \int_0^t \langle \mathcal{A}^*q_s, v\rangle ds + \int_0^t (h^\top q_s, v) dZ_s + \int_0^t \Big((\lambda - 1)q_{s-}, v\Big) d(Y_s - s), \\ q_0 = p_0 \in H. \end{cases} \tag{5.2.3}$$

Equivalent to Equation (5.2.1), we consider the following:

$$\begin{cases} q \in \mathcal{M}^2(0,T;V), \\ dq_t = \mathcal{A}^*q_t dt + h^\top q_t dZ_t + (\lambda - 1)q_{t-} d(Y_t - t), \quad t \in [0,T], \\ q_0 = p_0 \in H, \end{cases} \tag{5.2.4}$$

where p_0 is the density of the law of X_0. Equation (5.2.4) can be considered as an equation in V'.

5.3 Finding the unnormalized conditional density

The objective here is to show the idea of Theorem 5.3.

For the proof of Theorem 5.3, we follow the idea of Pardoux (1979b) and Pardoux (1979a). The so-called adjoint equations play a key role in dealing with the problem. Adjoint equations are, in general, backward equations with given terminal states.

We show how to find the density of the unnormalized conditional measure for the case that all the coefficients are 'sufficiently nice'.

The idea is to generalize the classical Feynman-Kac formula for second-order parabolic (deterministic) PDE's. Let $\{q_t, t \in [0,T]\}$ be a V-valued stochastic process. For any $0 \le \theta \le T$, suppose that there exist a V-valued stochastic process $\{r_t^\theta, t \in [0,\theta]\}$ and, for f bounded and square integrable,

i) $r_t^\theta(x) = \mathbb{E}^0\Big[f(X_\theta)\Lambda_\theta/\Lambda_t\Big|\sigma\{Z_s - Z_t, Y_s - Y_t, t \leq s \leq \theta\},\ X_t = x\Big]$, $\forall t \in [0,\theta]$, $x \in \mathbb{R}^d$, recalling that Λ is defined by Equation (2.4.5).

ii) q is adjoint to r^θ, that is almost all trajectories of the process $R_t := (q_t, r_t^\theta)$, $t \in [0,\theta]$ are constant. And $q_0 = p_0$.

These two properties guarantee that q is the unnormalized conditional density. To be precise, by property ii), q is adjoint to r^θ, then we have $R_0 = R_\theta$, that is

$$(q_0, r_0^\theta) = (q_\theta, r_\theta^\theta). \tag{5.3.1}$$

Notice that, by property i), $\forall x \in \mathbb{R}^d$, we get the starting and terminal values of r^θ,

$$\begin{aligned} r_\theta^\theta(x) &= f(x), \\ r_0^\theta(x) &= \mathbb{E}^0[f(X_\theta)\Lambda_\theta|\mathcal{F}_\theta^{Z,Y},\ X_0 = x]. \end{aligned}$$

Therefore (q_0, r_0^θ) and $(q_\theta, r_\theta^\theta)$ can be rewritten as

$$(q_0, r_0^\theta) = \int_{\mathbb{R}^d} p_0(x)\mathbb{E}^0\Big[f(X_\theta)\Lambda_\theta\Big|\mathcal{F}_\theta^{Z,Y},\ X_0 = x\Big]dx = \mathbb{E}^0[f(X_\theta)\Lambda_\theta|\mathcal{F}_\theta^{Z,Y}], \tag{5.3.2}$$

$$(q_\theta, r_\theta^\theta) = \int_{\mathbb{R}^d} q_\theta(x)f(x)dx. \tag{5.3.3}$$

To sum up, Equation (5.3.1), (5.3.2) and (5.3.3), give us

$$\int_{\mathbb{R}^d} q_\theta(x)f(x)dx = \mathbb{E}^0[f(X_\theta)\Lambda_\theta|\mathcal{F}_\theta^{Z,Y}].$$

Since θ is arbitrary, the fact that q is the unnormalized conditional density follows immediately.

Even so, this strategy is not the only one. For example, in Bain and Crisan (2009), page 165-179, the authors derive the result when the observation is only corrupted by a Wiener noise, under certain conditions. They show the following result step by step:

i Almost surely the unnormalised conditional distribution ρ_t has a density with respect to Lebesgue measure and this density is square integrable, that is, it is in H.

ii There exists a unique solution q of Equation (5.2.1).

iii Let $\tilde{\rho}$ be the measure with respect to Lebesgue measure with density q. Show that $\tilde{\rho}$ satisfies the Zakai equation (2.4.11). Although one cannot conclude that $\tilde{\rho}$ is equal to ρ, having not proven the uniqueness of the Zakai equation (2.4.11), Bain and Crisan (2009), on page 96 and page 177 take the PDE approach to solve the uniqueness problem.

iv By the PDE approach, to show that, a.s., for any $\phi \in C_k^\infty(\mathbb{R}^d)$,

$$\tilde{\rho}_t(\phi) = \rho_t(\phi).$$

v Combining the obtained result q is the unique density of the unnormalised conditional distribution.

However, these procedures require stronger regularity properties on the coefficients and we need more work for the PDE approach with jump observation. Instead, we use the same approach as that adopted in Pardoux (1979b), Pardoux (1979a) and Germani and Piccioni (1984) which works more with mild solutions.

5.4 Proof of Theorem 5.3

This section is devoted to proving Theorem 5.3. To find the unnormalized conditional density, we have the following key steps.

- In Section 5.4.1, we show the Zakai equation (5.2.1), the forward one, has a unique solution with starting state p_0, see Theorem 5.4.
- In Section 5.4.2, we construct a backward stochastic PDE with terminal states f. We show the backward one has a unique solution r from r_0 to f, See Theorem 5.6. An interesting fact is that r is the conditional statistics of the filtering problem, see Theorem 5.8.
- In Section 5.4.3, we show that the forward equation is the adjoint to the backward one, see Theorem 5.12.
- Combining the obtained results, we deduce that Zakai equation (5.2.1) describes the evolution of the unnormalized conditional density. See Theorem 5.13.

5.4.1 Some existence and uniqueness results on stochastic PDEs

In this section, we generalize Theorem 4.7 to the case of SPDEs driven by Lévy process. This enables us to study the existence and uniqueness for the solution of Zakai equation (5.2.1).

Our main result of this section is that Equation (5.2.4) has a unique solution with càdlàg trajectories in H. The existence of càdlàg solution helps us to obtain some boundedness of the solution which is useful for Lemma 5.10.

Theorem 5.4. *If Assumption 5.1 holds, Equation* (5.2.4) *has a unique solution* $q = \{q_t\}_{t\in[0,T]} \in \mathcal{S}^2(0,T;H) \cap \mathcal{M}^2(0,T;V)$. *Consequently, it satisfies*

$$\mathbb{E}^0\Big[\sup_{0\le t\le T} \|q_t\|_H^2 + \int_0^T \|q_t\|_V^2 dt\Big] < \infty. \tag{5.4.1}$$

Equation (5.2.4) is a linear parabolic SPDE. Various methods have been used to study linear SPDEs. One method, the variational approach, was developed from contributions made by Lions (1961) and others. Another is the semigroup approach. Both methods achieve similar results. These are reviewed in Chapter 4.

For the semigroup approach, this result can be obtained by an application of Peszat and Zabczyk (2007), Theorem 9.29, page 164, which gives the general result of the uniqueness and existence of weak solution of stochastic partial differential equation with Lévy noise. The existence and uniqueness can be obtained from the property, guaranteed by Proposition 5.1, that restriction $\mathcal{A}$ is the generator of a C_0-semigroup and multiplication by h, λ defines a bounded self-adjoint operator on H. Further, analytic semigroup is an important class of C_0 semigroup and variational generator is an important subclass of the generator of analytic semigroups. Under Assumption 5.1, $\mathcal{A}$ is variational, and in sequence G is a generalized contraction. The solution then has a càdlàg version.

In the following, we give the proof of Theorem 5.4 by generalizing the results of Pardoux (1979b), which is the existence and uniqueness result of SPEDs driven by a Brownian motion and is obtained by variational approach. In the proof, it will be convenient to model the processes

X, Y and Z on a product space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P}^0)$. We denote by $(\Omega_2, \mathcal{F}^2, (\mathcal{F}_t^2)_{0\le t\le T}, \mathbb{P}^{0,2})$ the Poisson space and by $(\Omega_1, \mathcal{F}^1, (\mathcal{F}_t^1)_{0\le t\le T}, \mathbb{P}^{0,1})$ the Wiener space. By the independent of X, Z, and Y, we have $\Omega = \Omega_1 \otimes \Omega_2$, $\mathcal{F} = \mathcal{F}^1 \otimes \mathcal{F}^2$, $\mathcal{F}_t = \mathcal{F}_t^1 \otimes \mathcal{F}_t^2$, $\mathbb{P}^0 = \mathbb{P}^{0,1} \otimes \mathbb{P}^{0,2}$, and for $\omega = (\omega_1, \omega_2) \in \Omega$,

$$X_t(\omega) = X_t(\omega_1),\ \ Z_t(\omega) = Z_t(\omega_1),\ \ Y_t(\omega) = Y_t(\omega_2),$$

and $\tau_i(\omega) = \tau_i(\omega_2)$, for $i = 1, 2, \ldots$, which are the jumping times of Y.

Proof of Theorem 5.4. We solve Equation (5.2.4) forward for each ω_2. Suppose $\tau_1(\omega_2) \le T$, we then solve Equation (5.4.25) from 0 to $\tau_1(\omega_2)$,

$$\begin{cases} dq_t = \mathcal{A}^* q_t dt + h^\top q_t dZ_t - (\lambda - 1) q_t dt, \\ q_0 = p_0. \end{cases} \tag{5.4.2}$$

By Theorem 4.7 or Pardoux (1979b), Theorem 1.3, Equation (5.4.2) has a unique solution which is a V-valued stochastic process with continuous trajectories in H, and, for $t < \tau_1(\omega_2)$,

$$\begin{aligned} &\|q_t\|_H^2 - 2\int_0^t \langle \mathcal{A}^* q_s, q_s\rangle ds \\ =&\|p_0\|_H^2 + 2\int_0^t (h^\top q_s, q_s) dZ_s + \int_0^t \|hq_s\|_H^2 ds - 2\int_0^t \Big((\lambda-1)q_s, q_s\Big) ds. \end{aligned} \tag{5.4.3}$$

Notice that, by Equation (5.2.4),

$$q_{\tau_1} - q_{\tau_1-} = (\lambda - 1) q_{\tau_1-}.$$

That is

$$q_{\tau_1} = \lambda q_{\tau_1-}. \tag{5.4.4}$$

Equation (5.4.4) defines q_{τ_1} as an element of H for λ bounded and $q_{\tau-} \in H$. Moreover, we have

$$\begin{aligned} \|q_{\tau_1}\|_H^2 - \|q_{\tau_1-}\|_H^2 =&(q_{\tau_1}, q_{\tau_1}) - (q_{\tau_1-}, q_{\tau_1-}) \\ =&(\lambda q_{\tau_1-}, \lambda q_{\tau_1-}) - (q_{\tau_1-}, q_{\tau_1-}) \\ =&\Big((\lambda^2 - 1) q_{\tau_1-}, q_{\tau_1-}\Big). \end{aligned} \tag{5.4.5}$$

Repeating this procedure, we define for each ω_2 a unique stochastic process $q(\omega_2)$ which is V-valued and with càdlàg trajectories in H. It now remains to show that Equation (5.4.1) holds. We first show that

$$\mathbb{E}^0\Big\{\int_0^T \|q_s\|_V^2\Big\} < \infty. \tag{5.4.6}$$

By Combining Equation (5.4.3) and (5.4.5), we get

$$\begin{aligned} &\|q_t\|_H^2 - 2\int_0^t \langle \mathcal{A}^* q_s, q_s\rangle ds \\ =&\|p_0\|_H^2 + 2\int_0^t (h^\top q_s, q_s) dZ_s + \int_0^t \|hq_s\|_H^2 ds \\ &+ \int_0^t \Big((\lambda^2-1) q_{s-}, q_{s-}\Big) d(Y_s - s) + \int_0^t \Big\|(\lambda-1)q_s\Big\|_H^2 ds. \end{aligned} \tag{5.4.7}$$

Let $t_n = \inf\{0 \le t \le T, \|q_t\|_H > n\}$. Using martingale properties with respect to $\{\mathcal{F}_t^{Z,Y}\}$, we get

$$\begin{aligned}&\mathbb{E}^0\|q_{t\wedge t_n}\|_H^2 - 2\mathbb{E}^0\int_0^{t\wedge t_n}\langle \mathcal{A}^* q_s, q_s\rangle ds\\ =&\|p_0\|_H^2 + \mathbb{E}^0\int_0^{t\wedge t_n}\|hq_s\|_H^2 ds + \mathbb{E}^0\int_0^{t\wedge t_n}\Big\|(\lambda-1)q_s\Big\|_H^2 ds.\end{aligned}$$

Notice that h and λ are bounded. For sake of simplicity, define

$$\bar{c} := \max\{\|h\|_\infty, \|\lambda\|_\infty\}. \tag{5.4.8}$$

Then we get

$$\mathbb{E}^0\|q_{t\wedge t_n}\|_H^2 \le \|p_0\|_H^2 + 2\mathbb{E}^0\int_0^{t\wedge t_n}\langle \mathcal{A}^* q_s, q_s\rangle ds + [\bar{c}^2 + (\bar{c}+1)^2]\mathbb{E}^0\int_0^{t\wedge t_n}\|q_s\|_H^2 ds.$$

Applying Lemma 5.2, there exists $c > 0$ and $\beta \ge 0$, such that

$$\langle \mathcal{A}^* q_s, q_s\rangle \le \beta\|q_s\|_H^2 - c\|q_s\|_V^2. \tag{5.4.9}$$

Therefore

$$\mathbb{E}^0\|q_{t\wedge t_n}\|_H^2 \le \|p_0\|_H^2 - 2c\mathbb{E}^0\int_0^{t\wedge t_n}\|q_s\|_V^2 ds + c_2\mathbb{E}^0\int_0^{t\wedge t_n}\|q_s\|_H^2 ds, \tag{5.4.10}$$

where

$$c_2 := \bar{c}^2 + (\bar{c}+1)^2 + 2\beta, \tag{5.4.11}$$

is a constant. When $n \to \infty$

$$t_n \to T,\ \ t\wedge t_n \downarrow t.$$

Notice that q is right-continuous,

$$q_{t\wedge t_n} \to q_t, a.s.$$

Notice that q has left-limit,

$$\begin{aligned}\int_0^{t\wedge t_n}\|q_s\|_V^2 ds &\to \int_0^t \|q_s\|_V^2 ds,\ \ a.s.\\ \int_0^{t\wedge t_n}\|q_s\|_H^2 ds &\to \int_0^t \|q_s\|_H^2 ds,\ \ a.s.\end{aligned}$$

By Fatou's Lemma, and the Monotone Convergence Theorem, see Folland (1999), page 52 and page 50,

$$\begin{aligned}&\mathbb{E}^0\|q_{t\wedge t}\|_H^2\\ \le&\lim_{n\to\infty}\mathbb{E}^0\|q_{t\wedge t_n}\|_H^2\\ \le&\|p_0\|_H^2 + \lim_{n\to\infty}\Big[-2c\mathbb{E}^0\int_0^{t\wedge t_n}\|q_s\|_V^2 ds + c_2\mathbb{E}^0\int_0^{t\wedge t_n}\|q_s\|_H^2 ds\Big]\\ \le&\|p_0\|_H^2 - 2c\mathbb{E}^0\int_0^t\|q_s\|_V^2 ds + c_2\mathbb{E}^0\int_0^t\|q_s\|_H^2 ds.\end{aligned}$$

That is

$$\mathbb{E}^0\|q_t\|_H^2 + 2c\mathbb{E}^0 \int_0^t \|q_s\|_V^2 ds \leq \|p_0\|_H^2 + c_2\mathbb{E}^0 \int_0^t \|q_s\|_H^2 ds. \tag{5.4.12}$$

Therefore we have, noting $c > 0$,

$$\mathbb{E}^0\|q_t\|_H^2 \leq \|p_0\|_H^2 + c_2 \int_0^t \mathbb{E}^0\|q_s\|_H^2 ds.$$

By Gronwall's lemma, we have

$$\mathbb{E}^0\|q_t\|_H^2 \leq \|p_0\|_H^2 \int_0^t e^{c_2 s} ds.$$

That is

$$\sup_{0\leq t\leq T} \mathbb{E}(\|q_t\|_H^2) \leq \|p_0\|_H^2 \int_0^T e^{c_2 s} ds < \infty. \tag{5.4.13}$$

Combining Formula (5.4.12) and (5.4.13), we deduce that

$$\begin{aligned}
\mathbb{E}^0\Big\{ \int_0^T \|q_s\|_V^2 \Big\} <& \frac{1}{2c}\Big[\|p_0\|_H^2 + c_2\mathbb{E}^0 \int_0^T \|q_s\|_H^2 ds\Big] && (5.4.14)\\
=& \frac{1}{2c}\Big[\|p_0\|_H^2 + c_2 \int_0^T \mathbb{E}^0\|q_s\|_H^2 ds\Big]\\
\leq& \frac{1}{2c}\Big[\|p_0\|_H^2 + c_2\cdot T\cdot \sup_{0\leq s\leq T} \mathbb{E}^0\|q_s\|_H^2\Big]\\
\leq& \frac{1}{2c}\Big[\|p_0\|_H^2 + c_2\cdot T\cdot \|p_0\|_H^2 \int_0^T e^{c_2 s} ds] && (5.4.15)\\
<& \infty.
\end{aligned}$$

Equation (5.4.6) is obtained. Compare to Equation (5.4.1), it remains to show that

$$\mathbb{E}^0\Big\{ \sup_{0\leq t\leq T} \|q_t\|_H^2 \Big\} \leq \infty. \tag{5.4.16}$$

Equation (5.4.7) implies

$$\begin{aligned}
\sup_{0\leq t\leq T} \|q(t)\|_H^2 \leq& \|p_0\|_H^2 + 2\int_0^T \Big|\langle \mathcal{A}^* q_s, q_s\rangle\Big| ds && (5.4.17)\\
&+ \int_0^T \|hq_s\|_H^2 ds + \int_0^T \|(\lambda - 1)q_s\|_H^2 ds\\
&+ 2\sup_{0\leq t\leq T} \Big| \int_0^t (h^\top q_s, q_s) dZ_s\Big|\\
&+ \sup_{0\leq t\leq T} \Big| \int_0^t \int_{\mathbb{R}^d} \Big(\lambda^2(x) - 1\Big) q_{s-}(x)^2 dx d(Y_s - s)\Big|.
\end{aligned}$$

By Equation (5.4.9),

$$\int_0^T \Big|\langle \mathcal{A}^* q_s, q_s\rangle\Big| ds \leq \beta \int_0^T \|q_s\|_H^2 ds + c \int_0^T \|q_s\|_V^2 ds. \tag{5.4.18}$$

By bounded of h and λ, we have

$$\int_0^T \|hq_s\|_H^2 ds + \int_0^T \|(\lambda - 1)q_s\|_H^2 ds \leq [\bar{c}^2 + (\bar{c}+1)^2] \int_0^T \|q_s\|_H^2 ds. \tag{5.4.19}$$

Applying Davis-Burkholder-Gundy inequality, see for instance Protter (2005), Theorem 48, page 193, there exists a universal constant C which is not depending on T, h, q and Z, such that

$$\begin{aligned}
\mathbb{E}^0\Big\{2 \sup_{0\leq t\leq T}\Big|\int_0^t (h^\top q_s, q_s) dZ_s\Big|\Big\} \leq & 2C\mathbb{E}^0\sqrt{\int_0^T (hq_s, q_s)^2 ds} \\
\leq & 2\bar{c}C\mathbb{E}^0\sqrt{\int_0^T \|q_s\|_H^2 \|q_s\|_H^2 ds} \\
\leq & 2\bar{c}C\mathbb{E}^0\Big[\sup_{0\leq t\leq T}\|q_t\|_H \sqrt{\int_0^T \|q_s\|_H^2 ds}\Big] \\
= & \mathbb{E}^0\Big[\frac{\sup_{0\leq t\leq T}\|q_t\|_H}{2}\cdot 4\bar{c}C\sqrt{\int_0^T \|q_s\|_H^2 ds}\Big] \\
\leq & \frac{1}{2}\Big\{\mathbb{E}^0 \frac{1}{4}\sup_{0\leq t\leq T}\|q_t\|_H^2 + 16(\bar{c}C)^2\mathbb{E}^0\int_0^T \|q_s\|_H^2 ds\Big\} \\
\leq & \frac{1}{8}\Big\{\mathbb{E}^0\sup_{0\leq t\leq T}\|q_t\|_H^2\Big\} + 8(\bar{c}C)^2\Big\{\mathbb{E}^0\int_0^T \|q_s\|_H^2 ds\Big\}.
\end{aligned} \tag{5.4.20}$$

Similarly,

$$\begin{aligned}
& \mathbb{E}^0\Big\{\sup_{0\leq t\leq T}\Big|\int_0^t \int_{\mathbb{R}^d}\Big(\lambda^2(x) - 1\Big) q_{s-}(x)^2 dx d(Y_s - s)\Big|\Big\} \\
\leq & C\mathbb{E}^0\Big\{\int_0^T \Big[\int_{\mathbb{R}^d}\Big(\lambda^2(x) - 1\Big) q_{s-}(x)^2 dx\Big]^2 ds\Big\}^{1/2} \\
\leq & C\|\lambda^2 - 1\|_\infty \mathbb{E}^0\sqrt{\int_0^T \|q_s\|_H^2 \|q_s\|_H^2} \\
\leq & C(\bar{c}^2 + 1)\mathbb{E}^0\sqrt{\int_0^T \|q_s\|_H^2 \|q_s\|_H^2} \\
\leq & \frac{1}{8}\Big\{\mathbb{E}^0\sup_{0\leq t\leq T}\|q_t\|_H^2\Big\} + 2[C(\bar{c}^2+1)]^2\Big\{\mathbb{E}^0\int_0^T \|q_s\|_H^2 ds\Big\}.
\end{aligned} \tag{5.4.21}$$

To sum up, combining Formula (5.4.6), (5.4.17), (5.4.18), (5.4.19), (5.4.20), (5.4.21), we obtain

$$\begin{aligned}
\mathbb{E}^0\Big[\sup_{0\leq t\leq T}\|q_t\|_H^2\Big] \leq & \|p_0\|_H^2 + 2\beta\int_0^T \|q_s\|_H^2 ds + 2c\int_0^T \|q_s\|_V^2 ds \\
& + [\bar{c}^2 + (\bar{c}+1)^2]\int_0^T \|q_s\|_H^2 ds \\
& + \frac{1}{8}\Big\{\mathbb{E}^0\sup_{0\leq t\leq T}\|q_t\|_H^2\Big\} + 8(\bar{c}C)^2\Big\{\mathbb{E}^0\int_0^T \|q_s\|_H^2 ds\Big\} \\
& + \frac{1}{8}\Big\{\mathbb{E}^0\sup_{0\leq t\leq T}\|q_t\|_H^2\Big\} + 2[C(\bar{c}^2+1)]^2\Big\{\mathbb{E}^0\int_0^T \|q_s\|_H^2 ds\Big\}.
\end{aligned}$$

Together with Equation (5.4.14), it is equivalent to

$$\begin{aligned}
&\frac{3}{4}\mathbb{E}^0\Big[\sup_{0\le t\le T}\|q_t\|_H^2\Big] \qquad (5.4.22)\\
\le& 2\|p_0\|_H^2+\Big\{2\beta+c_2+[\bar{c}^2+(\bar{c}+1)^2]+8(\bar{c}C)^2+2[C(\bar{c}^2+1)]^2\Big\}\mathbb{E}^0\int_0^T\|q_s\|_H^2ds\\
\le& 2\|p_0\|_H^2+\Big\{2\beta+c_2+[\bar{c}^2+(\bar{c}+1)^2]+8(\bar{c}C)^2+2[C(\bar{c}^2+1)]^2\Big\}T\|p_0\|_H^2\int_0^T e^{c_2 s}ds\\
<& \infty.
\end{aligned}$$

We have Equation (5.4.16). Combining Formula (5.4.6) and (5.4.16), we obtain the desired result. □

Remark 5.4. *From the proof of Theorem 5.4, by Equation* (5.4.8), (5.4.11), (5.4.15), (5.4.22), *there exists a constant C_q which only depends on $\|h\|_\infty$, $\|\lambda\|_\infty$, $\|p_0\|_H$, T, c, β, such that*

$$\mathbb{E}^0\Big[\sup_{0\le t\le T}\|q_t\|_H^2+\int_0^T\|q_t\|_V^2dt\Big]<C_q.$$

Here c, β are coefficients w.r.t. the coercivity assumption of $\mathcal{A}$, see Lemma 5.2. Moreover, by Equation (5.1.6), *β is depending on α and the upper bound of b, σ and their derivatives. To sum up, C_q is a constant which only depends on $\|h\|_\infty$, $\|\lambda\|_\infty$, $\|b\|_\infty$, $\|\nabla b\|_\infty$, $\|\sigma\|_\infty$, $\|\nabla\sigma\|_\infty$, $\|\nabla(\nabla\sigma)\|_\infty$, c, $\|p_0\|_H$ and T.*

Theorem 5.5. *Let n be any integer greater or equal to* 1. *Suppose, in addition to Assumption 5.1, that all coefficients a, b, h, and λ have bounded partial derivatives in x up to order n and that moreover*

$$p_0\in H^n(\mathbb{R}^d).$$

Then, each trajectory of q, solution of Equation (5.2.4), *belongs to $D\Big([0,T];H^n(\mathbb{R}^d)\Big)$, a.s..*

Proof. We solve Equation (5.2.4) forward for each ω_2. Suppose for instance that $\tau_1(\omega_2)\le T$. We then solve Equation (5.4.25) forward from 0 to $\tau_1(\omega_2)$,

$$\begin{cases} dq_t+\mathcal{A}^*q_tdt=hq_tdZ_t-(\lambda-1)q_tdt,\\ q_0=p_0\in H^n(\mathbb{R}^d).\end{cases} \qquad (5.4.23)$$

By Pardoux (1979b), Theorem 2.1, page 142, Equation (5.4.23) defines a unique element which is an $H^{n+1}(\mathbb{R}^d)$ valued stochastic process with continuous trajectories in $H^n(\mathbb{R}^d)$. By Equation (5.4.25), the increment of u at the time of the first jump is as follows

$$q_{\tau_1}-q_{\tau_1-}=(\lambda-1)q_{\tau_1-}.$$

Then, we have

$$q_{\tau_1}=\lambda q_{\tau_1-}. \qquad (5.4.24)$$

Equation (5.4.24) defines q_τ as an element of $H^n(\mathbb{R}^d)$ for $q_{\tau-}\in H^n(\mathbb{R}^d)$ and boundedness of λ.

Repeating this procedure, we define for each ω_2 a unique element $q(\omega_2)$ with trajectories belonging to $D\Big([0,T];H^n(\mathbb{R}^d)\Big)$.

□

5.4.2 The backward SPDEs

Suppose f is Borel measurable, bounded and square integrable. For $0 \leq \theta \leq T$, we consider the following backward stochastic PDE,

$$\begin{cases} dr_t + \mathcal{A}r_t dt + h^\top r_t dZ_t + (\lambda - 1) r_t d(Y_t - t) = 0, \quad t \in [0, \theta], \\ r_\theta = f. \end{cases} \tag{5.4.25}$$

Notice, r depends on θ and $r_t = r_t^\theta$. To simplify, we write r_t. Equation (5.4.25) is equivalent to a forward equation. Define

$$\mathcal{F}_{s,t}^{Z,Y} = \sigma\{Z_u - Z_s, Y_u - Y_s, s \leq u \leq t\}, \tag{5.4.26}$$

then $\mathcal{F}_t^{Z,Y} = \mathcal{F}_{0,t}^{Z,Y}$.

Define processes $\{\tilde{Z}_t^\theta, t \in [0, \theta]\}$ and $\{\tilde{Y}_t^\theta, t \in [0, \theta]\}$ by, for $t \in [0, \theta]$,

$$\begin{aligned} \tilde{Z}_t^\theta &= Z_\theta - Z_{\theta - t}, \\ \tilde{Y}_t^\theta &= Y_{\theta-} - Y_{(\theta - t)-}, \text{ with } Y_{0-} := 0. \end{aligned}$$

Define filtration $\{\mathcal{F}_t^{\tilde{Z},\tilde{Y},\theta}, t \in [0, \theta]\}$, by, for $t \in [0, \theta]$,

$$\mathcal{F}_t^{\tilde{Z},\tilde{Y},\theta} := \mathcal{F}_{(\theta-t)-,\theta-}^{Z,Y}, \text{ with } \mathcal{F}_{0-,\theta-}^{Z,Y} := \mathcal{F}_{0,\theta-}^{Z,Y}.$$

Here

$$\mathcal{F}_{(\theta-t)-,\theta-}^{Z,Y} := \cap_{u \geq 0} \mathcal{F}_{(\theta-t)-u,\theta-}^{Z,Y}, \text{ where } \mathcal{F}_{(\theta-t)-u,\theta-}^{Z,Y} := \cup_{s \geq 0} \mathcal{F}_{(\theta-t)-u,\theta-s}^{Z,Y}.$$

Then $\tilde{Z}^\theta$ is a $\{\mathcal{F}_t^{\tilde{Z},\tilde{Y},\theta}\}$-standard Brownian motion, $\tilde{Y}^\theta$ is a $\{\mathcal{F}_t^{\tilde{Z},\tilde{Y},\theta}\}$-Poisson process with intensity 1, and $\tilde{Z}^\theta$, $\tilde{Y}^\theta$ are independent.

Define $\tilde{r}_t^\theta := r_{(\theta-t)-}$, we have Equation (5.4.25) is equivalent to the forward equation

$$\begin{cases} d\tilde{r}_t^\theta = \mathcal{A}\tilde{r}_t^\theta dt + [h\tilde{r}_t^\theta, d\tilde{Z}_t] + (\lambda - 1)\tilde{r}_{t-}^\theta d(\tilde{Y}_t - t), \quad t \in [0, \theta], \\ \tilde{r}_0^\theta = f. \end{cases} \tag{5.4.27}$$

Theorem 5.4 guarantees Equation (5.4.27) has a unique solution, and so we have the following existence and uniqueness results for the backward stochastic PDEs.

Theorem 5.6. *If Assumption 5.1 holds and $f \in H$, then Equation* (5.4.25) *has a unique solution $r = \{r_t\}_{t \in [0,\theta]}$ which is an $\{\mathcal{F}_{t,\theta}^{Z,Y}\}_{t \in [0,\theta]}$-adapted, V-valued stochastic process with càdlàg trajectories with respect to H-norm, which means that $\forall \omega \in \Omega$, $r(\omega) \in D([0, \theta]; H)$, and*

$$\mathbb{E}^0 \Big[\sup_{0 \leq t \leq \theta} \|r_t\|_H^2 + \int_0^\theta \|r_t\|_V^2 dt \Big] < \infty. \tag{5.4.28}$$

Proof. Equation (5.4.25) is equivalent to forward equation (5.4.27) which coefficients satisfy Assumption 5.1. Then, the desired result is obtained by Theorem 5.4. ☐

Remark 5.5. *Notice, the term 'backward stochastic equation' is different from the notion used by Karoui, Peng, and Quenez (1997). One can rewrite it as a forward equation but, one cannot rewrite the backward equation mentioned in Karoui, Peng, and Quenez (1997) as a forward one.*

The corresponding regularity result of Equation (5.4.25) is as follows.

Corollary 5.7. *Let $n \geq 1$ be any integer. Suppose, in addition to Assumption 5.1, that all coefficients a, b, h, and λ have bounded partial derivatives in x up to order n and that moreover*

$$p_0 \in H^n(\mathbb{R}^d).$$

Then, each trajectory of r, solution of Equation (5.4.25), belongs to $D\Big([0,\theta]; H^n(\mathbb{R}^d)\Big)$, a.s..

Proof. This follows immediately by applying Theorem 5.5 to Equation (5.4.27), which is equivalent to Equation (5.4.25). ☐

We want to show that r, the solution of Equation (5.4.25), is the conditional law of X. To begin, we define

$$\Lambda_{s,t} = \Big\{ \prod_{s<\tau_n\leq t} \lambda(X_{\tau_n-})\Big\} \exp\Big[\int_s^t h(X_u)^\top dZ_u - \frac{1}{2}\int_s^t \|h(X_u)\|^2 du - \int_s^t \Big(\lambda(X_u)-1\Big)du\Big], \quad s\in[0,t], \tag{5.4.29}$$

then $\Lambda_t = \Lambda_{0,t}$, where Λ is defined by Equation (2.4.5). For sake of simplicity, we define

$$\mathbb{E}^0_{t,x}(\cdot) := \mathbb{E}^0(\cdot|X_t = x).$$

The following theorem shows r, the solution of Equation (5.4.25), is the conditional law of X.

Theorem 5.8. *If Assumption 5.1 holds and f is Borel measurable, bounded and square integrable, then, the solution of Equation (5.4.25) satisfies, $\forall t\in[0,\theta]$, $x\in\mathbb{R}^d$,*

$$r_t(x) = \mathbb{E}^0_{t,x}\Big[f(X_\theta)\Lambda_{t,\theta}\Big|\mathcal{F}^{Z,Y}_{t,\theta}\Big], \ \ a.e.. \tag{5.4.30}$$

Here $\mathcal{F}^{Z,Y}_{t,\theta}$ is defined by Equation (5.4.26), and $\Lambda_{t,\theta}$ is defined by Equation (5.4.29). The proof of Theorem 5.8 is organized as follows:

- Prove Equation (5.4.30) in the case of regular (in x) coefficients a, b, h, λ and f. See Lemma 5.9, which will be proved later.
- Take the limit of both sides of Equation (5.4.30) when a^n, b^n, h^n, λ^n and f^n converge.

Lemma 5.9. *Suppose Assumption 5.1 holds. Additionally, suppose that b, σ, h and $\lambda - 1$ are compact supported and have continuous partial derivations of any order, together with $f \in \cap_{n=1}^\infty H^n$. Then Equality (5.4.30) holds $\forall(t,x)$ a.s.*

Remark 5.6. *For the proof of Theorem 5.8, we follow the idea of Pardoux (1979b) which is only for the case with continuous observation. The basic tool of Pardoux (1979b) is the convergence result for diffusion process from Stroock (1975).*

Proof of Theorem 5.8. Let us first suppose that f is continuous, with compact, support. Let b^n_i, a^n_{ij}, h^n_k, $\lambda^n - 1$, f^n be a sequence of smooth and compact supported functions, such that:

i) b_i^n, σ_{ij}^n, $\partial a_{ij}^n/\partial x_j$, h_k^n, λ^n, f^n are all uniformly bounded by a constant independent of n, and $\mathcal{A}^n$ satisfies Assumption iii) with α independent of n.

ii) $\sigma_{ij}^n \to \sigma_{ij}$ and $f^n \to f$ uniformly on each compact set of $\mathbb{R}^d$.

iii) $\partial a_{ij}^n/\partial x_j \to \partial a_{ij}/\partial x_j$, $b_i^n \to b_i$, $h_k^n \to h_k$ and $\lambda^n \to \lambda$ in measure on each compact set of $\mathbb{R}^d$. Note that we can not assume uniform convergence here, for h is only measurable function.

Denote by $r_t^n(x)$ and $\Lambda_{t,\theta}^n$ the corresponding objects associated with b^n, σ^n, h^n, Λ^n and f^n. And $\mathbb{P}_{tx}^{0,n}$ is the probability measure corresponding to the law of X^n, Y^n and Z^n. We define by $\mathbb{P}_{tx}^n$ by

$$\mathbb{P}_{tx}^n(A) = \int_A \Lambda_{0,\theta}^n \mathbb{P}_{tx}^{0,n}(d\omega). \tag{5.4.31}$$

By Lemma 5.9, we have

$$r_t^n(x) = \mathbb{E}_{tx}^{0,n}\left[f^n(X_\theta)\Lambda_{t,\theta}^n \middle| \mathcal{F}_{t,\theta}^{Z,Y}\right]. \tag{5.4.32}$$

Then, it is suffice to show that there exists a subsequence such that[1]

$$\Big(x \mapsto \mathbb{E}_{tx}^{0,n_k}\left[f^{n_k}(X_\theta)\Lambda_{t,\theta}^{n_k} \middle| \mathcal{F}_{t,\theta}^{Z,Y}\right]\Big) \rightharpoonup \Big(x \mapsto \mathbb{E}_{tx}^{0}\left[f(X_\theta)\Lambda_{t,\theta} \middle| \mathcal{F}_{t,\theta}^{Z,Y}\right]\Big), \tag{5.4.33}$$

$$r_t^{n_k} \rightharpoonup r_t, \tag{5.4.34}$$

in Hilbert space $L^2\big((\Omega, \mathcal{F}_{t,\theta}^{Z,Y}, \mathbb{P}^0); H\big)$ weakly.

We show Equation (5.4.33) first. For sake of simplicity, define

$$\varsigma^n(t,x) := \mathbb{E}_{tx}^{0,n}\left[f^n(X_\theta)\Lambda_{t,\theta}^n \middle| \mathcal{F}_{t,\theta}^{Z,Y}\right], \tag{5.4.35}$$

$$\varsigma(t,x) := \mathbb{E}_{tx}^{0}\left[f(X_\theta)\Lambda_{t,\theta} \middle| \mathcal{F}_{t,\theta}^{Z,Y}\right]. \tag{5.4.36}$$

It is equivalent to show that,

$$\mathbb{E}^0\Big[\big(w, \varsigma^n(t,\cdot)\big)\cdot\varphi\Big] \to \mathbb{E}^0\Big[\big(w, \varsigma(t,\cdot)\big)\cdot\varphi\Big], \tag{5.4.37}$$

where $w \in H$ is nonnegative, compact supported, twice continuous differentiable, $\int_{\mathbb{R}^d} w(x)dx = 1$, and $\varphi : \Omega \to \mathbb{R}$ is continuous, bounded and $\mathcal{F}_{t,\theta}^{Z,Y}$-measurable.

---

[1]Notation $\rightharpoonup$ is used to denote the weak convergence in a Hilbert space. Notice that the term 'weak' refers to the weak convergence of a sequence in a Hilbert space and not to the weak convergence of random variables of a probability distribution.

On the one hand, we have

$$
\begin{aligned}
\mathbb{E}^0\Big[\big(w,\varsigma^n(t,\cdot)\big)\cdot\varphi\Big] =&\mathbb{E}^{0,n}\Big[\big(w,\varsigma^n(t,\cdot)\big)\cdot\varphi\Big]\\
=&\mathbb{E}^{0,n}\Big[\varphi\int_{\mathbb{R}^d}\varsigma^n(t,x)w(x)dx\Big]\\
=&\mathbb{E}^{0,n}\Big[\int_{\mathbb{R}^d}\varphi\varsigma^n(t,x)w(x)dx\Big]\\
=&\mathbb{E}^{0,n}\Big\{\int_{\mathbb{R}^d}\mathbb{E}^{0,n}_{tx}[\varphi f^n(X_\theta)\Lambda^n_{t,\theta}|\mathcal{F}^{Z,Y}_{t,\theta}]w(x)dx\Big\}\\
=&\int_{\mathbb{R}^d}\mathbb{E}^{0,n}\Big\{\mathbb{E}^{0,n}_{tx}[\varphi f^n(X_\theta)\Lambda^n_{t,\theta}|\mathcal{F}^{Z,Y}_{t,\theta}]\Big\}w(x)dx\\
=&\int_{\mathbb{R}^d}\mathbb{E}^{0,n}_{tx}\Big\{\mathbb{E}^{0,n}_{tx}[\varphi f^n(X_\theta)\Lambda^n_{t,\theta}|\mathcal{F}^{Z,Y}_{t,\theta}]\Big\}w(x)dx\\
=&\int_{\mathbb{R}^d}\mathbb{E}^{0,n}_{tx}\Big[\varphi f^n(X_\theta)\Lambda^n_{t,\theta}\Big]w(x)dx\\
=&\int_{\mathbb{R}^d}\mathbb{E}^{n}_{tx}\Big[\varphi f^n(X_\theta)\Big]w(x)dx.
\end{aligned}
\tag{5.4.38}
$$

Here, the first equality follows from $\varphi\big(w,\varsigma^n(t,x)\big)$ is $\mathcal{F}^{Z,Y}_{t,\theta}$-measurable, and the restriction of $\mathbb{P}^{0,n}$ to $\mathcal{F}^{Z,Y}_{t,\theta}$ does not depend on n. The fourth equality follow from Equality (5.4.35) and the property that φ is $\mathcal{F}^{Z,Y}_{t,\theta}$-measurable. Applying Fubini's theorem to exchange the order of expectation and integral, we obtain the fifth equality. The sixth equality follows from $\mathbb{E}^{0,n}_{tx}[\varphi f^n(X_\theta)\Lambda^n_{t,\theta}|\mathcal{F}^{Z,Y}_{t,\theta}]$ is $\mathcal{F}^{Z,Y}_{t,\theta}$-measurable and independent of X under $\mathbb{P}^0$. The seventh equality follows from iterated conditional expectation. The last equality follows from Equation (5.4.31).

On the other hand, similarly, we obtain

$$
\begin{aligned}
\mathbb{E}^0\Big[\varphi\big(w,\varsigma(t,x)\big)\Big] =&\mathbb{E}^0\Big[\varphi\int_{\mathbb{R}}\varsigma(t,x)w(x)dx\Big]\\
=&\mathbb{E}^0\Big[\int_{\mathbb{R}}\varphi\varsigma(t,x)w(x)dx\Big]\\
=&\mathbb{E}^0\Big\{\int_{\mathbb{R}}\mathbb{E}^0_{tx}[\varphi f(X_\theta)\Lambda_{t,\theta}|\mathcal{F}^{Z,Y}_{t,\theta}]w(x)dx\Big\}\\
=&\int_{\mathbb{R}}\mathbb{E}^0\Big\{\mathbb{E}^0_{tx}[\varphi f(X_\theta)\Lambda_{t,\theta}|\mathcal{F}^{Z,Y}_{t,\theta}]w(x)\Big\}dx\\
=&\int_{\mathbb{R}}\mathbb{E}^0_{tx}\Big\{\mathbb{E}^0_{tx}[\varphi f(X_\theta)\Lambda_{t,\theta}|\mathcal{F}^{Z,Y}_{t,\theta}]w(x)\Big\}dx\\
=&\int_{\mathbb{R}}\mathbb{E}^0_{tx}\Big[\varphi f(X_\theta)\Lambda_{t,\theta}\Big]w(x)dx\\
=&\int_{\mathbb{R}}\mathbb{E}_{tx}\Big[\varphi f(X_\theta)\Big]w(x)dx.
\end{aligned}
\tag{5.4.39}
$$

Combining Equation (5.4.38) and (5.4.39), in order to show Equation (5.4.37), it remains to show that

$$
\begin{aligned}
&\Big|\int_{\mathbb{R}} \mathbb{E}^n_{tx}[\varphi f^n(X_\theta)]w(x)dx - \int_{\mathbb{R}} \mathbb{E}_{tx}[\varphi f(X_\theta)]w(x)dx\Big| \\
\le &\Big|\int_{\mathbb{R}} \mathbb{E}^n_{tx}[\varphi f^n(X_\theta)]w(x)dx - \int_{\mathbb{R}} \mathbb{E}^n_{tx}[\varphi f(X_\theta)]w(x)dx\Big| \\
&+\Big|\int_{\mathbb{R}} \mathbb{E}^n_{tx}[\varphi f(X_\theta)]w(x)dx - \int_{\mathbb{R}} \mathbb{E}_{tx}[\varphi f(X_\theta)]w(x)dx\Big|
\end{aligned}
$$

converges to 0. It is true since

$$
\begin{aligned}
\Big|\int_{\mathbb{R}} \mathbb{E}^n_{tx}[\varphi f^n(X_\theta)]w(x)dx - \int_{\mathbb{R}} \mathbb{E}^n_{tx}[\varphi f(X_\theta)]w(x)dx\Big| =&\Big|\int_{\mathbb{R}} \mathbb{E}^n_{tx}\Big\{\varphi[f^n(X_\theta) - f(X_\theta)]\Big\}w(x)dx\Big| \\
\le&\|f_n - f\|_\infty \cdot \|\varphi\|_\infty \int_{\mathbb{R}} w(x)dx \\
=&\|f_n - f\|_\infty \cdot \|\varphi\|_\infty \cdot 1 \\
\to&0,
\end{aligned}
$$

and

$$
\Big|\int_{\mathbb{R}} \mathbb{E}^n_{tx}[\varphi f(X_\theta)]w(x)dx - \int_{\mathbb{R}} \mathbb{E}_{tx}[\varphi f(X_\theta)]w(x)dx\Big| \to 0,
$$

where the first convergence follows from Assumption ii) and boundedness of φ while the second convergence follows from Jacod and Shiryaev (2003), Theorem IX.4.8, page 556.

Now we have shown that Equation (5.4.34) holds. It follows from Assumption i) and Remark 5.4 that $(r^n_t, n \ge 1)$ is bounded in $L^2\Big((\Omega, \mathcal{F}^{Z,Y}_{t,\theta}, \mathbb{P}^0); H\Big)$ and $r^n, n \ge 1$ is bounded in $L^2\Big((\Omega, \mathcal{F}^{Z,Y}_{t,\theta}, \mathbb{P}^0, \{\mathcal{F}^{Z,Y}_{s,\theta}\}_{t\le s\le\theta}) \times [t,\theta]; V\Big)$. Notice that every bounded sequence in a Hilbert space contains a weakly convergent subsequence. Then there exist a subsequence r^{n_k}, such that:

$$
r^{n_k}_t \rightharpoonup \eta \;\text{ in }\; L^2\Big((\Omega, \mathcal{F}^{Z,Y}_{t,\theta}, \mathbb{P}^0); H\Big) \;,\; \text{weakly,} \tag{5.4.40}
$$

$$
r^{n_k} \rightharpoonup \xi \;\text{ in }\; L^2\Big((\Omega, \mathcal{F}^{Z,Y}_{t,\theta}, \mathbb{P}^0, \{\mathcal{F}^{Z,Y}_{s,\theta}\}_{t\le s\le\theta}) \times [t,\theta]; V\Big) \;,\; \text{weakly.} \tag{5.4.41}
$$

It remains to show that $\xi = r$ and $\eta = r_t$. Let $\kappa \in C^1([t,\theta])$. Notice that $\kappa(\theta) = \kappa(t) + \int_t^\theta \kappa'(s)ds$, together with Equation (5.4.25), we obtain, noting the boundary condition $r^{n_k}_\theta = f^{n_k}$,

$$
\begin{aligned}
&\kappa(\theta)(f^{n_k}, w) + \int_t^\theta \kappa(s)\langle \mathcal{A}^{n_k} r^{n_k}_s, w\rangle ds + \int_t^\theta \kappa(s)(h^{n_k} r^{n_k}_s, w)dZ_s \\
&+ \int_t^\theta \kappa(s)\Big((\lambda^{n_k} - 1)r^{n_k}_s, w\Big)d(Y_s - s) - \int_t^\theta \kappa'(s)(r^{n_k}_s, w)ds \\
=&\kappa(t)(r^{n_k}_t, w).
\end{aligned} \tag{5.4.42}
$$

First, we have the following convergence weakly in $L^2\Big((\Omega, \mathcal{F}^{Z,Y}_{t,\theta}, \mathbb{P}^0); \mathbb{R}\Big)$, which will be shown later,

$$\int_t^\theta \kappa(s)\langle \mathcal{A}^{n_k} r_s^{n_k}, w\rangle ds \rightharpoonup \int_t^\theta \kappa(s)\langle \mathcal{A}\xi_s, w\rangle ds, \tag{5.4.43}$$

$$\int_t^\theta \kappa(s)(h^{n_k} r_s^{n_k}, w) dZ_s \rightharpoonup \int_t^\theta \kappa(s)(h\xi_s, w) dZ_s, \tag{5.4.44}$$

$$\int_t^\theta \kappa(s)\Big((\lambda^{n_k}-1) r_s^{n_k}, w\Big) d(Y_s - s) \rightharpoonup \int_t^\theta \kappa(s)\Big((\lambda-1\Big)\xi_s, w) d(Y_s - s), \tag{5.4.45}$$

$$\int_t^\theta \kappa'(s)(r_s^{n_k}, w) ds \rightharpoonup \int_t^\theta \kappa'(s)(\xi_s, w) ds, \tag{5.4.46}$$

$$\kappa(\theta)(f^{n_k}, w) \rightharpoonup \kappa(\theta)(f, w), \tag{5.4.47}$$

$$\kappa(t)(r_t^{n_k}, w) \rightharpoonup \kappa(t)(\eta, w). \tag{5.4.48}$$

Take the weak limit in $L^2\Big((\Omega, \mathcal{F}_{t,\theta}^{Z,Y}, \mathbb{P}^0); \mathbb{R}\Big)$ from both sides of Equation (5.4.42), yielding:

$$\begin{aligned}
&\kappa(\theta)(f, w) + \int_t^\theta \kappa(s)\langle \mathcal{A}\xi_s, w\rangle ds + \int_t^\theta \kappa(s)(h\xi_s, w) dZ_s \\
&+ \int_t^\theta \kappa(s)\Big((\lambda-1)\xi_s, w\Big) d(Y_s - s) + \int_t^\theta \kappa'(s)(\xi_s, w) ds \\
=&\kappa(t)(\eta, w).
\end{aligned}$$

It is then easy to conclude, from the uniqueness of the solution of Equation (5.4.25), that $\xi = r$ and $\eta = r_t$. It now remains to demonstrate the convergence in Equation (5.4.43), (5.4.44), (5.4.45), (5.4.46), (5.4.47), and (5.4.48).

First, we present the following convergence holds in H strongly:

$$\begin{aligned}
&\bar{a}^{n_k} w \to \bar{a} w, \ \ a^{n_k}\frac{\partial w}{\partial x_i} \to a\frac{\partial w}{\partial x_i}, \ \ i = 1, \dots, N, \\
&h^{n_k} w \to h w, \ \ (\lambda^{n_k} - 1) w \to (\lambda - 1) w.
\end{aligned} \tag{5.4.49}$$

Here, we only show $h^{n_k} w \to hw$. The other convergences in Equation (5.4.49) can be obtained similarly. By definition, w has a compact support, let it be $A \in \mathbb{R}^d$. By iii), we have $h_j^n \to h_j$ in measure on each compact set of $\mathbb{R}^d$. That is, set $A_\epsilon^n = \{x \in A : \|h^n(x) - h(x)\| \geq \epsilon\}$,

$$\lim_{n\to\infty} \mu(A_\epsilon^n) = 0,$$

where μ is the Lebesgue measure on $\mathbb{R}^d$. Then, for $\epsilon > 0$, $\exists K$, such that, $\forall k > K$,

$$\mu(A_\epsilon^{n_k}) < \epsilon.$$

And

$$\begin{aligned}
&\|h^{n_k} w - hw\|_H^2 \\
=&\int_A \left\|\Big(h^{n_k}(x) - h(x)\Big) w(x)\right\|^2 dx \\
=&\int_{A\setminus A_\epsilon^{n_k}} \left\|\Big(h^{n_k}(x) - h(x)\Big) w(x)\right\|^2 dx + \int_{A_\epsilon^{n_k}} \left\|\Big(h^{n_k}(x) - h(x)\Big) w(x)\right\|^2 dx \\
\leq&\int_{A\setminus A_\epsilon^{n_k}} \|h^{n_k}(x) - h(x)\|^2 \|w(x)\|^2 dx + \Big(\|h^{n_k}\|_\infty + \|h\|_\infty\Big)^2 \|w\|_\infty^2 \int_{A_\epsilon^{n_k}} dx \\
\leq&\epsilon^2 \|w\|_\infty^2 \mu(A) + \epsilon\Big(\|h^{n_k}\|_\infty + \|h\|_\infty\Big)^2 \|w\|_\infty^2,
\end{aligned}$$

the last inequality follows from the uniform boundedness of h^{n_k}, boundedness of h and w. Therefore, we have that

$$\|h^{n_k}w - hw\|_H \to 0, \text{ that is } h^{n_k}w \to hw.$$

Now we show Equation (5.4.43). It is equivalent to show that, $\forall \varphi \in L^2\Big((\Omega, \mathcal{F}^{Z,Y}_{t,\theta}, \mathbb{P}^0); \mathbb{R}\Big)$ and bounded,

$$\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle \mathcal{A}^{n_k} r^{n_k}_s, w\rangle ds\Big] \to \mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle \mathcal{A} r_s, w\rangle ds\Big].$$

We have

$$\begin{aligned}
&\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle \mathcal{A}^{n_k} r^{n_k}_s, w\rangle ds\Big]\\
=&\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\big\langle (\mathcal{A}^{n_k} - \mathcal{A} + \mathcal{A}) r^{n_k}_s, w\big\rangle ds\Big]\\
=&\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\big\langle (\mathcal{A}^{n_k} - \mathcal{A}) r^{n_k}_s, w\big\rangle ds\Big] + \mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle \mathcal{A} r^{n_k}_s, w\rangle ds\Big]\\
\to&\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle \mathcal{A}\xi_s, w\rangle ds\Big].
\end{aligned}$$

The last convergence follows from, by Equation (5.4.41),

$$\begin{aligned}
&\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle \mathcal{A} r^{n_k}_s, w\rangle ds\Big] - \mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle \mathcal{A}\xi_s, w\rangle ds\Big]\\
=&\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle \mathcal{A} r^{n_k}_s - \mathcal{A}\xi_s, w\rangle ds\Big]\\
=&\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle \mathcal{A}(r^{n_k}_s - \xi_s), w\rangle ds\Big]\\
=&\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)\langle r^{n_k}_s - \xi_s, \mathcal{A}^* w\rangle ds\Big]\\
\to&0
\end{aligned}$$

and meanwhile, for sake of simplicity, consider for $d = 1$, by Equation (5.1.4),

$$\begin{aligned}
&\Big|\mathbb{E}^0\big[\varphi \int_t^\theta \kappa(s)\big\langle (\mathcal{A}^{n_k} - \mathcal{A}) r^{n_k}_s, w\big\rangle ds\big]\Big|\\
\le&\mathbb{E}^0\Big[|\varphi| \int_t^\theta |\kappa(s)| \cdot \|r^{n_k}_s\|_V \cdot \Big(\|(a^{n_k} - a)w'\|_H + \|(\bar{a}^{n_k} - \bar{a})w\|_H\Big) ds\Big]\\
\le&\|\varphi\|_\infty \|\kappa\|_\infty \Big(\|(a^{n_k} - a)w'\|_H + \|(\bar{a}^{n_k} - \bar{a})w\|_H\Big)\mathbb{E}^0\Big[\int_t^\theta \|r^{n_k}(s)\|_V ds\Big]\\
\le&\|\varphi\|_\infty \|\kappa\|_\infty \Big(\|a^{n_k} - a\|_\infty \|w'\|_H + \|\bar{a}^{n_k} - \bar{a}\|_\infty \|w\|_H\Big)(\theta - t)\Big(|r^{n_k}|_{\theta - t} + 1\Big)\\
\to&0.
\end{aligned}$$

Here, the last convergence follows boundedness assumptions of φ, κ, and Equation (5.4.49). The high dimensional case is handled exactly in the same way. Therefore, we obtain Equation (5.4.43). Now we show the convergence (5.4.44). Define

$$\tilde{\varphi}_s = \mathbb{E}^0(\varphi | \mathcal{F}^{Z,Y}_{s,\theta}), \quad s \in [0, \theta].$$

Then $\tilde{\varphi}$ is a martingale. Due to the independence of Z and Y, we have a martingale representation of the $\mathbb{P}^0$-martingale $\tilde{\varphi}$ using the martingale representation theory, see Jacod and Shiryaev (2003), Theorem III.4.34, page 189,

$$\tilde{\varphi}_s = \mathbb{E}^0(\varphi) + \int_s^\theta \alpha_\nu dZ_\nu + \int_s^\theta \gamma_\nu d(Y_\nu - \nu), \quad s \in [0,\theta]. \tag{5.4.50}$$

Notice that φ is bounded, Protter (2005),Corollary 3, page 73, implies that coefficients $\{\alpha_\nu\}_{\nu\in[0,\theta]}$ and $\{\gamma_\nu\}_{\nu\in[0,\theta]}$ are square integrable. That is

$$\mathbb{E}^0\Big(\int_0^\theta \alpha_\nu^2 d\nu\Big) < \infty, \quad \mathbb{E}^0\Big(\int_0^\theta \gamma_\nu^2 d\nu\Big) < \infty. \tag{5.4.51}$$

Then, we get

$$\begin{aligned}
&\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)(h^{n_k} r_s^{n_k}, w) dZ_s\Big] \\
=&\mathbb{E}^0\Big\{\mathbb{E}^0\Big[\varphi \int_t^\theta \kappa(s)(h^{n_k} r_s^{n_k}, w) dZ_s | \mathcal{F}_{t\theta}^{Z,Y}\Big]\Big\} \\
=&\mathbb{E}^0\Big[\tilde{\varphi}_t \int_t^\theta \kappa(s)(h^{n_k} r_s^{n_k}, w) dZ_s\Big] \\
=&\mathbb{E}^0\Big\{\Big[\mathbb{E}^0(\varphi) + \int_t^\theta \alpha_\nu dZ_\nu + \int_t^\theta \gamma_\nu d(Y_\nu - \nu)\Big] \int_t^\theta \kappa(s)(h^{n_k} r_s^{n_k}, w) dZ_s\Big\} \\
=&\mathbb{E}^0\Big\{\int_t^\theta \alpha_\nu dZ_\nu \int_t^\theta \kappa(s)(h^{n_k} r_s^{n_k}, w) dZ_s\Big\} \\
=&\mathbb{E}^0\Big\{\int_t^\theta \alpha_s \kappa(s)(h^{n_k} r_s^{n_k}, w) ds\Big\} \\
\to&\mathbb{E}^0[\int_t^\theta \alpha_s \kappa(s)(h\xi_s, w) ds].
\end{aligned}$$

The first equality follows from iterated conditional expectation. The third equality follows from Equation (5.4.50). The fourth equality follows from independence of Z and Y. The last convergence is obtained as follows

$$\begin{aligned}
&\mathbb{E}^0\Big[\int_t^\theta \alpha_s \kappa(s)(h^{n_k} r_s^{n_k}, w) ds\Big] \\
=&\mathbb{E}^0\Big[\int_t^\theta \alpha_s \kappa(s)(r_s^{n_k}, h^{n_k} w) ds\Big] \\
=&\mathbb{E}^0\Big[\int_t^\theta \alpha_s \kappa(s)\Big(r_s^{n_k}, (h^{n_k} - h)w + hw\Big) ds\Big] \\
=&\mathbb{E}^0\Big[\int_t^\theta \alpha_s \kappa(s)\Big(r_s^{n_k}, (h^{n_k} - h)w\Big) ds] + \mathbb{E}^0[\int_t^\theta \alpha_s \kappa(s)(r_s^{n_k}, hw) ds\Big] \\
\to&\mathbb{E}^0\Big[\int_t^\theta \alpha_s \kappa(s)(h\xi_s, w) ds\Big],
\end{aligned}$$

where the last convergence follows from Equation (5.4.41) and

$$\begin{aligned}
&\mathbb{E}^0\Big[\int_t^\theta \alpha_s\kappa(s)\Big(r_s^{n_k},(h^{n_k}-h)w\Big)ds\Big]\\
\le&\mathbb{E}^0\Big[\int_t^\theta |\alpha_s|\cdot|\kappa(s)|\cdot\Big|\Big(r_s^{n_k},(h^{n_k}-h)w\Big)\Big|ds\Big]\\
\le&\mathbb{E}^0\Big[\int_t^\theta |\alpha_s|\cdot|\kappa(s)|\cdot\|r_s^{n_k}\|_H\|(h^{n_k}-h)w\|_H ds\Big]\\
\le&\|(h^{n_k}-h)w\|_H\|\kappa\|_\infty\mathbb{E}^0\int_t^\theta |\alpha_s|\|r_s^{n_k}\|_H ds\\
\le&\|(h^{n_k}-h)w\|_H\|\kappa\|_\infty\Big\{\mathbb{E}^0\int_t^\theta \alpha_s^2 ds\Big\}^{1/2}\cdot\Big\{\mathbb{E}^0\int_t^\theta \|r_s^{n_k}\|_H^2 ds\Big\}^{1/2}\\
\le&\|(h^{n_k}-h)w\|_H\|\kappa\|_\infty\Big\{\mathbb{E}^0\int_t^\theta \alpha_s^2 ds\Big\}^{1/2}(\theta-t)|r^{n_k}|_{\theta-t}\\
\to&0.
\end{aligned}$$

Here, the convergence follows from Equation (5.4.49), bounded assumptions of κ, and that α is square integrable, by Equation (5.4.51).

Meanwhile, we have

$$\mathbb{E}^0\Big[\varphi\int_t^\theta \kappa(s)(hr_s,w)dZ_s\Big]=\mathbb{E}^0\Big[\int_t^\theta \alpha_s\kappa(s)(h\xi_s,w)ds\Big].$$

Equation (5.4.44) is obtained.

Similarly, we obtain Equation (5.4.45) by

$$\begin{aligned}
&\mathbb{E}^0\Big\{\varphi\int_t^\theta \kappa(s)\Big((\lambda^{n_k}-1)r_s^{n_k},w\Big)d(Y_s-s)\Big\}\\
=&\mathbb{E}^0\Big\{\mathbb{E}^0\Big[\varphi\int_t^\theta \kappa(s)\Big((\lambda^{n_k}-1)r_s^{n_k},w\Big)d(Y_s-s)|\mathcal{F}_{t\theta}^{Z,Y}\Big]\Big\}\\
=&\mathbb{E}^0\Big\{\tilde{\varphi}_t\int_t^\theta \kappa(s)\Big((\lambda^{n_k}-1)r_s^{n_k},w\Big)d(Y_s-s)\Big\}\\
=&\mathbb{E}^0\Big\{\Big[\mathbb{E}^0(\varphi)+\int_t^\theta \alpha_\nu dZ_\nu+\int_t^\theta \gamma_\nu d(Y_\nu-\nu)\Big]\int_t^\theta \kappa(s)\Big((\lambda^{n_k}-1)r_s^{n_k},w\Big)d(Y_s-s)\Big\}\\
=&\mathbb{E}^0\Big\{\int_t^\theta \gamma_\nu d(Y_\nu-\nu)\int_t^\theta \kappa(s)\Big((\lambda^{n_k}-1)r_s^{n_k},w\Big)d(Y_s-s)\Big\}\\
=&\mathbb{E}^0\Big\{\int_t^\theta \gamma_s\kappa(s)\Big((\lambda^{n_k}-1)r_s^{n_k},w\Big)ds\Big\}\\
\to&\mathbb{E}^0\Big[\int_t^\theta \gamma_s\kappa(s)\Big((\lambda-1)\xi_s,w\Big)ds\Big]
\end{aligned}$$

and

$$\mathbb{E}^0\Big\{\varphi\int_t^\theta \kappa(s)\Big((\lambda^{n_k}-1)r_s^{n_k},w\Big)d(Y_s-s)\Big\}=\mathbb{E}^0\Big[\int_t^\theta \gamma_s\kappa(s)\Big((\lambda-1)\xi_s,w\Big)ds\Big].$$

Likewise, we get Equation (5.4.46), (5.4.47) and (5.4.48).

□

Now, we propose to prove Lemma 5.9. In order to simplify the notations, let $t = 0$. Define $\phi_s = r_s(X_s)\Lambda_s$. Then, given $X_0 = x$,

$$\begin{aligned}\phi_0 &= r_0(X_0)\Lambda_0 = r_t(x),\\ \phi_\theta &= r_\theta(X_\theta)\Lambda_\theta = f(X_\theta)\Lambda_\theta.\end{aligned}$$

Notice that

$$\begin{aligned}\mathbb{E}^0(\phi_0|\mathcal{F}_\theta^{Z,Y}, X_0 = x) &= r_t(x),\\ \mathbb{E}^0(\phi_\theta|\mathcal{F}_\theta^{Z,Y}, X_0 = x) &= r_t(x),\end{aligned}$$

the last equality follows from Lemma 5.9. Therefore, in order to prove Lemma 5.9, one approach is to show that

$$\mathbb{E}^0_{0,x}\Big[\phi_\theta - \phi_0\Big|\mathcal{F}_\theta^{Z,Y}\Big] = \mathbb{E}^0_{0,x}\Big[\int_0^\theta d\phi_s\Big|\mathcal{F}_\theta^{Z,Y}\Big] = 0.$$

It is suffice to show that the conditional expectation of the increment of ϕ with respect to $\mathcal{F}_\theta^{Z,Y}$ is zero, by Itô's formula, Equation (5.4.25) and (5.4.29). However, one cannot differentiate $\phi_s = r_s(X_s)\Lambda_s$ by Itô's formula. The reason being, Equation (5.4.25) is a backward SPDE while Equation (5.4.29) is a forward one. ϕ's differential could involve terms which do not make sense. Instead, we use time discretisation approach as that adopted in Pardoux (1979b), page 154.

Proof of Lemma 5.9. In the proof, for convenience, we define $r(t,x) := r_t(x)$. In the following, we will apply stochastic Fubini's theorem and central limit theorem. To do so, we need the following boundedness of r and Λ.

Lemma 5.10. *Under the assumption of Lemma 5.9, for each compact set A of $\mathbb{R}^d$, we have*

$$P^0\Big\{\exists c(\omega) \in \mathbb{R}^+, s.t. \sup_{0\le t\le\theta, x\in A}\{|r(t,x,\omega)|, |r_x(t,x,\omega)|, |r_{xx}(t,x,\omega)|\} \le c(\omega)\Big\} = 1.$$

Proof. It follows from the hypotheses and Lemma 5.7 that each trajectory of r belongs to $\cap_n D\Big(0,T;H^n(\mathbb{R}^d)\Big)$. By the properties of sobolev space, see Folland (1999), 9.18 Corollary, page 304, we have $\forall t$, $r(t,\cdot) \in C^\infty(\mathbb{R}^d)$. Therefore, $\forall t$, $r(t,\cdot)$ is uniformly bounded on each compact set of $\mathbb{R}^d$. By Theorem 5.6, every trajectory of r is right-continuous having left limits, applying Lemma 9.17, Peszat and Zabczyk (2007), page 154, we obtain the desired result. ☐

Lemma 5.11. *Under the assumption of Lemma 5.9, we have*

$$\sup_{0\le t\le T} \mathbb{E}^0[\Lambda_t^2|X_0 = x] < \infty, \tag{5.4.52}$$

$$\sup_{0\le t\le T} \mathbb{E}^0[\Lambda_t^4|X_0 = x] < \infty. \tag{5.4.53}$$

Proof. By the definition of Λ, see Equation (2.4.5) and (2.4.1),

$$\begin{aligned}\Lambda_t^2 = &\prod_{\tau_n\le t}[\lambda(X_{\tau_n-})]^2 \cdot \exp\Big[2\int_0^t h(X_u)^\top dZ_u - \int_0^t \|h(X_u)\|^2du - 2\int_0^t \Big(\lambda(X_u) - 1\Big)du\Big]\\ \le&\|\lambda\|_\infty^{2Y_t} \cdot \exp\Big[2\int_0^t h(X_u)^\top dZ_u\Big] \cdot e^{2t} \cdot \exp\Big[-\int_0^t \|h(X_u)\|^2du - 2\int_0^t \lambda(X_u)du\Big]\\ \le&e^{2t}\|\lambda\|_\infty^{2Y_t} \exp\Big[2\int_0^t h(X_u)^\top dZ_u\Big].\end{aligned}$$

Notice, Y is a Poisson process with intensity 1 and is independent with X, we have

$$\mathbb{E}^0[\Lambda_t^2|X_0=x] \leq e^{2t}\mathbb{E}^0\Big[\|\lambda\|_\infty^{2Y_t}\Big]\mathbb{E}^0\Big[\exp\Big(2\int_0^t h(X_u)^\top dZ_u\Big)\Big|X_0=x\Big].$$

On one side,

$$\begin{aligned}\mathbb{E}^0\Big[\|\lambda\|_\infty^{2Y_t}\Big] &= \sum_{i=0}^\infty \|\lambda\|_\infty^{2i}\frac{t^i}{i!}e^{-t} \\ &= e^{t(\|\lambda\|_\infty^2-1)}\sum_{i=0}^\infty \frac{(t\|\lambda\|_\infty^2)^i}{i!}e^{-t\|\lambda\|_\infty^2} \\ &= e^{t(\|\lambda\|_\infty^2-1)}.\end{aligned}$$

On the other side,

$$\begin{aligned}&\mathbb{E}^0\Big[\exp\Big(2\int_0^t h(X_u)^\top dZ_u\Big)\Big|X_0=x]\\ =&\mathbb{E}^0\Big[\mathbb{E}^0\{\exp(2\int_0^t h(X_u)^\top dZ_u)|\mathcal{F}_t^X, X_0=x\}\Big|X_0=x\Big]\\ =&\mathbb{E}^0\Big[\exp\Big(2\int_0^t \|h(X_u)\|^2 du\Big)\Big|X_0=x\Big]\\ \leq& e^{2t\|h\|_\infty^2}.\end{aligned}$$

Combining the results, we obtain

$$\sup_{0\leq t\leq T}\mathbb{E}^0\Big[\Lambda_t^2\Big|X_0=x\Big] \leq e^{2t}\cdot e^{t(\|\lambda\|_\infty^2-1)}\cdot e^{2t\|h\|_\infty^2} < \infty.$$

Similarly, we obtain Equation (5.4.53).

□

In order to simplify the notations, let us prove Equation (5.4.30) for t=0, write $\mathbb{E}_x^0(\cdot)$ for $\mathbb{E}_{0,x}^0(\cdot|\mathcal{F}_\theta^{Z,Y})$. We will suppose that $d=1$. The more general case is handled in exactly the same way.

For reasons previously mentioned, we cannot differentiate $\Lambda_t r(t,X_t)$ with respect to t, instead, we consider the following time discretisation. For each ω_2, suppose that $Y_\theta = N$ and then $0 < \tau_1(\omega_2) < \ldots < \tau_N(\omega_2) \leq \theta$. For a integer m, let $0 = s_0 < s_1 < s_2 < \ldots s_m = \theta$ be a mesh with $s_{i+1} - s_i = \theta/m =: k$. Now we have less than $N+m+1$ time points. Assume we have $n+1$ time points and obviously, $n < M+N$. Re-index the positions for the time points as $0 = t_0 < t_1 < t_2 \ldots < t_n = \theta$. Finally, the interval $[0,\theta]$ is partitioned into subinterval of length less than or equal to k with $0 = t_0 < t_1 < t_2 \ldots < t_n = \theta$. Then, we have the following decomposition. Notice that $f(\cdot) = r(\theta, \cdot)$ and $\Lambda_0 = 1$,

$$\begin{aligned}
\mathbb{E}_x^0\Big[\Lambda_\theta f(X_\theta)\Big] - r(0,x) =& \mathbb{E}_x^0\Big[\Lambda_\theta r(\theta, X_\theta)\Big] - \mathbb{E}_x^0\Big[\Lambda(0) r(0,x)\Big] \\
=& \sum_{i=1}^{n-1} \mathbb{E}_x^0\Big[\Lambda_{t_{i+1}} r(t_{i+1}, X_{t_{i+1}}) - \Lambda_{t_i} r(t_i, X_{t_i})\Big] \\
=& \sum_{i=1}^{n-1} \mathbb{E}_x^0\Big[\Lambda_{t_{i+1}} r(t_{i+1}, X_{t_{i+1}}) - \Lambda_{t_{i+1}-} r(t_{i+1}-, X_{t_{i+1}-}) \\
& + \Lambda_{t_{i+1}-} r(t_{i+1}-, X_{t_{i+1}-}) - \Lambda_{t_i} r(t_i, X_{t_i})\Big] \\
=& \sum_{i=1}^{n-1} \Big\{\mathbb{E}_x^0\Big[\Lambda_{t_{i+1}} r(t_{i+1}, X_{t_{i+1}}) - \Lambda_{t_{i+1}-} r(t_{i+1}-, X_{t_{i+1}-})\Big] \\
& + \mathbb{E}_x^0\Big[\Lambda_{t_{i+1}-} r(t_{i+1}-, X_{t_{i+1}-}) - \Lambda_{t_i} r(t_i, X_{t_i})\Big]\Big\}.
\end{aligned}$$

In the following, we will show that, for $0 < t \leq \theta$,

$$\Lambda_t r(t, X_t) - \Lambda_{t-} r(t-, X_{t-}) = 0. \tag{5.4.54}$$

Then consequently,

$$\mathbb{E}_x^0\Big[f(X_\theta)\Lambda_\theta\Big] - r(0,x) = \sum_{i=1}^{n-1} \triangle_i, \tag{5.4.55}$$

where, for $i = 1, 2, \ldots, n-1$,

$$\mathbb{E}_x^0\Big[\Lambda_{t_{i+1}-} r(t_{i+1}-, X_{t_{i+1}-}) - \Lambda_{t_i} r(t_i, X_{t_i})\Big] := \triangle_i.$$

In fact, when t is not the time of jump, then Equation (5.4.54) is obtain by continuity. Otherwise, by Equation (5.4.29),

$$\Lambda_t - \Lambda_{t-} = [\lambda(X_{t-}) - 1]\Lambda_{t-}.$$

That is

$$\Lambda_t = \lambda(X_{t-})\Lambda_{t-}. \tag{5.4.56}$$

Similarly, apply Equation (5.4.25),

$$r(t-, X_{t-}) = \lambda(X_{t-}) r(t, X_{t-}) = \lambda(X_{t-}) r(t, X_t), \tag{5.4.57}$$

where the last equality follows from the continuity of X, that is $X_t = X_{t-}$. Combining Equation (5.4.56) and (5.4.57) together,

$$\Lambda_t r(t, X_t) - \Lambda_{t-} r(t-, X_{t-}) = \lambda(X_{t-})\Lambda_{t-} r(t, X_t) - \Lambda_{t-}\lambda(X_{t-}) r(t, X_t) = 0.$$

By Equation (5.4.55), in order to prove Equation (5.4.30), it suffices to show that, as $m \to \infty$,

$$\sum_{i=1}^{n-1} \triangle_i \to 0 \;\; a.s..$$

As previously mentioned, we can not apply Itô formula to $\triangle_i$, by Equation (2.4.12) and (5.4.25). Instead we have the following decomposition,

$$\begin{aligned}\triangle_i =&\mathbb{E}^0_x\Big[\Lambda_{t_{i+1}-}r(t_{i+1}-,X_{t_{i+1}-})-\Lambda_{t_i}r(t_{i+1}-,X_{t_i})\Big]\\&+\mathbb{E}^0_x\Big[\Lambda_{t_i}r(t_{i+1}-,X_{t_i})-\Lambda_{t_i}r(t_i,X_{t_i})\Big],\end{aligned}\tag{5.4.58}$$

such that one part composes the increment cased by Λ_t and X_t, whose dynamics is described by forward equation, while the other part composes the increment cased by r with respect to t, whose dynamics is described by backward equation. On the one hand, we express the second term of $\triangle_i$ using Equation (5.4.25) at $x = X_t$, which make sense because of the regularity of the solution. Notice that there is no jump between the time interval $[t_i, t_{i+1})$,

$$\begin{aligned}&\Lambda_{t_i}r(t_{i+1}-,X_{t_i})-\Lambda_{t_i}r(t_i,X_{t_i})\\=&\int_{t_i}^{t_{i+1}}\Lambda_{t_i}\Big[-\mathcal{A}r(s,X_{t_i})ds-h(X_{t_i})r(s,X_{t_i})dZ_s+(\lambda(X_{t_i})-1)r(s,X_{t_i})ds\Big].\end{aligned}\tag{5.4.59}$$

On the other hand, we express the first term in $\triangle_i$ by means of Itô formula, by Equation (2.4.12) (2.2.1), applied to $\Lambda_s r(t_{i+1}, X_s)$.

$$\begin{aligned}&\Lambda_{t_{i+1}-}r(t_{i+1}-,X_{t_{i+1}-})-\Lambda_{t_i}r(t_{i+1}-,X_{t_i})\\=&\int_{t_i}^{t_{i+1}}\Lambda_s r(t_{i+1}-,X_s)h(X_s)dZ_s-\int_{t_i}^{t_{i+1}}\Lambda_s r(t_{i+1}-,X_s)\Big(\lambda(X_{s-})-1\Big)ds\\&+\int_{t_i}^{t_{i+1}}\Lambda_s\mathcal{A}r(t_{i+1}-,X_s)ds+\int_{t_i}^{t_{i+1}}\Lambda_s r_x(t_{i+1}-,X_s)\sigma(X_s)dV_s.\end{aligned}\tag{5.4.60}$$

Now we show that

$$\mathbb{E}^0_x\Big\{\int_{t_i}^{t_{i+1}}\Lambda_s r_x(t_{i+1}-,X_s)\sigma(X_s)dV_s\Big\}=0.\tag{5.4.61}$$

By assumption σ is smooth with compact support. Together with the boundedness of r, by Lemma 5.10, $r_x\sigma$ is bounded, then we have

$$\mathbb{E}^0_x\Big\{\int_{t_i}^{t_{i+1}}\Big[\Lambda_s r_x(t_{i+1}-,X_s)\sigma(X_s)\Big]^2ds\Big\}\le\|r_x\sigma\|^2_\infty\mathbb{E}^0_x\Big\{\int_{t_i}^{t_{i+1}}\Lambda_s^2ds\Big\}.\tag{5.4.62}$$

Notice that, by Lemma 5.11,

$$\begin{aligned}\mathbb{E}^0\Big[\int_{t_i}^{t_{i+1}}\Lambda_s^2ds|X_0=x\Big]=&\int_{t_i}^{t_{i+1}}\mathbb{E}^0[\Lambda_s^2|X_0=x]ds\\\le&(t_{i+1}-t_i)\sup_{0\le s\le T}\mathbb{E}^0[\Lambda_s^2|X_0=x]<\infty.\end{aligned}$$

Therefore, the left hand side of Equation (5.4.62) is finite *a.s.*. Together with V_t, Y_t, Z_t are $\mathbb{P}^0$-independent, we have Equation (5.4.61).

Combining with Equation (5.4.58), (5.4.59), (5.4.60) and (5.4.61), we have

$$\begin{aligned}
\triangle_i =&\mathbb{E}_x^0\Big\{\int_{t_i}^{t_{i+1}}\Lambda_s r(t_{i+1}-,X_s)h(X_s)dZ_s\Big\}\\
&-\mathbb{E}_x^0\Big\{\int_{t_i}^{t_{i+1}}\Lambda_s r(t_{i+1}-,X_s)[\lambda(X_{s-})-1)]ds\Big\}\\
&+\mathbb{E}_x^0\Big\{\int_{t_i}^{t_{i+1}}\Lambda_s \mathcal{A}r(t_{i+1}-,X_s)ds\Big\}\\
+&\mathbb{E}_x^0\Big\{\int_{t_i}^{t_{i+1}}\Lambda_{t_i}\Big[-\mathcal{A}r(s,X_{t_i})ds-h(X_{t_i})r(s,X_{t_i})dZ_s+(\lambda(X_{t_i})-1)r(s,X_{t_i})ds\Big]\Big\}\\
=&\mathbb{E}_x^0\Big\{\int_{t_i}^{t_{i+1}}\Lambda_s r(t_{i+1}-,X_s)h(X_s)dZ_s-\int_{t_i}^{t_{i+1}}\Lambda_{t_i}h(X_{t_i})r(s,X_{t_i})dZ_s\Big\}\\
&+\mathbb{E}_x^0\Big\{-\int_{t_i}^{t_{i+1}}\Lambda_s r(t_{i+1}-,X_s)[\lambda(X_{s-})-1)]ds+\int_{t_i}^{t_{i+1}}\Lambda_{t_i}[\lambda(X_{t_i})-1]r(s,X_{t_i})ds\Big\}\\
&+\mathbb{E}_x^0\Big\{\int_{t_i}^{t_{i+1}}\Lambda_s \mathcal{A}r(t_{i+1}-,X_s)ds-\int_{t_i}^{t_{i+1}}\Lambda_{t_i}\mathcal{A}r(s,X_{t_i})ds\Big\}. \qquad (5.4.63)
\end{aligned}$$

Λ has right continuous path with finite discontinuous time points, it is bounded for each trajectory. By assumption, b, σ, $\lambda-1$ is smooth and compact supported. Together with boundedness of r, by Lebesgue dominated convergence theorem, see Folland (1999), Theorem 2.24, page 54, we have that

$$\begin{aligned}
&\lim_{m\to\infty}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\Lambda_s r(t_{i+1}-,X_s)\Big[\lambda(X_{s-})-1\Big]ds\\
=&\lim_{m\to\infty}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\Lambda_{t_i}\Big[\lambda(X_{t_i})-1\Big]r(s,X_{t_i})ds\\
=&\int_0^\theta \Lambda_s\Big[\lambda(X_s)-1\Big]r(s,X_s)ds \qquad (5.4.64)
\end{aligned}$$

and

$$\begin{aligned}
\lim_{m\to\infty}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\Lambda_s\mathcal{A}r(t_{i+1},X_s)ds=&\lim_{m\to\infty}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\Lambda_{t_i}\mathcal{A}r(t_s,X_{t_i})ds\\
=&\int_0^\theta\Lambda_s\mathcal{A}r(s,X_s)ds. \qquad (5.4.65)
\end{aligned}$$

It is not possible to take the limit in the same way in the stochastic integrals, because the limit should not make sense.

$$\int_{t_i}^{t_{i+1}}\Lambda_s r(t_{i+1}-,X_s)h(X_s)dZ_s-\int_{t_i}^{t_{i+1}}\Lambda_{t_i}h(X_{t_i})r(t_s,X_{t_i})dZ_s=:\alpha_i+\beta_i, \qquad (5.4.66)$$

where

$$\begin{aligned}
\alpha_i=&\int_{t_i}^{t_{i+1}}\Big[\Lambda_s r(t_{i+1}-,X_s)h(X_s)-\Lambda_{t_i}h(X_{t_i})r(t_{i+1}-,X_{t_i})\Big]dZ_s,\\
\beta_i=&\int_{t_i}^{t_{i+1}}\Lambda_{t_i}\Big[h(X_{t_i})r(t_{i+1}-,X_{t_i})-h(X_{t_i})r(s,X_{t_i})\Big]dZ_s.
\end{aligned}$$

Combining the obtained result, Equation (5.4.63), (5.4.64), (5.4.65) and (5.4.66), it remains to show that $\mathbb{E}_x^0 \sum_{i=1}^{n-1} \alpha_i$ and $\mathbb{E}_x^0 \sum_{i=0}^{n-1} \beta_i$ converge to 0. The main tool for obtaining the convergence is the central limit theorem for stochastic integrals, see Revuz and Yor (1999), page 152. In order to show that the sequence of stochastic integrals converges to 0, by the central limit theorem, it is suffice to show that it converges in distribution to a normal random variable with variance 0. Let us first consider $\mathbb{E}_x^0 \sum_{i=1}^{n-1} \alpha_i$. It follows from Itô's formula, by Equation (2.2.1) and (2.4.12),

$$\begin{aligned}
&\Lambda_s r(t_{i+1}-, X_s)h(X_s) - \Lambda_{t_i} h(X_{t_i}) r(t_{i+1}, X_{t_i}) \\
=& \int_{t_i}^s \Lambda_u \mathcal{A}[h(\cdot)r(t_{i+1},\cdot)](X_u)du + \int_{t_i}^s \Lambda_u \frac{\partial[h(x)r(t_{i+1},x)]}{\partial x}\Big|_{x=X_u} \sigma(X_u)dV_u \\
&+ \int_{t_i}^s \Lambda_u h^2(X_u) r(t_{i+1}, X_u)dZ_u - \int_{t_i}^s \Lambda_u(\lambda(X_u)-1)h(X_u)r(t_{i+1},X_u)du \\
=& \int_{t_i}^s \Lambda_u \Big[\mathcal{A}[h(\cdot)r(t_{i+1},\cdot)](X_u) - (\lambda(X_u)-1)h(X_u)r(t_{i+1},X_u)\Big]du \\
&+ \int_{t_i}^s \Lambda_u h^2(X_u) r(t_{i+1}, X_u)dZ_u \\
&+ \int_{t_i}^s \Lambda_u \frac{\partial[h(x)r(t_{i+1},x)]}{\partial x}\Big|_{x=X_u} \sigma(X_u)dV_u.
\end{aligned}$$

For sake of simplicity, define

$$\begin{aligned}
\rho^\alpha(t,x) :=& \mathcal{A}[h(\cdot)r(t,\cdot)](x) - (\lambda(x)-1)h(x)r(t,x), \\
\eta^\alpha(t,x) :=& h^2(x)r(t,x), \\
\gamma^\alpha(t,x) :=& \frac{\partial[h(x)r(t,x)]}{\partial x}\sigma(x).
\end{aligned}$$

By assumption, coefficients in our model are smooth and compacted supported. Together, with the boundedness of r, Lemma 5.10 implies

$$\|\rho^\alpha\|_\infty, \|\eta^\alpha\|_\infty, \|\gamma^\alpha\|_\infty < \infty.$$

Then we have

$$\begin{aligned}
\mathbb{E}_x^0\Big\{\sum_{i=1}^n \alpha_i\Big\} =& \mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} dZ_s \int_{t_i}^s \Lambda_u \rho^\alpha(t_{i+1}, X_u)du\Big\} \\
&+ \mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} dZ_s \int_{t_i}^s \Lambda_u \eta^\alpha(t_{i+1}, X_u)dZ_u\Big\} \\
&+ \mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} dZ_s \int_{t_i}^s \Lambda_u \gamma^\alpha(t_{i+1}, X_u)dV_u\Big\} \\
:=& \alpha^{(1)} + \alpha^{(2)} + \alpha^{(3)}.
\end{aligned} \tag{5.4.67}$$

By stochastic Fubini's theorem, Protter (2005), page 208, we have

$$\begin{aligned}
\alpha^{(3)} =& \mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} dZ_s \int_{t_i}^s \Lambda_u \gamma^\alpha(t_{i+1}, X_u)\sigma(X_u)dV_u\Big\} \\
=& \sum_{i=1}^n \int_{t_i}^{t_{i+1}} dZ_s \mathbb{E}_x^0\Big\{\int_{t_i}^s \Lambda_u \gamma^\alpha(t_{i+1}, X_u)\sigma(X_u)dV_u\Big\}.
\end{aligned}$$

Notice that

$$\mathbb{E}_x^0 \int_{t_i}^s du\{\Lambda_u \gamma^\alpha(t_{i+1}, X_u)\sigma(X_u)\}^2 \leq \|\gamma^\alpha\|_\infty^2 \|\sigma\|_\infty^2 \mathbb{E}_x^0 \int_{t_i}^s \Lambda_u^2 du < \infty.$$

And the last inequality follows from

$$\mathbb{E}^0\Big[\int_{t_i}^s \Lambda_u^2 du | x_0 = x\Big] \leq (s - t_i) \sup_{0 \leq u \leq T} \mathbb{E}^0(\Lambda_u^2 | x_0 = x) < \infty.$$

Together with V_t, Y_t and Z_t are independent, we have

$$\mathbb{E}_x^0\Big\{\int_{t_i}^s \Lambda_u \gamma^\alpha(t_{i+1}, X_u)\sigma(X_u) dV_u\Big\} = 0.$$

Therefore,

$$\alpha^{(3)} = 0. \tag{5.4.68}$$

Now we consider $\alpha^{(1)}$.

$$\alpha^{(1)} = \mathbb{E}_x^0\Big\{\int_0^\theta dZ_s \sum_{i=1}^n \{\mathbf{1}_{\{t_i \leq s < t_{i+1}\}} \int_{t_i}^s \Lambda_u \rho^\alpha(t_{i+1}, X_u) du\}\Big\}.$$

In order to show that $\alpha^{(1)} \to 0$, by central limit theorem, it suffice to show that

$$\mathbb{E}_x^0\Big\{\int_0^\theta ds \Big[\sum_{i=1}^n \{\mathbf{1}_{\{t_i \leq s < t_{i+1}\}} \int_{t_i}^s \Lambda_u \rho^\alpha(t_{i+1}, X_u) du\}\Big]^2\Big\} \to 0.$$

That is

$$\mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} ds \Big[\int_{t_i}^s \Lambda_u \rho^\alpha(t_{i+1}, X_u) du\Big]^2\Big\} \to 0.$$

By Jensen's Inequality, the left hand side of the last formula is less then or equal to

$$\begin{aligned}
&\mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} ds(s - t_i) \int_{t_i}^s [\Lambda_u \rho^\alpha(t_{i+1}, X_u)]^2 du\Big\} \\
\leq &\|\rho^\alpha\|_\infty^2 \mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} ds(s - t_i) \int_{t_i}^s \Lambda_u^2 du\Big\}.
\end{aligned}$$

Notice that

$$\begin{aligned}
&\mathbb{E}^0\Big[\sum_{i=1}^n \int_{t_i}^{t_{i+1}} ds(s - t_i) \int_{t_i}^s \Lambda_u^2 du \big| X_0 = x\Big] \\
= &\sum_{i=1}^n \int_{t_i}^{t_{i+1}} ds(s - t_i) \int_{t_i}^s \mathbb{E}^0[\Lambda_u^2 | X_0 = x] du \\
\leq &\sup_{0 \leq u \leq T} \mathbb{E}^0[\Lambda_u^2 | X_0 = x] \sum_{i=1}^n \int_{t_i}^{t_{i+1}} ds(s - t_i)^2 \\
= &\sup_{0 \leq u \leq T} \mathbb{E}^0[\Lambda_u^2 | X_0 = x] \sum_{i=1}^n \frac{(t_{i+1} - t_i)^3}{3} \\
\leq &\sup_{0 \leq u \leq T} \mathbb{E}^0[\Lambda_u^2 | X_0 = x] \cdot \theta \cdot \frac{k^2}{3} \to 0,
\end{aligned}$$

as $m \to \infty$. Therefore

$$\lim_{m\to 0} \alpha^{(1)} \to 0, \quad a.s.. \tag{5.4.69}$$

In fact, the other convergence can be derived in the same way. However, for the reader, we show the proof in detail.

$$\begin{aligned}
\alpha^{(2)} =& \mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} dZ_s \int_{t_i}^{s} \Lambda_u \eta^\alpha(t_{i+1}, X_u) dZ_u\Big\} \\
=& \mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} \Lambda_u \eta^\alpha(t_{i+1}, X_u) dZ_u \int_{u}^{t_{i+1}} dZ_s\Big\} \\
=& \mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} \Lambda_u \eta^\alpha(t_{i+1}, X_u)(Z_{t_{i+1}} - Z_u) dZ_u\Big\} \\
=& \mathbb{E}_x^0\Big\{\int_0^\theta dZ_u \sum_{i=1}^n \mathbf{1}_{\{t_i \le u < t_{i+1}\}} \Lambda_u \eta^\alpha(t_{i+1}, X_u)(Z_{t_{i+1}} - Z_u)\Big\}.
\end{aligned}$$

Again, by central limit theorem, it suffice to show that

$$\mathbb{E}_x^0\Big\{\int_0^T du\{\sum_{i=1}^n \mathbf{1}_{\{t_i \le u < t_{i+1}\}} \Lambda_u \eta^\alpha(t_{i+1}, X_u)(Z_{t_{i+1}} - Z_u)\}^2\Big\} \to 0.$$

That is

$$\mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} du[\Lambda_u \eta^\alpha(t_{i+1}, X_u)(Z_{t_{i+1}} - Z_u)]^2\Big\} \to 0.$$

The left hand side of the last formula is less than or equal to

$$\|\eta^\alpha\|_\infty^2 \mathbb{E}_x^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} du[\Lambda_u (Z_{t_{i+1}} - Z_u)]^2\Big\}.$$

Notice that, by Cauchy-Schwartz inequality,

$$\begin{aligned}
& \mathbb{E}^0\Big\{\sum_{i=1}^n \int_{t_i}^{t_{i+1}} du[\Lambda_u (Z_{t_{i+1}} - Z_u)]^2 | X_0 = x\Big\} \\
=& \sum_{i=1}^n \int_{t_i}^{t_{i+1}} du \mathbb{E}^0\{[\Lambda_u (Z_{t_{i+1}} - Z_u)]^2 | X_0 = x\} \\
\le& \sum_{i=1}^n \int_{t_i}^{t_{i+1}} du[\mathbb{E}^0(\Lambda_u^4 | X_0 = x)]^{\frac{1}{2}} [\mathbb{E}^0((Z_{t_{i+1}} - Z_u)^4 | X_0 = x)]^{\frac{1}{2}}] \\
\le& \sup_{0\le u\le T} [\mathbb{E}^0(\Lambda_u^4 | X_0 = x)]^{\frac{1}{2}} \sum_{i=1}^n \int_{t_i}^{t_{i+1}} \sqrt{3}(t_{i+1} - u)^2 du \\
\le& \sup_{0\le u\le T} [\mathbb{E}^0(\Lambda_u^4 | X_0 = x)]^{\frac{1}{2}} \sum_{i=1}^n \frac{(t_{i+1} - t_i)^3}{\sqrt{3}} \\
\le& \sup_{0\le u\le T} [\mathbb{E}^0(\Lambda_u^4 | X_0 = x)]^{\frac{1}{2}} \cdot \theta \cdot \frac{k^2}{\sqrt{3}} \to 0.
\end{aligned}$$

Therefore,

$$\lim_{l\to 0} \alpha^{(2)} \to 0, \quad a.s.. \tag{5.4.70}$$

Now we show that $\mathbb{E}_x^0 \sum_{i=1}^n \beta_i \to 0$. By Equation (5.4.25), we have

$$\begin{aligned} & r(t_{i+1}-, X_{t_i}) - r(s, X_{t_i}) \\ =& \int_s^{t_{i+1}} -\mathcal{A}r(u, X_{t_i})du - h(X_{t_i})r(u, X_{t_i})dZ_u + (\lambda(X_{t_i}) - 1)r(u, X_{t_i})du \\ =& \int_s^{t_{i+1}} [-\mathcal{A}r(u, X_{t_i}) + (\lambda(X_{t_i}) - 1)r(u, X_{t_i})]du + \int_s^{t_{i+1}} -h(X_{t_i})r(u, X_{t_i})dZ_u. \end{aligned}$$

For sake of simplicity, define

$$\begin{aligned} \rho^\beta(t, x) :=& h(x)[-\mathcal{A}r(t, x) + (\lambda(x) - 1)r(t, x)], \\ \eta^\beta(t, x) :=& - h(x)^2 r(t, x). \end{aligned}$$

By assumption coefficients in our model are smooth and compacted supported. Together with the bounded of r, by Lemma 5.10, we have

$$\|\rho^\beta\|_\infty, \ \|\eta^\beta\|_\infty < \infty.$$

Then we have

$$\begin{aligned} \mathbb{E}_x^0 \Big\{ \sum_{i=1}^n \beta_i \Big\} =& \mathbb{E}_x^0 \Big\{ \sum_{i=1}^n \int_{t_i}^{t_{i+1}} dZ_s \Lambda_{t_i} \int_s^{t_{i+1}} \rho^\beta(u, X_{t_i})du \Big\} \\ &+ \mathbb{E}_x^0 \Big\{ \sum_{i=1}^n \int_{t_i}^{t_{i+1}} dZ_s \Lambda_{t_i} \int_s^{t_{i+1}} \eta^\beta(u, X_{t_i})dZ_u \Big\} \\ :=& \beta^{(1)} + \beta^{(2)}. \end{aligned} \tag{5.4.71}$$

Again, we rewrite $\beta^{(1)}$ as

$$\beta^{(1)} = \mathbb{E}_x^0 \int_0^\theta dZ_s \sum_{i=1}^n \mathbf{1}_{\{t_i \le s < t_{i+1}\}} \Lambda_{t_i} \int_s^{t_{i+1}} \rho^\beta(u, X_{t_i})du.$$

By the central limit theorem, it suffice to show that

$$\mathbb{E}_x^0 \Big\{ \int_0^\theta ds \{ \sum_{i=1}^n \mathbf{1}_{\{t_i \le s < t_{i+1}\}} \Lambda_{t_i} \int_s^{t_{i+1}} \rho^\beta(u, X_{t_i})du \}^2 \Big\} \to 0.$$

That is

$$\mathbb{E}_x^0 \Big\{ \sum_{i=1}^n \int_{t_i}^{t_{i+1}} ds (\Lambda_{t_i} \int_s^{t_{i+1}} \rho^\beta(u, X_{t_i})du)^2 \Big\} \to 0.$$

In fact, the left hand side of the last formula is less than or equal to

$$\|\rho^\beta\|_\infty^2 \mathbb{E}_x^0 \Big\{ \sum_{i=1}^n \int_{t_i}^{t_{i+1}} ds \Lambda_{t_i}^2 (t_{i+1} - s)^2 \Big\} = \|\rho^\beta\|_\infty^2 \mathbb{E}_x^0 \Big\{ \sum_{i=1}^n \Lambda_{t_i}^2 \frac{(t_{i+1} - t_i)^3}{3} \Big\}.$$

Notice that

$$\mathbb{E}^0\Big[\sum_{i=1}^n \Lambda_{t_i}^2 \frac{(t_{i+1}-t_i)^3}{3}|X_0=x\Big] \le \sup_{0\le t\le T}\mathbb{E}^0[\Lambda_t^2|X_0=x\Big]\sum_{i=1}^n \frac{(t_{i+1}-t)^3}{3}$$
$$\le \sup_{0\le t\le T}\mathbb{E}^0[\Lambda_t^2|X_0=x]\cdot\theta\cdot\frac{k^2}{3}\to 0.$$

Therefore,

$$\beta^{(1)}\to 0,\ \ a.s.. \tag{5.4.72}$$

We rewrite $\beta^{(2)}$ as

$$\begin{aligned}\beta^{(2)} =&\mathbb{E}_x^0\Big\{\sum_{i=1}^n\int_{t_i}^{t_{i+1}} dZ_s\Lambda_{t_i}\int_s^{t_{i+1}}\eta^\beta(u,X_{t_i})dZ_u\Big\}\\ =&\mathbb{E}_x^0\Big\{\sum_{i=1}^n\int_{t_i}^{t_{i+1}} dZ_u\Lambda_{t_i}\eta^\beta(u,X_{t_i})\int_{t_i}^u dZ_s\Big\}\\ =&\mathbb{E}_x^0\Big\{\sum_{i=1}^n\int_{t_i}^{t_{i+1}} dZ_u\Lambda_{t_i}\eta^\beta(u,X_{t_i})(Z_u-Z_{t_i})\Big\}\\ =&\mathbb{E}_x^0\Big\{\int_0^T dZ_u\sum_{i=1}^n \mathbf{1}_{\{t_i\le u<t_{i+1}\}}\Lambda_{t_i}\eta^\beta(u,X_{t_i})(Z_u-Z_{t_i})\Big\}.\end{aligned}$$

Again, by central limit theorem, it suffice to show that

$$\mathbb{E}_x^0\Big\{\int_0^T du\{\sum_{i=1}^n \mathbf{1}_{\{t_i\le u<t_{i+1}\}}\Lambda_{t_i}\eta^\beta(u,X_{t_i})(Z_u-Z_{t_i})\}^2\Big\}\to 0.$$

That is

$$\mathbb{E}_x^0\Big\{\sum_{i=1}^n\int_{t_i}^{t_{i+1}} du[\Lambda_{t_i}\eta^\beta(u,X_{t_i})(Z_u-Z_{t_i})]^2\Big\}\to 0.$$

In fact, the left hand side of the last formula is less than or equal to

$$\|\eta^\beta\|_\infty^2\mathbb{E}_x^0\Big\{\sum_{i=1}^n\int_{t_i}^{t_{i+1}} du[\Lambda_{t_i}(Z_u-Z_{t_i})]^2\Big\}.$$

Notice that, by Cauchy-Schwartz inequality,

$$\begin{aligned}&\mathbb{E}^0\Big\{\sum_{i=1}^n\int_{t_i}^{t_{i+1}} du[\Lambda_{t_i}(Z_u-Z_{t_i})]^2|X_0=x\Big\}\\ \le&\sum_{i=1}^n\int_{t_i}^{t_{i+1}} du\Big(\mathbb{E}^0[\Lambda_{t_i}^4|X_0=x]\Big)^{\frac12}\Big(\mathbb{E}^0[(Z_u-Z_{t_i})^4|X_0=x]\Big)^{\frac12}\\ \le&\sup_{0\le t\le T}\Big(\mathbb{E}^0[\Lambda_t^4|X_0=x]\Big)^{\frac12}\sum_{i=1}^n\int_{t_i}^{t_{i+1}} du\sqrt{3}(u-t_i)^2\\ \le&\sup_{0\le t\le T}\Big(\mathbb{E}^0[\Lambda_t^4|X_0=x]\Big)^{\frac12}\sum_{i=1}^n\frac{(t_{i+1}-t_i)}{\sqrt{3}}\\ \le&\sup_{0\le t\le T}\Big(\mathbb{E}^0[\Lambda_t^4|X_0=x]\Big)^{\frac12}\cdot\theta\cdot\frac{k^2}{\sqrt{3}}\to 0.\end{aligned}$$

Therefore,

$$\beta^{(2)} \to 0, \ \ a.s.. \tag{5.4.73}$$

Combining Equation (5.4.67), (5.4.69), (5.4.70), (5.4.68), (5.4.71), (5.4.72) and (5.4.73), we obtain the desired result. □

Remark 5.7. *For the proof of Lemma 5.9, the basic idea is borrowed from Pardoux (1979a) and Pardoux (1979b). The reason we show the proof here is as follows.*

- *Because our model have both jump and continuous observations, we cannot directly cite the result of Pardoux (1979b).*
- *There are some assertions in Pardoux (1979b) without proof.*
  - $v \in \bigcap_n H^n(\mathbb{R}^d) \Rightarrow v \in C_b^\infty(\mathbb{R}^d)$, *page 154.*
  - $\mathbb{E}^0\Big\{ \sup_{0\le s\le T} \Big[h^2(X_s)v(s,X_s)\Lambda_s\Big]^2 \Big| X_0 = x\Big\} < \infty$, *page 157.*

5.4.3 The unnormalized conditional density

One objective of this section is to show that Equation (5.2.1) is adjoint of Equation (5.4.25) in the sense specified by the following:

Theorem 5.12. *Assume Assumption 5.1 holds. Let $\{q_t\}_{t\in[0,T]}$ be the solution of Equation* (5.2.1) *and $\{r_t\}_{t\in[0,\theta]}$ be the solution of Equation* (5.4.25)*. Then, the following holds a.s.*

$$(q_t, r_t) = (q_s, r_s), \quad \forall s,t \in [0,\theta]. \tag{5.4.74}$$

Proof. Set $R_t = (q_t, r_t)$. Then it is suffice to show that

$$dR_t = 0.$$

It holds between jump times, by Pardoux (1979b), Theorem 3.1. At the time of jump, by Equations (5.2.1) and (5.4.25), for any jump time $0 < \tau_n \le \theta$,

$$\begin{aligned} q_{\tau_n} - q_{\tau_n-} &= q_{\tau_n-}(\lambda - 1), \\ r_{\tau_n} - r_{\tau_n-} &= -r_{\tau_n}(\lambda - 1). \end{aligned}$$

That is

$$q_{\tau_n} = q_{\tau_n-}\lambda, \ r_{\tau_n-} = r_{\tau_n}\lambda.$$

Therefore,

$$\triangle R_{\tau_n} = R_{\tau_n} - R_{\tau_n-} = (q_{\tau_n}, r_{\tau_n}) - (q_{\tau_n-}, r_{\tau_n-}) = (q_{\tau_n-}\lambda, r_{\tau_n}) - (q_{\tau_n-}, r_{\tau_n}\lambda) = 0.$$

□

Remark 5.8. *In the proof, we cannot differentiate (q_t, r_t), because its differential could involve terms which do not make sense.*

Another objective of this section is to show that p, the solution of Equation (5.2.1), is the unnormalized conditional density.

Theorem 5.13. *Assume Assumption 5.1 holds. Let* $\{q_t\}_{t\in[0,T]}$ *be the solution of Equation* (5.2.1)*. Then, for* f *bounded and square integrable,*

$$\mathbb{E}^0\Big[f(X_\theta)\Lambda_\theta\Big|\mathcal{F}_\theta^{Z,Y}\Big] = (q_\theta, f), \quad \theta \in [0,T]. \tag{5.4.75}$$

Proof. Let $\{r_t\}_{t\in[0,\theta]}$ be the solution of Equation (5.4.25). By Equation (5.4.30),

$$(q_0, r_0) = (q_\theta, r_\theta).$$

Applying Equation (5.4.74),

$$\int_{\mathbb{R}^d} p_0(x)\mathbb{E}^0_{0,x}(f(X_\theta)\Lambda_\theta|\mathcal{F}_\theta^{Z,Y})dx = (q_\theta, f).$$

□

θ is arbitrary, we have proved that Equation (5.2.4), which is as same as Equation (5.2.1) describes the evolution of the unnormalized conditional density.

The following theorem identifies the density of the conditional distribution of the signal.

Theorem 5.14. *Assume Assumption 5.1 holds. Let* $\{q_t\}_{t\in[0,T]}$ *be the solution of Equation* (5.2.1)*. Define* $p_\theta := \frac{q_\theta}{\int_{\mathbb{R}^d} q_\theta(x)dx}$, $\forall\theta \in [0,T]$*, then, for* f *bounded and square integrable,*

$$\mathbb{E}\Big[f(X_\theta)\Big|\mathcal{F}_\theta^{Z,Y}\Big] = (p_\theta, f), \quad \theta \in [0,T].$$

Proof. Immediate from Theorem 5.13 and Proposition 2.8. □

Chapter 6

Convergence results

The Galerkin method is one of the well-known methods in the theory of partial differential equations used to prove existence properties and to obtain finite dimensional approximations for the solutions of equations.

Since the Galerkin approximation is successful in solving deterministic partial differential equations, it seems natural to extend these methods to stochastic equations. In recent years, the deterministic Galerkin method, adapted to the stochastic case, has been examined by many authors.

Pardoux (1979b) proves the existence of the weak solution of the stochastic parabolic linear PDE's, approximating it by means of the Galerkin method.

Gyöngy (1988) studies the stability of stochastic partial differential equations with respect to simultaneous perturbation of the driving processes and the differential operators. Gyöngy (1989) applied the result in nonlinear filtering.

Grecksch and Kloeden (1996) study the convergence of the numerical methods for parabolic stochastic partial differential equations with time and space discretisations using the Galerkin method.

The Galerkin approximation for the Zakai equation, with continuous noisy observations, was considered in the literature by Ahmed and Radaideh (1997) and Germani and Piccioni (1984). Ahmed and Radaideh (1997) used the Galerkin method to approximate the unnormalized, conditional density of the filtered diffusion process. They illustrated the method with numerical examples. Their work mainly concentrated on the technique, and they didn't show the convergence of the Galerkin approximation theoretically. Germani and Piccioni (1984) show the convergence of the Galerkin approximation in mean square norm under appropriate conditions on the data. Twardowska, Marnik, and Paslawska-Poludniak (2003) apply the numerical method, mentioned by Ahmed and Radaideh (1997), to approximate the solution of Zakai equation with a delay.

These authors investigated stochastic partial differential equations with gaussian noise. However, the Galerkin method can also be used for stochastic partial differential equations with Lévy noise.

In this chapter, we apply the Galerkin method to nonlinear filtering problem w.r.t. Model (2.2.9) with Lévy noise. The solution of the Zakai equation is approximated by a finite combination of orthogonal series. Then, the solution of the Zakai equation is approximated by the solution of a family of finite dimensional stochastic ordinary differential equations. These can be solved numerically or analytically.

The objective of Chapter 6 is to show the convergence of the Galerkin approximation in mean square. Analogous results have been proven in the case with continuous noisy observations by Germani and Piccioni (1984). They have provided an idea to handle such a problem. We apply their idea to our case where the observation processed with both diffusion and jump noise.

In Section 6.1, we present the Galerkin approximation for solutions of Zakai equation (5.2.1) by solving a sequence of finite dimensional stochastic differential equations. A question then concerns the convergence and so, we will proof the Galerkin approximation convergent in mean square norm.

In the proof, the key tool is the mild solution of Zakai equation. In Section 6.2, we show the equivalence of weak and mild solution of the Zakai equation.

In Section 6.3, we give a definition of stochastic integral in Hilbert space. With such a definition, the Zakai equation can be rewritten as a mild equation. We then have a decomposition of the mild one, which is important to obtain the main result.

In order to study Galerkin approximation for equations of Zakai type, the main tool is a continuity theorem. From it, we know that the convergence of Galerkin approximation is guaranteed by the convergence of the linear deterministic equation. It is shown in Section 6.4.

Finally, in Section 6.5, we deal with the necessary and sufficient condition for the convergence of the linear deterministic equation. Together, with the continuity theorem, we get the necessary and sufficient condition for the mean square convergence of Galerkin approximation.

In this chapter, a large class of linear stochastic partial differential equations are studied. Although the techniques we introduce here are used for the Zakai equation, they can also be applied to those classes of linear stochastic PDEs that include the Zakai equation as a special example.

6.1 Introduction

In this section, we will show that by the Galerkin approximation, the solution of Zakai equation (5.2.1) is approximated by a finite combination of orthogonal series. The solution of the Zakai equation is then approximated by the solution of a family of finite dimensional stochastic ordinary differential equations, which can be solved numerically or analytically.

We make the following definition of the Galerkin approximation of Equation (5.2.3) which is equivalent of Equation (5.2.1).

Definition 6.1. Let $\{e_i, i = 0, 1, \ldots\}$ denote a Hilbert-basis of H, made of elements of $D(\mathcal{A})$ which is defined by Equation (5.1.2). Then $\{e_i\}$ is linearly independent and complete in H. For each natural n we consider n-dimensional subspace of H defined by $V_n = \operatorname{span}\{e_0, e_1, \ldots, e_{n-1}\}$ equipped with the norm induced from H. Let P_n be the orthogonal projections from H to V_n, that is $P_n H = V_n \subset H$, and we have, $\forall v \in H$,

$$P_n v = \sum_{i=0}^{n-1} (v, e_i) e_i \to v, \quad \text{as} \quad n \to \infty.$$

Let $(\mathcal{A}^*)^{(n)} : H \to V_n$, $\mathcal{B}_h^{(n)} : H \to V_n$, and $\mathcal{B}_\lambda^{(n)} : H \to V_n$ be defined respectively by

$$(\mathcal{A}^*)^{(n)} := P_n \mathcal{A}^* P_n, \;\; \mathcal{B}_h^{(n)} := P_n h^\top(\cdot) P_n, \;\; \mathcal{B}_\lambda^{(n)} := P_n [\lambda(\cdot) - 1] P_n.$$

The n-dimensional Galerkin approximation $\{q_t^{(n)}\}_{t\in[0,T]}$ to SPDE (5.2.3) is then, for any $v \in V_n$,

$$\begin{cases} (q_t^{(n)}, v) = (q_0^{(n)}, v) + \int_0^t \langle (\mathcal{A}^*)^{(n)} q_s^{(n)}, v\rangle ds + \int_0^t (\mathcal{B}_h^{(n)} q_s^{(n)}, v) dZ_s \\ \qquad\qquad + \int_0^t (\mathcal{B}_\lambda^{(n)} q_{s-}^{(n)}, v) d(Y_s - s), \quad t \in [0,T], \\ q_0^{(n)} = P_n p_0. \end{cases} \tag{6.1.1}$$

We will write the approximating equation coordinatewise. First, we can approximate the solution of the Zakai equation (5.2.1) in the form

$$q_t^{(n)}(x) = \sum_{i=0}^{n-1} \psi_i^{(n)}(t) e_i(x), \quad t \in [0,T], \tag{6.1.2}$$

where $\{\psi_i^{(n)}\}$ are Fourier coefficients to be chosen as follows. By projecting Equation (5.2.3) into space V_n spanned by $\{e_i, i = 0, 1, \ldots, n-1\}$, we obtain, for $j = 0, 1, \ldots, n-1$, $t \in [0,T]$,

$$\begin{aligned} \sum_{i=0}^{n-1} (e_i, e_j) d\psi_i^{(n)}(t) =& \Big\{ \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \langle \mathcal{A}^* e_i, e_j\rangle \Big\} dt + \sum_{k=1}^{l} \Big\{ \sum_{i=0}^{n-1} \psi_i^{(n)}(t) (h^k e_i, e_j) \Big\} dZ_t^k \\ &+ \Big\{ \sum_{i=0}^{n-1} \psi_i^{(n)}(t-) \Big((\lambda - 1) e_i, e_j \Big) \Big\} d(Y_t - t). \end{aligned}$$

This is an n-dimensional SDE for $\Upsilon^{(n)} = (\psi_0^{(n)}, \psi_1^{(n)}, \ldots, \psi_{n-1}^{(n)})^\top$ given by for $j = 0, 1, \ldots, n-1$,

$$\begin{cases} \sum_{i=0}^{n-1} a_{ji} d\psi_i^{(n)}(t) = \Big\{ \sum_{i=0}^{n-1} b_{ji} \psi_i^{(n)}(t) \Big\} dt + \sum_{k=1}^{l} \Big\{ \sum_{i=0}^{n-1} c_{ji}^k \psi_i^{(n)}(t) \Big\} dZ_t^k \\ \qquad\qquad + \Big\{ \sum_{i=0}^{n-1} g_{ji} \psi_i^{(n)}(t-) \Big\} d(Y_t - t), \\ \sum_{i=0}^{n-1} a_{ji} \psi_i^{(n)}(0) = q_{0j}, \end{cases}$$

where, for $i, j = 0, 1, \ldots, n-1$,

$$a_{ji} := (e_i, e_j), \tag{6.1.3}$$
$$b_{ji} := \langle \mathcal{A}^* e_i, e_j \rangle, \tag{6.1.4}$$
$$c_{ji}^k := (h^k e_i, e_j), \tag{6.1.5}$$
$$g_{ji} := \Big((\lambda - 1) e_i, e_j \Big), \tag{6.1.6}$$
$$q_{0j} := (p_0, e_j). \tag{6.1.7}$$

This is a finite-dimensional approximation of the Zakai equation (5.2.1).

Remark 6.1. *If $\{e_i, i = 0, 1, \ldots\} \subset D(\mathcal{L})$, then, by the definition of b_{ji},*

$$b_{ji} = \langle \mathcal{A}^* e_i, e_j \rangle = \langle e_i, \mathcal{A} e_j \rangle = \langle e_i, \mathcal{L} e_j \rangle = (e_i, \mathcal{L} e_j),$$

where $\mathcal{L}$ is the generator of the state process X defined by Equation (2.2.4). *Since, the basis functions used in the Galerkin approximation, for example Hermit polynomials and Gaussian functions, are usually subsets of $D(\mathcal{L})$, we compute b_{ji} by $(e_i, \mathcal{L} e_j)$ which is more convenient to deal with.*

For sake of brevity, define $A^{(n)} := (a_{ij})_{n\times n}$, $B^{(n)} := (b_{ij})_{n\times n}$, $C^{(n),k} := (c^k_{ij})_{n\times n}$, $G^{(n)} := (g_{ij})_{n\times n}$, $\mathbf{q}_0^{(n)} = (q_{00}, \dots, q_{0(n-1)})^\top$. We have the matrix notation equation, see Ahmed and Radaideh (1997),

$$\begin{cases} A^{(n)} d\Upsilon^{(n)}(t) = B^{(n)}\Upsilon^{(n)}(t)dt + \sum_{k=1}^{l} \left\{C^{(n),k}\Upsilon^{(n)}(t)\right\}dZ_t^k \\ \qquad\qquad +G^{(n)}\Upsilon^{(n)}(t-)d(Y_t - t), \quad t \in [0,T], \\ A^{(n)}\Upsilon^{(n)}(0) = \mathbf{q}_0^{(n)}. \end{cases} \tag{6.1.8}$$

Notice, $\{e_i\}$ is a basis of H whose components are linearly independent. Therefore $A^{(n)}$ has full rank, and is invertible. Thus, Equation (6.1.8) can be rewritten as

$$\begin{cases} d\Upsilon^{(n)}(t) = (A^{(n)})^{-1}[B^{(n)}\Upsilon^{(n)}(t)dt + \sum_{k=1}^{l} C^{(n),k}\Upsilon^{(n)}(t)dZ_t^k \\ \qquad\qquad +G^{(n)}\Upsilon^{(n)}(t-)d(Y_t - t)], \quad t \in [0,T], \\ \Upsilon^{(n)}(0) = (A^{(n)})^{-1}\mathbf{q}_0^{(n)}, \end{cases} \tag{6.1.9}$$

which can be solved numerically or analytically.

Now, there is a question which concerns the convergence of $\{q^{(n)}\}$ which is defined by Equation (6.1.2). We will show that $\{q^{(n)}\}$ converges to q in mean square sense.

6.2 The mild form of the Zakai equation

By Equation (5.1.3), if Assumption 5.1 satisfies, the restriction of $\mathcal{A}$ to $D(\mathcal{A})$ which is defined by Equation (5.1.2), is the infinitesimal generator of a contraction semigroup G of bounded operators on H, such that, for $t \in [0,T]$,

$$\|G_t\| \le e^{\beta t}. \tag{6.2.1}$$

Notice, the generator of G and operator $\mathcal{A}$, which is defined by Equation (5.1.1), is different. And the generator of G is the restriction of $\mathcal{A}$ to $D(\mathcal{A})$.

By the following theorem, we show the equivalence of weak and mild solutions.

Theorem 6.1. *Assume Assumption 5.1 is fulfilled. Then, the process $\{q_t, t \in [0,T]\}$ is the unique solution of Equation* (5.2.1), *if and only if it is the solution of the following equation, for $t \in [0,T]$,*

$$\begin{aligned}(q_t, v) =&(G_t^* q_0, v) + \int_0^t \left(q_\tau, h^\top G_{t-\tau} v\right) dZ_\tau \\ &+ \int_0^t \left(q_{\tau-}, (\lambda-1)G_{t-\tau}v\right) d(Y_\tau - \tau), \;\; v \in H.\end{aligned} \tag{6.2.2}$$

Proof. Notice, the solution of the Zakai equation (5.2.1) satisfies that, for $v \in D(\mathcal{A})$, $t \in [0,T]$,

$$\begin{cases} (v, q_t) = (v, q_0) + \int_0^t (\mathcal{A}v, q_s)ds + \int_0^t \left(h^\top v, q_s\right)dZ_s \\ \qquad\qquad + \int_0^t \left((\lambda-1)v, q_{s-}\right)d(Y_s - s), \\ q_0 = p_0 \in H. \end{cases} \tag{6.2.3}$$

The desired result is then obtained from the existence and uniqueness result of the Zakai equation, see Theorem 5.4, the existence and uniqueness result of Equation (6.2.3), see Peszat and

Zabczyk (2007), Theorem 9.29, page 164, and the equivalence of Equation (6.2.3) and Equation (6.2.2), see Peszat and Zabczyk (2007), Theorem 9.15, page 151.

□

Zakai equation (5.2.1) can be expressed in a mild form as soon as a suitable definition of a stochastic integral is given.

6.3 The stochastic integral

This section is devoted to construct a stochastic Itô integration of Hilbert-valued functions with respect to a finite dimensional Lévy process.

We define

$$M_t^1 = (M_t^{1,1}, M_t^{1,2}, \dots, M_t^{1,l})^\top := Z_t, \quad M_t^2 := Y_t - t, \quad t \in [0,T]. \tag{6.3.1}$$

By Proposition 2.5, under measure $\mathbb{P}^0$, Z is an l-dimensional standard Brownian motion and Y is a one-dimensional Poisson process with intensity 1. Then M^1 and M^2 are martingales.

Let $\mathcal{N}^2(0,T;H)$ be a space of stochastic processes which are continuous in the mean square norm. Lemma A.8 shows that $\mathcal{N}^2(0,T;H)$ is a Banach space with norm $|\cdot|_T$. For details of its definition, see Section A.1.2. We have the following definition of stochastic Itô integration:

Lemma 6.2. *Let $\{\phi_i, i \in \mathbb{N}\}$ be any complete orthonormal system(CONs) in H, then, for any $\xi = \Big((\xi^1)^\top, \xi^2\Big)^\top = (\xi^{1,1}, \dots, \xi^{1,l}, \xi^2)^\top$ with $\xi^{1,1}, \dots, \xi^{1,l}, \xi^2 \in \mathcal{N}^2(0,T;H)$, the sequence $\{I^{(n)}\} \subset \mathcal{N}^2(0,T;H)$ defined by*

$$I_t^{(n)} := \sum_{i=1}^{n} \Big\{ \sum_{j=1}^{l} \Big[\int_0^t (\xi_s^{1,j}, \phi_i) dM_s^{1,j} \Big] + \int_0^t (\xi_{s-}^2, \phi_i) dM_s^2 \Big\} \phi_i, \;\; t \in [0,T], \tag{6.3.2}$$

converges to a limit in $\mathcal{N}^2(0,T;H)$ which is independent of $\{\phi_i, i \in \mathbb{N}\}$ denoted by

$$I_t = \int_0^t (\xi_s^1)^\top dM_s^1 + \int_0^t \xi_{s-}^2 dM_s^2, \quad t \in [0,T]. \tag{6.3.3}$$

Proof. The convergence is obtained if one can show, for $n > m$

$$|I^{(n)} - I^{(m)}|_T \to 0, \;\; \text{as } m,n \to \infty. \tag{6.3.4}$$

By the definition of $|\cdot|_T$, we get

$$\begin{aligned}
|I^{(n)} - I^{(m)}|_T^2 &= \sup_{t \in [0,T]} \mathbb{E}^0 \Big\| \sum_{i=m+1}^{n} \Big[\sum_{j=1}^{l} \int_0^t (\xi_s^{1,j}, \phi_i) dM_s^{1,j} + \int_0^t (\xi_{s-}^2, \phi_i) dM_s^2 \Big] \phi_i \Big\|_H^2 \\
&= \sup_{t \in [0,T]} \mathbb{E}^0 \Big\{ \sum_{i=m+1}^{n} \Big[\sum_{j=1}^{l} \int_0^t (\xi_s^{1,j}, \phi_i) dM_s^{1,j} + \int_0^t (\xi_{s-}^2, \phi_i) dM_s^2 \Big]^2 \Big\} \\
&= \sup_{t \in [0,T]} \mathbb{E}^0 \Big\{ \sum_{i=m+1}^{n} \Big[\sum_{j=1}^{l} \int_0^t (\xi_s^{1,j}, \phi_i)^2 ds + \int_0^t (\xi_s^2, \phi_i)^2 ds \Big] \Big\} \\
&= \mathbb{E}^0 \Big\{ \sum_{j=1}^{l} \int_0^T \Big[\sum_{i=m+1}^{n} (\xi_s^{1,j}, \phi_i)^2 \Big] ds + \int_0^T \Big[\sum_{i=m+1}^{n} (\xi_s^2, \phi_i)^2 \Big] ds \Big\}.
\end{aligned}$$

The first equality follows from the definition of $I^{(n)}$ and $|\cdot|_T$. The second equality follows from the orthonomal property of $\{\phi_i\}$. The third equality follows from Itô Isometry.

On the one hand,

$$\sum_{i=m+1}^{n}(\xi_s^{1,j},\phi_i)^2 \le \sum_{i=1}^{\infty}(\xi_s^{1,j},\phi_i)^2 = \|\xi_s^{1,j}\|_H^2, \text{ for } j=1,\dots,l,$$
$$\sum_{i=m+1}^{n}(\xi_s^{2},\phi_i)^2 \le \sum_{i=1}^{\infty}(\xi_s^{2},\phi_i)^2 = \|\xi_s^{2}\|_H^2.$$

Therefore,

$$\begin{aligned}
|I^{(n)} - I^{(m)}|_T^2 \le& \mathbb{E}^0\Big\{\sum_{j=1}^{l}\int_0^T \|\xi_s^{1,j}\|_H^2 ds + \int_0^T \|\xi_s^2\|_H^2 ds\Big\} \qquad (6.3.5)\\
=& \mathbb{E}^0\Big\{\int_0^T \Big[\sum_{j=1}^{l}\|\xi_s^{1,j}\|_H^2 + \|\xi_s^2\|_H^2\Big] ds\Big\}\\
=& \mathbb{E}^0\Big\{\int_0^T \|\xi_s\|_{H^{l+1}}^2 ds\Big\}\\
=& \int_0^T \mathbb{E}^0\|\xi_s\|_{H^{l+1}}^2 ds\\
\le& T\cdot \sup_{s\in[0,T]} \mathbb{E}^0\|\xi_s\|_{H^{l+1}}^2\\
\le& T\cdot\Big\{\sum_{j=1}^{l}|\xi^{1,j}|_T^2 + |\xi^2|_T^2\Big\} < \infty.
\end{aligned}$$

On the other hand,

$$\sum_{i=m+1}^{n}(\xi_s^{1,j},\phi_i)^2 \to 0, \text{ for } j=1,\dots,l,$$
$$\sum_{i=m+1}^{n}(\xi_s^{2},\phi_i)^2 \to 0,$$

a.s. and a.e.. Applying stochastic dominated convergence theorem, Equation (6.3.4) is obtained.

Now, we show the convergence is independent of the CONs $\{\phi_i, i\in\mathbb{N}\}$. Let $\{\tilde{\phi}_i, i\in\mathbb{N}\}$ be another CONs basis in H. Define

$$\tilde{I}_t^{(n)} := \sum_{i=1}^{n}\Big\{\sum_{j=1}^{l}\Big[\int_0^t(\xi_s^{1,j},\tilde{\phi}_i)dM_s^{1,j}\Big] + \int_0^t(\xi_{s-}^2,\tilde{\phi}_i)dM_s^2\Big\}\tilde{\phi}_i, \ \ t\in[0,T].$$

Then it is suffice to show that

$$|I^{(n)} - \tilde{I}^{(n)}|_T \to 0. \qquad (6.3.6)$$

In fact, by the definition of $|\cdot|_T$, we get

$$
\begin{aligned}
|I^{(n)}-\tilde{I}^{(n)}|_T = \sup_{t\in[0,T]} \mathbb{E}^0 \Big\| \sum_{i=1}^{n} \Big\{ \sum_{j=1}^{l} \Big[\int_0^t (\xi_s^{1,j},\phi_i) dM_s^{1,j} \Big] + \int_0^t (\xi_{s-}^2,\phi_i) dM_s^2 \Big\} \phi_i \\
- \sum_{i=1}^{n} \Big\{ \sum_{j=1}^{l} \Big[\int_0^t (\xi_s^{1,j},\tilde{\phi}_i) dM_s^{1,j} \Big] + \int_0^t (\xi_{s-}^2,\tilde{\phi}_i) dM_s^2 \Big\} \tilde{\phi}_i \Big\|_H^2 \\
= \sup_{t\in[0,T]} \mathbb{E}^0 \Big\| \int_0^t \{ \sum_{i=1}^{n}\sum_{j=1}^{l} (\xi_s^{1,j},\phi_i)\phi_i - \sum_{i=1}^{n}\sum_{j=1}^{l} (\xi_s^{1,j},\tilde{\phi}_i)\tilde{\phi}_i \} dM_s^{1,j} \\
+ \int_0^t \{ \sum_{i=1}^{n} (\xi_{s-}^2,\phi_i)\phi_i - \sum_{i=1}^{n} (\xi_{s-}^2,\tilde{\phi}_i)\tilde{\phi}_i \} dM_s^2 \Big\|_H^2 \\
= \sup_{t\in[0,T]} \mathbb{E}^0 \Big\{ \int_0^t \Big\| \sum_{i=1}^{n}\sum_{j=1}^{l} (\xi_s^{1,j},\phi_i)\phi_i - \sum_{i=1}^{n}\sum_{j=1}^{l} (\xi_s^{1,j},\tilde{\phi}_i)\tilde{\phi}_i \Big\|_H^2 ds \\
+ \int_0^t \Big\| \sum_{i=1}^{n} (\xi_{s-}^2,\phi_i)\phi_i - \sum_{i=1}^{n} (\xi_{s-}^2,\tilde{\phi}_i)\tilde{\phi}_i \Big\|_H^2 ds \Big\} \\
= \mathbb{E}^0 \Big\{ \int_0^T \Big\| \sum_{i=1}^{n}\sum_{j=1}^{l} (\xi_s^{1,j},\phi_i)\phi_i - \sum_{i=1}^{n}\sum_{j=1}^{l} (\xi_s^{1,j},\tilde{\phi}_i)\tilde{\phi}_i \Big\|_H^2 ds \\
+ \int_0^T \Big\| \sum_{i=1}^{n} (\xi_{s-}^2,\phi_i)\phi_i - \sum_{i=1}^{n} (\xi_{s-}^2,\tilde{\phi}_i)\tilde{\phi}_i \Big\|_H^2 ds \Big\}.
\end{aligned}
$$

The third equality follows from Lemma 6.3 which will be shown later. On the one hand, since the orthonormal decompensation in Hilbert H does not depend on the basis, we have

$$
\begin{aligned}
&\Big\| \sum_{i=1}^{n}\sum_{j=1}^{l} (\xi_s^{1,j},\phi_i)\phi_i - \sum_{i=1}^{n}\sum_{j=1}^{l} (\xi_s^{1,j},\tilde{\phi}_i)\tilde{\phi}_i \Big\|_H^2 \to 0, \\
&\Big\| \sum_{i=1}^{n} (\xi_{s-}^2,\phi_i)\phi_i - \sum_{i=1}^{n} (\xi_{s-}^2,\tilde{\phi}_i)\tilde{\phi}_i \Big\|_H^2 \to 0.
\end{aligned}
$$

On the other hand, similarly as Equation (6.3.5), we have the following boundedness,

$$
|I^{(n)}-\tilde{I}^{(n)}|_T \le 2\mathbb{E}^0 \Big\{ \sum_{j=1}^{l} \int_0^T \|\xi_s^{1,j}\|_H^2 ds + \int_0^T \|\xi_s^2\|_H^2 ds \Big\} \le 2T \cdot \Big\{ \sum_{j=1}^{l} |\xi^{1,j}|_T^2 + |\xi^2|_T^2 \Big\} < \infty.
$$

To sum up, dominated convergence implies Equation (6.3.6). □

Lemma 6.3 (Itô Isometry)**.** *We have the following properties for stochastic integral I which is defined by Equation* (6.3.3),

$$
\mathbb{E}^0 \Big\{ \|I_t\|_H^2 \Big\} = \mathbb{E}^0 \Big\{ \int_0^t \|\xi_s\|_{H^{l+1}}^2 ds \Big\}, \quad t\in[0,T]. \tag{6.3.7}
$$

Proof. On the one hand, we have the following convergence of $I^{(n)}$ which is defined by Equation (6.3.2),

$$
\mathbb{E}^0 \Big\{ \|I_t^{(n)}\|_H^2 \Big\} \to \mathbb{E}^0 \Big\{ \|I_t\|_H^2 \Big\}, \quad \text{as} \quad n\to\infty. \tag{6.3.8}
$$

In fact, by Lemma 6.2,

$$|I^{(n)} - I|_T \to 0.$$

By the definition of norm $|\cdot|_T$, it is equivalent to

$$\Big\{ \sup_{t\in[0,T]} \mathbb{E}^0 \|I_t^{(n)} - I_t\|_H^2 \Big\}^{\frac{1}{2}} \to 0.$$

Then, for $t \in [0,T]$, we get

$$\Big\{ \mathbb{E}^0 \|I_t^{(n)} - I_t\|_H^2 \Big\}^{1/2} \to 0.$$

Notice that

$$\Big| \mathbb{E}^0 \Big\{ \|I_t^{(n)}\|_H^2 \Big\}^{1/2} - \mathbb{E}^0 \Big\{ \|I_t\|_H^2 \Big\}^{1/2} \Big| \leq \Big\{ \mathbb{E}^0 \|I_t^{(n)} - I_t\|_H^2 \Big\}^{1/2} \to 0,$$

Equation (6.3.8) is obtained.

On the other hand, we have

$$\mathbb{E}^0 \Big\{ \|I_t^{(n)}\|_H^2 \Big\} = \mathbb{E}^0 \Big\{ \sum_{j=1}^{l} \int_0^t \Big[\sum_{i=1}^{n} (\xi_s^{1,j}, \phi_i)^2 \Big] ds + \int_0^t \Big[\sum_{i=1}^{n} (\xi_s^2, \phi_i)^2 \Big] ds \Big\} \tag{6.3.9}$$

$$\to \mathbb{E}^0 \Big\{ \int_0^t \|\xi_s\|_{H^{l+1}}^2 ds \Big\}, \tag{6.3.10}$$

where the equality follows from Itô Isometry and the convergence obtained with the same method we used in Lemma 6.2. Combining Equation (6.3.8) and (6.3.9) together, we obtain the desired result. □

Lemma 6.4. *Let I be the stochastic integral which is defined by Equation* (6.3.3). *Then,*

$$\Big\| \mathbb{E}^0 (I_t) \Big\|_H = 0, \quad t \in [0,T]. \tag{6.3.11}$$

Proof. Similarly, as with the proof of Lemma 6.3, we get

$$\Big\| \mathbb{E}^0 (I_t) \Big\|_H = \lim_{n\to\infty} \Big\| \mathbb{E}^0 (I_t^{(n)}) \Big\|_H, \quad t \in [0,T].$$

Notice that $\Big\| \mathbb{E}^0 (I_t^{(n)}) \Big\|_H = 0$, by Shreve (2004), Theorem 4.3.1, page 134, the desired result is obtained. □

With such a definition, the Zakai equation (5.2.1) can be rewritten as a mild equation in $\mathcal{N}^2(0,T;H)$

$$q_t = G_t^* q_0 + \int_0^t G_{t-\tau}^* (h^\top q_\tau) dM_\tau^1 + \int_0^t G_{t-\tau}^* \Big[(\lambda - 1) q_{\tau-} \Big] dM_\tau^2, \; t \in [0,T]. \tag{6.3.12}$$

Equations of this kind have been largely studied in literature, in particular the uniqueness of the solution in $\mathcal{N}^2(0,T;H)$ is well known.

6.4 Continuity theorem

The objective of this section, is to introduce a continuity theorem, which says the convergence of Galerkin approximation is guaranteed by the convergence of the deterministic operators of the corresponding SDEs.

This section is organized as follows: In Section 6.4.1, we first introduce a large class of linear SPDEs, in mild form, which includes the Zakai equation as a special example. Then, we introduce a decomposition of these linear SPDEs which are important for our main result. We also introduce a mapping which associates the deterministic operators to the unique solution of the mild equation. Finally, we show that the mapping is continuous. For lucidity, we postpone the proofs of the related formulas to Section 6.4.2.

6.4.1 Continuity theorem

We now introduce some spaces of operators.

Definition 6.2. Let $\mathcal{S}$ be the space of all C_0-semigroups of linear bounded operators from H to H such that, $\forall S \in \mathcal{S}$, there exists $\bar{S} \in \mathbb{R}^+$, which is not depending on the choice of S, such that

$$\sup_{t\in[0,T]} \|S_t\| \leq \bar{S}. \tag{6.4.1}$$

It is endowed with the topology of the uniform strong convergence on $[0,T]$. That is, a sequence $\{S^{(n)}\}$ in $\mathcal{S}$ converges to a C_0-semigroup $S \in \mathcal{S}$ if, for any $x \in H$,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \left\| (S_t^{(n)} - S_t)x \right\|_H = 0.$$

Let $\mathcal{U}$ be the space of linear bounded operators from H to H, endowed with the strong topology. That is a sequence $\{B_2^{(n)}\}$ in $\mathcal{U}$ converges to a bounded linear operator $B_2 \in \mathcal{U}$ if, for any $x \in H$,

$$\lim_{n\to\infty} \left\| (B_2^{(n)} - B_2)x \right\|_H = 0.$$

Let $\mathcal{U}^l$ be the space of linear bounded operators from H to H^l. More precisely, it is the space of all linear bounded operator B_1 of the following form

$$B_1 x = (B^{1,1}x, \dots, B^{1,l}x), \ \forall x \in H,$$

where $B^{1,i} : H \to H, i = 1, \dots, l$, are some bounded linear operators in $\mathcal{U}$. It is endowed with the strong topology. Then, a sequence $\{B_1^{(n)}\}$ in $\mathcal{U}^l$ converges to a bounded linear operator $B_1 \in \mathcal{U}^l$ if, for any $x \in H$,

$$\lim_{n\to\infty} \left\| (B_1^{(n)} - B_1)x \right\|_{H^l} = 0.$$

A large class of linear stochastic partial differential equations

In this section, a large class of linear stochastic partial differential equations are studied. They include the Zakai equation as a special example.

Let $\{S_t, 0 \le t \le T\} \in \mathcal{S}$ be a C_0-semigroup of bounded linear operators from H to H and denote by the same symbol the semigroup of operators on H^l obtained by applying S_t componentwise, for each $t \ge 0$. Let $B_1 \in \mathcal{U}^l$, $B_2 \in \mathcal{U}$, we consider the following equation in $\mathcal{N}^2(0,T;H)$, for any f in H,

$$\chi_t = S_t f + \int_0^t S_{t-\tau} B_1(\chi_\tau) dM_\tau^1 + \int_0^t S_{t-\tau} B_2(\chi_{\tau-}) dM_\tau^2, \quad t \in [0,T]. \tag{6.4.2}$$

Here, defined by Equation (6.3.1), M^1 is an l-dimensional martingale and M^2 is a one-dimensional martingale. Notice that Zakai equation (6.3.12) is a special case of Equation (6.4.2).

An important decomposition

The following decomposition of Equation (6.4.2) is the starting point to obtain the main theorem. First, we introduce a linear operator L on $\mathcal{N}^2(0,T;H)$ defined by, $\forall \chi \in \mathcal{N}^2(0,T;H)$,

$$(L\chi)(t) := \int_0^t S_{t-\tau} B_1(\chi_\tau) dM_\tau^1 + \int_0^t S_{t-\tau} B_2(\chi_{\tau-}) dM_\tau^2, \ \ t \in [0,T]. \tag{6.4.3}$$

We then have the following properties of L.

Lemma 6.5. *Let $S \in \mathcal{S}$, $B_1 \in \mathcal{U}^l$, $B_2 \in \mathcal{U}$. Then, L, defined by Equation* (6.4.3)*, is a bounded linear operator from $\mathcal{N}^2(0,T;H)$ to $\mathcal{N}^2(0,T;H)$. Moreover, there exists a constant*

$$\gamma = \gamma\Big(\bar{S}, \|B_1\|, \|B_2\|, T\Big) = \bar{S}\Big(\|B_1\|^2 + \|B_2\|^2\Big)^{\frac{1}{2}} \sqrt{T}, \tag{6.4.4}$$

such that, $\forall n \in \mathbb{N}$,

$$\|L^n\|^{\frac{1}{n}} \le \frac{\gamma}{(n!)^{\frac{1}{2n}}}. \tag{6.4.5}$$

Lemma 6.6. *Let $S \in \mathcal{S}$, $f \in H$. Denote by $\chi^{[0]}$ the (nonrandom) element,*

$$\chi_t^{[0]} := S_t f, \quad t \in [0,T].$$

Then, $\chi^{[0]} \in \mathcal{N}^2(0,T;H)$.

For proofs, see Section 6.4.2. Together with Equation (6.4.3), Equation (6.4.2) can be rewritten as the abstract equation in $\mathcal{N}^2(0,T;H)$

$$\chi = \chi^{[0]} + L\chi. \tag{6.4.6}$$

The estimate (6.4.5) shows that L is a quasinilpotent operator so that the following theorem is stated.

Theorem 6.7. *Let $S \in \mathcal{S}$, $B_1 \in \mathcal{U}^l$, $B_2 \in \mathcal{U}$, and $f \in H$. Then, Equation* (6.4.6) *has a unique solution in $\mathcal{N}^2(0,T;H)$,*

$$\chi = (I-L)^{-1}\chi^{[0]} := \sum_{i=0}^{\infty} L^i \chi^{[0]}, \tag{6.4.7}$$

and we have $(I-L)^{-1}: \mathcal{N}^2(0,T;H) \to \mathcal{N}^2(0,T;H)$ *is a bounded linear operator, that is there exists a constant*

$$\kappa = \kappa\Big(\bar{S}, \|B_1\|, \|B_2\|, T\Big) = \frac{2}{\sqrt{3}} e^{2\gamma^2} \text{ with } \gamma = \bar{S}\Big(\|B_1\|^2 + \|B_2\|^2\Big)^{\frac{1}{2}} \sqrt{T}, \tag{6.4.8}$$

such that

$$\|(I-L)^{-1}\| < \kappa,$$

and $|\chi|_T < \infty$.

For the proof, see Section 6.4.2.

A mapping

The solution of Equation (6.4.2) is unique. Therefore, there is mapping which assigns every quad (f, B_1, B_2, S) the unique solution of Equation (6.4.2). More precisely, the definition of the mapping is as follows.

Recalling that spaces $\mathcal{S}$, $\mathcal{U}^l$ and $\mathcal{U}$ are defined by Definition 6.2, $\forall S \in \mathcal{S}$, $B_1 \in \mathcal{U}^l$, $B_2 \in \mathcal{U}$, and $f \in H$, Theorem 6.7 guarantees the uniqueness of the solution of Equation (6.4.2). So we get the definition of the following map. The map F from $H \times \mathcal{U}^l \times \mathcal{U} \times \mathcal{S}$ into $\mathcal{N}^2(0,T;H)$ is defined by mapping (f, B_1, B_2, S) into the unique solution χ in $\mathcal{N}^2(0,T;H)$ of Equation (6.4.2). In other words,

$$\chi = F(f, B_1, B_2, S). \tag{6.4.9}$$

In the next section, we will show the map is continuous.

Continuity theorem

In order to study Galerkin approximation for equations of Zakai type, we use a continuity theorem to show map F is continuous. It is as follows:

Theorem 6.8 (Mean Square Convergence). *Map F, defined by Equation* (6.4.9), *is continuous. More precisely, let $\{f^{(n)}\}$ be a sequence in H converging to $f \in H$, $\{B_1^{(n)}\}$ be a sequence in $\mathcal{U}^l$ converging to $B_1 \in \mathcal{U}^l$, $\{B_2^{(n)}\}$ be a sequence in $\mathcal{U}$ converging to $B_2 \in \mathcal{U}$, $\{S^{(n)}\}$ be a sequence in $\mathcal{S}$ converging to $S \in \mathcal{S}$. Then,*

$$\Big|F(f^{(n)}, B_1^{(n)}, B_2^{(n)}, S^{(n)}) - F(f, B_1, B_2, S)\Big|_T \to 0, \quad \text{as} \quad n \to \infty.$$

Remark 6.2. *In fact, with some more work, we can obtain the continuity theorem w.r.t. norm $\|\cdot\|_T$. Notice that the stochastic convolution integral*

$$\int_0^t S_{t-\tau} B_1(\chi_\tau) dM_\tau^1 + \int_0^t S_{t-\tau} B_2(\chi_\tau) dM_\tau^2.$$

is not a martingale. And standard tools of the martingale theory, like the Burkholder-Davis-Gundy inequality, are not available. Therefore, to obtain the continuity theorem w.r.t. norm $\|\cdot\|_T$, we need the inequality for stochastic convolutions driven by a Gaussian process, see Hausenblas and Seidler (2001), and by a Poisson process.

However, for practical point of view, uniform bounded obtained by $\|\cdot\|_T$ is not necessary. We prefer norm $|\cdot|_T$.

Remark 6.3. *Our proof of Theorem 6.8 follows the idea of Germani and Piccioni (1984). And we proof Equation* (6.4.17) *with another method.*

6.4.2 Proofs

Proof of Lemma 6.5. The desired result is obtained, if we can show $\forall \chi \in \mathcal{N}^2(0,T;H)$, the mean square continuous of $L\chi$ and Equation (6.4.5). Since mean square continuous together with Equation (6.4.5) means that L is a bounded linear operator from $\mathcal{N}^2(0,T;H)$ to $\mathcal{N}^2(0,T;H)$.

First of all, $\forall \chi \in \mathcal{N}^2(0,T;H)$, we show the mean square continuous of $L\chi$, that is $\forall t$,

$$\mathbb{E}^0\|(L\chi)(t+\Delta)-(L\chi)(t)\|_H^2 \to 0, \quad \text{as} \quad \Delta \to 0. \tag{6.4.10}$$

We then show the above convergence is true as $\Delta \to 0^+$. Similarly, the convergence as $\Delta \to 0^-$ can be obtained. By the definition of L, see Equation (6.4.3), we get

$$\begin{aligned}(L\chi)(t+\Delta) =& \int_0^{t+\Delta} S_{t+\Delta-\tau}B_1(\chi_\tau)dM_\tau^1 + \int_0^{t+\Delta} S_{t+\Delta-\tau}B_2(\chi_{\tau-})dM_\tau^2\\
=& \int_0^{\Delta} S_{t+\Delta-\tau}B_1(\chi_\tau)dM_\tau^1 + \int_0^{\Delta} S_{t+\Delta-\tau}B_2(\chi_{\tau-})dM_\tau^2\\
&+ \int_\Delta^{t+\Delta} S_{t+\Delta-\tau}B_1(\chi_\tau)dM_\tau^1 + \int_\Delta^{t+\Delta} S_{t+\Delta-\tau}B_2(\chi_{\tau-})dM_\tau^2\\
=& \int_0^{\Delta} S_{t+\Delta-\tau}B_1(\chi_\tau)dM_\tau^1 + \int_0^{\Delta} S_{t+\Delta-\tau}B_2(\chi_{\tau-})dM_\tau^2\\
&+ \int_0^{t} S_{t-\tau}B_1(\chi_{\tau+\Delta})dM_\tau^1 + \int_0^{t} S_{t-\tau}B_2(\chi_{(\tau+\Delta)-})dM_\tau^2,\end{aligned}$$

then

$$\begin{aligned}(L\chi)(t+\Delta)-(L\chi)(t) =& \Big\{ \int_0^{\Delta} S_{t+\Delta-\tau}B_1(\chi_\tau)dM_\tau^1 + \int_0^{\Delta} S_{t+\Delta-\tau}B_2(\chi_{\tau-})dM_\tau^2 \Big\}\\
&+ \Big\{ \int_0^{t} S_{t-\tau}B_1(\chi_{\tau+\Delta}-\chi_\tau)dM_\tau^1 + \int_0^{t} S_{t-\tau}B_2(\chi_{(\tau+\Delta)-}-\chi_{\tau-})dM_\tau^2 \Big\}\\
:=& I_1 + I_2,\end{aligned}$$

correspondingly and consequently,

$$\mathbb{E}^0\|(L\chi)(t+\Delta)-(L\chi)(t)\|_H^2 = \mathbb{E}^0\|I_1+I_2\|_H^2 \le 2\mathbb{E}^0\|I_1\|_H^2 + 2\mathbb{E}^0\|I_2\|_H^2. \tag{6.4.11}$$

On the one hand, Itô Isometry, see Lemma 6.3, implies that

$$\begin{aligned}\mathbb{E}^0\|I_1\|_H^2 =& \mathbb{E}^0\Big\{ \int_0^{\Delta} \|S_{t+\Delta-\tau}B_1(\chi_\tau)\|_{H^l}^2 d\tau + \int_0^{\Delta} \|S_{t+\Delta-\tau}B_2(\chi_\tau)\|_H^2 d\tau \Big\}\\
\le& \bar{S}^2(\|B_1\|^2 + \|B_2\|^2) \int_0^{\Delta} \mathbb{E}^0\|\chi_\tau\|_H^2 d\tau\\
\le& \bar{S}^2(\|B_1\|^2 + \|B_2\|^2) \int_0^{\Delta} \sup_{s\in[0,T]} \mathbb{E}^0\|\chi_s\|_H^2 d\tau\\
=& \bar{S}^2(\|B_1\|^2 + \|B_2\|^2)|\chi|_T^2\Delta\\
\to& 0,\end{aligned} \tag{6.4.12}$$

as $\Delta \to 0^+$. On the other hand, apply Itô Isometry again,

$$\begin{aligned}\mathbb{E}^0\|I_2\|_H^2 =&\mathbb{E}^0\Big\{\int_0^t \|S_{t-\tau}B_1(\chi_{\tau+\Delta}-\chi_\tau)\|_{H^l}^2 d\tau + \int_0^t \|S_{t-\tau}B_2(\chi_{(\tau+\Delta)}-\chi_\tau)\|_H^2 d\tau\Big\}\\ \le&\bar{S}^2(\|B_1\|^2+\|B_2\|^2)\int_0^t \mathbb{E}^0\|\chi_{\tau+\Delta}-\chi_\tau\|_H^2 d\tau\\ \to&0, \end{aligned}\tag{6.4.13}$$

applying dominated convergence theorem together with $\forall\tau\in[0,T]$,

$$\mathbb{E}^0\|\chi_{\tau+\Delta}-\chi_\tau\|_H^2 \to 0, \quad \text{as} \quad \Delta\to 0^+,$$

and

$$\int_0^t \mathbb{E}^0\|\chi_{\tau+\Delta}-\chi_\tau\|_H^2 d\tau \le 2\int_0^T \mathbb{E}^0 \sup_{s\in[0,T]}\|\chi_s\|_H^2 d\tau \le 2T|\chi|_T^2 < \infty.$$

Combining Equation (6.4.11), (6.4.12) and (6.4.13), the desired convergence (6.4.10) is obtained. It remains to show Equation (6.4.5). Notice that

$$\begin{aligned}&(L^n\chi)(t)\\ =&\int_0^t S_{t-t_1}B_1\Big\{\sum_{j_2,\dots,j_n\in\{1,2\}}\Big[\int_0^{t_1} S_{t_1-t_2}B_{j_2}\Big(\dots\int_0^{t_{n-1}} S_{t_{n-1}-t_n}B_{j_n}(\chi_{t_n})dM_{t_n}^{j_n}\dots\Big)dM_{t_2}^{j_2}\Big]\Big\}dM_{t_1}^1\\ &+\int_0^t S_{t-t_1}B_2\Big\{\sum_{j_2,\dots,j_n\in\{1,2\}}\Big[\int_0^{t_1} S_{t_1-t_2}B_{j_2}\Big(\dots\int_0^{t_{n-1}} S_{t_{n-1}-t_n}B_{j_n}(\chi_{t_n})dM_{t_n}^{j_n}\dots\Big)dM_{t_2}^{j_2}\Big]\Big\}dM_{t_1}^2.\end{aligned}$$

By the definition of $|\cdot|_T$, we have

$$\begin{aligned}&|L^n\chi|_T\\ =&\sup_{t\in[0,T]}\Big\{\mathbb{E}^0\Big[\|(L^n\chi)(t)\|_H^2\Big]\Big\}^{\frac12}\\ \le&\Big\{\mathbb{E}^0\int_0^T\Big\|S_{T-t_1}B_1\Big[\sum_{j_2,\dots,j_n\in\{1,2\}}\int_0^{t_1} S_{t_1-t_2}B_{j_2}\Big(\dots\int_0^{t_{n-1}} S_{t_{n-1}-t_n}B_{j_n}(\chi_{t_n})dM_{t_n}^{j_n}\dots\Big)dM_{t_2}^{j_2}\Big]\Big\|_{H^l}^2 dt_1\\ &+\mathbb{E}^0\int_0^T\Big\|S_{T-t_1}B_2\Big[\sum_{j_2,\dots,j_n\in\{1,2\}}\int_0^{t_1} S_{t_1-t_2}B_{j_2}\Big(\dots\int_0^{t_{n-1}} S_{t_{n-1}-t_n}B_{j_n}(\chi_{t_n})dM_{t_n}^{j_n}\dots\Big)dM_{t_2}^{j_2}\Big]\Big\|_H^2 dt_1\Big\}^{\frac12}\\ \le&\Big\{\bar{S}^2\|B_1\|^2\int_0^T\mathbb{E}^0\Big\|\sum_{j_2,\dots,j_n\in\{1,2\}}\int_0^{t_1} S_{t_1-t_2}B_{j_2}\Big(\dots\int_0^{t_{n-1}} S_{t_{n-1}-t_n}B_{j_n}(\chi_{t_n})dM_{t_n}^{j_n}\dots\Big)dM_{t_2}^{j_2}\Big\|_H^2 dt_1\\ &+\bar{S}^2\|B_2\|^2\int_0^T\mathbb{E}^0\Big\|\sum_{j_2,\dots,j_n\in\{1,2\}}\int_0^{t_1} S_{t_1-t_2}B_{j_2}\Big(\dots\int_0^{t_{n-1}} S_{t_{n-1}-t_n}B_{j_n}(\chi_{t_n})dM_{t_n}^{j_n}\dots\Big)dM_{t_2}^{j_2}\Big\|_H^2 dt_1\Big\}^{\frac12}\\ \le&\Big\{\bar{S}^{2n}\Big(\|B_1\|^2+\|B_2\|^2\Big)\sum_{j_2,\dots,j_n\in\{1,2\}}\Big(\|B_{j_2}\|^2\dots\|B_{j_n}\|^2\Big)\int_0^T\int_0^{t_1}\dots\int_0^{t_{n-1}}\mathbb{E}^0\|\chi_{t_n}\|_H^2 dt_n,\dots,dt_1\Big\}^{\frac12}\\ \le&\Big\{\bar{S}^{2n}\Big(\|B_1\|^2+\|B_2\|^2\Big)^n\cdot|\chi|_T^2\cdot\int_0^T\int_0^{t_1}\dots\int_0^{t_{n-1}} dt_n,\dots,dt_1\Big\}^{\frac12}\\ \le&\bar{S}^n\Big(\|B_1\|^2+\|B_2\|^2\Big)^{\frac{n}{2}}\sqrt{\frac{T^n}{n!}}\cdot|\chi|_T.\end{aligned}$$

Then we have

$$\|L^n\| \leq \bar{S}^n \Big(\|B_1\|^2 + \|B_2\|^2 \Big)^{\frac{n}{2}} \frac{\sqrt{T}^n}{(n!)^{\frac{1}{2}}}.$$

Set $\gamma = \bar{S}\Big(\|B_1\|^2 + \|B_2\|^2 \Big)^{\frac{1}{2}} \sqrt{T}$, we obtain the desired result. □

Proof of Lemma 6.6. By definition of $\chi^{[0]}$ and norm $|\cdot|_T$,

$$|\chi^{[0]}|_T^2 = \sup_{t\in[0,T]} \mathbb{E}^0 \|\chi_t^{[0]}\|_H^2 = \sup_{t\in[0,T]} \|S_t f\|_H^2 \leq \bar{S}^2 \|f\|_H^2 < \infty.$$

Notice that S is a C_0-semigroup, the continuity follows from strongly continuous of C_0-semigroup, see Definition 4.1. □

Proof of Theorem 6.7. Define $\xi^n = \sum_{i=0}^n L^i \chi^{[0]}$, then, by Lemma 6.5 and 6.6, we have that $\{\xi^n, n \geq 0\}$ is a Cauchy sequence in $\mathcal{N}^2(0,T;H)$. In fact, for $n > m \geq 0$,

$$|\xi^n - \xi^m|_T = \Big| \sum_{i=m}^n L^i \chi^{[0]} \Big|_T \leq \sum_{i=m}^n \Big| L^i \chi^{[0]} \Big|_T \leq \sum_{i=m}^n \|L^i\| \cdot |\chi^{[0]}|_T \leq |\chi^{[0]}|_T \cdot \sum_{i=m}^n \|L^i\|.$$

Together with Formula (6.4.5),

$$\begin{aligned} |\xi^n - \xi^m|_T \leq & |\chi^{[0]}|_T \cdot \sum_{i=m}^n \frac{\gamma^i}{(i!)^{\frac{1}{2}}} \\ \leq & |\chi^{[0]}|_T \cdot \frac{\gamma^m}{(m!)^{\frac{1}{2}}} \sum_{i=0}^\infty \frac{\gamma^i}{(i!)^{\frac{1}{2}}} \\ \leq & |\chi^{[0]}|_T \cdot \frac{\gamma^m}{(m!)^{\frac{1}{2}}} \cdot \frac{2}{\sqrt{3}} e^{2\gamma^2}, \end{aligned}$$

the last inequality follows from Lemma A.3. Notice that $\frac{\gamma^m}{(m!)^{\frac{1}{2}}} \to 0$, as $m \to \infty$, $\{\xi^n\}$ is Cauchy. $\mathcal{N}^2(0,T;H)$ is complete, therefore, $\{\xi^n\}$ has a limit. That is $\exists \xi \in \mathcal{N}^2(0,T;H)$, such that $\xi = \sum_{i=0}^\infty L^i \chi^{[0]}$. Now, we show that ξ is the unique solution of Equation (6.4.6). We have

$$\Big| L\xi - L\xi^n \Big|_T = \Big| L(\xi - \xi^n) \Big|_T \leq \|L\| \cdot \Big| \xi - \xi^n \Big|_T \to 0,$$

where the equality follow from the linearity of L. The inequality equality follows from boundedness of L by Equation (6.4.5). That is

$$L\xi = \lim_{n\to\infty} L\xi^n.$$

Therefore,

$$\begin{aligned} \chi^{[0]} + L\xi &= \lim_{n\to\infty} \chi^{[0]} + L\xi^n = \lim_{n\to\infty} \chi^{[0]} + L\sum_{i=0}^n L^i \chi^{[0]} = \lim_{n\to\infty} \chi^{[0]} + \sum_{i=0}^n L^{i+1} \chi^{[0]} \\ &= \lim_{n\to\infty} \sum_{i=0}^{n+1} L^i \chi^{[0]} = \xi. \end{aligned}$$

It remains to prove uniqueness. Let η be a solution of Equation (6.4.6). Set $u := \xi - \eta$. It is suffice to show that $|u|_T = 0$. By Equation (6.4.6),

$$u_t = (Lu)(t) = \int_0^t S_{t-\tau} B_1(u_\tau) dM^1_\tau + \int_0^t S_{t-\tau} B_2(u_\tau) dM^2_\tau.$$

Then we have,

$$\begin{aligned}
|u|_t^2 &= \sup_{s\in[0,t]} \mathbb{E}^0 \Big\| (Lu)(s) \Big\|_H^2 \\
&= \sup_{s\in[0,t]} \left\{ \mathbb{E}^0 \int_0^s \Big\| S_{s-\tau} B_1(u_\tau) \Big\|_H^2 d\tau + \mathbb{E}^0 \int_0^s \Big\| S_{s-\tau} B_2(u_\tau) \Big\|_H^2 d\tau \right\} \\
&\le \|\bar{S}\|^2 \Big(\|B_1\|^2 + \|B_2\|^2 \Big) \int_0^t \mathbb{E}^0 \|u(\tau)\|_H^2 d\tau \\
&\le \|\bar{S}\|^2 \Big(\|B_1\|^2 + \|B_2\|^2 \Big) \int_0^t |u|_\tau^2 d\tau.
\end{aligned}$$

It follows from Gronwall's inequality that

$$|u|_t = 0, \quad t \in [0, T].$$

The uniqueness is obtained. By Lemma 6.5 and Lemma A.3,

$$\Big\| (I - L)^{-1} \Big\| \le \sum_{n=1}^\infty \frac{\gamma^n}{(n!)^{\frac{1}{2}}} \le \frac{2}{\sqrt{3}} e^{2\gamma^2} < \infty.$$

Therefore, by the definition of $\chi^{[0]}$,

$$|\chi|_T \le \|(I - L)^{-1}\| \cdot \Big| \chi^{[0]} \Big|_T \le \|(I - L)^{-1}\| \cdot \bar{S} \cdot \|f\|_H.$$

□

Proof of Theorem 6.8. Since $S^{(n)} \to S$, $B_1^{(n)} \to B_1$ and $B_2^{(n)} \to B_2$, by the uniform boundedness principle there exist $\bar{N}$ and a constant $\bar{\gamma}$ such that

$$\sup_{t\in[0,T], n\ge\bar{N}} \left\{ \|S_t\| \vee \|S_t^{(n)}\| \vee \|B_1\| \vee \|B_1^{(n)}\| \vee \|B_2\| \vee \|B_2^{(n)}\| \right\} \le \bar{\gamma}. \tag{6.4.14}$$

In the following, we considered $\left\{ (f^{(n)}, B_1^{(n)}, B_2^{(n)}, S^{(n)}) \right\}$ restricted to $n > \bar{N}$. We define

$$\chi := F(f, B_1, B_2, S), \quad \chi^{(n)} := F(f^{(n)}, B_1^{(n)}, B_2^{(n)}, S^{(n)}),$$

and

$$\begin{aligned}
\chi_t^{[0,(n)]} &:= S_t^{(n)} f^{(n)}, \\
(L^{(n)}\xi)(t) &:= \int_0^t S_{t-\tau}^{(n)} B_1^{(n)}(\xi_\tau) dM^1_\tau + \int_0^t S_{t-\tau}^{(n)} B_2^{(n)}(\xi_{\tau-}) dM^2_\tau, \quad \forall \xi \in \mathcal{N}^2(0, T; H).
\end{aligned}$$

By the definition of F, we have

$$\chi^{(n)} = \chi^{[0,(n)]} + L^{(n)}\chi^{(n)}, \quad \chi = \chi^{[0]} + L\chi.$$

So

$$\begin{aligned}\chi^{(n)}-\chi &= \chi^{[0,(n)]}-\chi^{[0]}+L^{(n)}\chi^{(n)}-L\chi\\ &= (\chi^{[0,(n)]}-\chi^{[0]})+L^{(n)}(\chi^{(n)}-\chi)+(L^{(n)}-L)\chi\end{aligned}$$

from which

$$\chi^{(n)}-\chi=(I-L^{(n)})^{-1}\Big\{(\chi^{[0,(n)]}-\chi^{[0]})+(L^{(n)}-L)\chi\Big\}. \tag{6.4.15}$$

Notice, by Theorem 6.7, there exists a constant $\kappa=\kappa(\bar{\gamma})$, such that

$$\Big\|(I-L^{(n)})^{-1}\Big\|\le\kappa. \tag{6.4.16}$$

Moreover,

$$\begin{aligned}\Big|\chi^{[0,(n)]}-\chi^{[0]}\Big|_T^2 &= \sup_{t\in[0,T]}\Big\|S_t^{(n)}f^{(n)}-S_tf\Big\|_H^2\\ &= \sup_{t\in[0,T]}\Big\|S_t^{(n)}f^{(n)}-S_t^{(n)}f+S_t^{(n)}f-S_tf\Big\|_H^2\\ &=2\sup_{t\in[0,T]}\Big\|S_t^{(n)}(f^{(n)}-f)\Big\|_H^2+2\sup_{t\in[0,T]}\Big\|(S_t^{(n)}-S_t)f\Big\|_H^2\\ &\le 2\bar{\gamma}^2\|f^{(n)}-f\|_H^2+2\sup_{t\in[0,T]}\Big\|(S_t^{(n)}-S_t)f\Big\|_H^2\\ &\to 0,\end{aligned}$$

where the convergence follows from $(f^{(n)},S^{(n)})$ convergence to (f,S). By Formula (6.4.15) the theorem is proved once it is shown that $(L^{(n)}-L)\chi$ goes to zero. That is

$$\Big|(L^{(n)}-L)\chi\Big|_T\to 0,\ \text{ as } n\to\infty. \tag{6.4.17}$$

By the definition of L, see Equation (6.4.3),

$$\begin{aligned}\Big[(L^{(n)}-L)\chi\Big](t)=&\int_0^t(S_{t-\tau}^{(n)}B_1^{(n)}-S_{t-\tau}B_1)\chi_\tau dM_\tau^1\\ &+\int_0^t(S_{t-\tau}^{(n)}B_2^{(n)}-S_{t-\tau}B_2)\chi_\tau dM_\tau^2.\end{aligned}$$

Then we have,

$$
\begin{aligned}
&\Big|(L^{(n)}-L)\chi\Big|_T^2\\
=&\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\Big\|\big[(L^{(n)}-L)\chi\big](t)\Big\|_H^2\Big\}\\
=&\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\int_0^t\Big\|(S^{(n)}_{t-\tau}B^{(n)}_1-S_{t-\tau}B_1)\chi_\tau\Big\|_{H^l}^2d\tau+\int_0^t\Big\|(S^{(n)}_{t-\tau}B^{(n)}_2-S_{t-\tau}B_2)\chi_\tau\Big\|_H^2d\tau\Big\}\\
=&\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\int_0^t\Big\|(S^{(n)}_{t-\tau}B^{(n)}_1-S^{(n)}_{t-\tau}B_1+S^{(n)}_{t-\tau}B_1-S_{t-\tau}B_1)\chi_\tau\Big\|_{H^l}^2d\tau\\
&+\int_0^t\Big\|(S^{(n)}_{t-\tau}B^{(n)}_2-S^{(n)}_{t-\tau}B_2+S^{(n)}_{t-\tau}B_2-S_{t-\tau}B_2)\chi_\tau\Big\|_H^2d\tau\Big\}\\
\le&2\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\int_0^t\Big\|(S^{(n)}_{t-\tau}B^{(n)}_1-S^{(n)}_{t-\tau}B_1)\chi_\tau\Big\|_{H^l}^2d\tau\Big\}+2\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\int_0^t\Big\|(S^{(n)}_{t-\tau}B_1-S_{t-\tau}B_1)\chi_\tau\Big\|_{H^l}^2d\tau\Big\}\\
&+2\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\int_0^t\Big\|(S^{(n)}_{t-\tau}B^{(n)}_2-S^{(n)}_{t-\tau}B_2)\chi_\tau\Big\|_H^2d\tau\Big\}+2\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\int_0^t\Big\|(S^{(n)}_{t-\tau}B_2-S_{t-\tau}B_2)\chi_\tau\Big\|_H^2d\tau\Big\}\\
:=&2E^1+2E^2+2E^3+2E^4,
\end{aligned}
$$

correspondingly. Here, the first equality follows from the definition of $|\cdot|_T$, the second equality follows from Itô Isometry. It remains to show that, as $n\to\infty$,

$$E^1\to 0,\quad E^2\to 0,\quad E^3\to 0,\quad E^4\to 0.$$

Since the convergence of E^3 and E^4 can be obtained similarly, we only give the proof of the convergence of E^1 and E^2 in the following. It is obtained by

$$
\begin{aligned}
E^1=&\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\int_0^t\Big\|S^{(n)}_{t-\tau}\Big(B^{(n)}_1-B_1\Big)(\chi_\tau)\Big\|_{H^l}^2d\tau\Big\}\\
\le&\bar{\gamma}^2\mathbb{E}^0\Big\{\int_0^T\Big\|\Big(B^{(n)}_1-B_1\Big)(\chi_\tau)\Big\|_{H^l}^2d\tau\Big\}\\
\to&0,
\end{aligned}
\tag{6.4.18}
$$

and

$$
\begin{aligned}
E^2\le&\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\int_0^t\sup_{\tau\le s\le T}\Big\|\Big(S^{(n)}_{s-\tau}-S_{s-\tau}\Big)B_1(\chi_\tau)\Big\|_{H^l}^2d\tau\Big\}\\
\le&\sup_{t\in[0,T]}\mathbb{E}^0\Big\{\int_0^T\sup_{\tau\le s\le T}\Big\|\Big(S^{(n)}_{s-\tau}-S_{s-\tau}\Big)B_1(\chi_\tau)\Big\|_{H^l}^2d\tau\Big\}\\
=&\mathbb{E}^0\Big\{\int_0^T\Big[\sup_{\tau\le s\le T}\Big\|\Big(S^{(n)}_{s-\tau}-S_{s-\tau}\Big)B_1(\chi_\tau)\Big\|_{H^l}^2\Big]d\tau\Big\}\\
\to&0,
\end{aligned}
\tag{6.4.19}
$$

where Equation (6.4.18) and (6.4.19) are obtained as follows. We have $\forall\tau\in[0,T]$, $\forall\omega\in\Omega$, by strong convergence of $\{B^{(n)}_1\}$, $\{B^{(n)}_2\}$, and uniform strong convergence of $\{S^{(n)}\}$,

$$\Big\|\Big(B^{(n)}_1-B_1\Big)(\chi_\tau)\Big\|_{H^l}^2\to 0,\ \sup_{\tau\le s\le T}\Big\|\Big(S^{(n)}_{s-\tau}-S_{s-\tau}\Big)B_1(\chi_\tau)\Big\|_{H^l}^2\to 0.$$

Moreover, we have the following boundedness,

$$\mathbb{E}^0\Big\{\int_0^T \Big\|\Big(B_1^{(n)} - B_1\Big)(\chi_\tau)\Big\|_{H^l}^2 d\tau \le 4\bar{\gamma}^2\mathbb{E}^0\int_0^T \|\chi_\tau\|_H^2 d\tau \le 4\bar{\gamma}^2 T\cdot|\chi|_T^2 < \infty,$$

and

$$\mathbb{E}^0\Big\{\int_0^T \Big[\sup_{\tau\le s\le T}\Big\|\Big(S_{s-\tau}^{(n)} - S_{s-\tau}\Big)B_1(\chi_\tau)\Big\|_{H^l}^2\Big]d\tau\Big\} \le 4\bar{\gamma}^4\mathbb{E}^0\int_0^T \|\chi_\tau\|_H^2 d\tau \le 4\bar{\gamma}^4 T\cdot|\chi|_T^2 < \infty.$$

The last inequality follows from $|\chi|_T < \infty$, by Theorem 6.7. Now, by dominated convergence theorem, we obtain Equation (6.4.18) and (6.4.19). The proof of the theorem is obtained. □

6.5 Galerkin approximation

We now study the convergence of the finite-dimensional Galerkin approximation for Equation (5.2.1).

Recall that P_n is a sequence of finite-dimensional orthogonal projections on H defined by Definition 6.1. For each n, Equation (6.1.1) is equivalent to the flowing ordinary stochastic differential equation

$$\begin{cases} dq_t^{(n)} = P_n\mathcal{A}^*P_nq_t^{(n)}dt + \sum_{i=1}^l P_nh^iP_nq_t^{(n)}dM_t^{1,i} + P_n(\lambda-1)P_nq_{t-}^{(n)}dM_t^2, \quad t\in[0,T] \\ q_0^{(n)} = P_nq_0. \end{cases}$$

It is equivalent to the following mild stochastic differential equation on $[0,T]$,

$$\begin{aligned} q_t^{(n)} =& \exp(P_n\mathcal{A}^*P_nt)(P_nq_0) + \int_0^t \exp\Big(P_n\mathcal{A}^*P_n(t-\tau)\Big)(P_nh^\top P_n)q_\tau^{(n)}dM_\tau^1 \\ &+ \int_0^t \exp\Big(P_n\mathcal{A}^*P_n(t-\tau)\Big)\Big(P_n(\lambda-1)P_n\Big)q_{\tau-}^{(n)}dM_\tau^2. \end{aligned} \tag{6.5.1}$$

Notice that, if Assumption 5.1 is fulfilled and $q_0\in H$, then Theorem 6.7 implies that the above equation has a unique solution in $\mathcal{N}^2(0,T;H)$.

We are presented with an important question when the solution converges to the solution χ_t of Equation (6.3.12), that is

$$\lim_{n\to\infty} F\Big(P_nq_0, P_nhP_n, P_n(\lambda-1)P_n, \exp(P_n\mathcal{A}^*P_n\cdot)\Big) = F(q_0,h,(\lambda-1),G^*)$$

in the norm defined by Equation (A.1.3). The answer is given by the following corollary. Similar results can be found in Germani and Piccioni (1984).

Corollary 6.9. *Assume Assumption 5.1 is fulfilled. Let $\{q_t^{(n)}\}_{t\in[0,T]}$ be the solution of Equation* (6.5.1)*. Then, the sequence of processes $\{q^{(n)}\}$ converges to the process $\{q_t\}_{t\in[0,T]}$, the solution of Equation* (6.3.12)*, for any choice of the initial state $q_0\in H$ if and only if for any $x\in H$,*

$$\lim_{n\to\infty}\sup_{t\in[0,T]}\Big\|\Big(\exp(P_n\mathcal{A}^*P_nt) - G_t^*\Big)x\Big\|_H = 0. \tag{6.5.2}$$

Proof. Notice that $\forall x \in H$, $P_n x \to x$, $P_n h P_n x \to hx$ and $P_n(\lambda - 1)P_n x \to (\lambda - 1)x$. In fact,

$$\begin{aligned} P_n h P_n x - hx =& (P_n h P_n x - P_n hx) + (P_n hx - hx) \\ =& P_n h(P_n x - x) + (P_n hx - hx) \to 0, \end{aligned}$$

by

$$\Big\| P_n h(P_n x - x) \Big\|_{H^l} \le \|h\|_\infty \|P_n x - x\|_H \to 0.$$

Then, the sufficiency is obtained by Theorem 6.8.

For necessity, the convergence of $q_t^{(n)}$ to q_t uniformly in mean square norm implies the uniform convergence of their mean vectors. To be precise, first we have

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \mathbb{E}^0 \|q_t^{(n)} - q_t\|_H \to 0.$$

Notice,

$$\mathbb{E}^0 \|q_t^{(n)} - q_t\|_H \ge \Big\| \mathbb{E}^0 (q_t^{(n)} - q_t) \Big\|_H,$$

we then get

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \Big\| \mathbb{E}^0 (q_t^{(n)} - q_t) \Big\|_H \to 0.$$

By Equation (6.5.1) (6.3.12), and Lemma 6.4, we obtain

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \mathbb{E}^0 \Big\| \exp(P_n \mathcal{A}^* P_n t) P_n x - G_t^* x \Big\|_H = 0,$$

for each $x \in H$. Moreover,

$$\begin{aligned} & \| \exp(P_n \mathcal{A}^* P_n t) - G_t^*) x\|_H \\ =& \| \exp(P_n \mathcal{A}^* P_n t) x - \exp(P_n \mathcal{A}^* P_n t) P_n x + \exp(P_n \mathcal{A}^* P_n t) P_n x - G_t^* x\|_H \\ =& \| \exp(P_n \mathcal{A}^* P_n t)(x - P_n x) + \exp(P_n \mathcal{A}^* P_n t) P_n x - G_t^* x\|_H \\ \le& \| \exp(P_n \mathcal{A}^* P_n t)(x - P_n x)\|_H + \| \exp(P_n \mathcal{A}^* P_n t) P_n x - G_t^* x\|_H. \end{aligned}$$

Notice, $x - P_n x \perp V_n$ and $\exp(P_n \mathcal{A}^* P_n t) y = 0$ for $y \perp V_n$, the necessity is obtained. □

Remark 6.4. *Applying Theorem 6.8 and Corollary 6.9 to the case $h = 0$ and $\lambda - 1 = 0$, Equation* (6.5.2) *guarantees that the convergence of the corresponding deterministic partial differential equation*

$$\begin{cases} dq_t^{(n)} = P_n \mathcal{A}^* P_n q_t^{(n)} dt, & t \in [0, T], \\ q_0^{(n)} = P_n q_0. \end{cases}$$

Therefore, Theorem 6.8 and Corollary 6.9 reduce the convergence problem for a stochastic partial differential equation to the corresponding deterministic one.

The necessary and sufficient conditions for Equation (6.5.2) to hold are obtained by means of the Trotter-Kato Theorem.

Theorem 6.10. *Suppose there exist M and γ such that for any n*

$$\| \exp(P_n \mathcal{A}^* P_n t) \leq M \exp(\gamma t)\|. \tag{6.5.3}$$

Then, each of the following conditions is equivalent to Equation (6.5.2)*:*

(i) For some $\alpha > \gamma$ and any $x \in H$ the (unique) solution y_n of the following equation in V_n

$$\alpha y_n - P_n \mathcal{A}^* P_n y_n = P_n x \tag{6.5.4}$$

converges as $n \to \infty$ to the unique solution y of the equation in H

$$\alpha y - \mathcal{A}^* y = x. \tag{6.5.5}$$

(ii) Let D be a core for $\mathcal{A}^*$. For any $x \in D$, there exist $x_n \in V_n$ such that $x_n \to x$, $P_n \mathcal{A}^* P_n x_n \to \mathcal{A}^* x$, as $n \to \infty$.

The uniform growth condition (6.5.3) is satisfied if there exists β such that $\|G_t\| \leq \exp(\beta t)$. In this case $M = 1$ and $\gamma = \beta$. And this is true if Assumption 5.1 holds, see Equation (5.1.3).

Proof. The result that Condition (i) is equivalent to Equation (6.5.2) is obtained by Germani and Piccioni (1987), Theorem 6.2, page 66. And the result that Condition (ii) is equivalent to Equation (6.5.2) is obtained by Ethier and Kurtz (1986), Theorem 6.1, page 28. □

Remark 6.5. *The second condition of Germani and Piccioni (1987), Theorem 6.2, page 66, is another equivalent condition for Equation* (6.5.2) *to hold. But condition (ii) of Theorem 6.10 is convenient to deal with.*

Assumption 6.1. *Assume $\cup_n V_n$ is dense in V, where V_n are defined by Definition 6.1.*

In the next Chapter, by Corollary 7.3, we will present a example for which this assumption holds. This is a sufficient condition for condition (i) of Theorem 6.10 to hold.

Theorem 6.11. *Assume Assumption 5.1 and 6.1 are fulfilled. Let $\{q_t^{(n)}\}_{t\in[0,T]}$ be the solution of Equation* (6.5.1)*. Then, the sequence of processes $\{q^{(n)}\}$ converges to the process q, the solution of Equation* (6.3.12)*, for any choice of the initial state $q_0 \in H$, as $n \to \infty$.*

Proof. Apply Corollary 6.9 and Theorem 6.10, it is enough to prove the first condition or the second condition in Theorem 6.10. And apply Assumption 6.1, the first one holds which is proved in Theorem 4, Germani and Piccioni (1984), page 420. □

Corollary 6.12. *Assume Assumption 5.1 and 6.1 are fulfilled. Let $\{q_t^{(n)}\}_{t\in[0,T]}$ be the solution of Equation* (6.5.1) *and $\{q_t\}_{t\in[0,T]}$ be the solution of Equation* (6.3.12)*. Then, for $q_0 \in H$ and for any φ bounded and square integrable, we have*

$$\sup_{t\in[0,T]} \mathbb{E}^0 \left| (q_t^{(n)}, \varphi) - (q_t, \varphi) \right|^2 \to 0, \quad \text{as } n \to \infty.$$

Proof. By Cauchy-Schwartz inequality,

$$\begin{aligned}\Big|(q_t^{(n)},\varphi)-(q_t,\varphi)\Big|^2 =&\Big\{\int_{\mathbb{R}^d}\varphi(x)[q_t^{(n)}(x)-q_t(x)]dx\Big\}^2\\ \leq&\int_{\mathbb{R}^d}\varphi(x)^2dx\int_{\mathbb{R}^d}[q_t^{(n)}(x)-q_t(x)]^2dx\\ =&\|\varphi\|_H^2\cdot\|q_t^{(n)}-q_t\|_H^2.\end{aligned}$$

Applying Theorem 6.11, we obtain the desired result. □

Corollary 6.13. *Assume Assumption 5.1 and 6.1 are fulfilled and $q_0\in H$. Let $\{q_t^{(n)}\}_{t\in[0,T]}$ be the solution of Equation (6.5.1) and $\{q_t\}_{t\in[0,T]}$ be the solution of Equation (6.3.12). Moreover, $\forall t\in[0,T]$,*

$$\mathbb{E}^0\Big[\frac{1}{(q_t,\mathbf{1})^2}\Big]<\infty,\quad \text{and}\quad \mathbb{E}^0\Big|(q_t^{(n)},\mathbf{1})-(q_t,\mathbf{1})\Big|^2\to 0,\ \text{as}\ n\to\infty.$$

Then, for any φ bounded and square integrable, we have

$$\mathbb{E}^0\Big|(p_t^{(n)},\varphi)-(p_t,\varphi)\Big|^2\to 0,\ \text{as}\ n\to\infty.$$

Proof. To minimize, we define $p_t(\varphi):=(p_t,\varphi)$ and similarly for $p_t^{(n)}(\varphi)$, $q_t^{(n)}(\varphi)$, $q_t(\varphi)$, $q_t^{(n)}(\mathbf{1})$ $q_t(\mathbf{1})$. Notice that

$$\begin{aligned}\Big|p_t^{(n)}(\varphi)-p_t(\varphi)\Big| =&\Big|\frac{q_t^{(n)}(\varphi)}{q_t^{(n)}(\mathbf{1})}-\frac{q_t(\varphi)}{q_t(\mathbf{1})}\Big|\\ =&\Big|\frac{q_t^{(n)}(\varphi)}{q_t^{(n)}(\mathbf{1})q_t(\mathbf{1})}[q_t(\mathbf{1})-q_t^{(n)}(\mathbf{1})]+\frac{1}{q_t(\mathbf{1})}[q_t^{(n)}(\varphi)-q_t(\varphi)]\Big|\\ \leq&\Big|\frac{q_t^{(n)}(\varphi)}{q_t^{(n)}(\mathbf{1})q_t(\mathbf{1})}[q_t(\mathbf{1})-q_t^{(n)}(\mathbf{1})]\Big|+\Big|\frac{1}{q_t(\mathbf{1})}[q_t^{(n)}(\varphi)-q_t(\varphi)]\Big|\\ \leq&\|\varphi\|_\infty\Big|\frac{1}{q_t(\mathbf{1})}(q_t(\mathbf{1})-q_t^{(n)}(\mathbf{1}))\Big|+\Big|\frac{1}{q_t(\mathbf{1})}(q_t^{(n)}(\varphi)-q_t(\varphi))\Big|.\end{aligned}$$

Then Cauchy-Schwarz inequality implies

$$\begin{aligned}\mathbb{E}^0\Big|p_t^{(n)}(\varphi)-p_t(\varphi)\Big|^2\leq&\|\varphi\|_\infty\Big\{\mathbb{E}^0[q_t(\mathbf{1})]^{-2}\mathbb{E}^0|q_t(\mathbf{1})-q_t^{(n)}(\mathbf{1})|^2\Big\}^{\frac{1}{2}}\\ &+\Big\{\mathbb{E}^0[q_t(\mathbf{1})]^{-2}\mathbb{E}^0|q_t^{(n)}(\varphi)-q_t(\varphi)|^2\Big\}^{\frac{1}{2}}.\end{aligned}$$

Applying Corollary 6.12, we obtain the desired result. □

Chapter 7

Galerkin approximation

In the previous chapter, we show the convergence of the Galerkin approximation in mean square norm. Again, we have a reasonable question concerning the rate of convergence. In this chapter, we design an adaptive Galerkin approximation with a basis of Hermite polynomials and present numerical examples to illustrate the effectiveness of the proposed method. In simulation study, we compare the proposed method with particle methods. We show the Galerkin approximation converges well.

This chapter is organized as follows: Recall that, in Section 6.1, we show the Galerkin approximation for Zakai equation (5.2.1) is the solution of a finite-dimensional SDE (6.1.9). Now, in Section 7.1, we introduce the analytically and numerical solutions for SDE (6.1.9). The solution of the Zakai equation can be constructed by the Galerkin method using any suitable set of basis functions from Hilbert space. It is possible to choose a complete set of basis functions, like Gaussian series and Hermite functions. These are introduced in Section 7.2. In order to increase the effectiveness of the Galerkin approximation, we design an adaptive Galerkin approximation in Section 7.3. We introduce the motivation, the adaptive Galerkin approximation with normal basis, the one with a basis of Hermite polynomials and discuss the advantage of using Hermite functions. For sake of lucidity, we postpone proofs of the related results for Hermite polynomials to Section 7.4. Examples and the corresponding simulation results are presented in Section 7.5.

7.1 SDEs

In Section 6.1, we have shown that by the Galerkin approximation, the solution of Zakai equation (5.2.1) is approximated by a finite combination of orthogonal series. The solution of the Zakai equation is then approximated by the solution of finite dimensional ordinary SDE (6.1.9).

For its solution, we have two possibilities: a numerical approach or an explicit analytical solution. For the numerical approach, one may proceed with time discretisation, according to Euler-Maruzama scheme, or splitting-up approximation which will be introduced in Section 7.1.2. Following the idea of Frey and Runggaldier (2010), Proposition 5.4, we have the following result for the analytical solution.

7.1.1 Analytical solution

When we keep the assumptions and notations in Section 6.1, then, the following proposition provides the analytical solution for Equation (6.1.9).

Proposition 7.1. *For sake of simplicity, further assume that $\{e_i\}$ is an orthonormal basis, then*[1] $A^{(n)} = I_n$. *Consequently, the solution of Equation* (6.1.9) *is given by* $\Upsilon^{(n)}(t) := L_t^{(n)}\tilde{L}_t^{(n)}\mathbf{r}_t^{(n)}$, *where, for* $t \in [0,T]$,

$$
\begin{aligned}
L_t^{(n)} &:= \exp\Big\{\sum_{k=1}^{l}\Big(C^{(n),k}Z_t^k - \frac{1}{2}(C^{(n),k})^2 t\Big)\Big\},\\
\tilde{L}_t^{(n)} &:= \{I_n + (L_{\tau_{Y_t}-}^{(n)})^{-1}G^{(n)}L_{\tau_{Y_t}-}^{(n)}\}\dots\{I_n + (L_{\tau_2-}^{(n)})^{-1}G^{(n)}L_{\tau_2-}^{(n)}\}\{I_n + (L_{\tau_1-}^{(n)})^{-1}G^{(n)}L_{\tau_1-}^{(n)}\}I_n,
\end{aligned}
$$

and $\mathbf{r}^{(n)}$ *is the solution of the following ordinary linear differential equation*

$$
\begin{cases}
\frac{d\mathbf{r}_t^{(n)}}{dt} = (\tilde{L}_t^{(n)})^{-1}(L_t^{(n)})^{-1}[B^{(n)} - G^{(n)}]L_t^{(n)}\tilde{L}_t^{(n)}\mathbf{r}_t^{(n)}, \quad t \in [0,T],\\
\mathbf{r}_0^{(n)} = \mathbf{q_0^{(n)}}.
\end{cases}
\tag{7.1.1}
$$

Proof. For sake of simplicity, we consider the case of $l = 1$, and we write A, B, C, G, Υ, L, $\tilde{L}$, $\mathbf{r}$, $\mathbf{q}_0$ instead of $A^{(n)}$, $B^{(n)}$, $C^{(n)}$, $G^{(n)}$, $\Upsilon^{(n)}$, $L^{(n)}$, $\tilde{L}^{(n)}$, $\mathbf{r}^{(n)}$, $\mathbf{q}_0^{(n)}$.
First of all, it is easy to see that $L_0 = I_n$, $\tilde{L}_0 = I_n$, and by the definition of Υ,

$$\Upsilon(0) = L_0\tilde{L}_0\mathbf{r}_0 = I_nI_n\mathbf{q}_0 = \mathbf{q}_0.$$

Compared to Equation (6.1.9), it remains to show that Υ has the desired dynamics. The key step is to derive the dynamics of $\Upsilon(t) = L_t\tilde{L}_t\mathbf{r}_t$ by Itô's formula. For this, we need the dynamic of $L, \tilde{L}$ first, noting that the dynamic of $\mathbf{r}$ is given. From the definition of $\tilde{L}$, we have

$$d\tilde{L}_t = (L_{t-})^{-1}GL_{t-}\tilde{L}_{t-}dY_t. \tag{7.1.2}$$

Now, we derive the dynamics of L. For any $t \in \mathbb{R}^+$, matrices CZ_t and $-\frac{1}{2}C^2t$ commute (meaning that $(CZ_t)(-\frac{1}{2}C^2t) = (-\frac{1}{2}C^2t)(CZ_t)$), so, by properties of exponential of sums,

$$L_t = e^{CZ_t - \frac{1}{2}C^2t} = e^{CZ_t}\cdot e^{-\frac{1}{2}C^2t}. \tag{7.1.3}$$

One the one hand, the matrix-valued power series

$$e^{CZ_t} = \sum_{k=0}^{\infty}\frac{(CZ_t)^k}{k!} = \sum_{k=0}^{\infty}\frac{C^kZ_t^k}{k!}$$

has infinite radius of convergence. So, each component is differentiable with derivative given by term by term differentiation:

$$de^{CZ_t} = \sum_{k=1}^{\infty}\frac{C^kZ_t^{k-1}}{(k-1)!}dZ_t + \frac{1}{2}\sum_{k=2}^{\infty}\frac{C^kZ_t^{k-2}}{(k-2)!}dt = \Big[CdZ_t + \frac{1}{2}C^2dt\Big]e^{CZ_t}. \tag{7.1.4}$$

[1] I_n denotes the n-by-n identity matrix .

On the other hand, by the expression for the derivative of matrix exponent,

$$de^{-\frac{1}{2}C^2t} = -\frac{1}{2}C^2e^{-\frac{1}{2}C^2t}dt. \tag{7.1.5}$$

Combining Equation (7.1.3), (7.1.4) and (7.1.5), applying Itô formula, we have

$$dL_t = (de^{CZ_t})e^{-\frac{1}{2}C^2t} + e^{CZ_t}(de^{-\frac{1}{2}C^2t}) = CL_tdZ_t. \tag{7.1.6}$$

It remains to derive dynamics of Υ. First, by Equation (7.1.2) and (7.1.1), applying Itô formula, we obtain

$$\begin{aligned} d(\tilde{L}_t\mathbf{r}_t) =&(d\tilde{L}_t)\mathbf{r}_{t-} + \tilde{L}_td\mathbf{r}_t \\ =&(L_{t-})^{-1}GL_{t-}\tilde{L}_{t-}\mathbf{r}_{t-}dY_t + (L_t)^{-1}(B-G)L_t\tilde{L}_t\mathbf{r}_tdt \\ =&L_t^{-1}BL_t\tilde{L}_t\mathbf{r}_tdt + L_{t-}^{-1}GL_{t-}\tilde{L}_{t-}\mathbf{r}_{t-}d(Y_t-t). \end{aligned} \tag{7.1.7}$$

Further, combining Equation (7.1.6) and (7.1.7) together, applying Itô formula again, we have

$$\begin{aligned} d\Upsilon(t) &= d\Big[L_t\cdot(\tilde{L}_t\mathbf{r}_t)\Big] \\ &= dL_t(\tilde{L}_t\mathbf{r}_t) + L_td(\tilde{L}_t\mathbf{r}_t) \\ &= CL_t\tilde{L}_t\mathbf{r}_tdZ_t + BL_t\tilde{L}_t\mathbf{r}_tdt + GL_{t-}\tilde{L}_{t-}\mathbf{r}_{t-}d(Y_t-t) \\ &= B\Upsilon(t)dt + C\Upsilon(t)dZ_t + G\Upsilon(t-)d(Y_t-t). \end{aligned}$$

The desired result is obtained. $\square$

7.1.2 Numerical methods for SDEs

By the Galerkin method, we obtain the discretisation of the Zakai equation with respect to the space parameter. The solution of Zakai equation (5.2.1), which is a SPDE, is approximated by the solution of SDE (6.1.9). In order to solve the SDE numerically, discretisation of time variable is required. The objective of this section is to introduce some numerical methods discretising the time variable.

Let $0 = t_0 < t_1 < \cdots < t_k < \cdots < t_L = T$ be a uniform partition of the interval $[0,T]$ with time step $\Delta = t_k - t_{k-1} = \frac{T}{L}$. Assume that $\{Z_{t_k}\}$, $\{Y_{t_k}\}$, $k = 0,1,\ldots,L$, is the sampled trajectories of Z and Y at discrete times. Recall that Z and Y are observation processes which are introduced in Section 2.2 in detail.

$q_{t_k}^{(n,\Delta)}$, the approximation of the unnormalized conditional density at discrete times ($t_k, k = 0,1,\ldots,L$) is as follows,

$$q_{t_k}^{(n,\Delta)}(x) := \sum_{i=0}^{n-1}\psi_{k,i}^{(n,\Delta)}e_i(x),$$

where $\{e_i\}$ is the basis in the Galerkin approximation and $\Upsilon_k^{(n,\Delta)} := (\psi_{k,0}^{(n,\Delta)}, \psi_{k,1}^{(n,\Delta)}, \ldots, \psi_{k,n-1}^{(n,\Delta)})^\top$ are the Fourier coefficients and the numerical approximation to the true solution of Equation (6.1.9) which is obtained by Euler-Maruyama approximation or splitting-up approximation. We introduce as follows.

Euler-Maruyama approximation

In this section, we apply the Euler-Maruyama method to obtain the numerical solution of the SDE (6.1.9). It is a generalization of the Euler method, for ordinary differential equations to stochastic differential equations. It is a numerical method by discretisation of the time variable. For a general introduction and reference, see, for instance McLachlan and Krishnan (1997).

Algorithm 7.1. *The Euler-Maruyama approximation to the true solution of Equation* (6.1.9), *is obtained as follows.*

i Compute the coefficients in the Galerkin approximation $A^{(n)} := (a_{ij})_{n\times n}$, $B^{(n)} := (b_{ij})_{n\times n}$, $C^{(n),k} := (c^k_{ij})_{n\times n}$, $G^{(n)} := (g_{ij})_{n\times n}$, $\mathbf{q}_0^{(n)} = [q_{00}, \dots, q_{0(n-1)}]^T$ *by Equation* (6.1.3), (6.1.4), (6.1.5), (6.1.6) *and* (6.1.7).

ii Get the starting point by $\Upsilon_0^{(n,\Delta)} = (A^{(n)})^{-1}\mathbf{q}_0^{(n)}$.

iii Solve the ordinary SDE (6.1.9) *numerically to obtain* $\{\Upsilon_k^{(n,\Delta)}, i = 1, 2, \dots, L\}$. *For* $k = 0, 1, \dots, L-1$,

$$\begin{aligned}\Upsilon_{k+1}^{(n,\Delta)} =& \Upsilon_k^{(n,\Delta)} + (A^{(n)})^{-1}\Big\{B^{(n)}\Upsilon_k^{(n,\Delta)}\Delta + \sum_{m=1}^{l} C^{(n),m}\Upsilon_k^{(n,\Delta)}(Z^m_{t_{k+1}} - Z^m_{t_k}) \\ &+ G^{(n)}\Upsilon_k^{(n,\Delta)}(Y_{t_{k+1}} - Y_{t_k} - \Delta)\Big\}.\end{aligned}$$

Splitting-up approximation

The objective of this section is to apply splitting-up approximation to solve Equation (6.1.9) numerically.

Applying the Runge-Kutta method to obtain $\Upsilon^{(n)}$ with Proposition 7.1, one should inverse $\tilde{L}^{(n)}$ and $L^{(n)}$ at each time point. Not only is it time consuming, it also leads to an additive computation error. Instead, we consider the splitting-up algorithm.

Splitting-up approximations are a numerical method for both SDE and SPDE based on semi-group theory introduced in Bensoussan, Glowinski, and Rascanu (1990) and further developed in Le Gland (1992). Using this approach, one can decompose the original equation into both a stochastic and deterministic one. This is much simpler than handling the original problem.

Now we generalize this method to Equation (6.1.9) which is with Lévy noise. For sake of simplicity, we assume $l = 1$. Assume $A^{(n)} = I_n$. We consider the following decomposition of Equation (6.1.9),

$$d\Upsilon^{(n)}(t) = (B^{(n)} - G^{(n)})\Upsilon^{(n)}(t-)dt + C^{(n)}\Upsilon^{(n)}(t-)dZ_t + G^{(n)}\Upsilon^{(n)}(t-)dY_t.$$

The splitting-up approximation $\Upsilon_k^{(n,\Delta)}$, for $k = 0, 1, \dots, L$, is obtained by $\Upsilon_0^{(n,\Delta)} = \Upsilon^{(n)}(0)$ and, for $k = 1, 2, \dots, L$,

$$\Upsilon_k^{(n,\Delta)} = R_{t_{k-1},t_k} Q_{t_{k-1},t_k} P_\Delta \Upsilon_{k-1}^{(n,\Delta)}.$$

Here $\{P_t : 0 \leq t \leq T\}$, $\{Q_{st} : 0 \leq s \leq t \leq T\}$ and $\{R_{st} : 0 \leq s \leq t \leq T\}$ denote the solution operators corresponding to the equations

$$\begin{aligned} d\Upsilon_1(t) =&(B^{(n)} - G^{(n)})\Upsilon_1(t)dt, \\ d\Upsilon_2(t) =&C^{(n)}\Upsilon_2(t)dZ_t, \\ d\Upsilon_3(t) =&G^{(n)}\Upsilon_3(t-)dY_t, \end{aligned}$$

respectively.

We have the following algorithm of the splitting-up approximation.

Algorithm 7.2. *Each interval $[t_{k-1}, t_k]$ the transition from $\Upsilon_{k-1}^{(n,\Delta)}$ to $\Upsilon_k^{(n,\Delta)}$ is divided into the following three steps:*

- *The first step, called the prediction step, consists in solving the following Focker-Planck equation, which is a deterministic differential equation, for the time interval $[t_{k-1}, t_k]$,*

$$\begin{cases} d\Upsilon_1(t) = (B^{(n)} - G^{(n)})\Upsilon_1(t)dt, \\ \Upsilon_1(t_{k-1}) = \Upsilon_{k-1}^{(n,\Delta)}. \end{cases} \tag{7.1.8}$$

  The Fokker-Planck equation can be solved analytically. Its solution at time t_k is the prior estimate defined as $\bar{\Upsilon}_k^{(n,\Delta)}$,

$$\bar{\Upsilon}_k^{(n,\Delta)} := \Upsilon_1(t_k) = \exp\left\{(B^{(n)} - G^{(n)})\Delta\right\}\Upsilon_{k-1}^{(n,\Delta)}.$$

- *The second step, called the correction step 1, using the new observation Z, consists in solving the following SDE, for the time interval $[t_{k-1}, t_k]$,*

$$\begin{cases} d\Upsilon_2(t) = C^{(n)}\Upsilon_2(t)dZ_t, \\ \Upsilon_2(t_{k-1}) = \bar{\Upsilon}_k^{(n,\Delta)}. \end{cases} \tag{7.1.9}$$

  The SDE can be solved analytically. Its solution at time t_k is the second prior estimate defined as $\tilde{\Upsilon}_k^{(n,\Delta)}$,

$$\tilde{\Upsilon}_k^{(n,\Delta)} := \Upsilon_2(t_k) = \exp\left\{C^{(n)}(Z_{t_k} - Z_{t_{k-1}}) - \frac{1}{2}(C^{(n)})^2\Delta\right\}\bar{\Upsilon}_k^{(n,\Delta)}.$$

- *The third step, called the correction step 2, use the new observation Y, consists in solving the following SDE, for the time interval $[t_{k-1}, t_k]$,*

$$\begin{cases} d\Upsilon_3(t) = G^{(n)}\Upsilon_3(t-)dY_t, \\ \Upsilon_3(t_{k-1}) = \tilde{\Upsilon}_k^{(n,\Delta)}. \end{cases} \tag{7.1.10}$$

  The SDE can be solved analytically. Its solution at time t_k is the estimate $\Upsilon_{\Delta,k}^{(n)}$,

$$\Upsilon_k^{(n,\Delta)} := \Upsilon_3(t_k) = (I_n + G^{(n)})^{(Y_{t_k} - Y_{t_{k-1}})}\tilde{\Upsilon}_k^{(n,\Delta)}. \tag{7.1.11}$$

We have the following convergence results for continuous case. Recall that λ is the jump intensity of observation Y. Assume that $\lambda \equiv 1$, then $G^{(n)} = 0$. That is Equation (6.1.9) driven only by continuous noise. Then it is shown by Le Gland (1992), Proposition 3.1, under suitable regularity assumption on the coefficients, the approximating process $\{\Upsilon_k^{(n,\Delta)}\}_{k=0,1,\ldots,L}$ obtained as above , converges to $\{\Upsilon^{(n)}(t)\}_{t\in[0,T]}$, the solution of the Equation (6.1.9) with error, for $k = 0, 1, \ldots, L$,

$$\left\{\mathbb{E}^0|\Upsilon^{(n)}(t_k) - \Upsilon_k^{(n,\Delta)}|^2\right\}^{\frac{1}{2}} \leq c\sqrt{\Delta},$$

where c is a constant. The convergence can be obtained similarly for SDEs driven by Lévy noise.

7.2 Basis functions

The basis functions control the quality of the Galerkin approximation and in order to compute the Galerkin approximation of the unnormalized conditional density, we must first choose the basis.

In Section 7.2.1, we introduce Gaussian series and discuss the advantages of using it. In Section 7.2.2, we introduce Hermite functions and discuss the advantages of using Hermite polynomials.

7.2.1 Gaussian series

Ahmed and Radaideh (1997) use Gaussian series in the Galerkin approximation. The definition for the family of function, $\{e_i(x), x \in \mathbb{R}^d, i = 1, 2, \ldots\}$, is as follows

$$e_i(x) = e(x, m_i, B_i) := \frac{1}{\sqrt{(2\pi)^n \det(B_i)}} e^{-\frac{1}{2}(x-m_i)' B_i^{-1}(x-m_i)}, \quad x \in \mathbb{R}^d,$$

parameterized by m_i, B_i where $m_i \in \mathbb{R}^d$, $B_i \in \mathbb{R}^{d\times d}$ be any positive symmetric matrices, $m_i \neq m_j$, for $i \neq j$, $\det(B_i)$ is the determinant of B_i, recalling that d is the dimension of the state process X. The system of functions $\{e_i(x), x \in \mathbb{R}, i = 1, 2, \ldots\}$ is linearly independent and complete in H, see Theorem 3, Ahmed and Radaideh (1997).

Using the Galerkin approximation with a basis of Gaussian series, the conditional mean $\pi_t(X)$ and its associated error covariance matrix $P_t = \mathbb{E}\Big\{\Big(X_t - \pi_t(X)\Big)\Big(X_t - \pi_t(X)\Big)'\Big|\mathcal{F}_t^{Z,Y}\Big\}$ can be calculated as, for $t \in [0, T]$,

$$\begin{aligned}
\pi_t(X) = \mathbb{E}[X_t|\mathcal{F}_t^{Z,Y}] &\cong \frac{\sum_{i=1}^n \psi_i^{(n)}(t) m_i}{\sum_{i=1}^n \psi_i^{(n)}(t)},\\
P_t &\cong \frac{\sum_{i=1}^n \psi_i^{(n)}(t)\Big[B_i + \Big(\pi_t(X) - m_i\Big)\Big(\pi_t(X) - m_i\Big)'\Big]}{\sum_{i=1}^n \psi_i^{(n)}(t)},
\end{aligned}$$

where $\{\psi_i^{(n)}(t)\}$, $i = 1, 2, \ldots, n$, are the Fourier coefficients obtained by solving Equation (6.1.9). This is an advantage of using Gaussian series. In fact, the above equations are obtained as follows, consider $d = 1$, notice that $\{e_i\}$ are Gaussian series and, for $i = 1, 2, \ldots$,

$$\int_{\mathbb{R}} e_i(x)dx = 1, \quad \int_{\mathbb{R}} x e_i(x)dx = m_i, \quad \int_{\mathbb{R}} (x - m_i)^2 e_i(x)dx = B_i.$$

Then, by Kallianpur-Striebel formula (2.4.8), recalling that by Galerkin approximation $q_t(x) \approx q_t^{(n)}(x) = \sum_{i=1}^n \psi_i^{(n)}(t) e_i(x)$, we get

$$\begin{aligned}
\mathbb{E}[X_t|\mathcal{F}_t^{Z,Y}] \cong& \frac{\int_{\mathbb{R}} x q_t^{(n)}(x)dx}{\int_{\mathbb{R}} q_t^{(n)}(x)dx} = \frac{\int_{\mathbb{R}} x[\sum_{i=1}^n \psi_i^{(n)}(t) e_i(x)]dx}{\int_{\mathbb{R}}[\sum_{i=1}^n \psi_i^{(n)}(t) e_i(x)]dx}\\
=& \frac{\sum_{i=1}^n \psi_i^{(n)}(t)[\int_{\mathbb{R}} x e_i(x)dx]}{\sum_{i=1}^n \psi_i^{(n)}(t)[\int_{\mathbb{R}} e_i(x)dx]} = \frac{\sum_{i=1}^n \psi_i^{(n)}(t) m_i}{\sum_{i=1}^n \psi_i^{(n)}(t)},
\end{aligned}$$

and similarly,

$$\mathbb{E}\Big[\Big(X_t - \pi_t(X)\Big)^2\Big|\mathcal{F}_t^{Z,Y}\Big] \simeq \frac{\int_{\mathbb{R}}[x-\pi_t(X)]^2 q_t^{(n)}(x)dx}{\int_{\mathbb{R}} q_t^{(n)}(x)dx}$$
$$= \frac{\sum_{i=1}^n \psi_i^{(n)}(t)\Big[B_i + \Big(\pi_t(X) - m_i\Big)^2\Big]}{\sum_{i=1}^n \psi_i^{(n)}(t)},$$

where the last equality follows from

$$\int_{\mathbb{R}}[x-\pi_t(X)]^2 e_i(x)dx = \int_{\mathbb{R}}\Big\{(x-m_i)^2 + 2xm_i - m_i^2 - 2x\pi_t(X) + \pi_t(X)^2\Big\}e_i(x)dx$$
$$= B_i + [\pi_t(X) - m_i]^2.$$

7.2.2 Hermite expansion

The Hermite polynomials are classical orthogonal polynomial sequences. Hermite polynomials are used to derive expressions for the moments of univariate and multivariate normal distributions, see for example Willink (2005). Hermite expansions are used to compute the moments of the approximated density, see for example Singer (2006). We take the Hermite polynomials as the basis functions in the Galerkin approximation and show its efficiency.

This section is first organized with an introduction to the definition and properties of the Hermite polynomials. Then, we build a complete orthonormal basis of $L^2(\mathbb{R})$ from the Hermite polynomials and show the advantages of using the basis in the Galerkin approximation.

Hermite polynomials

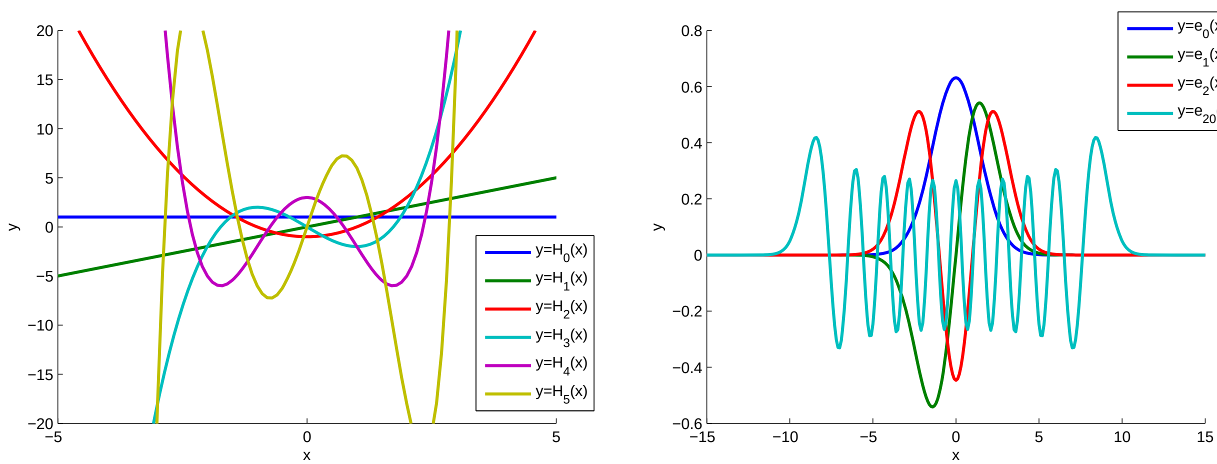

Figure 7.1: Hermite polynomials $\{H_i\}$ and orthonomal basis $\{e_i\}$.

The definition and well-known properties of the Hermite polynomials, and the details can be found in, for instance, Courant and Hilbert (1968). We show the definition first and then, we introduce the orthogonality and completeness of the Hermite polynomials. These properties help us to built a complete orthonormal basis of $L^2(\mathbb{R})$. Finally, we introduce the recurrence relation of the Hermite polynomials. This property helps us to compute the coefficients of the Galerkin approximation quickly.

The Hermite polynomials are classical orthogonal polynomial sequence. The Hermite polynomials $H_i(x)$ are defined by, for $i = 0, 1, 2, \ldots$,

$$H_i(x) = (-1)^i e^{x^2/2} \frac{d^i}{dx^i} e^{-x^2/2}, \quad x \in \mathbb{R}. \tag{7.2.1}$$

$H_i(x)$ is an ith-degree polynomial with leading coefficient 1 for $i = 0, 1, 2, 3, \ldots$. The first few of these polynomials are given by, for $x \in \mathbb{R}$,

$$\begin{aligned}
H_0(x) &= 1, \\
H_1(x) &= x, \\
H_2(x) &= x^2 - 1, \\
H_3(x) &= x^3 - 3x, \\
H_4(x) &= x^4 - 6x^2 + 3, \\
H_5(x) &= x^5 - 10x^3 + 15x, \\
H_6(x) &= x^6 - 15x^4 + 45x^2 - 15, \\
H_7(x) &= x^7 - 21x^5 + 105x^3 - 105x, \\
H_8(x) &= x^8 - 28x^6 + 210x^4 - 420x^2 + 105, \\
H_9(x) &= x^9 - 36x^7 + 378x^5 - 1260x^3 + 945x, \\
H_{10}(x) &= x^{10} - 45x^8 + 630x^6 - 3150x^4 + 4725x^2 - 945.
\end{aligned}$$

Inversely, it is easy to see any power of x can be represented by the Hermite polynomials, for $x \in \mathbb{R}$,

$$\begin{aligned}
1 &= H_0(x), \\
x &= H_1(x), \\
x^2 &= H_2(x) + 1, \\
x^3 &= H_3(x) + 3H_1(x), \\
x^4 &= H_4(x) + 6H_2(x) + 3, \\
x^5 &= H_5(x) + 10H_3(x) + 15H_1(x), \\
x^6 &= H_6(x) + 15H_4(x) + 45H_2(x) + 15, \\
x^7 &= H_7(x) + 21H_5(x) + 105H_3(x) + 105H_1(x), \\
x^8 &= H_8(x) + 28H_6(x) + 210H_4(x) + 420H_2(x) + 105, \\
x^9 &= H_9(x) + 36H_7(x) + 378H_5(x) + 1260H_3(x) + 945H_1(x), \\
x^{10} &= H_{10}(x) + 45H_8(x) + 630H_6(x) + 3150H_4(x) + 4725H_2(x) + 945.
\end{aligned}$$

Now, let ϑ^i_k be the coefficient at x^k of $H_i(x)$ and ι^i_k be the coefficient at H_k of x^i, $i = 0, 1, 2, \ldots$,

$k = 0, 1, \dots, i$, we then have the following expression for $x \in \mathbb{R}$, $i = 0, 1, 2, \dots$,

$$H_i(x) = \sum_{k=0}^{i} \vartheta_k^i x^k, \tag{7.2.2}$$

$$x^i = \sum_{k=0}^{i} \iota_k^i H_k(x). \tag{7.2.3}$$

The Hermite polynomials form a complete orthogonal set on the interval $-\infty < x < \infty$ with respect to the weighting function ϕ, see Courant and Hilbert (1968),

$$\int_{\infty}^{\infty} H_i(x) H_j(x) \phi(x) dx = i! \delta_{i,j}, \tag{7.2.4}$$

with weighting function

$$\phi(x) := (2\pi)^{-\frac{1}{2}} \exp(-\frac{x^2}{2}), \quad x \in \mathbb{R}, \tag{7.2.5}$$

where $i!$ stands for the factorial of i and

$$\delta_{i,j} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

is the Kronecker delta.

In addition, the sequence of Hermite polynomials satisfies the following recurrence relations, for $x \in \mathbb{R}$, $i = 1, 2, \dots$,

$$H_{i+1}(x) = x H_i(x) - i H_{i-1}(x), \tag{7.2.6}$$

$$H_i'(x) = i H_{i-1}(x). \tag{7.2.7}$$

Expressions (7.2.2), (7.2.3), (7.2.6), and (7.2.7) are useful to derive the coefficients of the Galerkin approximation.

Hermite expansion

In this part, we build a complete orthonormal basis of $L^2(\mathbb{R})$ from the Hermite polynomials first and show the advantages of using the basis in the Galerkin approximation.

We define a family of function, $\{e_i(x), x \in \mathbb{R}, i = 0, 1, 2, \dots\}$ from the Hermite polynomials, as follows

$$e_i(x) := \sqrt{\frac{\phi(x)}{i!}} H_i(x), \quad x \in \mathbb{R}, \tag{7.2.8}$$

with weighting function ϕ given by Equation (7.2.5) and $H_i(x)$ are Hermite polynomials defined by Equation (7.2.1). By the orthogonality and completeness of the Hermite polynomials, see Equation (7.2.4), Hermite functions $\{e_i(x), x \in \mathbb{R}, i = 0, 1, 2, \dots\}$ form a complete orthonormal basis of $L^2(\mathbb{R})$.

Then, q_t, the density function of the unnormalized conditional distribution of the state process X given past observation of Z and Y of Model (2.2.9) with $d = 1$, can be expanded as

$$q_t(x) = \sum_{i=0}^{\infty} \psi_i(t)\, e_i(x), \quad t \in [0, T], \quad x \in \mathbb{R},$$

and the Fourier coefficients are given by, for $i = 0, 1, \ldots$,

$$\psi_i(t) := \int_{-\infty}^{\infty} e_i(x)\, q_t(x) dx, \quad t \in [0, T].$$

Based on the recurrence relations of Hermite polynomials, we have the following results for Hermite functions.

Lemma 7.2. *Sequence $\{e_i(x), x \in \mathbb{R}, i = 0, 1, 2, \ldots\}$, defined by Equation* (7.2.8)*, satisfies, for $x \in \mathbb{R}$, $i = 1, 2, \ldots$,*

$$xe_i(x) = \sqrt{i+1}\, e_{i+1}(x) + \sqrt{i}\, e_{i-1}(x), \tag{7.2.9}$$

$$e_i'(x) = -\frac{\sqrt{i+1}}{2}\, e_{i+1}(x) + \frac{\sqrt{i}}{2}\, e_{i-1}(x), \tag{7.2.10}$$

and

$$xe_0(x) = e_1(x), \quad e_0'(x) = -\frac{1}{2}\, e_1(x).$$

In what follows, we show the convergence of Galerkin approximation with a basis of Hermite functions.

Corollary 7.3. *Assume that Assumption 5.1 is fulfilled. Let $\{e_i\}$, defined by Equation* (7.2.8)*, is the basis used in the Galerkin approximation defined by Definition 6.1. Let $\{q_t^{(n)}\}_{t\in[0,T]}$ be the solution of Equation* (6.5.1)*. Then, the sequence of processes $\{q^{(n)}\}$ converges to the process q, the solution of Equation* (6.3.12)*, for any choice of the initial state $q_0 \in H$, as $n \to \infty$.*

There are some advantages of using Hermite functions in the Galerkin approximation. One advantage, as matter of fact, is the conditional mean and the conditional error can be expressed directly by the Fourier coefficients.

As we introduced in Section 6.1, for a natural n, by the Galerkin approximation with $\{e_i\}$ as a basis we have the following n-dimensional approximation of the conditional density

$$q_t(x) \approx q_t^{(n)}(x) = \sum_{i=0}^{n-1} \psi_i^{(n)}(t)\, e_i(x), \quad t \in [0, T], \quad x \in \mathbb{R}, \tag{7.2.11}$$

where $\{\psi_i^{(n)}(t)\}$ are the expansion coefficients obtained by solving Equation (6.1.9). Moreover, by the approximation, the conditional mean and its associated error can be calculated as follows, for the proof, see Section 7.4. For sake of short, define constants, for $i = 0, 1, \ldots, n-1$,

$$r_{0i}^{(n)} := \frac{1}{\sqrt{i!}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^k, \quad r_{1i}^{(n)} := \frac{1}{\sqrt{i!}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^{k+1}, \quad r_{2i}^{(n)} := \frac{1}{\sqrt{i!}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^{k+2}, \tag{7.2.12}$$

recall that ϑ_k^i and ι_0^k are the coefficients of linear combinations (7.2.2) and (7.2.3). And define n-dimensional vectors $\mathbf{r}_0^{(n)} := (r_{00}^{(n)}, r_{01}^{(n)}, \ldots, r_{0(n-1)}^{(n)})^\top$, $\mathbf{r}_1^{(n)} := (r_{10}^{(n)}, r_{11}^{(n)}, \ldots, r_{1(n-1)}^{(n)})^\top$, $\mathbf{r}_2^{(n)} := (r_{20}^{(n)}, r_{21}^{(n)}, \ldots, r_{2(n-1)}^{(n)})^\top$. Then, for $0 \le t \le T$,

$$\pi_t(X) := \mathbb{E}[X_t | \mathcal{F}_t^{Z,Y}] \cong \sqrt{2} \frac{(\mathbf{r}_1^{(n)})^\top \Upsilon^{(n)}(t)}{(\mathbf{r}_0^{(n)})^\top \Upsilon^{(n)}(t)}, \tag{7.2.13}$$

$$\mathbb{E}\Big[\big(X_t - \pi_t(X)\big)^2 \Big| \mathcal{F}_t^{Z,Y}\Big] \cong 2 \frac{(\mathbf{r}_2^{(n)})^\top \Upsilon^{(n)}(t)}{(\mathbf{r}_0^{(n)})^\top \Upsilon^{(n)}(t)} - [\pi_t(X)]^2. \tag{7.2.14}$$

Notice, the components of $\mathbf{r}_0^{(n)}$, $\mathbf{r}_1^{(n)}$, $\mathbf{r}_2^{(n)}$ are constants, which are independent of X, Z and Y. They can be computed before hand. Therefore, the conditional mean and its associated error can be calculated directly by the expansion coefficients. This is one advantage of using Hermite functions.

Another advantage is the coefficients in the Galerkin approximation can be computed explicitly for some special cases. We can see this in the following examples for Model (2.2.9). For the proof of related formulas, see Section 7.4.

Example 7.1. *Suppose, for simplicity, the state process $\{X_t\}_{t\in[0,T]}$, continuous observation $\{Z_t\}_{t\in[0,T]}$ are given by the following scale linear Gaussian equations, for $t \in [0,T]$,*

$$\begin{cases} X_t = X_0 + \int_0^t bX_s ds + \int_0^t \sigma dV_s, \\ Z_t = \int_0^t hX_s ds + B_t, \end{cases}$$

where $b, h \in \mathbb{R}$ are constants, and $\sigma \in \mathbb{R}^+$ is a positive constant. X_0 is normally distributed with mean μ_0 and variance σ_0^2. V and B are independent standard one-dimensional Brownian motions. Another observation $\{Y_t\}_{t\in[0,T]}$ is a stochastic Poisson process with intensity λX_t^2, where $\lambda \in \mathbb{R}^+$ is a positive constant. Then the coefficients in the Galerkin approximation, which are defined by Equation (6.1.3), (6.1.4), (6.1.5), (6.1.6), (6.1.7), *are as follows, for $i = 0, 1, \ldots, n-1$, $j = 0, 1, \ldots, n-1$,*

$$a_{ji} = \delta_{i,j}, \tag{7.2.15}$$

$$\begin{aligned} b_{ji} =& (-\frac{b}{2} + \frac{\sigma^2}{8})\sqrt{(j+2)(j+1)}\delta_{i,j+2} - (\frac{b}{2} + \frac{\sigma^2(2j+1)}{8})\delta_{i,j} \\ &+ (\frac{b}{2} + \frac{\sigma^2}{8})\sqrt{j(j-1)}\delta_{i,j-2}, \end{aligned} \tag{7.2.16}$$

$$c_{ji} = h\sqrt{j}\delta_{i+1,j} + h\sqrt{j+1}\delta_{i-1,j}, \tag{7.2.17}$$

$$g_{ji} = \lambda\Big[\sqrt{j(j-1)}\delta_{i+2,j} + (2j+1)\delta_{i,j} + \sqrt{(j+2)(j+1)}\delta_{i-2,j}\Big] - \delta_{i,j}, \tag{7.2.18}$$

$$q_{0j} = \frac{1}{\sqrt{j!}}\sum_{k=0}^{j} \vartheta_k^j \sum_{i=0}^{k} C_k^i c_1^i c_2^{(k-i)} \iota_0^i, \quad \text{with } c_1 = \sqrt{\frac{2\sigma_0^2}{2+\sigma_0^2}}, \quad c_2 = \frac{2\mu_0}{2+\sigma_0^2}. \tag{7.2.19}$$

Example 7.2. *Suppose that state process X which is a geometric Brownian motion and continuous observation Z are given by the following linear equations, for $t \in [0,T]$,*

$$\begin{cases} X_t = X_0 + \int_0^t bX_s ds + \int_0^t \sigma X_s dV_s, \\ Z_t = \int_0^t hX_s ds + B_t, \end{cases}$$

and jump observation $\{Y_t\}_{t\in[0,T]}$ is a stochastic Poisson process with intensity λX_t. Here $b, h \in \mathbb{R}$ are constants, and $\sigma, \lambda \in \mathbb{R}^+$ are positive constants. V and B are independent standard one-dimensional Brownian motions. $X_0 \sim \ln N(\mu_0, \sigma_0^2)$ is distributed log-normally with parameters μ_0 and σ_0. Then the coefficients in the Galerkin approximation which are defined by Equation

(6.1.3), (6.1.4), (6.1.5), (6.1.6), (6.1.7), *are as follows, for* $i = 0, 1, \ldots, n-1$, $j = 0, 1, \ldots, n-1$,

$$
\begin{aligned}
a_{ji} =& \delta_{i,j}, \\
b_{ji} =& \frac{\sigma^2}{8}\sqrt{(j+4)(j+3)(j+2)(j+1)}\delta_{i,j+4} + (-\frac{b}{2} + \frac{\sigma^2}{2})\sqrt{(j+2)(j+1)}\delta_{i,j+2} \\
& - [\frac{b}{2} + \frac{\sigma^2}{8}(2j^2 + 2j - 1)]\delta_{i,j} + \frac{b}{2}\sqrt{j(j-1)}\delta_{i,j-2} + \frac{\sigma^2}{8}\sqrt{j(j-1)(j-2)(j-3)}\delta_{i,j-4}, \\
c_{ji} =& h\sqrt{j}\delta_{i+1,j} + h\sqrt{j+1}\delta_{i-1,j}, \\
g_{ji} =& \lambda\sqrt{j}\delta_{i+1,j} + \lambda\sqrt{j+1}\delta_{i-1,j} - \delta_{i,j}.
\end{aligned}
$$

Here q_{0j}, $j = 0, 1, \ldots, n-1$ *can not be obtained analytically, and they can be computed numerically.*

Example 7.3 (Cox-Ingersoll-Ross model)**.** *The Cox-Ingersoll-Ross model for the state process* X *is*

$$dX_t = (\alpha - \beta X_t)dt + \sigma\sqrt{X_t}dV_t, \quad t \in [0, T], \tag{7.2.20}$$

where α, β, *and* σ *are positive constants.* X_0 *is normally distributed with mean* μ_0 *and variance* σ_0^2. *The advantage of CIR model is that the state in the model does not become negative. If* X_t *reaches zero, the term multiplying* dV_t *vanishes and the positive drift term* αdt *in Equation* (7.2.20) *drives the state back into positive territory.*

Continuous observation Z *is given by the following linear equation*

$$Z_t = \int_0^t hX_s ds + B_t, \quad t \in [0, T],$$

and jump observation Y *is a stochastic Poisson process with intensity* λX_t. *Here* $h \in \mathbb{R}$, $\lambda \in \mathbb{R}^+$ *are constants.* V *and* B *are independent stand Brownian motions. Then the coefficients in the Galerkin approximation which are defined by Equation* (6.1.3), (6.1.4), (6.1.5), (6.1.6), (6.1.7), *are as follows, for* $i = 0, 1, \ldots, n-1$, $j = 0, 1, \ldots, n-1$,

$$
\begin{aligned}
a_{ji} =& \delta_{i,j}, \\
b_{ji} =& \frac{\sigma^2}{8}\sqrt{(j+3)(j+2)(j+1)}\delta_{i,j+3} + \frac{\beta}{2}\sqrt{(j+2)(j+1)}\delta_{i,j+2} - [\frac{\alpha}{2} + \frac{\sigma^2}{8}(j-1)]\sqrt{j+1}\delta_{i,j+1} \\
& + \frac{\beta}{2}\delta_{i,j} + [\frac{\alpha}{2} - \frac{\sigma^2}{8}(j+2)]\sqrt{j}\delta_{i,j-1} - \frac{\beta}{2}\sqrt{j(j-1)}\delta_{i,j-2} + \frac{\sigma^2}{8}\sqrt{j(j-1)(j-2)}\delta_{i,j-3}, \\
c_{ji} =& h\sqrt{j}\delta_{i+1,j} + h\sqrt{j+1}\delta_{i-1,j}, \\
g_{ji} =& \lambda\sqrt{j}\delta_{i+1,j} + \lambda\sqrt{j+1}\delta_{i-1,j} - \delta_{i,j}, \\
q_{0j} =& \frac{1}{\sqrt{j!}}\sum_{k=0}^{j}\vartheta_k^j\sum_{i=0}^{k} C_k^i c_1^i c_2^{(k-i)}\iota_0^i, \quad \text{with } c_1 = \sqrt{\frac{2\sigma_0^2}{2+\sigma_0^2}}, \quad c_2 = \frac{2\mu_0}{2+\sigma_0^2}.
\end{aligned}
$$

Remark 7.1. *As a conclusion, the advantages of the Galerkin approximation with a basis of Hermite polynomials are:*

i *The conditional mean and its associated error can be calculated in terms of the expansion coefficients, see Equation* (7.2.13) *and* (7.2.14).

ii The sequence of Hermite polynomials satisfies the recurrence relations, therefore $A^{(n)}$, $B^{(n)}$, $C^{(n)}$, $G^{(n)}$ *and* $\mathbf{q}_0^{(n)}$, *which are the coefficients in Galerkin approximation, can be computed explicitly, for the case of* b, σ^2, h *and* λ, *which are the coefficients of Model* (2.2.9), *are taking the following forms, for* $N = 0, 1, 2, \cdots$,

$$\sum_{i=0}^{N} c_i x^i e^{-a_i x^2 + b_i x}, \; c_i, b_i \in \mathbb{R}, \; a_i \in \mathbb{R}^+, \; i = 0, 1, \cdots N. \tag{7.2.21}$$

This is a large class of functions, including a lot of interesting models, see Example 7.1, 7.2 and 7.3.

7.3 The adaptive Galerkin approximation

There is difficulty for the Galerkin approximation similar to that of the particle filter. After a few steps , the majority of the Fourier coefficients are close to zero, and all the weights tend to concentrate on a very few coefficients. This reduces the effectiveness of the Galerkin approximation. In this section, we design an adaptive Galerkin approximation to overcome this. For the sake of lucidity, we postpone the proof of related results of this section in Section 7.4.

First, in Section 7.3.1, we introduce the motivation for this adaptive method. Then, in Section 7.3.2, we introduce the adaptive Galerkin approximation with normal basis. In general, this method provides a better estimator then the normal Galerkin approximation, but it is really time consuming. Finally, in Section 7.3.3, we introduce the adaptive Galerkin approximation for a large class of models with a basis of Hermite polynomials which is more efficient.

7.3.1 Introduction

In most nonlinear filtering problems, and in particular when the observation noise is small, the conditional density is well-localized in some small region of the state space and generally could not be predicted in advance. In the following, we present an example to show that, in this case, the effective of the approximation depends on the locations of the density.

Example 7.4. *In this example, we take* $\{e_i\}$ *as the basis of Hermite polynomials defined by Equation* (7.2.8). *We present the corresponding results in Figure 7.2 and 7.3.*

For a fixed real number μ *and a positive number* σ, *let* $\xi \sim N(\mu, \sigma^2)$ *be a normal random variable with mean* μ, *variance* σ^2 *and density* $p(\mu, \sigma, x)$, $x \in \mathbb{R}$. *For* $n \in \mathbb{N}$, *we use* n *basis functions* $\{e_i, i = 0, \ldots, n-1\}$ *to approximate the density* $p(\mu, \sigma, x)$. *Then, we obtain* $\hat{p}(\mu, \sigma, x)$, *which is the approximated density, defined by*

$$\hat{p}(\mu, \sigma, x) := \sum_{i=0}^{n-1} \Big(p(\mu, \sigma, \cdot), e_i \Big) e_i.$$

And we obtain $\hat{\mu}(\mu, \sigma)$, *which is the mean of the approximated density and* $\hat{\sigma}(\mu, \sigma)$, *which is the square root of the variance of the approximated density,*

$$\hat{\mu}(\mu, \sigma) = \int_{\mathbb{R}} x \hat{p}(\mu, \sigma, x) dx,$$
$$\hat{\sigma}(\mu, \sigma) = \Big\{ \int_{\mathbb{R}} \Big(x - \hat{\mu}(\mu, \sigma) \Big)^2 \hat{p}(\mu, \sigma, x) dx \Big\}^{\frac{1}{2}}.$$

To compare the difference between the estimated distribution and the true one, we compute the following terms. For fixed σ, we compute

$$\mu \mapsto d_1^\sigma(\mu) := \hat{\mu}(\mu,\sigma) - \mu, \tag{7.3.1}$$
$$\mu \mapsto d_2^\sigma(\mu) := \hat{\sigma}(\mu,\sigma) - \sigma, \tag{7.3.2}$$
$$\mu \mapsto d_3^\sigma(\mu) := \left\{ \int_{\mathbb{R}} \Big[\hat{p}(\mu,\sigma,x) - p(\mu,\sigma,x)\Big]^2 dx \right\}^{\frac{1}{2}}, \tag{7.3.3}$$

and we show the figures of that μ maps to $d_1^\sigma(\mu)$, μ maps to $d_2^\sigma(\mu)$, and μ maps to $d_3^\sigma(\mu)$ in Figure 7.2.

For fixed μ, we compute

$$\sigma \mapsto d_1^\mu(\sigma) := \hat{\mu}(\mu,\sigma) - \mu, \tag{7.3.4}$$
$$\sigma \mapsto d_2^\mu(\sigma) := \hat{\sigma}(\mu,\sigma) - \sigma, \tag{7.3.5}$$
$$\sigma \mapsto d_3^\mu(\sigma) := \left\{ \int_{\mathbb{R}} \Big[\hat{p}(\mu,\sigma,x) - p(\mu,\sigma,x)\Big]^2 dx \right\}^{\frac{1}{2}}, \tag{7.3.6}$$

and we show the results of that σ maps to $d_1^\mu(\sigma)$, σ maps to $d_2^\mu(\sigma)$, and σ maps to $d_3^\mu(\sigma)$ in Figure 7.3.

From the results, we see that the difference of the estimated result and the true one varies significantly with respect to difference μ and σ. We obtain a good approximation when μ is close to 0 and σ^2 is close to 2. We obtain bad approximation otherwise. The reason is that, for weighting function ϕ of the basis function $\{e_i\}$, which is defined by Equation (7.2.5),

$$\frac{\int_{\mathbb{R}} x\sqrt{\phi(x)}dx}{\int_{\mathbb{R}} \sqrt{\phi(x)}dx} = \frac{\int_{\mathbb{R}} xe^{-\frac{x^2}{4}}dx}{\int_{\mathbb{R}} e^{-\frac{x^2}{4}}dx} = 0,$$
$$\frac{\int_{\mathbb{R}} x^2\sqrt{\phi(x)}dx}{\int_{\mathbb{R}} \sqrt{\phi(x)}dx} = \frac{\int_{\mathbb{R}} x^2e^{-\frac{x^2}{4}}dx}{\int_{\mathbb{R}} e^{-\frac{x^2}{4}}dx} = \frac{2\sqrt{2}\int_{\mathbb{R}} x^2e^{-\frac{x^2}{2}}dx}{\sqrt{2}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}dx} = 2.$$

In other words, the mean and the variance of the normalized weighting function is 0 and 2.

From this example, we see that the approximation depends on the location of the density which generally could not to be predicted in advance. This is our motivation to change the center and scale of the basis functions adaptively.

7.3.2 The adaptive Galerkin approximation

In the following, we introduce the basic computation steps of the adaptive Galerkin approximation for general case. Here, we assume that $d = 1$.

Let $\{e_i\}$ be a basis of Hilbert space H, made of elements of $\{v \in V | \mathcal{A}v \in H\}$. Recall, in the normal Galerkin approximation introduced in Section 6.1, we take $\{e_i\}$ as the basis which is independent of time. Now, we introduce the adaptive Galerkin approximation. We take adaptive basis $\{\bar{e}_i^t, t \in [0,T], i = 0,1\ldots\}$ introduced below at each time t instead of $\{e_i\}$. And $\{\bar{e}_i^t, t \in [0,T], i = 0,1\ldots\}$ are obtained by adapted shift and scale of $\{e_i\}$ at some time points.

We start by choosing parameters n, N and M. We use n basis functions in the Galerkin approximation. Let $0 \le t_0 < \ldots < t_k < \ldots < t_{NM} = T$ be a uniform partition of the interval

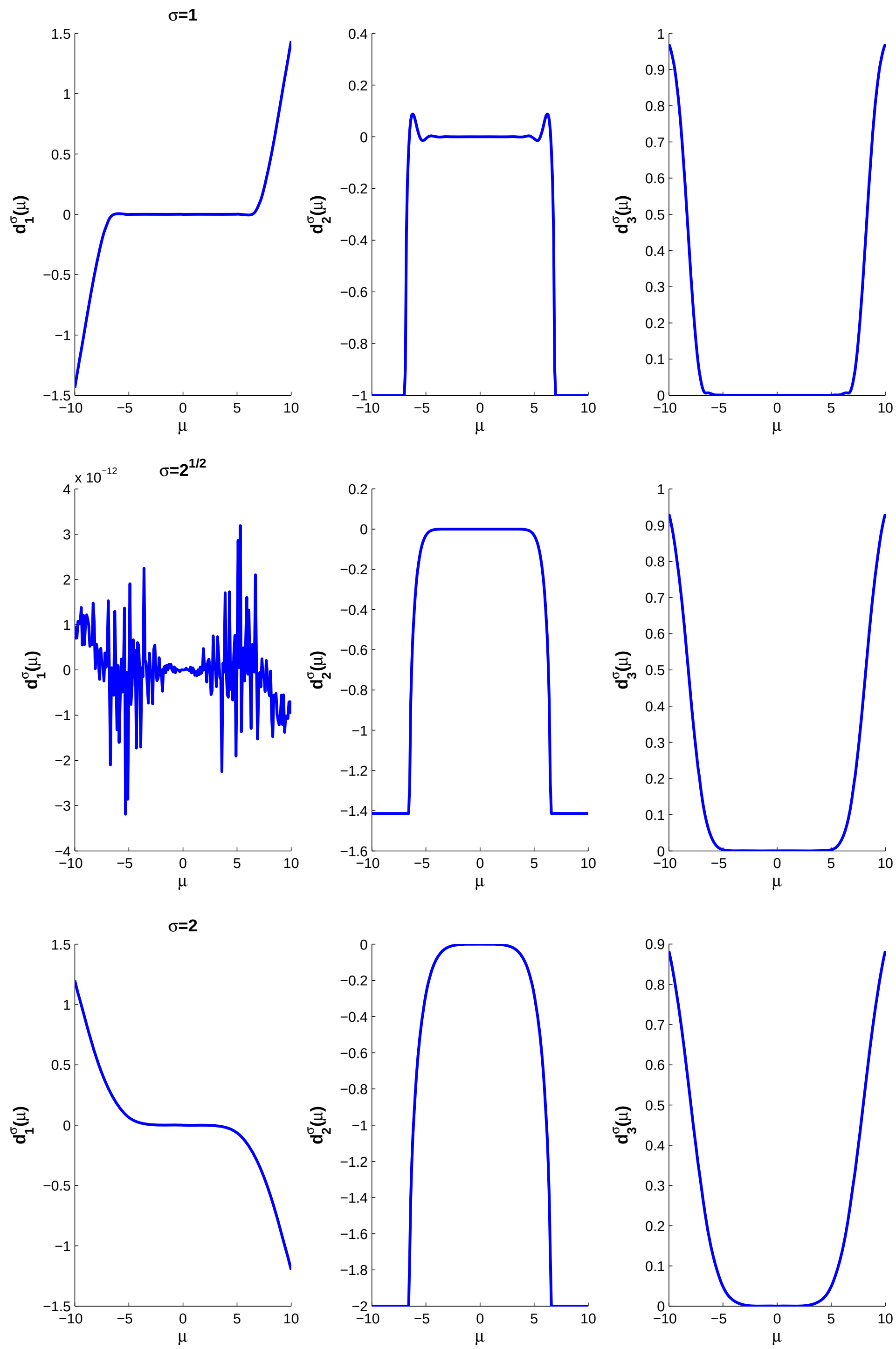

Figure 7.2: Error of the approximation: We approximate the distribution of a normal random variable $\xi \sim N(\mu, \sigma^2)$ by a linear combination of $n = 20$ basis functions $\{e_i, i = 0, \ldots, 19\}$. To compare the difference between the estimated distribution and the true one, we present the figures of that μ maps to $d_1^\sigma(\mu)$, μ maps to $d_2^\sigma(\mu)$, and μ maps to $d_3^\sigma(\mu)$ which are given by Equation (7.3.1) (7.3.2) and (7.3.3). We present the results for different σ, $\sigma = 1$(top row), $\sigma = 2^{1/2}$(middle row), $\sigma = 2$(bottom row).

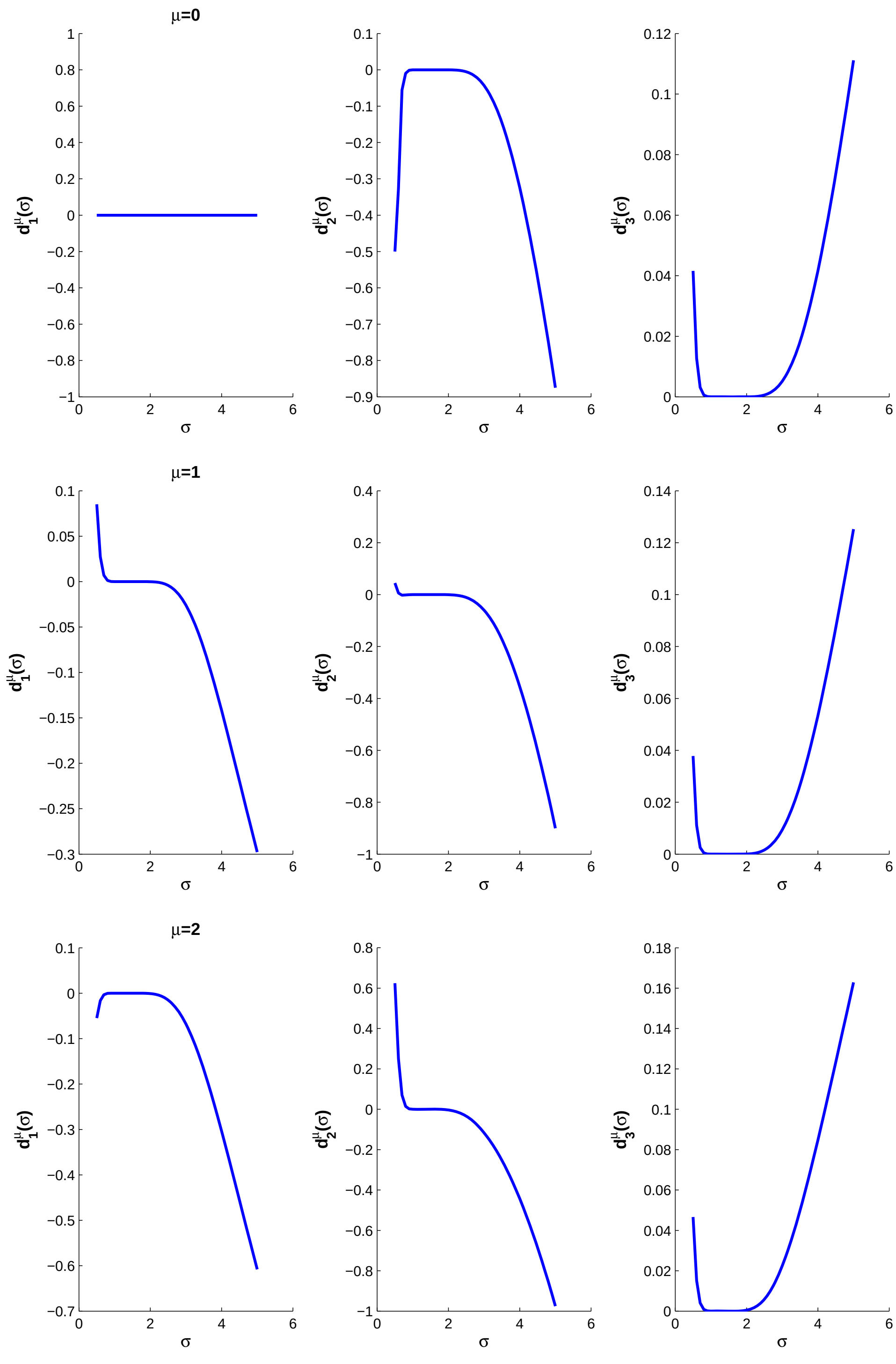

Figure 7.3: Error of the approximation: We approximate the distribution of a normal random variable $\xi \sim N(\mu, \sigma^2)$ by a linear combination of $n = 20$ basis functions $\{e_i, i = 0, \ldots, 19\}$. To compare the difference between the estimated distribution and the true one, we present the figures of that σ maps to $d_1^\mu(\sigma)$, σ maps to $d_2^\mu(\sigma)$, and σ maps to $d_3^\mu(\sigma)$ which are given by Equation (7.3.4) (7.3.5) and (7.3.6). We present the results for different μ, $\mu = 0$(top row), $\mu = 1$(middle row), $\mu = 2$(bottom row).

$[0,T]$ with time step $\Delta = t_k - t_{k-1} = \frac{T}{NM}$. We apply the iteration procedure at each subinterval $[t_{k-1}, t_k]$. And we apply the transition procedure to changing the center and scale of the basis functions adaptively at times $kM\Delta$, for $k = 1, \ldots, N$. The choice of the parameters n, N and M depends on the nonlinear filtering problems. In addition, we also change the center and scale of the basis at times of jumps if the the density varies significantly.

The adaptive basis functions at time t_k is

$$\bar{e}_i^{t_k}(x) := \frac{1}{\sqrt{\sigma_{t_k}^b}} e_i\left(\frac{x - \mu_{t_k}^b}{\sigma_{t_k}^b}\right), \quad x \in \mathbb{R}, \quad i = 0, 1, \ldots, n-1, \tag{7.3.7}$$

where $\mu_{t_k}^b \in \mathbb{R}$ define the center and $\sigma_{t_k}^b \in \mathbb{R}^+$ define the scale of the adaptive basis $\{\bar{e}_i^{t_k}\}$. They are chosen according to some algorithms introduced below.

Then, the approximation of the conditional density at discrete times $(t_k, k = 0, 1, \ldots, NM)$ is

$$p_{t_k}^{(n,\Delta)}(x) := \frac{q_{t_k}^{(n,\Delta)}(x)}{\int_{\mathbb{R}} q_{t_k}^{(n,\Delta)}(x)dx}, \quad x \in \mathbb{R}, \tag{7.3.8}$$

with $q_{t_k}^{(n,\Delta)}$ which is the approximated unnormalized conditional density defined as

$$q_{t_k}^{(n,\Delta)}(x) := \sum_{i=0}^{n-1} \psi_{k,i}^{(n,\Delta)}(t)\bar{e}_i^{t_k}(x) = \Big(\bar{e}_0^{t_k}(x), \bar{e}_1^{t_k}(x), \ldots, \bar{e}_{n-1}^{t_k}(x)\Big)\Upsilon_k^{(n,\Delta)}, \quad x \in \mathbb{R}.$$

Here, $\Upsilon_k^{(n,\Delta)} := \Big(\psi_{k,0}^{(n,\Delta)}, \psi_{k,1}^{(n,\Delta)}, \ldots, \psi_{k,n-1}^{(n,\Delta)}\Big)^\top$ is the vector of Fourier coefficients of $q_{t_k}^{(n,\Delta)}$ w.r.t. adaptive basis $\{\bar{e}_i^{t_k}\}$, obtained by algorithms, introduced below.

To sum up, in order to obtain the approximated conditional density at time t_k, defined by Equation (7.3.8), the key point it to get $\mu_{t_k}^b$, $\sigma_{t_k}^b$, which determine the adaptive basis $\{\bar{e}_i^{t_k}\}$ at time t_k, and the corresponding Fourier coefficients $\Upsilon_k^{(n,\Delta)}$. They are obtained by initialization step, iteration step and transition step introduced below. Notice, for normal Galerkin approximation, it is suffice to determine $\Upsilon_k^{(n,\Delta)}$, since the basis is independent of time.

Compare to Equation (6.1.3), (6.1.4), (6.1.5) and (6.1.6), which are the coefficients of the normal Galerkin approximation, the coefficients of the adaptive Galerkin approximation $A^{(n)} = (a_{ij})_{n\times n}$, $B^{(n)} = (b_{ij})_{n\times n}$, $C^{(n)} = (c_{ij})_{n\times n}$ and $G^{(n)} = (g_{ij})_{n\times n}$ w.r.t. basis $\{\bar{e}_i^{t_k}\}_{i=0,1,\ldots,n-1}$, are as following, for $i, j = 0, 1, \ldots, n-1$,

$$a_{ji} = (\bar{e}_i^{t_k}, \bar{e}_j^{t_k}), \quad b_{ji} = \langle \mathcal{A}^*\bar{e}_i^{t_k}, \bar{e}_j^{t_k}\rangle, \quad c_{ji} = (h\bar{e}_i^{t_k}, \bar{e}_j^{t_k}), \quad g_{ji} = \Big((\lambda - 1)\bar{e}_i^{t_k}, \bar{e}_j^{t_k}\Big). \tag{7.3.9}$$

Notice, in this case, the coefficients depend on t_k. For example, $A^{(n)} = A_{t_k}^{(n)}$ and we write by $A^{(n)}$ for short.

The following is the initialization step:

Algorithm 7.3 (Initialization). • *Compute the mean and the variance of the distribution of X_0 and take the mean and the square root of the variance as the center and scale of the adaptive basis functions at time $t_0 = 0$,*

$$\mu_{t_0}^b := \int_{\mathbb{R}} x p_0(x)dx, \quad \sigma_{t_0}^b := \Big\{\int_{\mathbb{R}} (x - \mu_{t_0}^b)^2 p_0(x)dx\Big\}^{\frac{1}{2}}.$$

Recall that p_0 *is the law of the state process* X *at time* 0. *Consequently, the adaptive basis functions* $\{\bar{e}_i^{t_0}\}$ *at time* t_0 *are, by Equation* (7.3.7),

$$\bar{e}_i^{t_0}(x) = \frac{1}{\sqrt{\sigma_{t_0}^b}} e_i(\frac{x-\mu_{t_0}^b}{\sigma_{t_0}^b}), \quad x \in \mathbb{R}, \quad i = 0, 1, \ldots, n-1.$$

- *Compute the coefficients of the Galerkin approximation* $A^{(n)}$, $B^{(n)}$, $C^{(n)}$ *and* $G^{(n)}$ *w.r.t. basis* $\{\bar{e}_i^{t_0}\}_{i=0,1,\ldots,n-1}$, *as following, applying Equation* (7.3.9),

$$a_{ji} = (\bar{e}_i^{t_0}, \bar{e}_j^{t_0}), \quad b_{ji} = \langle \mathcal{A}^* \bar{e}_i^{t_0}, \bar{e}_j^{t_0} \rangle, \quad c_{ji} = (h\bar{e}_i^{t_0}, \bar{e}_j^{t_0}), \quad g_{ji} = \Big((\lambda - 1)\bar{e}_i^{t_0}, \bar{e}_j^{t_0} \Big).$$

- *Applying Equation* (6.1.7), *obtain the starting point of* $\Upsilon^{(n,\Delta)}$, *for* $i = 0, 1, \ldots, n-1$,

$$\psi_{0,i}^{(n,\Delta)} = (p_0, \bar{e}_i^{t_0}).$$

Now, we present the iteration step. Assume, $\{Z_{t_k}\}$, $\{Y_{t_k}\}$, $k = 0, 1, \ldots, NM$, are the sampled trajectories of the observation processes Z and Y at discrete times. The aim of this step is to obtain the the numerical solution of Equation (6.1.9) in subinterval $[(k-1)\Delta, k\Delta]$ based on the splitting up approximation, see Algorithm 7.2. It can be replaced by other time-discretisation method. Moreover, at times of jumps, the figure of the density may vary significantly. This is our motivation to change the center and scale of the basis functions adaptively at times of jumps. That is, if the density vary significantly, then we move the center and scale of the adaptive basis, otherwise, they remain.

Algorithm 7.4 (Iteration $(k-1)\Delta$ to $k\Delta$)**.**

- *Keep the center and scale of the adaptive basis as before, that is*

$$\mu_{t_k}^b := \mu_{t_{k-1}}^b, \quad \sigma_{t_k}^b := \sigma_{t_{k-1}}^b.$$

Consequently, the adaptive basis $\{\bar{e}_i^{t_k}\}$ *at time* t_k *is obtained by Equation* (7.3.7). *Then, compute the corresponding Fourier coefficients* $\Upsilon_k^{(n,\Delta)}$ *as follows.*

- *The prediction step:*

$$\bar{\Upsilon}_k^{(n,\Delta)} = \exp\Big[(B^{(n)} - G^{(n)})\Delta\Big]\Upsilon_{k-1}^{(n,\Delta)}.$$

- *The correction step with* Z*:*

$$\tilde{\Upsilon}_k^{(n,\Delta)} = \exp\Big\{C^{(n)}(Z_{t_k} - Z_{t_{k-1}}) - \frac{1}{2}(C^{(n)})^2\Delta\Big\}\bar{\Upsilon}_k^{(n,\Delta)}. \tag{7.3.10}$$

- *The correction step with* Y*: notice, in this step, we may change the center and scale of the adaptive basis.*

  i Compute the predicted Fourier coefficients as follows

$$\tilde{\tilde{\Upsilon}}_k^{(n,\Delta)} := (I_n + G^{(n)})^{(Y_{t_k} - Y_{t_{k-1}})}\tilde{\Upsilon}_k^{(n,\Delta)}. \tag{7.3.11}$$

ii *Compute the mean and variance of the predicted approximated conditional density obtained by* $\tilde{\tilde{\Upsilon}}_k^{(n,\Delta)}$,

$$\hat{p}_{t_k}^{(n,\Delta)} := \frac{\hat{q}_{t_k}^{(n,\Delta)}}{\int_{\mathbb{R}} \hat{q}_{t_k}^{(n,\Delta)}(x)dx} \quad \text{with} \quad \hat{q}_{t_k}^{(n,\Delta)}(x) := \left(\bar{e}_0^{t_k}(x), \bar{e}_1^{t_k}(x), \ldots, \bar{e}_{n-1}^{t_k}(x)\right) \tilde{\tilde{\Upsilon}}_k^{(n,\Delta)}$$

and set the mean and the square root of the variance to be $\tilde{\mu}^b$ *and* $\tilde{\sigma}^b$,

$$\tilde{\mu}^b := \int_{\mathbb{R}} x\hat{p}_{t_k}^{(n,\Delta)}(x)dx, \quad \tilde{\sigma}^b := \left\{ \int_{\mathbb{R}} (x-\tilde{\mu}^b)^2 \hat{p}_{t_k}^{(n,\Delta)}(x)dx \right\}^{\frac{1}{2}}.$$

iii *If* $\tilde{\mu}^b$ *is close to* $\mu_{t_k}^b$ *and* $\tilde{\sigma}^b$ *is closed to* $\sigma_{t_k}^b$, *keep the current basis functions, and set the Fourier coefficients as the predicted Fourier coefficients obtained by Equation* (7.3.11). *That is*

$$\Upsilon_k^{(n,\Delta)} := \tilde{\tilde{\Upsilon}}_k^{(n,\Delta)}.$$

Otherwise, set $\Upsilon_k^{(n,\Delta)}$ *as the Fourier coefficient of the approximated unnormalized conditional density w.r.t. basis* $\left\{\frac{1}{\sqrt{\tilde{\sigma}^b}} e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\right\}$ *and move the center and scale of the basis correspondingly. To be precise:*

1) *Compute* q_{t_k-}, *the approximated unnormalized conditional density before jump, with Fourier coefficients* $\tilde{\Upsilon}_k^{(n,\Delta)}$, *which are obtained by Equation* (7.3.10), *w.r.t. basis* $\{\bar{e}_i^{t_k}\}$,

$$q_{t_k-}(x) := \left(\bar{e}_0^{t_k}(x), \bar{e}_1^{t_k}(x), \ldots, \bar{e}_{n-1}^{t_k}(x)\right) \tilde{\Upsilon}_k^{(n,\Delta)}, \quad x \in \mathbb{R}. \tag{7.3.12}$$

2) *Compute* q_{t_k}, *the approximated unnormalized conditional density after jump,*

$$q_{t_k}(x) := \lambda^{Y_{t_k} - Y_{t_{k-1}}} q_{t_k-}(x), \quad x \in \mathbb{R}. \tag{7.3.13}$$

3) *Compute Fourier coefficients of* q_{t_k} *w.r.t. basis* $\left\{\frac{1}{\sqrt{\tilde{\sigma}^b}} e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\right\}_{i=0,1,\ldots,n-1}$. *In other words, project* q_{t_k} *to subspace* $\operatorname{span}\{\frac{1}{\sqrt{\tilde{\sigma}^b}} e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b}), i = 0, 1, \ldots, n-1\}$. *And set the Fourier coefficients to be* $\Upsilon_k^{(n,\Delta)} = (\psi_{k,0}^{(n,\Delta)}, \psi_{k,1}^{(n,\Delta)}, \ldots, \psi_{k,n-1}^{(n,\Delta)})^\top$. *That is,*

$$\psi_{k,i}^{(n,\Delta)} := \left(q_{t_k}, \frac{1}{\sqrt{\tilde{\sigma}^b}} e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\right), \quad i = 0, 1, \ldots, n-1. \tag{7.3.14}$$

4) *Move the center and scale of the basis functions, that is,*

$$\begin{aligned} &\text{reset} \quad \mu_{t_k}^b \quad \text{as} \quad \tilde{\mu}^b, \\ &\text{reset} \quad \sigma_{t_k}^b \quad \text{as} \quad \tilde{\sigma}^b, \end{aligned}$$

and, consequently, the adaptive basis $\{\bar{e}_i^{t_k}\}$ *at time* t_k *is reset by Equation* (7.3.7). *Reset* $A^{(n)}, B^{(n)}$, $C^{(n)}$, $G^{(n)}$ *by the coefficients of the Galerkin approximation w.r.t. the current basis* $\{\bar{e}_i^{t_k}\}$, *by Equation* (7.3.9).

Remark 7.2. *By Step 1), 2), and 3), we obtain $\Upsilon_k^{(n,\Delta)}$, the Fourier coefficients after jumps defined by Equation* (7.3.14)*, from $\tilde{\Upsilon}_k^{(n,\Delta)}$, the Fourier coefficients before jumps which are defined by Equation* (7.3.10)*. Notice, they are Fourier coefficients w.r.t. different basis functions. Step 1), 2), and 3) are equivalent as follows, set $R := (r_{ij})_{n\times n}$, with*

$$r_{ji} := \Big(\lambda^{Y_{t_k}-Y_{t_{k-1}}}\frac{1}{\sqrt{\sigma_{t_k}^b}}e_i(\frac{x-\mu_{t_k}^b}{\sigma_{t_k}^b}),\ \frac{1}{\sqrt{\tilde{\sigma}^b}}e_j(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\Big), \quad i,j=0,1,\ldots,n-1.$$

Then we have

$$\Upsilon_k^{(n,\Delta)} = R\tilde{\Upsilon}_k^{(n,\Delta)}.$$

Proof. By Equation (7.3.14), (7.3.13) and (7.3.12), and the definition of R,

$$\begin{aligned}\psi_{k,i}^{(n,\Delta)} =&\Big(q_{t_k},\ \frac{1}{\sqrt{\tilde{\sigma}^b}}e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\Big)\\ =&\Big(\lambda^{Y_{t_k}-Y_{t_{k-1}}}q_{t_k-}(x),\ \frac{1}{\sqrt{\tilde{\sigma}^b}}e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\Big)\\ =&(r_{i0},r_{i1},\ldots,r_{i,n-1})\tilde{\Upsilon}_k^{(n,\Delta)}.\end{aligned}$$

□

As time passes by, the mean and variance of the approximated conditional density may vary significantly. Perhaps they are far away from the center and scale of the basis and this reduces the effective of the Galerkin approximation. This is the motivation to change the center and scale of the basis functions adaptively after some times. The following is the transition step. The aim of this step is to change the center and scale of the basis adaptively at times t_k, for $k = M, 2M, 3M, \ldots, NM$.

Algorithm 7.5 (Transition)**.**

- *With the current Fourier coefficients $\Upsilon_k^{(n,\Delta)}$, compute the mean and variance of current conditional density $p_{t_k}^{(n,\Delta)}$, which is obtained by Equation* (7.3.8)*, and set the mean and the square root of the variance to be $\tilde{\mu}^b$ and $\tilde{\sigma}^b$,*

$$\tilde{\mu}^b := \int_{\mathbb{R}} xp_{t_k}^{(n,\Delta)}(x)dx, \quad \tilde{\sigma}^b := \Big\{\int_{\mathbb{R}}(x-\tilde{\mu}^b)^2p_{t_k}^{(n,\Delta)}(x)dx\Big\}^{\frac{1}{2}}.$$

- *If they are close to $\mu_{t_k}^b$ and $\sigma_{t_k}^b$, keep the current basis functions, and keep the current Fourier coefficients.*

  Otherwise, reset $\Upsilon_k^{(n,\Delta)}$ as the Fourier coefficients of the unnormalized conditional density w.r.t. basis $\Big\{\frac{1}{\sqrt{\tilde{\sigma}^b}}e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\Big\}$ and move the center and scale of the basis correspondingly. To be precise:

  1) *Compute $q_{t_k}^{(n,\Delta)}$, the current unnormalized conditional density obtained by Fourier coefficients $\Upsilon_k^{(n,\Delta)}$ w.r.t. basis $\{\bar{e}_i^{t_k}\}$,*

$$q_{t_k}^{(n,\Delta)}(x) := \Big(\bar{e}_0^{t_k}(x),\bar{e}_1^{t_k}(x),\ldots,\bar{e}_{n-1}^{t_k}(x)\Big)\Upsilon_k^{(n,\Delta)}, \quad x\in\mathbb{R}. \tag{7.3.15}$$

2) *Compute Fourier coefficients of* $q_{t_k}^{(n,\Delta)}$ *w.r.t. basis* $\left\{\frac{1}{\sqrt{\tilde{\sigma}^b}} e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\right\}_{i=0,1,\dots,n-1}$. *In other words, project* $q_{t_k}^{(n,\Delta)}$ *to subspace* $\mathrm{span}\{\frac{1}{\sqrt{\tilde{\sigma}^b}} e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b}), i = 0, 1, \dots, n-1\}$. *And reset the corresponding Fourier coefficients to be* $\Upsilon_k^{(n,\Delta)} = (\psi_{k,0}^{(n,\Delta)}, \psi_{k,1}^{(n,\Delta)}, \dots, \psi_{k,n-1}^{(n,\Delta)})^\top$. *That is, for* $i = 0, 1, \dots, n-1$,

$$\textit{reset} \quad \psi_{k,i}^{(n,\Delta)} \textit{as} \quad \left(q_{t_k}^{(n,\Delta)}, \frac{1}{\sqrt{\tilde{\sigma}^b}} e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\right). \tag{7.3.16}$$

3) *Move the center and scale of the basis functions, that is*

$$\begin{aligned} &\textit{reset} \quad \mu_{t_k}^b \quad \textit{as} \quad \tilde{\mu}^b, \\ &\textit{reset} \quad \sigma_{t_k}^b \quad \textit{as} \quad \tilde{\sigma}^b, \end{aligned}$$

and consequently the adaptive basis $\{\bar{e}_i^{t_k}\}$ *at time* t_k *is reset by Equation* (7.3.7). *Reset* $A^{(n)}$, $B^{(n)}$, $C^{(n)}$, $G^{(n)}$ *by the coefficients of the Galerkin approximation w.r.t. the current basis* $\{\bar{e}_i^{t_k}\}$, *by Equation* (7.3.9).

Similar to Remark 7.2, we have

Remark 7.3. *Step 1) and 2) are equivalent as follows, set* $K := (k_{ij})_{n\times n}$, *with*

$$k_{ji} := \left(\frac{1}{\sqrt{\sigma_{t_k}^b}} e_i(\frac{x-\mu_{t_k}^b}{\sigma_{t_k}^b}), \frac{1}{\sqrt{\tilde{\sigma}^b}} e_j(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\right), \quad i, j = 0, 1, \dots, n-1.$$

Then

$$\textit{reset} \quad \Upsilon_k^{(n,\Delta)} \quad \textit{as} \quad K\Upsilon_k^{(n,\Delta)}.$$

Proof. By Equation (7.3.16) and (7.3.15), and the definition of K,

$$\left(q_{t_k}^{(n,\Delta)}, \frac{1}{\sqrt{\tilde{\sigma}^b}} e_i(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\right) = (k_{i0}, k_{i1}, \dots, k_{i,n-1})\Upsilon_k^{(n,\Delta)}.$$

□

Remark 7.4. *To conclude, in this section, we introduce the adaptive Galerkin approximation. Based on the normal Galerkin approximation for which the basis is independent of time, this method is obtained by adapted shift and sale of basis functions. The center and scale of basis are changed adaptively,*

- *at times* $kM\Delta$, *for* $k = 1, 2, \dots, N$, *if the mean and the square root of the variance for the approximated conditional density is far away from the center and scale of the adaptive basis.*
- *at times* $k\Delta$, *for* $k = 1, 2, \dots, NM$, *if observation* Y *jumps and the variation of the conditional density caused by the jump is significant.*

Therefore, compare to the normal Galerkin approximation, we should record not only the Fourier coefficients but also the center and scale of the adaptive basis functions.

The adaptive Galerkin approximation provides better results compared to the normal one. But, it is really time consuming since $A^{(n)}$, $B^{(n)}$, $C^{(n)}$ and $G^{(n)}$, which are the coefficients of the Galerkin approximation, and the mean and variance of conditional density should be recomputed many times. In next section, we will introduce the adaptive Galerkin approximation with Hermite polynomials for which the corresponding terms can be obtained explicitly. And consequently, the adaptive Galerkin approximation with this special basis is quite efficient.

7.3.3 The adaptive Galerkin approximation with Hermite polynomials

The objective of this section is to introduce, for some special cases, the adaptive Galerkin approximation with a basis of Hermite polynomials.

The basis computation step is the same as the one introduced in the previous section. Moreover, for reasons mentioned in Remark 7.1, using the basis functions obtained from Hermite polynomials, the conditional mean and conditional variance can be computed directly by the Fourier coefficients and the coefficients in Galerkin approximation can be computed explicitly for a large class of functions. Therefore, for the case of coefficients b, σ^2, h and λ are taking forms (7.2.21), we have the following efficient adaptive Galerkin approximation with Hermite polynomials for which the corresponding coefficients can be computed explicitly.

The general setting is the same as the setting in Section 7.3.2. Moreover, compare Section 7.3.2, in this section, we assume b, σ^2, h and $\lambda - 1$ are polynomials. Let non-negative integers b^n, α^n, h^n and λ^n be the corresponding degrees. Let b_κ, α_κ, h_κ and λ_κ be the corresponding coefficients at x^κ. Then we have the following representations

$$b(x) = \sum_{\kappa=0}^{b^n} b_\kappa x^\kappa, \;\; \sigma^2(x) = \sum_{\kappa=0}^{\alpha^n} \alpha_\kappa x^\kappa, \;\; h(x) = \sum_{\kappa=0}^{h^n} h_\kappa x^\kappa, \;\; \lambda(x) - 1 = \sum_{\kappa=0}^{\lambda^n} \lambda_\kappa x^\kappa, \quad x \in \mathbb{R}. \tag{7.3.17}$$

And we take $\{e_i\}$ as the basis of Hermite functions, defined by Equation (7.2.8).

Let $\mu^b_{t_k}$, $\sigma^b_{t_k}$ be the center and scale of the adaptive basis $\{\bar{e}^{t_k}_i\}$ we choose at time t_k. The computational advantages of this case are as follows: First, applying Equation (6.1.3), (6.1.4), (6.1.5) and (6.1.6), the coefficients of the Galerkin approximation $A^{(n)} = (a_{ij})_{n\times n}$, $B^{(n)} = (b_{ij})_{n\times n}$, $C^{(n)} = (c_{ij})_{n\times n}$ and $G^{(n)} = (g_{ij})_{n\times n}$ w.r.t. basis $\{\bar{e}^{t_k}_i\}_{i=0,1,\dots,n-1}$ defined by Equation

(7.3.7) can be computed explicitly as follows, for $i, j = 0, 1, \ldots, n-1$,

$$a_{ji} =(\bar{e}_i^{t_k}, \bar{e}_j^{t_k}) = \delta_{i,j}, \tag{7.3.18}$$

$$\begin{aligned} b_{ji} =&\langle \mathcal{A}^* \bar{e}_i^{t_k}, \bar{e}_j^{t_k} \rangle \\ =&\frac{\sqrt{i!}}{\sqrt{j!}} \Big\{ -\frac{1}{2\sigma_{t_k}^b} \sum_{m=i}^{b^n+j+1} \sum_{\kappa=0\vee(m-j-1)}^{m\wedge b^n} \sum_{r=\kappa}^{b^n} b_r C_r^\kappa (\mu_{t_k}^b)^{r-\kappa} (\sigma_{t_k}^b)^\kappa \vartheta_{m-\kappa}^{j+1} \iota_i^m \\ &+ \frac{j}{2\sigma_{t_k}^b} \sum_{m=i}^{b^n+j-1} \sum_{\kappa=0\vee(m-j+1)}^{m\wedge b^n} \sum_{r=\kappa}^{b^n} b_r C_r^\kappa (\mu_{t_k}^b)^{r-\kappa} (\sigma_{t_k}^b)^\kappa \vartheta_{m-\kappa}^{j-1} \iota_i^m \\ &+ \frac{1}{8(\sigma_{t_k}^b)^2} \sum_{m=i}^{\sigma^n+j+2} \sum_{\kappa=0\vee(m-j-2)}^{m\wedge \sigma^n} \sum_{r=\kappa}^{\sigma^n} \sigma_r C_r^\kappa (\mu_{t_k}^b)^{r-\kappa} (\sigma_{t_k}^b)^\kappa \vartheta_{m-\kappa}^{j+2} \iota_i^m \\ &- \frac{2j+1}{8(\sigma_{t_k}^b)^2} \sum_{m=i}^{\sigma^n+j} \sum_{\kappa=0\vee(m-j)}^{m\wedge \sigma^n} \sum_{r=\kappa}^{\sigma^n} \sigma_r C_r^\kappa (\mu_{t_k}^b)^{r-\kappa} (\sigma_{t_k}^b)^\kappa \vartheta_{m-\kappa}^{j} \iota_i^m \\ &+ \frac{j(j-1)}{8(\sigma_{t_k}^b)^2} \sum_{m=i}^{\sigma^n+j-2} \sum_{\kappa=0\vee(m-j+2)}^{m\wedge \sigma^n} \sum_{r=\kappa}^{\sigma^n} \sigma_r C_r^\kappa (\mu_{t_k}^b)^{r-\kappa} (\sigma_{t_k}^b)^\kappa \vartheta_{m-\kappa}^{j-2} \iota_i^m \Big\}, \end{aligned} \tag{7.3.19}$$

$$c_{ji} =(h\bar{e}_i^{t_k}, \bar{e}_j^{t_k}) = \frac{\sqrt{j!}}{\sqrt{i!}} \sum_{m=j}^{h^n+i} \sum_{\kappa=0\vee(m-i)}^{m\wedge h^n} \sum_{r=\kappa}^{h^n} h_r C_r^\kappa (\mu_{t_k}^b)^{r-\kappa} (\sigma_{t_k}^b)^\kappa \vartheta_{m-\kappa}^{i} \iota_j^m, \tag{7.3.20}$$

$$g_{ji} =\Big((\lambda-1)\bar{e}_i^{t_k}, \bar{e}_j^{t_k}\Big) = \frac{\sqrt{j!}}{\sqrt{i!}} \sum_{m=j}^{\lambda^n+i} \sum_{\kappa=0\vee(m-i)}^{m\wedge \lambda^n} \sum_{r=\kappa}^{\lambda^n} \lambda_r C_r^\kappa (\mu_{t_k}^b)^{r-\kappa} (\sigma_{t_k}^b)^\kappa \vartheta_{m-\kappa}^{i} \iota_j^m. \tag{7.3.21}$$

Second, the conditional mean and variance can be computed explicitly by the Fourier coefficients $\Upsilon_k^{(n,\Delta)}$ similarly as in Equation (7.2.13) and (7.2.14). To be precise, define $c_1 : \mathbb{R}^n \to \mathbb{R}$ and $c_2 : \mathbb{R}^n \to \mathbb{R}$ by

$$c_1(\Upsilon_k^{(n,\Delta)}) := \sqrt{2}\frac{(\mathbf{r}_1^{(n)})^\top \Upsilon_k^{(n,\Delta)}}{(\mathbf{r}_0^{(n)})^\top \Upsilon_k^{(n,\Delta)}}, \quad c_2(\Upsilon_k^{(n,\Delta)}) := 2\frac{(\mathbf{r}_2^{(n)})^\top \Upsilon_k^{(n,\Delta)}}{(\mathbf{r}_0^{(n)})^\top \Upsilon_k^{(n,\Delta)}}, \tag{7.3.22}$$

where $\mathbf{r}_0^{(n)}$, $\mathbf{r}_1^{(n)}$ and $\mathbf{r}_2^{(n)}$ are n-dimensional vectors whose components are constants defined by Equation (7.2.12), then we have

$$\pi_{t_k}(X) := \mathbb{E}[X_{t_k} | \mathcal{F}_{t_k}^{Z,Y}] \cong \sigma_{t_k}^b c_1(\Upsilon_k^{(n,\Delta)}) + \mu_{t_k}^b, \tag{7.3.23}$$

$$\begin{aligned} \mathbb{E}\Big[\big(X_{t_k} - \pi_{t_k}(X)\big)^2 \Big| \mathcal{F}_{t_k}^{Z,Y}\Big] \cong{}& (\sigma_{t_k}^b)^2 c_2(\Upsilon_k^{(n,\Delta)}) + 2\sigma_{t_k}^b \mu_{t_k}^b c_1(\Upsilon_k^{(n,\Delta)}) + (\mu_{t_k}^b)^2 \\ &- (\pi_{t_k}(X))^2. \end{aligned} \tag{7.3.24}$$

Finally, matrix R, defined in Remark 7.2, and matrix K, defined in Remark 7.3, can be computed explicitly. This will be shown in the following algorithms.

We now introduce the basis computation steps of the adaptive Galerkin approximation with a basis of Hermite polynomials, where the corresponding results are proven in Section 7.4.

Algorithm 7.6 (Initialization)**.** *The initialization step is the same as the one introduced in Section 7.3.2. Furthermore, in this case, the coefficients of the Galerkin approximation $B^{(n)}$, $C^{(n)}$ and $G^{(n)}$ can be computed explicitly by Equation* (7.3.18)*,* (7.3.19)*,* (7.3.20) *and* (7.3.21)*.*

Algorithm 7.7 ($(k-1)\Delta$ to $k\Delta$). • *Keep the center and scale of the adaptive basis as before, that is*

$$\mu^b_{t_k} := \mu^b_{t_{k-1}}, \quad \sigma^b_{t_k} := \sigma^b_{t_{k-1}}.$$

And consequently, the adaptive basis $\{\bar{e}^{t_k}_i\}$ *at time* t_k *is obtained by Equation* (7.3.7)*. Then, compute the corresponding Fourier coefficients* $\Upsilon^{(n,\Delta)}_k$ *as follows.*

- *The prediction step:*

$$\bar{\Upsilon}^{(n,\Delta)}_k = \exp\Big[(B^{(n)} - G^{(n)})\Delta\Big]\Upsilon^{(n,\Delta)}_{k-1}.$$

- *The correction step with* Z*:*

$$\tilde{\Upsilon}^{(n,\Delta)}_k = \exp\Big\{C^{(n)}(Z_{t_k} - Z_{t_{k-1}}) - \frac{1}{2}(C^{(n)})^2\Delta\Big\}\bar{\Upsilon}^{(n,\Delta)}_k.$$

- *The correction step with* Y*:*

  i Compute the predicted Fourier coefficients as follows

$$\tilde{\tilde{\Upsilon}}^{(n,\Delta)}_k := (I_n + G^{(n)})^{(Y_{t_k} - Y_{t_{k-1}})}\tilde{\Upsilon}^{(n,\Delta)}_k. \tag{7.3.25}$$

  ii Compute the mean and variance of the predicted conditional density and set the mean and the square root of the variance to be $\tilde{\mu}^b$ *and* $\tilde{\sigma}^b$*, which can be obtained explicitly, applying Equation* (7.3.23) *and* (7.3.24)*,*

$$\begin{aligned}
\tilde{\mu}^b &:= \sigma^b_{t_k} c_1(\tilde{\tilde{\Upsilon}}^{(n,\Delta)}_k) + \mu^b_{t_k}, \\
\tilde{\sigma}^b &:= \Big\{(\sigma^b_{t_k})^2 c_2(\tilde{\tilde{\Upsilon}}^{(n,\Delta)}_k) + 2\sigma^b_{t_k}\mu^b_{t_k} c_1(\tilde{\tilde{\Upsilon}}^{(n,\Delta)}_k) + (\mu^b_{t_k})^2 - (\tilde{\mu}^b)^2\Big\}^{\frac{1}{2}},
\end{aligned}$$

  where c_1 *and* c_2 *defined by Equation* (7.3.22)*.*

  iii If they are close to $\mu^b_{t_k}$ *and* $\sigma^b_{t_k}$*, keep the basis functions, and set the Fourier coefficients as the predicted Fourier coefficients,*

$$\Upsilon^{(n,\Delta)}_k := \tilde{\tilde{\Upsilon}}^{(n,\Delta)}_k. \tag{7.3.26}$$

  Else , set $R := (r_{ij})_{n\times n}$*, with, for* $i, j = 0, 1, \ldots, n-1$*,*

$$r_{ji} := \Big(\lambda^{Y_{t_k} - Y_{t_{k-1}}} \frac{1}{\sqrt{\sigma^b_{t_k}}} e_i(\frac{x - \mu^b_{t_k}}{\sigma^b_{t_k}}),\ \frac{1}{\sqrt{\tilde{\sigma}^b}} e_j(\frac{x - \tilde{\mu}^b}{\tilde{\sigma}^b})\Big).$$

  Notice that λ *is a polynomial defined by Equation* (7.3.17)*,* $\lambda^{Y_{t_k} - Y_{t_{k-1}}}$ *is a polynomial. Then there exist* $\bar{\lambda}^n \in \mathbb{N}$ *and* $\bar{\lambda}_0$*,* $\bar{\lambda}_1$*, …,* $\bar{\lambda}_{\bar{\lambda}^n}$*, such that*

$$\bar{\lambda} := \lambda^{Y_{t_k} - Y_{t_{k-1}}} = \sum_{\kappa=0}^{\bar{\lambda}^n} \bar{\lambda}_\kappa x^\kappa, \quad x \in \mathbb{R}.$$

Then for $i, j = 0, 1, \ldots, n-1$,

$$r_{ji} = \frac{1}{\sqrt{i!j!}} \frac{1}{\sqrt{\sigma_{t_k}^b \tilde{\sigma}^b}} e^{\gamma_c} \frac{1}{\gamma_a} \sum_{r=0}^{i} \sum_{\nu=0}^{j} \kappa_r^i \kappa_\nu^j r! \sum_{m=r}^{\bar{\lambda}^n+\nu} \sum_{\kappa=0\vee(m-\nu)}^{m\wedge\bar{\lambda}^n} \bar{\lambda}_\kappa \vartheta_{m-\kappa}^\nu, \tag{7.3.27}$$

where

$$\begin{aligned} \gamma_a :=& \frac{1}{\sqrt{2}} \sqrt{\frac{(\sigma_{t_k}^b)^2 + (\tilde{\sigma}^b)^2}{(\sigma_{t_k}^b)^2 (\tilde{\sigma}^b)^2}}, \\ \gamma_b :=& \frac{1}{\sqrt{2}\sqrt{\frac{(\sigma_{t_k}^b)^2+(\tilde{\sigma}^b)^2}{(\sigma_{t_k}^b)^2(\tilde{\sigma}^b)^2}}} \left(\frac{\mu_{t_k}^b}{(\sigma_{t_k}^b)^2} + \frac{\tilde{\mu}^b}{(\tilde{\sigma}^b)^2}\right), \\ \gamma_c :=& -\frac{1}{4}\left[\left(\frac{(\mu_{t_k}^b)^2}{(\sigma_{t_k}^b)^2} + \frac{(\tilde{\mu}^b)^2}{(\tilde{\sigma}^b)^2}\right) - 2\gamma_b^2\right], \end{aligned} \tag{7.3.28}$$

and for $i = 0, 1, \ldots, n-1$, $r = 0, 1, \ldots, i$,

$$\kappa_r^i := \sum_{m=i}^{m} \Big\{ \sum_{\kappa=m}^{i} \vartheta_\kappa^i C_\kappa^m \Big(\frac{\gamma_b - \gamma_a \mu_{t_k}^b}{\gamma_a \sigma_{t_k}^b}\Big)^{\kappa-m} \Big(\frac{1}{\gamma_a \sigma_{t_k}^b}\Big)^m \Big\} \iota_r^m.$$

And compute

$$\Upsilon_k^{(n,\Delta)} := R \tilde{\Upsilon}_k^{(n,\Delta)}.$$

Move the center and scale of the basis functions, that is

$$\begin{aligned} &\textit{reset} \quad \mu_{t_k}^b \quad \textit{as} \quad \tilde{\mu}^b, \\ &\textit{reset} \quad \sigma_{t_k}^b \quad \textit{as} \quad \tilde{\sigma}^b, \end{aligned}$$

and consequently the adaptive basis $\{\bar{e}_i^{t_k}\}$ at time t_k is reset by Equation (7.3.7) and reset $A^{(n)}, B^{(n)}$, $C^{(n)}$, $G^{(n)}$ by the coefficients of the Galerkin approximation w.r.t. the current basis $\{\bar{e}_i^{t_k}\}$, by Equation (7.3.18), (7.3.19), (7.3.20), and (7.3.21).

The following is the transition step:

Algorithm 7.8 (Transition).

- *Compute the mean and variance of current conditional density $p_{t_k}^{(n,\Delta)}$ and set the mean the the square root of the variance to be $\tilde{\mu}^b$ and $\tilde{\sigma}^b$, applying Equation (7.3.23) and (7.3.24),*

$$\begin{aligned} \tilde{\mu}^b &:= \sigma_{t_k}^b c_1(\Upsilon_k^{(n,\Delta)}) + \mu_{t_k}^b, \\ \tilde{\sigma}^b &:= \Big\{ (\sigma_{t_k}^b)^2 c_2(\Upsilon_k^{(n,\Delta)}) + 2\sigma_{t_k}^b \mu_{t_k}^b c_1(\Upsilon_k^{(n,\Delta)}) + (\mu_{t_k}^b)^2 - (\tilde{\mu}^b)^2 \Big\}^{\frac{1}{2}}, \end{aligned}$$

  where c_1, c_2 are defined by Equation (7.3.22).

- *If they are close to μ^b and σ^b, keep the basis functions, and keep the Fourier coefficients. Otherwise, change the center and scale of the basis functions as follows. Set $K := (k_{ij})_{n\times n}$,*

Table 7.1: Proofs and corresponding pages.

with, for $i, j = 0, 1, \dots, n-1$,

$$k_{ji} := \Big(\frac{1}{\sqrt{\sigma^b_{t_k}}} e_i\big(\frac{x-\mu^b_{t_k}}{\sigma^b_{t_k}}\big),\ \frac{1}{\sqrt{\tilde\sigma^b}} e_j\big(\frac{x-\tilde\mu^b}{\tilde\sigma^b}\big) \Big) \tag{7.3.29}$$

$$= \frac{1}{\sqrt{i!j!}} \frac{1}{\sqrt{\sigma^b_{t_k}\tilde\sigma^b}} e^{\gamma_c} \frac{1}{\gamma_a} \sum_{r=0}^{i\wedge j} \Big\{ \sum_{m=r}^{i} [\sum_{\kappa=m}^{i} \vartheta^i_\kappa C^m_\kappa (\frac{\gamma_b - \gamma_a \mu^b_{t_k}}{\gamma_a \sigma^b_{t_k}})^{\kappa-m} (\frac{1}{\gamma_a \sigma^b_{t_k}})^m] \iota^m_r \Big\}$$

$$\cdot \Big\{ \sum_{m=r}^{j} [\sum_{\kappa=m}^{j} \vartheta^j_\kappa C^m_\kappa (\frac{\gamma_b - \gamma_a \tilde\mu^b}{\gamma_a \tilde\sigma^b})^{\kappa-m} (\frac{1}{\gamma_a \tilde\sigma^b})^m] \iota^m_r \Big\} r!,$$

where γ_a, γ_b *and* γ_c *are defined by Equation* (7.3.28).

Then,

$$reset \quad \Upsilon^{(n,\Delta)}_k \quad as \quad K\Upsilon^{(n,\Delta)}_k.$$

Move the center and scale of the basis functions, that is

$$\begin{aligned} reset \quad \mu^b_{t_k} \quad as \quad \tilde\mu^b, \\ reset \quad \sigma^b_{t_k} \quad as \quad \tilde\sigma^b, \end{aligned}$$

and consequently the adaptive basis $\{\bar e^{t_k}_i\}$ *at time* t_k *is reset by Equation* (7.3.7). *Reset* $A^{(n)}$, $B^{(n)}$, $C^{(n)}$, $G^{(n)}$ *by the coefficients of the Galerkin approximation w.r.t. the current basis* $\{\bar e^{t_k}_i\}$, *by Equation* (7.3.18), (7.3.19), (7.3.20), *and* (7.3.21).

7.4 Proofs

In this section, we present the proofs related to Hermite polynomials, see Table 7.1.

For starters, we simplify the proof, define $\forall x \in \mathbb{R}$, $i = 0, 1, \dots$,

$$E_i(x) := \varphi(x) H_i(x), \quad \text{with} \quad \varphi(x) := [\phi(x)]^{\frac{1}{2}} = (2\pi)^{-\frac{1}{4}} e^{-\frac{x^2}{4}}. \tag{7.4.1}$$

It is easy to see that, by the definition of Hermite functions, see Equation (7.2.8),

$$e_i = \frac{1}{\sqrt{i!}} E_i.$$

Lemma 7.4. *We have the following recurrence relations of E_i, which is defined by Equation (7.4.1),*

$$xE_i(x) = E_{i+1}(x) + iE_{i-1}(x), \quad i \geq 1, \tag{7.4.2}$$

$$(E_i)' = -\frac{1}{2}E_{i+1} + \frac{i}{2}E_{i-1}, \quad i \geq 1, \tag{7.4.3}$$

$$(E_i)'' = \frac{1}{4}E_{i+2} - \frac{2i+1}{4}E_i + \frac{i(i-1)}{4}E_{i-2}, \quad i \geq 2. \tag{7.4.4}$$

Proof. By the recurrent relation of hermite polynomials, see Equation (7.2.6) and (7.2.7), we get

$$\begin{aligned}
xE_i(x) =& \varphi(x)xH_i(x) = \varphi[H_{i+1}(x) + iH_{i-1}(x)] = E_{i+1}(x) + iE_{i-1}(x),\\
(E_i)' =& (\varphi H_i)' = -\frac{x}{2}\varphi H_i + i\varphi H_{i-1} = -\frac{x}{2}E_i + iE_{i-1}\\
=& -\frac{1}{2}(E_{i+1} + iE_{i-1}) + iE_{i-1} = -\frac{1}{2}E_{i+1} + \frac{i}{2}E_{i-1},\\
(E_i)'' =& [(E_i)']' = (-\frac{1}{2}E_{i+1} + \frac{i}{2}E_{i-1})' = \frac{1}{4}E_{i+2} - \frac{2i+1}{4}E_i + \frac{i(i-1)}{4}E_{i-2}.
\end{aligned}$$

□

In the following proofs, we have written μ^b, σ^b, $\bar{e}_i$ instead of $\mu^b_{t_k}$, $\sigma^b_{t_k}$, $\bar{e}^{t_k}_i$. We have the following results which are useful for our proofs.

Lemma 7.5. *Let $f(x) = \sum_{k=0}^{a^n} a_k x^k$ be a polynomial, where a^n, a non-negative integer, is the degree of f and $a_k \in \mathbb{R}$ is the coefficient at f of x^k for $k = 0, 1, \ldots, a^n$. Then, for any $\mu, \sigma \in \mathbb{R}$ and $\sigma \neq 0$, we have $f(\mu + \sigma x)$ is a polynomial and $\forall x \in \mathbb{R}$,*

$$f(\mu + \sigma x) = \sum_{m=0}^{a^n} \tilde{a}_m x^m, \ \textit{with} \ \tilde{a}_m := \sum_{k=m}^{a^n} a_k C_k^m \mu^{k-m}\sigma^m. \tag{7.4.5}$$

Proof. By the definition of f,

$$\begin{aligned}
f(\mu + \sigma x) =& \sum_{k=0}^{a^n} a_k(\mu + \sigma x)^k\\
=& \sum_{k=0}^{a^n} a_k \sum_{m=0}^{k} C_k^m \mu^{k-m}\sigma^m x^m\\
=& \sum_{k=0}^{a^n}\sum_{m=0}^{k} a_k C_k^m \mu^{k-m}\sigma^m x^m\\
=& \sum_{m=0}^{a^n}\Big\{\sum_{k=m}^{a^n} a_k C_k^m \mu^{k-m}\sigma^m\Big\}x^m\\
=& \sum_{m=0}^{a^n} \tilde{a}_m x^m.
\end{aligned}$$

□

Lemma 7.6. *For any $\mu, \sigma \in \mathbb{R}$ and $\sigma \neq 0$, we have $H_i(\mu+\sigma x)$ is linear combination of Hermite polynomials as follows, $\forall x \in \mathbb{R}$,*

$$H_i(\mu+\sigma x) = \sum_{r=0}^{i} \Big\{ \sum_{m=r}^{i} \Big[\sum_{k=m}^{i} \vartheta_k^i C_k^m \mu^{k-m} \sigma^m \Big] \iota_r^m \Big\} H_r(x). \tag{7.4.6}$$

Proof. Apply Equation (7.2.2), (7.2.3) and Lemma 7.5,

$$\begin{aligned} H_i(\mu+\sigma x) &= \sum_{k=0}^{i} \vartheta_k^i (\mu+\sigma x)^k \\ &= \sum_{m=0}^{i} \Big[\sum_{k=m}^{i} \vartheta_k^i C_k^m \mu^{k-m} \sigma^m \Big] x^m \\ &= \sum_{m=0}^{i} \Big[\sum_{k=m}^{i} \vartheta_k^i C_k^m \mu^{k-m} \sigma^m \Big] \Big(\sum_{r=0}^{m} \iota_r^m H_r(x) \Big) \\ &= \sum_{m=0}^{i} \sum_{r=0}^{m} \Big[\sum_{k=m}^{i} \vartheta_k^i C_k^m \mu^{k-m} \sigma^m \Big] \iota_r^m H_r(x) \\ &= \sum_{r=0}^{i} \Big\{ \sum_{m=r}^{m} \Big[\sum_{k=m}^{i} \vartheta_k^i C_k^m \mu^{k-m} \sigma^m \Big] \iota_r^m \Big\} H_r(x). \end{aligned}$$

□

Lemma 7.7. *Let $f(x) = \sum_{k=0}^{a^n} a_k x^k$ be a polynomial, where a^n, a non-negative integer, is the degree of f and $a_k \in \mathbb{R}$ is the coefficient at f of x^k for $k = 0, 1, \dots, a^n$. Then we have[2], for $i, j = 0, 1, \dots,$*

$$(f, E_i E_j) = j! \sum_{m=j}^{a^n+i} \sum_{k=0 \vee (m-i)}^{m \wedge a^n} a_k \vartheta_{m-k}^i \iota_j^m. \tag{7.4.7}$$

Proof. By definition of E_i, see Equation (7.4.1),

$$(f, E_i E_j) = (f, \phi H_i H_j) = (f H_i, \phi H_j).$$

Notice, the product of two polynomials is a polynomial, then fH_i is a polynomial as follows, by polynomial multiplication, $\forall x \in \mathbb{R}$,

$$f(x)H_i(x) = \Big(\sum_{m=0}^{a^n} a_m x^m \Big) \Big(\sum_{k=0}^{i} \vartheta_k^i x^k \Big) = \sum_{m=0}^{a^n+i} \Big\{ \sum_{k=0 \vee (m-i)}^{m \wedge a^n} a_k \vartheta_{m-k}^i \Big\} x^m = \sum_{m=0}^{a^n+i} \tilde{a}_m x^m,$$

with coefficients, for $m = 0, 1, \dots, a^n + i$,

$$\tilde{a}_m := \sum_{k=0 \vee (m-i)}^{m \wedge a^n} a_k \vartheta_{m-k}^i.$$

[2]The summation $\sum_{i=m}^{n}$ is taken to be 0 if the lower bound of summation m greater than the upper bound of summation n.

Then, by Equation (7.2.3), $f(x)H_i(x)$ is a linear combination of Hermite polynomials

$$f(x)H_i(x) = \sum_{m=0}^{a^n+i} \tilde{a}_m x^m = \sum_{m=0}^{a^n+i} \tilde{a}_m \Big\{ \sum_{k=0}^{m} \iota_k^m H_k(x) \Big\}.$$

By the orthogonal property of Hermite polynomials

$$\begin{aligned}
(fH_i, \phi H_j) =& \Big(\sum_{m=0}^{a^n+i} \tilde{a}_m \sum_{k=0}^{m} \iota_k^m H_k, \phi H_j \Big) = \sum_{m=0}^{a^n+i} \tilde{a}_m \sum_{k=0}^{m} \iota_k^m \Big(H_k, \phi H_j \Big) \\
=& \sum_{m=0}^{a^n+i} \tilde{a}_m \sum_{k=0}^{m} \iota_k^m \delta_{k,j} j! = j! \sum_{m=j}^{a^n+i} \tilde{a}_m \iota_j^m.
\end{aligned}$$

Combining the obtained results, we obtain Equation (7.4.7). $\square$

Proof of Equation (7.2.13) *and* (7.2.14).

$$\pi_t(X) = \mathbb{E}[X_t | \mathcal{F}_t^{Z,Y}] = \frac{\int_{-\infty}^{\infty} x q_t(x) dx}{\int_{-\infty}^{\infty} q_t(x) dx}, \tag{7.4.8}$$

$$\mathbb{E}\Big[\big(X_t - \pi_t(X) \big)^2 | \mathcal{F}_t^{Z,Y} \Big] = \frac{\int_{-\infty}^{\infty} [x - \pi_t(X)]^2 q_t(x) dx}{\int_{-\infty}^{\infty} q_t(x) dx} = \frac{\int_{-\infty}^{\infty} x^2 q_t(x) dx}{\int_{-\infty}^{\infty} q_t(x) dx} - [\pi_t(X)]^2. \tag{7.4.9}$$

By the approximation (7.2.11),

$$\int_{-\infty}^{\infty} q_t(x) dx \cong \int_{-\infty}^{\infty} q_t^n(x) dx = \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} e_i(x) dx \tag{7.4.10}$$

and by the definition of e_i,

$$\begin{aligned}
\int_{-\infty}^{\infty} e_i(x) dx =& \frac{1}{\sqrt{i!}} \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{4}} e^{-\frac{x^2}{4}} H_i(x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{-\frac{1}{4}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} H_i(\sqrt{2}x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{-\frac{1}{4}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} x^k dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \sum_{j=0}^{k} \iota_j^k H_j(x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \sum_{j=0}^{k} \iota_j^k \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} H_j(x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \sum_{j=0}^{k} \iota_j^k \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} H_j(x) H_0(x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \sum_{j=0}^{k} \iota_j^k \delta_{0,j} \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^k.
\end{aligned}$$

Therefore,

$$\int_{-\infty}^{\infty} q_t(x)dx \cong \sqrt{2}(2\pi)^{\frac{1}{4}} \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \frac{1}{\sqrt{i!}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^k.$$

Likewise, we get

$$\begin{aligned}\int_{-\infty}^{\infty} xe_i(x)dx &= \frac{1}{\sqrt{i!}} 2(2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^{k+1},\\ \int_{-\infty}^{\infty} x^2 e_i(x)dx &= \frac{1}{\sqrt{i!}} 2^{\frac{3}{2}} (2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^{k+2},\end{aligned}$$

and

$$\begin{aligned}\int_{-\infty}^{\infty} xq_t(x)dx &\cong \int_{-\infty}^{\infty} xq_t^{(n)}(x)dx = 2(2\pi)^{\frac{1}{4}} \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \frac{1}{\sqrt{i!}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^{k+1},\\ \int_{-\infty}^{\infty} x^2 q_t(x)dx &\cong \int_{-\infty}^{\infty} x^2 q_t^{(n)}(x)dx = 2^{\frac{3}{2}} (2\pi)^{\frac{1}{4}} \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \frac{1}{\sqrt{i!}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^{k+2}.\end{aligned}$$

Combining the obtained results, we get Equation (7.2.13) and (7.2.14). □

Proof of Equation (7.3.23) *and* (7.3.24)*.* The desired result can be obtained similarly as the proof of Equation (7.2.13) and (7.2.14).

By definition of $\bar{e}_i$, Equation (7.3.7),

$$\int_{-\infty}^{\infty} \bar{e}_i(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\sigma^b}} e_i(\frac{x-\mu^b}{\sigma^b})dx = \sqrt{\sigma^b} \int_{-\infty}^{\infty} e_i(x)dx.$$

Similarly, we get

$$\begin{aligned}\int_{-\infty}^{\infty} x\bar{e}_i(x)dx &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{\sigma^b}} e_i(\frac{x-\mu^b}{\sigma^b})dx\\ &= \int_{-\infty}^{\infty} (\sigma^b x + \mu^b) \frac{1}{\sqrt{\sigma^b}} e_i(x)\sigma^b dx\\ &= \sqrt{\sigma^b} \Big\{ \sigma^b \int_{-\infty}^{\infty} xe_i(x)dx + \mu^b \int_{-\infty}^{\infty} e_i(x)dx \Big\},\end{aligned}$$

and

$$\begin{aligned}\int_{-\infty}^{\infty} x^2 \bar{e}_i(x)dx &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{\sigma^b}} e_i(\frac{x-\mu^b}{\sigma^b})dx\\ &= \int_{-\infty}^{\infty} (\sigma^b x + \mu^b)^2 \frac{1}{\sqrt{\sigma^b}} e_i(x)\sigma^b dx\\ &= \sqrt{\sigma^b} \Big\{ (\sigma^b)^2 \int_{-\infty}^{\infty} x^2 e_i(x)dx + 2\sigma^b \mu^b \int_{-\infty}^{\infty} xe_i(x)dx + (\mu^b)^2 \int_{-\infty}^{\infty} e_i(x)dx \Big\}.\end{aligned}$$

Then, applying Equation (7.4.8), (7.4.9), (7.4.10), and Equation (7.3.22) which are the definitions of c_1, c_2, we get

$$\begin{aligned}\pi_t(X) \approx& \frac{\int_{-\infty}^{\infty} x q_t^{(n)}(x)dx}{\int_{-\infty}^{\infty} q_t^{(n)}(x)dx}\\ =&\frac{\sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} x\bar{e}_i(x)dx}{\sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} \bar{e}_i(x)dx}\\ =&\frac{\sqrt{\sigma^b}\Big\{\sigma^b \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} x e_i(x)dx + \mu^b \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} e_i(x)dx\Big\}}{\sqrt{\sigma^b} \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} e_i(x)dx}\\ =&\sigma^b c_1 + \mu^b,\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}\Big[\big(X_t - \pi_t(X)\big)^2 | \mathcal{F}_t^{Z,Y}\Big]\\ \approx&\frac{\int_{-\infty}^{\infty} x^2 q_t^{(n)}(x)dx}{\int_{-\infty}^{\infty} q_t^{(n)}(x)dx} - (\tilde{\mu}^b)^2\\ =&\frac{\sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} x^2\bar{e}_i(x)dx}{\sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} \bar{e}_i(x)dx} - (\tilde{\mu}^b)^2\\ =&\frac{\sqrt{\sigma^b}\Big\{(\sigma^b)^2 \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} x^2 e_i(x)dx + 2\sigma^b\mu^b \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} x e_i(x)dx\Big\}}{\sqrt{\sigma^b} \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} e_i(x)dx}\\ &+ \frac{(\mu^b)^2 \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} e_i(x)dx}{\sqrt{\sigma^b} \sum_{i=0}^{n-1} \psi_i^{(n)}(t) \int_{-\infty}^{\infty} e_i(x)dx} - (\tilde{\mu}^b)^2\\ =&(\sigma^b)^2 c_2 + 2\sigma^b\mu^b c_1 + (\mu^b)^2 - (\tilde{\mu}^b)^2.\end{aligned}$$

The desired result is obtained. □

Proof of Corollary 7.3. Applying Theorem 6.11, the desired result is obtained if we can show that $\cup_n V_n$ is dense in V, where $V_n = \text{span}\{e_0, e_1, \ldots, e_{n-1}\}$ and $V = H^1(\mathbb{R}^d)$. One the one hand, let C_0^∞ be the set of infinitely differentiable functions with compact support. Then, by Bongioanni and Torrea (2006), Proposition 1, page 339, and Bongioanni and Torrea (2006), Theorem 4, page 348, we get $\forall u \in C_0^\infty$, there exist a sequence $u_n \in \cup_n V_n$, s.t.,

$$\|u - u_n\|_V \to 0, \quad \text{as} \quad n \to \infty.$$

Therefore, we obtain

$$\overline{\cup_n V_n} \supset C_0^\infty.$$

Moreover, by Frey (2008), Theorem 2.17, page 31, C_0^∞ is dense in V. Then,

$$\overline{\cup_n V_n} \supset \overline{C_0^\infty} \supset V.$$

On the other hand, by the properties of Hermite functions, we have

$$V_n \subset V, \quad \forall n.$$

Notice that V is complete, we get

$$\overline{\cup_n V_n} \subset V.$$

To sum up, we get

$$\overline{\cup_n V_n} = V.$$

□

Proof of Example 7.1. Equation (7.2.15) is obtained from the orthogonality of $\{e_i\}$. Now, it is time to compute $B^{(n)}$. By the definition of $B^{(n)}$, and Remark 6.1,

$$b_{ji} = \langle \mathcal{A}^* e_i, e_j \rangle = \langle \mathcal{A} e_j, e_i \rangle = (\mathcal{L} e_j, e_i) = \frac{1}{\sqrt{i!}\sqrt{j!}}(\mathcal{L} E_j, E_i).$$

Recall that $\mathcal{L}$ is a second order differential operator defined by Equation (2.2.4), then, we get,

$$\begin{aligned}
\mathcal{L} E_j =& bx(E_j)' + \frac{\sigma^2}{2}(E_j)'' \\
=& -\frac{b}{2}E_{j+2} - \frac{b}{2}E_j + \frac{bj(j-1)}{2}E_{j-2} + \frac{\sigma^2}{2}(\frac{1}{4}E_{j+2} - \frac{2j+1}{4}E_j + \frac{j(j-1)}{4}E_{j-2} \\
=& (-\frac{b}{2} + \frac{\sigma^2}{8})E_{j+2} - (\frac{b}{2} + \frac{\sigma^2(2j+1)}{8})E_j + (\frac{b}{2} + \frac{\sigma^2}{8})j(j-1)E_{j-2}.
\end{aligned}$$

Finally, we obtain Equation (7.2.16).

Similarly, Equation (7.2.17) and (7.2.18) are obtained from Equation (6.1.5) and (6.1.6) applying Equation (7.4.2) repeatedly. It remains to compute q_{0j}. By the definition

$$\begin{aligned}
q_{0j} = (q_0, e_j) = \int \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}} \frac{1}{\sqrt{j!}}(2\pi)^{-\frac{1}{4}} e^{-\frac{x^2}{4}} H_j(x) dx \\
= \frac{1}{\sqrt{j!}}(2\pi)^{-\frac{1}{4}} \frac{1}{\sqrt{2\pi}\sigma_0} \int e^{-\left[\frac{(x-\mu_0)^2}{2\sigma_0^2} + \frac{x^2}{4}\right]} H_j(x) dx.
\end{aligned}$$

Notice that

$$\frac{(x-\mu_0)^2}{2\sigma_0^2} + \frac{x^2}{4} = \frac{(x-c_2)^2}{2c_1^2} - c_3,$$

then, we get

$$\begin{aligned}
q_{0j} =& \frac{1}{\sqrt{j!}}(2\pi)^{-\frac{1}{4}} \frac{1}{\sqrt{2\pi}\sigma_0} \int e^{-\frac{(x-c_2)^2}{2c_1^2} + c_3} H_j(x) dx \\
=& \frac{1}{\sqrt{j!}}(2\pi)^{-\frac{1}{4}} \frac{1}{\sqrt{2\pi}\sigma_0} e^{c_3} \int e^{-\frac{(x-c_2)^2}{2c_1^2}} H_j(x) dx \\
=& (2\pi)^{-\frac{1}{4}} \frac{1}{\sigma_0} e^{c_3} c_1 \int \frac{1}{\sqrt{j!}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} H_j(c_1 x + c_2) dx.
\end{aligned}$$

Finally, the desired result is obtained by Lemma 7.6 and the orthonormal property of Hermite polynomials.

□

Proof of Example 7.2. Notice that in this case

$$\begin{aligned}\mathcal{L}E_j =& bx(E_j)' + \frac{\sigma^2}{2}x^2(E_j)'' \\ =& \frac{\sigma^2}{8}E_{j+4} + (-\frac{b}{2} + \frac{\sigma^2}{2})E_{j+2} - \Big[\frac{b}{2} + \frac{\sigma^2}{8}(2j^2 + 2j - 1)\Big]E_j \\ &+ \frac{b}{2}j(j-1)E_{j-2} + \frac{\sigma^2}{8}j(j-1)(j-2)(j-3)E_{j-4},\end{aligned}$$

and the desired result is obtained similarly as in Example 7.1. □

Proof of Example 7.3. Notice that in this case

$$\begin{aligned}\mathcal{L}E_j =& (\alpha - \beta x)(E_j)' + \frac{\sigma^2}{2}x(E_j)'' \\ =& \frac{\sigma^2}{8}E_{j+3} + \frac{\beta}{2}E_{j+2} - \Big[\frac{\alpha}{2} + \frac{\sigma^2}{8}(j-1)\Big]E_{j+1} + \frac{\beta}{2}E_j \\ &+ \Big[\frac{\alpha}{2}j - \frac{\sigma^2}{8}(j+2)j\Big]E_{j-1} - \frac{\beta}{2}j(j-1)E_{j-2} + \frac{\sigma^2}{8}j(j-1)(j-2)E_{j-3},\end{aligned}$$

and the desired result is obtained similarly as in Example 7.1. □

Proof of Equation (7.3.19). By the definition of b_{ji} and Remark 6.1,

$$b_{ji} = \langle \mathcal{A}^* \bar{e}_i, \bar{e}_j \rangle = \langle \mathcal{A}\bar{e}_j, \bar{e}_i \rangle = (\mathcal{L}\bar{e}_j, \bar{e}_i) = \frac{1}{\sqrt{i!}\sqrt{j!}}(\mathcal{L}\bar{E}_j, \bar{E}_i).$$

Notice that, by Equation (7.4.3) and (7.4.4), then,

$$\begin{aligned}\mathcal{L}\bar{E}_j =& b(x)(\bar{E}_j)' + \frac{[\sigma(x)]^2}{2}(\bar{E}_j)'' \\ =& b(x)\frac{1}{\sigma^b}\frac{1}{\sqrt{\sigma^b}}E_j'(\frac{x-\mu^b}{\sigma^b}) + \frac{[\sigma(x)]^2}{2}\frac{1}{(\sigma^b)^2}\frac{1}{\sqrt{\sigma^b}}E_j''(\frac{x-\mu^b}{\sigma^b}) \\ =& \frac{b(x)}{\sigma^b}\frac{1}{\sqrt{\sigma^b}}E_j'(\frac{x-\mu^b}{\sigma^b}) + \frac{[\sigma(x)]^2}{2(\sigma^b)^2}\frac{1}{\sqrt{\sigma^b}}E_j''(\frac{x-\mu^b}{\sigma^b}) \\ =& \frac{b(x)}{\sigma^b}\frac{1}{\sqrt{\sigma^b}}[-\frac{1}{2}E_{j+1}(\frac{x-\mu^b}{\sigma^b}) + \frac{j}{2}E_{j-1}(\frac{x-\mu^b}{\sigma^b})] \\ &+ \frac{[\sigma(x)]^2}{2(\sigma^b)^2}\frac{1}{\sqrt{\sigma^b}}[\frac{1}{4}E_{j+2}(\frac{x-\mu^b}{\sigma^b}) - \frac{2j+1}{4}E_j(\frac{x-\mu^b}{\sigma^b}) + \frac{j(j-1)}{4}E_{j-2}(\frac{x-\mu^b}{\sigma^b})] \\ =& \frac{b(x)}{\sigma^b}\Big[-\frac{1}{2}\bar{E}_{j+1} + \frac{j}{2}\bar{E}_{j-1}\Big] \\ &+ \frac{[\sigma(x)]^2}{2(\sigma^b)^2}\Big[\frac{1}{4}\bar{E}_{j+2} - \frac{2j+1}{4}\bar{E}_j + \frac{j(j-1)}{4}\bar{E}_{j-2}\Big].\end{aligned}$$

Then, we have

$$\begin{aligned}&(\mathcal{L}\bar{E}_j, \bar{E}_i) \\ =& \int \Big\{\frac{b(x)}{\sigma^b}[-\frac{1}{2}\bar{E}_{j+1} + \frac{j}{2}\bar{E}_{j-1}] + \frac{[\sigma(x)]^2}{2(\sigma^b)^2}[\frac{1}{4}\bar{E}_{j+2} - \frac{2j+1}{4}\bar{E}_j + \frac{j(j-1)}{4}\bar{E}_{j-2}]\Big\}\bar{E}_i dx \\ =& \int \Big\{\frac{b(\mu^b + \sigma^b x)}{\sigma^b}[-\frac{1}{2}E_{j+1} + \frac{j}{2}E_{j-1}] + \frac{[\sigma(\mu^b + \sigma^b x)]^2}{2(\sigma^b)^2}[\frac{1}{4}E_{j+2} - \frac{2j+1}{4}E_j + \frac{j(j-1)}{4}E_{j-2}]\Big\}E_i dx,\end{aligned}$$

where the last equality is obtained by changing the variable in the integration. Applying Equation (7.4.5), $b(\mu^b + \sigma^b x)$ and $[\sigma(\mu^b + \sigma^b x)]^2$ are polynomials as follows

$$\begin{aligned} b(\mu^b + \sigma^b x) &= \sum_{m=0}^{b^n} \Big[\sum_{k=m}^{b^n} b_k C_k^m (\mu^b)^{k-m} (\sigma^b)^m \Big] x^m := \tilde{b}(x), \\ \Big[\sigma(\mu^b + \sigma^b x)\Big]^2 &= \sum_{m=0}^{\sigma^n} \Big[\sum_{k=m}^{\sigma^n} \sigma_k C_k^m (\mu^b)^{k-m} (\sigma^b)^m \Big] x^m := \tilde{\sigma}(x). \end{aligned}$$

Then we have

$$\begin{aligned} (\mathcal{L}\bar{E}_j, \bar{E}_i) = &- \frac{1}{2\sigma^b}(\tilde{b}E_{j+1}, E_i) + \frac{j}{2\sigma^b}(\tilde{b}E_{j-1}, E_i) \\ &+ \frac{1}{8(\sigma^b)^2}(\tilde{\sigma}E_{j+2}, E_i) - \frac{2j+1}{8(\sigma^b)^2}(\tilde{\sigma}E_j, E_i) + \frac{j(j-1)}{8(\sigma^b)^2}(\tilde{\sigma}E_{j-2}, E_i). \end{aligned}$$

Applying Equation (7.4.7), we obtain the desired results. □

Proof of Equation (7.3.20) *and* (7.3.21). By the definition of c_{ji},

$$\begin{aligned} c_{ji} =& (h\bar{e}_i, \bar{e}_j) \\ =& \frac{1}{\sqrt{i!}} \frac{1}{\sqrt{j!}} \int \Big(\sum_{k=0}^{h^n} h_k x^k \Big) \frac{1}{\sqrt{2\pi}\sigma^b} e^{-\frac{(x-\mu^b)^2}{2(\sigma^b)^2}} H_i(\frac{x-\mu^b}{\sigma^b}) H_j(\frac{x-\mu^b}{\sigma^b}) dx \\ =& \frac{1}{\sqrt{i!}} \frac{1}{\sqrt{j!}} \int \Big[\sum_{k=0}^{h^n} h_k (\mu^b + \sigma^b x)^k \Big] \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} H_i(x) H_j(x) dx, \end{aligned}$$

where the last equality is obtained by changing the variable in the integration. Notice that $\sum_{k=0}^{h^n} h_k(\mu^b + \sigma^b x)^k$ is a polynomials and its standard form is as follows, by Equation (7.4.5),

$$\sum_{k=0}^{h^n} h_k \Big(\mu^b + \sigma^b x\Big)^k = \sum_{m=0}^{h^n} \Big[\sum_{k=m}^{h^n} h_k C_k^m (\mu^b)^{k-m} (\sigma^b)^m \Big] x^m := \tilde{h}(x).$$

Then by Equation (7.4.7),

$$\begin{aligned} c_{ji} =& \frac{1}{\sqrt{i!}} \frac{1}{\sqrt{j!}} (\tilde{h}, E_i E_j) \\ =& \frac{\sqrt{j!}}{\sqrt{i!}} \sum_{m=j}^{h^n+i} \sum_{k=0\vee(m-i)}^{m\wedge h^n} \sum_{r=k}^{h^n} h_r C_r^k (\mu^b)^{r-k} (\sigma^b)^k \vartheta_{m-k}^i \iota_j^m. \end{aligned}$$

Equation (7.3.21) can be obtained similarly. □

Proof of Equation (7.3.29). By definition of k_{ji}

$$\begin{aligned} k_{ji} =& \Big(\frac{1}{\sqrt{\sigma^b}} e_i(\frac{x-\mu^b}{\sigma^b}), \frac{1}{\sqrt{\tilde{\sigma}^b}} e_j(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b}) \Big) \\ =& \frac{1}{\sqrt{i!j!}} \int \frac{1}{\sqrt{2\pi}\sqrt{\sigma^b\tilde{\sigma}^b}} \exp\Big[-\frac{1}{4}(\frac{x-\mu^b}{\sigma^b})^2 - \frac{1}{4}(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})^2 \Big] H_i(\frac{x-\mu^b}{\sigma^b}) H_j(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b}) dx. \end{aligned}$$

For sake of simplicity, define

$$\begin{aligned}
\gamma_a :=& \frac{1}{\sqrt{2}}\sqrt{\frac{(\sigma^b)^2+(\tilde{\sigma}^b)^2}{(\sigma^b)^2(\tilde{\sigma}^b)^2}},\\
\gamma_b :=& \frac{1}{\sqrt{2}\sqrt{\frac{(\sigma^b)^2+(\tilde{\sigma}^b)^2}{(\sigma^b)^2(\tilde{\sigma}^b)^2}}}\Big(\frac{\mu^b}{(\sigma^b)^2}+\frac{\tilde{\mu}^b}{(\tilde{\sigma}^b)^2}\Big),\\
\gamma_c :=& -\frac{1}{4}\Big[(\frac{(\mu^b)^2}{(\sigma^b)^2}+\frac{(\tilde{\mu}^b)^2}{(\tilde{\sigma}^b)^2})-2\gamma_b^2\Big].
\end{aligned}$$

Then,

$$\begin{aligned}
&\exp\Big[-\frac{1}{4}(\frac{x-\mu^b}{\sigma^b})^2-\frac{1}{4}(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})^2\Big]\\
=&\exp\Big\{-\frac{1}{4}\Big[\frac{(\sigma^b)^2+(\tilde{\sigma}^b)^2}{(\sigma^b)^2(\tilde{\sigma}^b)^2}x^2-2(\frac{\mu^b}{(\sigma^b)^2}+\frac{\tilde{\mu}^b}{(\tilde{\sigma}^b)^2})x+(\frac{(\mu^b)^2}{(\sigma^b)^2}+\frac{(\tilde{\mu}^b)^2}{(\tilde{\sigma}^b)^2})\Big]\Big\}\\
=&e^{\gamma_c}e^{-\frac{1}{2}(\gamma_a x-\gamma_b)^2},
\end{aligned}$$

and

$$\begin{aligned}
k_{ji} =& \frac{1}{\sqrt{i!j!}}\frac{1}{\sqrt{\sigma^b\tilde{\sigma}^b}}e^{\gamma_c}\int\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\gamma_a x-\gamma_b)^2}H_i(\frac{x-\mu^b}{\sigma^b})H_j(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})dx\\
=& \frac{1}{\sqrt{i!j!}}\frac{1}{\sqrt{\sigma^b\tilde{\sigma}^b}}e^{\gamma_c}\frac{1}{\gamma_a}\int\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}H_i(\frac{\gamma_b-\gamma_a\mu^b}{\gamma_a\sigma^b}+\frac{1}{\gamma_a\sigma^b}x)H_j(\frac{\gamma_b-\gamma_a\tilde{\mu}^b}{\gamma_a\tilde{\sigma}^b}+\frac{1}{\gamma_a\tilde{\sigma}^b}x)dx,
\end{aligned}$$

where, by Equation (7.4.6),

$$\begin{aligned}
H_i(\frac{\gamma_b-\mu^b}{\sigma^b}+\frac{1}{\gamma_a\sigma^b}x) &= \sum_{r=0}^{i}\{\sum_{m=r}^{i}[\sum_{k=m}^{i}\vartheta_k^i C_k^m(\frac{\gamma_b-\gamma_a\mu^b}{\gamma_a\sigma^b})^{k-m}(\frac{1}{\gamma_a\sigma^b})^m]\iota_r^m\}H_r(x) = \sum_{r=0}^{i}\kappa_r^i H_r(x),\\
H_j(\frac{\gamma_b-\tilde{\mu}^b}{\tilde{\sigma}^b}+\frac{1}{\gamma_a\tilde{\sigma}^b}x) &= \sum_{r=0}^{j}\{\sum_{m=r}^{i}[\sum_{k=m}^{j}\vartheta_k^j C_k^m(\frac{\gamma_b-\gamma_a\tilde{\mu}^b}{\gamma_a\tilde{\sigma}^b})^{k-m}(\frac{1}{\gamma_a\tilde{\sigma}^b})^m]\iota_r^m\}H_r(x) = \sum_{r=0}^{j}\kappa_r^j H_r(x),
\end{aligned}$$

with, for $i=0,1,\ldots,n-1$, $r=0,1,\ldots,i$,

$$\kappa_r^i := \sum_{m=r}^{i}\Big[\sum_{k=m}^{i}\vartheta_k^i C_k^m(\frac{\gamma_b-\gamma_a\mu^b}{\gamma_a\sigma^b})^{k-m}(\frac{1}{\gamma_a\sigma^b})^m\Big]\iota_r^m.$$

By the orthogonality of the Hermite polynomials

$$\begin{aligned}
k_{ji} =& \frac{1}{\sqrt{i!j!}}\frac{1}{\sqrt{\sigma^b\tilde{\sigma}^b}}e^{\gamma_c}\frac{1}{\gamma_a}\sum_{r=0}^{i}\sum_{\nu=0}^{j}\kappa_r^i\kappa_\nu^j(E_r,E_\nu)\\
=& \frac{1}{\sqrt{i!j!}}\frac{1}{\sqrt{\sigma^b\tilde{\sigma}^b}}e^{\gamma_c}\frac{1}{\gamma_a}\sum_{r=0}^{i}\sum_{\nu=0}^{j}\kappa_r^i\kappa_\nu^j\delta_{r,\mu}r!\\
=& \frac{1}{\sqrt{i!j!}}\frac{1}{\sqrt{\sigma^b\tilde{\sigma}^b}}e^{\gamma_c}\frac{1}{\gamma_a}\sum_{r=0}^{\min(i,j)}\kappa_r^i\kappa_r^j r!.
\end{aligned}$$

Combining the obtained results, we obtain Equation (7.3.29). □

Proof of Equation (7.3.27). By definition of r_{ji},

$$\begin{aligned}
r_{ji} =& \big(\bar{\lambda} e_i \frac{1}{\sqrt{\sigma^b}}(\frac{x-\mu^b}{\sigma^b}), \frac{1}{\sqrt{\tilde{\sigma}^b}} e_j(\frac{x-\tilde{\mu}^b}{\tilde{\sigma}^b})\big) \\
=& \frac{1}{\sqrt{i!j!}} \frac{1}{\sqrt{\sigma^b \tilde{\sigma}^b}} e^{\gamma_c} \frac{1}{\gamma_a} \\
&\cdot \int \bar{\lambda}(\frac{\gamma_b}{\gamma_a} + \frac{x}{\gamma_a}) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} H_i(\frac{\gamma_b - \gamma_a \mu^b}{\gamma_a \sigma^b} + \frac{1}{\gamma_a \sigma^b} x) H_j(\frac{\gamma_b - \gamma_a \tilde{\mu}^b}{\gamma_a \tilde{\sigma}^b} + \frac{1}{\gamma_a \tilde{\sigma}^b} x) dx.
\end{aligned}$$

Notice that $\bar{\lambda}(\frac{\gamma_b}{\gamma_a} + \frac{x}{\gamma_a})$ is a polynomial and its standard form is as follows, by Equation (7.4.5),

$$\sum_{k=0}^{\bar{\lambda}^n} \bar{\lambda}_k (\frac{\gamma_b}{\gamma_a} + \frac{x}{\gamma_a})^k = \sum_{m=0}^{\bar{\lambda}^n} [\sum_{k=m}^{\bar{\lambda}^n} \bar{\lambda}_k C_k^m (\frac{\gamma_b}{\gamma_a})^{k-m} (\frac{1}{\gamma_a})^m] x^m := \tilde{\lambda}(x),$$

then applying Equation (7.4.7),

$$\begin{aligned}
r_{ji} =& \frac{1}{\sqrt{i!j!}} \frac{1}{\sqrt{\sigma^b \tilde{\sigma}^b}} e^{\gamma_c} \frac{1}{\gamma_a} \sum_{r=0}^{i} \sum_{\nu=0}^{j} \kappa_r^i \kappa_\nu^j (\tilde{\lambda} E_r, E_\nu) \\
=& \frac{1}{\sqrt{i!j!}} \frac{1}{\sqrt{\sigma^b \tilde{\sigma}^b}} e^{\gamma_c} \frac{1}{\gamma_a} \sum_{r=0}^{i} \sum_{\nu=0}^{j} \kappa_r^i \kappa_\nu^j r! \sum_{m=r}^{\bar{\lambda}^n + \nu} \sum_{k=0 \vee (m-\nu)}^{m \wedge \bar{\lambda}^n} \tilde{\lambda}_k \vartheta_{m-k}^{\nu}.
\end{aligned}$$

Combining the obtained result, we get Equation (7.3.27). □

7.5 Simulation studies

The Galerkin approximation of Model (2.2.9) is tested in simulation studies and compared with particle filters. In Section 7.5.1, we present the simulated results, by tables, from which we can see the results in average sense. In Section 7.5.2, we present the simulated results by figures from which we can see the results for some specific examples.

First, by simulation, we obtain sampled trajectory of X, Z and Y. Here X and Z are simulated by Euler-Maruyama method, see Algorithm 7.1, while Y is simulated according to Algorithm 9.14, McNeil, Frey, and Embrechts (2005), page 399. The detailed algorithm is as follows.

Algorithm 7.9. *The sampled trajectory of the state process X, the observation processes Z and Y of Model* (2.2.9) *at discrete equidistance times $\{t_k, k = 1, 2, \ldots, L\}$, with $\delta = t_k - t_{k-1}$, are as follows*

i Simulate the driven processes V and B.

- *Simulate ΔV_i, ΔB_i, $i = 1, 2, \ldots, L$ using the MATLAB package, where ΔV_i and ΔB_i are independent and identically $N(0, \delta)$-normally distributed random variables.*
- *Then the path of the process V and B can be realized by $V(0) = 0$, $V(t_k) = \sum_{i=1}^{k} \Delta V_i$ and $B(0) = 0$, $B(t_k) = \sum_{i=1}^{k} \Delta B_i$.*

ii Simulate a trajectory of the state process X.

- *Generate X_0 by the law of X_0 using the MATLAB package.*

- *Given X_0, solve the stochastic differential equation $X_t = X_0+\int_0^t b(X_s)ds+\int_0^t \sigma(X_s)dV_s$ using Euler-Maruyama method, Algorithm 7.1, or splitting up method, Algorithm 7.2, to obtain the values of the state process X at discrete times $\{X_{t_k}, k = 1, 2, \dots, L\}$ recursively,*

$$X_{t_{k+1}} = X_{t_k} + b(X_{t_k})\delta + \sigma(X_{t_k})\Delta V_k.$$

iii Simulate a trajectory of the observation process Z.

- $Z_0 = 0$,
- *Again using Euler-Maruyama method, Algorithm 7.1, to obtain the values of Z at discrete times $\{Z_{t_k}, k = 1, 2, \dots, L\}$,*

$$Z_{t_{k+1}} = Z_{t_k} + h(X_{t_k})\delta + \Delta B_k. \tag{7.5.1}$$

iv Simulate a trajectory of the observation process Y, which is equivalent to simulate the jumping times of Y, $\tau_1, \tau_2, \dots$,

- *Compute $\lambda(X_t)$ at discrete times by $\{X_{t_k}, k = 1, 2, \dots, L\}$ obtained before.*
- *Compute for $j \le k$, $\Gamma_j(t_k) = \delta \sum_{i=j}^{k-1} \lambda(X_{t_i})$.*
- *Set $\tau_0 = 0$.*
- *For i=1,2..., generate a unit exponential random variable E independent of X and set*

$$\tau_i = \tau_{i-1} + \Gamma_{i-1}^{-1}(E).$$

7.5.1 Tables

In this section, we show the simulation results by tables such that we can compare with different method in average sense. We obtained the conditional distribution of X with respect to past observation with different methods and compare their performance. The simulation procedure is as follows.

1) Obtain the artificially simulated data for X, Y and Z by Algorithm 7.9 at discrete times $\{t_k, k = 1, 2, \dots, L\}$.

2) Compute the filter estimates with different methods based on the algorithms introduced in Chapter 3 and Chapter 7.

3) Repeat procedure 1) and 2) m times. In the simulation, we take $m = 100$. As a measure to compare performance, we computed the length of time used in the computation and the root mean square error(rms error for short) which is defined by, over all simulated trajectories and over all times,

$$d := \left\{ \frac{1}{mL} \sum_{j=1}^{m} \sum_{i=1}^{L} \|X^j(t_i) - \hat{X}^j(t_i)\|^2 \right\}^{\frac{1}{2}},$$

where $X^j(t_i)$ is the j-th simulated trajectory of X at time t_i and $\hat{X}^j(t_i)$ is the filtering estimate at time t_i in the j-th simulation.

In Table 7.3 and Table 7.4, we present some numerical examples to illustrate the effectiveness of the proposed filter. It is seen that Galerkin method is much faster than the particle method while the rms errors of both method are almost the same. From this point of view, the Galerkin method performs better.

7.5.2 Figures

In this section, we present some figures to show the numerical results of the nonlinear problem. We show some numerical results for one-dimensional cases and multi-dimensional cases.

We have two observations, one is continuous, another is a Poisson process whose stochastic intensity is a given function of the signal X. The objective is to compute the conditional distribution of the state X given the past observation. We obtained the conditional distribution of X with respect to past observation with different methods and compared their performance. As a measure for the compared performance, we computed the conditional mean and conditional variance of X by the conditional distribution. We see from the results, the conditional expectation with respect to both observation processes does much better.

We show some numerical results for Example 7.1 from Figure 7.4 to Figure 7.14. Here we take $b = 0.5$, $\sigma = 2$, $\mu_0 = 5$, $\sigma_0 = 0.0001$. And by Algorithm 7.9, simulated trajectories of X, Y and Z are obtained at discrete equidistance times $\{t_k, k = 1, 2, \ldots, L\}$, where $t_0 = 0$, $t_L = T = 1$, $t_k - t_{k-1} = 10^{-6}$. We compute the filter estimates with different methods based on algorithms introduced in Chapter 3 and Chapter 7. As a measure to compare performance, we compute conditional mean and conditional root mean square deviation(RMSD) which is the square root of the conditional variance. For the Galerkin filter, they are obtained by Equation (7.2.13), (7.2.14) and Equation (7.3.23), (7.3.24). For the particle filter, they are obtained by Equation (3.3.1).

We obtain an artificial trajectory of X at discrete times. Notice, in order to compare the performance efficiently, we keep it as the trajectory of X from Figure 7.4 to Figure 7.13. These figures are obtained as follows:

Figure 7.4: In this figure , we compare the performance of Galerkin filter and particle filter. The procedure is as follows:

i Set $h = 0.5$, $\lambda = 0.05$, obtain the simulated trajectories for Y and Z by Algorithm 7.9.

ii We compute the filter estimates with different methods. For Galerkin filter, we apply Algorithm 7.2 with $n = 20$ basis functions to approximate the conditional density. For this case, h and λ are small. That means the observation noise weight heavily. We need not change the center and scale of the basis functions adaptively.

For particle filter, we apply Algorithm 3.1 with 10^3 particles.

Figure 7.5 and Figure 7.6: Here, we show the result obtained by the Galerkin filter with more information compared to Figure 7.4. In other words, we take larger h and λ in this case. The procedure is as follows:

i Set $h = 5.5$, $\lambda = 10$, obtain the trajectories for Y and Z by Algorithm 7.9.

ii Compute the estimates by Galerkin filter. Here we apply Algorithm 7.6, 7.7 and 7.8 with $n = 20$ Hermite basis functions to approximate the conditional density. For this case, h and λ are larger. That means the observation noise is small. We change the center and scale of the basis functions adaptively by algorithms introduced in Section 7.3.3.

For particle filter, we apply Algorithm 3.1 with 10^3 particles.

It is seen, in figures 7.4, 7.5, and 7.6, more information improves the estimation procedure. It is clear that the results obtained from observations Y an Z are much better than the results obtained from only one observation Z. And the approximation obtained by the Galerkin filter is very close to that obtained by particle filter.

Figure 7.7, 7.8, and 7.9: Here, we compute conditional density with only continuous observation Z. In this case, the filter problem is the so called Kalman-Bucy filter and has an explicit solution. Then, we compare the results obtained by Galerkin filter to the exact solution. The procedure is as follows:

i Set $h = 5.5$, obtain an artificially trajectory for Z.

ii Compute the estimates with different methods. For Galerkin filter, we use $n = 20$ Hermite basis functions to approximate the conditional density and change the center and scale of the basis functions adaptively according to Algorithm 7.6, 7.7 and 7.8. For the Kalman-bucy filter, it can be solved explicitly by Equation (2.1.5).

In Figure 7.7, we compare the conditional mean and conditional root mean square deviation. In Figure 7.8 and 7.9, we compare the conditional density and show the difference of the conditional density obtained by Galerkin filter and Kalman-Bucy filter. It is seen that approximations are very close to the explicit solutions. Those filters with more basis functions improve the estimation procedure.

Figure 7.10: Here, we study the convergence with respect to h. Fix λ, and increase h, then we obtain different trajectories of Z with corresponding information. The larger h is, the more information Z has. Finally, we compare the results obtained with different information. The procedure is as follows:

i Set $\lambda = 0.05$, obtain a simulated trajectory for Y by Algorithm 7.9.

ii For $h = 0.5$, $h = 5.5$, $h = 15.5$, obtain 3 artificially trajectories $Z^{(1)}$, $Z^{(2)}$ and $Z^{(3)}$ of Z with corresponding h by Algorithm 7.9.

iii Compute the filter estimates with $(Y, Z^{(1)})$, $(Y, Z^{(2)})$ and $(Y, Z^{(3)})$ using Galerkin filter.

For Galerkin filter, we use $n = 20$ Hermite basis functions to approximate the conditional density and change the center and scale of the basis functions adaptively according Algorithm 7.6, 7.7 and 7.8.

Again, more information improves the estimation procedure. The estimator converges to the trajectory of X as h increases.

Figure 7.11: Here, we study the convergence with respect to λ. Fix h, and increase λ, then we obtain different trajectories of Y with corresponding information. The larger λ is, the more information Y has. Finally, we compare the results obtained with different information. The procedure is as follows:

i Set $h = 0.5$, obtain the artificially trajectory for Z by Algorithm 7.9.

ii For $\lambda = 0.05$, $\lambda = 5.05$, $\lambda = 15.05$, obtain 3 artificially trajectories $Y^{(1)}$, $Y^{(2)}$ and $Y^{(3)}$ of Y with corresponding λ, by Algorithm 7.9.

iii Compute the filter estimates with $(Y^{(1)}, Z)$, $(Y^{(2)}, Z)$ and $(Y^{(3)}, Z)$ using Galerkin filter.

For Galerkin filter, we use $n = 20$ Hermite basis functions to approximate the conditional density and change the center and scale of the basis functions adaptively according to Algorithm 7.6, 7.7 and 7.8.

Once again, we see more information improves the estimation procedure. The estimator converges to the trajectory of X as λ increases.

Figure 7.12: Here, we study the convergence with respect to n, where n is the number of the basis functions in the Galerkin filter. The procedure is as follows:

i Set $h = 5.5$, $\lambda = 10$, obtain the artificially trojectories for Z and Y by Algorithm 7.9.

ii Compute the filter estimates with Y and Z using Galerkin filter with $n = 4, 8, 16$ basis functions. In the computation, we change the center and scale of the basis functions adaptively according to Algorithm 7.6, 7.7 and 7.8.

It is seen that the filter with 8 and 16 basis functions improves the estimation procedure. Although the filter with 8 basis functions is no more precise than the filter with 16, it still provides a good estimate. Moreover, it is faster than the filter with 16 basis functions. Realistically, we can choose the number of basis functions according to the weight of time and precision.

Figure 7.13: Here, we show the efficiency of Galerkin filter with changing the center and scale of the basis functions adaptively. The procedure is as follows:

i Set $h = 20$, $\lambda = 10$, obtain the simulated trajectories of Z and Y by Algorithm 7.9.

ii Compute the filter estimates with Y and Z using Galerkin filter with $n = 20$ Hermite basis functions. One result is obtained by change the center and scale of the basis functions adaptively according Algorithm 7.6, 7.7 and 7.8. Another result is obtained without change by Algorithm 7.2.

It is seen that adaptive Galerkin filter improves the estimation procedure. The simple Galerkin filter without change does not do much better for this case. This poor performance can be attributed to the small observation noise. Since we have more information, the conditional density is well-localized in some small region of the state space and generally could not be predicted in advance. Adaptive Galerkin filter is seen to drastically improved the performance.

Figure 7.14: Here, we compare the performance of Galerkin filter with a basis of Hermite polynomials and particle filter for multi-dimensional case.

There are some methods to construct a multi-dimensional basis by Hermite polynomials. For the introduction of multi-dimensional Hermite polynomials, see for example Berkowitz and Garner (1970). Now, we construct the basis of $L^2(\mathbb{R}^d)$ by another method. Let $\{e_i\}$ be the basis of one-dimensional Hermite polynomials defined by Equation (7.2.8) which is a basis of $L^2(\mathbb{R})$, then

$$\left\{e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_d} : \quad i_1, i_2, \ldots i_d = 0, 1, 2, \ldots\right\}$$

is a basis of $L^2(\mathbb{R}^d)$.

Here, we take $d = 5$, then $X = \{X_t\}_{t\in[0,T]} = \left\{\left(X_t^1, X_t^2, X_t^3, X_t^4, X_t^5\right)\right\}_{t\in[0,T]}$ is a 5-dimensional stochastic process driven by 3-dimensional Brownian motion. X_0^1, X_0^2, X_0^3, X_0^4 and X_0^5 are independent, normally distributed with mean 0 and variance 0.01^2. Z is a 3-dimensional noisy nonlinear observations of the state process X. Y is a one-dimensional stochastic Poisson process with intensity $0.1(X_t^1)^2 + 0.2(X_t^2)^2 + 0.3(X_t^3)^2 + 0.1(X_t^4)^2 + 0.1(X_t^5)^2$. The corresponding coefficients are as follows:

$$b = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 & 1 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & -1 & -1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 1 \end{pmatrix}, \ \sigma = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \ h = \begin{pmatrix} 0.2 & 0.3 & 0.2 & 0.3 & 0.4 \\ 0.2 & 0.1 & 0.2 & 0.1 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 & 0.2 \end{pmatrix}.$$

In the Galerkin approximation of this case, we take a linear combination of

$$\left\{e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_5} : \quad i_1, i_2, \ldots i_5 = 0, 1, 2, 3\right\} \tag{7.5.2}$$

to approximate the unnormalized conditional density.

The procedure is as follows:

i Obtain the simulated trajectories for X, Y and Z according to Algorithm 7.9.

ii Compute the filter estimates with different methods. For Galerkin filter, we apply Algorithm 7.2 with basis functions defined by Equation (7.5.2) to approximate the conditional density. Then, the number of basis functions is $n = 4^5 = 1024$. For particle filter, we apply Algorithm 3.1 use 10^3 particles.

In the simulation study, it takes 12 seconds for particle filter and 9 seconds for the Galerkin filter. It is seen that the approximation obtained by the Galerkin filter is very close to that obtained by particle filter.

7.5.3 Summary

In Chapter 3 and Chapter 7 we surveyed numerical approximations for the nonlinear filtering problem w.r.t Model (2.2.9).

The finite dimensional filter, introduced in Section 3.1, is easy to implement. The conditional mean and variance can be computed explicitly, since the approximated conditional density is a linear combination of Gaussian functions. However, this method performs poorly if the nonlinearities are strong. Furthermore, the approximation conditional density at time t, for $\tau_n \le t < \tau_{n+1}$ is a linear combination of 2^n Gaussian functions, recalling that τ_n is the n-th jumping time of the jump observation. Therefore, the computation becomes expensive for higher jump intensity.

X	the trajectory of X
PF	by particle method with observations Z and Y
GF	by Galerkin method with observations Z and Y
PF^C	by particle method with only continuous observation Z
GF^C	by Galerkin method with only continuous observation Z
KBF	the result obtained by Kalman-Bucy filter
AGF	the adaptive Galerkin filter
h	the coefficient w.r.t. the continuous observation Z of Model 7.1
λ	the coefficient w.r.t. the jump observation Y of Model 7.1
n	the number of basis functions used in the Galerkin filter

Table 7.2: Notations interpretations: From Figure 7.4 to Figure 7.14.

The finite-state Markov chain approximation, introduced in Section 3.2, is fast because, the approximated conditional probabilities are finite dimensional. The solution of a finite-dimensional SDE and the coefficient matrices in the SDE are diagonal. Unfortunately, this method is not flexible. In practice, one uses the information from the filtering results to dynamically move the grid in a suitable manner.

The particle methods introduced in Section 3.3 are very flexible and easy to implement. The basic idea is to approximate the expectation by Monto Carlo methods.

The Galerkin approximation introduced in Chapter 7 is easy to implement. In this case, the conditional density is approximation by a linear combination of finite number of basis functions, the corresponding Fourier coefficients are the solutions of an ordinary SDE, and the coefficients matrices of the SDE can be computed before hand. In addition, the Galerkin approximation is parsimonious, since, see for example Figure 7.12, it provides good results with only 8 parameters. The key point of the Galerkin approximation is how to choose the basis functions. It will provide perfect results with suitable basis functions, see for example Figure 7.7, 7.8, and 7.9. As with the finite-state Markov chain approximation, the normal Galerkin approximation is not flexible. For this, we design an adaptive Galerkin approximation. Moreover, in Section 7.3.3, we derive explicit coefficient in the adaptive Galerkin approximations for a large class of coefficients functions of Model (2.2.9).

Rms error

N_G/N_P	5/20	10/50	15/100	20/1000
GF(EM)	0.6326	0.4245	0.4234	0.4232
GF(SU)	0.6541	0.4281	0.4259	0.4259
PF	0.4580	0.4663	0.4138	0.4277

Time(second)

N_G/N_P	5/20	10/50	15/100	20/1000
GF(EM)	$0.10s$	$0.11s$	$0.12s$	$0.13s$
GF(SU)	$2.4s$	$3.08s$	$3.94s$	$4.29s$
PF	$9s$	$22s$	$46s$	$472s$

Table 7.3: Performance comparison: This table Shows the performance comparison for Example 7.1. Here $b = 1$, $\sigma = 1$, $h = 0.1$, $\lambda = 0.1$, $X_0 \sim N(2, 1)$, $T = 0.5$. For details, see Section 7.5.1. See Table 7.5 for notation interpretations.

Rms error

N_G/N_P	5/20	10/50	15/100	20/1000
GF(EM)	0.4036	0.3996	0.3996	0.3996
GF(SU)	0.6541	0.3993	0.3993	0.3993
PF	0.4663	0.4148	0.4052	0.3928

Time(second)

N_G/N_P	5/20	10/50	15/100	20/1000
GF(EM)	$0.10s$	$0.11s$	$0.12s$	$0.13s$
GF(SU)	$2.4s$	$3.08s$	$3.94s$	$4.29s$
PF	$9s$	$22s$	$46s$	$472s$

Table 7.4: Performance comparison: This table shows the performance comparison for Example 7.2. Here $b = 1$, $\sigma = 0.4$, $h = 0.2$, $\lambda = 0.2$, $p_0 \sim \ln N(2, 1)$, $T = 0.5$. For details, see Section 7.5.1. See Table 7.5 for notation interpretations.

N_G	the number of basis functions in the Galerkin filter
N_P	the number of particle in the particle filter
GF(EM)	Galerkin filter with Euler-Maruyama approximation for Equation (6.1.9)
GF(SU)	Galerkin filter with splitting-up approximation for Equation (6.1.9)
PF	particle filter

Table 7.5: Notation interpretations: For Table 7.3 and Tabel 7.4.

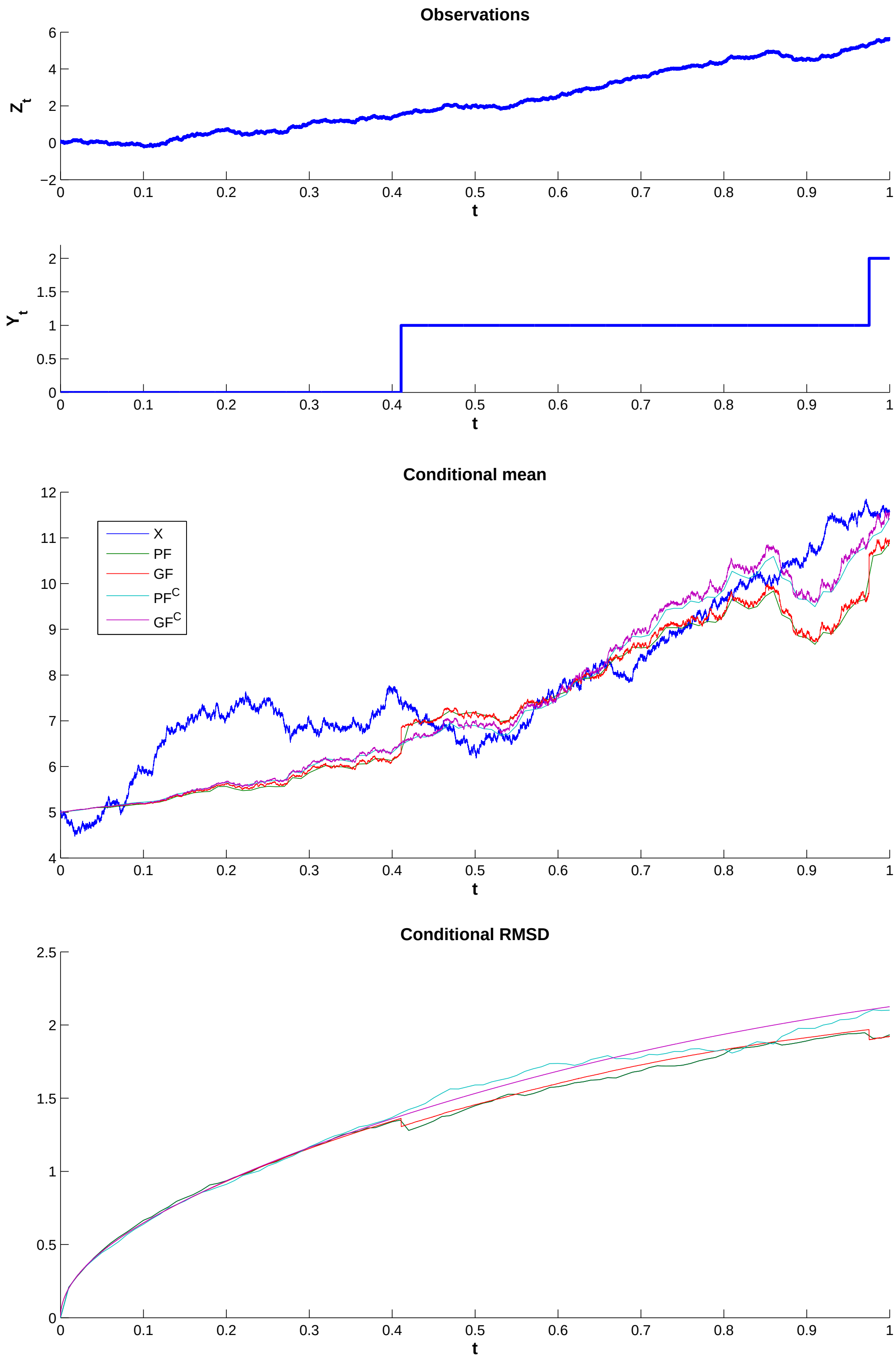

Figure 7.4: Comparison of Galerkin filter with particle filter: A trajectory of Z (the first row), A trajectory of Y (the second row), Conditional mean (the third row), Conditional root mean square deviation(bottom row). See Table 7.2 for notations interpretations.

It is seen that the approximation obtained by Galerkin filter is very close to that obtained by particle filter.

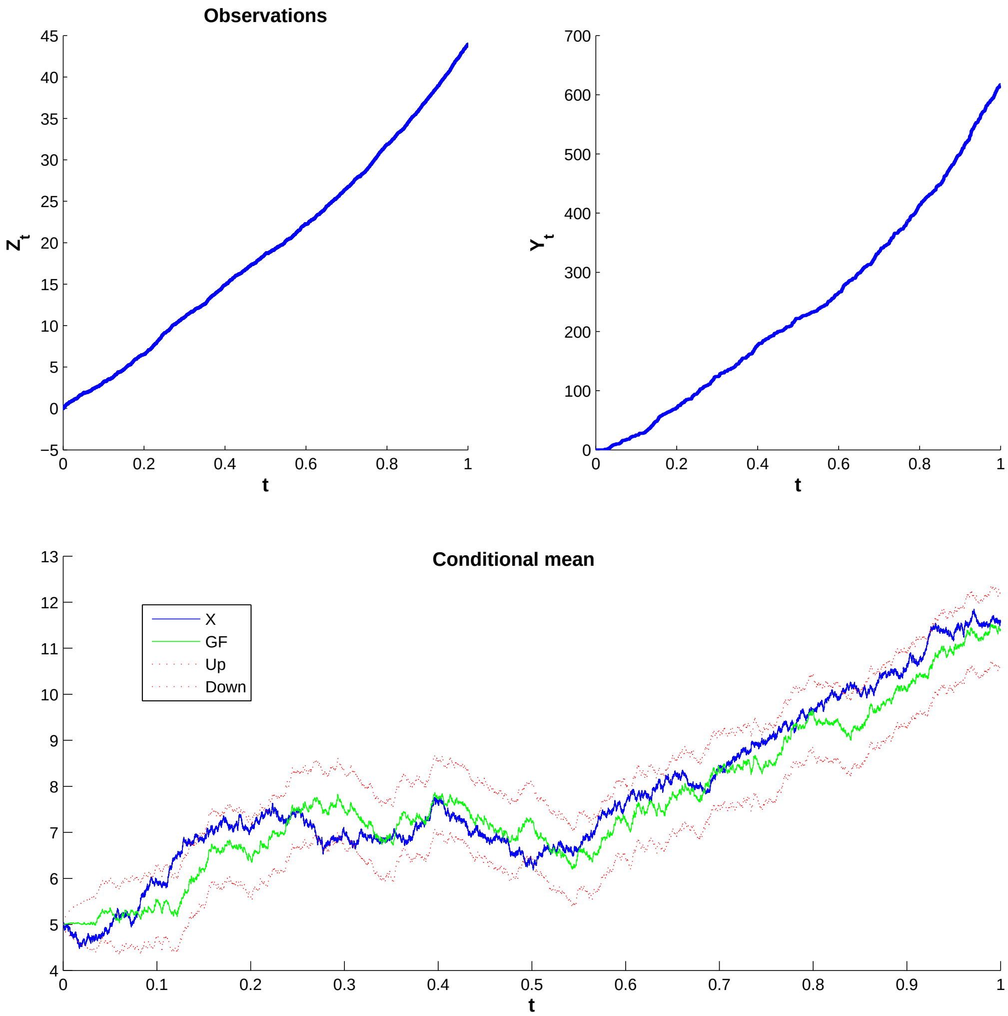

Figure 7.5: Galerkin filter with more information: A trajectory of Z (top row left), A trajectory of Y (top row right), Conditional mean (bottom figure). See Table 7.2 for notation interpretations.

Here 'up' means conditional mean $+1.64\times$ CRMSD, and 'down' means conditional mean $-1.64\times$ CRMSD. It is seen that with more information the results obtained by Galerkin filter is very close to the trajectory of state process X.

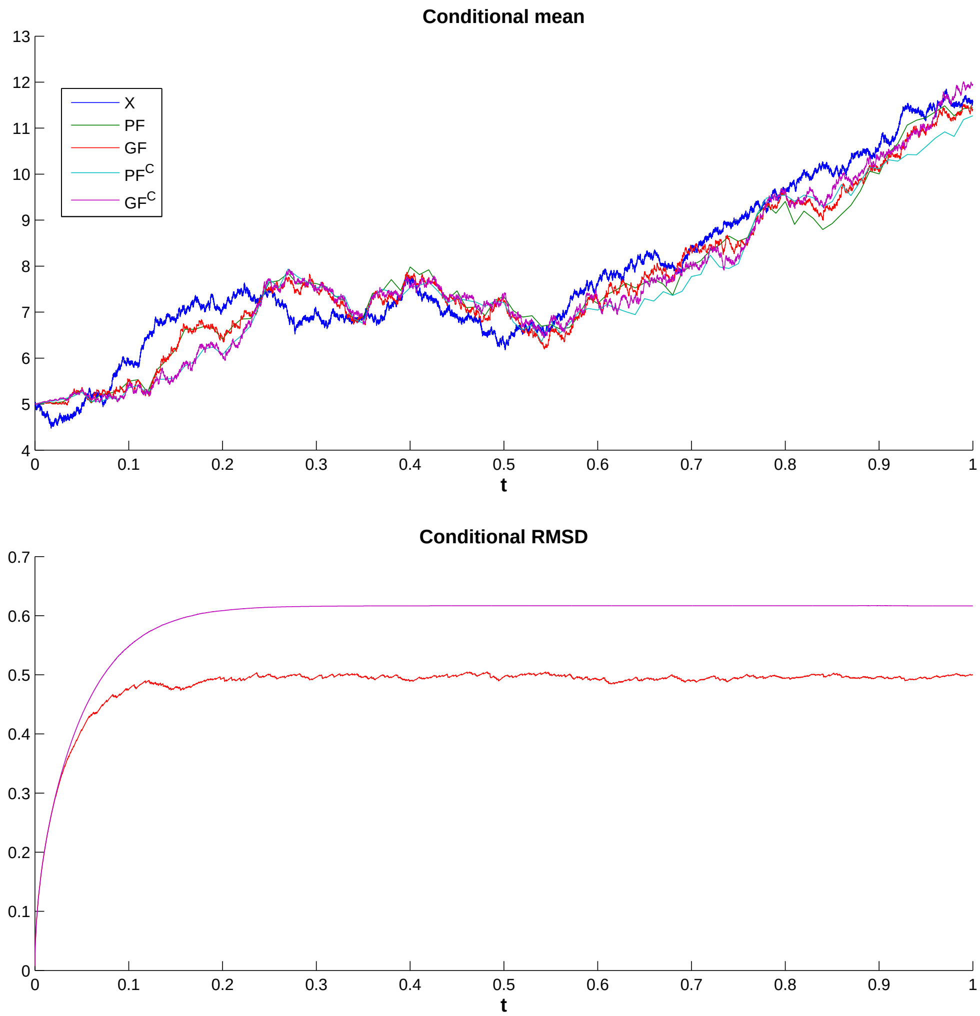

Figure 7.6: Comparison of Galerkin filter with particle filter with more information: Conditional mean (top figure), Conditional root mean square deviation(bottom figure). See Table 7.2 for notation interpretations.

It is seen that the approximation obtained by Galerkin filter is very close to that obtained by particle filter.

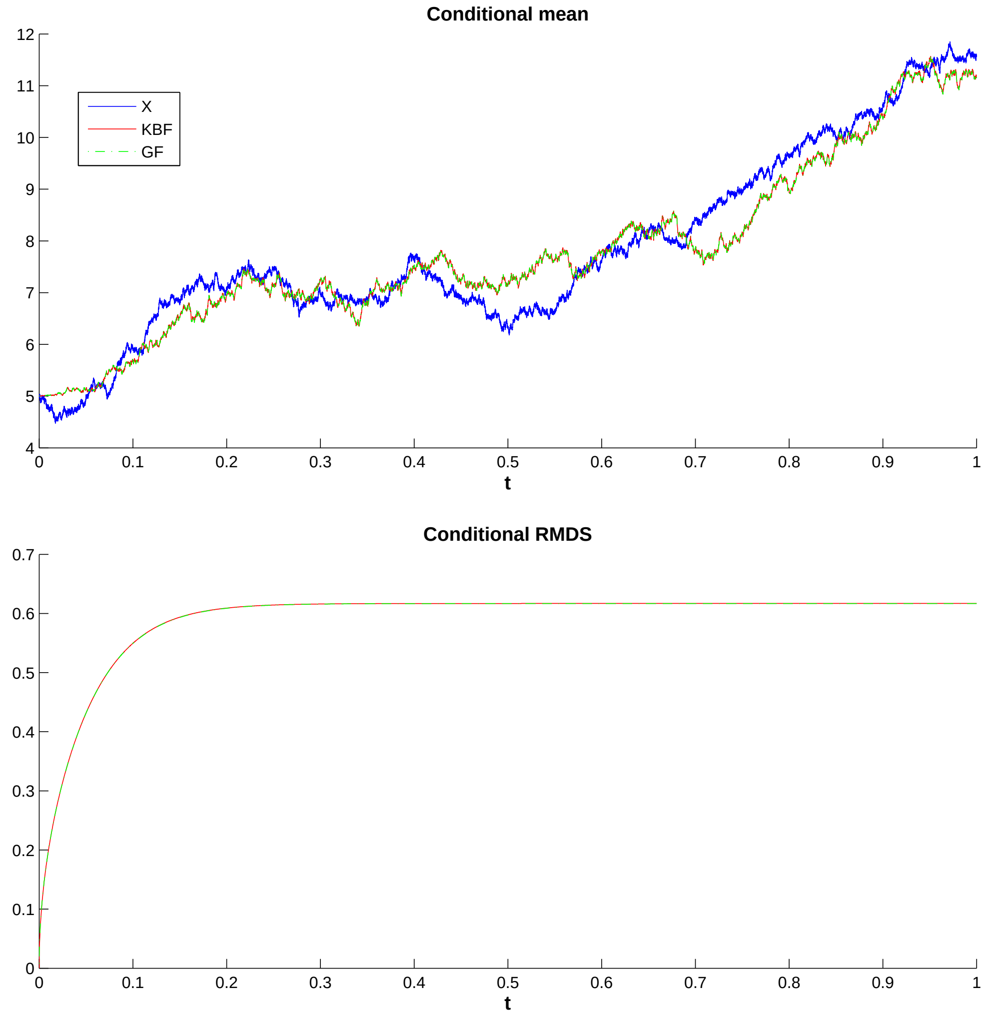

Figure 7.7: Comparison of Galerkin filter with Kalman-Bucy filter: Conditional mean (top figure), Conditional root mean square deviation(bottom figure). See Table 7.2 for notation interpretations.

It is seen that the approximations are very close to the explicit solutions.

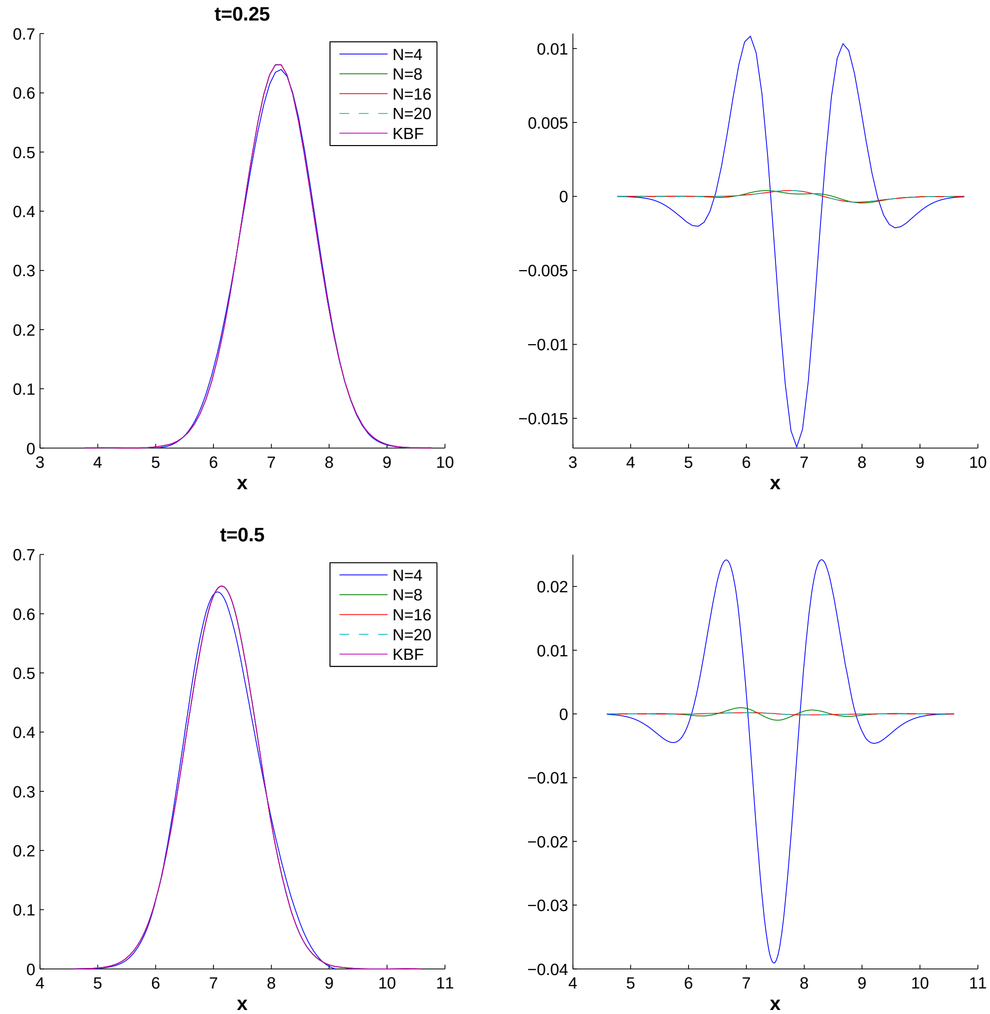

Figure 7.8: Comparison of Galerkin filter with Kalman-Bucy filter: Conditional density (figures left), the difference of conditional density obtained by Galerkin filter and Kalman-bucy filter(figures right). We present the results for different time t, $t = 0.25$(top row), $t = 0.5$(bottom row).

In this case, we take 20 Hermite basis functions in the Galerkin filter. In this figure, we present the approximated conditional density obtained by the linear combination of the first N basis functions. It is seen that the approximations are very close to the explicit solutions and filters with more basis functions improved the estimation procedure.

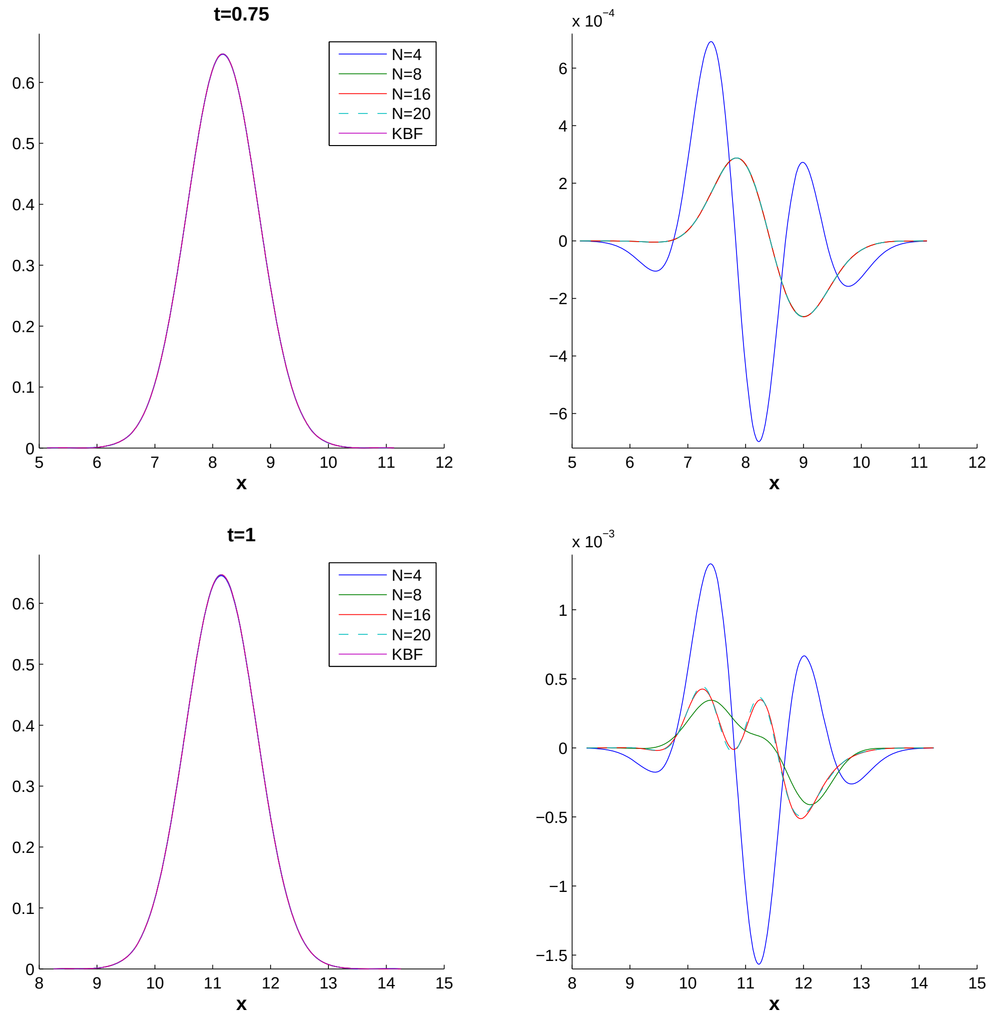

Figure 7.9: Comparison of Galerkin filter with Kalman-Bucy filter: Conditional density (figures left), the difference of conditional density obtained by Galerkin filter and Kalman-bucy filter(figures right). We present the results for different time t, $t = 0.75$(top row), $t = 1$(bottom row).

In this case, we take 20 Hermite basis functions in the Galerkin filter. In this figure, we present the approximated conditional density obtained by the linear combination of the first N basis functions. It is seen that the approximations are very close to the explicit solutions and filters with more basis functions improved the estimation procedure.

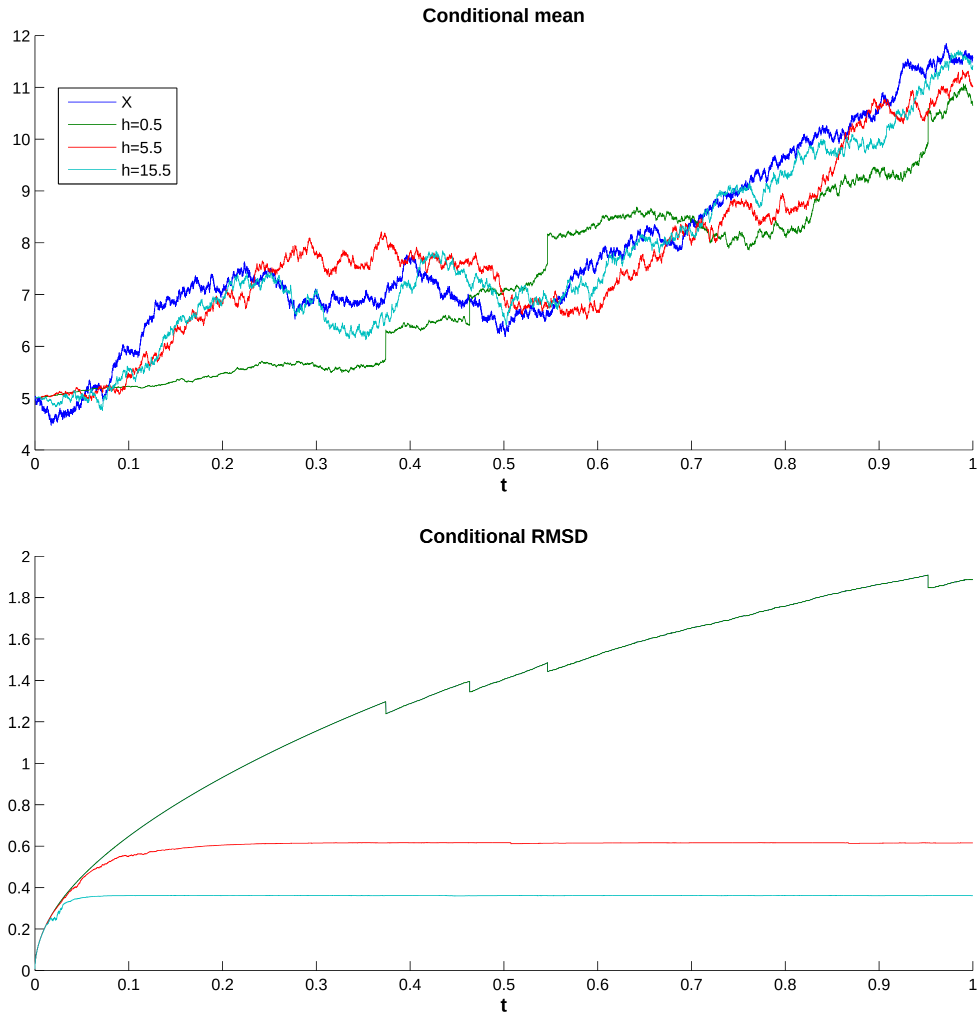

Figure 7.10: Convergence study for h: Conditional mean (top figure), Conditional root mean square deviation(bottom figure). See Table 7.2 for notations interpretations.

The results are obtained by Galerkin filter with observation Y and Z with different h, where h is the coefficient w.r.t. continuous observation Z for Example 7.1. The larger h is, the more information Z has. It is seen that the estimator converges to the trajectory of X as h increases.

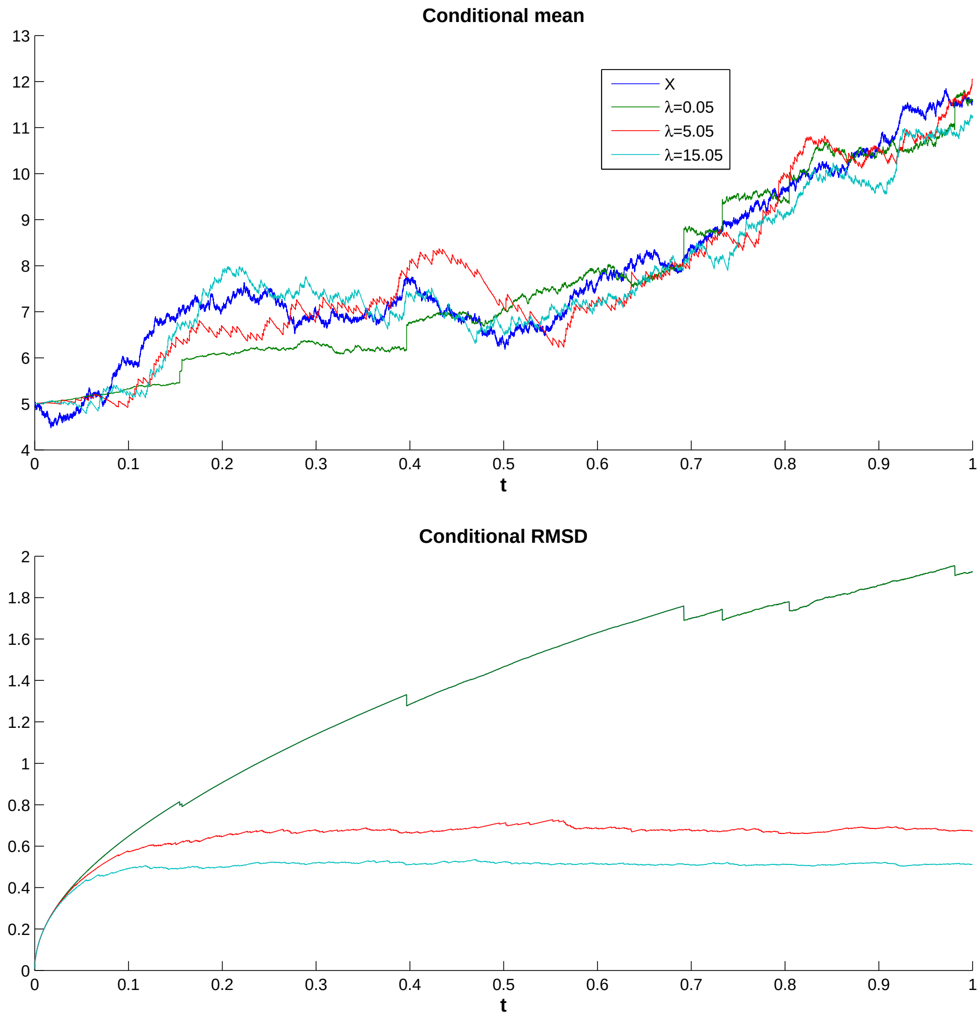

Figure 7.11: Convergence study for λ: Conditional mean (top figure), Conditional root mean square deviation(bottom figure). See Table 7.2 for notations interpretations.

The results are obtained by Galerkin filter with observation Y and Z with different λ, where λ is the intensity coefficient of the jump observation Y for Example 7.1. The larger λ is, the more information Y has. It is seen that the estimator converges to the trajectory of X as λ increases.

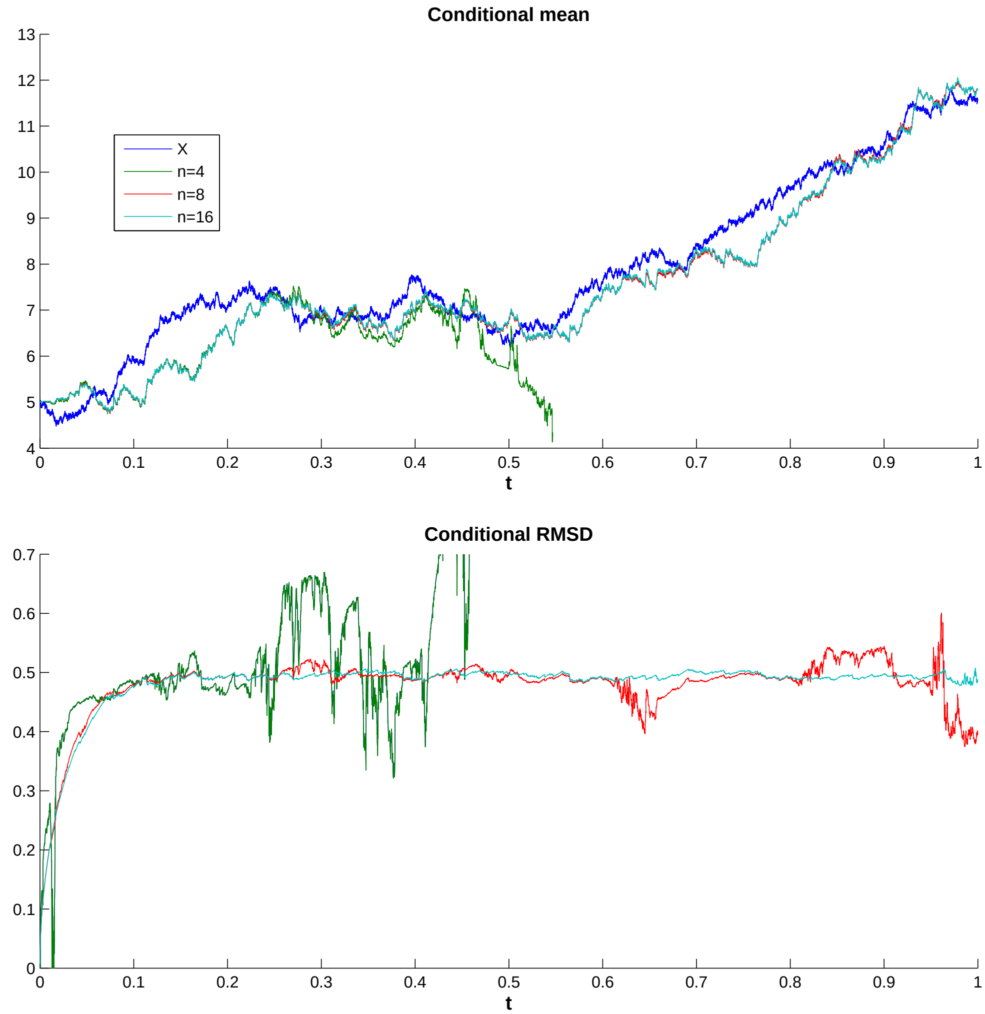

Figure 7.12: Convergence study for n: Conditional mean (top figure), Conditional root mean square deviation(bottom figure). See Table 7.2 for notations interpretations.

The results are obtained by Galerkin filter with observation Y and Z with different n, where n is the number of basis functions used in the Galerkin filter. The case $n = 4$ shows a bad performance as the number of the basis elements is too low. It is necessary to consider a satisfied high number of basis elements.

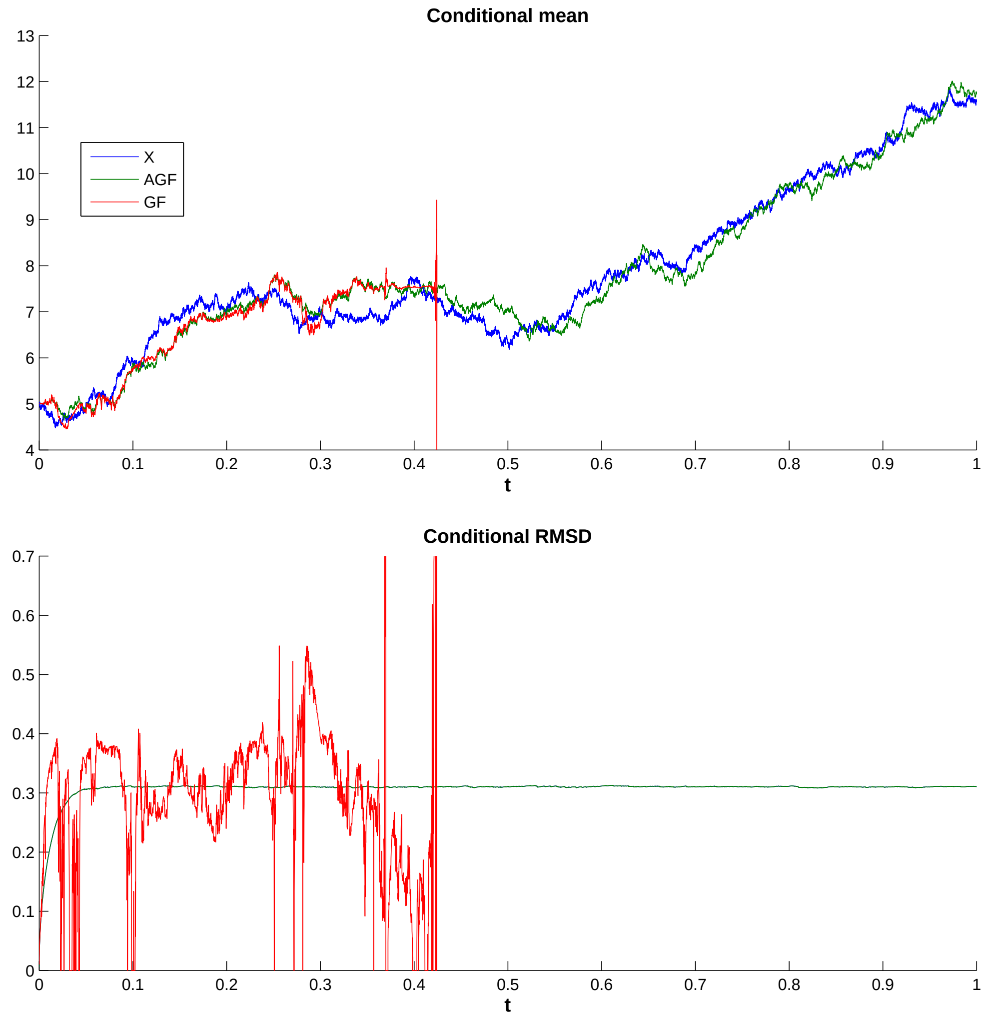

Figure 7.13: Galerkin filter: Conditional mean (top figure), Conditional root mean square deviation(bottom figure). See Table 7.2 for notations interpretations.

The results are obtained by Galerkin filter with observation Y and Z. One is adaptive Galerkin filter, the other is the usual Galerkin filter both with a basis of Hermite polynomials. It is seen that the adaptive Galerkin filter improves the estimation procedure.

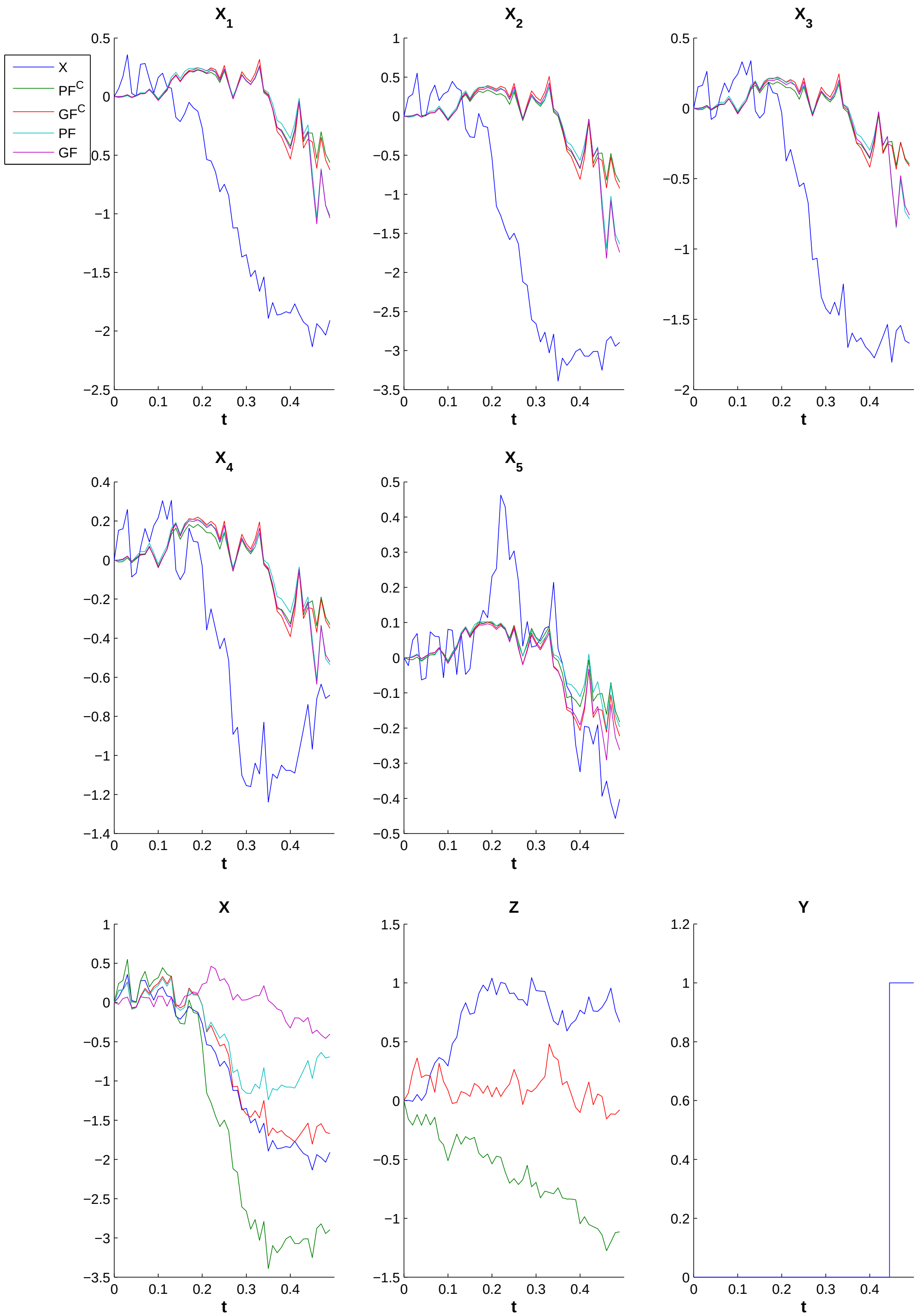

Figure 7.14: Comparison of Galerkin filter with particle filter (muti-dimensional case): Conditional mean (top figures and middle figures), trajectories of X (bottom row left), trajectories of Z (bottom row middle), a trajectory of Y (bottom row right). See Table 7.2 for notation interpretations.

It is seen that the approximation obtained by Galerkin filter is very close to that obtained by particle filter.

Appendix A

Appendix

For the reader's convenience we proof some well known results.

Lemma A.1. *Let Y be a stochastic random time with jump intensity 1 and ξ be the corresponding jump indicator process, then the moment-generating function of ξ is, $\forall u \in \mathbb{R}$,*

$$\mathbb{E}(e^{u\xi}) = e^{u}(1 - e^{-t}) + e^{-t}.$$

Proof. We have, by assumption,

$$\mathbb{E}(e^{u\xi}) = e^{u\cdot 1}\mathbb{P}(Y \leq t) + e^{u\cdot 0}\mathbb{P}(Y > t) = e^{u}(1 - e^{-t}) + e^{-t}.$$

□

Lemma A.2. *Let $X \in \mathbb{R}^+$ be a random variable, then, if $\mathbb{E}(X) < \infty$, $\mathbb{P}(X < \infty) = 1$.*

Lemma A.3. *For any $x \in \mathbb{R}$,*

$$\sum_{n=0}^{\infty} \frac{x^n}{(n!)^{\frac{1}{2}}} \leq \frac{2}{\sqrt{3}} e^{2x^2}.$$

Proof.

$$\begin{aligned}
\sum_{n=0}^{\infty} \frac{x^n}{(n!)^{\frac{1}{2}}} =& (\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{x^n}{(n!)^{\frac{1}{2}}} \frac{x^m}{(m!)^{\frac{1}{2}}})^{\frac{1}{2}} \\
=& (\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{2^{-m}(2x)^n}{(n!)^{\frac{1}{2}}} \frac{2^{-n}(2x)^m}{(m!)^{\frac{1}{2}}})^{\frac{1}{2}} \\
\leq& (\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{1}{2}(\frac{2^{-2m}(2x)^{2n}}{n!} + \frac{2^{-2n}(2x)^{2m}}{m!}))^{\frac{1}{2}} \\
=& (\sum_{n=0}^{\infty} 2^{-2n} \sum_{m=0}^{\infty} \frac{(2x)^{2m}}{m!})^{\frac{1}{2}} \\
=& (\frac{1}{1 - 2^{-2}} e^{4x^2})^{\frac{1}{2}} = \frac{2}{\sqrt{3}} e^{2x^2}.
\end{aligned}$$

□

Lemma A.4 (Gronwall's Lemma). *Let A and B are positive constants. If x is a non-negative function such that, for all $t \geq 0$,*

$$x_t \leq A + B \int_0^t x_s ds,$$

then, for all $t \geq 0$,

$$x_t \leq Ae^{Bt}.$$

Lemma A.5 (Burkholder-Davis-Gundy). *For a local martingale M starting at zero, with maximum denoted by $M_t = \sup_{s\in[0,T]} |M_s|$, and any real number $p \geq 1$, the inequality is*

$$c_p \mathbb{E}([M]_t^{p/2}) \leq \mathbb{E}((M_t^*)^p) \leq C_p \mathbb{E}([M]_t^{p/2}).$$

Here, $c_p < C_p$ are constants depending on the choice of p, but not depending on the martingale M or time t used. If M is a continuous local martingale, then the Burkholder-Davis-Gundy inequality holds for any positive value of p.

Theorem A.6 (Uniform boundedness principle). *Let U be a Banach space and V be a normed vector space. Suppose that F is a collection of continuous linear operators from U to V. The uniform boundedness principle states that if for all x in U we have*

$$\sup_{T\in F} \|T(x)\| < \infty,$$

then

$$\sup_{T\in F} \|T\| < \infty.$$

A.1 Preliminaries

In this section we introduce some basic concepts and properties that will be need in the development of the theory of SPEDs. The aim is to prepare the reader with necessary material for the study of SPDEs.

A.1.1 Sobolev spaces

In this section, the Sobolev spaces are introduced. The importance of Sobolev spaces comes from the fact that solutions of partial differential equations are often easier to be found in Sobolev spaces, rather than in spaces of continuous functions and with the derivatives understood in the classical sense.

A Sobolev space is a vector space of functions equipped with a norm that is a combination of L^2 norms of the function itself as well as its derivatives up to a given order. The derivatives are understood in a suitable weak sense to make the space complete. Let $\alpha := (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ with $|\alpha| := |\alpha_1| + \ldots + |\alpha_d|$ be an arbitrary multi-index. Given two function f and $g \in L^2(\mathbb{R}^d)$, we say that $\partial^\alpha f = g$ in the weak sense if for all $\phi \in C_0^\infty(\mathbb{R}^d)$, we have

$$\int_{\mathbb{R}^d} f(x)\partial^\alpha \phi(x)dx = (-1)^{|\alpha|} \int_{\mathbb{R}^d} g(x)\phi(x)dx,$$

where

$$\partial^\alpha \phi = \frac{\partial^\alpha \phi}{\partial x_1^{\alpha_1} \cdot \partial x_d^{\alpha_d}}.$$

Let us now recall the definition of the Sobolev spaces. For $m = 0, 1, 2, \ldots$, the Sobolev space, denote by $W_m^p(\mathbb{R}^d)$, is the space of all functions $v \in L^p(\mathbb{R}^d)$ such that the partial derivatives $\partial^\alpha v$ exist in the weak sense and are in $L^p(\mathbb{R}^d)$ whenever $|\alpha| \leq m$, that is,

$$W_m^p(\mathbb{R}^d) := \{v : \partial^\alpha v \in L^p(\mathbb{R}^d), \ \ |\alpha| \leq m\},$$

with the norm

$$\|v\|_{W_m^p} := \Big\{ \sum_{|\alpha| \leq m} \int_{\mathbb{R}^d} |\partial^\alpha v(x)|^p dx \Big\}^{1/p}. \tag{A.1.1}$$

$W_m^p(\mathbb{R}^d)$ is complete with respect to the norm defined by Equation (A.1.1), hence it is a Banach space.

Sobolev spaces with $p = 2$, defined as $H^m(\mathbb{R}^d) := W_m^2(\mathbb{R}^d)$, are especially important because they form a Hilbert space with the inner product

$$(f, g)_{H^m} = \sum_{|\alpha| \leq m} (\partial^\alpha f, \partial^\alpha g), \ \ \forall f, g \in H^m(\mathbb{R}^d),$$

where $(\cdot, \cdot)$ is the usual inner product on $L^2(\mathbb{R}^d)$,

$$(f, g) := \int_{\mathbb{R}^d} f(x) g(x) dx, \ \ \forall f, g \in L^2(\mathbb{R}^d).$$

For $m = -1, -2, \ldots$, define $H^m(\mathbb{R}^d) := (H^{-m}(\mathbb{R}^d))^*$. For any $m \in \mathbb{Z}$, $H^m(\mathbb{R}^d)$ is a Hilbert space, see Folland (1999), page 302, with the inner product, $\forall f, g \in H^m(\mathbb{R}^d)$,

$$(f, g)_{H^m} := \int_{\mathbb{R}^d} \widehat{f}(\xi)(1 + |\xi|^2)^m \overline{\widehat{g}(\xi)} d\xi,$$

where $\widehat{\cdot}$ defined as the Fourier transform and $\bar{\cdot}$ defined as the conjugate of the complex number.

A.1.2 Some spaces of processes

We look for solutions of SPDEs in a suitably chosen processes space. From now on, we define some spaces of stochastic processes with values in Hilbert spaces. In this section, let $\mathcal{H}$ be a real separable Hilbert space.

Let $D(0, T; \mathcal{H})$ be the space of $\mathcal{H}$-valued functions ξ on $[0, T]$ that are right-continuous and have left-hand limits:

- i For $0 \leq t < T$, $\xi(t+) = \lim_{s \downarrow t} \xi(s)$ and $\xi(t+) = \xi(t)$.
- ii For $0 < t \leq T$, $\xi(t-) = \lim_{s \uparrow t} \xi(s)$ exists.

Lemma A.7. *$D(0, T; \mathcal{H})$ is a Banach space with norm $\|x\| := \sup_{t \in [0,T]} \|x_t\|_{\mathcal{H}}$.*

Proof. See Billingsley (1999), page 124. □

A stochastic process ξ is said to be càdlàg if it a.s. has sample paths in $D(0,T;\mathcal{H})$. That is ξ maps Ω into $D(0,T;\mathcal{H})$.

Let $\mathcal{M}^2(0,T;\mathcal{H})$, more simply $\mathcal{M}^2(0,T)$ or even $\mathcal{M}^2$, denote the space of $\mathcal{H}$-valued processes ϕ which satisfy:

i ϕ is $\mathbb{F}^{Z,Y}$-adapted,

ii $\mathbb{E}^0\Big\{\int_0^T \|\phi(t)\|^2_{\mathcal{H}}\Big\} < \infty$.

Let $\mathcal{S}^2(0,T;\mathcal{H})$ denote the set of $\mathbb{F}^{Z,Y}$-adapted càdlàg processes $\{\xi(t), 0 \le t \le T\}$ which are such that

$$\|\xi\|_T := \Big\{\mathbb{E}^0 \sup_{t\in[0,T]} \|\xi(t)\|^2_{\mathcal{H}}\Big\}^{1/2} < \infty. \tag{A.1.2}$$

Let $\mathcal{S}^2_w(0,T;\mathcal{H})$, more simply $\mathcal{S}^2_w(0,T)$ or even $\mathcal{S}^2_w$, denote the set of $\mathbb{F}^{Z,Y}$-adapted càdlàg processes $\{\xi(t), 0 \le t \le T\}$ which are such that

$$|\xi|_T := \Big\{\sup_{t\in[0,T]} \mathbb{E}^0\|\xi(t)\|^2_{\mathcal{H}}\Big\}^{1/2} < \infty. \tag{A.1.3}$$

Let $\mathcal{C}^2(0,T;\mathcal{H})$ denote the set of $\mathbb{F}^{Z,Y}$-adapted continuous processes $\{\xi(t), 0 \le t \le T\}$ which are such that

$$\|\xi\|_T := \Big\{\mathbb{E}^0 \sup_{t\in[0,T]} \|\xi(t)\|^2_{\mathcal{H}}\Big\}^{1/2} < \infty.$$

Let $\mathcal{C}^2_w(0,T;\mathcal{H})$, denote the set of $\mathbb{F}^{Z,Y}$-adapted continuous processes $\{\xi(t), 0 \le t \le T\}$ which are such that

$$|\xi|_T := \Big\{\sup_{t\in[0,T]} \mathbb{E}^0\|\xi(t)\|^2_{\mathcal{H}}\Big\}^{1/2} < \infty.$$

Let $\mathcal{N}^2(0,T;\mathcal{H})$ denote the set of $\mathbb{F}^{Z,Y}$-adapted, $\mathcal{H}$-valued processes $\{\xi(t), 0 \le t \le T\}$, continuous in the mean square norm, which are such that

$$|\xi|_T := \Big\{\sup_{t\in[0,T]} \mathbb{E}^0\|\xi(t)\|^2_{\mathcal{H}}\Big\}^{1/2} < \infty. \tag{A.1.4}$$

It is easy to see that $|\cdot|_T < \|\cdot\|_T$. Therefore $\mathcal{S}^2(0,T;\mathcal{H}) \subset \mathcal{S}^2_w(0,T;\mathcal{H})$ and $\mathcal{C}^2(0,T;\mathcal{H}) \subset \mathcal{C}^2_w(0,T;\mathcal{H})$. And we have the following results:

Lemma A.8. *$\mathcal{N}^2(0,T;\mathcal{H})$ is a Banach space with norm $|\cdot|_T$.*

Proof. See Germani and Piccioni (1984). □

Lemma A.9. *$\mathcal{S}^2(0,T;\mathcal{H})$ is a Banach space with norm $\|\cdot\|_T$.*

Proof. Assume $\{\xi_n\}$ is a Cauchy sequence, i.e.,

$$\mathbb{E}^0\Big\{\sup_{t\in[0,T]}\|\xi_m(t)-\xi_n(t)\|^2_{\mathcal{H}}\Big\}\to 0,\ \text{ as } n,m\to\infty.$$

It follows that one can find a subsequence ξ_{n_k} such that

$$\mathbb{P}^0\Big\{\sup_{t\in[0,T]}\|\xi_{n_{k+1}}(t)-\xi_{n_k}(t)\|_{\mathcal{H}}\geq 2^{-k}\Big\}\leq 2^{-k}.$$

The Borel-Cantelli lemma implies that $\{\xi_{n_k}\}$ converges $\mathbb{P}^0$-a.s. to a process $\{\xi(t),\ t\in[0,T]\}$, uniformly on $[0,T]$. By Lemma A.7, ξ is a càdlàg process. □

Lemma A.10. *$C^2(0,T;\mathcal{H})$ is a Banach space.*

Proof. The proof is analogous to the proof of Lemma A.9. □

Remark A.1. *$C^2_w(0,T;\mathcal{H})$ is not a Banach space.*

Proof. It is sufficient to give a counter example. For reason of simplicity, set $T=1$. Let U is a $\mathbb{R}$-valued random variable with uniform distribution $U(0,1)$. Let $\{\sigma_n, n\geq 1\}$ is a $\mathbb{R}^+$-valued sequence and $\sigma_n\to 0$ as $n\to\infty$. For a fixed $v\in\mathcal{H}$, set

$$\begin{aligned}&\xi_n(t,U)=v\Big(\mathbf{1}_{\{0\leq t\leq U\}}+\mathbf{1}_{\{U<t\leq 1\}}e^{-\frac{(t-U)^2}{4\sigma_n^2}}\Big),\\&\xi(t,U)=v\mathbf{1}_{\{t<U\}}.\end{aligned}$$

Notice, $\xi_n(\cdot,U)$ is continuous in t, while $\xi(\cdot,U)$ is not. Then,

$$\begin{aligned}\mathbb{E}^0\|\xi_n(t)-\xi(t)\|^2_{\mathcal{H}}&=\int_0^1\|\xi_n(t,\nu)-\xi(t,\nu)\|^2_{\mathcal{H}}d\nu\\&=\int_t^1\|\xi_n(t,\nu)\|^2_{\mathcal{H}}d\nu\\&=\|v\|^2_{\mathcal{H}}\int_t^1 e^{-\frac{(t-\nu)^2}{2\sigma_n^2}}d\nu\\&\leq\|v\|^2_{\mathcal{H}}\int_{\mathbb{R}}e^{-\frac{(t-\nu)^2}{2\sigma_n^2}}d\nu\\&=\|v\|^2_{\mathcal{H}}\sqrt{2\pi}\sigma_n\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}\sigma_n}e^{-\frac{(t-\nu)^2}{2\sigma_n^2}}d\nu\\&=\|v\|^2_{\mathcal{H}}\sqrt{2\pi}\sigma_n.\end{aligned}$$

By the definition of $|\cdot|_T$,

$$|\xi_n-\xi|_T=\Big\{\sup_{t\in[0,T]}\mathbb{E}^0\|\xi_n(t)-\xi(t)\|^2_{\mathcal{H}}\Big\}^{\frac{1}{2}}\leq\|v\|_{\mathcal{H}}\sqrt{\sqrt{2\pi}\sigma_n}\to 0,$$

as $n\to\infty$.

Therefore, ξ_n converges to ξ in $S^2_w(0,T;\mathcal{H})$. But ξ is a non-continuous function and not in $\mathcal{C}^2_w(0,T;\mathcal{H})$. So $\mathcal{C}^2_w(0,T;\mathcal{H})$ is not complete and in sequence it is not a Banach space.

But this is not a counter example for Lemma A.10. Notice that

$$\|\xi_n - \xi\|_T = \left\{\mathbb{E}^0 \sup_{t\in[0,T]} \|\xi_n(t) - \xi(t)\|^2_{\mathcal{H}}\right\}^{\frac{1}{2}} = 1,$$

ξ_n doesn't converge to ξ in $\mathcal{C}^2(0,T;H)$. □

A.2 Some Hilbert spaces

Let $\mathcal{H}$ be a real separable Hilbert space. $(\bar{\Omega}, \mathcal{G}, \mathbb{Q})$ be a probability space. Let $\{\mathcal{G}_t\}$, which satisfies the usual conditions, is the filtration.

For $0 \le s \le \theta \le T$, let $L^2\Big((\bar{\Omega}, \mathcal{G}, \mathbb{Q}, \{\mathcal{G}_t\}) \times [s,\theta]; \mathcal{H}\Big)$ denote the space of $\mathcal{H}$-valued processes $\{\xi(t), s \le t \le \theta\}$ which satisfy

i ξ is $\{\mathcal{G}_t\}$-adapted.

ii $\mathbb{E}^{\mathbb{Q}}\Big\{\int_s^\theta \|\xi(t)\|^2_{\mathcal{H}} dt\Big\} < \infty$.

Let $L^2\Big((\bar{\Omega}, \mathcal{G}, \mathbb{Q}); \mathcal{H}\Big)$ denote the space of $\mathcal{H}$ random variables φ which satisfy

i φ is $\mathcal{G}$-measurable.

ii $\mathbb{E}^{\mathbb{Q}}\Big\{\|\varphi\|^2_{\mathcal{H}}\Big\} < \infty$.

Appendix B

List of frequently used notation and symbols

$\mathbb{R}^k$	k-dimensional Euclidean space
$\mathbb{R}^+$	the non-negative real numbers
$\mathbb{R}^{n\times m}$	the $n \times m$ matrices
C_0^∞	the set of infinitely differentiable functions with compact support
$L^k(\mathbb{R}^d)$	$\{u : \mathbb{R}^d \to \mathbb{R}$ measurable, with $\|u\|_k := \|u\|_{L^k(\mathbb{R}^d)} := [\int_{\mathbb{R}^d} \|u(x)\|^k dx]^{\frac{1}{k}} < \infty\}$
$L^\infty(\mathbb{R}^d)$	$\{u : \mathbb{R}^d \to \mathbb{R}$ measurable, $\|u\|_{L^\infty(\mathbb{R}^d)} : \sup_{x\in\mathbb{R}^d} \|u(x)\| < \infty\}$
$s \wedge t$	the minimum of s and t
$s \vee t$	the maximum of s and t
$D(A)$	the domain of definition of the operator A
w.r.t.	with respect to
s.t.	such that
a.e.	almost everywhere
a.s.	almost surely
$\sim$	coincides in law with
$\perp$	orthogonal to
$:=$	equal to by definition
$\|\cdot\|_\infty$	the supremum norm of a function
$\|\cdot\|$	the Euclidean norm
$\mathcal{F}_t^\xi$	the σ-algebra generated by $\{\xi_s, 0 \le s \le t\}$
$\mathcal{H}$	a real separable Hilbert space
$\mathcal{B}$	a Banach space
H	$L^2(\mathbb{R}^d)$
V	$H^1(\mathbb{R}^d)$
V'	$H^{-1}(\mathbb{R}^d)$
$C([0,T];\mathcal{H})$	the continuous functions $[0,T] \to \mathcal{H}$
$C^1\Big((0,T);\mathcal{H}\Big)$	the once continuous differentiable functions $(0,T) \to \mathcal{H}$
$D([0,T];\mathcal{H})$	the right-continuous functions $[0,T] \to \mathcal{H}$ have left limits in $\mathcal{H}$

$\mathbb{E}$	the expectation wrt. the probability measure $\mathbb{P}$
$\mathbb{E}^0$	the expectation wrt. the probability measure $\mathbb{P}^0$
I_n	the n-by-n identity matrix
C_k^m	the number of m-combinations from a given set of k elements
$\mathbf{1}_G$	the indicator function of set G, ($\mathbf{1}_G(x) = 1$ if $x \in G$, $\mathbf{1}_G(x) = 0$ otherwise)
$N(\mu, \sigma^2)$	a normal distribution with mean μ and variance σ^2
$\ln N(\mu, \sigma^2)$	a log-normal distribution where μ and σ are the mean and standard deviation of the variable's natural logarithm.

Bibliography

Ahmed, N. U. and S. M. Radaideh (1997). A powerful numerical technique solving Zakai equation for nonlinear filtering. *Dynamics and Control 7*(3), 293–308.

Bain, A. and D. Crisan (2009). *Fundamentals of Stochastic filtering*. New York: Springer.

Bensoussan, A., R. Glowinski, and A. Rascanu (1990). Approximation of the Zakai equation by the splitting up method. *SIAM J. Control Optim. 28*(6), 1420–1431.

Berkowitz, S. and F. J. Garner (1970). The calculation of multidimensional Hermite polynomials and Gram-Charlier coefficients. *Mathematics of Computation 24*, 537–545.

Billingsley, P. (1999). *Convergence of Probability Measures* (2nd ed.). New York: John Wiley & Sons, Inc.

Black, F. and J. C. Cox (1976, May). Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance 31*(2), 351–367.

Black, F. and M. Scholes (1973, May/June). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Blanchet-Scalliet, C. and M. Jeanblanc (2004, Jan.). Hazard rate for credit risk and hedging defaultable contingent claims. *Finance and stochastics 8*(1), 145–159.

Bongioanni, B. and J. L. Torrea (2006). Sobolev spaces associated to the harmonic oscillator. *Proc. Indian Acad. Sci. Math. Sci. 116*(3), 337–360.

Brémaud, P. (1981). *Point Processes and Queues*. New York-Berlin: Springer.

Brémaud, P. (1972). *A Martingale Approach to Point Processes*. Ph. D. thesis, University of California, Berkeley, Department of EECS.

Budhiraja, A., L. Chen, and C. Lee (2007). A survey of numerical methods for nonlinear filtering problems. *Phys. D 230*(1-2), 27–36.

Cai, Z., F. L. Gland, and H. Zhang (1995, Oct.). An adaptive local grid refinement method for nonlinear filtering. Technical Report 2679, IRISA Campus de Beaulieu Internal.

Carpenter, J., P. Clifford, and P. Fearnhead (1999, Feb.). Improved particle filter for nonlinear problems. *IEE Proc. Radar, Sonar and Navigation 146*(1), 2–7.

Coculescu, D., H. Geman, and M. Jeanblanc (2008). Valuation of default-sensitive claims under imperfect information. *Finance Stoch. 12*(2), 195–218.

Collin-dufresne, P., R. S. Goldstein, and J. Helwege (2003). Is credit event risk priced? Modeling contagion via the updating of beliefs. Working paper.

Courant, R. and D. Hilbert (1968). *Methoden der Mathematischen Physik*. Springer.

Crisan, D., P. D. Moral, and T. Lyons (1999). Discrete filtering using branching and interacting particle systems. *Markov Processes and Related Fields 5*(3), 293–318.

Davis, M. H. A. and S. I. Marcus (1981). An introduction to nonlinear filtering. *Stochastic Systems : the Mathematics of Filtering and Identification and Applications*, 53–75.

Duffie, D. and D. Lando (2001, May). Term structure of credit risk with incomplete accounting observations. *Econometrica 69*(3), 633–664.

Duffie, D. and K. Singleton (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies 12*(4), 687–720.

Duncan, T. E. (1967). *Probability Densities for Diffusion Processes*. Ph. D. thesis, Stanford University, Department of Electrical. Engineering.

Dupuis, P. G. and H. J. Kushner (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*. New York: Springer.

Ethier, S. and T. G. Kurtz (1986). *Markov Processes: Characterization and Convergence*. New York: John Wiley & Sons, Inc.

Filipovic, D., S. Tappe, and J. Teichmann (2010, Oct.). Jump-diffusions in Hilbert spaces: Existence, stability and numerics. *Stochastics 82*(5), 475–520.

Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications* (2nd ed.). New York: John Wiley & Sons, Inc.

Fontana, C. and W. J. Runggaldier (2010). Credit risk and incomplete information: filtering and EM parameter estimation. *International Journal of Theoretical and Applied Finance 13*(1), 683–715.

Frey, P. (2008). Numerical simulation of complex PDE problems. Lecture note.

Frey, R. (2000). Risk minimization with incomplete information in a model for high-frequency data. *Math. Finance 10*(2), 215–225.

Frey, R. and W. J. Runggaldier (2010). Credit risk and incomplete information : a nonlinear-filtering approach. *Finance and Stochastics*.

Frey, R. and T. Schmidt (2009). Pricing corporate securities under noisy asset information. *Mathematical Finance 19*(3), 403–421.

Frey, R. and T. Schmidt (2010). Pricing and hedging of credit derivatives via the innovations approach to nonlinear filtering. *Finance and Stochastics*.

Gawarecki, L. and V. Mandrekar (2001). Stochastic differential equations with discontinuous drift in Hilbert space with applications to interacting particle systems. *J. Math. Sci. (New York) 105*(6), 2550–2554.

Germani, A. and M. Piccioni (1984). A Galerkin approximation for the Zakai equation. *Lecture Notes in Control and Information Sciences 59*, 415–423.

Germani, A. and M. Piccioni (1987). Finite-dimensional approximations for the equation of nonlinear filtering derived in mild form. *Applied Mathematics and Optimization 16*(1), 51–72.

Giesecke, K. (2004, July). Correlated default with incomplete information. *Journal of Banking and Finance. 28*(7), 1521–1545.

Grecksch, W. and P. E. Kloeden (1996). Time discretized Galerkin approximations of stochastic parapolic PDEs. *Bull. Austral. Math. Soc. 54*, 79–85.

Gyöngy (1988). On the approximation of stochastic partial differential equations. I. *Stochastics 25*(2), 59–85.

Gyöngy (1989). On the approximation of stochastic partial differential equations. II. *Stoch. Stoch. Rep. 26*(3), 129–164.

Hairer, M. (2009). An introduction to stochastic PDEs. Lecture note. Lecture note.

Hausenblas, E. and J. Seidler (2001). A note on maximal inequality for stochastic convolutions. *51(126)*(4), 785–790.

Jacod, J. and A. Shiryaev (2003). *Limit Theorems for Stochastic Processes* (2nd ed.). Berlin: Springer.

Jarrow, R. and P. Protter (2004, May). Structural versus reduced form models: a new information based perspective. *Journal of Investment Management 2*(2), 1–10.

Jarrow, R. A. and S. M. Turnbull (1995, March). Pricing derivatives on financial securities subject to credit risk. *Journal of Finance 50*(1), 53–85.

Kallianpur, G. and C. Striebel (1968). Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors. *Ann. Math. Statist. 39*(3), 785–801.

Kallianpur, G. and J. Xiong (1995). *Stochastic Differential Equations in Infinite Dimensional Spaces*, Volume 26 of *Lecture Notes-Monograph Series.* Institute of Mathematical Statistics, Hayward.

Karoui, N. E., S. Peng, and M. C. Quenez (1997). Backward stochastic differential equations in finance. *Math. Finance 7*(1), 1–71.

Knoche, C. (2004). SPDEs in infinite dimension with Poisson noise. *C. R. Math. Acad. Sci. Paris 339*(9), 647–652.

Knoche, C. (2005). Existence, uniqueness and regularity w.r.t. the initial condition of mild solutions to SPDEs driven by Poisson noise. Working paper.

Kusuoka, S. (1999). A remark on default risk models. *Advances in Mathematical Economics 1*, 69–82.

L. Gawarecki, V. Mandrekar, P. R. (1999). Existence of weak solutions for stochastic differential equations and martingale solutions for martingale solutions for stochastic semilinear equations. *Random Oper. Stochastic Equations 7*(3), 215–240.

Lando, D. (1998, Dec.). On Cox processes and credit risky securities. *Derivatives Research 2*(2-3), 99–120.

Le Gland, F. (1992). Splitting-up approximation for SPDEs and SDEs with application to nonlinear filtering. *Lecture Notes in Control and Information Sciences 176*, 177–187.

Lions, J. L. (1961). *Equations Différentielles Opérationnelles et Problèmes aux Limites.* Berlin-Göttingen-Heidelberg: Springer.

Liptser, R. S. and A. N. Shiryaev (1974a). *Stochastic of Random Processes. I* (2nd ed.). Springer Verlag. Berlin Heidelberg New York.

Liptser, R. S. and A. N. Shiryaev (1974b). *Stochastic of Random Processes. II* (2nd ed.). Springer Verlag. Berlin Heidelberg New York.

Lototsky, S. V. (2006). Wiener chaos and nonlinear filtering. *Appl. Math. Optim. 54*(3), 265–291.

McLachlan, G. J. and T. Krishnan (1997). *The EM Algorithm and Extensions* (2nd ed.). John Wiley & Sons. New York.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management*. Princeton University Press.

Merton, R. C. (1974, May). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance 29*(2), 449–470.

Mortensen, R. E. (1966). *Optimal Control of Continuous-time Stochastic Systems*. Ph. D. thesis, university of California, Berkeley, Department of Electrical. Engineering.

Øksendal, B. (1980). *Stochastic Differential Equations* (3nd ed.). Berlin-Heidelberg: Springer.

Pardoux, E. (1979a). Filtering of a diffusion process with Poission-type observation. In *Stochastic Control Theory and Stochastic Differential Systems*, Volume 16 of *Lecture Notes in Contr. and Inf. Sci.*, pp. 510–518. Berlin- Heidelberg: Springer.

Pardoux, E. (1979b). Stochastic partial differential equations and filtering of diffusion process. *Stochastics 3*(2), 127–167.

Peszat, P. and J. Zabczyk (2007). *Stochastic Partial Differential Equations with Lévy Noise* (1nd ed.). Cambridge University Press.

Prato, G. D. and J. Zabczyk (1992). *Stochastic Equations in Infinte Dimensions* (1nd ed.). Cambridge University Press.

Protter, P. E. (2005). *Stochastic Integration and Differential Equations* (2nd ed.). Berlin: Springer.

Revuz, D. and M. Yor (1999). *Continuous martingales and Brownian Motion* (3nd ed.). Berlin: Springer.

Schönbucher, P. (2004). Information-driven default contagion. Working paper.

Showalter, R. E. (1977). *Hilbert space Methods for Partial Differential Equations*, Volume 1 of *Monographs and Studies in Mathematics*. Pitman.

Shreve, S. E. (2004). *Stochastic Calculus for Finance. II*. New York: Springer.

Singer, H. (2006). Moment equations and Hermite expansion for nonlinear stochastic differential equations with application to stock price models. *Comput. Statist. 21*(3-4), 385–397.

Stroock, D. W. (1975). Diffusion processes associated with Lévy generators. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 32*(3), 209–244.

Twardowska, K., T. Marnik, and M. Paslawska-Poludniak (2003). Approximation of the Zakai equation in a nonlinear filtering problem with delay. *Int. J. Appl. Math. Comput. Sci. 13*(2), 151–160.

Willink, R. (2005). Normal moments and Hermite polynomials. *Statistics and Probability Letters 73*(3), 271–275.

Zakai, M. (1969). On the optimal filtering of diffusion processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 11*, 230–243.

Daten zum Autor

Name:	Ling Xu
Geburtsdatum:	22. December 1977 in Herbei
09/1996 - 06/2000	Studium der Mathematik in Tianjin University, Tianjin
09/2000 - 06/2003	Master in Mathematik Institut Tianjin University, Tianjin
09/2003 - 06/2004	Mitarbeiter der Mathematik Institut Tianjin University, Tianjin
12/2004 - 09/2010	Doktorand am Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig

Bibliographische Daten

A Galerkin Approximation for the Zakai Equation
Xu, Ling
Universität Leipzig, Dissertation, 2010
170 Seiten, 14 Abbildungen, 5 Tables, 76 Referenzen

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

Leipzig, den 23.11.2010

. . . .
(Ling Xu)