PROBLEMS IN STATISTICS OF STOCHASTIC DIFFERENTIAL EQUATIONS

by

Sergey Vladimir Lototsky

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(Applied Mathematics)

August 1996

Copyright 1997 Sergey Vladimir Lototsky

Acknowledgments

First, I would like to thank my academic advisor prof. B. L. Rozovskii for suggesting the interesting topic of the research and guiding me through it. Without his help and encouragement this dissertation could hardly be written.

I also appreciate the help of the members of my committee — prof. P. Baxendale and prof. I. Reed.

My special thanks go to prof. R. Sh. Liptser, who introduced me into the Theory of Probability and Stochastic Processes and supervised my first research in the area.

Contents

Abstract

Two problems involving stochastic partial differential equations (SPDEs) are considered: nonlinear filtering of diffusion processes and parameter estimation for SPDEs.

A new approach to nonlinear filtering is developed using the Wiener chaos decomposition of the optimal filter. Based on this approach, two recursive algorithms are suggested for computing an approximation of the optimal filter. While both approximations converge to the optimal filter, one provides high resolution in space and the other in time. The computational complexity of the algorithms is studied and the approximation errors are derived. The existing algorithms require on-line solution of partial differential equations and evaluation of integrals, which is time consuming. The on-line performance of both proposed algorithms can be significantly improved by performing all the time consuming operations off line.

In the second part of the dissertation, the parameter estimation problem for a stochastic partial differential equation is considered. The estimate is based on a finite dimensional projection of the solution, and the dimension of the projection is the only asymptotic parameter to guarantee consistency and asymptotic normality of the estimate (both the observation time interval and the amplitude of noise are fixed). So far the problem has been studied under the assumption that all the operators in the equation have a common system of eigenfunctions, which effectively reduces the model to the equation with constant coefficients. It is proved that the estimate is consistent and asymptotically normal without any restrictions on the eigenvalues of the operators.

Chapter 1

Stochastic Partial Differential Equations and Problems in Statistics

1.1 Forward and Inverse Problems for Differential Equations

The problems in statistics of stochastic differential equations studied in this dissertation are related to nonlinear filtering and parameter estimation.

The objective of filtering is estimation of a component of a partially observed stochastic system. Under certain assumptions, this problem is reduced to the determination of a special random field called the unnormalized filtering density, which satisfies a stochastic partial differential equation driven by the observable component of the system; the equation is usually called the Zakai equation. In the canonical setting, the coefficients of the equation are known, but in most cases the solution cannot be computed explicitly. The filtering problem then becomes a forward problem for the stochastic partial differential equation meaning that an approximate solution of the equation must be computed.

The parameter estimation is an example of an inverse problem when the solution of the equation is observed and the conclusions must be made about the coefficients of the equation. In the deterministic setting, numerous examples of such problems in ecology, material sciences, biology, etc. are given in the book by Banks and Kunisch [4]. The stochastic term is usually introduced to take into account those components of the model that cannot be described exactly. A typical example is the heat balance equation describing the sea surface temperature anomalies. According to Frankignoul [19], the equation is

$$du(t,x) = (D\Delta u(t,x) - (\vec{v}(x), \nabla)u(t,x) - \lambda u(t,x))dt + dW(t,x), \tag{1.1}$$

where u represents the deviation of the temperature form the given average value, $x = (x_1, x_2) \in \mathbf{R}^2$, $\vec{v}(x) = (v_1(x_1,x_2), v_2(x_1,x_2))$ is the velocity field of the top layer of the ocean, D is thermodiffusivity, λ is the cooling coefficient, and the random perturbation $W(t,x)$ represents the short time atmospheric changes. A priory, D, λ, and $\vec{v}$ are unknown, but usually some assumptions are made about one or two of them in order to estimate the rest.

It is worth mentioning that in its original setting, the nonlinear filtering problem can be viewed as an inverse problem of nonparametric estimation for nonlinear stochastic ordinary differential equations and the Zakai equation reduces it to the forward problem for a linear partial differential equation. More detailed discussion of the filtering problem is given in the next section.

1.2 Nonlinear Filtering and Wiener Chaos Decomposition

If the unobserved state process $X = X(t)$ is a Markov process with the generator $\mathcal{L}$ and the observation process $Y = Y(t)$ is

$$Y(t) = \int_0^t h(X(s))ds + W(t),$$

where the Wiener process W is independent of X, then the optimal filter $\hat{f}_t$, i.e. the best mean square estimate of $f(X(t))$ given the trajectory of $Y(s)$, $0 \leq s \leq t$, is given by

$$\hat{f}(x(t)) = \frac{\int f(x)p(t,x)dx}{\int p(t,x)dx} \tag{1.2}$$

with the unnormalized filtering density $p = p(t,x)$ satisfying the Zakai equation

$$dp(t,x) = \mathcal{L}^* p(t,x) + h(x)p(t,x)dY(t). \tag{1.3}$$

This equation was studied by many authors: Baras [5], Benesh [6], Bennaton [7], Bensoussan et al. [8], Clark [13], Elliott and Glowinski [17], Florchinger and LeGland [18], Ito [31], Krylov and Rozovskii [38], Kunita [41], Kushner [42], Rozovskii [61], Zakai [66], etc.

The *computable* explicit solution of the Zakai equation exists only in a few special cases when the function p admits a finite dimensional sufficient statistics. One such example is when both the state and the observations are linear diffusion processes, in which case the solution is given by the Kalman - Bucy filter [34, 47]. Another example, the diffusion with a cubic drift, was discovered by Benesh [6]. The general description of the unnormalized filtering density admitting a finite dimensional sufficient statistics is given in Daum [14], but there are many problems that cannot be solved using this approach.

In the general nonlinear setting, an approximate solution of the Zakai equation must be computed, and there are two major approaches to doing this.

The first approach uses the existing numerical methods for deterministic partial and stochastic ordinary equations. These methods are applied either directly to equation (1.3) or to its *robust* form [7, 13, 31], which means that the original Zakai equation is first reduced to a random partial differential equation (a deterministic partial differential equation with random coefficients). Among the proposed algorithms are the splitting-up approximation based on the Trotter type product formula [8, 17, 18], the implicit Euler scheme [31], the Galerkin approximation [7, 31], the finite element method [21], and the hidden Markov model [16, 42]. All these algorithms have to deal with the same difficulty — the fast growth of computational complexity when the dimension of the state process is increased. As a result, the on-line solution of the Zakai equation, which is necessary in many applications, is almost impossible using this approach if the dimension of the state process is more than three.

The idea of the second approach is to separate the deterministic and stochastic components of the equation. The result is a representation of the solution either as a sum of multiple Wiener integrals (i.e. Ito integrals with deterministic integrands) or as an orthogonal series with respect to the Cameron - Martin basis in the space of square integrable functional of the Winer process. In both cases the representation is known as the Wiener chaos decomposition and an important property of this representation is that it allows a part of the computations (usually the most time consuming part involving partial differential equations) to be performed off line before the observations become available. By shifting part of the computations off line, it should be possible to improve the on-line performance of the corresponding algorithm.

As an analytical tool to study stochastic differential equations, the Wiener chaos decomposition was first used by Krylov and Veretennikov [39] to derive a representation of a functional of the solution of a nonlinear stochastic ordinary differential equation as a sum of multiple Wiener integrals with deterministic kernels. Kunita [41] used this method to prove the uniqueness of the solution of the Zakai equation, Wong [64] — to derive an explicit solution to certain class of nonlinear filtering problems (this solution, however, is not, in general, computable), and Ocone [59] — to study the original optimal filter (1.2). A numerical algorithm for computing a recursive approximation of the unnormalized filtering density and the optimal filter independently of each other was suggested by Lototsky and Rozovskii [51]. Budhiraja and Kallianpur [9] studied the expansion of the unnormalized filtering density using multiple Stratonovich integrals; they also considered a computable approximation of the expansion.

An equivalent form of the Wiener chaos decomposition uses the result of Cameron and Martin [11] about the orthonormal basis in the space of square integrable functionals of the Wiener process. This form, however, can have computational advantages because it requires only simple Wiener integrals. Lo and Ng [48] used the Cameron - Martin version of the decomposition to study the optimal filter (1.2) *given* the expansions of the numerator and the denominator. The expansion of the unnormalized filtering density using the Cameron - Martin theory was introduced by Mikulevicius and Rozovskii [53], and a numerical algorithm for computing a recursive approximation of the unnormalized filtering density using this expansion was suggested and investigated in Lototsky et al. [50]. Further investigation of the various algorithms based on this expansion, and numerical simulations were done by Fung [20].

1.3 Parameter Estimation for Stochastic Differential Equations

Parameter estimation is a particular case of the inverse problem when the coefficients of the equation depend on a certain number of unknown scalar parameters that must be estimated from the observations of the solution. An estimate is a functional of the observations, and the objective is to find a consistent estimate, i.e. the estimate that it in some sense close to the true value of the parameters.

The formal statement of the problem leads to the notion of a family of statistical experiments indexed by a parameter ϵ. Each experiment gives rise to an estimate, and the properties of the estimate are studied in the limit $\epsilon \to 0$. The notions of consistency, asymptotic normality, and asymptotic efficiency are introduced to characterize and compare the limiting behavior of different estimates. The general asymptotic theory of statistical estimation was developed by Ibragimov and Khasminskii [29].

Parameter estimation problems for stochastic differential equations are usually formulated for evolution equations of the form

$$du(t) = \mathbf{A}(t, \theta, u)dt + \mathbf{B}(t, u)dW(t), \tag{1.4}$$

where u takes values in some Hilbert space $\mathbf{H}$ (finite or infinite dimensional), $\mathbf{A}$ and $\mathbf{B}$ are operators in that space, θ is the unknown parameter belonging to some domain in $\mathbf{R}^d$, and $W(t)$ is a random perturbation.

The following questions usually must be addressed before any estimate of θ is considered:

1. What are the properties of the solution of (1.4)? This question is especially important if the space $\mathbf{H}$ is infinite dimensional.

2. What is the observation process? It can be u itself or some transformation of u, possibly with additional noise.

3. What is the asymptotic parameter of the estimate? It can be decreasing amplitude of noise, increasing time of observations, or, in the case of infinite dimensional observations, the dimension of the projection used to construct the estimate.

In the case of ordinary differential equations (i.e. when the space $\mathbf{H}$ is finite dimensional) the parameter estimation problems were studied by Kutoyants [43, 44] in both direct and partial observation settings and in both small noise and large observation time asymptotics. A somewhat intermediate case is when the original process is infinite dimensional, but the observation process is finite dimensional of fixed dimension. Aihara [2] and Loges [49] studied the asymptotic properties of the estimates for this model as the observation time interval tends to infinity.

In the case of infinite dimensional observations it is possible to get consistent estimates on a fixed time interval with a fixed amplitude of noise. To achieve this, the estimate is constructed using a finite dimensional projection of the observation process, and the asymptotic properties of the estimate are studied as the dimension of the projection tends to infinity. This approach was first suggested by Huebner et al. [26] and further developed by Huebner [27], Huebner and Rozovskii [28], Piterbarg and Rozovskii [60].

1.4 Summary of Main Results

The major part of the dissertation deals with orthogonal expansions for stochastic evolution equations with applications to nonlinear filtering of diffusion processes. In the last chapter, a parameter estimation problem for a partial differential equation is considered.

The general theory of the Wiener chaos decomposition of the solution of a stochastic evolution equation is presented in **Chapter 2**. Consider a random process $u = u(t)$ with values in a Hilbert space $\mathbf{H}$, satisfying the evolution equation

$$u(t) = u_0 + \int_{T_0}^{t} \mathcal{A}u(s)ds + \sum_{l=1}^{r} \int_{T_0}^{t} \mathcal{B}_l u(s)dW_l(s), \tag{1.5}$$

where $\mathcal{A}$ and $\mathcal{B}_l$, $l = 1, \ldots, r$, are linear (possibly unbounded) operators on $\mathbf{H}$ and $W = (W_1(t), \ldots, W_r(t))$ is a standard Wiener process. Then for every fixed $t^* \in [T_0, T]$ and all $t \in [T_0, t^*]$ the solution u can be written as

$$u(t) = \sum_{\alpha \in J} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(t)\xi_\alpha, \tag{1.6}$$

where the coefficients φ_α satisfy a recursive system of deterministic evolution equations and the random variables ξ_α are Wick products (certain products of Hermite polynomials) of the Wiener integrals $\int_{T_0}^{t^*} m_k(t)dW_l(t)$, $\{m_k\}_{k\geq 1}$ is an orthonormal basis in $L_2([T_0, t^*])$; summation in (1.6) is over all r-dimensional multi-indices. The rate of convergence of the series in (1.6) is established and a recursive (step-by-step) version of the expansion is studied. The relation to the multiple Wiener integral expansion is also discussed.

In [50, 51, 53], the Wiener chaos decomposition was studied for the Zakai equation with independent noise, which essentially corresponds to equation (1.5) with commuting operators $\mathcal{B}_l$. Chapter 2 provides generalization and further development of those results, in particular, the analysis of the approximation error and of the multiple Wiener integral expansion.

In **Chapter 3**, the solution of equation (1.5) is expanded with respect to an orthonormal basis in the space $\mathbf{H}$. Two approximations of the solution are constructed using this expansion and the corresponding errors are estimated. In a very special form, one of the approximations was introduced in [50]. The other is the Galerkin approximation, for which, in spite of frequent use in nonlinear filtering algorithms, no reliable error bound has been known.

In **Chapter 4**, the results of the previous two chapters are used to study the filtering problem for a diffusion filtering model. Two algorithms are suggested for computing a recursive approximation of the optimal filter. The first algorithm was essentially suggested in Lototsky et al. [50], and the idea of the second algorithm belongs to Fung [20]. The novelty of the approach is that the major part of the computations in both algorithms can be done off line making the algorithms suitable for on-line implementation. The rate of convergence of both algorithms is established and certain computational aspects are discussed.

In **Chapter 5**, a parameter estimation problem is considered for a stochastic partial differential equation on a compact smooth manifold M. The equation is

$$du(t,x) + (\mathcal{A}_0 + \theta\mathcal{A}_1)u(t,x)dt = dW(t,x),\ 0 < t \leq T,\ x \in M;\ u(0,x) = 0,$$

where $\mathcal{A}_0$ and $\mathcal{A}_1$ are known differential operators and W is a cylindrical Brownian motion. An estimate $\hat{\theta}^K$ of θ is constructed using finite dimensional projections of the solution:

$$\hat{\theta}^K = \frac{\int_0^T (\Pi^K \mathcal{A}_1 u(t), d\Pi^K u(t) - \Pi^K \mathcal{A}_0 u(t) dt)_0}{\int_0^T \|\Pi^K \mathcal{A}_1 u(t)\|_0^2 dt}.$$

If the operators $\mathcal{A}_0$ and $\mathcal{A}_1$ have a common system of eigenfunctions and the projection operator Π^K commutes with both $\mathcal{A}_0$ and $\mathcal{A}_1$, then $\hat{\theta}^K$ is the maximum likelihood estimate of θ studied in [26, 27, 28]. It is proved that the estimate remains consistent and asymptotically normal, as $K \to \infty$, without any assumptions about the eigenfunctions of the operators $\mathcal{A}_0$ and $\mathcal{A}_1$. The obtained results are applied to the estimation problem for the heat balance equation (1.1) in which the velocity field $\vec{v}$ no longer has to be constant.

Chapter 2

Wiener Chaos Decomposition for Stochastic Evolution Equations

2.1 Introduction

One of the principal tools in the study of partial differential equations is separation of variables. The idea of the method for the deterministic partial differential equations is to separate the time and the space variables, and the result is usually a Fourier type series representation of the solution.

For a stochastic equation, the same idea can also be used to separate the deterministic and stochastic components of the solution. This separation is possible if the underlying probability space has a *computable* orthonormal basis. If the only source of randomness in the equation is the Wiener process, then, under the conditions of existence and uniqueness, the solution is usually a square integrable functional of that process. Then it is possible to use the classical result of Cameron and Martin [11], which provides an explicit construction of the orthonormal basis in the space of square integrable functionals of a Wiener process. The corresponding representation of the solution is called the Wiener chaos decomposition. The objective of this chapter is the study of this decomposition for a stochastic evolution equation in a Hilbert space.

In **Section 2.2**, the general results are presented about the properties of the solution of a general parabolic equation in a Hilbert space. These results are used throughout the rest of the dissertation.

In **Section 2.3**, the construction of the orthonormal basis in the space of square integrable Wiener functionals is presented and then the solution of the equation

$$u(t) = u_0 + \int_{T_0}^{t} \mathcal{A}u(s)ds + \int_{T_0}^{t} \sum_{l=1}^{r} \mathcal{B}_l u(s)dW_l(s) \qquad (2.1)$$

is expanded with respect to the basis for $t \in [T_0, t^*]$, where $t^* \leq T$ is fixed. After that, the expansion is truncated to a finite sum and the error of the truncation is derived.

In **Section 2.4**, a recursive version of the expansion is constructed on the fixed time grid $T_0 = t_0 < \ldots < t_M = T$. As in Section 2.3, a finite truncation of the expansion is studied. This step-by-step expansion reduces the overall approximation error over the interval $[T_0, T]$ as compared to the one step expansion.

In **Section 2.5**, an alternative version of the Wiener chaos decomposition is considered using a theorem by Ito [30]. This expansion is based on the multiple Wiener integrals and, when applied recursively, can be used to construct *computable* approximations of the solution with the prescribed rate of convergence in time.

2.2 Stochastic Evolution Equations: An Overview

In this section, the main results are presented concerning the existence, uniqueness, and other properties of the solutions of stochastic evolution equations in a Hilbert space. All definitions and theorems are stated in the form most suitable for future references. The more general versions of all presented results can be found in [61, Chapter 3].

Let $W = W(t)$ be an r - dimensional Wiener process on a complete probability space $(\Omega, \mathcal{F}, I\!\!F, \mathbf{P})$ with a right continuous filtration $I\!\!F = \{\mathcal{F}_t\}_{t\geq 0}$, and let $\mathcal{A}$ and $\mathcal{B}_l$, $l = 1, \ldots, r$, be linear operators defined on a dense subset of a real separable Hilbert space $\mathbf{H}$.

Consider the stochastic evolution equation

$$u(t) = u_0 + \int_{T_0}^{t} \mathcal{A}u(s)ds + \int_{T_0}^{t} \sum_{l=1}^{r} \mathcal{B}_l u(s)dW_l(s) + M(t) - M(T_0),\ t \in [T_0, T],\ T_0 \geq 0, \quad (2.2)$$

where u_0 is an $\mathcal{F}_0$-measurable random element with values in $\mathbf{H}$ and $M = M(t)$ is a continuous square integrable martingale on $(\Omega, \mathcal{F}, I\!\!F, \mathbf{P})$ with values in $\mathbf{H}$.

Let $\{\mathbf{H}^a\}_{a\in\mathbf{R}}$ be a scale of Hilbert spaces (or a Hilbert scale) so that $\mathbf{H}^0 = \mathbf{H}$. The main facts about the Hilbert scales can be found in [37, Section IV.1.10] or [61, Sections 2.4 and 3.2]. If $a_0 \in \mathbf{R}$, $b > 0$, then the collection $\{\mathbf{H}^{a_0+b}, \mathbf{H}^{a_0}, \mathbf{H}^{a_0-b}\}$ is a normal triple. The canonical bilinear functional of the triple will be denoted by $\langle u, v\rangle_{a_0,b}$, $u \in \mathbf{H}^{a_0-b}$, $v \in \mathbf{H}^{a_0+b}$. Also, the norm and the inner product in $\mathbf{H}^a$ are denoted by $\|\cdot\|_a$ and $(\cdot,\cdot)_a$ respectively.

2.2.1. Definition. Assume that $\mathcal{A}$ is a linear bounded operator from $\mathbf{H}^1$ to $\mathbf{H}^{-1}$ and each $\mathcal{B}_l$, $l = 1, \ldots, r$, is a linear bounded operator from $\mathbf{H}^1$ to $\mathbf{H}^0$. A random element $U = U(t; T_0; u_0)$ with values in $L_2([T_0, T]; \mathbf{H}^1) \cap \mathbf{C}([T_0, T]; \mathbf{H}^0)$ is called **a solution** of (2.2) if there exists a subset Ω' of Ω with $\mathbf{P}(\Omega') = 1$ such that on the set $\Omega' \times [T_0, T]$

$$\begin{aligned}(U(t;T_0;u_0), v)_0 &= (u_0, v)_0 + \int_{T_0}^{t} \langle \mathcal{A}U(s;T_0;u_0), v\rangle_{0,1} ds + \\ &\int_{T_0}^{t} \sum_{l=1}^{r} (\mathcal{B}_l U(s;T_0;u_0), v)_0 dW_l(s) + (M(t) - M(T_0), v)_0\end{aligned} \quad (2.3)$$

for every $v \in \mathbf{H}^1$.

To establish existence of such a solution, some additional assumptions must be made.

2.2.2. Definition. The stochastic evolution equation (2.2) is called **coercive** if the following conditions are fulfilled:

(C1) The operator $\mathcal{A}$ is linear and bounded from $\mathbf{H}^1$ to $\mathbf{H}^{-1}$, and the operators $\mathcal{B}_l$, $l = 1, \ldots, r$, are linear and bounded from $\mathbf{H}^1$ to $\mathbf{H}^0$;

(C2) $M = M(t)$ is a continuous square integrable martingale with values in $\mathbf{H}^0$; u_0 is $\mathcal{F}_{T_0}$ - measurable and $u_0 \in L_2(\Omega; \mathbf{H}^0)$;

(C3) There exist positive numbers N and δ so that

$$2\langle \mathcal{A}v, v\rangle_{0,1} + \sum_{l=1}^{r} \|\mathcal{B}_l v\|_0^2 + \delta\|v\|_1^2 \leq N\|v\|_0^2 \text{ for all } v \in \mathbf{H}^1.$$

The operator $\mathcal{A}$ is called **coercive** if it satisfies (C3) with $\mathcal{B}_l = 0$.

2.2.3. Remark. Condition (C3) is trivially fulfilled if the operators $\mathcal{A}$ and $\mathcal{B}_l$ are linear and bounded from $\mathbf{H}$ to $\mathbf{H}$ (in this case, the Hilbert scale is degenerate: $\mathbf{H}^a = \mathbf{H}$ for all $a \in \mathbf{R}$).

For the rest of this section, C denotes a real constant independent of U, t, and M.

2.2.4. Theorem. *If equation (2.2) is coercive, then there exists a unique solution $U = U(t; T_0; u_0)$ with the following property: $U \in L_2(\Omega \times [T_0, T]; \mathbf{H}^1) \cap L_2(\Omega; \mathbf{C}([T_0, T]; \mathbf{H}^0))$ and*

$$\mathbf{E}\Big(\sup_{T_0 \leq t \leq T} \|U(t; T_0; u_0)\|_0^2 + \int_{T_0}^{T} \|U(t; T_0; u_0)\|_1^2 dt\Big) \leq e^{C(T-T_0)}\mathbf{E}\Big(\|u_0\|_0^2 + \|M(T)\|_0^2\Big); \quad (2.4)$$

Proof. This follows from Theorem 3.1.4 and Remark 3.1.3 in [61].

□

2.2.5. Definition. The stochastic evolution equation (2.2) is called **dissipative** if the following conditions are fulfilled:

(D1) The operator $\mathcal{A}$ is linear and bounded from $\mathbf{H}^2$ to $\mathbf{H}^0$ and from $\mathbf{H}^1$ to $\mathbf{H}^{-1}$, and the operators $\mathcal{B}_l$, $l = 1, \ldots, r$, are linear and bounded from $\mathbf{H}^2$ to $\mathbf{H}^1$ and from $\mathbf{H}^1$ to $\mathbf{H}^0$;

(D2) $M = M(t)$ is a continuous square integrable martingale with values in $\mathbf{H}^2$; u_0 is $\mathcal{F}_{T_0}$ - measurable and $u_0 \in L_2(\Omega; \mathbf{H}^1)$;

(D3) There exists a positive number N so that

$$2\langle \mathcal{A}v, v\rangle_{a,1} + \sum_{l=1}^{r} \|\mathcal{B}_l v\|_a^2 \leq N\|v\|_a^2 \text{ for all } v \in \mathbf{H}^{a+1},\ a = 0, 1.$$

The operator $\mathcal{A}$ is called **dissipative** if it satisfies (D3) with $\mathcal{B}_l = 0$ and $a = 0$.

2.2.6. Theorem. *If equation (2.2) is dissipative, then there exists a unique solution $U = U(t; T_0; u_0)$ with the following property: $U \in L_2(\Omega \times [T_0, T]; \mathbf{H}^1) \cap L_2(\Omega; \mathbf{C}([T_0, T]; \mathbf{H}^0))$ and*

$$\mathbf{E} \sup_{T_0 \leq t \leq T} \|U(t; T_0; u_0)\|_0^2 \leq e^{C(T-T_0)}\mathbf{E}\Big(\|u_0\|_0^2 + \|M(T)\|_0^2\Big), \quad (2.5)$$

$$\mathbf{E}\int_{T_0}^{T}\|U(t;T_0;u_0)\|_1^2dt \le e^{C(T-T_0)}\mathbf{E}\Big(\|u_0\|_1^2+\|M(T)\|_2^2\Big). \tag{2.6}$$

Proof. This follows from Theorem 3.2.2, Remark 3.1.3, and Proposition 3.1.7 in [61].

□

2.2.7. Theorem. *Assume that equation (2.2) is either coercive or dissipative. Then its solution $U(t;T_0;u_0)$ has the following properties:*

(P1) If $s\in[T_0,t],\ t\le T$, then $U(t;T_0;u_0)=U(t;s;U(s;T_0;u_0))$;

(P2) If g is deterministic and $V(t;T_0;g)$ is the solution of (2.2) with the initial condition $u_0=g$, then $U(t;T_0;u_0)=V(t;T_0;u_0)$, and if $M(t)\equiv 0$, then $\mathbf{E}\|U(t;T_0;u_0)\|_0^2=\mathbf{E}F(u_0)$, where $F(g)=\mathbf{E}\|V(t;T_0;g)\|_0^2$;

(P3) If u_0 is deterministic and $M(t)\equiv 0$, then for every $s\in[T_0,t],\ t\le T$, $U(t;T_0;u_0)$ is an $\mathcal{F}^W_{T_0,t}$-measurable random element with values in $L_2(\Omega;\mathbf{H}^0)$, where $\mathcal{F}^W_{T_0,t}$ is the σ-algebra generated by $W(s)-W(T_0),\ T_0\le s\le t$.

Proof. Property (P1) and the first part of property (P2) follow directly from the uniqueness and the definition of the solution, property (P3) is implied by the construction of the solution, and the second part of property (P2) is the consequence of the Markov property of the solution U. The details of the proof can be found in [61, Chapter 3].

□

The homogeneous deterministic evolution equation is a special case of (2.2) with a deterministic initial condition, $\mathcal{B}_l=0,\ l=1,\dots,r$, and $M(t)\equiv 0$, and therefore all the above results are valid for such equations. In particular, the following can be said about the semigroup generated by the operator $\mathcal{A}$.

2.2.8. Theorem. *If equation (2.2) is either coercive or dissipative, then the operator $\mathcal{A}$ generates a time homogeneous semigroup $\{\Phi_t\}_{t\ge 0}$ and there is a real constant C so that*

$$\|\Phi_t v\|_0 \le e^{Ct}\|v\|_0 \tag{2.7}$$

for every $t\ge 0,\ v\in\mathbf{H}^0$.

Proof. Assume that $M(t)\equiv 0$ and all $\mathcal{B}_l,\ l=1,\dots,r$, are zero operators. Then the existence of the semigroup follows from the existence statements of Theorems 2.2.4 and 2.2.6. The semigroup is time homogeneous because the operator $\mathcal{A}$ does not depend on time. Finally, (2.7) follows from either (2.4) or (2.5) (in the case of a dissipative equation, (2.5) is combined with the fact that $\mathbf{H}^1$ is dense in $\mathbf{H}^0$, so that Φ_t can be extended to $\mathbf{H}^0$ by continuity).

□

2.3 One Step Expansion

Let $W = W(t)$ be an r - dimensional Wiener process on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and let $\mathcal{A}$ and $\mathcal{B}_l$, $l = 1, \dots, r$, be linear operators in the Hilbert scale $\{\mathbf{H}^a\}_{a \in \mathbf{R}}$. It is assumed that the operator $\mathcal{A}$ is linear and bounded from $\mathbf{H}^1$ to $\mathbf{H}^{-1}$ and each $\mathcal{B}_l$ is a linear bounded operator from $\mathbf{H}^1$ to $\mathbf{H}^0$. The σ- algebra generated by $W(s) - W(t_1)$, $t_1 \le s \le t_2$, will be denoted by $\mathcal{F}^W_{t_1,t_2}$.

Consider the following stochastic evolution equation:

$$u(t) = u_0 + \int_{T_0}^t \mathcal{A}u(s)ds + \int_{T_0}^t \sum_{l=1}^r \mathcal{B}_l u(s) dW_l(s), \tag{2.8}$$

or in the differential form

$$\begin{aligned} du(t) &= \mathcal{A}u(t)dt + \textstyle\sum_{l=1}^r \mathcal{B}_l u(t) dW_l(t),\ T_0 < t \le T; \\ u(T_0) &= u_0, \end{aligned} \tag{2.9}$$

where u_0 is independent of $\mathcal{F}^W_{T_0,T}$. Assume that equation (2.8) is either coercive or dissipative. Then there is a unique solution of (2.8); this solution is denoted by $u(t; T_0; u_0)$.

To begin with, assume that $u_0 = g$ is deterministic. In what follows, the Wiener chaos decomposition of $u(t; T_0; u_0)$ will be derived and the properties of the decomposition studied.

As a first step, recall the construction of an orthonormal basis in $L_2(\Omega, \mathcal{F}^W_{T_0,t}, \mathbf{P})$. Let α be an r-**dimensional multi-index**, i.e. a collection $\alpha = (\alpha_k^l)_{1 \le l \le r,\ k \ge 1}$ of nonnegative integers such that only finitely many of α_k^l are different from zero. The set of all such multi-indices will be denoted by J. For $\alpha \in J$ define $\alpha! := \prod_{k,l}(\alpha_k^l!)$.

For a fixed $t^* \in [T_0, T]$ choose a complete orthonormal system $\{m_k\} = \{m_k(s)\}_{k \ge 1}$ in $L_2([T_0, t^*])$ and define

$$\xi_{k,l} = \int_{T_0}^{t^*} m_k(s) dW_l(s)$$

so that $\xi_{k,l}$ are independent Gaussian random variables with zero mean and unit variance.

If

$$H(x) := (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} \tag{2.10}$$

is the n-th Hermite polynomial, then the collection

$$\left\{ \xi_\alpha(W_{T_0,t^*}) := \prod_{k,l} \left(\frac{H_{\alpha_k^l}(\xi_{k,l})}{\sqrt{\alpha_k^l!}} \right), \quad \alpha \in J \right\}$$

is an orthonormal system in $L_2(\Omega, \mathcal{F}^W_{T_0,t^*}, \mathbf{P})$. A theorem by Cameron and Martin [11] shows that $\{\xi_\alpha(W_{T_0,t^*})\}_{\alpha \in J}$ is actually a basis in that space.

2.3.1. Theorem. *If* $\eta \in L_2(\Omega, \mathcal{F}^W_{T_0,t^*}, \mathbf{P})$*, then*

$$\eta = \sum_{\alpha \in J} \mathbf{E}[\eta \xi_\alpha(W_{T_0,t^*})] \xi_\alpha(W_{T_0,t^*}) \tag{2.11}$$

and

$$\mathbf{E}|\eta|^2 = \sum_{\alpha \in J} |\mathbf{E}\eta\xi_\alpha(W_{T_0,t^*})|^2.$$

Proof. This theorem is proved in [11] and [23].

□

2.3.2. Remark. The representation (2.11) is known as (the Cameron - Martin version of) the Wiener chaos decomposition of the random variable η.

By property (P3) from Theorem 2.2.7 and the Pettis theorem [15, Theorem III.6.11], if $s \in [T_0, t^*]$, $t^* \le T$, and $v \in \mathbf{H}^1$, then the random variable $(u(s;T_0;g), v)_0$ belongs to $L_2(\Omega, \mathcal{F}^W_{T_0,t^*}, \mathbf{P})$. Therefore it follows from Theorem 2.3.1 that

$$(u(s;T_0;g), v)_0 = \sum_{\alpha \in J} \mathbf{E}[(u(s;T_0;g), v)_0 \xi_\alpha(W_{T_0,t^*})] \xi_\alpha(W_{T_0,t^*}) \tag{2.12}$$

and

$$\mathbf{E}|(u(s;T_0;g), v)_0|^2 = \sum_{\alpha \in J} |\mathbf{E}[(u(s;T_0;g), v)_0 \xi_\alpha(W_{T_0,t^*})]|^2. \tag{2.13}$$

The properties of the solution of (2.8) imply that for every $\alpha \in J$ the expectation $\mathbf{E}[u(t;T_0;g)\xi_\alpha(W_{T_0,t^*})]$, as a function of t, is a well defined element of $L_2([T_0,t^*];\mathbf{H}^1) \cap \mathbf{C}([T_0,t^*];\mathbf{H}^0)$; this element will be denoted by $\frac{1}{\sqrt{\alpha!}}\varphi_\alpha(t;T_0;g)$ (the normalizing factor $\frac{1}{\sqrt{\alpha!}}$ is introduced for technical reasons). It is shown in the following theorem that the functions $\varphi_\alpha(t;T_0;g)$, $\alpha \in J$, satisfy a recursive system of deterministic evolution equations.

2.3.3. Theorem. *Suppose that* $u_0 = g$ *is deterministic and equation (2.8) is either coercive or dissipative. If* $t^* \in (T_0, T]$ *is fixed, then for every* $s \in [T_0, t^*]$ *the solution* $u(s;T_0;g)$*, viewed as an element of* $\mathbf{H}^0$*, can be written as*

$$u(s;T_0;g) = \sum_{\alpha \in J} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(s;T_0;g) \xi_\alpha(W_{T_0,t^*}), \tag{2.14}$$

so that the series converges in $L_2(\Omega;\mathbf{H}^0)$ *and the following Parseval's equality holds:*

$$\mathbf{E}\|u(s;T_0;g)\|_0^2 = \sum_{\alpha \in J} \frac{1}{\alpha!} \|\varphi_\alpha(s;T_0;g)\|_0^2. \tag{2.15}$$

The coefficients of the expansion satisfy the recursive system of deterministic equations

$$\begin{aligned} \frac{\partial \varphi_\alpha(s;T_0;g)}{\partial s} &= \mathcal{A}\varphi_\alpha(s;T_0;g) + \sum_{k,l} \alpha_k^l m_k(s)\mathcal{B}_l \varphi_{\alpha(k,l)}(s;T_0;g), \quad T_0 < s \le t^*; \\ \varphi_\alpha(T_0;T_0;g) &= g 1_{\{|\alpha|=0\}}, \end{aligned} \tag{2.16}$$

where $\alpha = (\alpha_k^l)_{1\le l\le r,\ k\ge 1} \in J$ *and* $\alpha(i,j)$ *stands for the multi-index* $\tilde{\alpha} = (\tilde{\alpha}_k^l)_{1\le l\le r,\ k\ge 1}$ *with*

$$\tilde{\alpha}_k^l = \begin{cases} \alpha_k^l & \text{if } k \ne i \text{ or } l \ne j \text{ or both} \\ \max(0, \alpha_i^j - 1) & \text{if } k = i \text{ and } l = j. \end{cases} \tag{2.17}$$

Proof. Let $\{e_k\}_{k\ge 1}$ be an orthonormal basis in $\mathbf{H}^0$. Then (2.13) and the Fubini theorem imply

$$\mathbf{E}\|u(s;T_0;g)\|_0^2 = \sum_{\alpha\in J} \frac{1}{\alpha!} \sum_{k\ge 1} |(\varphi_\alpha(s;T_0;g), e_k)_0|^2 = \sum_{\alpha\in J} \frac{1}{\alpha!} \|\varphi_\alpha(s;T_0;g)\|_0^2,$$

which proves (2.15). After that,

$$\sum_{\alpha\in J} \frac{1}{\sqrt{\alpha!}} (\varphi_\alpha(s;T_0;g), v)_0 = \Big(\sum_{\alpha\in J} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(s;T_0;g), v \Big)_0$$

for all $v \in \mathbf{H}^0$ and (2.14) follows from (2.12).

To prove that the coefficients satisfy (2.16), define

$$P_t(z) = \exp\Big\{ \int_{T_0}^t \sum_{l=1}^r m_z^l(s) dW_l(s) - \frac{1}{2}\int_{T_0}^t \sum_{l=1}^r |m_z^l(s)|^2 ds \Big\},\ T_0 \le t \le t^*,$$

where $m_z^l = \sum_{k\ge 1} m_k(s) z_k^l$ and $\{z_k^l\},\ l = 1,\ldots,r,\ k = 1,2,\ldots,$ is a sequence of real numbers such that $\sum_{k,l} |z_k^l|^2 < \infty$. Then direct computations show that

$$\xi_\alpha(W_{T_0,t^*}) = \frac{1}{\sqrt{\alpha!}} \frac{\partial^\alpha}{\partial z^\alpha} P_{t^*}(z)\Big|_{z=0},$$

where

$$\frac{\partial^\alpha}{\partial z^\alpha} = \prod_{k,l} \frac{\partial^{\alpha_k^l}}{(\partial z_k^l)^{\alpha_k^l}},$$

and also, that

$$\mathbf{E}[\eta \xi_\alpha(W_{T_0,t^*})] = \frac{\partial^\alpha}{\partial z^\alpha} \mathbf{E}[\eta P_{t^*}(z)]\Big|_{z=0}$$

for every $\eta \in L_2(\Omega, \mathcal{F}^W_{T_0,t^*}, \mathbf{P})$. Consequently,

$$(\varphi_\alpha(s;T_0;g), v)_0 = \frac{\partial^\alpha}{\partial z^\alpha} \mathbf{E}[(u(s;T_0;g), v) P_{t^*}(z)]\Big|_{z=0} = \frac{\partial^\alpha}{\partial z^\alpha} \mathbf{E}[(u(s;T_0;g), v)_0 P_s(z)]\Big|_{z=0},$$

where the second equality follows from the martingale property of $P_s(z)$ on $(\Omega, \{\mathcal{F}^W_{T_0,t}\}_{T_0\le t\le t^*}, \mathbf{P})$. It follows from the definition of $P_s(z)$ that

$$dP_s(z) = \sum_{l=1}^{r} m_z^l(s)P_s(z)dW_l(s),\ T_0 \le\ s \le t;\ P_{T_0}(z) = 1.$$

Then (2.2) and the Ito formula imply that

$$\begin{aligned}(u(s;T_0;g),v)_0 P_s(z) = g+\\ \int_{T_0}^{s}\Big(\langle \mathcal{A}u(\tau;T_0;g),v\rangle_{0,1} + \sum_{l=1}^{r}(\mathcal{B}_l u(\tau;T_0;g),v)_0\Big)m_z^l(\tau)P_\tau(z)d\tau+\\ \int_{T_0}^{s}\sum_{l=1}^{r}\Big((\mathcal{B}_l u(\tau;T_0;g),v)_0 + (u(\tau;T_0;g),v)_0 m_z^l(s)\Big)P_s(z)dW_l(\tau).\end{aligned}$$

Taking the expectation on both sides of the last equality and setting $\varphi(s,z;T_0;g) := \mathbf{E}u(s;T_0;g)P_s(z)$ results in

$$\begin{aligned}(\varphi(s,z;T_0;g),v)_0 =\ & (g,v)_0 + \int_{T_0}^{s}\Big(\langle\mathcal{A}\varphi(\tau,z;T_0;g),v\rangle_{0,1}+\\ & \sum_{l=1}^{r} m_z^l(\tau)(\mathcal{B}_l\varphi(\tau,z;T_0;g),v)_0\Big)d\tau.\end{aligned}$$

Applying the operator $\dfrac{1}{\sqrt{\alpha!}}\dfrac{\partial^\alpha}{\partial z^\alpha}$ and setting $z=0$ yields that the functions $\varphi_\alpha(s;T_0;g)$ satisfy (2.16).

□

2.3.4. Corollary. *If equation (2.8) is either coercive or dissipative and u_0 is independent of $\mathcal{F}^W_{T_0,T}$, then $u(s;T_0,u_0)$ is independent of $\mathcal{F}^W_{t^*,T}$ and*

$$u(s;T_0;u_0) = \sum_{\alpha\in J}\frac{1}{\sqrt{\alpha!}}\varphi_\alpha(s;T_0;u_0)\xi_\alpha(W_{T_0,t^*}), \tag{2.18}$$

$$\mathbf{E}\|u(s;T_0;u_0)\|_0^2 = \sum_{\alpha\in J}\frac{1}{\alpha!}\mathbf{E}\|\varphi_\alpha(s;T_0;u_0)\|_0^2, \tag{2.19}$$

Proof. Equality (2.18) follows directly from (2.14) and the first part of property (P2) form Theorem 2.2.7. To prove (2.19), define $F_s(g) = \mathbf{E}\|u(s;T_0;g)\|_0^2$ so that

$$F_s(g) = \sum_{\alpha\in J}\frac{1}{\alpha!}\|\varphi_\alpha(s;T_0;g)\|_0^2.$$

By the second part of property (P2) from Theorem 2.2.7,

$$\mathbf{E}\|u(s;T_0;u_0)\|_0^2 = \mathbf{E}F_s(u_0),$$

which proves the result.

$\square$

2.3.5. Corollary. *Assume that* $\mathbf{X}$ *is a real separable Hilbert space with the norm* $\|\cdot\|_{\mathbf{X}}$. *If* g *is deterministic and* $u(s;T_0;g)\in L_2(\Omega,\mathcal{F}^W_{T_0,t^*},\mathbf{P};\mathbf{X})$, *then each* $\varphi_\alpha(s;T_0;g)$ *belongs to* $\mathbf{X}$ *and*

$$\mathbf{E}\|u(s;T_0;g)\|^2_{\mathbf{X}}=\sum_{\alpha\in J}\frac{1}{\alpha!}\|\varphi_\alpha(s;T_0;g)\|^2_{\mathbf{X}}. \tag{2.20}$$

Proof. Since by definition

$$\varphi_\alpha(s;T_0;g)=\sqrt{\alpha!}\mathbf{E}u(s;T_0;g)\xi_\alpha(W_{T_0,t^*}),$$

the properties of the expectation imply that $\varphi_\alpha(s;T_0;g)\in\mathbf{X}$. If $\{e_k\}_{k\geq 1}$ is an orthonormal basis in $\mathbf{X}$, then

$$(u,e_k)_{\mathbf{X}}\in L_2(\Omega,\mathcal{F}^W_{T_0,t^*},\mathbf{P})$$

for every $k\geq 1$, and by Theorem 2.3.1,

$$\mathbf{E}\|u(s;T_0;g)\|^2_{\mathbf{X}}=\sum_{\alpha\in J}\frac{1}{\alpha!}\sum_{k\geq 1}|\mathbf{E}(u(s;T_0;g)\xi_\alpha(W_{T_0,t^*}),e_k)_{\mathbf{X}}|^2=$$
$$\sum_{\alpha\in J}\frac{1}{\alpha!}\sum_{k\geq 1}|(\varphi_\alpha(s;T_0;g),e_k)_{\mathbf{X}}|^2=\sum_{\alpha\in J}\frac{1}{\alpha!}\|\varphi_\alpha(s;T_0;g)\|^2_{\mathbf{X}}\ ,$$

whence (2.20).

$\square$

2.3.6. Remark. Clearly, existence and uniqueness of the solution of (2.16) is implied by the existence and uniqueness of the solution of (2.8). On the other hand, existence and uniqueness for (2.16) can often be established under less restrictive assumptions than the conditions of Theorems 2.2.4 or 2.2.6. Then representation (2.14) can be taken as a *definition* of the solution of (2.8), leading to the notion of the **soft solutions** of stochastic evolution equations [54, 56].

The next natural question is the rate of convergence of (2.14). To answer this question, it is necessary to have an explicit formula for the solution of (2.16). To derive this formula, some additional notations and definitions are introduced.

For a multi-index $\alpha\in J$ define

$|\alpha|:=\sum_{l,k}\alpha_l^k$ (length of α);

$d(\alpha):=\max\{k\geq 1:\ \alpha_k^l>0$ for some $1\ \leq l\leq r\}$ (order of α).

Every multi-index α with $|\alpha|=k$ can be identified with the set $K_\alpha=\{(i_1^\alpha,q_1^\alpha),\ldots,(i_k^\alpha,q_k^\alpha)\}$ so that $i_1^\alpha\leq i_2^\alpha\leq\ldots\leq i_k^\alpha$ and if $i_j^\alpha=i_{j+1}^\alpha$, then $q_j^\alpha\leq q_{j+1}^\alpha$. The first pair (i_1^α,q_1^α) in K_α is the position numbers of the first nonzero element of α. The second pair is the same as the first if the first nonzero element of α is greater than one; otherwise, the second pair is the position numbers of the second nonzero element of α and so on. As a result, if $\alpha_j^q > 0$, then exactly α_j^q pairs in K_α are (j, q). The set K_α will be referred to as **the characteristic set** of the multi-index α. For example, if $r = 2$ and

$$\alpha = \begin{pmatrix} 0 & 1 & 0 & 2 & 3 & 0 & 0 & \dots \\ 1 & 2 & 0 & 0 & 0 & 1 & 0 & \dots \end{pmatrix},$$

then the nonzero elements are $\alpha_1^2 = \alpha_2^1 = \alpha_1^6 = 1,\ \alpha_2^2 = \alpha_4^1 = 2,\ \alpha_5^1 = 3$, and the characteristic set is
$K_\alpha = \{(1,2), (2,1), (2,2), (2,2), (4,1), (4,1), (5,1), (5,1), (5,1), (6,2)\}$. In the future, when there is no danger of confusion, the superscript α in i and q will be omitted (i.e. (i_j, q_j) will be used instead of (i_j^α, q_j^α)).

Let $\mathcal{P}^k$ be the permutation group of the set $\{1, \dots, k\}$. For a given $\alpha \in J$ with $|\alpha| = k$ and the characteristic set $\{(i_1, q_1), \dots, (i_k, q_k)\}$ define

$$E_\alpha(s^k; l^k) := \sum_{\sigma \in \mathcal{P}^k} m_{i_1}(s_{\sigma(1)}) 1_{\{l_{\sigma(1)} = q_1\}} \cdots m_{i_k}(s_{\sigma(k)}) 1_{\{l_{\sigma(k)} = q_k\}}.$$

The following notations are introduced to simplify the further presentation:

s^k, the ordered set $(s_1, \dots, s_k)$; $ds^k := ds_1 \dots ds_k$;

l^k, the ordered set $(l_1, \dots, l_k)$;

$F(t; s^k; l^k; g) := \Phi_{t-s_k} \mathcal{B}_{l_k} \Phi_{s_k - s_{k-1}} \dots \mathcal{B}_{l_1} \Phi_{s_1 - T_0} g,\ k \geq 1,\ g \in \mathbf{H}^0$;

$\int_{T_0}^{(k,t)} (\cdots) ds^k := \int_{T_0}^{t} \int_{T_0}^{s_k} \dots \int_{T_0}^{s_2} (\cdots) ds_1 \dots ds_k$;

$\sum_{l^k} := \sum_{l_1, \dots, l_k = 1}^{r}$.

2.3.7. Theorem. *Assume that equation (2.8) is either coercive or dissipative with a deterministic initial condition $u_0 = g$, and the operators $\mathcal{B}_l$, $l = 1, \dots, r$ are linear and bounded from $\mathbf{H}^0$ to $\mathbf{H}^0$. If $t^* \in (T_0, T]$ is fixed and $\alpha \in J$ is a multi-index with $|\alpha| = k$ and the characteristic set $\{(i_1, q_1), \dots, (i_k, q_k)\}$, then for $t \in [T_0, t^*]$ the corresponding solution $\varphi_\alpha(t; T_0; g)$ of (2.16) is given by*

$$\begin{aligned} &\varphi_\alpha(t; T_0; g) = \\ &\sum_{\sigma \in \mathcal{P}^k} \sum_{l^k} \int_{T_0}^{(k,t)} F^k(t; s^k; l^k; g) m_{i_{\sigma(k)}}(s_k) 1_{\{l_k = q_{\sigma(k)}\}} \cdots m_{i_{\sigma(1)}}(s_1) 1_{\{l_1 = q_{\sigma(1)}\}} ds^k,\ k > 1; \\ &\varphi_\alpha(t; T_0; g) = \int_{T_0}^{t} \Phi_{t-s_1} \mathcal{B}_{q_1} \Phi_{s_1 - T_0} g m_{i_1}(s_1) ds_1,\ k = 1; \\ &\varphi_\alpha(t; T_0; g) = \Phi_{t-T_0} g,\ k = 0, \end{aligned} \tag{2.21}$$

and

$$\sum_{|\alpha| = k} \frac{\|\varphi_\alpha(t; T_0; g)\|_0^2}{\alpha!} = \sum_{l^k} \int_{T_0}^{(k,t)} \|F(t; s^k; l^k; g)\|_0^2 ds^k. \tag{2.22}$$

Proof. For the sake of simplicity, the arguments T_0 and g will be omitted wherever possible. Representation (2.21) is obviously true for $|\alpha| = 0$. Then the general case $|\alpha| \geq 1$ follows by induction from the variation of parameters formula.

To prove (2.22), first of all note that

$$\sum_{\sigma \in \mathcal{P}^k} m_{i_{\sigma(k)}}(s_k) 1_{\{l_k = q_{\sigma(k)}\}} \cdots m_{i_{\sigma(1)}}(s_1) 1_{\{l_1 = q_{\sigma(1)}\}} = \sum_{\sigma \in \mathcal{P}^k} m_{i_k}(s_{\sigma(k)}) 1_{\{l_{\sigma(k)} = q_k\}} \cdots m_{i_1}(s_{\sigma(1)}) 1_{\{l_{\sigma(1)} = q_1\}}.$$

Indeed, every term on the left corresponding to a given $\sigma_0 \in \mathcal{P}^k$ coincides with the term on the right corresponding to $\sigma_0^{-1} \in \mathcal{P}^k$.

Then (2.21) can be written as

$$\varphi_\alpha(t) = \sum_{l^k} \int_{T_0}^{(k,t)} F(t; s^k; l^k) E_\alpha(s^k; l^k) ds^k.$$

Using the notation

$$G(t; s^k; l^k) := \sum_{\sigma \in \mathcal{P}^k} \Phi_{t - s_{\sigma(k)}} \mathcal{B}_{l_{\sigma(k)}} \ldots \Phi_{s_{\sigma(2)} - s_{\sigma(1)}} \mathcal{B}_{l_{\sigma(1)}} \Phi_{s_{\sigma(1)} - T_0} g 1_{s_{\sigma(1)} < \ldots < s_{\sigma(k)} < t},$$

it can be rewritten as

$$\varphi_\alpha(t) = \frac{1}{k!} \sum_{l^k} \int_{[T_0, t^*]^k} G(t; s^k; l^k) E_\alpha(s^k; l^k) ds^k. \tag{2.23}$$

Since for every $t \in [T_0, t^*]$, $G(t; s^k; l^k)$ is a symmetric function from $L_2(([T_0, t^*] \times \{1, \ldots, r\})^k; \mathbf{H}^0)$ with $\mathbf{H}^0$ separable, and the collection $\{E_\alpha / \sqrt{\alpha! k!}, |\alpha| = k\}$ is a CONS for the symmetric part of the space, it is possible to write

$$G(t; s^k; l^k) = \sum_{|\beta| = k} \frac{c_\beta(t) E_\beta(s^k; l^k)}{\sqrt{\beta! k!}}$$

with some $c_\beta(t) \in \mathbf{H}^0$. This and (2.23) imply $\|\varphi_\alpha(t)\|_0^2 / \alpha! = \|c_\alpha\|_0^2 / k!$ and so

$$\sum_{|\alpha| = k} \frac{\|\varphi_\alpha(t)\|_0^2}{\alpha!} = \frac{1}{k!} \sum_{|\alpha| = k} \|c_\alpha(t)\|_0^2 = \frac{1}{k!} \int_{[T_0, t^*]^k} \|G(t; s^k; l^k)\|_0^2 ds^k =$$
$$\frac{1}{k!} \sum_{l^k} \int_{[T_0, t^*]^k} \Big\| \sum_{\sigma \in \mathcal{P}^k} \Phi_{t - s_{\sigma(k)}} \mathcal{B}_{l_{\sigma(k)}} \ldots \Phi_{s_{\sigma(2)} - s_{\sigma(1)}} \mathcal{B}_{l_{\sigma(1)}} \Phi_{s_{\sigma(1)} - T_0} g 1_{s_{\sigma(1)} < \ldots < s_{\sigma(k)} < t} \Big\|_0^2 ds^k =$$
$$\sum_{l^k} \int_{T_0}^{(k,t)} \|F(t; s^k; l^k)\|_0^2 ds^k,$$

which proves (2.22).

□

2.3.8. Corollary. *Assume that* $\mathbf{X}$ *is a real separable Hilbert space with norm* $\|\cdot\|_{\mathbf{X}}$, $g \in \mathbf{X}$, *the operators* Φ_t *and* $\mathcal{B}_l$ *are linear and bounded from* $\mathbf{X}$ *to* $\mathbf{X}$ *so that* $\|\Phi_t v\|^2_{\mathbf{X}} \le e^{C_1 t}\|v\|^2_{\mathbf{X}}$, $\|\mathcal{B}v\|^2_{\mathbf{X}} \le C_2\|v\|^2_{\mathbf{X}}$, *and* $\bar{C} := C_1 + C_2 r$. *Then*

$$\sum_{|\alpha|=k} \frac{\|\varphi_\alpha(t;T_0;g)\|^2_{\mathbf{X}}}{\alpha!} \le e^{C_1(t-T_0)}\,\frac{[C_2 r(t-T_0)]^k}{k!}\|g\|^2_{\mathbf{X}}, \tag{2.24}$$

$$\sum_{\alpha\in J} \frac{\|\varphi_\alpha(t;T_0;g)\|^2_{\mathbf{X}}}{\alpha!} \le e^{\bar{C}(t-T_0)}\;\|g\|^2_{\mathbf{X}}, \tag{2.25}$$

and, if $u_0 \in L_2(\Omega;\mathbf{X})$ *and is independent of* $\mathcal{F}^W_{T_0,T}$,

$$\mathbf{E}\|u(t;T_0;u_0)\|^2_{\mathbf{X}} \le e^{\bar{C}(t-T_0)}\;\mathbf{E}\|u_0\|^2_{\mathbf{X}}. \tag{2.26}$$

Proof. The same arguments as in the proof of (2.22) show that

$$\sum_{|\alpha|=k} \frac{\|\varphi_\alpha(t;T_0;g)\|^2_{\mathbf{X}}}{\alpha!} = \sum_{l^k}\int_{T_0}^{(k,t)} \|F(t;s^k;l^k;g)\|^2_{\mathbf{X}} ds^k.$$

The assumptions and the definition of $F(t;s^k;l^k;g)$ imply

$$\|F(t;s^k;l^k;g)\|^2_{\mathbf{X}} \le e^{C_1(t-T_0)}(C_2)^k\|g\|^2_{\mathbf{X}}\ .$$

Therefore

$$\sum_{l^k}\int_{T_0}^{(k,t)} \|F(t;s^k;l^k;g)\|^2_{\mathbf{X}} ds^k \le e^{C_1(t-T_0)}\,\frac{[C_2 r(t-T_0)]^k}{k!}\|g\|^2_{\mathbf{X}}\ ,$$

whence (2.24) and (2.25). After that, inequality (2.26) follows from (2.25), Corollary 2.3.5, and the Markov property of u, because (2.25) implies the convergence of (2.14) in $L_2(\Omega,\mathcal{F}^W_{T_0,t^*},\mathbf{P};\mathbf{X})$ so that $u(t;T_0;g) \in L_2(\Omega,\mathcal{F}^W_{T_0,t^*},\mathbf{P};\mathbf{X})$ for every $g \in \mathbf{X}$.

□

To study the truncation of (2.14), it is necessary to note that the summation $\sum_{\alpha\in J}$ is double infinite:

$$\sum_{\alpha\in J} = \sum_{k=0}^{\infty}\sum_{|\alpha|=k} \tag{2.27}$$

and there are infinitely many multi-indices α with $|\alpha| = k > 0$. To begin with, the truncation of the length of α is considered.

Define $J_N = \{\alpha \in J : |\alpha| \le N\}$ and

$$u_N(s;T_0;u_0) = \sum_{\alpha\in J_N} \frac{1}{\sqrt{\alpha!}}\varphi_\alpha(s;T_0;u_0)\xi_\alpha(W_{T_0,t^*}). \tag{2.28}$$

Note that the summation in (2.28) is still infinite.

2.3.9. Theorem. *Assume that equation (2.8) is either coercive or dissipative, the initial condition u_0 is independent of $\mathcal{F}^W_{T_0,T}$, $t^* \in (T_0,T]$ is fixed, and $s \in [T_0,t^*]$. If $\|\Phi_t v\|_0^2 \le e^{C_1 t}\|v\|_0^2$, $\|B_l v\|_0^2 \le C_2\|v\|_0^2$, and $\bar{C} = C_1 + C_2 r$, then*

$$\sup_{s\in[T_0,t^*]} \mathbf{E}\|u(s;T_0;u_0) - u_N(s;T_0;u_0)\|_0^2 \le \frac{[C_2 r(t^*-T_0)]^{N+1}}{(N+1)!} e^{\bar{C}(t^*-T_0)}\, \mathbf{E}\|u_0\|_0^2. \tag{2.29}$$

Proof. To simplify the presentation, the arguments T_0 and u_0 will be omitted wherever possible.

By Theorem 2.3.7,

$$\sum_{|\alpha|=k} \frac{\|\varphi_\alpha(s)\|_0^2}{\alpha!} = \sum_{l^k} \int_{T_0}^{(k,s)} \|F(s;s^k;l^k)\|_0^2 ds^k. \tag{2.30}$$

Since the random variables $\xi_\alpha(W_{T_0,t})$ are uncorrelated and are independent of u_0, formulas (2.18) and (2.28) imply

$$\mathbf{E}\|u(s) - u_N(s)\|_0^2 = \sum_{k>N} \sum_{|\alpha|=k} \frac{\mathbf{E}\|\varphi_\alpha(s)\|_0^2}{\alpha!}.$$

By inequality (2.24),

$$\sum_{k>N} \sum_{|\alpha|=k} \frac{\mathbf{E}\|\varphi_\alpha(s)\|_0^2}{\alpha!} \le e^{C_1(s-T_0)}\, \mathbf{E}\|u_0\|_0^2 \sum_{k>N} \frac{(C_2 r(s-T_0))^k}{k!} \le \frac{(C_2 r(t^*-T_0))^{N+1}}{(N+1)!} e^{\bar{C}(t^*-T_0)}\, \mathbf{E}\|u_0\|_0^2,$$

which completes the proof.

□

For $k \le N$, the truncation of the sum $\sum_{|\alpha|=k}$ can be performed as follows. Define $J_N^n = \{\alpha \in J : |\alpha| \le N,\ d(\alpha) \le n\}$ and then

$$u_N^n(s;T_0;u_0) = \sum_{\alpha\in J_N^n} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(s;T_0;u_0)\xi_\alpha(W_{T_0,t^*}). \tag{2.31}$$

Now the summation in (2.31) is finite: if $d(\alpha) \le n$, then there are at most $(nr)^k$ multi-indices α with $|\alpha| = k$.

The error analysis for this truncation is more delicate.

2.3.10. Theorem. *Let the following conditions be fulfilled:*

(1) Equation (2.8) is either coercive or dissipative;

(2) The initial condition u_0 is independent of $\mathcal{F}^W_{T_0,T}$ and belongs to $L_2(\Omega;\mathbf{H}^2)$;

(3) The operators $\mathcal{B}_l$, $l = 1, \ldots, r$, are linear and bounded from $\mathbf{H}^a$ to $\mathbf{H}^a$, $a = 0, 2$, the operator $\mathcal{A}$ is linear and bounded from $\mathbf{H}^2$ to $\mathbf{H}^0$, and the operator Φ_t is linear and bounded from $\mathbf{H}^2$ to $\mathbf{H}^2$ for every $t \geq 0$ so that $\|\mathcal{A}v\|_0^2 \leq C_0\|v\|_2^2$, $\|\Phi_t v\|_a^2 \leq e^{C_1 t}\|v\|_a^2$, $\|\mathcal{B}_l v\|_a^2 \leq C_2\|v\|_a^2$, $a = 0, 2$;

(4) The basis $\{m_k\}$ is the Fourier cosine basis

$$m_1(s)=\frac{1}{\sqrt{t^*-T_0}};\ m_k(s)=\sqrt{\frac{2}{t^*-T_0}}\cos\left(\frac{\pi(k-1)(s-T_0)}{t^*-T_0}\right),\ k>1;\ T_0 \leq s \leq t^*. \quad (2.32)$$

Then

$$\begin{aligned}\mathbf{E}\|u_N(t^*;T_0;u_0) - u_N^n(t^*;T_0;u_0)\|_0^2 \leq\ & C_2 r\ e^{\bar{C}(t^*-T_0)}\Big(\epsilon(B)\frac{(t^*-T_0)^2}{n}\mathbf{E}\|u_0\|_0^2 + \\ & C_0\frac{(t^*-T_0)^3}{n}\mathbf{E}\|u_0\|_2^2\Big),\end{aligned} \quad (2.33)$$

where $\bar{C} = C_1 + C_2 r$ and $\epsilon(B) = 0$ if the operators $\mathcal{B}_l$ commute with one another (in particular, if $r = 1$); $1 \leq \epsilon(B) \leq 4$ otherwise.

Proof. To simplify the presentation, the arguments T_0 and u_0 will be omitted wherever possible and all constants are determined up to an absolute constant factor (e.g., $C_1 = 2C_1$).

If α is a multi-index with $|\alpha| = k$ and the characteristic set $\{(i_1^\alpha, q_1^\alpha) \ldots, (i_k^\alpha, q_k^\alpha)\}$, then $i_k^\alpha = d(\alpha)$, the order of α, and so the set J_N^n can be described as $\{\alpha \in J : |\alpha| \leq N;\ i_{|\alpha|}^\alpha \leq n\}$. Since the random variables ξ_α are uncorrelated and are independent of u_0,

$$\mathbf{E}\|u_N^n(t^*) - u_N(t^*)\|_0^2 = \sum_{b=n+1}^{\infty}\sum_{k=1}^{N}\sum_{|\alpha|=k;i_k^\alpha=b}\frac{\mathbf{E}\|\varphi_\alpha(t^*)\|_0^2}{\alpha!}.$$

The problem is thus to estimate $\sum_{b=n+1}^{\infty}\sum_{k=1}^{N}\sum_{|\alpha|=k;i_k^\alpha=b}\frac{\|\varphi_\alpha(t^*)\|_0^2}{\alpha!}$.

By Theorem 2.3.7 the corresponding solution φ_α of (2.16) can be written as

$$\varphi_\alpha(t^*) = \sum_{l^k}\int_{T_0}^{(k,t^*)} F(t^*; s^k; l^k)E_\alpha(s^k, l^k)ds^k. \quad (2.34)$$

According to (2.17), the characteristic set of $\alpha(i_k, q_k)$ is
$\{(i_1, q_1), \ldots, (i_{k-1}, q_{k-1})\}$, therefore it is possible to write

$$E_\alpha(s^k) = \sum_{j=1}^{k} m_{i_k}(s_j)1_{\{l_j=q_k\}}E_{\alpha(i_k,q_k)}(s_j^k; l_j^k),$$

where s_j^k (resp. l_j^k) denotes the same set $(s_1, \ldots, s_k)$ (resp. $(l_1, \ldots, l_k)$) with omitted s_j (resp. l_j); for example, $s_1^k = (s_2, \ldots, s_k)$.

As a result, (2.34) can be rewritten as

$$\varphi_\alpha(t^*)=\sum_{j=1}^{k}\sum_{l_j^k}\int\limits_{T_0}^{(k-1,t^*)}\Big(\int_{s_{j-1}}^{s_{j+1}}F(t^*;s^k;l^k)m_{i_k}(s_j)1_{\{l_j=q_k\}}ds_j\Big)E_{\alpha(i_k,q_k)}(s_j^k;l_j^k)ds_j^k, \qquad (2.35)$$

where $s_0 := T_0;\ s_{k+1} := t^*$. (Change the order of integration in the multiple integral.)

Denote

$$M_k(s) := \frac{\sqrt{2(t^*-T_0)}}{\pi(k-1)}\sin\Big(\frac{\pi(k-1)(s-T_0)}{(t^*-T_0)}\Big);\ k>1,\ T_0\le s\le t^*,$$

and also $F_j := \dfrac{\partial F(t^*;s^k;l^k)}{\partial s_j}$. Then, as long as $i_k = b > 1$, integration by parts in the inner integral on the right hand side of (2.35) yields:

$$\int_{s_{j-1}}^{s_{j+1}} F(t^*;s^k;l^k)m_b(s_j)ds_j =$$
$$F(t^*;s^k;l^k)M_b(s_j)\Big|_{s_j=s_{j-1}}^{s_j=s_{j+1}} - \int_{s_{j-1}}^{s_{j+1}} F_j(t^*;s^k;l^k)M_b(s_j)ds_j.$$

For each j, let us rename the remaining variables s_j^k in (2.35) as follows: $t_i := s_i, i \le j-1;\ t_i := s_{i+1}, i > j-1$, or, symbolically, $t^{k-1} := s_j^k$. We will set $t_0 := T_0$, $t_k := t^*$ and denote by $t^{k-1,j}, j = 1,\dots,k-1$, the set t^{k-1} in which t_j is repeated twice (e.g. $t^{k-1,1} = (t_1,t_1,\dots,t_{k-1})$, etc.); also $t^{k-1,0} := (t_0,t_1,t_2,\dots,t_{k-1})$, $t^{k-1,k} := (t_1,\dots,t_{k-1},t_k)$.

The similar changes will also be made with the set l^k: for fixed j, there are $k-1$ free indices $l_1,\dots,l_{j-1},l_{j+1},\dots,l_k$ and they are renamed just like s^k to form the set l^{k-1} (in this case, the same symbols are used). Similarly, $l^{k-1,j}$ denotes the set $(l_1,\dots,l_{j-1},q_k,l_j,\dots,l_{k-1})$. After these transformations, $E_{\alpha(i_k,q_k)}(s_j^k;l_j^k)$ becomes $E_{\alpha(i_k,q_k)}(t^{k-1};l^{k-1})$ - independent of j, and

$$F(t^*;s^k;l^k)1_{\{l_j=q_k\}}M_b(s_j)\Big|_{s_j=s_{j-1}}^{s_j=s_{j+1}} =$$
$$F(t^*;t^{k-1,j};l^{k-1,j})M_b(t_j) - F(t^*;t^{k-1,j-1};l^{k-1,j})M_b(t_{j-1}),\ j=1,\dots,k.$$

Therefore, if $d(\alpha) = b > 1$ and $|\alpha| = k > 0$, then

$$\varphi_\alpha(t^*) = \sum_{l^{k-1}}\int\limits_{T_0}^{(k-1,t^*)}\Big(f_b^{(1)}(t^*;t^{k-1};l^{k-1})+ f_b^{(2)}(t^*;t^{k-1};l^{k-1})\Big)E_{\alpha(i_k,q_k)}(t^{k-1};l^{k-1})dt^{k-1},$$

where

$$f_b^{(1)}(t^*;t^{k-1};l^{k-1}) = \sum_{j=1}^{k}\Big(F(t^*;t^{k-1,j};l^{k-1,j})M_b(t_j)- F(t^*;t^{k-1,j-1};l^{k-1,j})M_b(t_{j-1})\Big)\quad \text{if } k>1,$$

$f_b^{(1)} = 0$ if $k=1$ – because $M_b(t_0) = M_b(t_k) = 0$ (this is the only place where the choice of $\{m_k\}$ really makes the difference), and

$$\begin{aligned} f_b^{(2)}(t^*; t^{k-1}; l^{k-1}) = & -\int_{T_0}^{t_1} F_1(t^*; s, t^{k-1}; q_k, l^{k-1}) M_b(s) ds - \\ & \sum_{j=2}^{k-1} \int_{t_{j-1}}^{t_j} F_j(t^*; \ldots, t_{j-1}, s, t_j, \ldots; l^{k-1,j}) M_b(s) ds - \\ & \int_{t_{k-1}}^{t_k} F_k(t^*; t^{k-1}, s; l^{k-1}, q_k) M_b(s) ds. \end{aligned}$$

Note that if the operators $\mathcal{B}_l$ commute with one another, then $f_b^{(1)}(t^*; t^{k-1}; l^{k-1})$ is identically equal to zero for all k.

Since $|\alpha(i_{|\alpha|}, q_{|\alpha|})| = |\alpha| - 1$ and $\alpha! \geq \alpha(i_{|\alpha|}, q_{|\alpha|})!$, it now follows from (2.35) that

$$\begin{aligned} & \sum_{|\alpha|=k; i_k^\alpha = b} \frac{\|\varphi_\alpha(t^*)\|_0^2}{\alpha!} = \\ & \sum_{|\alpha|=k; i_k^\alpha = b} \sum_{q_k=1}^{r} \left\| \frac{1}{\sqrt{\alpha!}} \sum_{l^{k-1}} \int_{T_0}^{(k-1,t^*)} (f_b^{(1)} + f_b^{(2)}) E_{\alpha(b,q_k)} dt^{k-1} \right\|_0^2 \leq \\ & \sum_{q_k=1}^{r} \sum_{|\beta|=k-1} \left\| \frac{1}{\sqrt{\beta!}} \sum_{l^{k-1}} \int_{T_0}^{(k-1,t^*)} (f_b^{(1)} + f_b^{(2)}) E_\beta dt^{k-1} \right\|_0^2, \end{aligned}$$

and arguments similar to those used in the proof of Theorem 2.3.7 show that the last expression is equal to

$$\sum_{q_k=1}^{r} \sum_{l^{k-1}} \int_{T_0}^{(k-1,t^*)} \left\| f_b^{(1)}(t^*; t^{k-1}; l^{k-1}) + f_b^{(2)}(t^*; t^{k-1}; l^{k-1}) \right\|_0^2 dt^{k-1}. \tag{2.36}$$

Definition of $f_b^{(1)}$ and the assumptions of the theorem imply

$$\|f_b^{(1)}\|_0^2 = 0,\ k=1; \quad \|f_b^{(1)}\|_0^2 \leq \frac{(C_2)^k \epsilon(B) k (t^* - T_0)}{(b-1)^2} e^{C_1(t^*-T_0)} \|u_0\|_0^2,\ k \geq 2. \tag{2.37}$$

Next, direct computations yield

$$\begin{aligned} F_j(t^*; s^k; l^k) = & \ \Phi_{t^*-s_k} \mathcal{B}_{l_k} \ldots \Phi_{s_{j+1}-s_j} \mathcal{B}_{l_j} \mathcal{A} \Phi_{s_j - s_{j-1}} \ldots \Phi_{s_1 - T_0} u_0 - \\ & \ \Phi_{t^*-s_k} \mathcal{B}_{l_k} \ldots \mathcal{A} \Phi_{s_{j+1}-s_j} \mathcal{B}_{l_j} \Phi_{s_j - s_{j-1}} \ldots \Phi_{s_1 - T_0} u_0, \end{aligned}$$

so that by assumption (3) of the theorem,

$$\|F_j(t^*; s^k; l^k)\|_0^2 \leq C_0 (C_2)^k e^{C_1(t^*-T_0)} \|u_0\|_2^2.$$

After that the definition of $f_b^{(2)}$ and the obvious inequalities

$$(a_1 + \ldots + a_k)^2 \leq k(a_1^2 + \ldots + a_k^2)$$

and

$$\Big(\int_{T_0}^{t} f(y)dy\Big)^2 \le (t-T_0)\int_{T_0}^{t}(f(y))^2 dy$$

imply:

$$\begin{aligned}\|f_b^{(2)}\|_0^2 \le\ & C_0 k(C_2)^k e^{C_1(t^*-T_0)}\|u_0\|_2^2(t^*-T_0)\int_{T_0}^{t^*}(M_b(s))^2 ds \le \\ & \frac{C_0 k(C_2)^k(t^*-T_0)^3}{(b-1)^2}e^{C_1(t^*-T_0)}\ \|u_0\|_2^2\ ;\end{aligned}$$

so, since $\int\limits_{T_0}^{(k-1,t^*)} dt^{k-1} = (t^*-T_0)^{k-1}/(k-1)!$, (2.36), (2.37) and the last inequality yield

$$\begin{aligned}&\mathbf{E}\|u_N(t^*)-u_N^n(t^*)\|_0^2 = \sum_{b\ge n+1}\sum_{k=1}^{N}\sum_{|\alpha|=k;i_k^\alpha=b}\frac{\mathbf{E}\|\varphi_\alpha(t^*)\|_0^2}{\alpha!}\le \\ &C_2 r\ e^{C_1(t^*-T_0)}\Big[\epsilon(B)(t^*-T_0)^2\mathbf{E}\|u_0\|_0^2\sum_{k\ge 0}\frac{k+2}{k+1}\frac{(C_2 r(t^*-T_0))^k}{k!}+ \\ &C_0(t^*-T_0)^3\mathbf{E}\|u_0\|_2^2\sum_{k\ge0}\frac{(k+1)(C_2 r(t^*-T_0))^k}{k!}\Big]\sum_{b\ge n}\frac{1}{b^2}\le \\ &C_2 r\ e^{\bar{C}(t^*-T_0)}\Big[\epsilon(B)\frac{(t^*-T_0)^2}{n}\mathbf{E}\|u_0\|_0^2 + C_0\frac{(t^*-T_0)^3}{n}\mathbf{E}\|u_0\|_2^2\Big].\end{aligned}$$

This completes the proof of the theorem.

□

2.3.11. Corollary. *Under the assumptions of Theorem 2.3.10,*

$$\begin{aligned}\mathbf{E}\|u(t^*;T_0;u_0)-u_N^n(t^*;T_0;u_0)\|_0^2 \le e^{\bar{C}(t^*-T_0)}\bigg(&\frac{[C_2 r(t^*-T_0)]^{N+1}}{(N+1)!}\mathbf{E}\|u_0\|_0^2+ \\ C_2 r\ \epsilon(B)\frac{(t^*-T_0)^2}{n}\mathbf{E}\|u_0\|_0^2 + C_2 C_0 r&\frac{(t^*-T_0)^3}{n}\mathbf{E}\|u_0\|_2^2\bigg)\end{aligned} \tag{2.38}$$

Proof. It follows from (2.18), (2.28), (2.31), and the orthogonality of ξ_α that

$$\begin{aligned}\mathbf{E}\|u(t^*;T_0;u_0)-u_N^n(t^*;T_0;u_0)\|_0^2 = \mathbf{E}\|u(t^*;T_0;u_0)-u_N(t^*;T_0;u_0)\|_0^2+ \\ \mathbf{E}\|u_N(t^*;T_0;u_0)-u_N^n(t^*;T_0;u_0)\|_0^2,\end{aligned}$$

and then (2.38) follows from (2.29) and (2.33).

□

Recall that if $\alpha=(\alpha_k^l)_{1\le l\le r,\ k\ge1}$, then α_k^l defines the degree of the Hermite polynomial of $\int_{T_0}^{t^*} m_k(s)dW_l(s)$ used in the construction of $\xi_\alpha(W_{T_0,t^*})$. If $d(\alpha)\le n$, then $\alpha_k^l=0$ for all $k>n$, so the truncation of the order of α is equivalent to keeping only the first n elements of the (deterministic) basis $\{m_k(s)\}_{k\ge1}$.

On the other hand, by restricting the length of α, we eliminate a number of elements of the stochastic basis $\{\xi_\alpha(W_{T_0,t^*})\}$, which are otherwise available with the retained collection of $\{m_k\}$.

Thus, restriction of the order of α makes the inner sum in (2.27) finite and is equivalent to the truncation of the deterministic basis $\{m_k\}$ while restriction of the length of α makes the outer sum in (2.27) finite and is equivalent to the truncation of the stochastic basis $\{\xi_\alpha(W_{T_0,t^*})\}$.

Note that the error bound in (2.29) holds for every point $s \in [T_0, t^*]$ and does not depend on the choice of the basis $\{m_k\}$. The reason is that the only property used in the proof of (2.29) was $\mathbf{E}\xi_\alpha(W_{T_0,t^*})\xi_\beta(W_{T_0,t^*}) = 0$ for $\alpha \neq \beta$, which is true for every orthogonal system $\{m_k\}$. On the other hand, the error bound in (2.33) is valid only at the end point t^*, depends crucially on the choice of the basis $\{m_k\}$, and requires extra regularity assumptions. Since the truncation in the order of α is related to the truncation of the Fourier series of the function $F(t^*; s^k; l^k; u_0)$, these additional restrictions are not surprising, and neither is the slow rate of convergence in n.

2.4 Step-by-Step Expansion

In this section, a recursive version of the Wiener chaos decomposition is studied for equation (2.8). There are two main reasons to consider the recursive expansion:

1. Representation (2.18) cannot be applied when $t > t^*$, so, to extend the time interval, it is necessary to choose another basis $\{m_k\}$ and again solve system (2.16);

2. It follows from (2.38) that the error of the truncation of (2.18) increases rapidly as $t^* - T_0$ increases, so the application of (2.18) to solving (2.2) is limited to short time intervals.

Let $T_0 = t_0 < t_1 < \ldots < t_M = T$ be a partition of the interval $[T_0, T]$ and assume that $\{m_k^i\} = \{m_k^i(s)\}_{k\geq 1}$ is a CONS in $L_2([t_{i-1}, t_i])$. In principle, the partition can be not uniform and the collections $\{m_k^i\}$ can be different for different i. Define random variables

$$\left\{ \xi_\alpha^i(W_{t_{i-1},t_i}) := \prod_{k,l} \left(\frac{H_{\alpha_k^l}(\xi_{k,l}^i)}{\sqrt{\alpha_k^l!}} \right), \quad \alpha \in J \right\}, \tag{2.39}$$

where $\xi_{k,l}^i = \int_{t_{i-1}}^{t_i} m_k^i(s) dW_l(s)$ and H_n is the nth Hermite polynomial (2.10).

The following is the recursive analog of Theorem 2.3.3.

2.4.1. Theorem. *Assume that equation (2.8) is either coercive or dissipative and the initial condition u_0 is independent of $\mathcal{F}^W_{T_0,T}$. Then for every $t \in [t_{i-1}, t_i]$ the solution $u = u(t; T_0; u_0)$ can be written as*

$$u(t; T_0; u_0) = \sum_{\alpha \in J} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha^i(t; t_{i-1}; u(t_{i-1}; T_0; u_0)) \xi_\alpha^i(W_{t_{i-1},t_i}), \quad i = 1, \ldots, M, \tag{2.40}$$

where the coefficients $\varphi_\alpha^i = \varphi_\alpha^i(t; t_{i-1}; g)$ are the solution of the system of equations

$$\begin{aligned} \frac{\partial \varphi_\alpha^i(t; t_{i-1}; g)}{\partial t} &= \mathcal{A}\varphi_\alpha^i(t; t_{i-1}; g) + \sum_{k,l} \alpha_k^l m_k^i(t) \mathcal{B}_l \varphi^i_{\alpha(k,l)}(t; t_{i-1}; g), \ t_{i-1} < t \leq t_i; \\ \varphi_\alpha^i(t_{i-1}; t_{i-1}; g) &= g 1_{\{|\alpha|=0\}}. \end{aligned} \tag{2.41}$$

Proof. If $t \in [t_{i-1}, t_i]$, then by property (P1) from Theorem 2.2.7,

$$u(t; T_0; u_0) = u(t; t_{i-1}; u(t_{i-1}; T_0; u_0))$$

and by Corollary 2.3.4, $u(t_{i-1}; T_0; u_0)$ is independent of $\mathcal{F}^W_{t_i,T}$, so that (2.40) follows directly from (2.18).

□

2.4.2. Corollary. *If* $t_i - t_{i-1} = \Delta_i$ *and*

$$m^i_k(s) = \frac{1}{\sqrt{\Delta_i}} m_k\left(\frac{s - t_{i-1}}{\Delta_i}\right), \quad s \in [t_{i-1}, t_i],$$

where $m_k = m_k(t)$ *form a CONS in* $L_2([0,1])$*, then*

$$u(t_i; T_0; u_0) = \sum_{\alpha \in J} \frac{1}{\sqrt{\alpha!}} \varphi^i_\alpha(\Delta_i; u(t_{i-1}; T_0; u_0)) \xi^i_\alpha(W_{t_{i-1},t_i}), \quad i = 1, \ldots, M, \tag{2.42}$$

and the coefficients $\varphi^i_\alpha = \varphi^i_\alpha(t; g)$ *are the solution of the system of equations*

$$\begin{aligned} \frac{\partial \varphi^i_\alpha(t; g)}{\partial s} &= \mathcal{A}\varphi^i_\alpha(t; g) + \frac{1}{\sqrt{\Delta_i}} \sum_{k,l} \alpha^l_k m_k(t/\Delta_i) \mathcal{B}_l \varphi^i_{\alpha(k,l)}(t; g), \ 0 < t \le \Delta_i; \\ \varphi^i_\alpha(0; g) &= g 1_{\{|\alpha|=0\}}. \end{aligned} \tag{2.43}$$

Proof. With this special choice of the basis $\{m^i_k\}$, Theorem 2.3.7 implies that $\varphi^i_\alpha(\Delta_i; g) = \varphi^i_\alpha(t_i; t_{i-1}; g)$ and then (2.42) follows from (2.40).

□

It will be shown next that the recursive application of the truncated expansion (2.31) results in an approximation that converges to the exact solution on the given time interval as the size of the partition tends to zero.

Assume that

$$m^i_1(s) = \frac{1}{\sqrt{\Delta_i}}; \ m^i_k(s) = \sqrt{\frac{2}{\Delta_i}} \cos\left(\frac{\pi(k-1)(s - t_{i-1})}{\Delta_i}\right), \ k > 1; \ T_{i-1} \le s \le t_i. \tag{2.44}$$

Define $u^n_N(t_0) = u_0$ and then by induction

$$u^n_N(t_i) = \sum_{\alpha \in J_N} \frac{1}{\sqrt{\alpha!}} \varphi^i_\alpha(\Delta_i; u^n_N(t_{i-1})) \xi_\alpha(W_{t_{i-1},t_i}), \tag{2.45}$$

where $\varphi^i_\alpha(\Delta_i; \cdot)$ is the solution of (2.43) with the corresponding initial condition and the functions m^i_k are chosen according to (2.44). As before, $J^n_N = \{\alpha \in J : |\alpha| \le N, \ d(\alpha) \le n\}$.

2.4.3. Theorem. *Let the following conditions be fulfilled:*

(1) Equation (2.8) is either coercive or dissipative;

(2) The initial condition u_0 is independent of $\mathcal{F}^W_{T_0,T}$ and belongs to $L_2(\Omega;\mathbf{H}^2)$;

(3) The operators $\mathcal{B}_l$, $l=1,\dots,r$, are linear and bounded from $\mathbf{H}^a$ to $\mathbf{H}^a$, $a=0,2$, the operator $\mathcal{A}$ is linear and bounded from $\mathbf{H}^2$ to $\mathbf{H}^0$, and the operator Φ_t is linear and bounded from $\mathbf{H}^2$ to $\mathbf{H}^2$ for every $t\ge 0$ so that $\|\mathcal{A}v\|_0^2\le C_0\|v\|_2^2$, $\|\Phi_t v\|_a^2\le e^{C_1 t}\|v\|_a^2$, $\|\mathcal{B}_l v\|_a^2\le C_2\|v\|_a^2$, $a=0,2$.

With no loss of generality, assume that $\bar{C}:=C_1+C_2 r>0$. Then

$$\max_{1\le i\le M}\mathbf{E}\|u^n_N(t_i)-u(t_i;T_0;u_0)\|_0^2\le\frac{e^{\bar{C}(T-T_0)}}{\bar{C}}\Big(\frac{[C_2 r\Delta]^N}{(N+1)!}\mathbf{E}\|u_0\|_0^2+ C_2 r\epsilon(B)\frac{\Delta}{n}\mathbf{E}\|u_0\|_0^2+C_0C_b r\frac{\Delta^2}{n}\mathbf{E}\|u_0\|_2^2\Big), \tag{2.46}$$

where $\Delta=\max_{1\le i\le M}\Delta_i$, and $\epsilon(B)=0$ if the operators $\mathcal{B}_l$ commute with one another (in particular, if $r=1$), $1\le\epsilon(B)\le 4$ otherwise.

Proof. To simplify the presentation, the arguments T_0 and u_0 will be omitted wherever possible and all constants are determined up to an absolute constant factor.

It follows from (2.42), (2.45), and the linearity of (2.43) that

$$\begin{aligned}\mathbf{E}\|u^n_N(t_i)-u(t_i)\|_0^2=&\sum_{\alpha\in J^n_N}\frac{1}{\alpha!}\mathbf{E}\|\varphi^i_\alpha(\Delta_i;u^n_N(t_{i-1})-u(t_{i-1}))\|_0^2+\\ &\sum_{\alpha\notin J^n_N}\frac{1}{\alpha!}\mathbf{E}\|\varphi^i_\alpha(\Delta_i;u(t_{i-1}))\|_0^2\le\\ &\sum_{\alpha\in J}\frac{1}{\alpha!}\mathbf{E}\|\varphi^i_\alpha(\Delta_i;u^n_N(t_{i-1})-u(t_{i-1}))\|_0^2+\\ &\sum_{\alpha\notin J^n_N}\frac{1}{\alpha!}\mathbf{E}\|\varphi^i_\alpha(\Delta_i;u(t_{i-1}))\|_0^2.\end{aligned}\tag{2.47}$$

By (2.42) from Corollary 2.4.2,

$$\begin{aligned}&\sum_{\alpha\in J}\frac{1}{\alpha!}\mathbf{E}\|\varphi^i_\alpha(\Delta_i;u^n_N(t_{i-1})-u(t_{i-1}))\|_0^2=\\ &\mathbf{E}\|u(t_i;t_{i-1};u^n_N(t_{i-1})-u(t_{i-1}))\|_0^2,\end{aligned}\tag{2.48}$$

and by (2.26) from Corollary 2.3.8,

$$\mathbf{E}\|u(t_i;u^n_N(t_{i-1})-u(t_{i-1}))\|_0^2\le e^{\bar{C}\Delta_i}\mathbf{E}\|u^n_N(t_{i-1})-u(t_{i-1})\|_0^2. \tag{2.49}$$

Next, by Corollary 2.3.11,

$$\sum_{\alpha\notin J^n_N}\frac{1}{\alpha!}\mathbf{E}\|\varphi^i_\alpha(\Delta_i;u(t_{i-1}))\|_0^2\le e^{\bar{C}(T-T_0)}\Big(\frac{[C_2 r\Delta_i]^{N+1}}{(N+1)!}\mathbf{E}\|u_0\|_0^2+ C_2 r\epsilon(B)\frac{\Delta_i^2}{n}\mathbf{E}\|u_0\|_0^2+C_0C_2 r\frac{\Delta_i^3}{n}\mathbf{E}\|u_0\|_2^2\Big), \tag{2.50}$$

and by (2.26),

$$\mathbf{E}\|u(t_{i-1})\|_a^2 \le e^{\bar{C}(T-T_0)}\|u_0\|_a^2, \quad a=0,2. \tag{2.51}$$

If $\mathbf{E}\|u_N^n(t_i)-u(t_i)\|_0^2$ is denoted by ε_i, then (2.47) – (2.51) imply

$$\begin{aligned}\varepsilon_i \le e^{\bar{C}\Delta_i}\varepsilon_{i-1} + e^{\bar{C}(T-T_0)}\Big(\frac{[C_2 r\Delta_i]^{N+1}}{(N+1)!}\mathbf{E}\|u_0\|_0^2 + \\ C_2 r\epsilon(B)\frac{\Delta_i^2}{n}\mathbf{E}\|u_0\|_0^2 + C_0 C_b r\frac{\Delta_i^3}{n}\mathbf{E}\|u_0\|_2^2\Big) \le \\ e^{\bar{C}\Delta}\varepsilon_{i-1} + e^{\bar{C}(T-T_0)}\Big(\frac{[C_2 r\Delta]^{N+1}}{(N+1)!}\mathbf{E}\|u_0\|_0^2 + \\ C_2 r\epsilon(B)\frac{\Delta^2}{n}\mathbf{E}\|u_0\|_0^2 + C_0 C_2 r\frac{\Delta^3}{n}\mathbf{E}\|u_0\|_2^2\Big)\end{aligned}$$

Since $\varepsilon_0=0$, the statement of the theorem follows from the discrete Gronwall lemma.

□

Using the properties of the solution of (2.41), the recursive relation (2.40) can be applied repeatedly resulting in

$$\begin{aligned}&u(t;T_0;u_0) = \\ &\sum_{\alpha,\beta\in J}\frac{1}{\sqrt{\alpha!\beta!}}\varphi_\alpha^i(t;t_{i-1};\varphi_\beta^{i-1}(t_{i-1};t_{i-2};u(t_{i-2};T_0;u_0)))\xi_\alpha^i(W_{t_{i-1},t_i})\xi_\beta^{i-1}(W_{t_{i-2},t_{i-1}}) = \\ &\cdots = \sum_{\alpha,\beta,\dots,\gamma\in J}\frac{1}{\sqrt{\alpha!\beta!\cdots\gamma!}}\varphi_\alpha^i(t;t_{i-1};\varphi_\beta^{i-1}(t_{i-1};t_{i-2}\dots\varphi_\gamma^1(t_1;T_0;u_0)\dots)) \\ &\xi_\alpha^i(W_{t_{i-1},t_i})\xi_\beta^{i-1}(W_{t_{i-2},t_{i-1}})\cdots\xi_\gamma^1(W_{t_1,T_0}).\end{aligned} \tag{2.52}$$

Continue $m_k^i(t)$ to $[T_0,T]$ by setting $m_k^i(t)=0$ if $t\notin[t_{i-1},t_i]$. Then the collection $\{m_k^i\}$ becomes a CONS in $L_2([T_0,T])$ and it is possible to consider the one step expansion (2.18) using this CONS.

2.4.4. Theorem. *If the CONS in $L_2([T_0,T])$ is given by $\{m_k^i\}$, then for every $t\in[t_{i-1},t_i]$ the one step representation (2.18) is equivalent to (2.52) in the following sense: every nonzero term $\varphi_\alpha(t;T_0;u_0)\xi_\alpha(W_{T_0,T})$ from (2.18) coincides with some term on the right hand side of (2.52) and conversely, every term on the right hand side of (2.52) coincides with some nonzero term $\varphi_\alpha(t;T_0;u_0)\xi_\alpha(W_{T_0,T})$ from (2.18).*

Proof. Even though representation (2.40) is recursive, representation (2.52) is not, and neither is (2.18), so the natural proof by induction (anchor step – induction step) is not applicable. The equivalence can be verified directly by considering consequetively $i=1,2,\dots,M$. To simplify the presentation, the proof will be given for the partition $T_0=t_0<t_1<t_2=T$. For the general partition, the same approach is used but the notations are far more complicated.

To begin with, consider system (2.16) when the basis $\{m_k\}$ is given by $\{m_{k'}^1, m_{k''}^2\}$, where m_k^i is supported on $[t_{i-1},t_i]$, $i=1,2$. Recall that if $\alpha_k^l\ne 0$ for some $l=1,\dots,r$, then the function m_k is present on the right hand side of (2.16).

Let us break the multi-index α into two parts: $\alpha = \{\beta, \gamma\}$, where both β and γ are multi-indices so that the nonzero elements of β and γ correspond to $m^1_{k'}$ and $m^2_{k''}$ respectively. Then equation (2.16) can be rewritten as

$$
\begin{aligned}
\frac{\partial \varphi_\alpha(t;T_0;u_0)}{\partial t} &= \mathcal{A}\varphi_\alpha(t;T_0;u_0) + \sum_{k,l} \beta^l_k m^1_k(t)\mathcal{B}_l \varphi_{\beta(k,l),\gamma}(t;T_0;u_0) + \\
&\quad \sum_{k,l} \gamma^l_k m^2_k(t)\mathcal{B}_l \varphi_{\beta,\gamma(k,l)}(t;T_0;u_0), \quad T_0 < t \le T; \\
\varphi_\alpha(T_0;T_0;u_0) &= u_0 1_{\{|\beta|+|\gamma|=0\}}.
\end{aligned} \tag{2.53}
$$

If $t < t_1$, then $m^2_k(s) = 0$ and (2.53) becomes

$$
\begin{aligned}
\frac{\partial \varphi_\alpha(t;T_0;u_0)}{\partial t} &= \mathcal{A}\varphi_\alpha(t;T_0;u_0) + \sum_{k,l} \beta^l_k m^1_k(t)\mathcal{B}_l \varphi_{\beta(k,l),\gamma}(t;T_0;u_0), \quad T_0 < t \le t_1; \\
\varphi_\alpha(T_0;T_0;u_0) &= u_0 1_{\{|\beta|+|\gamma|=0\}}.
\end{aligned}
$$

If $|\gamma| = 0$ and 0 denotes the zero multi-index, then $\varphi_{\beta,0}$ satisfies

$$
\begin{aligned}
\frac{\partial \varphi_{\beta,0}(t;T_0;u_0)}{\partial t} &= \mathcal{A}\varphi_{\beta,0}(t;T_0;u_0) + \sum_{k,l} \beta^l_k m^1_k(t)\mathcal{B}_l \varphi_{\beta(k,l),0}(t;T_0;u_0), \quad T_0 < t \le t_1; \\
\varphi_{\beta,0}(T_0;T_0;u_0) &= u_0 1_{\{|\beta|=0\}}.
\end{aligned}
$$

Comparing this with (2.41) yields

$$
\varphi_{\beta,0}(t;T_0;u_0) = \varphi^1_\beta(t;T_0;u_0), \ t \in [t_0, t_1].
$$

If $|\gamma| > 0$, then $|\beta| + |\gamma| > 0$ and (2.53) implies that $\varphi_{\beta,\gamma}$ satisfies

$$
\begin{aligned}
\frac{\partial \varphi_{\beta,\gamma}(t;T_0;u_0)}{\partial t} &= \mathcal{A}\varphi_\alpha(t;T_0;u_0) + \sum_{k,l} \beta^l_k m^1_k(t)\mathcal{B}_l \varphi_{\beta(k,l),\gamma}(t;T_0;u_0), \quad T_0 < t \le t_1; \\
\varphi_{\beta,\gamma}(T_0;T_0;u_0) &= 0,
\end{aligned}
$$

and therefore $\varphi_{\beta,\gamma}(t;T_0;u_0) = 0, \ t \in [t_0, t_1]$, by the uniqueness of the solution. As a result, if $t \in [t_0, t_1]$, then

$$
u(t;T_0;u_0) = \sum_{\beta \in J} \frac{1}{\sqrt{\beta!}} \varphi_{\beta,0}(t;T_0;u_0)\xi_{\beta,0}(W_{T_0,T}) = \sum_{\beta \in J} \frac{1}{\sqrt{\beta!}} \varphi^1_\beta(t;T_0;u_0)\xi^1_\beta(W_{T_0,t_1}),
$$

which is the same as (2.52) with $i = 1$.

Next, consider system (2.53) for $t \in [t_1, t_2]$. In this case, $m^1_k(t) = 0$ and

$$
\begin{aligned}
\frac{\partial \varphi_\alpha(t;T_0;u_0)}{\partial t} &= \mathcal{A}\varphi_{\beta,\gamma}(t;T_0;u_0) + \sum_{k,l} \beta^l_k m^1_k(t)\mathcal{B}_l \varphi_{\beta,\gamma(l,k)}(t;T_0;u_0), \quad t_1 < t \le T; \\
\varphi_\alpha(t_1;T_0;u_0) &= \begin{cases} \varphi^1_\beta(t_1;T_0;u_0), & \text{if } |\gamma| = 0, \\ 0, & \text{if } |\gamma| > 0. \end{cases}
\end{aligned}
$$

For every fixed β this system is of the type (2.41) so that the solution is given by

$$\varphi_\alpha(t;T_0;u_0) = \varphi^2_\gamma(t;t_1;\varphi^1_\beta(t_1;T_0;u_0)).$$

Since by definition

$$\xi_\alpha(W_{T_0,T}) = \prod_{k,l}\left(\frac{H_{\beta^l_k}(\xi^1_{k,l})}{\sqrt{\beta^l_k!}}\right)\prod_{k,l}\left(\frac{H_{\gamma^l_k}(\xi^2_{k,l})}{\sqrt{\gamma^l_k!}}\right) = \xi^1_\beta(W_{T_0,t_1})\xi^2_\gamma(W_{t_1,T}),$$

it follows that for $t \in [t_1, t_2]$,

$$u(t;T_0;u_0) = \sum_{\alpha\in J}\frac{1}{\sqrt{\alpha!}}\varphi_\alpha(t;T_0;u_0)\xi_\alpha(W_{T_0,T}) =$$
$$\sum_{\beta,\gamma\in J}\frac{1}{\sqrt{\beta!\gamma!}}\varphi^2_\gamma(t;t_1;\varphi^1_\beta(t_1;T_0;u_0))\xi^1_\beta(W_{T_0,t_1})\xi^2_\gamma(W_{t_1,T}),$$

and the last expression is the same as (2.52) with $i = 2$.

□

The advantages of (2.52) as compared to (2.40) are:

1. Representation (2.52) provides complete separation of the deterministic and stochastic data of equation (2.8). This separation can be used for solving some filtering problems.
2. Representation (2.52) provides more flexibility in truncating the infinite sum.

The main disadvantage, which effectively rules out any practical application of (2.52), is that the representation of the solution on (2.52) is not recursive and the number of terms in any reasonably truncated sum becomes too large after very few steps. As a result, both numerical schemes for the Zakai filtering equation that are presented below in Chapter 4 are based on representation (2.40).

2.5 Multiple Wiener Integral Expansion

In this section, an alternative representation of the solution of (2.8) will be presented using the relation between the Cameron - Martin basis in $L_2(\Omega, \mathcal{F}_{T_0,t}, \mathbf{P})$ and the multiple (or, more precisely, iterated) Wiener integrals. Recall that equation (2.8) is

$$u(t;T_0;u_0) = u_0 + \int_{T_0}^t \mathcal{A}u(s;T_0;u_0)ds + \sum_{l=1}^r \int_{T_0}^t \mathcal{B}_l u(s;T_0;u_0)dW_l(s),\ t \in [T_0, T], \tag{2.54}$$

considered in the Hilbert scale $\{\mathbf{H}^a\}_{a\in\mathbf{R}}$, and u_0 is independent of $\mathcal{F}^W_{T_0,T}$.

Assume that $t^* \in [T_0, T]$ is fixed and recall that the Cameron - Martin basis in $L_2(\Omega, \mathcal{F}_{T_0,t^*}, \mathbf{P})$ is $\{\xi_\alpha(W_{T_0,t^*})\}_{\alpha\in J}$, where J is the set of r-dimensional multi-indices,

$$\xi_\alpha(W_{T_0,t^*}) = \frac{1}{\sqrt{\alpha!}}\prod_{k,l} H_{\alpha^l_k}\left(\int_{T_0}^{t^*} m_k(s)dW_l(s)\right),$$

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2},$$

and $\{m_k\}$ is a CONS in $L_2([T_0, t^*])$. If $|\alpha| = k$, then by Theorem 3.1 in [30],

$$\xi_\alpha(W_{T_0,t^*}) = \frac{1}{\sqrt{\alpha!}} \sum_{l^k} \int_{T_0}^{t^*} \int_{T_0}^{s_k} \cdots \int_{T_0}^{s_2} E_\alpha(s^k; l^k) dW_{l_1}(s_1) \cdots dW_{l_k}(s_k), \tag{2.55}$$

where

$$E_\alpha(s^k; l^k) := \sum_{\sigma \in \mathcal{P}^k} m_{i_1}(s_{\sigma(1)}) 1_{\{l_{\sigma(1)} = q_1\}} \cdots m_{i_k}(s_{\sigma(k)}) 1_{\{l_{\sigma(k)} = q_k\}}. \tag{2.56}$$

(the notations were introduced before Theorem 2.3.7).

2.5.1. Theorem. *Assume that equation (2.54) is either coercive or dissipative and each operator $\mathcal{B}_l$, $l = 1, \ldots, r$ is linear and bounded from $\mathbf{H}^0$ to $\mathbf{H}^0$. Then for every $t \in [T_0, T]$ the solution $u(t; T_0; u_0)$ can be written as*

$$u(t; T_0; u_0) = \Phi_{t-T_0} u_0 + \sum_{k \geq 1} \sum_{l^k} \int_{T_0}^{t} \int_{T_0}^{s_k} \cdots \int_{T_0}^{s_2} F(t; s^k; l^k; u_0) dW_{l_1}(s_1) \cdots dW_{l_k}(s_k), \tag{2.57}$$

where $F(t; s^k; l^k; u_0) := \Phi_{t-s_k} \mathcal{B}_{l_k} \Phi_{s_k - s_{k-1}} \ldots \mathcal{B}_{l_1} \Phi_{s_1 - T_0} u_0$ and Φ_t is the semigroup generated by $\mathcal{A}$.

Proof. For simplicity, the arguments T_0 and u_0 will be omitted wherever possible.

It follows from (2.18) and (2.21) that for $t \in [T_0, t^*]$,

$$u(t) = \Phi_{t-T_0} u_0 + \sum_{k \geq 1} \sum_{|\alpha| = k} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(t) \xi_\alpha(W_{T_0,t^*}) \tag{2.58}$$

and by (2.23),

$$\varphi_\alpha(t) = \frac{1}{k!} \sum_{l^k} \int_{[T_0,t^*]^k} G(t; s^k; l^k) E_\alpha(s^k; l^k) ds^k, \tag{2.59}$$

where

$$G(t; s^k; l^k) := \sum_{\sigma \in \mathcal{P}^k} \Phi_{t - s_{\sigma(k)}} \mathcal{B}_{l_{\sigma(k)}} \ldots \Phi_{s_{\sigma(2)} - s_{\sigma(1)}} \mathcal{B}_{l_{\sigma(1)}} \Phi_{s_{\sigma(1)} - T_0} u_0 1_{s_{\sigma(1)} < \ldots < s_{\sigma(k)} < t}, \tag{2.60}$$

and E_α is defined in (2.56). Next, it was shown in the proof of Theorem 2.3.7, that

$$G(t; s^k; l^k) = \sum_{|\alpha| = k} \frac{1}{\alpha! k!} \left(\int_{[T_0,t^*]^k} G(t; s^k; l^k) E_\alpha(s^k; l^k) ds^k \right) E_\alpha(s^k; l^k). \tag{2.61}$$

Substitution of (2.55) and (2.59) for $\xi_\alpha(W_{T_0,t^*})$ and $\varphi_\alpha(t)$, respectively, into (2.58) results in

$$u(t) = \Phi_{t-T_0}u_0 + \sum_{k\geq 1}\sum_{l^k}\int_{T_0}^{t}\int_{T_0}^{s_k}\cdots\int_{T_0}^{s_2}\sum_{|\alpha|=k}\frac{1}{\alpha!k!}$$
$$\left(\int\limits_{[T_0,t^*]^k} G(t;s^k;l^k)E_\alpha(s^k;l^k)ds^k\right)E_\alpha(s^k;l^k)dW_{l_1}(s_1)\cdots dW_{l_k}(s_k).$$

Together with (2.60) and (2.61), the last equality implies (2.57).

Note that

$$\sum_{|\alpha|=k}\frac{1}{\sqrt{\alpha!}}\varphi_\alpha(t;T_0;u_0)\xi_\alpha(W_{T_0,t^*}) =$$
$$\sum_{l^k}\int_{T_0}^{t}\int_{T_0}^{s_k}\cdots\int_{T_0}^{s_2}F(t;s^k;l^k;u_0)dW_{l_1}(s_1)\cdots dW_{l_k}(s_k),\ k\geq 1. \tag{2.62}$$

□

With the notations

$$\Psi_0(t;T_0;u_0) = \Phi_{t-T_0}u_0;$$
$$\Psi_k(t;T_0;u_0) = \sum_{l^k}\int_{T_0}^{t}\int_{T_0}^{s_k}\cdots\int_{T_0}^{s_2}F(t;s^k;l^k;u_0)dW_{l_1}(s_1)\cdots dW_{l_k}(s_k),\ k\geq 1,$$

representation (2.57) becomes

$$u(t;T_0;u_0) = \sum_{k\geq 0}\Psi_k(t;T_0;u_0). \tag{2.63}$$

If $u_N(t;T_0;u_0)$ is defined by (2.28), then (2.62) implies

$$u_N(t;T_0;u_0) = \sum_{k=0}^{N}\Psi_k(t;T_0;u_0). \tag{2.64}$$

It follows from the definition of Ψ_k that, for $k\geq 1$,

$$\Psi_k(t;T_0;u_0) = \sum_{l=1}^{r}\int_{T_0}^{t}\Phi_{t-s}\mathcal{B}_l\Psi_{k-1}dW_l(s), \tag{2.65}$$

and consequently the collection $\{\Psi_k\}_{k\geq 1}$ is the solution of the recursive system of evolution equations

$$\begin{aligned} d\Psi_0(t;T_0;u_0) &= \mathcal{A}\Psi_0(t;T_0;u_0)dt, \quad \Psi_0(T_0;T_0;u_0) = u_0;\\ d\Psi_k(t;T_0;u_0) &= \mathcal{A}\Psi_k(t;T_0;u_0)dt+\\ &\sum_{l=1}^{r}\mathcal{B}_l\Psi_{k-1}(t;T_0;u_0)dW_l(t), \quad \Psi_k(T_0;T_0;u_0) = 0. \end{aligned} \tag{2.66}$$

This means that the approximation $u_N(t;T_0;u_0)$ becomes computable as long as the solution of (2.66) can be obtained.

As usual, the recursive versions of (2.63) and (2.64) are desirable. Let $T_0 = t_0 < t_1 < \ldots < t_M = T$ be a partition of $[T_0, T]$. The relation $u(t;T_0;u_0) = u(t;t_{i-1};u(t;t_{i-1};u_0))$, $t \in [t_{i-1}, t_i]$, implies that for such t

$$u(t;T_0;u_0) = \sum_{k\geq 0} \Psi_k(t;t_{i-1};u(t_{i-1};T_0;u_0)),$$

and then the recursive approximation $u_N(t)$ is defined as follows:

$$\begin{aligned}&u_N(t_0) = u_0;\\ &u_N(t) = \sum_{k=0}^{N} \Psi_k(t;t_{i-1};u_N(t_{i-1})),\ t \in [t_{i-1};t_i].\end{aligned}$$

2.5.2. Theorem. *Assume that equation (2.54) is either coercive or dissipative and* $\|\Phi_t v\|_0^2 \leq e^{C_1 t}\|v\|_0^2$, $\|\mathcal{B}_l v\|_0^2 \leq C_2\|v\|_0^2$. *If* $\Delta = \max_{1\leq i\leq M}|t_i - t_{i-1}|$ *and* $\bar{C} = C_1 + C_2 r > 0$, *then*

$$\sup_{t\in[T_0,T]} \mathbf{E}\|u(t;T_0;u_0) - u_N(t)\|_0^2 \leq \frac{[C_2 r\Delta]^{N+1}}{\bar{C}(N+1)!} e^{\bar{C}(T-T_0)}\, \mathbf{E}\|u_0\|_0^2\,. \tag{2.67}$$

Proof. The same type of arguments is used as in the proof of (2.46) from Theorem 2.4.3. For simplicity, u_0 and T_0 are omitted wherever possible.

Orthogonality of ξ_a and formula (2.62) imply $\mathbf{E}(\Psi_k, \Psi_l)_0 = 0,\ k \neq l$. Then for $t \in [t_{i-1}, t_i]$

$$\begin{aligned}\mathbf{E}\|u(t) - u_N(t)\|_0^2 = \sum_{k=0}^{N} \mathbf{E}\|\Psi_k(t;t_{i-1};u(t_i) - u_N(t_i))\|_0^2 +\\ \sum_{k>N} \mathbf{E}\|\Psi_k(t;t_{i-1};u(t_{i-1}))\|_0^2 \leq \sum_{k\geq 0} \mathbf{E}\|\Psi_k(t;t_{i-1};u(t_i) - u_N(t_i))\|_0^2 +\\ \sum_{k>N} \mathbf{E}\|\Psi_k(t;t_{i-1};u(t_{i-1}))\|_0^2\,.\end{aligned}$$

By (2.63),

$$\sum_{k\geq 0} \mathbf{E}\|\Psi_k(t;t_{i-1};u(t_i) - u_N(t_i))\|_0^2 = \mathbf{E}\|u(t;t_{i-1};u(t_{i-1}) - u_N(t_{i-1})\|_0^2,$$

and by Corollary 2.3.8,

$$\begin{aligned}\sup_{t\in[t_{i-1},t_i]} \mathbf{E}\|u(t;t_{i-1};u(t_{i-1}) - u_N(t_{i-1}))\|_0^2 \leq e^{C_1\Delta_i}\, \mathbf{E}\|u(t_{i-1}) - u_N(t_{i-1})\|_0^2 \leq\\ e^{C_1\Delta_i} \sup_{t\in[t_{i-2},t_{i-1}]} \mathbf{E}\|u(t;t_{i-2};u(t_{i-2}) - u_N(t))\|_0^2.\end{aligned}$$

Next, by Theorem 2.3.9,

$$\sup_{t\in[t_{i-1},t_i]}\sum_{k>N}\mathbf{E}\|\Psi_k(t;t_{i-1};u(t_{i-1}))\|_0^2 \le \frac{[C_2 r\Delta_i]^{N+1}}{(N+1)!}e^{\bar{C}(T-T_0)}\,\mathbf{E}\|u(t_{i-1})\|_0^2\,,$$

and by Corollary 2.3.8,

$$\mathbf{E}\|u(t_i)\|_0^2 \le e^{\bar{C}(T_0-T)}\mathbf{E}\|u_0\|_0^2.$$

As a result, $\varepsilon_i = \max_{1\le i\le M}\sup_{t\in[t_{i-1},t_i]}\mathbf{E}\|u(t;T_0;u_0)-u_N(t)\|_0^2$ satisfies

$$\varepsilon_i \le e^{\bar{C}\Delta_i}\varepsilon_{i-1} + e^{\bar{C}(T-T_0)}\frac{[C_2 r\Delta_i]^{N+1}}{(N+1)!}\mathbf{E}\|u_0\|_0^2 \le$$
$$e^{\bar{C}\Delta}\varepsilon_{i-1} + e^{\bar{C}(T-T_0)}\frac{[C_2 r\Delta]^{N+1}}{(N+1)!}\mathbf{E}\|u_0\|_0^2.$$

Since $\varepsilon_0 = 0$, the statement of the theorem follows from the discrete Gronwall lemma.

□

From now till the end of the section, all constants will be denoted by C. Approximation (2.64) can be used to construct higher order approximations of $u(t)$, i.e. approximations $\tilde{u}(t)$ for which $\sup_{t\in[T_0,T]}\mathbf{E}\|u(t)-\tilde{u}(t)\|_0^2 \le C\Delta^p$ for some $p>1$. A computable version of such approximations requires a high order approximation to the solution of (2.66). If the objective is an approximation that simplifies the computations involving the Wiener process (which is desirable in many applications), then the natural approach is to use (2.65) and approximate the integral on the right with the prescribed accuracy. In fact, approximation (2.45) is a particular case of this approach, when the integrand in (2.65) is expanded in a Fourier series on $[t_{i-1},t_i]$. The next lemma shows that if the Taylor series is used instead, then it is possible to obtain approximations of arbitrarily high order.

2.5.3. Lemma.

1. Assume that operators $\mathcal{A}$, Φ_t, and $\mathcal{B}$ are linear and bounded from $\mathbf{H}^a$ to $\mathbf{H}^{a-2}$, from $\mathbf{H}^a$ to $\mathbf{H}^a$, and from $\mathbf{H}^a$ to $\mathbf{H}^a$, respectively for all $2\le a\le a_0$, $a_0\ge 1$. If $g\in\mathbf{H}^{2a_0}$, then

$$\sup_{t\in[0,\Delta]}\|\Phi_{t-s}\mathcal{B}\Phi_s g - \Phi_t\mathcal{B}g - \sum_{k=1}^{a_0-1}\frac{t^k}{k!}\Phi_t[\mathcal{B},\mathcal{A}]^{(k)}g\|_0 \le \frac{C\Delta^{a_0}}{a_0!}\,\|g\|_{2a_0}, \tag{2.68}$$

where $[\mathcal{B},\mathcal{A}]^{(1)} = \mathcal{B}\mathcal{A}-\mathcal{A}\mathcal{B}$ and $[\mathcal{B},\mathcal{A}]^{(k)} = [[\mathcal{B},\mathcal{A}]^{(k-1)},\mathcal{A}]^{(1)}$.

2. Assume that $W=W(t)$ is a standard one-dimensional Wiener process and the functions $v_1=v_1(t)$ and $v_2=v_2(t)$ are adapted to the filtration generated by W and belong to $L_2(\Omega;\mathbf{C}([0,\Delta];\mathbf{H}^0))$. If $\sup_{t\in[0,\Delta]}\mathbf{E}\|v_1(t)-v_2(t)\|_0^2\le C\Delta^p$, then

$$\sup_{t\in[0,\Delta]}\mathbf{E}\|\int_0^t(v_1(s)-v_2(s))dW(s)\|_0^2 \le C\Delta^{p+1}. \tag{2.69}$$

Proof. Inequality (2.68) follows from the Taylor formula and the following equality:

$$\frac{d^k}{ds^k}\Phi_{t-s}\mathcal{B}\Phi_s g = \Phi_{t-s}[\mathcal{B},\mathcal{A}]^{(k)}\Phi_s g.$$

Inequality (2.69) is obvious.

□

Note that

$$\Psi_1(t;T_0;u_0) = \sum_{l=1}^{r}\int_{T_0}^{t}\Phi_{t-s}\mathcal{B}_l\Phi_s u_0 dW_l(s). \tag{2.70}$$

Using this observation and the previous lemma, it is possible to construct an approximation of u with the prescribed rate of convergence in time.

2.5.4. Theorem. *Assume that $p \ge 1$ is fixed, equation (2.54) is either coercive or dissipative, the operators $\mathcal{A}$, Φ_t, and $\mathcal{B}_l$ are linear and bounded from $\mathbf{H}^a$ to $\mathbf{H}^{a-2}$, from $\mathbf{H}^a$ to $\mathbf{H}^a$, and from $\mathbf{H}^a$ to $\mathbf{H}^a$, respectively for all $2 \le a \le a_0$, $2a_0 \ge p$, and $u_0 \in L_2(\Omega;\mathbf{H}^{2a_0})$. If $T_0 = t_0 < t_1 < \ldots < t_M = T$ is a partition of $[T_0,T]$ with $\max_{1\le i\le M}|t_i - t_{i-1}| = \Delta$, then there is an approximation $\tilde{u}(t)$ of $u(t;T_0;u_0)$ so that*

$$\sup_{t\in[T_0,T]} \mathbf{E}\|u(t;T_0;u_0) - \tilde{u}(t)\|_0^2 \le C\Delta^p. \tag{2.71}$$

Proof. It follows from the proof of Theorem 2.5.2 that, to achieve (2.71), it is sufficient to consider u_N with $N \ge p$. If $\Delta_i = t_i - t_{i-1}$ and it is assumed that the semigroup $\Phi_{t-t_{i-1}}$ can be computed with the prescribed accuracy, then each of $\Psi_k(t;t_{i-1};g)$ can be approximated by first using (2.68) to approximate (2.70) and then repeatedly applying (2.65) followed by (2.68). If the approximation is denoted by $\bar{\Psi}_k(t;t_{i-1};g)$, then, for sufficiently regular g,

$$\mathbf{E}\|\bar{\Psi}_k(t;t_{i-1};g)\|_0^2 \le \frac{(C\Delta_i)^k}{k!}\|g\|_a^2,$$

so that (2.69) implies that the approximation of the integral in (2.65) by (2.68) requires $2a_0 + k \ge p+1$ or $a_0 \le (p+1-k)/2$ meaning that the larger the k, the lower order approximation of the integral is needed. To have $\bar{\Psi}_k(t;t_{i-1};g)$ preserving the $\mathbf{H}^0$-norm, the operator $\mathcal{A}$ can be approximated (if necessary) by Φ_{Δ_i} from the relation

$$\mathcal{A} \approx \sum_{n=1}^{n_0}(-1)^n\frac{(\Phi_{\Delta_i} - I)^n}{n\Delta_i},$$

where I is the identity operator. If $\tilde{\Psi}_k(t;t_{i-1};g)$ is the resulting approximation of $\Psi_k(t;t_{i-1};g)$, then the approximation $\tilde{u}$ is given by

$$\tilde{u}(t_0) = u_0,$$
$$\tilde{u}(t_i) = \Phi_{t-t_{i-1}}\tilde{u}(t_{i-1}) + \sum_{k=1}^{N}\tilde{\Psi}_k(t_i;t_{i-1};\tilde{u}(t_{i-1})),\ t\in[t_{i-1},t_i].$$

The time integrals in $\tilde{\Psi}_k$ involve only the Wiener processes $W_l(t)$ and the integer powers of t. Some methods of computation of these integrals are discussed in [35, 36, 58]. The exact formulas that produce the approximation $\tilde{u}$ are complicated and hardly add anything to the subject of the discussion.

□

In reality, the approximation of the semigroup Φ_t can be computed only at finitely many time moments so that $\tilde{u}$ provides an approximation of u at the points t_i of the partition. The following is an example of such an approximation when the partition is uniform. Even though the proof of Theorem 2.5.4 describes the general procedure of constructing an approximation with the given rate of convergence in time, the actual computations must be performed for each problem separately.

Consider the equation

$$u(t;T_0;u_0) = u_0 + \int_{T_0}^{t} \mathcal{A}u(s;T_0;u_0)ds + \int_{T_0}^{t} \mathcal{B}u(s;T_0;u_0)dW(s),\ \ t \in [T_0,T],$$

which corresponds to (2.54) with $r=1$. Assume that conditions of Theorem 2.5.4 are fulfilled with $p=3$, $a_0=2$. Define $\tilde{u}(t_i)$ by

$$\begin{aligned}
\tilde{u}(t_0) &= u_0;\\
\tilde{u}(t_i) &= \Phi_\Delta \tilde{u}(t_{i-1}) + \Phi_\Delta \mathcal{B}\tilde{u}(t_{i-1})(W(t_i)-W(t_{i-1}))+\\
&(\mathcal{B}\Phi_\Delta - \Phi_\Delta\mathcal{B})\tilde{u}(t_{i-1})\Big(W(t_i) - \frac{1}{\Delta}\int_{t_{i-1}}^{t_i} W(t)dt\Big)+\\
&\frac{1}{2}\Phi_\Delta\mathcal{B}^2\tilde{u}(t_{i-1})\Big((W(t_i)-W(t_{i-1}))^2-\Delta\Big)+\\
\frac{1}{2}&\Phi_\Delta\mathcal{B}^3\tilde{u}(t_{i-1})\Big((W(t_i)-W(t_{i-1}))^3/3-(W(t_i)-W(t_{i-1}))\ \Delta\Big).
\end{aligned}$$

Direct computations show that

$$\max_{1\leq i\leq M} \mathbf{E}\|u(t_i;T_0;u_0)-\tilde{u}(t_i)\|_0^2 \leq C\Delta^3.$$

This example shows that, even for simple equations, the higher order approximation schemes can result in complicated expressions, which can lead to numerical instability.

Chapter 3

H-orthogonal Projections of Solutions of Stochastic Evolution Equations

3.1 Introduction

The results presented in this chapter describe the standard time-space separation of variables technique for partial differential equations, applied to the stochastic evolution equations. The idea of this separation of variables is to represent the solution of the equation as a Fourier series with respect to an orthonormal basis in the underlying Hilbert space. The orthonormal basis usually consists of the eigenfunctions of some operator that is not necessarily the operator from the equation. The truncated version of the expansion can be used to construct a computable approximation of the solution as far as the space discretization is concerned.

In **Section 3.2**, the general construction of an orthonormal basis in a Hilbert space is described and examples are given to illustrate the presentation.

In **Section 3.3**, the evolution equation is reduced to a discrete time recursive relation which is then expanded with respect to an orthonormal basis in the corresponding Hilbert space. The result is a recursive relation for the coefficients of the expansion. A truncated version of the expansion is considered and the error of the approximation is derived. The term *H-orthogonal projection of the first kind* is used to distinguish the result from the Galerkin approximation.

In **Section 3.4**, the standard Galerkin approximation of the solution of the evolution equation is considered and the error of the approximation is derived.

3.2 Orthonormal Bases in Hilbert Spaces

Consider the stochastic evolution system

$$u(t) = u_0 + \int_{T_0}^{t} \mathcal{A}u(s)ds + \int_{T_0}^{t} \sum_{l=1}^{r} \mathcal{B}_l u(s)dW_l(s). \tag{3.1}$$

The notations and main assumptions are the same as in Chapter 2. In particular, $W = W(t)$ is an r - dimensional standard Wiener process on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$, u_0 is independent of $\mathcal{F}^W_{T_0,T}$, and $\mathcal{A}$ and $\mathcal{B}_l$, $l = 1, \ldots, r$, are linear operators on a real separable Hilbert space $\mathbf{H}$. The corresponding Hilbert scale $\{\mathbf{H}^a\}_{a\in\mathbf{R}}$ with $\mathbf{H}^0 = \mathbf{H}$ is also given. Recall that the inner product and the norm in $\mathbf{H}^a$ are denoted by $(\cdot,\cdot)_a$ and $\|\cdot\|_a$, respectively. It will be assumed that equation (3.1) is either coercive or dissipative so that there is a unique solution $u = u(t; T_0; u_0)$ in the space $L_2(\Omega \times [T_0, T]; \mathbf{H}^1) \cap L_2(\Omega; \mathbf{C}([T_0, T]; \mathbf{H}^0))$.

In Chapter 2, various representations of the solution of (3.1) and various approximations of that solution were considered under the assumption that the semigroup $\Phi = \{\Phi_t\}_{t\geq 0}$ generated by the operator $\mathcal{A}$ can be computed exactly. However, in many problems where equation of the type (3.1) must be solved, including the nonlinear filtering, the operator $\mathcal{A}$ is a partial differential operator so that the explicit formula for Φ_t is available only in a very few examples. As a result, finite dimensional approximations of Φ_t and of the solution u are necessary.

Assume that there is an orthonormal basis $\{e_k\}_{k\geq 0}$ in the space $\mathbf{H}$ so that each e_k is an eigenfunction of a linear operator Λ:

$$\Lambda e_k = \lambda_k e_k,$$

and $\lambda_k \asymp k^\sigma$ for some $\sigma > 0$, i.e.

$$0 < \liminf_{k\to\infty} \frac{\lambda_k}{k^\sigma} \leq \limsup_{k\to\infty} \frac{\lambda_k}{k^\sigma} < \infty.$$

This in particular implies that Λ is an unbounded operator on $\mathbf{H}$. The adjoint of Λ will be denoted by Λ^*.

In the following examples of the bases $\{e_k\}_{k\geq 1}$, Γ is the set of multi-indices $\gamma = \{\gamma_1, \ldots, \gamma_d\}$ with γ_i non-negative integer. The ordering of the set Γ is defined by $\gamma < \tau$ if $\gamma_1 + \cdots + \gamma_d < \tau_1 + \cdots + \tau_d$ or if $\gamma_1 + \cdots + \gamma_d = \tau_1 + \cdots + \tau_d$ and $\gamma_i = \tau_i,\ i < j,\ \gamma_j < \tau_j$ for some $j \geq 1$.

Example 1.

For $\gamma \in \Gamma$ and $x \in \mathbf{R}^d$ define

$$e_\gamma(x) = \prod_{i=1}^{d} e_{\gamma_i}(x_i),$$

where

$$e_n(t) = \frac{(-1)^n}{\sqrt{2^n \pi^{1/2} n!}} e^{t^2/2} \frac{d^n}{dt^n} e^{-t^2}.$$

Then [22, 24] the collection $\{e_\gamma\}_{\gamma\in\Gamma}$ is an orthonormal basis, called the **Hermite basis** in the space $\mathbf{H} = L_2(\mathbf{R}^d)$. If $\Lambda = -\nabla^2 + (1 + |x|^2)$, where ∇^2 is the Laplace operator in $\mathbf{R}^d$, and $|x|^2 = x_1^2 + \cdots + x_d^2$, then

$$\Lambda e_\gamma = (2(\gamma_1 + \cdots + \gamma_d) + d + 1)e_\gamma.$$

With the above ordering of the set Γ, the eigenvalue λ_k satisfies $\lambda_k \asymp k^{1/d}$. The approximation of the Zakai filtering equation using the Hermite basis is studied in Chapter 4.

□

Example 2.

Let $T_n(t) = \cos(n \arccos t)$, $|t| \le 1$, be the nth Chebychev polynomial. For $\gamma \in \Gamma$ and $x \in [-1,1]^d$ define

$$e_\gamma(x) = \prod_{i=1}^{d} c_{\gamma_i} T_{\gamma_i}(x_i),$$

where $c_{\gamma_i} = \sqrt{\pi}$ if $\gamma_i = 0$ and $c_{\gamma_i} = \sqrt{\pi/2}$ if $\gamma_i > 0$. Then [22] the collection $\{e_\gamma\}_{\gamma\in\Gamma}$ is an orthonormal basis in the space $\mathbf{H} = L_2([-1,1]^d, d\mu)$ with the measure $d\mu(x) = \prod_{i=1}^{d}(1 - x_i^2)^{-1/2}dx_i$. If Λ is defined by

$$\Lambda = \sum_{i=1}^{d} \left((1 - x_i^2)\partial^2/\partial x_i^2 - x_i \partial/\partial x_i \right),$$

then

$$\Lambda e_\gamma = \left(\sum_{i=1}^{d} \gamma_i^2 \right) e_\gamma,$$

and, with the above ordering of the set Γ, the corresponding eigenvalue λ_k satisfies $\lambda_k \asymp k^{2/d}$. The domains of the operators Λ and Λ^* include $\mathbf{C}^2([-1,1]^d)$.

□

Example 3.

This is a generalization of the previous two examples. Consider the Hilbert space $\mathbf{H} = L_2(G, d\mu)$, where $G = G_1 \times \ldots \times G_d$ with $G_i \subseteq \mathbf{R}^1$, and $d\mu(x) = d\mu_1(x_1) \cdots d\mu_d(x_d)$. Assume that $\{e_k^i\}_{k\ge0}$ is an orthonormal basis in $L_2(G_i, d\mu_i)$ so that $\Lambda^i e_k^i = \lambda_k^i e_k^i$ for some linear operator Λ^i and $\lambda_k^i \asymp k^\sigma$ for all i. Define $\Lambda = \Lambda^1 + \cdots + \Lambda^d$ and $e_\gamma(x) = e_{\gamma_1}^1(x_1) \cdots e_{\gamma_d}^d(x_d)$ for $\gamma \in \Gamma$. Then the collection $\{e_\gamma\}_{\gamma\in\Gamma}$ is an orthonormal basis in $\mathbf{H}$, $\Lambda e_\gamma = (\lambda_{\gamma_1}^1 + \cdots + \lambda_{\gamma_d}^d)e_\gamma$, and with, the above ordering of the set Γ, $\lambda_k \asymp k^{\sigma/d}$.

□

In the next two sections, the solution of (3.1) is approximated by a finite linear combination of the first K basis functions $e_1, \ldots, e_K$ and the quality of this approximation is studied. As it might be expected, the rate of convergence (as $K \to \infty$) of such an approximation depends on the regularity of u relative to Λ^*, i.e. the larger the number n for which u is in the domain of $(\Lambda^*)^n$, the faster the convergence.

3.3 H-orthogonal Projection of the First Kind

Let $T_0 = t_0 < t_1 < \ldots < t_M = T$ be a partition of $[T_0, T]$ and $\Delta_i = t_i - t_{i-1}$. The solution of (3.1) satisfies $u(t_i; T_0; u_0) = u(t_i; t_{i-1}; u(t_{i-1}; T_0; u_0))$.

Let $\{e_k\}_{k\ge0}$ be an orthonormal basis in the space $\mathbf{H}$ and define

$$\psi_k(t_i) = (u(t_i; T_0; u_0), e_k)_0.$$

Then $u(t_i;T_0;u_0)=\sum_{k\geq 0}\psi_k(t_i)e_k$, and the properties of the solution of (3.1) imply that

$$\sum_{k\geq 0}\psi_k(t_i)e_k=\sum_{k\geq 0}u(t_i;t_{i-1};e_k)\psi_k(t_{i-1})$$

or

$$\psi_k(t_i)=\sum_{l\geq 0}(u(t_i;t_{i-1};e_l),e_k)_0\psi_l(t_{i-1}).$$

For a fixed $K>0$ denote by Π^K the projection operator on the span of $e_0,\ldots,e_K$. Define $U^K(t_i)$ by

$$U^K(t_0)=\Pi^K u_0,$$
$$U^K(t_i)=\Pi^K u(t_i;t_{i-1};U^K(t_{i-1})).$$

The sequence $\{U^K(t_i)\}_{0\leq i\leq M}$ will be referred to as the **H-orthogonal projection of the first kind** of the solution u. It follows from the definition that

$$U^K(t_i)=\sum_{k=0}^{K}\psi_k^K(t_i)e_k,$$

where

$$\psi_k^K(t_i)=\sum_{l=0}^{K}(u(t_i;t_{i-1};e_l),e_k)_0\psi_l^K(t_{i-1}).$$

Introduce the following notation: for $v\in L_2(\Omega;\mathbf{H})$, $|||v|||:=\sqrt{\mathbf{E}\|v\|_0^2}$. Recall that under the assumption of coercivity or dissipativity

$$|||u(t;T_0;u_0)|||\leq e^{C(t-T_0)}|||u_0||| \tag{3.2}$$

for some constant C.

To study the approximation $U^K(t_i)$, the following result is used.

3.3.1. Lemma. *The map $v\mapsto |||v|||$ defines a norm in $L_2(\Omega;\mathbf{H})$.*
Proof. It suffices to verify the triangle inequality. For $v,w\in L_2(\Omega;\mathbf{H})$,

$$\begin{aligned}|||v+w|||^2 &= \mathbf{E}\left(\|v\|_0^2+\|w\|_0^2+2(v,w)_0\right)\leq\\ &\mathbf{E}\left(\|v\|_0^2+\|w\|_0^2+2\|v\|_0\|w\|_0\right)\leq\\ &|||v|||^2+|||w|||^2+2|||v|||\cdot|||w|||=\left(|||v|||+|||w|||\right)^2,\end{aligned}$$

and the result follows.

□

3.3.2. Theorem. *Assume that equation (3.1) is either coercive or dissipative and the basis $\{e_k\}$ consists of the eigenfunctions of a linear operator Λ with the corresponding eigenvalues λ_k satisfying $\lambda_k\asymp k^\sigma$, $\sigma>0$. If $p\geq 1$ is an integer such that $p\sigma>1/2$, and $u(t_i;T_0;u_0)$ is in the domain of $(\Lambda^*)^p$ so that $|||(\Lambda^*)^p u(t_i;T_0;u_0)|||<\infty$ for all $i=0,\ldots,M$, then*

$$\max_{0\leq i\leq M}\mathbf{E}\|u(t_i;T_0;u_0)-U^K(t_i)\|_0^2\leq\max_{0\leq i\leq M}\mathbf{E}\|(\Lambda^*)^p u(t_i;T_0;u_0)\|_0^2\,\frac{C\,e^{C(T-T_0)}}{K^{2p\sigma-1}\,\Delta^2}\,, \tag{3.3}$$

where $\Delta = \max_{0\le i\le M}\Delta_i$ *and* $C>0$ *is a constant depending only on* p,σ, *and the constant from (3.2).*

Proof. By the definition of U^K and inequality (3.2),

$$\begin{aligned}&|||u(t_i;T_0;u_0)-U^K(t_i)||| \le |||\Pi^K u\left(t_i;t_{i-1};u(t_{i-1};T_0;u_0)-U^K(t_{i-1})\right)|||+\\&|||(I-\Pi^K)u(t_i;T_0;u_0)||| \le |||u\left(t_i;t_{i-1};u(t_{i-1};T_0;u_0)-U^K(t_{i-1})\right)|||+\\&|||(I-\Pi^K)u(t_i;T_0;u_0)||| \le e^{C\Delta_i}\,|||u(t_{i-1};T_0;u_0)-U^K(t_{i-1})|||+\\&|||(I-\Pi^K)u(t_i;T_0;u_0)|||.\end{aligned} \tag{3.4}$$

Next, using the assumptions of the theorem,

$$\begin{aligned}|\psi_k(t_i)| &= |(u(t_i;T_0;u_0),e_k)_0| = \frac{1}{\lambda_k}|(\Lambda^* u(t_i;T_0;u_0),e_k)_0| = \dots\\ &= \frac{1}{\lambda_k^p}|((\Lambda^*)^p u(t_i;T_0;u_0),e_k)_0| \le \frac{\|(\Lambda^*)^p u(t_i;T_0;u_0)\|_0}{\lambda_k^p},\end{aligned} \tag{3.5}$$

and consequently

$$\begin{aligned}|||(I-\Pi^K)u(t_i;T_0;u_0)||| = &\left(\sum_{k>K}\mathbf{E}\psi_k^2(t_i)\right)^{1/2} \le |||(\Lambda^*)^p u(t_i;T_0;u_0)|||\left(\sum_{k>K}\frac{1}{\lambda_k^{2p}}\right)^{1/2} \le\\ &\max_{0\le i\le M}|||(\Lambda^*)^p u(t_i;T_0;u_0)|||\frac{C}{K^{p\sigma-1/2}}.\end{aligned} \tag{3.6}$$

As a result, if $\varepsilon_i = |||u(t_i;T_0;u_0)-U^K(t_i)|||$, then (3.4) and (3.6) imply

$$\begin{aligned}\varepsilon_i &\le e^{C\Delta_i}\,\varepsilon_{i-1} + \max_{0\le i\le M}|||(\Lambda^*)^p u(t_i;T_0;u_0)|||\frac{C}{K^{p\sigma-1/2}} \le e^{C\Delta}\,\varepsilon_{i-1}+\\ &\max_{0\le i\le M}|||(\Lambda^*)^p u(t_i;T_0;u_0)|||\frac{C}{K^{p\sigma-1/2}},\end{aligned}$$

and (3.3) follows from the discrete Gronwall lemma.

□

3.3.3. Corollary. *If* $\max_{0\le i\le M}\mathbf{E}\|(\Lambda^*)^p u(t_i;T_0;u_0)\|_0^2<\infty$ *for all* $p\ge 1$, *then for every* $\nu>0$ *there is a constant* $C(\nu)$ *so that*

$$\mathbf{E}\|u(t_i;T_0;u_0)-U^K(t_i)\|_0^2 \le \frac{e^{C(T-T_0)}\,C(\nu)}{K^\nu\,\Delta^2}. \tag{3.7}$$

3.3.4. Remark. The proof of Theorem 3.3.2 shows that the following more general result is true.

Assume that $\{v(t_i)\}_{0\le i\le M}$ is a sequence of elements from $L_2(\Omega;\mathbf{H})$ with the properties:

1. $v(t_i)=F_i(v(t_{i-1}))$, where F_i is a linear bounded operator on $\mathbf{H}$ and $\mathbf{E}\|F_i(v(t_i))\|_0^2 \le e^{C\Delta_i}\,\mathbf{E}\|v(t_{i-1})\|_0^2$;

2. $\max_{0\le i\le M}\mathbf{E}\|(\Lambda^*)^p v(t_i)\|_0^2<\infty$ for every positive integer p.

If $V^K(t_i)$ is defined by

$$\begin{aligned}V^K(t_0) &= \Pi^K v(t_0),\\ V^K(t_i) &= \Pi^K F_i(V^K(t_{i-1})),\end{aligned}$$

then for every $\nu > 0$ there is a constant $C(\nu)$ so that

$$\mathbf{E}\|v(t_i) - V^K(t_i)\|_0^2 \leq \frac{e^{C(T-T_0)}\, C(\nu)}{K^\nu\, \Delta^2}. \tag{3.8}$$

This result will be used below in Section 4.3 to study the approximation error for the Zakai equation.

3.4 H-orthogonal Projection of the Second Kind (Galerkin Approximation)

The Galerkin approximation was used in [61] to prove the existence and the regularity of the solution of (3.1). Now that the properties of the solution are known, it is possible to use the approximation for computational purposes. In [20], the Galerkin approximation was used to study the filtering equation in a bounded domain. In what follows, the approximation error of the Galerkin approximation will be studied for (3.1).

Assume that the elements of the orthonormal basis $\{e_k\}$ in $\mathbf{H}$ belong to $\mathbf{H}^1$. Consider the following system of stochastic ordinary differential equations:

$$\begin{aligned} du_k^K(t) = &\sum_{n=0}^{K}\langle \mathcal{A}e_n, e_k\rangle_{0,1} u_n^K(t)dt + \\ &\sum_{l=1}^{r}\sum_{n=0}^{K}(\mathcal{B}_l e_n, e_k)_0 u_n^K(t)dW_l(t),\ T_0 < t \leq T, \\ u_k^K(T_0) = &(u_0, e_k)_0,\ k = 0, \ldots, K. \end{aligned} \tag{3.9}$$

Then the function

$$u^K(t) = \sum_{k=1}^{K} u_k^K(t) e_k$$

is called the **Galerkin approximation** or the **H-orthogonal projection of the second kind** of the solution of (3.1). It is proved in the following theorem that

$$\lim_{K\to\infty} \sup_{T_0 \leq t \leq T} \mathbf{E}\|u(t;T_0;u_0) - u^k(t)\|_0^2 = 0,$$

and the rate of convergence is determined.

3.4.1. Theorem. *Let the following conditions be fulfilled:*

1. *Equation (3.1) is either coercive or dissipative;*
2. *The basis $\{e_k\}$ consists of the eigenfunctions of a linear operator Λ with the corresponding eigenvalues λ_k satisfying $\lambda_k \asymp k^\sigma$, $\sigma > 0$;*
3. *$e_k \in \mathbf{H}^1$ and $\|e_k\|_1 \leq Ck^q$, $q \geq 0$;*
4. *$\sup_{T_0 \leq t \leq T} \mathbf{E}\|(\Lambda^*)^p u(t;T_0;u_0)\|_0^2 < \infty$ for some positive integer p so that $\sigma_1 := p\sigma - 2q >$ 1.*

Then

$$\sup_{T_0\le t\le T_0}\mathbf{E}\|u(t;T_0;u_0)-u^K(t)\|_0^2\le \sup_{T_0\le t\le T}\mathbf{E}\|(\Lambda^*)^p u(t;T_0;u_0)\|_0^2\,\frac{C(r+1)e^{C(T-T_0)}}{K^{2(\sigma_1-1)}}, \quad (3.10)$$

where C is a constant depending only on p, σ, q, and the operators $\mathcal{A}$ and $\mathcal{B}_l$, $l=1,\ldots,r$.

Proof. To simplify the notations, the arguments T_0 and u_0 will be omitted wherever possible.

If $\psi_k(t):=(u(t),e_\gamma)_0$, then

$$\mathbf{E}\|u(t)-u^K(t)\|_0^2=\sum_{k=0}^{K}\mathbf{E}|\psi_k(t)-u_k^K(t)|^2+\sum_{k>K}\mathbf{E}|\psi_k(t)|^2. \quad (3.11)$$

By (3.5) and assumption 4 of the theorem,

$$|\psi_k(t)|\le\frac{\|(\Lambda^*)^p u(t)\|_0}{\lambda_k^p} \quad (3.12)$$

so that

$$\begin{aligned}\sup_{T_0\le t\le T}\sum_{k>K}\mathbf{E}|\psi_k(t)|^2 \ &\le \sup_{T_0\le t\le T}\mathbf{E}\|(\Lambda^*)^p u(t)\|_0^2\frac{C}{K^{2p\sigma-1}}\le\\ &\sup_{T_0\le t\le T}\mathbf{E}\|(\Lambda^*)^p u(t)\|_0^2\frac{C}{K^{2(\sigma_1-1)}}.\end{aligned} \quad (3.13)$$

For $0\le k\le K$ define $\delta_k(t):=\psi_k(t)-u_k^K(t)$, so that $\sum_{k=0}^K\mathbf{E}|\psi_k(t)-u_k^K(t)|^2=\sum_{k=0}^K\mathbf{E}|\delta_k|^2$, and also define

$$\delta_{1,n}(t):=\sum_{k>K}\langle\mathcal{A}e_k,e_n\rangle_{0,1}\psi_k(t),\quad \delta_{2,n}^l(t):=\sum_{k>K}(\mathcal{B}_le_k,e_n)_0\psi_k(t).$$

Both $\delta_{1,n}(t)$ and $\delta_{2,n}^l(t)$ are well defined due to (3.12) and the assumptions 3 and 4 of the theorem. Then

$$\begin{aligned}d\delta_n(t)=\ &\sum_{k=0}^{K}\langle\mathcal{A}e_k,e_n\rangle_{0,1}\delta_k(t)dt+\sum_{l=1}^{r}\sum_{k=0}^{K}(\mathcal{B}_le_k,e_n)_0\delta_k(t)dW_l(t)+\\ &\delta_{1,n}(t)dt+\sum_{l=1}^{r}\delta_{2,n}^l dW_l(t),\ T_0<t\le T;\\ \delta_n(T_0)=\ &0,\ 0\le n\le K,\end{aligned} \quad (3.14)$$

and by the Ito formula,

$$\begin{aligned}&\sum_{n=0}^{K}\mathbf{E}|\delta_n(t)|^2=2\int_{T_0}^{t}\sum_{n,k=0}^{K}\langle\mathcal{A}e_k,e_n\rangle_{0,1}\mathbf{E}\delta_n(s)\delta_k(s)ds+\\ &\sum_{l=1}^{r}\sum_{n=0}^{K}\int_{T_0}^{t}\mathbf{E}\left(\sum_{k=0}^{K}(\mathcal{B}_le_k,e_n)_0\delta_k(s)\right)^2ds+2\sum_{n=0}^{K}\int_{T_0}^{t}\mathbf{E}\delta_{1,n}(s)\delta_n(s)ds+\\ &2\sum_{l=1}^{r}\sum_{n,k=0}^{K}\int_{T_0}^{t}(\mathcal{B}_le_k,e_n)_0\mathbf{E}\delta_{2,n}^l(s)\delta_k(s)ds+\sum_{l=1}^{r}\sum_{n=0}^{K}\int_{T_0}^{t}\mathbf{E}(\delta_{2,n}^l(s))^2ds.\end{aligned} \quad (3.15)$$

It follows from assumption 1 of the theorem that

$$2\int_{T_0}^{t}\sum_{n,k=0}^{K}\langle\mathcal{A}e_k,e_n\rangle_{0,1}\mathbf{E}\delta_n(s)\delta_k(s)ds+ \sum_{l=1}^{r}\sum_{n=0}^{K}\int_{T_0}^{t}\mathbf{E}\left(\sum_{k=0}^{K}(\mathcal{B}_le_k,e_n)_0\delta_k(s)\right)^2ds\le C\sum_{k=0}^{K}\int_0^t\mathbf{E}(\delta_k(s))^2ds \tag{3.16}$$

(by coersivity or dissipativity inequlaity) and also, using assumptions 2 and 3,

$$|\langle\mathcal{A}e_k,e_n\rangle_{0,1}|\le C\|e_k\|_1\|e_n\|_1\le Ck^qn^q,\quad |(\mathcal{B}_le_k,e_n)_0|\le C\|e_k\|_1\|e_n\|_0\le Ck^q,$$

so that by (3.12),

$$|\delta_{1,n}(t)|\le Cn^q\frac{\|(\Lambda^*)^pu(t)\|_0}{K^{p\sigma-q-1}},\quad |\delta_{2,n}^l|\le C\frac{\|(\Lambda^*)^pu(t)\|_0}{K^{p\sigma-q-1}}.$$

After that,

$$\sum_{n=0}^{K}\int_{T_0}^{T}\mathbf{E}(\delta_{1,n}(s))^2ds+\sum_{l=1}^{r}\sum_{n=0}^{K}\int_{T_0}^{T}\mathbf{E}(\delta_{2,n}^l(s))^2ds\le (T-T_0)\sup_{T_0\le t\le T}\mathbf{E}\|(\Lambda^*)^pu(t)\|_0^2\,\frac{(r+1)C}{K^{2(\sigma_1-1)}}, \tag{3.17}$$

and (3.15)–(3.17) and the obvious inequality $2|ab|\le a^2+b^2$ imply

$$\sum_{n=0}^{K}\mathbf{E}|\delta_n(t)|^2\le C\sum_{n=0}^{K}\int_{T_0}^{t}\mathbf{E}|\delta_n(s)|^2ds+(T-T_0)\sup_{T_0\le t\le T}\mathbf{E}\|(\Lambda^*)^pu(t)\|_0^2\,\frac{(r+1)C}{K^{2(\sigma_1-1)}}$$

so that by the Gronwall inequality

$$\sup_{T_0\le t\le T}\sum_{n=0}^{K}\mathbf{E}|\delta_n(t)|^2\le(T-T_0)\sup_{T_0\le t\le T}\mathbf{E}\|(\Lambda^*)^pu(t)\|_0^2\,e^{C(T-T_0)}\,\frac{(r+1)C}{K^{2(\sigma_1-1)}}.$$

Together with (3.11) and (3.13), the last inequality implies (3.10).

□

3.4.2. Corollary. *If $\sup_{T_0\le t\le T}\mathbf{E}\|(\Lambda^*)^pu(t;T_0;u_0)\|_0^2<\infty$ for all positive integers p, then for every $\nu>0$ there is a constant $C(\nu)$ so that*

$$\sup_{T_0\le t\le T_0}\mathbf{E}\|u(t;T_0;u_0)-u^K(t)\|_0^2\le\frac{e^{C(T-T_0)}\,C(\nu)}{K^\nu}.$$

Chapter 4

Nonlinear Filtering of Diffusion Processes: Spectral Separating Schemes for the Zakai Equation

4.1 Introduction

In a typical filtering model, a non-anticipative functional $f_t(X)$ of the unobserved signal process $X = (X(t))_{t\geq 0}$ is estimated from the observations $Y(s),\ s \leq t$. The best mean square estimate is known to be the conditional expectation
$\mathbf{E}[f_t(X)|Y(s),\ s \leq t]$, called the **optimal filter**. When the observation noise is additive, the Kallianpur-Striebel formula (Kallianpur [33], Liptser and Shiryayev [47]) provides the representation of the optimal filter as follows:

$$\mathbf{E}[f_t(X)|Y(s),\ s \leq t] = \frac{\phi_t[f]}{\phi_t[1]},$$

where $\phi_t[\cdot]$ is a functional called **the unnormalized optimal filter**. In the particular case $f_t(X) = f(X(t))$, there are two approaches to computing $\phi_t[f]$.

In the first approach (Lo and Ng [48], Mikulevicius and Rozovskii [55], Ocone [59]), the functional $\phi_t[f]$ is expanded in a series of multiple integrals with respect to the observation process. This approach can be used to obtain representations of general functionals, but these representations are not recursive in time. In fact, there is no closed form differential equation satisfied by $\phi_t[f]$.

In the second approach (Kallianpur [33], Liptser and Shiryayev [47], Rozovskii [61]), it is proved that, under certain regularity assumptions, the functional $\phi_t[f]$ can be written as

$$\phi_t[f] = \int f(x)p(t,x)dx \tag{4.1}$$

for some function $p(t,x)$, called **the unnormalized filtering density** (UFD). Even though the computation of $u(t,x)$ can be organized recursively in time, and there are many numerical algorithms to do this (Budhiraja and Kallianpur [9], Elliott and Glowinski [17], Florchinger and LeGland [18], Ito [31], Lototsky et al. [50], etc.), these algorithms are time consuming because they involve evaluation of $p(t,x)$ at many spatial points. Moreover, computation of $\phi_t[f]$ using this approach requires subsequent evaluation of the integral (4.1).

In this chapter, two algorithms are constructed to compute approximations of both the UFD $p(t,x)$ and the unnormalized optimal filter $\phi_t[f]$ recursively in time and, in some sense, independently of each other (meaning that the computations of $\phi_t[f]$ do not require the evaluation of $p(t,x)$ at the points of the spatial domain). The basic assumption is that both the state and the observation processes are stationary diffusions with smooth bounded coefficients so that the unnormalized conditional density satisfies a certain stochastic partial differential equation. The main results concerning this filtering model are summarized in **Section 4.2**.

One algorithm, called the spectral separating scheme of the first kind, is described in **Section 4.3**. In this algorithm, the solution of the original equation satisfied by $p(t,x)$ is approximated using the (Cameron-Martin version of the) Wiener chaos decomposition. The result is then approximated further using the H-orthogonal projection of the first kind. The error of the approximation is computed for both the UFD and the unnormalized optimal filter.

The other algorithm, called the spectral separating scheme of the second kind, is described in **Section 4.4**. In this algorithm, the equation for $p(t,x)$ is first reduced to a system of stochastic ordinary differential equations using the Galerkin approximation and after that the solution of the system is approximated using the Wiener chaos decomposition. The error of the approximation is computed for both the UFD and the unnormalized optimal filter.

The computational aspects of the algorithms are studied in **Section 4.5**, where both spectral separating schemes are also compared with the splitting-up approximation — an existing scheme for computing an approximation of the UFD.

4.2 The Diffusion Filtering Model

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a complete probability space with independent standard Wiener processes $W = W(t)$ and $V = V(t)$ of dimensions d_1 and r respectively. In the **diffusion filtering model**, the unobserved d - dimensional state (or signal) process $X = X(t)$ and the r-dimensional observation process $Y = Y(t)$ are defined by the stochastic ordinary differential equations

$$\begin{aligned} dX(t) &= b(X(t))dt + \sigma(X(t))dW(t) + \rho(X(t))dV(t), \\ dY(t) &= h(X(t))dt + dV(t),\ 0 < t \le T; \\ X(0) &= X_0, \quad Y(0) = 0. \end{aligned} \tag{4.2}$$

If $f = f(x)$ is a scalar measurable function on $\mathbf{R}^d$ so that $\sup_{0 \le t \le T} \mathbf{E}|f(X(t))|^2 < \infty$, then the **filtering problem** for (4.2) is to find the best mean square estimate $\hat{f}_t$ of $f(X(t)),\ t \le T$, given the observations $Y(s),\ 0 < s \le t$.

4.2.1. Definition. The **parameters** of the filtering model (4.2) are the functions $b,\ \sigma,\ \rho,\ h,\ f$, the random variable X_0, the dimensions d and r of the state and the observation processes, and the length T of the time interval.

4.2.2. Definition. The filtering model (4.2) is called **regular** if the following conditions are fulfilled:

(R1) The functions $b = b(x) \in \mathbf{R}^d$, $\sigma = \sigma(x) \in \mathbf{R}^{d\times d_1}$, $\rho = \rho(x) \in \mathbf{R}^{d\times r}$, and $h = h(x) \in \mathbf{R}^r$ are infinitely differentiable and bounded with all the derivatives;

(R2) The random variable X_0 and the Wiener processes W and V are independent of one another;

(R3) The random variable X_0 has a density $p_0(x)$, $x \in \mathbf{R}^d$, so that the function $p_0 = p_0(x)$ is infinitely differentiable and decays at infinity with all the derivatives faster than any power of $|x|$;

(R4) The function $f = f(x)$ is of polynomial growth, i.e. there exist positive numbers L and k_0 so that
$$|f(x)| \leq L(1 + |x|^{k_0}) \quad \text{for all } x \in \mathbf{R}^d.$$

4.2.3. Lemma. *If the filtering model is regular, then*

$$\sup_{0 \leq t \leq T} \mathbf{E}|f(X(t))|^n < \infty$$

for every positive integer n.

Proof. This follows from Theorem 4.6 in [47].

□

Throughout the rest of the chapter, the filtering model (4.2) will be considered, and the model is assumed to be regular in the sense of Definition 4.2.2. All constants will be denoted by C and are either absolute constants or depend only on the parameters of the model. The value of C can be different in different places and dependence of C on other parameters will be explicitly indicated.

Denote by $\mathcal{F}_t^Y$ the σ-algebra generated by $Y(s)$, $0 \leq s \leq t$. Then the properties of the conditional expectation imply that the solution of the filtering problem is

$$\hat{f}_t = \mathbf{E}\left(f(X(t))|\mathcal{F}_t^Y\right).$$

To derive an alternative representation of $\hat{f}_t$, some additional constructions will be necessary.

Define a new probability measure $\tilde{\mathbf{P}}$ on $(\Omega, \mathcal{F})$ as follows: for $A \in \mathcal{F}$,

$$\tilde{\mathbf{P}}(A) = \int_A Z_T^{-1} d\mathbf{P},$$

where
$$Z_t = \exp\left\{\int_0^t h^*(X(s))dY(s) - \frac{1}{2}\int_0^t |h(X(s))|^2 ds\right\}$$

(here and below, if $\zeta \in \mathbf{R}^k$, then ζ is a *column* vector, $\zeta^* = (\zeta_1, \ldots, \zeta_k)$, and $|\zeta|^2 = \zeta^*\zeta$). Since the function h is bounded, the measures $\mathbf{P}$ and $\tilde{\mathbf{P}}$ are equivalent. The expectation with respect to the measure $\tilde{\mathbf{P}}$ will be denoted by $\tilde{\mathbf{E}}$.

The following properties of the measure $\tilde{\mathbf{P}}$ are well known [33, 61]:

1. Under the measure $\tilde{\mathbf{P}}$, the distributions of the Wiener process W and the random variable X_0 are unchanged, the observation process Y is a standard Wiener process, and the state process X satisfies

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t) + \rho(X(t))\left(dY(t) - h(X(t))dt\right),\ 0 < t \le T;$$
$$X(0) = X_0;$$

2. Under the measure $\tilde{\mathbf{P}}$, the Wiener processes W and Y and the random variable X_0 are independent of one another;

3. The optimal filter $\hat{f}_t$ satisfies

$$\hat{f}_t = \frac{\tilde{\mathbf{E}}\left[f(X(t))Z_t|\mathcal{F}_t^Y\right]}{\tilde{\mathbf{E}}[Z_t|\mathcal{F}_t^Y]}. \tag{4.3}$$

Next, consider the partial differential operators

$$\mathcal{L}g(x) = \frac{1}{2}\sum_{i,j=1}^{d}\left((\sigma(x)\sigma^*(x))_{ij} + (\rho(x)\rho^*(x))_{ij}\right)\frac{\partial^2 g(x)}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x)\frac{\partial g(x)}{\partial x_i};$$

$$\mathcal{M}_l g(x) = h_l(x)g(x) + \sum_{i=1}^{d}\rho_{il}(x)\frac{\partial g(x)}{\partial x_i},\ l = 1, \ldots, r$$

and their adjoints

$$\mathcal{L}^* g(x) = \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}\left((\sigma(x)\sigma^*(x))_{ij}g(x) + (\rho(x)\rho^*(x))_{ij}g(x)\right) - \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\left(b_i(x)g(x)\right);$$

$$\mathcal{M}_l^* g(x) = h_l(x)g(x) - \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\left(\rho_{il}(x)g(x)\right),\ l = 1, \ldots, r.$$

The following result is well known [61, Theorem 6.2.1].

4.2.4. Theorem. *If the filtering model is regular, then there is a random field* $p = p(t,x),\ t \in [0,T],\ x \in \mathbf{R}^d$, *so that*

$$\tilde{\mathbf{E}}\left[f(X(t))Z_t|\mathcal{F}_t^Y\right] = \int_{\mathbf{R}^d} f(x)p(t,x)dx \tag{4.4}$$

and $p(t,x)$ *satisfies the stochastic partial differential equation*

$$\begin{aligned} dp(t,x) &= \mathcal{L}^* p(t,x)dt + \sum_{l=1}^{r}\mathcal{M}_l^* p(t,x)dY_l(t),\ 0 < t \le T,\ x \in \mathbf{R}^d; \\ p(0,x) &= p_0(x). \end{aligned} \tag{4.5}$$

4.2.5. Definition. Equation (4.5) is called the **Zakai equation**, the random field $p = p(t,x)$, the **unnormalized filtering density** (UFD), and the random variable $\phi_t[f] = \tilde{\mathbf{E}}\left[f(X(t))Z_t|\mathcal{F}_t^Y\right]$, the **unnormalized optimal filter**.

It follows from (4.3) and (4.4) that the optimal filter $\hat{f}_t$ satisfies

$$\hat{f}_t = \frac{\int_{\mathbf{R}^d} f(x)p(t,x)dx}{\int_{\mathbf{R}^d} p(t,x)dx} = \frac{\phi_t[f]}{\phi_t[1]},$$

and the filtering problem is solved once the UFD $p(t,x)$ (or an approximation of $p(t,x)$) is computed.

Let $\mathbf{H}^a$ be the Sobolev space $W_2^a(\mathbf{R}^d)$ [1, 37, 45, 61]. Under assumption (R1) from Definition 4.2.2, for every $a \in \mathbf{R}$ the operator $\mathcal{L}^*$ is linear and bounded from $\mathbf{H}^a$ to $\mathbf{H}^{a-2}$ and each operator $\mathcal{M}_l^*$, $l = 1,\ldots,r$, is linear and bounded from $\mathbf{H}^a$ to $\mathbf{H}^{a-1}$. The Zakai equation (4.5) will be considered on the probability space $(\Omega, \mathcal{F}, \tilde{\mathbf{P}})$, so that $Y = Y(t)$ is a standard Wiener process on this space. Therefore equation (4.5) is a particular case of the stochastic evolution equation studied in the previous two chapters. Direct computations show that equation (4.5) is (at least) dissipative, and then Theorem 2.2.6 implies that the solution $p = p(t,x)$ exists, is unique, and

$$p \in L_2(\Omega \times [0,T]; \mathbf{H}^1) \cap L_2(\Omega; \mathbf{C}([0,T]; \mathbf{H}^0))$$

(with $\mathbf{H}^0 = L_2(\mathbf{R}^d)$). In fact, the regularity assumptions (R1) and (R3) imply much higher regularity of the random field p. To state the corresponding result, another definition is necessary.

4.2.6. Definition. For $q \in \mathbf{R}$, the space $L_2(q, \mathbf{R}^d)$ is defined by

$$L_2(q, \mathbf{R}^d) = \left\{ g : \int_{\mathbf{R}^d} (1+|x|^2)^q |g(x)|^2 dx < \infty \right\}.$$

The norm $\|\cdot\|_{0,q}$ in the space $L_2(q, \mathbf{R}^d)$ is defined by

$$\|g\|_{0,q} = \left(\int_{\mathbf{R}^d} (1+|x|^2)^q |g(x)|^2 dx \right)^{1/2}.$$

The **weighted Sobolev space** $\mathbf{H}^n(q)$, $n = 0, 1, \ldots$, is the collection of (generalized) functions which belong to the space $L_2(q, \mathbf{R}^d)$ together with all their generalized derivatives up to order n. For every n and q, $\mathbf{H}^n(q)$ is a separable Hilbert space [61, Theorem 3.4.7]; the norm and the inner product in this space will be denoted by $\|\cdot\|_{n,q}$ and $(\cdot,\cdot)_{n,q}$ respectively. If $u,\ v \in \mathbf{H}^n(q)$, then

$$(u,v)_{n,q} = \sum_{|\gamma| \le n} \int_{\mathbf{R}^d} D^\gamma u(x) D^\gamma v(x) (1+|x|^2)^q dx,$$

where $\gamma = (\gamma_1, \ldots, \gamma_d)$, $|\gamma| = \sum_{i=1}^d \gamma_i$, $D^\gamma = \dfrac{\partial^{|\gamma|}}{\partial x_i^{\gamma_1} \ldots \partial x_d^{\gamma_d}}$.

Clearly, $\mathbf{H}^0(q) = L_2(q, \mathbf{R}^d)$ and $\mathbf{H}^n(0) = \mathbf{H}^n$.

4.2.7. Theorem. *If the filtering model is regular, then for every $q \in \mathbf{R}$ and every non-negative integer n*

$$p \in L_2(\Omega; \mathbf{C}([0,T]; \mathbf{H}^n(q))$$

and

$$\tilde{\mathbf{E}}\|p(t,\cdot)\|^2_{n,q} \le e^{C(n,q)t}\ \|p_0\|^2_{n,q}.$$

Proof. This follows from Theorem 4.3.2 in [61].

□

4.2.8. Corollary. *If $\{\Phi_t\}_{t\ge 0}$ is the semigroup generated by the operator $\mathcal{L}^*$, then for each $t \ge 0$ the operator Φ_t is linear and bounded from $\mathbf{H}^n(q)$ to $\mathbf{H}^n(q)$ and*

$$\|\Phi_t g\|_{n,q} \le e^{C(n,q)t}\ \|g\|_{n,q}.$$

4.3 The Spectral Separating Scheme of the First Kind

Consider the regular filtering model

$$\begin{aligned} &dX(t) = b(X(t))dt + \sigma(X(t))dW(t),\\ &dY(t) = h(X(t))dt + dV(t),\ 0 < t \le T;\\ &X(0) = X_0, \quad Y(0) = 0, \quad \text{estimate } f(X(t)); \end{aligned} \tag{4.6}$$

which is a particular case of (4.2) under an additional assumption that $\rho(x) = 0$ for all $x \in \mathbf{R}^d$. This assumption means that the observation noise is independent of the state process. Then the Zakai equation (4.5) becomes

$$\begin{aligned} dp(t,x) &= \mathcal{L}^* p(t,x)dt + \sum_{l=1}^{r} h_l(x)p(t,x)dY_l(t),\ 0 < t \le T,\ x \in \mathbf{R}^d;\\ p(0,x) &= p_0(x), \end{aligned} \tag{4.7}$$

so that the operators $\mathcal{M}_l = h_l(x),\ l = 1, \dots, r$, commute with each other and are linear and bounded from $\mathbf{H}^n(q)$ to $\mathbf{H}^n(q)$ for every $q \in \mathbf{R}$ and every non-negative integer n.

In this section, a recursive approximation of the solution of (4.7) will be constructed using the Cameron-Martin version of the Wiener chaos decomposition followed by the H-orthogonal projection of the first kind. This approach also provides a recursive approximation of the unnormalized optimal filter. The corresponding algorithm is referred to as the spectral separating scheme of the first kind (or S_1^3 for short). Equation (4.7) is considered on the probability space $(\Omega, \mathcal{F}, \tilde{\mathbf{P}})$ so that all error bounds are given in terms of the expectation $\tilde{\mathbf{E}}$. The properties of the approximation on the original probability space $(\Omega, \mathcal{F}, \mathbf{P})$ are studies below in
Section 4.5.

Let $0 = t_0 < t_1 \ldots < t_M = T$ be a uniform partition of the interval $[0, T]$ with step Δ, and define an orthonormal basis $\{m_k\}$ in $L_2([0, \Delta])$ by

$$m_1(s) = \frac{1}{\sqrt{\Delta}}; \quad m_k(s) = \sqrt{\frac{2}{\Delta}} \cos\left(\frac{\pi(k-1)s}{\Delta}\right), \; k > 1; \; 0 \le s \le \Delta. \tag{4.8}$$

Recall that J denotes the collection of r-dimensional multi-indices, and for $\alpha \in J$, $\alpha(i,j)$ stands for the multi-index $\tilde{\alpha} = (\tilde{\alpha}^l_k)_{1 \le l \le r, \; k \ge 1}$ with

$$\tilde{\alpha}^l_k = \begin{cases} \alpha^l_k & \text{if } k \ne i \text{ or } l \ne j \text{ or both} \\ \max(0, \alpha^j_i - 1) & \text{if } k = i \text{ and } l = j. \end{cases}$$

Also, for $\alpha \in J$, $|\alpha| = \sum_{k,l} \alpha^l_k$, $d(\alpha) = \max\{k : \alpha^l_k > 0 \text{ for some } l\}$, and $J^N_n = \{\alpha \in J : |\alpha| \le N, \; d(\alpha) \le n\}$. Define

$$\xi^i_\alpha = \frac{1}{\sqrt{\alpha!}} \prod_{k,l} H_{\alpha^l_k}\left(\int_{t_{i-1}}^{t_i} m_k(s - t_{i-1}) dY_l(s)\right), \tag{4.9}$$

where $H_n(t) = (-1)^n e^{t^2/2} \frac{d^n}{dt^n} e^{-t^2/2}$.

By Corollary 2.4.2, the solution of (4.7) can be written as

$$p(t_i, x) = \sum_{\alpha \in J} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(\Delta, x; p(t_{i-1}, \cdot)) \xi^i_\alpha, \quad i = 1, \ldots, M,$$

where the coefficients $\varphi_\alpha(\Delta, x; g)$ are the solutions of

$$\begin{aligned} \frac{\partial \varphi_\alpha(s, x; g)}{\partial s} &= \mathcal{L}^* \varphi_\alpha(s, x; g) + \sum_{k,l} \alpha^l_k m_k(s) h_l(x) \varphi_{\alpha(k,l)}(s, x; g), \; 0 < s \le \Delta; \\ \varphi_\alpha(0, x; g) &= g(x) 1_{\{|\alpha| = 0\}}. \end{aligned} \tag{4.10}$$

4.3.1. Lemma. *If $g \in \mathbf{H}^n(q)$, then $\varphi_a(t, \cdot; g) \in \mathbf{H}^n(q)$ for every $t \in [0, \Delta]$ and*

$$\sum_{|\alpha| = k} \frac{\|\varphi_\alpha(t, \cdot; g)\|^2_{n,q}}{\alpha!} \le e^{C(n,q)t} \frac{[C(n,q)t]^k}{k!} \|g\|^2_{n,q}.$$

Proof. This follows from Corollaries 2.3.8 and 4.2.8.

□

Let $\{e_\gamma\}_{\gamma \in \Gamma}$ be the Hermite basis in $\mathbf{R}^d$ so that

$$e_\gamma(x) = \prod_{i=1}^d e_{\gamma_i}(x_i),$$

where

$$e_n(t) = \frac{(-1)^n}{\sqrt{2^n \pi^{1/2} n!}} e^{t^2/2} \frac{d^n}{dt^n} e^{-t^2}.$$

The function e_γ is the eigenfunction of the self-adjoint operator $\Lambda = -\nabla^2 + (1 + |x|^2)$:

$$\Lambda e_\gamma = (2|\gamma| + d + 1) e_\gamma, \tag{4.11}$$

where ∇^2 is the Laplace operator. The domain of Λ^k (k-th power of Λ) includes $\mathbf{H}^{2k}(2k)$. Indeed,

$$\|\Lambda g\|_{k,k} \leq C(k) \left(\|g\|_{k+2,k} + \|g\|_{k,k+2} \right) \leq C(k) \|g\|_{k+2,k+2}$$

and so by induction

$$\|\Lambda^k g\|_0 \leq C(k) \|g\|_{2k,2k} \,. \tag{4.12}$$

Define $f_\gamma,\ \gamma \in \Gamma$, by

$$f_\gamma = \int_{\mathbf{R}^d} f(x) e_\gamma(x) dx.$$

The integral is well defined because the function f is of polynomial growth and $e_\gamma(x)$ decays exponentially fast as $|x| \to \infty$.

The following is a representation of the unnormalized optimal filter.

4.3.2. Theorem. *If the filtering model (4.6) is regular and*

$$\psi_\gamma(t_i) = (p(t_i, \cdot), e_\gamma)_0,$$

then

$$\phi_{t_i}[f] = \sum_{\gamma \in \Gamma} f_\gamma \psi_\gamma(t_i) \qquad (\mathbf{P}\text{–a.s.}) \tag{4.13}$$

Proof. By definition,

$$\phi_{t_i}[f] = \int_{\mathbf{R}^d} f(x) p(t_i, x) dx,$$

and by Theorem 4.2.7,

$$p(t_i, x) = \sum_{\gamma \in \Gamma} \psi_\gamma(t_i) e_\gamma(x) \qquad (\mathbf{P}\text{–a.s.})$$

Therefore, (4.13) will follow if the series $\sum_{\gamma \in \Gamma} f_\gamma \psi_\gamma(t_i)$ converges absolutely $\mathbf{P}$- a.s. so that it is possible to interchange the order of summation and integration. Since the measures $\mathbf{P}$ and $\tilde{\mathbf{P}}$ are equivalent, it is sufficient to establish $\tilde{\mathbf{P}}$ - a.s. convergence of the series, which will follow if

$$\sum_{\gamma \in \Gamma} |f_\gamma| \ \tilde{\mathbf{E}} |\psi_\gamma(t_i)| < \infty. \tag{4.14}$$

By the regularity assumption (R4),

$$|f_\gamma| \leq L \int_{\mathbf{R}^d} (1 + |x|^{k_0}) |e_\gamma(x)| dx \leq C \prod_{i=1}^{d} \int_{\mathbf{R}} (1 + |x_i|^{2k_0}) |e_{\gamma_i}(x_i)| dx_i,$$

and it follows from [25, paragraph (21.3.3)] that

$$\int_{\mathbf{R}} (1+|x_i|^{2k_0})|e_{\gamma_i}(x_i)|dx_i \le C|\gamma_i|^{(4k_0+1)/4},$$

whence

$$|f_\gamma| \le C|\gamma|^{(k_0+1)d}. \tag{4.15}$$

On the other hand, by (4.11), (4.12), and Theorem 4.2.7, for every $\nu > 0$ there is a $C(\nu)$ so that

$$\tilde{\mathbf{E}}|\psi_\gamma(t_i)| \le \sqrt{\tilde{\mathbf{E}}|(p(t_i,\cdot),e_\gamma)_0|^2} \le \frac{C(\nu)}{|\gamma|^\nu}.$$

Combining the last inequality with (4.15) and taking ν sufficiently large implies (4.14).

□

For a positive integer κ, the set $\{\gamma \in \Gamma : |\gamma| = \sum_{i=1}^d \gamma_i \le \kappa\}$ will be denoted by Γ_κ. If $K_\kappa = |\Gamma_\kappa|$ is the number of elements in Γ_κ, then K_κ is the number of non-negative integer solutions $\gamma_1, \ldots, \gamma_d$ of the inequality $\gamma_1 + \ldots + \gamma_d \le \kappa$ and is equal to $(\kappa+d)!/\kappa!d!$ (the proof can be found in [12]).

The approximation $P_N^{n,\kappa}(t_i,x)$ of $p(t_i,x)$ is defined by

$$P_N^{n,\kappa}(t_i,x) = \sum_{\gamma\in\Gamma_\kappa} \psi_\gamma(t_i;N,n,\kappa)e_\gamma(x),$$

where

$$\begin{aligned}
&\psi_\gamma(t_0;N,n,\kappa) = (p_0,e_\gamma)_0,\ \gamma \in \Gamma_\kappa,\\
&\psi_\gamma(t_{i+1};N,n,\kappa) = \sum_{\tau\in\Gamma_\kappa}\sum_{\alpha\in J_N^n} (\varphi_\alpha(\Delta,\cdot;e_\tau),e_\gamma)_0\psi_\tau(t_i;N,n,\kappa)\xi_\alpha^i.
\end{aligned}$$

Also define an approximation $\bar{\phi}_{t_i,\kappa}[f]$ of the unnormalized optimal filter $\phi_t[f]$ by

$$\bar{\phi}_{t_i,\kappa}[f] = \sum_{\gamma\in\Gamma_\kappa} f_\gamma\psi_\gamma(t_i;N,n,\kappa).$$

4.3.3. Theorem. *It the model (4.6) is regular, then for every positive integer ν there is a constant*[1] *$C(\nu)$ so that*

$$\max_{0\le i\le M} \tilde{\mathbf{E}}\|p(t_i,\cdot) - P_N^{n,\kappa}(t_i,\cdot)\|_0^2 \le \frac{(C\Delta)^N}{(N+1)!} + \frac{C\Delta^2}{n} + \frac{C(\nu)}{\kappa^\nu\Delta^2}, \tag{4.16}$$

and

$$\max_{0\le i\le M} \tilde{\mathbf{E}}|\phi_{t_i}[f] - \bar{\phi}_{t_i,\kappa}[f]|^2 \le \frac{(C\Delta)^N}{(N+1)!} + \frac{C\Delta^2}{n} + \frac{C(\nu)}{\kappa^\nu\Delta^2}. \tag{4.17}$$

[1] Recall that all constants are denoted by C and can depend on the parameters of the model, which include the coefficients of the diffusion equations and the length of the time interval; the value of the constants can be different in different places.

Proof. Define operators F_i by

$$F_i(g)(x) = \sum_{\alpha \in J_N^n} \varphi_\alpha(\Delta, x; g)\xi_\alpha^i, \; i = 1, \ldots, M,$$

and then the sequence $\{p_N^n(t_i, x)\}_{0 \le i \le M}$,

$$p_N^n(t_0, x) = p_0(x), \; p_N^n(t_{i+1}, x) = F_{i+1}(p_N^n(t_i, \cdot))(x), \; i = 1, \ldots, M.$$

By Theorem 2.4.3 and Corollary 4.2.8,

$$\max_{0 \le i \le M} \tilde{\mathbf{E}} \|p(t_i, \cdot) - p_N^n(t_i, \cdot)\|_0^2 \le \frac{(C\Delta)^N}{(N+1)!} + \frac{C\Delta^2}{n}. \tag{4.18}$$

Next, if Π^{K_κ} is the orthogonal projection on the span of $\{e_\gamma\}_{\gamma \in \Gamma_\kappa}$, then

$$P_N^{n,\kappa}(t_{i+1}, x) = \Pi^{K_\kappa} F_{i+1}(P_N^{n,\kappa}(t_i, \cdot))(x).$$

By Lemma 4.3.1,

$$\tilde{\mathbf{E}} \|F_{i+1}(p_N^n(t_i, \cdot))\|_0^2 \le e^{C\Delta} \tilde{\mathbf{E}} \|p_N^n(t_i, \cdot)\|_0^2,$$

and by (4.12) and Lemma 4.3.1,

$$\tilde{\mathbf{E}} \|\Lambda^k p_N^n(t_i, \cdot)\|_0^2 \le C(k) \tilde{\mathbf{E}} \|p_N^n(t_i, \cdot)\|_{2k,2k}^2 < \infty.$$

Then, according to Remark 3.3.4, for every $\nu > 0$ there is a $C(\nu)$ so that

$$\max_{0 \le i \le M} \tilde{\mathbf{E}} \|p_N^n(t_i, \cdot) - P_N^{n,\kappa}(t_i, \cdot)\|_0^2 \le \frac{C(\nu)}{K_\kappa^\nu \, \Delta^2}.$$

Since $K_\kappa \asymp \kappa^d$, the last inequality implies

$$\max_{0 \le i \le M} \tilde{\mathbf{E}} \|p_N^n(t_i, \cdot) - P_N^{n,\kappa}(t_i, \cdot)\|_0^2 \le \frac{C(\nu)}{\kappa^\nu \Delta^2}. \tag{4.19}$$

Combining (4.18) and (4.19) results in (4.16).

To prove (4.17) note that the assumption

$$|f(x)| \le L(1 + |x|^{k_0})$$

implies

$$f \in L_2(-q, \mathbf{R}^d) \;\text{ for every } q > \frac{d + 2k_0}{2}.$$

Fix an even integer q satisfying $q > d/2 + k_0$. Since both d and k_0 are the parameters of the model, the dependence of all constants on q will be suppressed. For every $g \in L_2(q, \mathbf{R}^2)$ the integral $\int_{R^d} f(x)g(x)dx$ is well defined and will be denoted by $(g, f)_0$. The Cauchy-Schwartz inequality implies that

$$|(g, f)_0| \le \|g\|_{0,q} \|f\|_{0,-q} \, ,$$

and then

$$\tilde{\mathbf{E}}|\bar{\phi}_{t_i,\kappa}[f]-\phi_{t_i}[f]|^2\equiv\tilde{\mathbf{E}}(p(t_i,\cdot)-P_N^{n,\kappa}(t_i,\cdot),f)_0^2\le \\ \|f\|_{0,-q}^2\tilde{\mathbf{E}}\|p(t_i,\cdot)-P_N^{n,\kappa}(t_i,\cdot)\|_{0,q}^2\le 2\|f\|_{0,-q}^2\Big(\tilde{\mathbf{E}}\|p(t_i,\cdot)-p_N^n(t_i,\cdot)\|_{0,q}^2+ \\ \tilde{\mathbf{E}}\|p_N^n(t_i,\cdot)-P_N^{n,\kappa}(t_i,\cdot)\|_{0,q}^2\Big). \tag{4.20}$$

The regularity assumption (R1) implies that for $a=0,2$ the operator $\mathcal{L}^*$ is linear and bounded from $\mathbf{H}^a(q)$ to $\mathbf{H}^{a-2}(q)$, the operators $\mathcal{M}_l^*,\ l=1,\ldots,r$, are linear and bounded from $\mathbf{H}^a(q)$ to $\mathbf{H}^a(q)$, and by Corollary 4.2.8 the operator Φ_t is linear and bounded from $\mathbf{H}^a(q)$ to $\mathbf{H}^a(q)$. Therefore by Theorem 2.4.3,

$$\max_{0\le i\le M}\tilde{\mathbf{E}}\|p(t_i,\cdot)-p_N^n(t_i,\cdot)\|_{0,q}^2\le\frac{(C\Delta)^N}{(N+1)!}+\frac{C\Delta^2}{n},$$

and, according to (4.20), inequality (4.17) will follow if for every $\nu>0$ there is a $C(\nu)$ so that

$$\tilde{\mathbf{E}}\|p_N^n(t_i,\cdot)-P_N^{n,\kappa}(t_i,\cdot)\|_{0,q}^2\le\frac{C(\nu)}{\kappa^\nu\Delta^2}. \tag{4.21}$$

To verify (4.21) note that if $g\in L_2(q,\mathbf{R}^d)$ and $\beta(x):=(1+|x|^2)^{q/2}$, then

$$\|g\|_{0,q}^2=\|g\beta\|_0^2=\sum_\gamma(g,\beta e_\gamma)_0^2.$$

The properties of e_γ imply that

$$(1+|x|^2)e_\gamma(x)=\sum_{\tau\in\Gamma(\gamma)}C_\tau(\gamma)e_\tau(x), \tag{4.22}$$

where $\Gamma(\gamma)$ is a finite subset of Γ with the number of elements bounded from above by a constant independent of γ, and $C_\tau(\gamma)$ is a constant with $|C_\tau(\gamma)|\le C|\tau|$. Indeed, $(1+|x|^2)e_\gamma(x)=(2|\gamma|+d+1)e_\gamma(x)+\nabla^2e_\gamma(x)$, $e_\gamma(x)=\prod_{i=1}^d e_{\gamma_i}(x_i)$, and

$$\frac{de_{\gamma_i}(x_i)}{dx_i}=\frac{1}{\sqrt{2}}\left(\sqrt{\gamma_i}e_{\gamma_i-1}(x_i)-\sqrt{\gamma_i+1}e_{\gamma_i+1}(x_i)\right), \tag{4.23}$$

whence (4.22). Since $q/2$ is an integer, proceeding by induction results in

$$\beta(x)e_\gamma(x)=\sum_{\tau\in\Gamma(\gamma)}C_\tau(\gamma)e_\tau(x),$$

where $\Gamma(\gamma)$ is a finite subset of Γ with the number of elements bounded from above by a constant independent of γ, and $C_\tau(\gamma)$ is a constant with $|C_\tau(\gamma)|\le C|\tau|^{q/2}$. Therefore,

$$(g,\beta e_\gamma)_0^2\le C\sum_{\tau\in\Gamma(\gamma)}|\tau|^q(q,e_\gamma)_0^2$$

and

$$\|g\|_{0,q}^2\equiv\sum_{\gamma\in\Gamma}(g,\beta e_\gamma)_0^2\le C\sum_{\gamma\in\Gamma}|\gamma|^q(g,e_\gamma)_0^2.$$

Thus,

$$\begin{aligned}
&\tilde{\mathbf{E}}\|p_N^n(t_i,\cdot)-P_N^{n,\kappa}(t_i,\cdot)\|_{0,q}^2 \le C\sum_{\gamma\in\Gamma}|\gamma|^q\,\tilde{\mathbf{E}}\left(p_N^n(t_i,\cdot)-P_N^{n,\kappa}(t_i,\cdot),e_\gamma\right)_0^2 = \\
&C\sum_{\gamma\in\Gamma_\kappa}|\gamma|^q\,\tilde{\mathbf{E}}\left(p_N^n(t_i,\cdot)-P_N^{n,\kappa}(t_i,\cdot),e_\gamma\right)_0^2 + C\sum_{\gamma\notin\Gamma_\kappa}|\gamma|^q\,\tilde{\mathbf{E}}\left(p_N^n(t_i,\cdot),e_\gamma\right)_0^2 \le \\
&C\kappa^q\tilde{\mathbf{E}}\|p_N^n(t_i,\cdot)-P_N^{n,\kappa}(t_i,\cdot)\|_0^2 + C\sum_{\gamma\notin\Gamma_\kappa}|\gamma|^q\,\tilde{\mathbf{E}}\left(p_N^n(t_i,\cdot),e_\gamma\right)_0^2
\end{aligned} \tag{4.24}$$

By (4.19),

$$\max_{0\le i\le M}\kappa^q\tilde{\mathbf{E}}\|p_N^n(t_i,\cdot)-P_N^{n,\kappa}(t_i,\cdot)\|_0^2 \le \frac{C(\nu)}{\kappa^\nu}, \tag{4.25}$$

and by Lemma 4.3.1, since $p_N^n(t_i,\cdot)$ is in the domain of every power of Λ,

$$\tilde{\mathbf{E}}\left(p_N^n(t_i,\cdot),e_\gamma\right)_0^2 \le \frac{C(\nu)}{|\gamma|^\nu},$$

so that taking ν sufficiently large results in

$$\max_{0\le i\le M}\sum_{\gamma\notin\Gamma_\kappa}|\gamma|^q\,\tilde{\mathbf{E}}\left(p_N^n(t_i,\cdot),e_\gamma\right)_0^2 \le \frac{C(\nu)}{|\gamma|^\nu}. \tag{4.26}$$

After that, (4.21) follows from (4.24), (4.25), and (4.26). This completes the proof of the theorem.

□

The algorithm for computing $P_N^{n,\kappa}$ and $\bar{\phi}_{t_i,\kappa}[f]$ is presented below.

1. Off line (before the observations are available):

for $\alpha\in J_N^n$ *and* $\gamma,\tau\in\Gamma_\kappa$ *compute*
$q_{\gamma\tau}^\alpha=(\varphi_\alpha(\Delta,\cdot;e_\tau),e_\gamma)_0$, $f_\gamma=(f,e_\gamma)_0$, *and* $\psi_\gamma(t_0;N,n,\kappa)=(p_0,e_\gamma)_0$;

set $P_N^{n,\kappa}(t_0,x)=\sum_{\gamma\in\Gamma_\kappa}\psi_\gamma(t_0;N,n,\kappa)e_\gamma(x)$ *and* $\bar{\phi}_{t_0,\kappa}[f]=\sum_{\gamma\in\Gamma_\kappa}f_\gamma\psi_\gamma(t_0;N,n,\kappa)$.

2. On line, step i (as the observations become available): compute

$$\psi_\gamma(t_i;N,n,\kappa)=\sum_{\tau\in\Gamma_\kappa}Q_{\gamma\tau}(\xi^i)\psi_\tau(t_{i-1};N,n,\kappa),\quad \gamma\in\Gamma_\kappa, \tag{4.27}$$

where

$$Q_{\gamma\tau}(\xi^i)=\sum_{\alpha\in J_N^n}q_{\gamma\tau}^\alpha\xi_\alpha^i,$$

then, if necessary, compute

$$P_N^{n,\kappa}(t_i,x)=\sum_{\gamma\in\Gamma_\kappa}\psi_\gamma(t_i;N,n,\kappa)e_\gamma(x), \tag{4.28}$$

$$\bar{\phi}_{t_i,\kappa}[f]=\sum_{\gamma\in\Gamma_\kappa}f_\gamma\psi_\gamma(t_i;N,n,\kappa), \tag{4.29}$$

and

$$\bar{f}_{t_i,\kappa} = \frac{\bar{\phi}_{t_i,\kappa}[f]}{\bar{\phi}_{t_i,\kappa}[1]}. \tag{4.30}$$

This algorithm will be referred to as the **spectral separating scheme of the first kind** (or S_1^3 for short).

4.3.4. Remark. The main advantage of S_1^3 as compared to the existing schemes for solving the Zakai equation is that the time consuming computations, including solving partial differential equations and computing integrals, are performed off line, while the on-line part is relatively simple even when the dimension d of the state process is large. Here are some other features of the algorithm:

(1) The overall amount of the off-line computations does not depend on the number of the on-line time steps;

(2) Formulas (4.29) and (4.30) can be used to compute an approximation to $\hat{f}_{t_i}$ (e.g. conditional moments) without the time consuming computations of $P_N^{n,\kappa}(t_i, x)$ and the related integrals;

(3) Only the coefficients $\psi_\gamma(t_i; N, n, \kappa)$ must be computed at every time step while the approximate filter $\bar{f}_{t_i}$ and/or UFD $P_N^{n,\kappa}(t_i, x)$ can be computed as needed, e.g. at the final time moment.

(4) The on-line part of the algorithm can be easily parallelized.

(5) An obvious limitation is that all the parameters of the model must be known in advance. Also, in the above form, the algorithm cannot be used if $\rho \neq 0$ or the coefficients of the equations (4.6) depend on time or are random.

4.3.5. Remark. By Theorem 4.3.3, the algorithm converges in the limit $\lim_{\Delta\to 0}\lim_{\kappa\to\infty}$ for all positive integers N and n. If $n = 1$, then each ξ_α^i depends only on the increments $Y_l(t_i) - Y_l(t_{i-1})$ of the observation process. In this form the algorithm can be used in the model with discrete time observations [52]. The question of approximating the integrals $\int_{t_{i-1}}^{t_i} m_k(s - t_{i-1})dY_l(s)$ for $k > 1$ and other computational aspects of the algorithm, including the complexity, are studied in Section 4.5.

4.4 The Spectral Separating Scheme of the Second Kind

Consider the regular filtering model

$$\begin{aligned} &dX(t) = b(X(t))dt + \sigma(X(t))dW(t) + \rho(X(t))dV(t),\\ &dY(t) = h(X(t))dt + dV(t),\ 0 < t \leq T;\\ &X(0) = X_0, \quad Y(0) = 0, \quad \text{estimate } f(X(t)). \end{aligned} \tag{4.31}$$

The Zakai equation describing the time evolution of the UFD $p(t,x)$ is

$$\begin{aligned} dp(t,x) &= \mathcal{L}^* p(t,x)dt + \sum_{l=1}^{r} \mathcal{M}_l^* p(t,x)dY_l(t),\ 0 < t \le T,\ x \in \mathbf{R}^d; \\ p(0,x) &= p_0(x), \end{aligned} \tag{4.32}$$

where

$$\mathcal{L}^* g(x) = \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}\left((\sigma(x)\sigma^*(x))_{ij} g(x) + (\rho(x)\rho^*(x))_{ij} g(x)\right) - \sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left(b_i(x) g(x)\right);$$

$$\mathcal{M}_l^* g(x) = h_l(x) g(x) - \sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left(\rho_{il}(x) g(x)\right),\ l = 1, \ldots, r.$$

In this section, a recurcive approximation of the solution of (4.32) will be constructed using the Galerkin approximation followed by the Cameron-Martin version of the Wiener chaos decomposition applied to the resulting system of stochastic ordinary differential equations. The idea was first used by Fung [20] to study the Zakai equation in a bounded domain of $\mathbf{R}^d$. This approach also provides a recursive approximation of the unnormalized optimal filter. The corresponding algorithm is referred to as the spectral separating scheme of the second kind (or S_2^3 for short). Equation (4.32) is considered on the probability space $(\Omega, \mathcal{F}, \tilde{\mathbf{P}})$ so that all error bounds are given in terms of the expectation $\tilde{\mathbf{E}}$. The properties of the approximation on the original probability space $(\Omega, \mathcal{F}, \mathbf{P})$ are studies below in Section 4.5.

Let $\{e_\gamma\}_{\gamma \in \Gamma}$ be the Hermite basis in $\mathbf{R}^d$ so that e_γ is the eigenfunction of the self-adjoint operator $\Lambda = -\nabla^2 + (1 + |x|^2)$:

$$\Lambda e_\gamma = (2|\gamma| + d + 1)e_\gamma,$$

where ∇^2 is the Laplace operator. For a positive integer κ define the set $\Gamma_\kappa = \{\gamma \in \Gamma : |\gamma| = \sum_{i=1}^{d} \gamma_i \le \kappa\}$. The number of elements in Γ_κ is denoted by K_κ. Define the matrices $A^\kappa = (A^\kappa_{\gamma\tau})_{\gamma,\tau \in \Gamma_\kappa}$ and $B^\kappa_l = (B^\kappa_{l,\gamma\tau})_{\gamma,\tau \in \Gamma_\kappa}$, $l = 1, \ldots, r$, by

$$A^\kappa_{\gamma\tau} = (\mathcal{L}^* e_\tau, e_\gamma)_0, \quad B^\kappa_{l,\gamma\tau} = (\mathcal{M}_l^* e_\tau, e_\gamma)_0.$$

Since $e_\gamma \in \mathbf{H}^2$ for all γ, the scalar products are well defined. The Galerkin approximation $p^\kappa(t,x)$ of $p(t,x)$ is given by

$$p^\kappa(t,x) = \sum_{\gamma \in \Gamma_\kappa} p^\kappa(t) e_\gamma(x),$$

where the vector $p^\kappa(t) = (p^\kappa_\gamma(t))_{\gamma \in \Gamma_\kappa}$ is the solution of the system of stochastic ordinary differential equations

$$dp^\kappa(t) = A^\kappa p^\kappa(t)dt + \sum_{l=1}^{r} B^\kappa_l p^\kappa(t) dY_l(t) \tag{4.33}$$

with the initial condition $p^\kappa_\gamma(0) = (p_0, e_\gamma)_0$. System (4.33) is a particular case of the stochastic evolution equation studied in Chapters 2 and 3 with the operators $\mathcal{A}$ and $\mathcal{B}_l$ linear and bounded in a finite dimensional space. Note that the matrices B^κ_l, $l = 1, \ldots, r$, do not, in general, commute with each other even if $\rho(x) \equiv 0$.

Let $0 = t_0 < t_1 \ldots < t_M = T$ be a uniform partition of the interval $[0,T]$ with step Δ. Define an orthonormal basis $\{m_k\}$ in $L_2([0,\Delta])$ by

$$m_1(s) = \frac{1}{\sqrt{\Delta}}; \quad m_k(s) = \sqrt{\frac{2}{\Delta}} \cos\left(\frac{\pi(k-1)s}{\Delta}\right), \ k > 1; \ 0 \le s \le \Delta,$$

and

$$\xi_\alpha^i = \frac{1}{\sqrt{\alpha!}} \prod_{k,l} H_{\alpha_k^l}\left(\int_{t_{i-1}}^{t_i} m_k(s - t_{i-1}) dY_l(s)\right),$$

where $H_n(t) = (-1)^n e^{t^2/2} \dfrac{d^n}{dt^n} e^{-t^2/2}$.

By Corollary 2.4.2, the solution of (4.33) can be written as

$$p^\kappa(t_i) = \sum_{\alpha \in J} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha^\kappa(\Delta; p^\kappa(t_{i-1}))\xi_\alpha^i, \quad i = 1, \ldots, M,$$

where the coefficients $\varphi_\alpha^\kappa(\Delta;\zeta)$ are the solutions of

$$\begin{aligned}\frac{\partial \varphi_\alpha^\kappa(s;\zeta)}{\partial s} &= A^\kappa \varphi_\alpha^\kappa(s;\zeta) + \sum_{k,l} \alpha_k^l m_k(s) B_l^\kappa \varphi_{\alpha(k,l)}^\kappa(s;\zeta), \ 0 < s \le \Delta \\ \varphi_\alpha^\kappa(0;\zeta) &= \zeta 1_{\{|\alpha|=0\}}, \quad \zeta \in \mathbf{R}^{K_\kappa}.\end{aligned}$$

For fixed positive integers N and n define the approximation $p_N^{\kappa,n}(t_i)$ of $p^\kappa(t)$ by

$$p_N^{\kappa,n}(t_0) = p^\kappa(0), \quad p_N^{\kappa,n}(t_i) = \sum_{\alpha \in J_N^n} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha^\kappa(\Delta; p_N^{\kappa,n}(t_{i-1}))\xi_\alpha^i, \quad i = 1, \ldots, M,$$

and also

$$p_N^{\kappa,n}(t_i, x) = \sum_{\gamma \in \Gamma_\kappa} p_{N,\gamma}^{\kappa,n}(t_i) e_\gamma(x),$$

$$\tilde{\phi}_{t_i,\kappa}[f] = \sum_{\gamma \in \Gamma_\kappa} f_\gamma p_{N,\gamma}^{\kappa,n}(t_i),$$

where $f_\gamma = \int_{\mathbf{R}^d} f(x) e_\gamma(x) dx$.

4.4.1. Theorem. *If the filtering model (4.31) is regular in the sense of Definition 4.2.2 and*

$$C_\rho = \max_{i,l} \sup_{x \in \mathbf{R}^d} |\rho_{il}(x)|^2,$$

then for every $\nu > 0$ there is a $C(\nu)$ so that

$$\max_{0 \le i \le M} \tilde{\mathbf{E}} \|p(t_i, \cdot) - p_N^{\kappa,n}(t_i, \cdot)\|_0^2 \le \frac{C(\nu)}{\kappa^\nu} + \left(C\frac{(1 + C_\rho\kappa)\Delta + (\kappa^2 + C_\rho\kappa^3)\Delta^2}{n} + \frac{(C(1 + C_\rho\kappa))^{N+1}\Delta^N}{(N+1)!}\right) e^{C(1+C_\rho\kappa)T} \tag{4.34}$$

and

$$\max_{0\le i\le M}\tilde{\mathbf{E}}|\phi_{t_i}[f]-\tilde{\phi}_{t_i,\kappa}[f]|^2\le\frac{C(\nu)}{\kappa^\nu}+ \\ \left(C\frac{(1+C_\rho\kappa)\Delta+(\kappa^2+C_\rho\kappa^3)\Delta^2}{n}+\frac{(C(1+C_\rho\kappa))^{N+1}\Delta^N}{(N+1)!}\right)e^{C(1+C_\rho\kappa)T}. \tag{4.35}$$

Proof. To simplify the presentation, denote $1+C_\rho\kappa$ by C_κ. It follows from (4.23) that $e_\gamma\in\mathbf{H}^1$ and

$$\|e_\gamma\|_1\le C|\gamma|^{1/2}.$$

Consequently, by Theorem 3.4.1, for ever $\nu>0$ there is a $C(\nu)$ so that

$$\sup_{0\le t\le T}\tilde{\mathbf{E}}\|p(t,\cdot)-p^\kappa(t,\cdot)\|_0^2\le\frac{C(\nu)}{\kappa^\nu}. \tag{4.36}$$

To prove (4.34), it remains to show that

$$\tilde{\mathbf{E}}|p^\kappa(t)-p_N^{\kappa,n}(t)|^2\le\left(C\frac{C_\kappa\Delta+\kappa^2C_\kappa\Delta^2}{n}+\frac{(CC_\kappa)^{N+1}\Delta^N}{(N+1)!}\right)e^{CC_\kappa T},$$

and by Theorem 2.4.3 this inequality holds if for every vector $\zeta\in\mathbf{R}^{K_\kappa}$,

$$|A^\kappa\zeta|^2\le C\kappa^2|\zeta|^2, \tag{4.37}$$

$$|B_l^\kappa\zeta|^2\le CC_\kappa|\zeta|^2, \tag{4.38}$$

and

$$|\Phi_t^\kappa\zeta|^2\le e^{Ct}|\zeta|^2, \tag{4.39}$$

where $\Phi_t^\kappa=e^{A^kt}$.

These inequalities are verified below. For the vector ζ, $g=g(x)$ denotes the function

$$g(x)=\sum_{\gamma\in\Gamma_\kappa}\zeta_\gamma e_\gamma(x),$$

and Π^{K_κ} is the orthogonal projection on the span of $\{e_\gamma\}_{\gamma\in\Gamma_\kappa}$.
Proof of (4.37). By the definition of the matrix A^κ,

$$|A^\kappa\zeta|^2=\sum_{\gamma\in\Gamma_\kappa}\left(\sum_{\tau\in\Gamma_\kappa}(\mathcal{L}^*e_\tau,e_\gamma)_0\zeta_\tau\right)^2=\|\Pi^{K_\kappa}\mathcal{L}^*g\|_0^2\le\|\mathcal{L}^*g\|_0^2\le \\ C\|g\|_2^2\le C\|(1-\nabla^2)g\|_0^2\le C\|\Lambda g\|_0^2=C\sum_{\gamma\in\Gamma_\kappa}(1+d+2|\gamma|)^2|\zeta_\gamma|^2\le C\kappa^2|\zeta|^2,$$

whence (4.37).

Proof of (4.38). By the definition of the matrix B_l^κ,

$$
\begin{aligned}
|B_l^\kappa\zeta|^2 &= \sum_{\gamma\in\Gamma_\kappa}\left(\sum_{\tau\in\Gamma_\kappa}(\mathcal{M}_l^* e_\tau, e_\gamma)_0\zeta_\tau\right)^2 = \|\Pi^{K_\kappa}\mathcal{M}_l^* g\|_0^2 \le \|\mathcal{M}_l^* g\|_0^2 = \\
&C(\mathcal{M}_l\mathcal{M}_l^* g, g)_0 \le C\|g\|_0^2 + CC_\rho\left((1-\nabla^2)g, g\right)_0 \le C\|g\|_0^2 + CC_\rho(\Lambda g, g)_0 = \\
&C\|g\|_0^2 + CC_\rho \sum_{\gamma\in\Gamma_\kappa}(1+d+2|\gamma|)|\zeta_\gamma|^2 \le C(1+C_\rho\kappa)\ |\zeta|^2,
\end{aligned}
$$

whence (4.38).
Proof of (4.39). The dissipativity of the operator $\mathcal{L}^*$ implies

$$
\sum_{\gamma,\tau\in\Gamma_\kappa} A_{\gamma\tau}^\kappa \zeta_\gamma\zeta_\tau = (\mathcal{L}^* g, g)_0 \le C\|g\|_0^2 = C|\zeta|^2.
$$

The vector $v(t) = \Phi_t^\kappa\zeta$ is the solution of the ordinary differential equation

$$
dv(t) = A^\kappa v(t)dt,\ v(0) = \zeta,
$$

so that

$$
d|v(t)|^2 = 2\sum_{\gamma,\tau\in\Gamma_\kappa} A_{\gamma\tau}^\kappa v_\gamma(t)v_\tau(t)dt \le C|v(t)|^2 dt,
$$

and by the Gronwall lemma,

$$
|v(t)|^2 \le e^{Ct}|v(0)|^2,
$$

whence (4.39).

This completes the proof of (4.34).

To prove (4.35), the same arguments are used as in the proof of (4.17) from Theorem 4.3.3. Choose an even integer $q > k_0 + d/2$ so that

$$
\begin{aligned}
&\tilde{\mathbf{E}}|\tilde{\phi}_{t_i,\kappa}[f] - \phi_{t_i}[f]|^2 \le 2\|f\|_{0,-q}^2\Big(\tilde{\mathbf{E}}\|p(t_i,\cdot) - p^\kappa(t_i,\cdot)\|_{0,q}^2 + \\
&\tilde{\mathbf{E}}\|p^\kappa(t_i,\cdot) - p_N^{\kappa,n}(t_i,\cdot)\|_{0,q}^2\Big).
\end{aligned}
$$

Next, by the inequality

$$
\|g\|_{0,q}^2 \le C\sum_{\gamma\in\Gamma}|\gamma|^q(g,e_\gamma)_0^2, \tag{4.40}
$$

which was established during the proof of Theorem 4.3.3,

$$
\begin{aligned}
&\tilde{\mathbf{E}}\|p(t_i,\cdot) - p^\kappa(t_i,\cdot)\|_{0,q}^2 \le C\sum_{\gamma\in\Gamma}|\gamma|^q\ \tilde{\mathbf{E}}\left(p(t_i,\cdot) - p^\kappa(t_i,\cdot), e_\gamma\right)_0^2 = \\
&C\sum_{\gamma\in\Gamma_\kappa}|\gamma|^q\ \tilde{\mathbf{E}}|p_\gamma(t_i) - p_\gamma^\kappa(t_i)|^2 + C\sum_{\gamma\notin\Gamma_\kappa}|\gamma|^q\ \tilde{\mathbf{E}}|p_\gamma(t_i)|^2 \le \\
&C\kappa^q\tilde{\mathbf{E}}|p(t_i) - p^\kappa(t_i)|^2 + C\sum_{\gamma\notin\Gamma_\kappa}|\gamma|^q\ \tilde{\mathbf{E}}|p_\gamma(t_i)|^2.
\end{aligned}
\tag{4.41}
$$

By (4.36),

$$\max_{0\le i\le M} \kappa^q \tilde{\mathbf{E}}|p(t_i)-p^\kappa(t_i)|^2 \le \frac{C(\nu)}{\kappa^\nu}, \tag{4.42}$$

and since $p(t_i,\cdot)$ is in the domain of every power of Λ,

$$\tilde{\mathbf{E}}|p_\gamma(t_i)|^2 \le \frac{C(\nu)}{|\gamma|^\nu},$$

so that taking ν sufficiently large results in

$$\max_{0\le i\le M} \sum_{\gamma\notin\Gamma_\kappa} |\gamma|^q\, \tilde{\mathbf{E}}|p_\gamma(t_i)|^2 \le \frac{C(\nu)}{|\gamma|^\nu}. \tag{4.43}$$

After that, (4.41), (4.42), and (4.43) imply

$$\max_{0\le i\le M} \tilde{\mathbf{E}}\|p(t_i,\cdot)-p^\kappa(t_i,\cdot)\|^2_{0,q} \le \frac{C(\nu)}{\kappa^\nu}.$$

It remains to show that

$$\max\nolimits_{0\le i\le M} \tilde{\mathbf{E}}\|p^\kappa(t_i,\cdot)-p^{\kappa,n}_N(t_i,\cdot)\|^2_{0,q} \le \left(C\frac{C_\kappa\Delta+\kappa^2 C_\kappa\Delta^2}{n}+\frac{(CC_\kappa)^{N+1}\Delta^N}{(N+1)!}\right)e^{CC_\kappa T}. \tag{4.44}$$

Since $\Lambda e_\gamma=(1+d+2|\gamma|)e_\gamma := \lambda_\gamma e_\gamma$, inequality (4.40) implies

$$\tilde{\mathbf{E}}\|p^\kappa(t_i,\cdot)-p^{\kappa,n}_N(t_i,\cdot)\|^2_{0,q} \le C\sum_{\gamma\in\Gamma_\kappa} \lambda_\gamma^q \tilde{\mathbf{E}}|p^\kappa_\gamma - p^{\kappa,n}_{N,\gamma}|^2.$$

Define the diagonal matrix $\hat\Lambda=(\hat\Lambda_{\gamma\tau})_{\gamma,\tau\in\Gamma_k}$ by $\hat\Lambda_{\gamma\gamma}=\lambda_\gamma^{q/2}$. Then define the matrices

$$\hat A=\hat\Lambda A^\kappa\hat\Lambda^{-1},\;\; \hat B_l=\hat\Lambda B^\kappa_l\hat\Lambda^{-1}, \hat\Phi_t=e^{\hat A t},$$

and the vectors $\hat p(t)=\hat\Lambda p^\kappa(t)$, $\hat p^n_N(t_i)=\hat\Lambda p^{n,\kappa}_N(t_i)$. With these definitions, the vector $\hat p(t)$ is the solution of

$$d\hat p(t)=\hat A\hat p(t)dt+\sum_{l=1}^r \hat B_l\hat p(t)dY_l(t)$$

with the initial condition $\hat p_\gamma(0)=\lambda_\gamma^{q/2}(p_0,e_\gamma)_0$, and the vector $\hat p^n_N(t)$ satisfies

$$\hat p^n_N(t_0)=\hat p(0),\;\; \hat p^n_N(t_i)=\sum_{\alpha\in J^n_N}\frac{1}{\sqrt{\alpha!}}\varphi^\kappa_\alpha(\Delta;\hat p^n_N(t_{i-1}))\xi^i_\alpha,\quad i=1,\dots,M.$$

Direct computations show that the matrices $\hat A$, $\hat B_l$, and $\hat\Phi_t$ satisfy

$$|\hat A^\kappa\zeta|^2 \le C\kappa^2|\zeta|^2, \tag{4.45}$$

$$|\hat{\mathcal{B}}^\kappa_l\zeta|^2 \le CC_\kappa|\zeta|^2, \tag{4.46}$$

and

$$|\hat{\Phi}_t^\kappa \zeta|^2 \le e^{Ct}|\zeta|^2 \tag{4.47}$$

for all $\zeta \in \mathbf{R}^{K_\kappa}$. One possible way to verify this is by using the calculus of pseudo-differential operators in $\mathbf{R}^d$ [62, Chapter 4]. The (left) principal symbols of the operators $\mathcal{L}^*$, $\mathcal{M}_l^*$, and $\Lambda^{q/2}$ are

$$\begin{aligned}
a_{\mathcal{L}}(x,\eta) &= -\sum_{i,j=1}^{d}((\sigma(x)\sigma^*(x))_{ij} + (\rho(x)\rho^*(x))_{ij})\eta_i\eta_j,\\
a_l &= -\sqrt{-1}\sum_{i=1}^{d}\rho_{il}(x)\eta_i,\\
a_\Lambda &= (1+|x|^2+|\eta|^2)^{q/2}.
\end{aligned}$$

These symbols determine the continuity properties of the corresponding operators, and it follows from formulas (5.17) and (23.56) in [62] that the operators $\Lambda^{q/2}\mathcal{L}^*\Lambda^{-q/2}$ and $\Lambda^{q/2}\mathcal{M}_l^*\Lambda^{-q/2}$ have the same principal symbols as the operators $\mathcal{L}^*$ and $\mathcal{M}_l^*$ respectively. After that, the arguments are the same as in the proof of (4.37)–(4.39). In particular, the operator $\Lambda^{q/2}\mathcal{L}^*\Lambda^{-q/2}$ is dissipative (because $\mathcal{L}^*$ is), which implies (4.47). Also,

$$\hat{A}_{\gamma\tau} = (\Lambda^{q/2}\mathcal{L}^*\Lambda^{-q/2}e_\tau, e_\gamma)_0,\ \hat{B}_{l,\gamma\tau} = (\Lambda^{q/2}\mathcal{M}_l^*\Lambda^{-q/2}e_\tau, e_\gamma)_0,$$

so that

$$\begin{aligned}
|\hat{A}\zeta|^2 &\le \|\Lambda^{q/2}\mathcal{L}^*\Lambda^{-q/2}g\|_0^2 \le C\|\Lambda g\|_0^2 \le C\kappa^2|\zeta|^2,\\
|\hat{B}_l\zeta|^2 &\le \Lambda^{q/2}\mathcal{M}^*\Lambda^{-q/2}g\|_0^2 \le (\Lambda^{-q/2}\mathcal{M}_l\Lambda^q\mathcal{M}_l^*\Lambda^{-q/2}g,g)_0 \le CC_\kappa|\zeta|^2,
\end{aligned}$$

which proves (4.45) and (4.46).

As a result, Theorem 2.4.3 implies

$$\max_{0\le i\le M}\tilde{\mathbf{E}}|\hat{p}(t_i)-\hat{p}_N^n(t_i)|_{0,q}^2 \le \left(C\frac{C_\kappa\Delta + (\kappa^2C_\kappa)\Delta^2}{n} + \frac{(CC_\kappa)^{N+1}\Delta^N}{(N+1)!}\right)e^{CC_\kappa T}|\hat{p}(0)|^2,$$

and

$$|\hat{p}(0)|^2 = \sum_{\gamma\in\Gamma_k}\lambda_\gamma^q(p_0,e_\gamma)_0^2 \le \|\Lambda^{q/2}p_0\|_0^2 \le C,$$

which proves (4.44) and the theorem as a whole.

□

Let $\{\zeta^\gamma\}_{\gamma\in\Gamma}$ be the standard unit basis in $\mathbf{R}^{K_\kappa}$, i.e. $\zeta_\tau^\gamma = 1$ if $\tau=\gamma$ and $\zeta_\tau^\gamma = 0$ otherwise. The vector $p_N^{\kappa,n}(t_i)$ can then be written as

$$p_N^{\kappa,n}(t_i) = \sum_{\gamma\in\Gamma_\kappa} p_{N,\gamma}^{\kappa,n}(t_i)\zeta^\gamma,$$

and by the recursive definition of $p_N^{\kappa,n}(t_i)$,

$$p_N^{\kappa,n}(t_{i+1}) = \sum_{\alpha\in J_N^n}\varphi_\alpha^\kappa(\Delta;p_N^{\kappa,n}(t_i))\xi_\alpha^i = \sum_{\gamma\in\Gamma_\kappa}\sum_{\tau\in\Gamma_\kappa}\varphi_\alpha^\kappa(\Delta;\zeta^\tau)p_{N,\tau}^{\kappa,n}(t_i)\xi_\alpha^i$$

so that

$$p_{N,\gamma}^{\kappa,n}(t_{i+1}) = \sum_{\gamma\in\Gamma_\kappa}\sum_{\tau\in\Gamma_\kappa} q_{\gamma\tau}^{\kappa,\alpha} p_{N,\tau}^{\kappa,n}(t_i)\xi_\alpha^i, \tag{4.48}$$

where $q_{\gamma\tau}^{\kappa,\alpha} = \varphi_{\alpha,\gamma}^{\kappa}(\Delta;\zeta^\tau)$ is the component of the vector $\varphi_\alpha^\kappa(\Delta;\zeta^\tau)$ in the direction of the basis vector ζ^γ. Formula (4.48) is the foundation of the algorithm for computing $p_N^{\kappa,n}(t_i)$ and the approximation $\tilde{\phi}_{t_i,\kappa}[f]$ of the unnormalized optimal filter. The following is the description of the algorithm.

1. Off line (before the observations are available):

for $\alpha \in J_N^n$ *and* $\gamma, \tau \in \Gamma_\kappa$ *compute*
$q_{\gamma\tau}^{\kappa,\alpha} = \varphi_{\alpha,\gamma}^{\kappa}(\Delta;\zeta^\tau)$, $f_\gamma = (f, e_\gamma)_0$, *and* $p_{N,\gamma}^{\kappa,n}(t_0) = (p_0, e_\gamma)_0$;

set $p_N^{\kappa,n}(t_0,x) = \sum_{\gamma\in\Gamma_\kappa} p_{N,\gamma}^{\kappa,n}(t_0)e_\gamma(x)$ *and* $\tilde{\phi}_{t_0,\kappa}[f] = \sum_{\gamma\in\Gamma_\kappa} f_\gamma p_{N,\gamma}^{\kappa,n}(t_0)$.

2. On line, $i-th$ step (as the observations become available): compute

$$p_{N,\gamma}^{\kappa,n}(t_i) = \sum_{\tau\in\Gamma_\kappa} Q_{\gamma\tau}^{\kappa}(\xi^i) p_{N,\tau}^{\kappa,n}(t_{i-1}), \quad \gamma\in\Gamma_\kappa, \tag{4.49}$$

where

$$Q_{\gamma\tau}^{\kappa}(\xi^i) = \sum_{\alpha\in J_N^n} q_{\gamma\tau}^{\kappa,\alpha}\xi_\alpha^i,$$

then, if necessary, compute

$$p_N^{\kappa,n}(t_i,x) = \sum_{\gamma\in\Gamma_\kappa} p_{N,\gamma}^{\kappa,n}(t_i)e_\gamma(x), \tag{4.50}$$

$$\tilde{\phi}_{t_i,\kappa}[f] = \sum_{\gamma\in\Gamma_\kappa} f_\gamma p_{N,\gamma}^{\kappa,n}(t_i), \tag{4.51}$$

and

$$\tilde{f}_{t_i,\kappa} = \frac{\tilde{\phi}_{t_i,\kappa}[f]}{\tilde{\phi}_{t_i,\kappa}[1]}. \tag{4.52}$$

This algorithm will be referred to as the **spectral separating scheme of the second kind** (or S_2^3 for short).

4.4.2. Remark. The main advantage of S_2^3 as compared to the existing schemes for solving the Zakai equation is that the time consuming computations, including solving partial differential equations and computing integrals, are performed off line, while the on-line part is relatively simple even when the dimension d of the state process is large. Here are some other features of the algorithm:

(1) The overall amount of the off-line computations does not depend on the number of the on-line time steps;

(2) Formulas (4.51) and (4.52) can be used to compute an approximation to $\hat{f}_{t_i}$ (e.g. conditional moments) without the time consuming computations of $p_N^{\kappa,n}(t_i,x)$ and the related integrals;

(3) Only the coefficients $p^{\kappa,n}_{N,\gamma}(t_i)$ must be computed at every time step while the approximate filter $\tilde{f}_{t_i,\kappa}$ and/or UFD $p^{\kappa,n}_N(t_i,x)$ can be computed as needed, e.g. at the final time moment.

(4) The on-line part of the algorithm can be easily parallelized.

(5) An obvious limitation is that all the parameters of the model must be known in advance. Unlike S^3_1, the spectral separating scheme of the second kind can be used when $\rho \neq 0$. Still, there is no substantiation of S^3_2 if the coefficients of the equations (4.31) are random and/or depend on time.

4.4.3. Remark. By Theorem 4.4.1, the algorithm converges in the limit $\lim\limits_{\kappa\to\infty}\lim\limits_{\Delta\to 0}$ for all positive integers N and n. If $n = 1$, then each ξ^i_α depends only on the increments $Y_l(t_i) - Y_l(t_{i-1})$ of the observation process. The question of approximating the integrals $\int_{t_{i-1}}^{t_i} m_k(s-t_{i-1})dY_l(s)$ for $k > 1$ and other computational aspects of the algorithm, including the complexity, are studied in Section 4.5.

4.5 Computational Aspects of the Spectral Separating Schemes

4.5.1 Approximation of the Wiener Integrals

If S^3_1 or S^3_2 is used with $n > 1$, then the integrals

$$\xi^i_{k,l} = \int_{t_{i-1}}^{t_i} m_k(s - t_{i-1})dY_l(s)$$

with

$$m_k(t) = \sqrt{\frac{2}{\Delta}}\cos\left(\frac{\pi(k-1)t}{\Delta}\right), \quad \Delta = t_i - t_{i-1}, \quad 1 < k \leq n,$$

must be computed approximately. As a result, the random variables

$$\xi^i_\alpha = \frac{1}{\sqrt{\alpha!}}\prod_{k,l} H_{\alpha^l_k}(\xi^i_{k,l})$$

are also computed with some error. The contribution of this error is studied below.

To approximate $\xi^i_{k,l}$, the integrals are reduced to the Riemann integrals by integrating by parts and the result is then approximated using the rectangular rule.

If

$$m'_k(t) = -\frac{\sqrt{2}\pi(k-1)}{\Delta^{3/2}}\sin\left(\frac{\pi(k-1)t}{\Delta}\right),$$

then

$$\xi^i_{k,l} = \sqrt{\frac{2}{\Delta}}\left((-1)^{k-1}Y_l(t_i) - Y_l(t_{i-1})\right) - \int_{t_{i-1}}^{t_i} m'_k(s - t_{i-1})Y_l(s)ds.$$

Let $t_{i-1} = t_{i,0} < \ldots < t_{i,L} = t_i$ be a uniform partition of the interval $[t_{i-1}, t_i]$ with step δ. Define the approximation $\bar{\xi}^i_{k,l}$ of $\xi^i_{k,l}$ by

$$\bar{\xi}^i_{k,l} = \sqrt{\frac{2}{\Delta}}\Big((-1)^{k-1}Y_l(t_i) - Y_l(t_{i-1})\Big) - \delta\sum_{j=0}^{L-1} m'_k(j\Delta)Y_l(t_{i,j})$$

and also

$$\bar{\xi}^i_\alpha = \frac{1}{\sqrt{\alpha!}}\prod_{k,l} H_{\alpha^l_k}(\bar{\xi}^i_{k,l}),$$

so that

$$\begin{aligned}\xi^i_\alpha - \bar{\xi}^i_\alpha = \sum_{k,l}\Big(H_{\alpha^l_k}(\xi^i_{k,l}) - H_{\alpha^l_k}(\bar{\xi}^i_{k,l})\Big)\mathcal{P}_{k,l}(\xi^i, \bar{\xi}^i) = \\ \sum_{k,l}(\xi^i_{k,l} - \bar{\xi}^i_{k,l})P_{k,l}(\xi^i, \bar{\xi}^i),\end{aligned}$$

where $\mathcal{P}_{k,l}(\xi^i, \bar{\xi}^i)$ and $P_{k,l}(\xi^i, \bar{\xi}^i)$ are polynomials in $\xi^i_{k,l}$ and $\bar{\xi}^i_{k,l}$ of degree at most $|\alpha|$. If $\alpha \in J^n_N$, then

$$\begin{aligned}\tilde{\mathbf{E}}|\xi_\alpha - \bar{\xi}_\alpha|^2 \le C(N,n)\sum_{k,l}\tilde{\mathbf{E}}\left|(\xi^i_{k,l} - \bar{\xi}^i_{k,l})P_{k,l}(\xi^i, \bar{\xi}^i)\right|^2 \le \\ C(N,n)\max_{k,l}\sqrt{\tilde{\mathbf{E}}|\xi^i_{k,l} - \bar{\xi}^i_{k,l}|^4}\ \max_{k,l}\sqrt{\tilde{\mathbf{E}}|P_{k,l}(\xi^i, \bar{\xi}^i)|^4},\end{aligned} \tag{4.53}$$

and the degree of $P_{k,l}$ does not exceed N. To estimate the right hand side of (4.53), the following result is used.

4.5.1. Lemma. *Assume that $f \in \mathbf{C}^2([T_0, T_1])$ is a deterministic function and $W = W(t)$ is a Wiener process on $(\Omega, \mathcal{F}, \mathbf{P})$. For a uniform partition $T_0 = t_0 < \ldots < t_L = T_1$ with step δ define*

$$I_f = f(T_1)W(T_1) - f(T_0)W(T_0) - \delta\sum_{i=0}^{L-1} f'(t_i)W(t_i).$$

Then

$$\mathbf{E}\left|\int_{T_0}^{T_1} f(t)dW(t) - I_f\right|^4 \le C(f_1\delta^2 + T_1^2 f_2\delta^4)(T_1 - T_0)^4,$$

where $f_1 = \max_{T_0 \le t \le T_1}|f'(t)|^4$, $f_2 = \max_{T_0 \le t \le T_1}|f''(t)|^4$, and C is an absolute constant.

Proof. Since f is continuously differentiable,

$$\int_{T_0}^{T_1} f(t)dW(t) = f(T_1)W(T_1) - f(T_0)W(T_0) - \int_{T_0}^{T_1} f'(t)W(t)dt.$$

By the Ito formula,

$$f'(t)W(t) = f'(t_i)W(t_i) + \int_{t_i}^{t} f''(s)W(s)ds + \int_{t_i}^{t} f'(s)dW(s),\ t_i \le t \le t_{i+1}.$$

Then

$$\int_{t_i}^{t_{i+1}} f'(t)W(t)dt = \delta f'(t_i)W(t_i) + \zeta_i + \eta_i,$$

where

$$\zeta_i = \int_{t_i}^{t_{i+1}} \int_{t_i}^{t} f'(s)dW(s)dt, \quad \eta_i = \int_{t_i}^{t_{i+1}} \int_{t_i}^{t} f''(s)W(s)dsdt,$$

so that

$$\mathbf{E}\left|\int_{T_0}^{T_1} f(t)dW(t) - I_f\right|^4 \leq (2L)^3 \left(\sum_{i=0}^{L-1} \mathbf{E}|\zeta_i|^4 + \sum_{i=0}^{L-1} \mathbf{E}|\eta_i|^4\right). \tag{4.54}$$

By direct computation,

$$\begin{aligned} \mathbf{E}|\zeta_i|^4 &\leq \delta^3 \int_{t_i}^{t_{i+1}} \mathbf{E}\left(\int_{t_i}^{t} f'(s)dW(s)ds\right)^4 dt \leq C\delta^3 \int_{t_i}^{t_{i+1}} \left(\mathbf{E}\left(\int_{t_i}^{t} f'(s)dW(s)ds\right)^2\right)^2 dt \\ &\leq C\delta^3 \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^{t} |f'(s)|^2 ds\right)^2 dt \leq Cf_1\delta^6, \end{aligned} \tag{4.55}$$

$$\begin{aligned} \mathbf{E}|\eta_i|^4 &\leq \delta^3 \int_{t_i}^{t_{i+1}} (t-t_i)^3 \int_{t_i}^{t} \mathbf{E}|f''(s)W(s)|^4 dsdt \\ & Cf_2\delta^7 \int_{t_i}^{t_{i+1}} s^2 ds \leq Cf_2\delta^8 T_1^2, \end{aligned} \tag{4.56}$$

and since $L = (T_1 - T_0)/\delta$, the statement of the lemma follows from (4.54)–(4.56).

□

4.5.2. Remark. Higher order approximations that are used for deterministic integrals (trapezoidal rule, Simpson's rule, etc.) will not improve the asymptotic quality of I_f because the integrand $f(t)W(t)$ is not smooth enough.

4.5.3. Theorem. *If $n > 1$ and $N > 1$ are fixed, then*

$$\max_{0 \leq i \leq M} \tilde{\mathbf{E}}|\xi_\alpha^i - \bar{\xi}_\alpha^i|^2 \leq \frac{C(N,n)\delta}{\Delta^{2N+2}}.$$

Proof. Use inequality (4.53). By the previous lemma, if $k \leq n$, then

$$\sqrt{\tilde{\mathbf{E}}|\xi_{k,l}^i - \bar{\xi}_{k,l}^i|^4} \leq C\left(\frac{n^4\delta^2}{\Delta^3} + \frac{n^2\delta}{\Delta}\right) \leq C(n)\frac{\delta}{\Delta^2}. \tag{4.57}$$

Indeed, $Y_l(t)$ is a Wiener process on $(\Omega, \mathcal{F}, \tilde{\mathbf{P}})$, $T_1 - T_0 = \Delta$,

$$|m_k'(t)|^2 \leq \frac{Cn^2}{\Delta^3}, \quad |m_k''(t)|^2 \leq \frac{Cn^4}{\Delta^5}.$$

Next, with respect to the measure $\tilde{\mathbf{P}}$, each of $\xi^i_{k,l}$ is a Gaussian random variable with zero mean and unit variance and $\bar{\xi}^i_{k,l}$ is a Gaussian random variable with zero mean and variance

$$\tilde{\mathbf{E}}|\bar{\xi}^i_{k,l}|^2 \leq \frac{C(n)}{\Delta}$$

so that for every integer q

$$\tilde{\mathbf{E}}|\xi^i_{k,l}|^q \leq \left(\tilde{\mathbf{E}}|\xi^i_{k,l}|^{2q}\right)^{1/2} \leq C(q), \qquad \tilde{\mathbf{E}}|\bar{\xi}^i_{k,l}|^q \leq \frac{C(n,q)}{\Delta^q}.$$

Consequently

$$\sqrt{\tilde{\mathbf{E}}|P_{k,l}(\xi^i,\bar{\xi}^i)|^4} \leq \frac{C(N,n)}{\Delta^{2N}},$$

which together with (4.57) proves the theorem.

□

4.5.4. Remark. The main asymptotic parameters of both S^3_1 and S^3_2 are Δ and κ. Approximation of the Wiener integrals introduces another asymptotic parameter, δ, which depends on Δ, N, and n. It follows from the discrete Gronwall lemma that if ξ^i_α is replaced by $\bar{\xi}^i_\alpha$, then the error bound in (4.16), (4.17), (4.34), and (4.35) will have an extra term

$$\frac{C(N,n)\delta}{\Delta^{2N+4}},$$

and this additional error can be controlled by choosing δ sufficiently small (after Δ has been chosen). It is also worth repeating that the approximation of ξ^i_α can be avoided altogether by choosing $n=1$, and the asymptotic rate of convergence in Δ is not affected by this choice of n. *To simplify the future presentation, the possible error due to the approximation of the Wiener integrals will not be considered.*

4.5.2 Choosing the Parameters

Both S^3_1 and S^3_2 have the same asymptotic parameters — κ and Δ — that assure the convergence of the algorithms. Still, all error bounds contain other parameters — N, n, and ν. It follows from (4.16), (4.17), (4.34), and (4.35) that, as long as $n \geq 1$ and $N \geq 2$, the particular choice of n and N does not change the rate of convergence. The construction of the algorithms also shows that the small values of N and n can improve the numerical stability by reducing the number of terms in the sums. Thus, $N=2$ and $n=1$ can be thought of as the typical values of these parameters.

On the other hand, ν is not used in the construction of the algorithms and the choice of ν is not so obvious: the larger the ν, the faster the convergence in κ, but the constant $C(\nu)$ also becomes large. The natural question to ask is the following: given $\varepsilon > 0$, what is the value of κ so that

$$\frac{C(\nu)}{\kappa^\nu} \leq \varepsilon$$

and the growth of κ as $\varepsilon \to 0$ is as slow as possible?

The error bound

$$err(\nu,\kappa) = \frac{C(\nu)}{\kappa^\nu} \quad \text{for every} \quad \nu > 0 \tag{4.58}$$

implies that, as $\varepsilon \to 0$, $\kappa(\varepsilon)$ will tend to infinity slower than any power of $1/\varepsilon$. The exact growth rate depends on how fast $C(\nu)$ grows with ν. Analysis of the proofs of Theorems 4.3.3 and 4.4.1 shows that

$$C(\nu) = C \max_{0\le i\le M} \tilde{\mathbf{E}} \|\Lambda^{\nu+\nu_0} p(t_i,\cdot)\|_0^2, \tag{4.59}$$

where ν_0 depends only on the parameters of the model.

To simplify the further analysis, assume that $r = d = 1$, $\rho = 0$, and the operator $\mathcal{L}$ is of the form

$$\mathcal{L}g(x) = ag''(x) + b(x)g'(x),$$

where $a > 0$ is a constant. The Zakai equation becomes

$$dp(t,x) = (ap_{xx}(t,x) - (b(x)p(t,x))_x)dt + h(x)p(t,x)dY(t). \tag{4.60}$$

Assume also that

$$p_0(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right), \tag{4.61}$$

and that the derivatives of b and h satisfy

$$\sup_{x\in\mathbf{R}} |b^{(k)}(x)| \le C^{k+1}k!, \quad \sup_{x\in\mathbf{R}} |h^{(k)}(x)| \le C^{k+1}k! \tag{4.62}$$

(here and below in this subsection, $g^{(k)}$ denotes the kth derivative of the function g with respect to x).

It follows by induction that

$$\Lambda^k = \sum_{i=0}^{2k} P_{2k-i}(x)\frac{d^i}{dx^i}\ ,$$

where P_{2k-i} is a polynomial of degree at most $2k - i$ with the coefficients not exceeding $4^k(2k)!$. Consequently, $|P_{2k-i}(x)| \le 2k \cdot 4^k(2k)!\ (1+x^2)^k$ and

$$\begin{aligned} \|\Lambda^k p(t,\cdot)\|_0^2 &\le 2k \sum_{i=0}^{2k} \|P_{2k-i}p^{(i)}(t,\cdot)\|_0^2 \\ &\le C^k k^2((2k)!)^2 \|p(t,\cdot)\|_{2k,2k}^2. \end{aligned} \tag{4.63}$$

Next, it follows from equation (4.60) that for every positive k, m

$$\begin{aligned} (1+x^2)^{m/2}p^{(k)}(t,x) &= (1+x^2)^{m/2}p_0^{(k)}(x) + \int_0^t (1+x^2)^{m/2} a p^{(k+2)}(s,x)ds - \\ \int_0^t (1+x^2)^{m/2}&(b(x)p(s,x))^{(k)}ds + \int_0^t (1+x^2)^{m/2}(h(x)p(s,x))^{(k)}dY(s). \end{aligned} \tag{4.64}$$

Assume that

$$\sup_{0\le t\le T}\tilde{\mathbf{E}}\|p^{(l)}(t,\cdot)\|_{0,n}\le e^{\bar{C}(l+m)^2}B_0(l,n) \tag{4.65}$$

for all $l\le k,\ n\le m$ and some constant $\bar{C}$ independent of l,n, where $B_0(l,n)$ is a bound on $\|p_0^{(l)}\|^2_{0,n}$:

$$\|p_0^{(l)}\|^2_{0,n}\le B_0(l,n).$$

Standard parabolic estimates imply that (4.65) is true for every *fixed* k,m; it will be shown by induction that it is possible to find a constant $\bar{C}$ independent of k and m so that (4.65) is true for *all* m and k.

By assumption (4.61),

$$\|p_0^{(l)}\|^2_{0,n}\le C^{l+n+1}(l+n)! = B_0(l,n)\le B_0(k,m). \tag{4.66}$$

Then (4.64), the Ito formula, and the induction assumption (4.65), together with the bounds (4.62) on the derivatives of b and h, imply

$$\begin{aligned}\tilde{\mathbf{E}}\|p^{(k)}(t,\cdot)\|^2_{0,m}\le C_1k\int_0^t\tilde{\mathbf{E}}\|p^{(k)}(s,\cdot)\|^2_{0,m}ds+B_0(k,m)\Bigg(1+\\ Cm^2\exp\left\{\bar{C}(k+m-1)^2\right\}+\sum_{l=1}^{k}\left(\frac{(k+1)!}{(k-l)!}\right)^2C^l\exp\left\{\bar{C}(k+m-l)^2\right\}\Bigg),\end{aligned} \tag{4.67}$$

where the constant C_1 is independent of k,m and can very independently of $\bar{C}$. Since e^{k^2} grows faster than $(k!)^n$ for every n, it follows from (4.67) and the Gronwall lemma that

$$\sup_{0\le t\le T}\tilde{\mathbf{E}}\|p^{(k)}(t,\cdot)\|^2_{0,m}\le\exp\left\{C_1k+C\ln m+\bar{C}(k+m-1)^2\right\}B_0(k,m).$$

This means that $\bar{C}$ should satisfy

$$C_1k+C\ln m+\bar{C}(k+m-1)^2\le\bar{C}(k+m)^2$$

or

$$2\bar{C}(k+m-1/2)\ge C_1k+C\ln m. \tag{4.68}$$

Given C_1 and C, it is clearly possible to choose $\bar{C}$ so that (4.68) holds for all $k,m\ge 1$.

Combining (4.63), (4.65), and (4.66) results in

$$\sup_{0\le t\le T}\|\Lambda^k p(t,\cdot)\|^2_0\le e^{Ck^2},\quad k\ge 1.$$

The computations also show that the slower growth, say $e^{Ck\ln k}$, cannot be achieved using the above method.

In view of (4.59), the error bound (4.58) can now be written as

$$err(\nu,\kappa)=\frac{e^{C(\nu+\nu_0)^2}}{\kappa^\nu}.$$

Minimizing the expression with respect to ν for fixed κ yields

$$\nu = \frac{\ln \kappa}{2C} - \nu_0.$$

Then the condition

$$err(\nu, \kappa) \leq \varepsilon$$

implies

$$\kappa > e^{C(\sqrt{\ln(1/\varepsilon)}+\nu_0)} \quad \text{or} \quad \ln \kappa \asymp \sqrt{\ln(1/\varepsilon)}.$$

Another way to express the result is

$$\kappa \asymp (1/\varepsilon)^{1/\nu}, \quad \text{where} \quad \nu \asymp \sqrt{\ln(1/\varepsilon)}.$$

Given $0 < \varepsilon \ll 1$, it is now possible to give the asymptotic values of Δ and κ which make the left hand sides of (4.16), (4.17), (4.34), and (4.35) less than ε. Assume that $N \geq 2$ and $n \geq 1$ are fixed.

Spectral Separating Scheme of the First Kind: since the left hand side of both (4.16) and (4.17) is bounded by $C(\nu)/(\kappa^\nu \Delta^2) + C\Delta^2$,

$$\Delta \asymp \sqrt{\varepsilon}, \qquad \kappa \asymp (1/\varepsilon^2)^{1/\nu}, \quad \text{where} \quad \nu \asymp \sqrt{\ln(1/\varepsilon)};$$

Spectral Separating Scheme of the Second Kind: assume that $\rho = 0$ and $r > 1$ so that the left hand side of both (4.34) and (4.35) is bounded by $C(\nu)/(\kappa^\nu) + C\kappa^2\Delta$. Then

$$\kappa \asymp (1/\varepsilon)^{1/\nu}, \quad \text{where} \quad \nu \asymp \sqrt{\ln(1/\varepsilon)}, \qquad \Delta \asymp \varepsilon/\kappa^2.$$

It is obvious that in the case of S_2^3 the required value of Δ is smaller than in the case of S_1^3, and is smaller yet if $\rho \neq 0$. The value of κ, on the other hand, can be expected to be larger for S_1^3. This means that S_3^1 provides better resolution in space and S_3^2 in time.

4.5.3 Computational Complexity

For fixed N, n, κ, and Δ, one on-line step of either S_1^3 or S_2^3 includes

1. Computation of the coefficients $\psi_\gamma(t_i; N, n, \kappa)$ or $p_{N,\gamma}^{\kappa,n}$ for $\gamma \in \Gamma_\kappa$;
2. Computation of the random variables ξ_α^i for $\alpha \in J_N^n$;
3. Possible evaluation of the approximate UFD $P_N^{n,\kappa}(t_i, x)$ or $p_N^{\kappa,n}(t_i, x)$;
4. Possible computation of the approximate optimal filter.

Computation of the coefficients $\psi_\gamma(t_i; N, n, \kappa)$ or $p_{N,\gamma}^{\kappa,n}$ for $\gamma \in \Gamma_\kappa$ involves $|J_N^n|$ matrix - vector multiplications and the same number of vector - scalar multiplications. Since the size of each vector is $K_\kappa = (\kappa+d)!/\kappa!d!$, the total number of operations to compute the coefficients is about $|J_N^n| K_\kappa^2$ flops. Here and below, a **flop** is a floating point operation consisting of one scalar multiplication and one scalar addition.

Computation of ξ_α is relatively simple. It can be easily parallelized or even implemented on the hardware level, and the number of operations is of order $|J^n_N|$, which is negligible compared to $|J^n_N|K^2_\kappa$.

Evaluation of the approximate UFD requires about $K_\kappa N_s$ flops, where N_s is the number of points on the corresponding spatial grid. Still, all available information about the UFD is contained in the coefficients $\psi_\gamma(t_i; N, n, \kappa)$ or $p^{\kappa,n}_{N,\gamma}$ for $\gamma \in \Gamma_\kappa$ and therefore this evaluation can be avoided altogether.

Computation of the approximate optimal filter requires about K_κ flops, which is again negligible compared to $|J^n_N|K^2_\kappa$. Thus, one on-line step of either S^3_1 or S^3_2 requires about $|J^n_N|K^2_\kappa$ flops, and the estimate is quite accurate. This number can be expected to be smaller for S^3_2 because the error bound shows that the algorithm requires smaller values of κ.

Assume now that for fixed $N \geq 2$, $n \geq 1$, and $T > 0$, a prescribed precision is to be achieved at the end of the interval $[0, T]$. As before, the precision is measured by the value ε of the $\tilde{\mathbf{E}}$-expectation of the square of the corresponding difference. Since the number of steps is $M = T/\Delta$, it follows that the total number of on-line operations is $N_{total} = |J^n_N|K^2_\kappa T/\Delta$. If $\rho = 0$ and $r > 1$, then, in the case of S^3_1, $\Delta \asymp \sqrt{\varepsilon}$ and in the case of S^3_2, $\Delta \asymp \varepsilon$, while K_κ grows slower than any power of $1/\varepsilon$ for both algorithms. This means that, on the fixed time interval, the total number of on-line operations required by S^3_2 can be expected to be larger than the corresponding number required by S^3_1.

4.5.4 Error Bounds on the Original Probability Space

So far the error analysis was carried out on the reference probability space $(\Omega, \mathcal{F}, \tilde{\mathbf{P}})$, where

$$\tilde{\mathbf{P}}(A) = \int_A Z_T^{-1} d\mathbf{P}$$

and

$$Z_T = \exp\left\{ \int_0^T h^*(X(s))dY(s) - \frac{1}{2}\int_0^T |h(X(s))|^2 ds \right\}.$$

The relation between the expectations $\mathbf{E}$ and $\tilde{\mathbf{E}}$ is given in the following lemma.

4.5.5. Lemma. *If the function h is bounded and $\tilde{\mathbf{E}}|\eta|^2 < \infty$, then*

$$\mathbf{E}|\eta| \leq e^{h_0 T/2}\sqrt{\tilde{\mathbf{E}}|\eta|^2},$$

where $h_0 = \sup_{x\in\mathbf{R}^d} |h(x)|^2$.

Proof. By the Cauchy - Schwartz inequality,

$$\mathbf{E}|\eta| = \tilde{\mathbf{E}}|\eta|Z_T \leq \sqrt{\tilde{\mathbf{E}}Z_T^2}\sqrt{\tilde{\mathbf{E}}|\eta|^2}.$$

Since $Y = Y(t)$ is a standard Wiener process on $(\Omega, \mathcal{F}, \tilde{\mathbf{P}})$,

$$\tilde{\mathbf{E}} Z_T^2 = \tilde{\mathbf{E}} \exp\left\{ 2\int_0^T h^*(X(t))dY(t) - \int_0^T |h(X(t))|^2 dt \right\} =$$
$$\tilde{\mathbf{E}}\left(\exp\left\{ \int_0^T |h(X(t))|^2 dt \right\} \exp\left\{ 2\int_0^T h^*(X(t))dY(t) - 2\int_0^T |h(X(t))|^2 dt \right\} \right) \le e^{h_0 T},$$

which competes the proof.

□

4.5.6. Corollary. *Both* S_1^3 *and* S_2^3 *converge in the mean with respect to the measure* $\mathbf{P}$.

Even though the UFD $p = p(t,x)$ is non-negative with probability one for all $t \in [0,T]$, $x \in \mathbf{R}^d$, both approximations $P_N^{n,\kappa}(t_i,x)$ and $p_N^{\kappa,n}(t_i,x)$ do not have this property and can be negative with positive probability. As a result, the value of $\bar{\phi}_{t_i,\kappa}[1]$ or $\tilde{\phi}_{t_i,\kappa}[1]$ can be zero, which can cause numerical instability when the approximate optimal filter is computed according to (4.30) or (4.52). On the other hand, since the (normalized) optimal filter is given by

$$\hat{f}_{t_i} = \frac{\phi_{t_i}[f]}{\phi_{t_i}[1]},$$

very small values of $\phi_{t_i}[1]$ can cause computational problems even if $\phi_{t_i}[1]$ is computed exactly and therefore it is desirable to have $\phi_{t_i}[1] > \epsilon$ for some $\epsilon > 0$. Define

$$\pi_\epsilon^i = \mathbf{P}(\phi_{t_i}[1] > \epsilon).$$

The number π_ϵ^i can be viewed as the probability of reliable work of the exact optimal filter.

Let $\hat{\phi}_{t_i}$ be some approximation of the unnormalized optimal filter. Consider

$$\hat{\pi}_\epsilon^i = \mathbf{P}(\hat{\phi}_{t_i}[1] > \epsilon),$$

the probability of reliable work of the approximate optimal filter. Elementary computations show that

$$\begin{aligned} &\hat{\pi}_\epsilon^i \ge \mathbf{P}\left(|\phi_{t_i}[1] - \hat{\phi}_{t_i}[1]| \le \epsilon,\ \phi_{t_i}[1] > 2\epsilon \right) \ge \\ &\mathbf{P}(\phi_{t_i}[1] > 2\epsilon) - \mathbf{P}(|\phi_{t_i}[1] - \hat{\phi}_{t_i}[1]| \ge \epsilon) \ge \pi_{2\epsilon}^i - \frac{\mathbf{E}|\phi_{t_i}[1] - \hat{\phi}_{t_i}[1]|}{\epsilon}. \end{aligned} \tag{4.69}$$

Consider the approximations $\bar{\phi}_{t_i,\kappa}$ and $\tilde{\phi}_{t_i,\kappa}$ that are produced by S_1^3 and S_2^3 respectively. Assume that for both algorithms $N \ge 2$ and $n \ge 1$ are fixed, $\rho = 0$, $r > 1$, and the random variables ξ_α^i are computed exactly. Then by (4.16), (4.34), and Lemma 4.5.5,

$$\mathbf{E}|\phi_{t_i}[1] - \bar{\phi}_{t_i,\kappa}[1]| \le \frac{C(\nu)}{\kappa^\nu\,\Delta} + C\Delta,$$

$$\mathbf{E}|\phi_{t_i}[1] - \tilde{\phi}_{t_i,\kappa}[1]| \le \frac{C(\nu)}{\kappa^\nu} + C\kappa\sqrt{\Delta}.$$

Inequality (4.69) then implies that for both S_1^3 and S_2^3 it is possible to choose the parameters κ and Δ so that the probability of the reliable work of the approximate filters is arbitrarily close to the probability of the reliable work of the exact filter.

4.5.5 Spectral Separating Schemes vs. Splitting-up Approximation

The splitting-up approximation is one of the well known time discretization algorithms for solving the Zakai equation. Assume that $0 = t_1 < \ldots < t_M = T$ is a uniform partition of the interval $[0, T]$ with step Δ. According to Florchinger and LeGland [18], the solution $p(t_i, x)$ of

$$\begin{aligned} dp(t,x) &= \mathcal{L}^* p(t,x)dt + \sum_{l=1}^{r} \mathcal{M}_l^* p(t,x) dY_l(t),\ 0 < t \leq T,\ x \in \mathbf{R}^d; \\ p(0,x) &= p_0(x) \end{aligned}$$

is approximated by

$$\begin{aligned} p^{sp-up}(t_0, x) &= p_0(x), \\ p^{sp-up}(t_{i+1}, x) &= \Phi_\Delta p^{sp-up}(t_i, \eta(x, t_{i+1})) G(x, t_{i+1}), \end{aligned} \tag{4.70}$$

where $x \in \mathbf{R}^d$, $\{\Phi_t\}_{t \geq 0}$ is the semigroup generated by the operator $\mathcal{L}^*$,

$$\begin{aligned} \eta(x, t_{i+1}) &= x - \rho(x)(Y(t_{i+1}) - Y(t_i) - h(x)\Delta) + R(x)\Delta, \\ R_i(x) &= \sum_{k=1}^{r} \sum_{j=1}^{d} \frac{\partial \rho_{ik}(x)}{\partial x_j} \rho_{jk}(x),\ i = 1, \ldots, d; \end{aligned}$$

$$\begin{aligned} G(x, t_{i+1}) &= \exp\Big\{ h^*(x)(Y(t_{i+1}) - Y(t_i)) - \tfrac{1}{2}|h(x)|^2 - \\ & A^*(x)(Y(t_{i+1}) - Y(t_i) - h(x)\Delta) + \bar{A}(x)\Delta + A_0(x)\Delta \Big\}, \\ & A_k(x) = \sum_{i=1}^{d} \frac{\partial \rho_{ik}(x)}{\partial x_i},\ k = 1, \ldots, r, \\ & \bar{A}(x) = \frac{1}{2} \sum_{k=1}^{r} \sum_{i,j=1}^{d} \frac{\partial \rho_{ik}(x)}{\partial x_j} \frac{\partial \rho_{jk}(x)}{\partial x_i}, \\ & A_0(x) = \sum_{k=1}^{r} \sum_{i,j=1}^{d} \frac{\partial^2 \rho_{jk}(x)}{\partial x_i \partial x_j} \rho_{ik}(x). \end{aligned}$$

If $\rho(x) = 0$, then (4.70) becomes

$$p^{sp-up,0}(t_{i+1}, x) = \Phi_\Delta \exp\left\{ h^*(x)(Y(t_{i+1}) - Y(t_i)) - \frac{1}{2}|h(x)|^2 \right\} p^{sp-up,0}(t_i, x).$$

It is proved in [18] that [2]

$$\max_{0 \leq i \leq M} \tilde{\mathbf{E}} \| p(t_i, x) - p^{sp-up}(t_i, x) \|_0^2 \leq C\Delta$$

[2]under an additional assumption that $\rho(x) = const$

and

$$\max_{0\le i\le M} \tilde{\mathbf{E}}\|p(t_i,x)-p^{sp-up,0}(t_i,x)\|_0^2 \le C\Delta^2.$$

Therefore the splitting-up approximation has the same rate of convergence, as $\Delta \to 0$, as the spectral separating schemes.

The main advantage of the splitting-up approximation is that $p^{sp-up}(t_{i+1},x)$ is positive for all t and x so that the algorithm is numerically more stable than S_1^3 or S_2^3. On the other hand, every on-line step of the splitting-up approximation includes solving the partial differential equation

$$\frac{\partial u(t,x)}{\partial t} = \mathcal{L}^* u(t,x) \tag{4.71}$$

with the initial condition depending on the computations from the previous step. So far, it is practically impossible to solve (4.71) on line if $d > 3$ because the required number of operations is too large. If N_s is the number of the points of the spatial grid, then the realistic lower bound on the number of operations to solve (4.71) is $N_s(\ln N_s)^{d-1}$. Indeed, assume that the equation is reduced to a system of linear equations, and the system is solved by an iterative procedure without preconditioning. The matrix of the system is of dimension $N_s \times N_s$, sparse and non-symmetric (since the operator $\mathcal{L}^*$ is not self-adjoint). Then one iteration requires about $C_d N_s$ flops, where $C_d > 1$ is a constant depending on d and on the particular numerical algorithm [3], and the total number of iterations is proportional to the condition number of the matrix [65]. For non-symmetric matrices, the condition number is proportional to at least $(\ln N_s)^{d-1}$ [3, 10]. Thus the total number of operations required to solve the equation is $C_d N_s(\ln N_s)^{d-1}$. All other computations can be done much faster, and so the total number of on-line operations per step is $N_{sp-up} > N_s(\ln N_s)^{d-1}$. If $d = 6$ and $N_s = 10^6$, i.e. only 10 points are taken in each direction, then the result is $N_{sp-up} > 5\cdot 10^{11}$.

It is worth mentioning that the theoretical bound on the number of operations required to solve (4.71) using the multi-grid methods is CN_s. On the other hand, according to [3], a *working* scheme for $d = 3$ requires $CN_s^{1.5}$ operations. Note also that the operator $\mathcal{L}^*$ is not self-adjoint and need not be uniformly elliptic, which further complicates the problem. Last but not least, the computation of the optimal filter requires the numerical evaluation of the integral

$$\int_{\mathbf{R}^d} p^{sp-up}(t_i,x) f(x)dx$$

for different functions $f = f(x)$, and this is also very difficult to do on line if d is large. In the spectral separating schemes, all these problems are dealt with off line and therefore S_1^3 and S_2^3 seem more appropriate for on-line implementation. The first numerical experiments [20] show that the implementation is possible.

The main features of S_1^3, S_2^3, and the splitting-up approximation are summarized below.

	S_1^3	S_2^3 $(\rho = 0,\ r > 1)$	Splitting-up $(\rho = 0)$
parameters	$\kappa, \quad \Delta$	$\kappa, \quad \Delta$	Δ
P - mean error ε	$\frac{C(\nu)}{\kappa^\nu \Delta} + C\Delta$	$\frac{C(\nu)}{\kappa^\nu} + C\kappa\sqrt{\Delta}$	$C\Delta$
on-line operations (per step)	$\lvert J_N^n \rvert \ K_\kappa^2$	$\lvert J_N^n \rvert \ K_\kappa^2$	$> N_s(\ln N_s)^{d-1}$
numerical values, $\kappa = 10$, $N_s = 10^d$, $\lvert J_N^n \rvert = 10$			
$d = 3$	$8.2 \cdot 10^5$	$8.2 \cdot 10^5$	$> 4.8 \cdot 10^4$
$d = 6$	$6.4 \cdot 10^8$	$6.4 \cdot 10^8$	$> 5.0 \cdot 10^{11}$
$d = 9$	$8.5 \cdot 10^{10}$	$8.5 \cdot 10^{10}$	$> 3.4 \cdot 10^{19}$

Chapter 5

Parameter Estimation for Stochastic Evolution Equations with Non-commuting Operators

5.1 Introduction

In this chapter, a parameter estimation problem is considered for a stochastic partial differential equation

$$du(t,x) + (\mathcal{A}_0 + \theta\mathcal{A}_1)u(t,x)dt = dW(t,x),\ 0 < t \le T\ x \in M;\ u(0,x) = 0, \tag{5.1}$$

where M is a compact smooth d - dimensional manifold (without boundary), $\mathcal{A}_0,\ \mathcal{A}_1$ are differential operators on M, $W = W(t,x)$ is a white noise process modeled by a cylindrical Brownian motion on M, and θ is the unknown parameter belonging to an open subset of the real line.

An estimate of θ is constructed using the finite dimensional projections of the solution u. Asymptotic properties of the estimate are studied as the dimension K of the projection tends to infinity while the length T of the time interval and the amplitude of noise remain constant.

In [27, 28], the maximum likelihood estimate (MLE) of θ was studied in the situation when equation (5.1) is considered on a bounded domain in $\mathbf{R}^d$ with zero boundary conditions and the operators $\mathcal{A}_0,\ \mathcal{A}_1$ are self-adjoint and elliptic with a common system of eigenfunctions. It was proved that the MLE of θ is consistent if and only if

$$order(\mathcal{A}_1) \ge \frac{1}{2}(order(\mathcal{A}_0 + \theta\mathcal{A}_1) - d), \tag{5.2}$$

in which case the estimate is also asymptotically normal and asymptotically efficient.

A similar problem was studied in [60], where the observations are in discrete time and the operators are not necessarily self-adjoint but still have a common system of eigenfunctions.

In this chapter, some of the results from [27, 28] are obtained for equation (5.1) under less restrictive assumptions on the operators. The main assumption used here is that the operators $\mathcal{A}_0$ and $\mathcal{A}_1$ are of different orders and the operator $\mathcal{A}_0 + \theta\mathcal{A}_1$ is elliptic for all admissible values of θ. The model is described in **Section 5.2** and the main results are presented in **Section 5.3**. If $\mathcal{A}_1$ is the leading operator, then the estimate of θ is consistent and asymptotically normal (as $K \to \infty$). On the other hand, if $\mathcal{A}_0$ is the leading operator, then the estimate of θ is consistent and asymptotically normal if condition (5.2) holds and operator $\mathcal{A}_1$ satisfies a certain non-degeneracy property. In particular, condition (5.2) is necessary for consistency. When (5.2) does not hold, the asymptotic shift of the estimate is computed. The proof of the main theorem about the consistency and asymptotic normality is given in **Section 5.5**.

In **Section 5.4** an example is presented, illustrating how the obtained results can be applied to the estimation of either thermodiffusivity or the cooling coefficient in the heat balance equation with a variable velocity field.

5.2 The Setting

Let M be a d-dimensional compact orientable $\mathbf{C}^\infty$ manifold with a smooth positive measure dx. If $\mathcal{L}$ is an elliptic positive definite self-adjoint differential operator of order $2m$ on M, then the operator $\Lambda = (\mathcal{L})^{1/(2m)}$ is elliptic of order 1 and generates the scale $\{\mathbf{H}^s\}_{s\in\mathbf{R}}$ of Sobolev spaces on M [40, 62]. All differential operators on M are assumed to be non-zero with real $\mathbf{C}^\infty(M)$ coefficients, and only real elements of $\mathbf{H}^s$ will be considered. The variable x will usually be omitted in the argument of functions defined on M.

In what follows, an alternative characterization of the spaces $\{\mathbf{H}^s\}$ will be used. By Theorem I.8.3 in [62], the operator $\mathcal{L}$ has a complete orthonormal system of eigenfunctions $\{e_k\}_{k\geq 1}$ in the space $L_2(M, dx)$ of square integrable functions on M. With no loss of generality it can be assumed that each $e_k(x)$ is real. Then for every $f \in L_2(M, dx)$ the representation

$$f = \sum_{k\geq 1} \psi_k(f) e_k$$

holds, where

$$\psi_k(f) = \int_M f(x) e_k(x) dx.$$

If $l_k > 0$ is the eigenvalue of $\mathcal{L}$ corresponding to e_k and $\lambda_k := l_k^{1/(2m)}$, then, for $s \geq 0$, $\mathbf{H}^s = \{f \in L_2(M, dx) : \sum_{k\geq 1} \lambda_k^{2s} |\psi_k(f)|^2 < \infty\}$ and for $s < 0$, $\mathbf{H}^s$ is the closure of $L_2(M, dx)$ in the norm $\|f\|_s = \sqrt{\sum_{k\geq 1} \lambda_k^{2s} |\psi_k(f)|^2}$. As a result, every element f of the space $\mathbf{H}^s$, $s \in \mathbf{R}$, can be identified with a sequence $\{\psi_k(f)\}_{k\geq 1}$ such that $\sum_{k\geq 1} \lambda_k^{2s} |\psi_k(f)|^2 < \infty$. The space $\mathbf{H}^s$, equipped with the inner product

$$(f, g)_s = \sum_{k\geq 1} \lambda_k^{2s} \psi_k(f) \psi_k(g), \;\; f, g \in \mathbf{H}^s, \tag{5.3}$$

is a Hilbert space.

A **cylindrical Brownian motion** $W = (W(t))_{0\leq t\leq T}$ on M is defined as follows: for every $t \in [0, T]$, $W(t)$ is the element of $\cup_s \mathbf{H}^s$ such that $\psi_k(W(t)) = w_k(t)$, where $\{w_k\}_{k\geq 1}$ is a collection of independent one dimensional Wiener processes on the given probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbf{P})$ with a right continuous filtration $\mathbb{F} = \{\mathcal{F}_t\}_{0\leq t\leq T}$. Since by Theorem

II.15.2 in [62] $\lambda_k \asymp k^{1/d},\ k \to \infty$,[1] it follows that $W(t) \in \mathbf{H}^s$ for every $s < -d/2$. Direct computations show that W is an $\mathbf{H}^s$ - valued Wiener process with the covariance operator Λ^{2s}. This definition of W agrees with the alternative definitions of the cylindrical Brownian motion [57, 63].

Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{N}$ be differential operators on M of orders $order(\mathcal{A})$, $order(\mathcal{B})$, and $order(\mathcal{N})$ respectively. It is assumed that $\max(order(\mathcal{A}), order(\mathcal{B}), order(\mathcal{N})) < 2m$.

Consider the random field u defined by the evolution equation

$$du(t) + [\theta_1(\mathcal{L} + \mathcal{A}) + \theta_2\mathcal{B} + \mathcal{N}]u(t)dt = dW(t),\ 0 < t \leq T,\ u(0) = 0. \tag{5.4}$$

Here $\theta_1 > 0,\ \theta_2 \in \mathbf{R}$, and the dependence of u and W on x and ω is suppressed.

If the trajectory $u(t),\ 0 \leq t \leq T$, is observed, then the following scalar parameter estimation problems can be stated:

1). estimate θ_1 assuming that θ_2 is known;

2). estimate θ_2 assuming that θ_1 is known.

5.2.1. Remark. The general model

$$du(t) + [\theta_1\mathcal{A}_0 + \theta_2\mathcal{A}_1 + \mathcal{N}]u(t)dt = dW(t),\ 0 < t \leq T,\ u(0) = 0$$

is reduced to (5.4) if operator $\theta_1\mathcal{A}_0 + \theta_2\mathcal{A}_1$ is elliptic of order $2m$ for all admissible values of parameters $\theta_1,\ \theta_2$ and $order(\mathcal{A}_0) \neq order(\mathcal{A}_1)$. For example, if $order(\mathcal{A}_1) = 2m$, then $\mathcal{L} = (\mathcal{A}_1 + \mathcal{A}_1^*)/2 + (c+1)I,\ \mathcal{A} = (\mathcal{A}_1 - \mathcal{A}_1^*)/2 - (c+1)I,\ \mathcal{B} = \mathcal{A}_0$, where c is the lower bound on eigenvalues of $(\mathcal{A}_1 + \mathcal{A}_1^*)/2$ and I is the identity operator. Indeed, by Corollary 2.1.1 in [40], if an operator $\mathcal{P}$ is of even order with real coefficients, then the operator $\mathcal{P} - \mathcal{P}^*$ is of lower order than $\mathcal{P}$.

Before discussing possible solutions to the above parameter estimation problems, it seems appropriate to mention the analytical properties of the field u. The following result will be used.

5.2.2. Lemma. *If $\mathcal{P}$ is a differential operator of order p on M, then for every $s \in \mathbf{R}$ there are positive constants C_1 and C_2, possibly depending on s, so that for every $f \in \mathbf{C}^\infty(M)$,*

$$((\mathcal{L} + \mathcal{P})f, f)_s \geq C_1\|f\|_{s+m}^2 - C_2\|f\|_s^2. \tag{5.5}$$

Proof. Clearly, $((\mathcal{L} + \mathcal{P})f, f)_s = (\mathcal{L}f, f)_s + (\mathcal{P}f, f)_s$. By the definition of the norm $\|\cdot\|_s$,

$$\|f\|_s^2 = (\Lambda^s f, \Lambda^s f)_0.$$

[1] Notation $a_k \asymp b_k,\ k \to \infty$ means

$$0 < c_1 \leq \liminf_k (a_k/b_k) \leq \limsup_k (a_k/b_k) \leq c_2 < \infty.$$

Since $\mathcal{L} = \Lambda^{2m}$,

$$(\mathcal{L}f, f)_s = \|f\|_{s+m}.$$

Next,

$$|(\mathcal{P}f, f)_s| = |(\Lambda^{s-m}\mathcal{P}f, \Lambda^{s+m}f)_0| \le \|f\|_{s+m}\ \|f\|_{s-m+p}.$$

If $p \le m$, then $\|f\|_{s-m+p} \le C\|f\|_s$ so that

$$|(\mathcal{P}f, f)_s| \le C\|f\|_{s+m}\ \|f\|_s \le C\epsilon\|f\|^2_{s+m} + C\epsilon^{-1}\|f\|^2_s,\ \epsilon > 0,$$

and (5.5) follows if ϵ is sufficiently small.

If $m < p < 2m$, then use the property of the Hilbert scale [37, Definition III.1.1], according to which

$$\|f\|_{s-p+m} \le \|f\|_{s+m}^{\frac{p-m}{m}}\ \|f\|_s^{\frac{2m-p}{m}},$$

and also the following inequality

$$|xy| \le \epsilon\ \frac{|x|^q}{q} + \epsilon^{-q'/q}\ \frac{|y|^{q'}}{q'},$$

which is valid for every $\epsilon > 0$ and $q, q' > 1$, $1/q + 1/q' = 1$. Taking $1/q = p/(2m)$, $1/q' = 1 - p/(2m)$ results in

$$|(\mathcal{P}f, f)_s| \le \|f\|^{2/q}_{s+m}\ \|f\|^{2/q'}_m \le \epsilon\ \frac{\|f\|^2_{s+m}}{q} + \epsilon^{-q/q'}\ \frac{\|f\|^2_s}{q'},$$

and (5.5) follows if ϵ is sufficiently small.

□

5.2.3. Remark. Inequality (5.5) is one of many forms of the Gårding inequality.

5.2.4. Theorem. *For every $s < -d/2$, equation (5.4) has a unique solution $u = u(t)$ so that*

$$u \in L_2(\Omega \times [0, T]; \mathbf{H}^{s+m}) \cap L_2(\Omega; \mathbf{C}([0, T]; \mathbf{H}^s)) \tag{5.6}$$

with

$$\mathbf{E} \sup_{t\in[0,T]} \|u(t)\|^2_s + \mathbf{E}\int_0^T \|u(t)\|^2_{s+m} dt \le CT \sum_{k\ge1} \lambda_k^{2s} < \infty. \tag{5.7}$$

Proof. By assumption, $\max(order(\mathcal{A}), order(\mathcal{B}), order(\mathcal{N})) < 2m$ and $\theta_1 > 0$. Then Lemma 5.2.2 implies that for every $s \in \mathbf{R}$ there exist positive constants C_1 and C_2 so that for every $f \in \mathbf{C}^\infty$

$$-((\theta_1(\mathcal{L} + \mathcal{A}) + \theta_2\mathcal{B} + \mathcal{N})f, f)_s \le -C_1\|f\|^2_{s+m} + C_2\|f\|^2_s,$$

which means that the operator $-(\theta_1(\mathcal{L} + \mathcal{A}) + \theta_2\mathcal{B} + \mathcal{N})$ is coercive in every normal triple $\{\mathbf{H}^{s+m}, \mathbf{H}^s, \mathbf{H}^{s-m}\}$. The statement of the theorem now follows if Theorem 2.2.4 is applied with $\mathbf{H} = \mathbf{H}^s$, $\mathcal{A} = -(\theta_1(\mathcal{L}+\mathcal{A})+\theta_2\mathcal{B}+\mathcal{N})$, $\mathcal{B}_l = 0$, $M(t) = W(t)$ (so that W is an $\mathbf{H}^s$-valued martingale for every $s < -d/2$).

□

5.3 The Estimate and Its Properties

Both parameter estimation problems for (5.4) can be stated as follows: estimate $\theta \in \Theta$ from the observations of

$$du^{\theta}(t) + (\mathcal{A}_0 + \theta \mathcal{A}_1)u^{\theta}(t)dt = dW(t). \tag{5.8}$$

Indeed, if θ_2 is known, then $\mathcal{A}_0 = \theta_2 \mathcal{B} + \mathcal{N}$, $\theta = \theta_1$, $\Theta = (0, +\infty)$, $\mathcal{A}_1 = \mathcal{L} + \mathcal{A}$ and if θ_1 is known, then $\mathcal{A}_0 = \theta_1(\mathcal{L} + \mathcal{A}) + \mathcal{N}$, $\theta = \theta_2$, $\Theta = \mathbf{R}$, $\mathcal{A}_1 = \mathcal{B}$. All main results will be stated in terms of (5.4), and (5.8) will play an auxiliary role.

It is assumed that the observed field u satisfies (5.8) for some unknown but fixed value θ^0 of the parameter θ. Depending on the circumstances, θ^0 can correspond to either θ_1 or θ_2 in (5.4), the other parameter being fixed and known. Even though the whole random field $u^{\theta^0}(t,x)$ is observed, the estimate of θ^0 will be computed using only finite dimensional processes $\Pi^K u^{\theta^0}$, $\Pi^K \mathcal{A}_0 u^{\theta^0}$, and $\Pi^K \mathcal{A}_1 u^{\theta^0}$. The operator Π^K used to construct the estimate is defined as follows: for every $f = \{\psi_k(f)\}_{k \geq 1} \in \cup_s \mathbf{H}^s$,

$$\Pi^K f = \sum_{k=1}^{K} \psi_k(f) e_k.$$

By (5.8),

$$d\Pi^K u^{\theta}(t) + \Pi^K(\mathcal{A}_0 + \theta \mathcal{A}_1)u^{\theta}(t)dt = dW^K(t), \tag{5.9}$$

where $W^K(t) = \Pi^K W(t)$. The process $\Pi^K u^{\theta} = (\Pi^K u^{\theta}(t), \mathcal{F}_t)_{0 \leq t \leq T}$ is finite dimensional, continuous in the mean, and Gaussian, but not, in general, a diffusion process because the operators $\mathcal{A}_0$ and $\mathcal{A}_1$ need not commute with Π^K. Denote by $\mathbf{P}^{\theta,K}$ the measure in $\mathbf{C}([0,T]; \Pi^K(\mathbf{H}^0))$, generated by the solution of (5.9). The measure $\mathbf{P}^{\theta,K}$ is absolutely continuous with respect to the measure $\mathbf{P}^{\theta^0,K}$ for all $\theta \in \Theta$ and $K \geq 1$. Indeed, denote by $\mathcal{F}_t^{K,\theta}$ the σ-algebra generated by $\Pi^K u^{\theta}(s)$, $0 \leq s \leq t$, and let $U_t^{\theta,K}(X)$ be the operator from $\mathbf{C}([0,T]; \Pi^K(\mathbf{H}^0))$ to $\mathbf{C}([0,T]; \Pi^K(\mathbf{H}^0))$ such that for all $t \in [0,T]$ and $\theta \in \Theta$,

$$U_t^{\theta,K}(\Pi^K u^{\theta}) = \mathbf{E}\left(\Pi^K(\mathcal{A}_0 + \theta\mathcal{A}_1)u^{\theta} | \mathcal{F}_t^{K,\theta}\right) \qquad (\mathbf{P}\text{–a.s.})$$

Then by Theorem 7.12 in [47] the process $\Pi^K u^{\theta}$ satisfies

$$d\Pi^K u^{\theta}(t) = U_t^{\theta,K}(\Pi^K u^{\theta})dt + d\tilde{W}^{\theta,K}(t),\ \Pi^K u^{\theta}(0) = 0,$$

where $\tilde{W}^{\theta,K}(t) = \sum_{k=1}^{K} \tilde{w}_k^{\theta}(t) e_k$ and $\tilde{w}_k^{\theta}(t)$, $k = 1, \ldots, K$, are independent one dimensional standard Wiener processes in general different for different θ. Since $(\Pi^K(\mathcal{A}_0 + \theta\mathcal{A}_1)u^{\theta}, W^K)$ is a Gaussian system for every $\theta \in \Theta$, it follows from Theorem 7.16 and Lemma 4.10 in [47] that

$$\frac{d\mathbf{P}^{\theta,K}}{d\mathbf{P}^{\theta^0,K}}(\Pi^K u^{\theta^0}) = \exp\left\{ \int_0^T \left(U_t^{\theta,K}(\Pi^K u^{\theta^0}) - U_t^{\theta^0,K}(\Pi^K u^{\theta^0}), d\Pi^K u^{\theta^0}(t)\right)_0 - \frac{1}{2}\int_0^T \left(\|U_t^{\theta,K}(\Pi^K u^{\theta^0})\|_0^2 - \|U_t^{\theta^0,K}(\Pi^K u^{\theta^0})\|_0^2\right) dt \right\}.$$

By definition, the maximum likelihood estimate (MLE) of θ^0 is then equal to $\arg\max_{\theta}\left(d\mathbf{P}^{\theta,K}/d\mathbf{P}^{\theta^0,K}\right)(\Pi^K u^{\theta^0})$, but since, in general, the functional $U_t^{\theta,K}(X)$ is not known explicitly, this estimate cannot be computed. The situation is much simpler if the operators $\mathcal{A}_0$ and $\mathcal{A}_1$ commute with Π^K so that $\Pi^K\mathcal{A}_i = \Pi^K\mathcal{A}_i\Pi^K$, $i=0,1$, and $U_t^{\theta,K}(X) = \Pi^K(\mathcal{A}_0 + \theta\mathcal{A}_1)X(t)$; in this case, the MLE $\hat{\theta}^K$ of θ^0 is computable and, as shown in [28],

$$\hat{\theta}^K = \frac{\int_0^T (\Pi^K\mathcal{A}_1 u^{\theta^0}(t), d\Pi^K u^{\theta^0}(t) - \Pi^K\mathcal{A}_0 u^{\theta^0}(t)dt)_0}{\int_0^T \|\Pi^K\mathcal{A}_1 u^{\theta^0}(t)\|_0^2 dt} \tag{5.10}$$

with the convention $0/0=0$.

Of course, expression (5.10) is well defined even when the operators $\mathcal{A}_0$ and $\mathcal{A}_1$ do not commute with Π^K, and if the whole trajectory u^{θ^0} is observed, then the values of $\Pi^K\mathcal{A}_0 u^{\theta^0}(t)$ and $\Pi^K\mathcal{A}_1 u^{\theta^0}(t)$ can be evaluated, making (5.10) computable. Even though (5.10) is not, in general, the maximum likelihood estimate of θ^0, it looks like a natural estimate to consider.

To simplify the notations, the superscript θ^0 will be omitted wherever possible so that $u(t)$ is the solution of (5.4) or (5.8), corresponding to the true value of the unknown parameter. To study the properties of (5.10), note first of all that for all sufficiently large K,

$$\mathbf{P}\{\int_0^T \|\Pi^K\mathcal{A}_1 u(t)\|_0^2 dt > 0\} = 1. \tag{5.11}$$

Indeed, by assumption, the operator $\mathcal{A}_1$ is not identical zero and therefore neither is $\Pi^K\mathcal{A}_1$ for all sufficiently large K. Then (5.11) is a direct consequence of the following result.

5.3.1. Lemma. *If $\mathcal{P}$ is a differential operator of order p on M, then*

$$\mathbf{P}\{\omega : \mathcal{P}u(t) = 0 \text{ for all } t\in[0,T]\} = 0. \tag{5.12}$$

Proof. On the set $\{\omega : \mathcal{P}u(t) = 0 \text{ for all } t\in[0,T]\}$,

$$\int_0^t \mathcal{P}[\theta_1(\mathcal{L}+\mathcal{A}) + \theta_2\mathcal{B}]u(s)ds = \mathcal{P}W(t),\quad 0\le t\le T, \tag{5.13}$$

and consequently, if $r+p+2m < -d/2$ and $0\neq f\in\mathbf{H}^r$, then

$$\int_0^t (\mathcal{P}[\theta_1(\mathcal{L}+\mathcal{A}) + \theta_2\mathcal{B}]u(s), f)_r ds = (W(t), \mathcal{P}^* f)_r.$$

According to (5.6), the left hand side of the last equality is a real valued process having, as a function of t, $\mathbf{P}$- a.s. bounded variation. On the other hand,

$$U(t) = \frac{(W(t), \mathcal{P}^* f)_r}{\|\Lambda^r\mathcal{P}^* f\|_r}$$

is a standard one dimensional Winer process and therefore has unbounded variation, as a function of t, with probability 1. This means that equality (5.13) is possible only on a set of $\mathbf{P}$ - measure 0.

□

It follows from (5.10) and (5.12) that

$$\hat{\theta}^K = \theta^0 + \frac{\int_0^T (\Pi^K \mathcal{A}_1 u(t), dW^K(t))_0}{\int_0^T \|\Pi^K \mathcal{A}_1 u(t)\|_0^2 dt} \qquad (\mathbf{P}\text{−a.s.}) \tag{5.14}$$

Representation (5.14) will be used to study the asymptotic properties of $\hat{\theta}^K$ as $K \to \infty$. To get a consistent estimate, it is intuitively clear that $\int_0^T \|\Pi^K \mathcal{A}_1 u(t)\|_0^2 dt$ should tend to infinity as $K \to \infty$, and this requires certain non-degeneracy of the operator $\mathcal{A}_1$.

5.3.2. Definition. A differential operator $\mathcal{P}$ of order p on M is called **essentially non-degenerate** if

$$\|\mathcal{P}f\|_s^2 \geq \varepsilon \|f\|_{s+p}^2 - L\|f\|_{s+p-\delta}^2 \tag{5.15}$$

for all $f \in \mathbf{C}^\infty(M)$, $s \in \mathbf{R}$, with some positive constants ε, L, δ.

If the operator $\mathcal{P}^*\mathcal{P}$ is elliptic of order $2p$, then, by Lemma 5.2.2, the operator $\mathcal{P}$ is essentially non-degenerate because in this case the operator $\mathcal{P}^*\mathcal{P}$ is positive definite and self-adjoint so that the operator $(\mathcal{P}^*\mathcal{P})^{1/(2p)}$ generates an equivalent scale of Sobolev spaces on M. In particular, every elliptic operator satisfies (5.15). Since, by Corollary 2.1.2 in [40], for every differential operator $\mathcal{P}$ the operator $\mathcal{P}^*\mathcal{P} - \mathcal{P}\mathcal{P}^*$ is of order $2p-1$, the operator $\mathcal{P}$ is essentially non-degenerate if and only if $\mathcal{P}^*$ is.

Let us now formulate the main result concerning the properties of the estimate (5.14). Recall that the observed field u satisfies

$$du(t) + [\theta_1(\mathcal{L} + \mathcal{A}) + \theta_2 \mathcal{B} + \mathcal{N}]u(t)dt = dW(t), \; 0 < t \leq T; \; u(0) = 0, \tag{5.16}$$

with one of $\theta_2 = \theta_2^0$ or $\theta_1 = \theta_1^0$ known. According to (5.14), the estimate of the remaining parameter is given by

$$\hat{\theta}_1^K = \frac{\int_0^T (\Pi^K(\mathcal{L} + \mathcal{A})u(t), d\Pi^K du(t) - d\Pi^K(\theta_2^0 \mathcal{B} + \mathcal{N})u(t))_0}{\int_0^T \|\Pi^K(\mathcal{L} + \mathcal{A})u(t)\|_0^2 dt}, \tag{5.17}$$

$$\hat{\theta}_2^K = \frac{\int_0^T (\Pi^K \mathcal{B}u(t), d\Pi^K du(t) - d\Pi^K(\theta_1^0(\mathcal{L} + \mathcal{A}) + \mathcal{N})u(t))_0}{\int_0^T \|\Pi^K \mathcal{B}u(t)\|_0^2 dt}. \tag{5.18}$$

5.3.3. Theorem. *Assume that equation (5.16) is considered on a compact d-dimensional smooth manifold M, $\theta_1^0 > 0$, $\mathcal{L}$ is a positive definite self-adjoint elliptic operator of order $2m$, and*

$$\max(order(\mathcal{A}), order(\mathcal{B}), order(\mathcal{N})) < 2m.$$

In the case θ_2 is known, the estimate (5.17) of θ_1^0 is consistent and asymptotically normal:

$$\mathbf{P} - \lim_{K \to \infty} |\hat{\theta}_1^K - \theta_1^0| = 0;$$
$$\Psi_{K,1}(\theta_1^0 - \hat{\theta}_1^K) \xrightarrow{d} \mathcal{N}(0,1),$$

where $\Psi_{K,1} = \sqrt{(T/(2\theta_1^0))\sum_{n=1}^K l_n}$.

In the case θ_1 *is known, the estimate (5.18) of* θ_2^0 *is consistent and asymptotically normal if the operator* $\mathcal{B}$ *is essentially non-degenerate and order*$(\mathcal{B}) = b \geq m - d/2$. *In that case,*

$$\mathbf{P} - \lim_{K\to\infty} |\hat{\theta}_2^K - \theta_2^0| = 0;$$
$$\Psi_{K,2}(\theta_2^0 - \hat{\theta}_2^K) \xrightarrow{d} \mathcal{N}(0,1),$$

where $\Psi_{K,2} \asymp \sqrt{\sum_{n=1}^K l_n^{(b-m)/m}}$.

The proof of this theorem is rather long and technical and is given below in Section 5.5.

5.3.4. Remark.

1. Since $l_k \asymp k^{2m/d}$, the rate of convergence for $\hat{\theta}_1^K$ is $\Psi_{K,1} \asymp K^{m/d+1/2}$, and for $\hat{\theta}_2^K$, it is

$$\Psi_{K,2} \asymp \begin{cases} K^{(b-m)/d+1/2} & \text{if } b > m - d/2, \\ \sqrt{\ln K} & \text{if } b = m - d/2. \end{cases}$$

2. All the statements of the theorem remain true if, instead of differential operators, pseudo-differential operators of class $S^{\cdot}_{\rho,\delta}$ are considered with $\rho > \delta$ [40, 62].

5.3.5. Theorem. *If* θ_1^0 *is known and order*$(\mathcal{B}) = b < m - d/2$, *then the measures generated in* $\mathbf{C}([0,T];\mathbf{H}^s)$, $s < -d/2$, *by the solutions of (5.16) are equivalent for all* $\theta_2 \in \mathbf{R}$ *and*

$$\mathbf{P} - \lim_{K\to\infty} \hat{\theta}_2^K = \theta_2^0 + \frac{\int_0^T (\mathcal{B}u(t), dW(t))_0}{\int_0^T \|\mathcal{B}u(t)\|_0^2 dt}. \tag{5.19}$$

Proof. By (5.7),

$$\mathbf{E}\int_0^T \|\mathcal{B}u(t)\|_0^2 dt < \infty \tag{5.20}$$

for all $\theta_2 \in \mathbf{R}$, and therefore the stochastic integral $\int_0^T (\mathcal{B}u(t), dW(t))_0$ is well defined [57, 63]. Then (5.19) follows from (5.18) and the properties of the stochastic integral.

Next, denote by P^{θ_2} the measure generated in $\mathbf{C}([0,T];\mathbf{H}^s)$, $s < -d/2$, by the solution of (5.16) corresponding to the given value of θ_2. Inequality (5.20) implies that

$$\int_0^T \|\mathcal{B}u(t)\|_0^2 dt < \infty \quad (\mathbf{P}\text{−a.s.}) \tag{5.21}$$

and therefore by Corollary 1 in [57] the measures P^{θ_2} are equivalent for all $\theta_2 \in \mathbf{R}$ with the likelihood ratio

$$\begin{aligned} &\frac{d\mathbf{P}^{\theta_2}}{d\mathbf{P}^{\theta_2^0}}(u) = \\ &\exp\Big((\theta_2 - \theta_2^0)\int_0^T (\mathcal{B}u(t), dW(t))_0 - (1/2)(\theta_2 - \theta_2^0)^2 \int_0^T \|\mathcal{B}u(t)\|_0^2 dt\Big), \end{aligned} \tag{5.22}$$

where $u(t)$ is the solution of (5.16) corresponding to θ_2^0. Note that

$$\hat{\theta}_2 = \theta_2^0 + \frac{\int_0^T (\mathcal{B}u(t), dW(t))_0}{\int_0^T \|\mathcal{B}u(t)\|_0^2 dt}$$

maximizes the likelihood ration (5.22).

□

If the operators $\mathcal{A}$, $\mathcal{B}$, $\mathcal{N}$ have the same eigenfunctions as $\mathcal{L}$, then the coefficients $\psi_k(u(t))$ are independent (for different k) Ornstien-Uhlenbeck processes and $\Pi^K \mathcal{A}u(t) = \Pi^K \mathcal{A}\Pi^K u(t)$, with similar relations for $\mathcal{B}$ and $\mathcal{N}$. As a result, other properties of (5.17) and (5.18) can be established, including strong consistency and asymptotic efficiency [27, 28, 60], and, in the case of the continuous time observations, all estimates are computable explicitly in terms of $\psi_k(u(t)), k = 1, \dots, K$.

In general, computation of $\hat{\theta}_1^K$ and $\hat{\theta}_2^K$ using (5.17) and (5.18) respectively requires the knowledge of the whole field u rather than its projection. Still, the operators $\Pi^K(\mathcal{L} + \mathcal{A})$, $\Pi^K \mathcal{B}$, and $\Pi^K \mathcal{N}$ have finite dimensional range, which should make the computations feasible. Another option is to *replace* u by $\Pi^K u$. This can simplify the computations, but the result is, in some sense, even further from the maximum likelihood estimate, because some information is lost, and the asymptotic properties of the resulting estimate are more difficult to study. In general, the construction of the estimate depending *only* on the projection $\Pi^K u(t)$ is equivalent to the parameter estimation for a partially observed system with observations being given by (5.9). Without special assumptions on the operators $\mathcal{A}_0$ and $\mathcal{A}_1$, this problem is extremely difficult even in the finite dimensional setting.

5.4 An Example

Consider the following stochastic partial differential equation:

$$du(t, x) = (D\nabla^2 u(t, x) - (\vec{v}(x), \nabla)u(t, x) - \lambda u(t, x))dt + dW(t, x). \tag{5.23}$$

It is called the **heat balance equation** and describes the dynamics of the sea surface temperature anomalies [19]. In (5.23), $x = (x_1, x_2) \in \mathbf{R}^2$, $\vec{v}(x) = (v_1(x_1, x_2), v_2(x_1, x_2))$ is the velocity field of the top layer of the ocean (it is assumed to be known), D is thermodiffusivity, λ is the cooling coefficient. The equation is considered on a rectangle $|x_1| \le a$; $|x_2| \le c$ with periodic boundary conditions $u(t, -a, x_2) = u(t, a, x_2)$, $u(t, x_1, -c) = u(t, x_1, c)$ and zero initial condition. This reduces (5.23) to the general model (5.16) with M being a torus, $d = 2$, $\mathcal{L} = -\Delta = -\partial^2/\partial x_1^2 - \partial^2/\partial x_2^2$, $\mathcal{A} = 0$, $\mathcal{B} = I$ (the identity operator), $\mathcal{N} = (\vec{v}, \nabla) = v_1(x_1, x_2)\partial/\partial x_1 + v_2(x_1, x_2)\partial/\partial x_2$, $\theta_1 = D$, $\theta_2 = \lambda$. Then $order(\mathcal{L}) = 2$ (so that $m = 1$), $order(\mathcal{A}) = 0$, $order(\mathcal{B}) = 0$ (so that $b = 0$), and $order(\mathcal{N}) = 1$. The basis $\{e_k\}_{k \ge 1}$ is the suitably ordered collection of real and imaginary parts of

$$g_{n_1,n_2}(x_1, x_2) = \frac{1}{\sqrt{4ac}} \exp\left\{\sqrt{-1}\pi(x_1 n_1/a + x_2 n_2/c)\right\},\ n_1, n_2 \ge 0.$$

By Theorem 5.3.3, the estimate of D is consistent and asymptotically normal, the rate of convergence is $\Psi_{K,1} \asymp K$; the estimate of λ is also consistent and asymptotically normal with the rate of convergence $\Psi_{K,2} \asymp \sqrt{\ln K}$, since $b = 0 = m - d/2$ and (5.15) holds.

Unlike the case of the commuting operators, the proposed approach allows non-constant velocity field. Still, a significant limitation is that the value of $\vec{v}(x)$ must be known.

5.5 Proof of Theorem 5.3.3.

Hereafter, $u(t)$ is the solution of (5.16) corresponding to the true value of the parameters (θ_1^0 and θ_2^0) and C is a generic constant with possibly different values in different places.

The following result can be found in [27] and is essential for the proof.

5.5.1. Lemma. *If $\mathcal{P}$ is a differential operator on M and*

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \|\Pi^K \mathcal{P} u(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P} u(t)\|_0^2 dt} = 1, \tag{5.24}$$

then

$$\lim_{K\to\infty} \frac{\int_0^T (\Pi^K \mathcal{P} u(t), dW^K(t))_0 dt}{\sqrt{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P} u\|_0^2 dt}} = \mathcal{N}(0,1) \tag{5.25}$$

in distribution.

Proof.

If

$$M_t^K := \frac{\int_0^t (\Pi^K \mathcal{P} u(s), dW^K(t))_0 ds}{\sqrt{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P} u(s)\|_0^2 ds}},$$

then $(M_t^K, \mathcal{F}_t)_{0\le t\le T}$ is a continuous square integrable martingale with quadratic characteristic

$$\langle M^K \rangle_t = \frac{\int_0^t \|\Pi^K \mathcal{P} u(s)\|_0^2 ds}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P} u(t)\|_0^2 dt}.$$

By assumption, $\mathbf{P} - \lim_{K\to\infty} \langle M^K \rangle_T = 1$.

On the other hand, if $(w_1(t), \mathcal{F}_t)_{0\le t\le T}$ is a one dimensional Wiener process (e.g., $w_1(t) = \psi_1(W(t))$) and $M_t := w_1(t)/\sqrt{T}$, then $(M_t, \mathcal{F}_t)_{0\le t\le T}$ is a continuous square integrable martingale, $\langle M \rangle_T = 1$.

As a result,

$$\lim_{K\to\infty} M_T^K = M_T$$

in distribution by [32, Theorem VIII.4.17] or [46, Theorem 5.5.4(II)]. Since M_T is a Gaussian random variable with zero mean and unit variance, (5.25) follows.

□

Once (5.24) and (5.25) hold and

$$\lim_{K\to\infty} \mathbf{E}\int_0^T \|\Pi^K \mathcal{P} u\|_0^2 dt = +\infty,$$

the convergence

$$\mathbf{P}-\lim_{K\to\infty}\frac{\int_0^T(\Pi^K\mathcal{P}u, dW^K(t)_0 dt}{\int_0^T\|\Pi^K\mathcal{P}u\|_0^2dt}=0$$

follows. Thus, it suffices to establish (5.24) and compute the asymptotics of $\mathbf{E}\int_0^T\|\Pi^K\mathcal{P}u\|_0^2dt$ for a suitable operator $\mathcal{P}$.

If $\psi_k(t):=\psi_k(u(t))$, then (5.16) implies

$$d\psi_k(t)=-\theta_1^0 l_k\psi_k(t)-\psi_k\Big((\theta_1^0\mathcal{A}+\theta_2^0\mathcal{B}+\mathcal{N})u(t)\Big)dt+dw_k(t),\ \psi_k(0)=0.$$

According to the variation of parameters formula, the solution of this equation is given by $\psi_k(t)=\xi_k(t)+\eta_k(t)$, where

$$\xi_k(t)=\int_0^t e^{-\theta_1^0 l_k(t-s)}dw_k(s),$$
$$\eta_k(t)=-\int_0^t e^{-\theta_1^0 l_k(t-s)}\psi_k\Big((\theta_1^0\mathcal{A}+\theta_2\mathcal{B}+\mathcal{N})u(s)\Big)ds.$$

If $\xi(t)$ and $\eta(t)$ are the elements of $\cup_s\mathbf{H}^s$ defined by the sequences $\{\xi_k(t)\}_{k\geq 1}$ and $\{\eta_k(t)\}_{k\geq 1}$ respectively, then the solution of (5.16) can be written as $u(t)=\xi(t)+\eta(t)$.

The following technical result will be used in the future.

5.5.2. Lemma. *If $a>0$ and $f(t)\geq 0$, then*

$$\int_0^T\Big(\int_0^t e^{-a(t-s)}f(s)ds\Big)^2dt\leq\frac{\int_0^T f^2(t)dt}{a^2}.$$

Proof. Note that

$$\Big(\int_0^t e^{as}f(s)ds\Big)^2=2\int_0^t\int_0^s e^{as}e^{au}f(u)f(s)duds.$$

If $U:=\int_0^T\Big(\int_0^t e^{-a(t-s)}f(s)ds\Big)^2dt$, then direct computations yield:

$$\begin{gathered}U=2\int_0^T\int_0^t\int_0^s e^{-a(2t-s-u)}f(u)f(s)dudsdt=\\ 2\int_0^T\Big(\int_0^s\Big(\int_s^T e^{-2at}dt\Big)e^{au}f(u)du\Big)e^{as}f(s)ds=\\ \int_0^T\Big(\int_0^s a^{-1}(e^{-2as}-e^{2aT})\ e^{au}f(u)du\Big)e^{as}f(s)ds\leq\\ a^{-1}\int_0^T\Big(\int_0^s e^{-a(s-u)}f(u)du\Big)f(s)ds\leq\\ a^{-1}\Big(\int_0^T f^2(s)ds\Big)^{1/2}\Big(\int_0^T\Big(\int_0^s e^{-a(s-u)}f(u)du\Big)^2ds\Big)^{1/2}=\\ a^{-1}\Big(\int_0^T f^2(s)ds\Big)^{1/2}U^{1/2},\end{gathered}$$

and the result follows.

□

It is shown in the next lemma that, under certain conditions on the operator $\mathcal{P}$, the asymptotics of $\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt$ is determined by the asymptotics of $\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt$.

5.5.3. Lemma. *If $\mathcal{P}$ is an essentially non-degenerate operator of order p on M and $p \geq m - d/2$, then*

$$\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt \asymp \sum_{k=1}^{N} l_k^{(p-m)/m}, \ K \to \infty, \tag{5.26}$$

$$\lim_{K\to\infty} \frac{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\eta(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt} = 0, \tag{5.27}$$

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \|\Pi^K \mathcal{P}\eta(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt} = 0, \tag{5.28}$$

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt} = 1. \tag{5.29}$$

Proof.
Proof of (5.26). It follows from the independence of $\xi_k(t)$ for different k that

$$\mathbf{E}\sum_{k=1}^{K} |\psi_k(\mathcal{P}\xi(t))|^2 = \mathbf{E}\sum_{k=1}^{K} \Big| \sum_{n\geq 1} \xi_n(t)(e_n, \mathcal{P}^* e_k)_0 \Big|^2 =$$
$$\sum_{k=1}^{K}\sum_{n\geq 1} \frac{1}{2\theta_1^0 l_n}(1 - e^{-2\theta_1^0 l_n t})|(e_n, \mathcal{P}^* e_k)_0|^2.$$

Integration yields:

$$\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt = \sum_{k=1}^{K}\sum_{n\geq 1} \frac{1}{2\theta_1^0 l_n}\left(T - \frac{1}{2\theta_1^0 l_n}(1 - e^{-2\theta_1^0 l_n T})\right)|(e_n, \mathcal{P}^* e_k)_0|^2.$$

Since $l_k \to \infty$ and only asymptotic behavior, as $K \to \infty$, of all expressions is studied, it can be assumed that $1 - e^{-2\theta_1^0 l_k T} > 0$ for all k. Then the last inequality and the definition of the norm $\|\cdot\|_s$ imply

$$\frac{T}{2\theta_1^0}\sum_{k=1}^{K} \|\mathcal{P}^* e_k\|_{-m}^2 - C\sum_{k=1}^{K} \|\mathcal{P}^* e_k\|_{-2m}^2 \leq \mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt \leq$$
$$\frac{T}{2\theta_1^0}\sum_{k=1}^{K} \|\mathcal{P}^* e_k\|_{-m}^2.$$

Since $\mathcal{P}$ satisfies (5.15),

$$\|\mathcal{P}^*\psi_k\|_{-m}^2 \geq \varepsilon\|e_k\|_{p-m}^2 - K\|e_k\|_{p-m-\delta}^2 = \varepsilon\lambda_k^{2(p-m)}(1 - (K/\varepsilon)\lambda_k^{-2\delta}).$$

In addition, $\|\mathcal{P}^* e_k\|_r^2 \leq C\|e_k\|_{r+p}^2$ and $\lambda_k = l_k^{1/(2m)}$. The result (5.26) follows.

Proof of (5.27). By assumptions,

$$c := \max(order(\mathcal{A}), order(\mathcal{B}), order(\mathcal{N})) < 2m.$$

By Lemma 5.5.2,

$$\int_0^T |\eta_n(t)|^2 dt \le \frac{1}{(\theta_1^0 l_n)^2} \int_0^T |\psi_n\Big((\theta_1^0 \mathcal{A} + \theta_2^0 \mathcal{B} + \mathcal{N})u(t)\Big)|^2 dt,$$

which implies that for every $r \in \mathbf{R}$,

$$\sum_{n\ge1} \lambda_n^{2r} \int_0^T |\psi_n(\mathcal{P}\eta(t))|^2 dt \equiv \int_0^T \|\mathcal{P}\eta(t)\|_r^2 dt \le C \int_0^T \|\eta(t)\|_{r+p}^2 dt \equiv$$
$$\sum_n \lambda_n^{2(r+p)} \int_0^T |\eta_n(t)|^2 dt \le C \int_0^T \|u(t)\|^2_{r-2m+c+p}.$$

If $c_1 := 2m - c > 0$ and $r = -x$ where $x = \max(0, d/2 + c_1/2 + p + c - 3m)$, then $-x - 2m + c + p = m - d/2 - c_1/2$ and, by (5.6), $E\int_0^T \|u(t)\|^2_{-x-2m+c+p} < \infty$. As a result, since $\lambda_k \asymp k^{1/d}$,

$$\frac{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\eta(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt} = \frac{\sum_{n=1}^K \lambda_n^{-2x}\lambda_n^{2x}\mathbf{E}\int_0^T |\psi_n(\mathcal{P}\eta(t))|^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt} \le$$
$$\frac{CK^{2x/d}\sum_{n\ge1}\lambda_n^{-2x}\mathbf{E}\int_0^T |\psi_n(\mathcal{P}\eta(t))|^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt} \le \frac{CK^{2x/d}}{\sum_{k=1}^K \lambda_k^{2(p-m)}} \to 0 \text{ as } K \to \infty,$$

because if $p - m = -d/2$, then $d/2 + c_1/2 + p + c - 3m = -c_1/2 < 0$ so that $x = 0$, while for $p - m > -d/2$ the sum $\sum_{k=1}^N \lambda_k^{2(p-m)}$ is of order $N^{2(p-m)/d+1}$ and $2(p-m)/d + 1 > (d + 2(p-m) - c_1/2) = 2x/d$. This proves (5.27). Then (5.28) follows from (5.27) and the Chebychev inequality.

Proof of (5.29). There are two steps in the proof. Writing $X_K(t) := \|\Pi^K \mathcal{P}\xi(t)\|_0^2$, the first step is to show that, for all $t \in [0, T]$,

$$var(X_K(t)) \le C \sum_{k=1}^K \lambda_k^{4(p-m)}. \tag{5.30}$$

This will imply that (5.29) holds (the second step), i.e. that

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T X_K(t)dt}{\mathbf{E}\int_0^T X_K(t)dt} = 1.$$

1). If $X_K^M(t) := \sum_{k=1}^K |\sum_{n=1}^M \xi_n(t)(e_n, \mathcal{P}^* e_k)_0|^2$, then $X_K^M(t)$ is a quadratic form of the Gaussian vector $(\xi_1(t), \ldots, \xi_M(t))$. The matrix of the quadratic form is $A = [A_{nn'}]_{n,n'=1,\ldots,M}$ with

$$A_{nn'} = \sum_{k=1}^K (e_n, \mathcal{P}^* e_k)_0 (e_{n'}, \mathcal{P}^* e_k)_0,$$

and the covariance matrix of the Gaussian vector is

$$R = diag\left(\frac{1-e^{-2\theta_1^0 l_n t}}{2\theta_1^0 l_k},\ n=1,\ldots,M\right).$$

It is still assumed that $1-e^{-2\theta_1^0 l_k T}>0$ for all k.

Direct computations yield

$$\mathbf{E}X_K^M(t)=\sum_{k=1}^{K}\sum_{n=1}^{M}\frac{1}{2\theta_1^0 l_n}(1-e^{-2\theta_1^0 l_n t})|(e_n,\mathcal{P}^*e_k)_0|^2=trace(AR).$$

Analysis of the proof of (5.26) shows that for every $t\in[0,T]$ and $k=1,\ldots,K$ the series $\sum_{n\geq 1}\xi_n(t)(e_n,\mathcal{P}^*e_k)_0$ converges with probability one and in the mean square. Consequently,

$$\begin{aligned}&\lim_{M\to\infty}X_K^M(t)=X_K(t)\qquad(\mathbf{P}\text{-a.s.});\\&\lim_{M\to\infty}\mathbf{E}X_K^M(t)=\sum_{k=1}^{K}\sum_{n\geq 1}\mathbf{E}|\xi_n(t)|^2\ |(e_n,\mathcal{P}^*e_k)_0|^2=\mathbf{E}X_K(t).\end{aligned}\tag{5.31}$$

Next,

$$\begin{aligned}&var(X_K^M(t))=2trace((AR)^2)\leq C\sum_{n,n'}\frac{1}{l_nl_{n'}}A_{nn'}^2=\\&\sum_{k,k'=1}^{K}|(\tilde{\mathcal{P}}e_k,e_{k'})_0|^2\lambda_k^{4(p-m)}\leq\sum_{k=1}^{K}\|\tilde{\mathcal{P}}e_k\|_0^2\lambda_k^{4(p-m)}\leq C\sum_{k=1}^{K}\lambda_k^{4(p-m)},\end{aligned}$$

where $\tilde{\mathcal{P}}:=\mathcal{P}\Lambda^{-2m}\mathcal{P}^*\Lambda^{2(m-p)}$ is a bounded operator in $\mathbf{H}^0$. After that, inequality (5.30) follows from (5.31) and the Fatou lemma:

$$\begin{aligned}&var(X_K(t))=\mathbf{E}\lim_{M\to\infty}|X_K^M(t)|^2-|\mathbf{E}\lim_{M\to\infty}X_K^M(t)|^2=\\&\mathbf{E}\lim_{M\to\infty}|X_K^M(t)|^2-\lim_{M\to\infty}|\mathbf{E}X_K^M(t)|^2\leq\liminf_{M\to\infty}\mathbf{E}|X_K^M(t)|^2-\lim_{M\to\infty}|\mathbf{E}X_K^M(t)|^2\leq\\&\liminf_{M\to\infty}var(X_K^M(t))\leq C\sum_{k=1}^{K}\lambda_k^{4(p-m)}.\end{aligned}$$

2). If $Y_K:=\int_0^T(X_K(t)-\mathbf{E}X_K(t))dt/\mathbf{E}\int_0^T X_K(t)dt$ then

$$\frac{\int_0^T X_K(t)dt}{\mathbf{E}\int_0^T X_K(t)dt}=1+Y_K$$

and

$$\mathbf{E}Y_K^2\leq\frac{T\int_0^T(var(X_K(t))dt}{\left(\mathbf{E}\int_0^T X_K(t)dt\right)^2}\leq C\frac{\sum_{k=1}^K\lambda_k^{4(p-m)}}{\left(\sum_{k=1}^K\lambda_k^{2(p-m)}\right)^2}\to 0\quad\text{as } K\to\infty.$$

By the Chebychev inequality, $\mathbf{P}-\lim_{K\to\infty}Y_K=0$, which implies (5.29).

□

5.5.4. Corollary. *If $\mathcal{P}$ is an essentially non-degenerate operator of order p on M and $p \geq m - d/2$, then*

$$\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|^2 dt \asymp \frac{\varepsilon T}{2\theta_1^0}\sum_{k=1}^{K} l_k^{(p-m)/m}, \ K \to \infty, \tag{5.32}$$

and

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt} = 1. \tag{5.33}$$

Proof. By the inequality $|2xy| \leq \epsilon x^2 + \epsilon^{-1}y^2$, which holds for every $\epsilon > 0$ and every real x, y,

$$\begin{aligned}
&(1-\epsilon)\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt + (1-\frac{1}{\epsilon})\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\eta(t)\|_0^2 dt \leq \\
&\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt \leq \\
&(1+\epsilon)\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt + (1+\frac{1}{\epsilon})\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\eta(t)\|_0^2 dt.
\end{aligned}$$

Since ϵ is arbitrary, (5.32) follows from (5.27) and (5.26). After that, (5.33) follows from (5.29).

□

To prove the first part of Theorem 5.3.3, it now suffices to apply Lemma 5.5.1 and Corollary 5.5.4 with $\mathcal{P} = \mathcal{L} + \mathcal{A}$; the non-degeneracy condition (5.15) holds with $p = 2m$, $\varepsilon = 1$, $\delta = m - order(\mathcal{A})/2$, because

$$\|\mathcal{L}f\|_s = \|f\|_{s+2m}$$

and, since the order of the operator $\mathcal{A}^*\mathcal{L}$ is $4m - 2\delta$,

$$\begin{aligned}
&(\mathcal{A}^*\mathcal{L}f, f)_s = (\Lambda^{-(2m-\delta)}\mathcal{A}^*\mathcal{L}f, \Lambda^{2m-\delta}f)_s \leq \\
&\|\Lambda^{-(2m-\delta)}\mathcal{A}^*\mathcal{L}f\|_s \, \|\Lambda^{2m-\delta}f\|_s \leq C\|f\|_{s+2m-\delta}^2.
\end{aligned}$$

Similarly, the second part of the theorem follows with $\mathcal{P} = \mathcal{B}$; now (5.15) is assumed. Analysis of the proof shows that

$$\lim_{K\to\infty} \frac{\Psi_{K,2}^2}{\sum_{k=1}^K l_k^{(b-m)/m}} \geq \frac{\varepsilon T}{2\theta_1^0}.$$

□

Chapter 6

Directions of Future Research

The open questions concerning the Wiener chaos decomposition of the stochastic evolution equations include:

1. Derivation and analysis of the expansion for equations with unbounded diffusion operator (this corresponds to the Zakai equation for the filtering model with correlated noise or unbounded observation function).

2. Derivation and analysis of the expansion for equations with infinite dimensional noise. This expansion can lead to new parameter estimation algorithms.

3. Further analysis of the multiple integral expansion, in particular, derivation of explicit formulas for higher order approximation schemes.

4. Analysis of the filtering equation with unbounded coefficients (in addition to the observation function, this includes unbounded drift and diffusion coefficients in the state equation).

5. Derivation of the filtering algorithms for models other than diffusion.

There are a number of open questions concerning the spectral separating schemes for the diffusion filtering model:

1. Is it possible to modify the algorithms so that the approximate unnormalized filtering density remains non-negative?

2. How to organize the off-line computations and the storage of data?

3. What is the efficient way to perform the on-line computations?

The next obvious step in the investigation of the parameter estimation problem is to consider the multi-dimensional parameter case (with the commuting operators, this problem was studied in [27]). Other directions include nonparametric estimation, nonlinear equations, and partially observed systems. Some problems of target tracking require the solution of the filtering problem with unknown coefficients of the state equation. The corresponding algorithm will have to include efficient on-line procedures for both filtering and estimation.

Reference List

[1] R. A. Adams. *Sobolev Spaces.* Academic Press, New York, 1975.

[2] S. I. Aihara. Regularized Maximum Likelihood Estimate for an Infinite Dimensional Parameter in Stochastic Parabolic Systems. *SIAM Journal on Control and Optimization*, 30(4):745–764, 1992.

[3] O. Axelsson and V. A. Barker. *Finite Element Solution of Boundary Value Problems.* Academic Press, New York, 1984.

[4] H. T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems.* Birkhäuser, Boston, 1989.

[5] J.S. Baras. Real-time Architectures for the Zakai Equations and Applications. In E. Mayer-Wolf et al., editor, *Stochastic Analysis*, Academic Press, Boston, 1991.

[6] V. E. Benesh. Exact Finite-dimensional Filters for Certain Diffusions with Nonlinear Drift. *Stochastics*, 5:65–92, 1981.

[7] J. F. Bennaton. Discrete Time Galerkin Approximation to the Nonlinear Filtering Solution. *Journal of Mathematical Analysis and Applications*, 110(2):364–383, 1985.

[8] A. Bensoussan, R. Glowinski, and R. Rascanu. Approximations of the Zakai Equation by Splitting-up Method. *SIAM Journal on Control and Optimization*, 28(6):1420–1431, 1990.

[9] A. Budhiraja and G. Kallianpur. Approximations to the Solution of the Zakai Equations Using Multiple Wiener and Stratonovich Integral Expansions. *Stochastics and Stochastics Reports*, 1996. (To appear).

[10] Z. Cai and Weinan E. Hierarchical Method for Elliptic Problems Using Wavelet. *Computations in Applied Numerical Methods*, 8:819–825, 1992.

[11] R. H. Cameron and W. T. Martin. The Orthogonal Development of Nonlinear Functionals in a Series of Fourier-Hermite Functions. *Annals of Mathematics*, 48(2):385–392, 1947.

[12] S. R. Cavior. A Master Key for Ten Locks. *The College Mathematical Journal*, 27(1):33–37, 1996.

[13] J. M. C. Clark. The Design of Robust Approximations to the Stochastic Differential Equations of Nonlinear Filtering. In J. K. Skwirzynski, editor, *Communication Systems and Random Processes Theory*, pages 721–734. Sijthoff & Noordhooff, 1978.

[14] F. E. Daum. New Exact Nonlinear Filters. In J. C. Spall, editor, *Bayesian Analysis of Time Series and Dynamic Models*, pages 199–226, Marsel Dekker, Inc., New York, 1988.

[15] N. Dunford and J. T. Schwartz. *Linear Operators, Part I: General Theory.* John Willey & Sons, New York, 1988.

[16] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models.* Springer, New York, 1995.

[17] R. J. Elliott and R. Glowinski. Approximations to Solutions of the Zakai Filtering Equation. *Stochastic Analysis and Applications*, 7(2):145–168, 1989.

[18] P. Florchinger and F. LeGland. Time Discretization of the Zakai Equation for Diffusion Processes Observed in Correlated Noise. *Stochastics and Stochastics Reports*, 35(4):233–256, 1991.

[19] C. Frankignoul. SST Anomalies, Planetary Waves and RC in the Middle Rectitudes. *Reviews of Geophysics*, 23(4):357–390, 1985.

[20] C. P. Fung. *New Numerical Algorithms for Nonlinear Filtering.* PhD thesis, University of Southern California, Los Angeles, CA, 90089, November 1995.

[21] A. Germani and M. Piccioni. Semi-discretisation of Stochastic Partial Differential Equations on $\mathbf{R}^d$ by a Finite Element Technique. *Stochastics*, 23(2):131–148, 1988.

[22] D. Gottlieb and S. A. Orszag. *Numerical Analysis of Spectral Methods: Theory and Applications.* CBMS-NSF Regional Conference, Series in Applied Mathematics, Vol.26, 1977.

[23] T. Hida. *Brownian Motion.* Springer, New York, 1979.

[24] T. Hida, H-H. Kuo, J. Potthoff, and L. Sreit. *White Noise.* Kluwer Academic Publishers, Boston, 1993.

[25] E. Hille and R. S. Phillips. *Functional Analysis and Semigroups.* Amer. Math. Soc. Colloq. Publ., Vol. XXXI, 1957.

[26] M. Hübner, R. Khasminskii, and B. L. Rozovskii. Two Examples of Parameter Estimation. In Cambanis, Ghosh, Karandikar, and Sen, editors, *Stochastic Processes*, Springer, New York, 1992.

[27] M. Hüebner. *Parameter Estimation for SPDE's.* PhD thesis, University of Southern California, Los Angeles, CA, 90089, June 1993.

[28] M. Hüebner and B. Rozovskii. On Asymptotic Properties of Maximum Likelihood Estimators for Parabolic Stochastic PDE's. *Probability Theory and Related Fields*, 103:143–163, 1995.

[29] I. A. Ibragimov and R. Z. Khasminskii. *Statistical Estimation (Asymptotic Theory).* Springer, New York, 1981.

[30] K. Ito. Multiple Wiener integral. *Journal of the Mathematical Society of Japan*, 3:157–169, 1951.

[31] Kazufumi Ito. Approximation of the Zakai Equation for Nonlinear Filtering. *SIAM Journal on Control and Optimization*, 34(2):620–634, 1996.

[32] J. Jacod and A. N. Shiryayev. *Limit Theorems for Stochastic Processes.* Springer, Berlin, 1987.

[33] G. Kallianpur. *Stochastic Filtering Theory.* Springer, New York, 1980.

[34] R. E. Kalman and R. S. Bucy. New Results in Linear Filtering and Prediction Problems. *ASME Transactions, Seires D, Journal of Basic Engineering*, 83:95–108, 1961.

[35] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 1992.

[36] P. E. Kloeden, E. Platen, and H. Schurz. *Numerical Solution of SDE Through Computer Experiments.* Springer, Berlin, 1994.

[37] S. G. Krein, Ju. I. Petunin, and E. M. Semenov. *Interpolation of Linear Operators.* American Mathematical Society, Providence, Rhode Island, 1981.

[38] N. V. Krylov and B. L. Rozovskii. On Conditional Distributions of Diffusion Processes. *Mathematics of the USSR Izvestija*, 12(2):336–356, 1978.

[39] N. V. Krylov and A.J. Veretennikov. On Explicit Formula for Solutions of Stochastic Equations. *Mathematics of the USSR Sbornik*, 29(2):239–256, 1976.

[40] H. Kumano-go. *Pseudo-Differential Operators.* The MIT Press, Cambridge, Massachusetts, 1981.

[41] H. Kunita. Cauchy Problem for Stochastic Partial Differential Equations Arising in Nonlinear Filtering Theory. *System & Control Letters*, 1(1):37–41, 1981.

[42] H. J. Kushner. *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations.* Academic Press, New York, 1977.

[43] Yu. A. Kutoyants. *Parameter Estimation for Stochastic Processes.* Heldermann Verlag, Berlin, 1984.

[44] Yu. A. Kutoyants. *Identification of Dynamical Systems With Small Noise.* Kluwer Academic Publisher, Dordrecht, 1994.

[45] O. A. Ladyzhenskaia, V. A. Solonikov, and N. N. Ural'tseva. *Linear and Quasi-linear Equations of Parabolic Type.* American Mathematical Society, Providence, Rhode Island, 1968.

[46] R. Sh. Liptser and A. N. Shiryayev. *Theory of Martingales.* Kluwer Academic Publishers, Boston, 1989.

[47] R. Sh. Liptser and A. N. Shiryayev. *Statistics of Random Processes.* Springer, New York, 1992.

[48] J. T.-H. Lo and S.-K. Ng. Optimal Orthogonal Expansion For Estimation I: Signal in White Gaussian Noise. In R.S. Bucy and J.M.F Moura, editors, *Nonlinear Stochastic Problems*, pages 291–309, D. Reidel Publ. Company, 1983.

[49] W. Loges. Girsanov's Theorem in Hilbert Space and an Application to Statistics of Hilbert Space Valued Stochastic Differential Equations. *Stochastic Processes and their Applications*, 17:243–263, 1984.

[50] S. Lototsky, R. Mikulevicius, and B. L. Rozovskii. Nonlinear Filtering Revisited: A Spectral Approach. *SIAM Journal on Control and Optimization*, 1997. To appear.

[51] S. Lototsky and B. L. Rozovskii. Recursive Multiple Wiener Integral Expansion for Nonlinear Filtering of Diffusion Processes. In N. Gretsky, editor, *Festschrift in honor of M. M. Rao*, Marsel Dekker publ. company, New York, 1996. (To appear).

[52] S. Lototsky and B. L. Rozovskii. Recursive Nonliner Filter for a Continuous - Discrete Time Model: Separation of Parametrs and Observations. *IEEE Transactions on Automatic Control*, 1996. (Submitted).

[53] R. Mikulevicius and B. L. Rozovskii. Separation of Observations and Parameters in Nonlinear Filtering. In *Proceedings of the 32nd IEEE Conference on Decision and Control, Part 2, San Antonio, 1993*, pages 1564–1569. IEEE Control Systems Society, 1993.

[54] R. Mikulevicius and B. L. Rozovskii. Soft Solutions of Linear Parabolic SPDE's and the Wiener Chaos Expansion. In *Stochastic Analysis on Infinite Dimensional Spaces*, volume 310, 1994.

[55] R. Mikulevicius and B. L. Rozovskii. Fourier-Hermite Expansion for Nonlinear Filtering. In A. A. Novikov, editor, *Festschrift in honor of A. N. Shiryayev*, 1995.

[56] R. Mikulevicius and B. L. Rozovskii. Linear Parabolic Stochastic PDE's and Wiener Chaos. *SIAM Journal on Mathematical Analysis*, 1996. (Submitted).

[57] R. Mikulevicius and B.L. Rozovskii. Uniqueness and Absolute Continuity of Weak Solutions for Parabolic SPDE's. *Acta Applicandae Mathematicae*, 35:179–192, 1994.

[58] G. Milstein. *Numerical Integration of Stochastic Differential Equations.* Kluwer Academic Publishers, Dordrecht, 1995.

[59] D. Ocone. Multiple Integral Expansions for Nonlinear Filtering. *Stochastics*, 10:1–30, 1983.

[60] L. Piterbarg and B. Rozovskii. Maximum Likelihood Estimators in SPDE's under Discrete Observations in Time. *Mathematical Methods of Statistics*, 1996. (To appear).

[61] B. L. Rozovskii. *Stochastic Evolution Systems.* Kluwer Academic Publishers, 1990.

[62] M.A. Shubin. *Pseudodifferential Operators and Spectral Theory.* Springer, New York, 1987.

[63] J. B. Walsh. An Introduction to Stochastic Partial Differential Equations. In P. L. Hennequin, editor, *Ecole d'été de Probabilités de Saint-Flour, XIV*, volume 1180, pages 265–439, Springer, Berlin, 1984.

[64] E. Wong. Explicit Solutions to a Class of Nonlinear Filtering Problems. *Stochastics*, 16(5):311–321, 1981.

[65] D. M. Young. *Iterative Solution of Large Linear Systems.* Academic Press, New York, 1971.

[66] M. Zakai. On the Optimal Filtering of Diffusion Processes. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 11(3):230–243, 1969.

Index