2013 (Heisei 25)

Doctoral Thesis

Essays on the discretization of stochastic differential equations and their applications

Doctoral Program in Integrated Science and Engineering
Graduate School of Science and Engineering
Ritsumeikan University

Hideyuki Tanaka

Acknowledgement

I would like to express my deepest gratitude to my supervisor Professor Arturo Kohatsu-Higa, from whom I learn not only how to solve the mathematical problem but also how to find a problem worth tackling. I must thank him for all of his continuous supports, for example, organizing seminars and conferences, giving me a chance to make presentations, and discussing mathematics in English with many postdocs from foreign countries etc, which greatly improved this thesis.

I am grateful to Professor Jirô Akahori for his boundless support for almost ten years, as well as Professor Setsuro Fujiie for his kindness, Professor Masatoshi Fujisaki for his generous advice on the basics of nonlinear filtering theory, and Professor Takahiro Aoyama for his encouragement.

I would like to thank Professor Akihiko Takahashi in University of Tokyo, and also his student Toshihiro Yamada. A part of this thesis (Chapter 5) is inspired by their innovative ideas. I would also like to thank several practitioners in Tokyo for discussing problems from the practical point of view throughout my career.

I would like to give my sincere thanks to all my colleagues, Tomonori Nakatsu, Gô Yûki, Takafumi Amaba, and Hidemi Aihara for their kindness and many interesting and usuful comments. I would like to acknowledge numerous members of our probability group in Ritsumeikan University, in particular, Hoang-Long Ngo, Azmi Makhlouf and Libo Li, who improved my English, Shunsuke Iwamoto and Yuya Tanaka for helping me revise my knowledge of mathematics through earnest and rigorous discussion of undergraduate/graduate-level textbooks.

Finally I must gratefully acknowledge my parents, who sadly passed away two and five years ago, for letting me walk my own path.

Abstract

The discretization of stochastic differential equations (SDEs) has been very important in many applications such as mathematical finance and nonlinear filtering. The aim of this thesis is to establish methods to construct a higher-order (or high accuracy) discretization scheme for general SDEs.

In the first chapter we give an overview of this research and briefly review the mathematical idea discussed throughout the thesis.

In the next three chapters, we propose several techniques for the construction of higher-order weak approximations of SDEs. Chapter 2 is devoted to an operator approach, often called the operator splitting method, which helps us to construct a higher-order scheme and to determine the rate of convergence. The discussion includes the analysis of approximations of Lévy-driven SDEs. Chapter 3 reviews the cubature formulas introduced by Lyons-Victoir (2004) and their relation with the operator splitting method. In Chapter 4, we introduce a space-time discretization scheme which can be applied to the computation of conditional expectations appeared in pricing American options and forward-backward SDEs.

In Chapter 5, several strong convergence results of an accelerated numerical scheme applied to perturbed SDEs are shown. The scheme introduced here was originally analyzed by Takahashi-Yoshida (2005) for weak approximations. We study the scheme from the viewpoint of strong convergence and the multi-level Monte Carlo method.

Finally in Chapter 6, we study a discrete-time approximation scheme for the nonlinear filtering problem. Picard (1984) showed that the scheme is a first-order approximation scheme under suitable conditions. We discuss a rigorous error analysis of the scheme using various techniques in infinite dimensional spaces, and in particular give a generalization of Picard's result.

Contents

Chapter 1

Introduction

1.1 Overview

1.1.1 Weak approximation problem

The discrete-time approximation of stochastic differential equations (SDEs) plays a crucial role in many applications. Historically, the approximation problem of SDEs has been developed in the fields of multidimensional partial differential equations via the Kolmogorov equation, and nonlinear filtering with countinuous-time observations in the latter half of the 20th century. Over the past few decades, there has been more importance on the approximation problem due to the development of mathematical finance.

We classify types of the error of approximations of SDEs. Let X_t be the solution of SDE and $\bar{X}_t^n$ be its discrete-time approximation with n steps. The convergence of an approximation $\bar{X}_t^n$ is basically twofold. We call by strong convergence (strong approximation)

$$E[|X_T - \bar{X}_T^n|^p]^{1/p} \to 0,$$

and by weak convergence (weak approximation)

$$E[f(X_T)] - E[f(\bar{X}_T^n)] \to 0.$$

If we obtain the exact rate $O(\frac{1}{n^\alpha})$ for the convergence "$\to 0$", the index α is called the strong (resp. weak) rate of convergence.

Our main interest in the present thesis is the analysis of weak rate of convergence. The weak approximation problem of stochastic differential equations has been studied by many authors. The Euler-Maruyama scheme ([64]) for stochastic differential equations is of weak rate $O(\frac{1}{n})$, which has been shown by [90] for smooth f, and by [6] for irregular f. A higher-order scheme based on the Itô-Taylor expansion has been also analyzed in e.g. [45], and however, the scheme is not always implementable due to the Lévy area whose distribution is unknown.

Recently, motivated by the financial industry, a new method of approximations of stochastic differential equations with high accuracy has been required, since practitioners (quantitative analysts) are going to develop more complicated models to control the financial risk. For the reason, we are concerned with the following two methods in this thesis.

1. Higher-order weak approximation + QMC (or Lattice): Our goal is to find a discrete approximation $\bar{X}_t^n$ so that

$$E[f(X_T)] - E[f(\bar{X}_T^n)] = O\left(\frac{1}{n^\alpha}\right)$$

   with a higher-order rate α than standard methods (e.g. Euler-Maruyama). The second order ($\alpha = 2$) schemes for stochastic differential equations have been developed by Lyons and Victoir [63] and Ninomiya and Victoir [69]. Higher-order schemes reduce the number of time partition n that is required for a given accuracy. In other words, the dimension of the numerical integration $E[f(\bar{X}_T^n)]$ decreases and such situation is preferable to the Quasi Monte Carlo (QMC) method. The efficiency of higher-order schemes with QMC was studied in [69], [68]. To make n small is also important in another idea by using recombining tree or lattice, which is considered in [60], [92].

2. Multi-level Monte Carlo (MLMC): The multi-level Monte Carlo (MLMC) is a kind of Romberg's extrapolation method for L^2-error in order to reduce the computational cost of simulations of Wiener functionals (Giles [27]). The purpose of the method is the same with weak approximations, that is, to compute $E[f(X_T)]$ with high accuracy. However, the computational efficiency of the MLMC is based on the strong rate of convergence of the approximation $\bar{X}^n$ rather than the weak rate of convergence.

In the following, we introduce mathematical aspects of higher-order weak approximations in Section 1.1.2 and of multi-level Monte Carlo in Section 1.1.3.

1.1.2 Three approaches in analysis of weak convergence

Throughout the present thesis, we often apply three methods introduced below to the analysis of weak rate of convergence. To understand these methods is important in both of the construction of new approximation schemes and the precise error estimates for them.

Let us now describe the basic methodology of each method. In what follows, we denote a Markov process by X_t, its infinitesimal generator by $\mathcal{L}$, and the associated semigroup by P_t.

(i) Short time expansion of semigroup: The short time expansion of the semigroup P_t is the most fundamental property in order to construct discrete time approximations for X_t. The expansion is formally expressed as

$$P_t = I + t\mathcal{L} + \frac{t^2}{2}\mathcal{L}^2 + \frac{t^3}{6}\mathcal{L}^3 + \cdots,$$

which can be considered as the formal exponential map $\exp(t\mathcal{L})$. This expansion motivates us to construct an operator Q_t which has the same short time expansion with P_t up to higher-order terms and also has a stochastic representation $Q_t f(x) = E[f(\bar{X}_t(x))]$ with some stochastic process $\bar{X}_t(x)$ starting at x.

This idea has been applied to (ordinary) differential equations by [81], [82], [85], [86] and to stochastic differential equations by [69], [24], [94].

(ii) Distance between two generators: Let $\tilde{P}_t$ be another Markov semigroup and $\tilde{\mathcal{L}}$ be its generator. Then we can derive the following formula

$$(P_t - \tilde{P}_t)f = \int_0^t \tilde{P}_{t-s}(\mathcal{L} - \tilde{\mathcal{L}})P_s f ds,$$

which is shown by taking the derivative of $s \mapsto \tilde{P}_{t-s}P_s f$. Roughly speaking, this means

$$P_t - \tilde{P}_t = O(t) \times \text{ distance between two generators.}$$

From the above expression, it is possible to consider numerical approximations for X_t through the analysis of generators. This essential idea plays a key role in the approximation of Lévy processes and more generally, Lévy-driven stochastic differential equations ([94], [47], [67], [46]).

(iii) Duality approach: It seems impossible to apply the above two approaches to the situations at which we cannot use the Markov property, such as discrete-time approximations of stochastic delay equations ([17]) and of nonlinear filters ([73], [93]).

Instead of the Markov property, we can use a duality formula for Wiener functionals in the following sense. Let F be a Wiener functional with Itô's representation $F = E[F] + \int_0^T \xi_s dB_s$, where ξ_s is an adapted process with finite moments and B_s is a Brownian motion. Then for any adapted process θ_s with finite moments, we have

$$E\Big[F\int_{t_i}^{t_{i+1}} \theta_s dBs\Big] = E\Big[\int_{t_i}^{t_{i+1}} \xi_s\theta_s ds\Big].$$

The left hand side tells us that this seems to be of order $O(\sqrt{t_{i+1} - t_i})$ from $\int_{t_i}^{t_{i+1}} \cdot dBs$ as it is. However from the representation of the right hand side, we might obtain the better rate $O(t_{i+1} - t_i)$. If F is smooth in Malliavin's sense, this formula can

be understood as the following duality between the stochastic derivative D and the stochastic integral dB

$$E\Big[F\int_{t_i}^{t_{i+1}}\theta_s dBs\Big] = E\Big[\int_{t_i}^{t_{i+1}}(D_sF)\theta_s ds\Big].$$

Since the Markov property is not assumed for F, the above duality formula is applicable to many situations. For example, nonlinear filtering problems deal with a conditional expectation of the form $E[f(X_T)|\mathcal{F}_T]$ for a given filtration $(\mathcal{F}_t)_{t\geq 0}$, then we are not able to apply the approach (i) even if X is a Markov process.

1.1.3 Multi-level Monte Carlo and strong convergence

The multi-level Monte Carlo method for Wiener functionals is formulated as follows. Let us denote a stochastic process at time T by X_T and its discretizaion (e.g. Euler scheme) by $\bar{X}_T^n$, where n is the number of time partition which is proportional to computational time. Then we define the MLMC as the following decomposition of Monte Carlo sampling by $m+1$ terms:

$$\frac{1}{N_0}\sum_{i_0=1}^{N_0} f(\bar{X}_T^{n_0,i_0}) + \sum_{\ell=1}^{m}\frac{1}{N_\ell}\sum_{i_\ell=1}^{N_\ell}\Big(f(\bar{X}_T^{n_\ell,i_\ell}) - f(\bar{X}_T^{n_{\ell-1},i_\ell})\Big)$$

with $n_\ell < n_{\ell+1}$ and N_ℓ i.i.d. sampling of $\bar{X}_T^{n,i_\ell}$ for each $\ell = 0,1,\dots,m$. Here, $f(\bar{X}_T^{n_\ell,i_\ell})$ and $f(\bar{X}_T^{n_{\ell-1},i_\ell})$ should be simulated to be pathwisely close to each other. Clearly the expectation of this sampling coincides with $E[f(\bar{X}_T^{n_m})]$. Hence the bias in the sense of expectation is equal to that of the usual sampling $E[f(\bar{X}_T^{n_m})] \approx \frac{1}{N}\sum_{i=1}^{N} f(\bar{X}_T^{n_m,i})$. On the other hand, the MLMC has a different structure in terms of the bias of Monte Carlo simulation. Let us summarize the key points of MLMC:

- The total computational cost (time) is of the order $O(\sum_{\ell=0}^{m} n_\ell N_\ell)$.
- The computational cost for generating $f(\bar{X}_T^{n_\ell,i_\ell}) - f(\bar{X}_T^{n_{\ell-1},i_\ell})$ *increases* as ℓ increases.
- The bias of Monte Carlo for $\{f(\bar{X}_T^{n_\ell,i_\ell}) - f(\bar{X}_T^{n_{\ell-1},i_\ell})\}_{i_\ell=1,\dots,N_\ell}$ *decreases* as ℓ increases.

We might control the computational efficiency by choosing m and N_ℓ. The optimal choice of these numbers is based on the L^2-error of $f(\bar{X}_T^{n_\ell}) - f(\bar{X}_T^{n_{\ell-1}})$, and by the triangle inequality we have

$$\|f(\bar{X}_T^{n_\ell}) - f(\bar{X}_T^{n_{\ell-1}})\|_2 \leq \|f(X_T) - f(\bar{X}_T^{n_\ell})\|_2 + \|f(X_T) - f(\bar{X}_T^{n_{\ell-1}})\|_2.$$

If f is Lipschitz continuous, the above quantity is bounded by

$$\|X_T - \bar{X}_T^{n_\ell}\|_2.$$

Therefore one of our interests in the MLMC method is to show the strong rate of convergence of $X_T - \bar{X}_T^n$ in L^2 sense. This is the key idea of the MLMC method for Wiener functionals introduced in [27]. The MLMC is an alternative and powerful numerical technique for general path-dependent functionals $E[f((X_t)_{0\leq t\leq T})]$ of the solution of a stochastic differential equation (X_t), since a higher-order weak approximation scheme for such functionals has not been found.

1.2 Outline of the thesis

In this thesis, we are interested in the several approximation methods of stochastic differential equations described in the previous section. This thesis consists of five chapters. In view of mathematical analysis, Chapter 2, 3 and 4 are based on the higher-order weak approximations of SDEs with two approaches (i) and (ii) in Section 1.1.2. Chapter 5 focuses on the strong rate of convergence related to Section 1.1.3. Finally, Chapter 6 relies on a duality formula (iii) in Section 1.1.2. The outline of each chapter is as follows.

In Chapter 2, we present a general framework, often called the operator splitting method, based on semigroup expansions for the construction of higher-order discretization schemes and analyze its rate of convergence. The error analysis essentially follows from the approaches (i) and (ii) in Section 1.1.2. We also apply the framework to approximate general Lévy-driven stochastic differential equations

Chapter 3 is devoted to the concept of cubature formulas on Wiener space and their connection to splitting methods for noncommutative exponential maps. More specifically, the relation between some higher-order weak approximation schemes (such as the Ninomiya-Victoir scheme) and cubature formulas is shown.

Chapter 4 presents a new class of higher-order space-time discretization schemes for multidimensional diffusions via lattice systems which involve space interpolation techniques. The key idea is to combine the weak approximation approach for stochastic differential equations and some techniques on high-dimensional spaces to break the curse of dimensionality. The first objective in this chapter is to investigate the error estimates derived from short time asymptotics of certain semigroup operators, together with the discussion of numerical stability. As the second objective, several computational experiments for some derivative pricing models are presented in one and three dimensional settings.

In Chapter 5, we determine the strong rate of convergence for an accelerated Euler-Maruyama scheme applied to perturbed stochastic differential equations. The theoretical results can be applied to analyzing the MLMC method. Several numerical experiments for the SABR stochastic volatility model are presented in order to confirm the efficiency of the schemes.

In Chapter 6, we study the concept of nonlinear filtering problems and a discrete-time approximation applied to them. Time discretizations for nonlinear filtering problems are related to both of strong and weak approximations of stochastic differ-

ential equations. We propose a new method of proof for the convergence of approximate nonlinear filter analyzed by Jean Picard, and show a more general result than the original one. The analysis for the error estimate is based on a kind of duality approach (iii) introduced in Section 1.1.2. For the proof, we develop an analysis of Hilbert space valued functionals on Wiener space.

Chapter 2

Operator splitting method

This chapter is based on the paper by Tanaka and Kohatsu-Higa [94] published in *Annals of Applied Probability* and includes some improvements of the proofs of theorems therein. Recent developments in this topics can be found in Kohatsu-Higa and Tankov [47], Kohatsu-Higa, Ortiz-Latorre and Tankov [46], Ngo and Kohatsu-Higa [67].

2.1 Introduction

Weak approximation problems play an important role in the numerical calculation of $E[f(X_t(x))]$ where $X_t(x)$ is the solution of the stochastic differential equation (SDE for short)

$$X_t(x) = x + \int_0^t \tilde{V}_0(X_{s-}(x))ds + \int_0^t V(X_{s-}(x))dB_s + \int_0^t h(X_{s-}(x))dY_s. \qquad (2.1)$$

with smooth coefficients $\tilde{V}_0 : \mathbf{R}^N \to \mathbf{R}^N, V = (V_1, \ldots, V_d), h : \mathbf{R}^N \to \mathbf{R}^N \otimes \mathbf{R}^d$. Here B_t is a d-dimensional standard Brownian motion and Y_t is an d-dimensional Lévy process associated with the Lévy triplet $(b, 0, \nu)$ satisfying the condition and which has finite L^p-moment for every $p \in \mathbf{N}$.

Our purpose is to find a discretization scheme $(X_t^{(n)}(x))_{t=0,T/n,\ldots,T}$ for given $T > 0$ such that

$$|E[f(X_T(x))] - E[f(X_T^{(n)}(x))]| \leq \frac{C(T, f, x)}{n^m}.$$

We denote briefly by $E[f(X_T(x))] - E[f(X_T^{(n)}(x))] = O(1/n^m)$ the above situation, and say that $X_T^{(n)}$ is a m-th order discretization scheme for X_t or that $X_T^{(n)}$ is an approximation scheme of order m. The Euler scheme is a 1st order scheme, and has been studied by many researchers. Talay-Tubaro [90] shows the 1st order convergence of the Euler scheme and 2nd order convergence with the Romberg extrapolation for continuous diffusions. The fact that the convergence rate of the Euler scheme also holds for certain irregular functions f under a Hörmander type condition

has been proved by Bally-Talay [6] using Malliavin calculus. For the general Lévy driven case, the Euler-Maruyama scheme was first studied in Protter- Talay [75], see also Jacod-Protter [40] and Jacod et al. [39] (for smooth f). The Itô-Taylor (weak-Taylor) high order scheme is a natural extension of the Euler scheme although is hard to simulate due to the use of multiple stochastic integrals. A discussion on the Itô-Taylor scheme with the Romberg extrapolation can be found in Kloeden-Platen [45].

In the continuous diffusion case, some new discretization schemes (also called Kusuoka type schemes) which are of order $m \geq 2$ without the Romberg extrapolation have been introduced by Kusuoka [53], Lyons-Victoir [63], Ninomiya-Victoir [69], Ninomiya-Ninomiya [68], Fujiwara [25] ($m = 6$) and Oshima-Teichmann-Veluscek [72] (m: even). The rate of convergence of these schemes is closely related to the stochastic Taylor expansion, or series expansion of exponential maps on a noncommutative algebra.

The actual simulation is carried out using Monte Carlo methods. That is, one computes $\frac{1}{M}\sum_{i=1}^{M} f(X_T^{(n),i}(x))$ where $X_T^{(n),i}(x)$, $i = 1, ..., M$ denotes M i.i.d. copies of $X_T^{(n)}(x)$. Therefore, the final error of L^2-convergence is:

$$\frac{1}{M}\sum_{i=1}^{M} f(X_T^{(n),i}(x)) - E[f(X_T(x))] = O\left(\frac{1}{\sqrt{M}} + \frac{1}{n^m}\right).$$

Then the optimal asymptotic choice of n is $O(n^m) = O(\sqrt{M})$.

The goal of the present chapter is two-fold. First, we introduce a general framework to study weak approximation problems from the standpoint of operator (semigroup) expansions. That is given two processes that have equal semigroup expansions up to some order lead after composition to two processes that are closed in law. This goal is not new. In fact, using PDE techniques, Milshtein and Talay between others proved various weak approximation results. Although our proof is essentially the same it gives a new viewpoint that will help in defining new approximation schemes.

The next idea, is to decompose the generator associated with (2.1) in $d + 2$ components where each component is associated with each component of the driving process (the whole Lévy process is considered as one component). Then we prove that if each of these components is approximated with an error of order $m + 1$ then the composition gives an error of order m. In the particular case that each component can be characterized as the semigroup of a flow-type process then the composition leads to a composition-type approximation scheme.

Secondly, using the above strategy we provide approximations for solutions of (2.1). In particular, our approximations are valid for infinite activity Lévy processes Y. We prove that in fact, if one uses the Asmussen-Rosiński idea of approximating the jumps of size smaller than ε with a Brownian motion and we only simulate one jump of size bigger than ε per each time interval in the approximation is enough to provide a first order approximation procedure. Furthermore we give the necessary

estimate to determine ε as a function of n. For this approximation, we found it better to decompose the generator in $d+4$ components.

This chapter is organized as follows. In Section 2.2, we introduce the main example and the goal for the first part of this article in explicit mathematical terms. The general framework is introduced in Section 2.3. In Section 2.4 we give the results of convergence rates of numerical discretization schemes in the general framework. In Section 2.5, we give a general result that states how to recombine the approximations to coordinate processes in order to approximate the semigroup associated to (2.1). Finally, in Section 2.6 we approximate each coordinate process and in particular, we define approximation schemes for Lévy driven SDEs.

2.2 Weak approximation problem

In order to better understand the abstract formulation in Section 3, we introduce here our main example. Let (Y_t) be a d-dimensional Lévy process characterized by Lévy-Khintchin formula:

$$E[e^{i\langle\theta,Y_t\rangle}] = \exp t\left(i\langle\theta,b\rangle - \frac{\langle\theta,c\theta\rangle}{2} + \int_{\mathbf{R}_0^d}(e^{i\langle\theta,y\rangle} - 1 - i\langle\theta,\tau(y)\rangle)\nu(dy)\right) \tag{2.2}$$

where $b \in \mathbf{R}^d$, $c \in \mathbf{R}^d \otimes \mathbf{R}^d$ (symmetric, semi-positive definite) and ν is a Borel measure on $\mathbf{R}_0^d := \mathbf{R}^d \setminus \{0\}$ satisfying that $\int_{|y|\le 1}|y|^2\nu(dy) < \infty$, which is called the Lévy measure.

Throughout this chapter, we assume that

$$\int_{|y|>1}|y|^p\nu(dy) < \infty, \quad \text{for all } p \ge 1. \tag{2.3}$$

It is well known that (2.3) implies that $Y_t \in \cap_{p\ge1}L^p$ for all t. We also recall that τ is a truncation function (e.g. $\tau(y) = y1_{\{|y|\le1\}}$, the constant b and τ depend on each other). The triplet (b,c,ν) is called the Lévy triplet.

The Lévy driven stochastic differential equation is given by

$$X_t(x) = x + \int_0^t \tilde{V}_0(X_{s-}(x))ds + \int_0^t V(X_{s-}(x))dB_s + \int_0^t h(X_{s-}(x))dY_s \tag{2.4}$$

with smooth coefficients $\tilde{V}_0 : \mathbf{R}^N \to \mathbf{R}^N, V = (V_1,\dots,V_d), h : \mathbf{R}^N \to \mathbf{R}^N \otimes \mathbf{R}^d$ whose derivatives of any order (≥ 1) are bounded. Here B_t and Y_t are independent d-dimensional standard Brownian motion and Y_t is a d-dimensional Lévy process associated with the Lévy triplet $(b,0,\nu)$ satisfying the condition (2.3). Using general semimartingale theory (see [74]) we have that the above equation has a unique solution. We define $V_0 := \tilde{V}_0 - \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^N \frac{\partial V_i}{\partial x_j}V_i^{(j)}$. Then (2.4) can be rewritten

in the following Stratonovich form:

$$X_t(x) = x + \sum_{i=0}^{d} \int_0^t V_i(X_{s-}(x)) \circ dB_s^i + \int_0^t h(X_{s-}(x)) dY_s$$

where $B_t^0 = t$.

Before introducing the general framework of approximation, let us explain in mathematical terms the goal in this article. Our main example corresponds to the approximation of the semigroup P_t defined as the semigroup associated to the Markov process X_t:

$$P_t f(x) = E[f(X_t(x))]$$

where $f : \mathbf{R}^N \to \mathbf{R}$ is a continuous function with polynomial growth at infinity.

Let $Q_t \equiv Q_t^n$ be an operator such that the semigroup property is satisfied in $\{kT/n; k = 0, ..., n\}$. Assume that Q_t approximates P_t in the sense that it satisfies the local error estimate $(P_t - Q_t)f(x) = O(t^{m+1})$. Then using the semigroup property of both P_t and $(Q_{kT/n})$, we notice that

$$P_T f(x) - (Q_{T/n})^n f(x) = \sum_{k=0}^{n-1} (Q_{T/n})^k (P_{T/n} - Q_{T/n}) P_{T - \frac{k+1}{n}T} f(x).$$

Therefore if we have good norm estimates of $(Q_{T/n})^k$ and $P_{T-\frac{k+1}{n}T}$ in a sense to be defined later (in particular the norm estimates have to be independent of n) then we can expect that $(Q_{T/n})^n$ is an approximation of order m to P_T. Finally in order to be able to perform Monte Carlo simulations we assume that Q has a *stochastic representation*. That is, there exists a stochastic process $M = M_t(x)$ starting at x such that $Q_t f(x) = E[f(M_t(x))]$. Then clearly, we have the following representation.

$$Q_T f(x) = (Q_{T/n})^n f(x) = E[f(M_{T/n}^1 \circ \cdots \circ M_{T/n}^n(x))]$$

where $M_{T/n}^i$ are independent copies of $M_{T/n}$ and $\circ$ is defined as $(M_t^i \circ M_t^j)(x) := M_t^i(M_t^j(x))$.

The above ideas are well known and have been already used to achieve proofs of weak convergence (for historical references, see [45]). Nevertheless, it seems to us that this is the first time it appears in this general framework. For example, if we take $M_t(x) := x + \tilde{V}_0(x)t + V(x)B_t + h(x)Y_t$ for $d = 1$, one obtains the Euler-Maruyama scheme.

Next to further simplify the procedure to obtain approximations we write the operator P_t as a composition of $d+2$ operators as follows. First define the following stochastic processes $X_{i,t}(x)$, $i = 0, ..., d+1$, usually called coordinate processes,

which are the unique solutions of

$$X_{0,t}(x) = x + \int_0^t V_0(X_{0,s}(x))ds$$
$$X_{i,t}(x) = x + \int_0^t V_i(X_{i,s}(x)) \circ dB_s^i \;\; 1 \le i \le d$$
$$X_{d+1,t}(x) = x + \int_0^t h(X_{d+1,s-}(x))dY_s.$$

Then we define

$$Q_{i,t}f(x) := E[f(X_{i,t}(x))] \tag{2.5}$$

for continuous function $f : \mathbf{R}^N \to \mathbf{R}$ with polynomial growth at infinity.

For notational convenience we identify a smooth function $V : \mathbf{R}^N \to \mathbf{R}^N$ with a smooth vector field $\sum_{i=1}^N V^{(i)} \frac{\partial}{\partial x_i}$ on $\mathbf{R}^N$. Let us define (integro-)differential operators L_i acting on C^2 by

$$L_0 f(x) := (V_0 f)(x), \quad L_i f(x) := \frac{1}{2}(V_i^2 f)(x), \; 1 \le i \le d \tag{2.6}$$
$$L_{d+1} f(x) := \nabla f(x) h(x) b + \int (f(x + h(x)y) - f(x) - \nabla f(x) h(x) \tau(y)) \nu(dy).$$

It is well known that $L := \sum_{i=0}^{d+1} L_i$ is the generator of X and similarly L_i is the generator of $X_{i,t}$. Also $e^{tL} := P_t$ and $e^{tL_i} := Q_{i,t}$ respectively where we consider these expressions as exponential maps on a noncommutative algebra. One notices that these operators have the form

$$e^{tL} = \sum_{k=0}^{m} \frac{t^k}{k!} L^k + O(t^{m+1}) \tag{2.7}$$

$$e^{tL_i} = \sum_{k=0}^{m} \frac{t^k}{k!} L_i^k + O(t^{m+1}) \tag{2.8}$$

To approximate e^{tL}, we would like to find some combination of operators satisfying

$$e^{tL} - \sum_{j=1}^{k} \xi_j e^{t_{1,j} A_{1,j}} \cdots e^{t_{\ell_j,j} A_{\ell_j,j}} = O(t^{m+1}) \tag{2.9}$$

with some $t_{i,j} > 0$, $A_{i,j} \in \{L_0, L_1, \ldots, L_{d+1}\}$ and weights $\{\xi_j\} \subset [0,1]$ with $\sum_{j=1}^k \xi_j = 1$. This will correspond to an m-th order discretization scheme.

To find such schemes, one can perform formal Taylor expansions for e^{tA} in each of the terms in (2.9). We remark that the result (2.9) will follow directly from (2.7) and (2.8) independent of the specific form of the decomposition $L := \sum_{i=0}^{d+1} L_i$. This algebraic calculation has lead to the introduction of the following approximation

schemes
Ninomiya-Victoir (a):

$$\frac{1}{2}e^{\frac{t}{2}L_0}e^{tL_1}\cdots e^{tL_{d+1}}e^{\frac{t}{2}L_0}+\frac{1}{2}e^{\frac{t}{2}L_0}e^{tL_{d+1}}\cdots e^{tL_1}e^{\frac{t}{2}L_0} \tag{2.10}$$

Ninomiya-Victoir (b):

$$\frac{1}{2}e^{tL_0}e^{tL_1}\cdots e^{tL_{d+1}}+\frac{1}{2}e^{tL_{d+1}}\cdots e^{tL_1}e^{tL_0} \tag{2.11}$$

Splitting method:

$$e^{\frac{t}{2}L_0}\cdots e^{\frac{t}{2}L_d}e^{tL_{d+1}}e^{\frac{t}{2}L_d}\cdots e^{\frac{t}{2}L_0} \tag{2.12}$$

The semigroups generated by these operators have a probabilistic representation. For example, Ninomiya-Victoir (a) corresponds to

$$1_{U<\frac{1}{2}}X_{0,t/2}\circ X_{d+1,t}\cdots X_{1,t}\circ X_{0,t/2}(x)+1_{\frac{1}{2}\le U}X_{0,t/2}\circ X_{1,t}\cdots X_{d+1,t}\circ X_{0,t/2}(x)$$

where U is a uniform random variable taking values in $[0,1]$, independent of $X_{i,t}$. However, since a closed-form solution $X_{i,t}$ is not always available, one has to replace $X_{i,t}$ with other approximations of order $m+1$ so that the final approximation result remains unchanged. Nevertheless the fact that there is only one driving process simplifies this task. This problem will be discussed in Section 2.5.

2.3 Preliminaries

2.3.1 Notation and assumptions

In this section, we consider a general framework for weak approximations following the arguments in Section 2.2, without using the specific form of the operator. We first define the following functional spaces.

- $C_p^m \equiv C_p^m(\mathbf{R}^N)$: the set of C^m functions $f:\mathbf{R}^N\to\mathbf{R}$ such that for each multi-index α with $0\le|\alpha|\le m$, $|\partial_x^\alpha f(x)|\le C(\alpha)(1+|x|^p)$ for some positive constant $C(\alpha)$.

We also let $C_p\equiv C_p^0$. Let us define a norm on C_p^m by

$$\|f\|_{C_p^m}:=\inf\{C\ge 0:|\partial_x^\alpha f(x)|\le C(1+|x|^p),0\le|\alpha|\le m,x\in\mathbf{R}^N\}$$

where we denote $|\alpha|:=\alpha_1+\cdots+\alpha_N$ for $\alpha=(\alpha_1,\ldots,\alpha_N)\in\mathbf{Z}_+^N$.

- $C_p^{1,m}([0,T]\times\mathbf{R}^N)$, $m\ge 2$: the set of functions $f:[0,T]\times\mathbf{R}^N\to\mathbf{R}$ such that $s\mapsto f(s,x)$ is continuous differentiable for all $x\in\mathbf{R}^N$ and satisfies that $f(s,\cdot)$, $\partial_s f(s,\cdot)\in C_{p+2}^{m-2}$ with $\sup_{s\in[0,T]}(\|f(s,\cdot)\|_{C_p^m}+\|\partial_s f(s,\cdot)\|_{C_{p+2}^{m-2}})<\infty$.

From now on, we denote by $Q_t : \cup_{p\geq 0} C_p(\mathbf{R}^N) \to \cup_{p\geq 0} C_p(\mathbf{R}^N)$ a linear operator for $0 \leq t \leq T$.

Assumption $(\mathcal{M}_0)$ **.** If $f \in C_p$ with $p \geq 2$, then $Q_t f \in C_p$ and

$$\sup_{t\in[0,T]} \|Q_t f\|_{C_p} \leq K \|f\|_{C_p}$$

for some constant $K > 0$ independent of t. Futhermore, we assume $0 \leq Q_t f(x) \leq Q_t g(x)$ whenever $0 \leq f \leq g$.

We now introduce two assumptions which are highly related to the convergence rate of approximation schemes.

Assumption $(\mathcal{M})$ **.** Q_t satisfies $(\mathcal{M}_0)$, and for each $f_p(x) := |x|^{2p}$ $(p \in \mathbf{N})$,

$$Q_t f_p(x) \leq (1 + Kt) f_p(x) + K't \tag{2.13}$$

for some constant $K = K(T, p)$, $K' = K'(T, p) > 0$.

For $m \in \mathbf{N}$, $\delta_m : [0, T] \to \mathbf{R}_+$ denotes a decreasing function which satisfies

$$\limsup_{t\to 0+} \frac{\delta_m(t)}{t^{m-1}} = 0.$$

Usually, we have $\delta_m(t) = t^m$.

Assumption $\mathcal{R}(m, \delta_m)$ **.** For each $p \geq 2$, there exists a constant $q = q(m, p) \geq p$ and linear operators $e_k : C_p^{2k} \to C_{p+2k}$ $(k = 0, 1, \ldots, m)$ such that
(A): For every $f \in C_p^{2(m'+1)}$ with $1 \leq m' \leq m$, the operator Q_t satisfies

$$Q_t f(x) = \sum_{k=0}^{m'} (e_k f)(x) t^k + (\mathrm{Err}_t^{(m')} f)(x), \ t \in [0, T], \tag{2.14}$$

where $e_0 f = f$, $\mathrm{Err}_t^{(m')} f \in C_q$, and satisfies the following condition:
(B): If $f \in C_p^{m''}$ with $m'' \geq 2k$, then $e_k f \in C_{p+2k}^{m''-2k}$ and there exists a constant constant $K = K(T, m) > 0$ such that

$$\|e_k f\|_{C_{p+2k}^{m''-2k}} \leq K \|f\|_{C_p^{m''}} \ k = 0, 1, \ldots, m. \tag{2.15}$$

Furthermore if $f \in C_p^{m''}$ with $m'' \geq 2m' + 2$,

$$\|\mathrm{Err}_t^{(m')} f\|_{C_q} \leq \begin{cases} K t^{m'+1} \|f\|_{C_p^{m''}} & \text{if } m' < m \\ K t \delta_m(t) \|f\|_{C_p^{m''}} & \text{if } m' = m \end{cases}$$

for all $0 \leq t \leq T$.
(C): For every $1 \leq k \leq m$ and $j \geq 2k$, if $f \in C_p^{1,j}([0, T] \times \mathbf{R}^N)$, then $e_k f \in C_{p+2k}^{1,j-2k}([0, T] \times \mathbf{R}^N)$ and $\partial_s e_k f = e_k \partial_s f$.

Remark 2.3.1. The condition (C) is only used for the Romberg extrapolation which is discussed in Theorem 2.4.4.

In order to compare the finite power expansions of different operators, we introduce the following notation.

$$J_{\leq m}(Q_t) := \sum_{k=0}^{m} t^k e_k$$

$$J_m(Q) := e_m.$$

$J_{\leq m}(Q_t)$ is a linear operator, which is related to the series expansion of $t \mapsto e^{tL_i}$ (cf. Proposition 2.3.6). The following Lemma comprises some basic properties related to the above definition. The proof is straightforward.

Lemma 2.3.2. *The following properties are satisfied:*

$$\mathcal{R}(m+1, \delta_{m+1}) \;\Rightarrow\; \mathcal{R}(m, t^m)$$

$$\mathcal{R}(m, \delta_m) \;\Rightarrow\; \mathcal{R}(m, \tilde{\delta}_m)$$

whenever $\delta_m(t) \leq K\tilde{\delta}_m(t)$ *and* $\limsup_{t\to 0+} \tilde{\delta}_m(t)/t^{m-1} = 0$.
(i) Let $\{\xi_i\}_{1\leq i\leq \ell}$ *be deterministic positive constants with* $\sum_i \xi_i = 1$, *and assume* $(\mathcal{M})$ *for* $Q_t^{(i)}$ $(i = 1, \ldots, \ell)$. *Then* $\sum_{i=1}^{\ell} \xi_i Q_t^{(i)}$ *also satisfies* $(\mathcal{M})$.
(ii) Let $\{\xi_i\}_{1\leq i\leq \ell} \subset \mathbf{R}$ *and assume* $\mathcal{R}(m, \delta_m)$ *for* $Q_t^{(i)}$ $(i = 1, \ldots, \ell)$. *Then* $\sum_{i=1}^{\ell} \xi_i Q_t^{(i)}$ *also satisfies* $\mathcal{R}(m, \delta_m)$.

2.3.2 Properties of Lévy driven SDEs

We start with the differentiability properties of $X_t(x)$ in x. The following material can be found in [40], [39], [49], [74] and [75]. We quote them here for completeness.

Lemma 2.3.3. *There exists a version of* $X_t(x)$ *such that a map* $x \mapsto X_t(x)$ *is infinite times continuous differentiable almost surely and in the* L^p*-sense. Moreover, we have for* $p \geq 2$,

$$E[\sup_{0\leq t\leq T} |X_t(x)|^p] \leq C(p, T)(1 + |x|^p) \tag{2.16}$$

and

$$\sup_{x\in\mathbf{R}^N} E[\sup_{0\leq t\leq T} |\partial_x^\alpha X_t(x)|^p] < \infty \tag{2.17}$$

for any multi-index α *with* $|\alpha| \geq 1$.

Proposition 2.3.4. *Let* $f \in C_p^m$ *with* $p \geq 2$.
(i)Then $P_t f \in C_p^m$ *for all* $t \geq 0$ *and*

$$\sup_{t\in[0,T]} \|P_t f\|_{C_p^m} \leq C\|f\|_{C_p^m} \tag{2.18}$$

(ii) If $m \geq 2$, then $Lf \in C_{p+2}^{m-2}$ and

$$\|Lf\|_{C_{p+2}^{m-2}} \leq C\|f\|_{C_p^m}.$$

(iii) If $f \in C_p^{1,m}([0,T] \times \mathbf{R}^N)$, then $(\partial_t Lf)(t,x) = (L\partial_t f)(t,x)$

Proof. The proof of (i) follows by interchange of derivation and expectation together with the moment estimates in Lemma 2.3.3. Recall that $L = \sum_{i=0}^{d+1} L_i$ as defined in (2.6).

(ii) We only do the proof for L_{d+1} with $m = 2$. We have

$$\begin{aligned}
&\Big| \int (f(x + h(x)y) - f(x) - \nabla f(x)h(x)\tau(y))\nu(dy) \Big| \\
&= \Big| \int \nabla f(x)h(x)(y - \tau(y))\nu(dy) + \int \int_0^1 (1-\theta)\frac{d^2}{d\theta^2} f(x + \theta h(x)y)d\theta\nu(dy) \Big| \\
&\leq C\|f\|_{C_p^2}(1 + |x|^{p+2}).
\end{aligned}$$

□

Proposition 2.3.5. *Let $f \in C_p^2$. Then P_t and L are commutative and $u_f(t,x) := P_t f(x)$ is the solution of the integro-differential equation:*

$$\begin{cases} \frac{d}{dt} u_f(t,x) = Lu_f(t,x) \\ u_f(0,x) = f(x). \end{cases}$$

Proof. 1. We first prove that $t \mapsto P_t g(x)$ is continuous when $g \in C_p(\mathbf{R}^N)$. Note that

$$E[g(X_t(x)) - g(X_{t-}(x))] = 0$$

since $P(|Y_t - Y_{t-}| > 0) = 0$ for a fixed time t. By this and Lemma 2.3.3, we deduce the continuity of $P_t g$.

2. By using Itô's formula (see e.g. [38]), for $g \in C_p^2(\mathbf{R}^N)$,

$$g(X_s(x)) = g(x) + \int_0^s Lg(X_{u-}(x))du + M_s \tag{2.19}$$

where M_s stands for some local martingale. By using Lemma 2.3.3 again, M_s is a martingale and hence $E[M_s] = 0$. Taking expectations of the above equation (2.19), we can show the continuous differentiability of $t \mapsto P_t g(x)$ and $\frac{d}{dt} P_t g(x) = P_t Lg(x)$.

3. Apply the above calculation for $g = P_t f$ and take the derivative of $s \mapsto P_{t+s} f = P_s P_t f$ around $s = 0$. Then we conclude that

$$P_t Lf(x) = \frac{d}{dt} P_t f(x) = \lim_{s \to 0+} s^{-1}(P_s P_t f(x) - P_t f(x)) = LP_t f(x).$$

□

Let $f \in C_p^{2m+2}$. Then the commutativity of P_t and L implies that $L^m u_f$ $(= u_{L^m f})$ is differentiable in t and the solution to similar integro-differential equations. That is,

$$\begin{cases} \frac{d}{dt}(L^m u_f)(t,x) = L(L^m u_f)(t,x) \\ (L^m u_f)(0,x) = (L^m f)(x). \end{cases}$$

for each $m \geq 0$. Consequently, applying Taylor's expansion to u_f, we have

Proposition 2.3.6. *For $f \in C_p^{2m+2}$,*

$$P_t f(x) = \sum_{k=0}^{m} \frac{t^k}{k!} L^k f(x) + \int_0^t \frac{(t-s)^m}{m!} P_s(L^{m+1} f)(x) ds$$

Furthermore, if $f \in C_p^m$ with $m \geq 2$. Then $P_t f \in C_p^{1,m}$.

Summarizing this section, we have

Corollary 2.3.7. *$P_t f(x) = E[f(X_t(x))]$ and $Q_t^i f(x) = E[f(X_t^i(x))]$ $(i = 0, 1, \ldots, d+1)$ satisfy the conditions $(\mathcal{M})$ and $\mathcal{R}(m, t^m)$. That is, for $p \in \mathbf{N}$,*

$$E[|X_t(x)|^{2p}] \leq (1 + Kt)|x|^{2p} + K't$$

for some constant $K = K(T,p)$, $K' = K'(T,p) > 0$ and

$$J_{\leq m}(P_t) = \sum_{k=0}^{m} \frac{t^k}{k!} L^k$$

$$J_{\leq m}(Q_t^i) = \sum_{k=0}^{m} \frac{t^k}{k!} L_i^k$$

for any $m \in \mathbf{N}$.

2.4 Weak rate of convergence

In this section, we prove the rate of convergence for the approximating operator Q under the assumptions $(\mathcal{M})$, $\mathcal{R}(m, \delta_m)$. Throughout this section, we assume the following assumption.

Assumption $(\mathcal{M}_P)$ **.** For all $f \in C_p^m$, $m \geq 2$, $p \geq 1$ then $P_\cdot f \in C_p^{1,m}$ and furthermore the following property is satisfied for some positive constant C:

$$\sup_{t \in [0,T]} \|P_t f\|_{C_p^m} \leq C \|f\|_{C_p^m}$$

for all $f \in C_p^m$, $m \geq 0$, $p \geq 1$.

Remark 2.4.1. The above assumption is satisfied for $P_t f(x) := E[f(X_t(x))]$ under the assumptions in Section 2.2.

Theorem 2.4.2. *Assume* $(\mathcal{M})$ *and* $\mathcal{R}(m, \delta_m)$ *for* P_t *and* Q_t *with* $J_{\leq m}(P_t - Q_t) = 0$. *Then for any* $f \in C_p^{2(m+1)}$, *there exists a constant* $K = K(T, x) > 0$ *such that*

$$\left| P_T f(x) - (Q_{T/n})^n f(x) \right| \leq K \delta_m \Big(\frac{T}{n} \Big) \|f\|_{C_p^{2(m+1)}}. \tag{2.20}$$

For the proof, we need the following lemma.

Lemma 2.4.3. *Under assumption* $(\mathcal{M})$, *the operators* P_t *and* Q_t *satisfy*

$$\sup_n \max_{0 \leq k \leq n} Q_{T/n}^k f(x) < \infty$$

for any positive function $f \in C_p$ *with* $p \geq 0$.

Proof. Let $f_p(x) = |x|^{2p}$ for $p \in \mathbf{N}$. By the assumption $(\mathcal{M})$, we have

$$\begin{aligned} (Q_{T/n})^k f_p(x) &= (Q_{T/n})^{k-1} (Q_{T/n} f_p)(x) \\ &\leq (1 + \frac{C}{n}) (Q_{T/n})^{k-1} f_p(x) + \frac{C'}{n} \end{aligned}$$

with some constant C, C' independent of t, x, k, n. Since $(1 + \frac{C}{n})^k \leq e^C$, one proves by induction that

$$\sup_n \max_{0 \leq k \leq n} (Q_{T/n})^k f_p(x) \leq e^C C' (1 + |x|^{2p}).$$

This completes the proof. □

Proof of Theorem 2.4.2. Let $f \in C_p^{2(m+1)}$. Using the semigroup property and assumption $\mathcal{R}(m, \delta_m)$, we have

$$\begin{aligned} P_T f(x) - (Q_{T/n})^n f(x) &= \sum_{k=0}^{n-1} (Q_{T/n})^k (P_{T/n} - Q_{T/n}) P_{T - \frac{k+1}{n} T} f(x) \\ &= \sum_{k=0}^{n-1} (Q_{T/n})^k (\mathrm{Err}_{T/n}^{(m)} P_{T - \frac{k+1}{n} T} f)(x) \end{aligned}$$

where $\mathrm{Err}_t^{(m)}$ is the error term of $(P - Q)$ defined in (2.14).

We obtain from assumptions $\mathcal{R}(m, \delta_m)$ and $(\mathcal{M}_P)$

$$\begin{aligned} |(\mathrm{Err}_{T/n}^{(m)} P_{T - \frac{k+1}{n} T} f)(x)| &\leq K_1 \frac{T}{n} \delta_m \Big(\frac{T}{n} \Big) (1 + |x|^q) \|P_{T - \frac{k+1}{n} T} f\|_{C_p^{2(m+1)}} \\ &\leq \frac{K_2 T}{n} \delta_m \Big(\frac{T}{n} \Big) (1 + |x|^q) \|f\|_{C_p^{2(m+1)}} \end{aligned}$$

and hence Lemma 2.4.3 leads to

$$\begin{aligned}|(Q_{T/n})^k(\mathrm{Err}^{(m)}_{T/n}P_{T-\frac{k+1}{n}T}f)(x)| &\le \frac{K_2T}{n}\delta_m\Big(\frac{T}{n}\Big)\|f\|_{C_p^{2(m+1)}}(Q_{T/n})^k((1+|x|^q)) \\ &\le \frac{K}{n}\delta_m\Big(\frac{T}{n}\Big)\|f\|_{C_p^{2(m+1)}}\end{aligned}$$

for some constant $K = K(T,x)$. This completes the proof. □

The following theorem is an extension of Theorem 2.4.2, and is analogous to Talay-Tubaro [90, Theorem 1].

Theorem 2.4.4. *Assume* $(\mathcal{M})$ *and* $\mathcal{R}(m+1,\delta_{m+1})$ *for* Q_t *with the conditions*

$$J_{\le m}(P_t - Q_t) = 0$$

and

$$\|P_tf - Q_tf\|_{C_{p+2}} \le Ct\|f\|_{C_p^2}. \tag{2.21}$$

Then for each $f \in C_p^{2(m+2)}$, *we have*

$$P_Tf(x) - (Q_{T/n})^nf(x) = \frac{K}{n^m} + O\Big(\Big(\frac{T}{n}\Big)^{m+1} \vee \delta_{m+1}\Big(\frac{T}{n}\Big)\Big) \tag{2.22}$$

where $K = T^m\int_0^T P_sJ_{m+1}(P-Q)P_{T-s}f(x)ds$.

Let us now prepare two auxiliary lemmas.

Lemma 2.4.5. *Let* $f = f_s(x) \in C_p^{1,2}([0,T]\times\mathbf{R}^N)$. *Then a map* $s \mapsto P_sf_s(x)$ *is Lipschitz continuous for all* $x \in \mathbf{R}^N$.

Proof. Note that

$$|P_tf_t(x) - P_sf_s(x)| \le |P_tf_t(x) - P_tf_s(x)| + |P_tf_s(x) - P_sf_s(x)|$$

Using the Lipschitz properties of $t \mapsto f_t(x)$ and $t \mapsto P_tf_s(x)$, the proof follows. □

Lemma 2.4.6. *Let* $g : [0,T] \to \mathbf{R}$ *be a Lipschitz continuous function. Then we have*

$$\Big|\frac{T}{n}\sum_{k=1}^n g(kT/n) - \int_0^T g(s)ds\Big| \le \frac{C(T,g)}{n}. \tag{2.23}$$

Proof. From the assumption we immediately obtain

$$\Big|\frac{T}{n}g(kT/n) - \int_{(k-1)T/n}^{kT/n} g(s)ds\Big| \le \frac{C}{n^2}$$

where C depends on T and the Lipschitz coefficient of g. This implies (2.23). □

Proof of Theorem 2.4.4. We start by noting that as in the proof of Theorem 2.4.2,

$$(P_{T/n} - Q_{T/n})P_{T-s}f(x) = \Big(\frac{T}{n}\Big)^{m+1} J_{m+1}(P-Q)P_{T-s}f(x) + (\mathrm{Err}^{(m+1)}_{T/n} P_{T-s}f)(x)$$

and therefore,

$$P_T f(x) - (Q_{T/n})^n f(x) = \Big(\frac{T}{n}\Big)^{m+1} \sum_{k=0}^{n-1} (Q_{T/n})^k J_{m+1}(P-Q)P_{T-\frac{k+1}{n}T} f(x) + O\Big(\delta_{m+1}\Big(\frac{T}{n}\Big)\Big).$$

Now applying the proof of Theorem 2.4.2 to with $J_{m+1}(P-Q)P_{T-\frac{k+1}{n}T} f \in C^2_{p+2(m+1)}$, we obtain from the inequality (2.21) and $(\mathcal{M}_P)$, for $k \geq 1$,

$$\begin{aligned}
&|((Q_{T/n})^k - P_{kT/n}) J_{m+1}(P-Q)P_{T-\frac{k+1}{n}T} f(x)| \\
&\leq \frac{C_1(T,x)}{n} \| J_{m+1}(P-Q)P_{T-\frac{k+1}{n}T} f\|_{C^2_{p+2(m+1)}} \\
&\leq \frac{C_2(T,x)}{n} \|f\|_{C_p^{2(m+2)}}.
\end{aligned}$$

Next, we have by hypothesis $(\mathcal{M}_P)$, for $0 \leq k \leq n-1$

$$\begin{aligned}
&|P_{kT/n} J_{m+1}(P-Q)P_{T-\frac{k+1}{n}T} f(x) - P_{\frac{k+1}{n}T} J_{m+1}(P-Q)P_{T-\frac{k+1}{n}T} f(x)| \\
&= |(I - P_{T/n})P_{kT/n} J_{m+1}(P-Q)P_{T-\frac{k+1}{n}T} f(x)| \\
&\leq \frac{C_3(T,x)}{n} \|P_{kT/n} J_{m+1}(P-Q)P_{T-\frac{k+1}{n}T} f\|_{C^2_{p+2(m+1)}} \\
&\leq \frac{C_4(T,x)}{n} \|f\|_{C_p^{2(m+2)}}.
\end{aligned}$$

Note that $J_{m+1}(P-Q)P_{T-s}f \in C^{1,2}_{p+2(m+1)}$ and its Lipschitz constant with respect to t is bounded by $J_{m+1}(P-Q)\partial_s P_{T-s} f$ (see the assumption (C) in $\mathcal{R}(m, \delta_m)$). Using Lemmas 2.4.5, 2.4.6, we have

$$\Big|\frac{T}{n}\sum_{k=0}^{n-1} P_{\frac{k+1}{n}T} J_{m+1}(P-Q)P_{T-\frac{k+1}{n}T} f(x) - \int_0^T P_s J_{m+1}(P-Q)P_{T-s} f(x) ds\Big| \leq \frac{C(T,f,x)}{n}.$$

Hence taking $K = T^m \int_0^T P_s J_{m+1}(P-Q)P_{T-s} f(x) ds$, we conclude that

$$P_T f(x) - (Q_{T/n})^n f(x) = \frac{K}{n^m} + O\Big(\Big(\frac{T}{n}\Big)^{m+1} \vee \delta_{m+1}\Big(\frac{T}{n}\Big)\Big).$$

This concludes the proof. □

2.5 Algebraic approximations of semigroup operators using coordinate operators

Throughout this section, we assume that P_t, $t \in [0,T]$ is a semigroup that satisfies $(\mathcal{M})$, $(\mathcal{M}_P)$ and $\mathcal{R}(m,\delta_m)$. Furthermore we suppose that

$$J_{\leq m}(P_t) = I + \sum_{j=1}^{m} \frac{t^j}{j!} e_j$$

with $e_j = \left(\sum_{i=0}^{d+1} L_i\right)^j$ satisfying the properties stated in $\mathcal{R}(m,\delta_m)$. Similarly, we assume that $Q_{i,t}$:$\cup_{p\geq 0} C_p(\mathbf{R}^N) \to \cup_{p\geq 0} C_p(\mathbf{R}^N)$, $i = 0,...,d+1$ be a sequence of operators such that they satisfy $(\mathcal{M})$, $(\mathcal{M}_P)$ and $\mathcal{R}(m,\delta_m)$ with

$$J_{\leq m}(Q_{i,t}) = I + \sum_{j=1}^{m} \frac{t^j}{j!} L_i^j.$$

In what follows, $\prod_{i=1}^{\ell} a_i := a_1 a_2 \cdots a_\ell$ denotes a noncommutative product.

Theorem 2.5.1. *Assume $m = 2$. That is, $(\mathcal{M})$ and $\mathcal{R}(2,\delta_2)$ are satisfied for $Q_{i,t}$ $(i = 0,1,\ldots,d+1)$. Then all the following operators satisfy $(\mathcal{M})$ and $\mathcal{R}(2,\delta_2)$:*

N-V(a) $Q_t^{(a)} = \frac{1}{2} Q_{0,t/2} \prod_{i=1}^{d+1} Q_{i,t} Q_{0,t/2} + \frac{1}{2} Q_{0,t/2} \prod_{i=1}^{d+1} Q_{d+2-i,t} Q_{0,t/2}$

N-V(b) $Q_t^{(b)} = \frac{1}{2} \prod_{i=0}^{d+1} Q_{i,t} + \frac{1}{2} \prod_{i=0}^{d+1} Q_{d+1-i,t}$

Splitting $Q_t^{(sp)} = Q_{0,t/2} \cdots Q_{d,t/2} Q_{d+1,t} Q_{d,t/2} \cdots Q_{0,t/2}$

Moreover, we have $J_{\leq 2}(Q_t^{(a)}) = J_{\leq 2}(Q_t^{(b)}) = J_{\leq 2}(Q_t^{(sp)}) = \sum_{k=0}^{2} \frac{t^k}{k!} L^k$. In particular, the above schemes define a second order approximation scheme.

The proof of Theorem 2.5.1 is an application of Theorem 2.4.2. The conditions follow from the next lemma, together with an algebraic calculation as pointed out at the end of Section 2.2.

This theorem can also be stated for third order approximation schemes.

Lemma 2.5.2. *Let Q_t^1 and $Q_t^2 : \cup_{p\geq 0} C_p(\mathbf{R}^N) \to \cup_{p\geq 0} C_p(\mathbf{R}^N)$ be two linear operators and let $Q_t^1 Q_t^2$ be the composite operator. Then*
(i) If $(\mathcal{M})$ holds for Q_t^1, Q_t^2, then it also holds for $Q_t^1 Q_t^2$.
(ii) If $\mathcal{R}(m,\delta_m)$ holds for Q_t^1, Q_t^2, then it also holds for $Q_t^1 Q_t^2$.

Proof. (i) is obvious. We now prove (ii). Let $m' \leq m$. We have by hypothesis that

$$Q_t^1 f(x) = \sum_{k=0}^{m'} (J_k Q_t^1 f)(x) t^k + (\mathrm{Err}_t^{(m',1)} f)(x)$$

$$Q_t^2 f(x) = \sum_{k=0}^{m'} (J_k Q_t^2 f)(x) t^k + (\mathrm{Err}_t^{(m',2)} f)(x)$$

for $f \in C_p^{2(m'+1)}$, $p \geq 2$. Furthermore there exists $q = q(m,p) > 0$ such that $\mathrm{Err}_t^{(m',1)} f$, $\mathrm{Err}_t^{(m',2)} f \in C_q$. Now we prove (A)-(C) in the definition of $\mathcal{R}(m, \delta_m)$.
(A): Note that for $f \in C_p^{2(m'+1)}(\mathbf{R}^N)$,

$$Q_t^1 Q_t^2 f(x) = Q_t^1 \left(\sum_{k=0}^{m'} (J_k Q_t^2 f)(x) t^k + (\mathrm{Err}_t^{(m',2)} f)(x) \right).$$

Since $J_k Q_t^2 f \in C_{p+2k}^{2(m'+1)-2k}$, $Q_t^1 (J_k Q_t^2 f)$ can be written as

$$(Q_t^1 (J_k Q_t^2 f))(x) = \sum_{\ell=0}^{m'-k} (J_\ell Q_t^1 (J_k Q_t^2 f))(x) t^\ell + (\mathrm{Err}_t^{(m'-k,1)} J_k Q_t^2 f)(x).$$

As a result, we have

$$Q_t^1 Q_t^2 f(x) = \sum_{k=0}^{m'} \sum_{\ell=0}^{m'-k} (J_\ell Q_t^1 (J_k Q_t^2 f))(x) t^{k+\ell} + (\mathrm{Err}_t^{(m',1,2)} f)(x)$$

where

$$(\mathrm{Err}_t^{(m',1,2)} f)(x) = (Q_t^1 \mathrm{Err}_t^{(m',2)} f)(x) + \sum_{k=0}^{m'} (\mathrm{Err}_t^{(m'-k,1)} J_k Q_t^2 f)(x) t^k. \tag{2.24}$$

We obtain from the properties of the error terms that $\mathrm{Err}_t^{(m',1,2)} f \in C_{q'}$ for some $q' = q'(m,p) > q$.
(B): For $f \in C_p^{m''}$ with $m'' \geq 2(m'+1)$, we can derive for $k + \ell \leq m'$,

$$\| J_\ell Q_t^1 (J_k Q_t^2 f) \|_{C_{p+2(k+\ell)}^{m''-2(k+\ell)}} \leq K_1 \| J_k Q_t^2 f \|_{C_{p+2k}^{m''-2k}} \leq K_2 \| f \|_{C_p^{m''}}$$

and by (2.24),

$$\begin{aligned} \| \mathrm{Err}_t^{(m',1,2)} f \|_{C_{q'}} &\leq K_3 \| \mathrm{Err}_t^{(m',2)} f \|_{C_q} + K_4 \| \mathrm{Err}_t^{(m',1)} J_0 Q_t^2 f \|_{C_{q'}} \\ &+ K_5 \sum_{k=1}^{m'} \| J_k Q_t^2 f \|_{C_{p+2k}^{m''-2k}} t^{m'+1} \\ &\leq \begin{cases} K t^{m'+1} \| f \|_{C_p^{m''}} & \text{if } m' < m \\ K t \delta_m(t) \| f \|_{C_p^{m''}} & \text{if } m' = m. \end{cases} \end{aligned}$$

Finally, the proof of (C) is straightforward. □

Proof of Theorem 2.5.1. Using this lemma, we end the proof, calculating $J_{\le m}$ for each numerical discretization scheme. For instance, in the case of N-V(b) (i.e. (2.11)), we obtain

$$\begin{aligned}
&J_{\le 2}\Big(\frac{1}{2}\prod_{i=0}^{d+1} Q_{i,t} + \frac{1}{2}\prod_{i=0}^{d+1} Q_{d+1-i,t}\Big)\\
&= \frac{1}{2}J_{\le 2}\Big(\prod_{i=0}^{d+1} J_{\le 2}\Big(Q_{it}\Big)\Big) + \frac{1}{2}J_{\le 2}\Big(\prod_{i=0}^{d+1} J_{\le 2}\Big(Q_{d+1-i,t}\Big)\Big)\\
&= \frac{1}{2}J_{\le 2}\Big(\prod_{i=0}^{d+1}\Big(\sum_{k=0}^{2}\frac{t^k}{k!}L_i^k\Big)\Big) + \frac{1}{2}J_{\le 2}\Big(\prod_{i=0}^{d+1}\Big(\sum_{k=0}^{2}\frac{t^k}{k!}L_{d+1-i}^k\Big)\Big)\\
&= \frac{1}{2}\Big(I + t\sum_{i=1}^{d+1} L_i + \frac{t^2}{2}\sum_{i=1}^{d+1} L_i^2 + \ t^2\sum_{i<j} L_iL_j\Big)\\
&\quad + \frac{1}{2}\Big(I + t\sum_{i=1}^{d+1} L_i + \frac{t^2}{2}\sum_{i=1}^{d+1} L_i^2 + \ t^2\sum_{i>j} L_iL_j\Big)\\
&= J_{\le 2}(P_t).
\end{aligned}$$

□

Another idea to construct construct higher order schemes is to use local Romberg extrapolation. In order to do this we need to weaken the assumption $\{\xi_i\} \subset [0,1]$. This is done in the next theorem.

Theorem 2.5.3. *Let $m = 1$ or 2. Assume $(\mathcal{M})$ and $\mathcal{R}(2m, t^{2m})$ for P_t and $Q_t^{[i]}$ $(i = 1, \dots, \ell)$ and $(\mathcal{M}_P)$ for P_t. Furthermore, we assume*
(1) $J_{\le 2m}\left(P_t - \sum_{i=1}^{\ell}\xi_i Q_t^{[i]}\right) = 0$ *for some real numbers $\{\xi_i\}_{i=1,\dots,\ell}$ with $\sum_{i=1}^{l}\xi_i = 1$*
(2) *There exists a constant $q = q(m,p) > 0$ such that for every $f \in C_p^{m'}$ with $m' \ge 2(m+1)$, $(P_t - Q_t^{[i]})f \in C_q^{m'-2(m+1)}$ and*

$$\sup_{t\in[0,T]} \|(P_t - Q_t^{[i]})f\|_{C_p^{m'-2(m+1)}} \le C_T\|f\|_{C_q^{m'}}t^{m+1}.$$

Then we have for any $f \in C_p^{4(m+1)}$,

$$\Big|P_T f(x) - \sum_{i=1}^{\ell}\xi_i (Q_{T/n}^{[i]})^n f(x)\Big| \le \frac{C(T,f,x)}{n^{2m}}.$$

Proof. We first remark that the operator $\sum_{i=1}^{\ell}\xi_i Q_t^{[i]}$ no longer satisfies the semigroup property, i.e. $\sum_{i=1}^{\ell}\xi_i (Q_{T/n}^{[i]})^n \neq (\sum_{i=1}^{\ell}\xi_i Q_{T/n}^{[i]})^n$. Thus the proof is nontrivial.

Note that for $f \in C_p^{4(m+1)}$,

$$\mathcal{E} := P_T f(x) - \sum_{i=1}^{\ell} \xi_i \Big(Q_{T/n}^{[i]}\Big)^n f(x) = \sum_{i=1}^{\ell} \xi_i \Big(P_T - \Big(Q_{T/n}^{[i]}\Big)^n\Big) f(x).$$

Using the semigroup property of P_t and $Q_{\frac{k}{n}T}^{[i]}$, we have

$$\begin{aligned}
\mathcal{E} &= \sum_{i=1}^{\ell} \xi_i \sum_{k=0}^{n-1} (Q_{T/n}^{[i]})^k \Big(P_{T/n} - Q_{T/n}^{[i]}\Big) P_{T-\frac{k+1}{n}T} f(x) \\
&= \sum_{i=1}^{\ell} \xi_i \sum_{k=0}^{n-1} P_{kT/n} \Big(P_{T/n} - Q_{T/n}^{[i]}\Big) P_{T-\frac{k+1}{n}T} f(x) \\
&+ \sum_{i=1}^{\ell} \xi_i \sum_{k=0}^{n-1} \Big((Q_{T/n}^{[i]})^k - P_{kT/n}\Big) \Big(P_{T/n} - Q_{T/n}^{[i]}\Big) P_{T-\frac{k+1}{n}T} f(x)
\end{aligned}$$

We expand $(Q_{T/n}^{[i]})^k - P_{kT/n}$ again, to obtain

$$\begin{aligned}
\mathcal{E} &= \sum_{k=0}^{n-1} (P_{T/n})^k \Big(P_{T/n} - \sum_{i=1}^{\ell} \xi_i Q_{T/n}^{[i]}\Big) P_{T-\frac{k+1}{n}T} f(x) \\
&+ \sum_{i=1}^{\ell} \xi_i \sum_{k=0}^{n-1} \sum_{l=0}^{k-1} \Big(Q_{T/n}^{[i]}\Big)^l \Big(Q_{T/n}^{[i]} - P_{T/n}\Big) P_{T-\frac{l+1}{n}T} \Big(P_{T/n} - Q_{T/n}^{[i]}\Big) P_{T-\frac{k+1}{n}T} f(x).
\end{aligned}$$

By the assumption (1), we have

$$\Big| \sum_{k=0}^{n-1} (P_{T/n})^k \Big(P_{T/n} - \sum_{i=1}^{\ell} \xi_i Q_{T/n}^{[i]}\Big) P_{T-\frac{k+1}{n}T} f(x) \Big| \le \frac{C_1(T,f,x)}{n^{2m}}.$$

Thus we end the proof by showing that

$$\begin{aligned}
&\Big| \sum_{i=1}^{\ell} \xi_i \sum_{k=0}^{n-1} \sum_{l=0}^{k-1} \Big(Q_{T/n}^{[i]}\Big)^l \Big(Q_{T/n}^{[i]} - P_{T/n}\Big) P_{T-\frac{l+1}{n}T} \Big(P_{T/n} - Q_{T/n}^{[i]}\Big) P_{T-\frac{k+1}{n}T} f(x) \Big| \\
&\le \frac{C_2(T,f,x)}{n^{2m}}.
\end{aligned}$$

Using here the assumption (2), we obtain

$$\begin{aligned}
&\Big\| \Big(Q_{T/n}^{[i]} - P_{T/n}\Big) P_{T-\frac{l+1}{n}T} \Big(P_{T/n} - Q_{T/n}^{[i]}\Big) P_{T-\frac{k+1}{n}T} f \Big\|_{C_{q'}} \\
&\le \frac{C(T)}{n^{m+1}} \Big\| \Big(P_{T/n} - Q_{T/n}^{[i]}\Big) P_{T-\frac{k+1}{n}T} f \Big\|_{C_q^{2(m+1)}} \\
&\le \frac{C'(T)}{n^{2(m+1)}} \|f\|_{C_p^{4(m+1)}}
\end{aligned}$$

and therefore

$$\Big|\sum_{i=1}^{\ell}\xi_i\sum_{k=0}^{n-1}\sum_{l=0}^{k-1}\Big(Q_{T/n}^{[i]}\Big)^l\Big(Q_{T/n}^{[i]}-P_{T/n}\Big)P_{T-\frac{l+1}{n}T}\Big(P_{T/n}-Q_{T/n}^{[i]}\Big)P_{T-\frac{k+1}{n}T}f(x)\Big|$$
$$\leq\sum_{k=0}^{n-1}\sum_{l=0}^{k-1}\frac{C_2(T,f,x)}{n^{2(m+1)}}\leq\frac{C_2(T,f,x)}{n^{2m}}.$$

This completes the proof. □

Example 2.5.4. It is known that the Ninomiya-Victoir scheme

$$\Big(\frac{1}{2}e^{\frac{T}{2n}L_0}\prod_{i=1}^{d+1}e^{\frac{T}{n}L_i}e^{\frac{T}{2n}L_0}+\frac{1}{2}e^{\frac{T}{2n}L_0}\prod_{i=1}^{d+1}e^{\frac{T}{n}L_{d+2-i}}e^{\frac{T}{2n}L_0}\Big)^n$$

is of order 2 ($m=2,\delta_2(t)=t^2$ in Theorem 2.4.2). By Theorem 2.5.3, the following modified Ninomiya-Victoir scheme

$$\frac{1}{2}\Big(e^{\frac{T}{2n}L_0}\prod_{i=1}^{d+1}e^{\frac{T}{n}L_i}e^{\frac{T}{2n}L_0}\Big)^n+\frac{1}{2}\Big(e^{\frac{T}{2n}L_0}\prod_{i=1}^{d+1}e^{\frac{T}{n}L_{d+2-i}}e^{\frac{T}{2n}L_0}\Big)^n$$

is also of order 2.

Example 2.5.5. Fujiwara [25] gives a proof of a similar version of the above theorem and some examples of 4th and 6th order. General even order schemes are given by [72]. We introduce the examples of 4th order which satisfies the conditions in Theorem 2.5.3 with $m=2$:

$$\frac{4}{3}\left(\frac{1}{2}\Big(\prod_{i=0}^{d+1}e^{\frac{t}{2}L_i}\Big)^2+\frac{1}{2}\Big(\prod_{i=0}^{d+1}e^{\frac{t}{2}L_{d+1-i}}\Big)^2\right)-\frac{1}{3}\left(\frac{1}{2}\prod_{i=0}^{d+1}e^{tL_i}+\frac{1}{2}\prod_{i=0}^{d+1}e^{tL_{d+1-i}}\right)$$

In order to complete the approximation procedure through (quasi) Monte Carlo methods we need to find a stochastic characterization of the operators $Q_{i,t}$.

Definition 2.5.6. Given a stochastic process $Y_t(x)\in\cap_{p\geq1}L^p$, we say that Y is the stochastic characterization of the linear operator Q_t if $Q_tf(x)=E\left[f(Y_t(x))\right]$ for $f\in\cup_{p\geq0}C_p$. In such as case we use the notation $Q_t\equiv Q_t^Y$.

Remark 2.5.7. Given the operators $Q_t^{Z^i}$ $(i=1,\ldots,\ell)$ and the deterministic positive weights $\{\xi_i\}_{1\leq i\leq\ell}$ with $\sum_{i=1}^{l}\xi_i=1$. Let U be a uniform random variable on $[0,1]$ independent of $(Z^i)_i$ and define $Z:=\sum_{i=1}^{\ell}1(\sum_{j=1}^{i-1}\xi_j\leq U<\sum_{j=1}^{i}\xi_j)Z^i$. Then

$$Q_t^Zf(x)\equiv E[f(Z_t(x))]=\sum_{i=1}^{\ell}\xi_iQ_t^{Z^i}f(x).$$

Therefore by Lemma 2.3.2 if $Q_t^{Z^i}$ satisfy $(\mathcal{M})$ and $\mathcal{R}(m,\delta_m)$ so does Q_t^Z. This property will be used repeatedly in what follows.

2.6 Applications

From this section on, we discuss the application of the previous approximation results to the case of solutions of the SDE (2.1). From the results in Section 2.3.2 (see Corollary 2.3.7), it is clear that the semigroup $P_t f(x) := E[f(X_t(x))]$ satisfies the hypotheses $(\mathcal{M})$ and $\mathcal{R}(m, \delta_m)$. We define various approximations generated via a stochastic process $\bar{X}_i$ with corresponding operator $Q_t^{\bar{X}_i}$ $(i = 0, 1, \ldots, d+1)$.

Due to the previous results and in particular, Theorem 2.5.1, we see that is enough to verify local conditions on the approximation operators to conclude global properties of approximation. In particular, we only need to verify that the operator associated with $\bar{X}_i$ (the approximation to the coordinate process) satisfies $(\mathcal{M})$ and $\mathcal{R}(m, \delta_m)$ and $J_{\leq m}(Q_t^{\bar{X}_i}) = I + \sum_{j=1}^{m} \frac{t^j}{j!} L_i^j$ for some $m \geq 2$ for L_i given by (2.6). This is the goal in most of the applications in this section.

Recall that the stochastic differential equation to be approximated is

$$X_t(x) = x + \sum_{i=0}^{d} \int_0^t V_i(X_{s-}(x)) \circ dB_s^i + \int_0^t h(X_{s-}(x)) dY_s.$$

In each of the following sections we consider different approximation processes for the coordinate processes $X_{i,t}$. In each section, the notation for the approximating process is always $\bar{X}_{i,t}$. We hope that this does not raise confusion as the framework in each section is clear.

2.6.1 Continuous diffusion component

a) Explicit solution: Let $V : \mathbf{R}^N \to \mathbf{R}^N$ be a smooth function satisfying the linear growth condition $|V(x)| \leq C(1 + |x|)$. The exponential map is defined as $\exp(V)x = z_1(x)$ where z denotes the solution of the ordinary differential equation

$$\frac{dz_t(x)}{dt} = V(z_t(x)), \; z_0(x) = x. \tag{2.25}$$

The solution of the coordinate sde is obtained in the following Proposition.

Proposition 2.6.1. *For $i = 0, 1, ..., d$, the stochastic differential equation*

$$X_{i,t}(x) = x + \int_0^t V_i(X_{i,s}(x)) \circ dB_s^i \tag{2.26}$$

has a unique solution given by

$$X_{i,t}(x) = \exp(B_t^i V_i)x.$$

Proof. Apply Itô's formula to $g_i(B_t^i)$ with $g_i(y) = \exp(yV_i)x$, $y \in \mathbf{R}$. □

$X_{i,t}(x)$ is called the i-th coordinate process and its semigroup is denoted by Q_t^i. This is a trivial example of the approximation of e^{tL_i}, $i = 0, 1, \ldots, d$ satisfying $(\mathcal{M})$ and $\mathcal{R}(m, t^m)$. However, sometimes it is not easy to obtain the closed-form solution to the ODE (2.25). In those cases, we shall approximate $\exp(tV)x$. Here we will do this with the Taylor expansion first and then the Runge-Kutta methods denoted by b_m and c_m respectively.

b) Taylor expansion: We first prove the following lemmas which help us to find the rate of convergence of the scheme to be defined later. The following Lemma follows easily from Gronwall's lemma.

Lemma 2.6.2. *Let V be a smooth function which satisfies the linear growth condition. Then $|\exp(tV)x| \leq C\ e^{K|t|}(1 + |x|)$ for $t \in \mathbf{R}$, $x \in \mathbf{R}^N$.*

From now on we denote by $e_j : \mathbf{R}^N \to \mathbf{R}$, the coordinate function $e_j(x) = x_j$ for $j = 1, ..., N$. Furthermore, we also denote by V the vector field operator defined from V.

Lemma 2.6.3. *Let $f \in C_p^{m+1}$. Then we have for $i = 0, 1, \ldots, d$,*

$$f\left(\exp(tV_i)x\right) = \sum_{k=0}^{m} \frac{t^k}{k!} V_i^k f(x) + \int_0^t \frac{(t-u)^m}{m!} V_i^{m+1} f(\exp(uV_i)x)du$$

and

$$\Big| \int_0^t \frac{(t-u)^m}{m!} V_i^{m+1} f(\exp(uV_i)x)du \Big| \leq C_m \|f\|_{C_p^{m+1}} e^{K|t|}(1 + |x|^{p+m+1})|t|^{m+1}.$$

for all $t \in \mathbf{R}$.

Proof. Assertion (2.6.3) follows application of Taylor expansion to the function $f(\exp(tV)x)$ around $t = 0$. Next, as $|V_i^{m+1} f(x)| \leq C(1 + |x|^{p+m+1})$, we obtain from Lemma 2.6.2,

$$\begin{aligned}
&\Big| \int_0^t \frac{(t-u)^m}{m!} V_i^{m+1} f(\exp(uV)x)du \Big| \\
&\leq C_m \|f\|_{C_p^{m+1}} \int_0^{|t|} |t|^m C e^{K|u|}(1 + |x|^{p+m+1}))du \\
&\leq C'_m \|f\|_{C_p^{m+1}} e^{K|t|}(1 + |x|^{p+m+1})|t|^{m+1}.
\end{aligned}$$

□

Based on this Lemma, we define the approximation to the solution of the coordinate equation (2.26) as follows

$$b_m^j(t, V)x = \sum_{k=0}^{m} \frac{t^k}{k!} (V^k e_j)(x),\ j = 1, ..., N.$$

Define

$$\bar{X}_{i,t}(x) = b_{2m+1}(B_t^i, V_i)x \text{ for } i = 0, ..., d.$$

Then we have the following approximation result.

Proposition 2.6.4. *(i) For every $p \geq 1$,*

$$\|X_{i,t}(x) - \bar{X}_{i,t}(x)\|_{L^p} \leq C(p, m, T)(1 + |x|^{2(m+1)})t^{m+1}.$$

(ii) Let $f \in C_p^1$. Then we have

$$E[|f(X_{i,t}(x)) - f(\bar{X}_{i,t}(x))|] \leq C(m, T)\|f\|_{C_p^1}(1 + |x|^{p+2(m+1)})t^{m+1}.$$

Proof. (i): Apply Proposition 2.6.1 and Lemma 2.6.3 with $f = e_i$. Then we have

$$\begin{aligned}\|X_{i,t}(x) - \bar{X}_{i,t}(x)\|_{L^p} &\leq E\left[|C_m e^{K|B_t|}(1 + |x|^{2(m+1)})|B_t|^{2(m+1)}|^p\right]^{1/p} \\ &\leq C(1 + |x|^{2(m+1)})t^{m+1}\end{aligned}$$

for some constant $C = C(p, m, T)$.
(ii): We first apply the mean value theorem to obtain

$$\begin{aligned}&E[|f(X_{i,t}(x)) - f(\bar{X}_{i,t}(x))|] \\ &\leq \|f\|_{C_p^1}\|1 + |\theta X_{i,t}(x) + (1-\theta)\bar{X}_{i,t}(x)|^p\|_{L^2}\|X_{i,t}(x) - \bar{X}_{i,t}(x)\|_{L^2} \\ &\leq C\|f\|_{C_p^1}\|1 + |X_{i,t}(x)|^p + |\bar{X}_{i,t}(x)|^p\|_{L^2}(1 + |x|^{2(m+1)})t^{m+1}.\end{aligned}$$

We see by Lemma 2.6.2 that

$$\sup_{t\in[0,T]} \|1 + |X_{i,t}(x)|^p + |\bar{X}_{i,t}(x)|^p\|_{L^2} \leq C'(1 + |x|^p)$$

from which the proof follows. □

As a result of this proposition we can see that $\mathcal{R}(m, t^m)$ holds for the operators associated with $b_m(t, V_0)x$ and $b_{2m+1}(B_t^i, V_i)x$, $1 \leq i \leq d$. Indeed, we have for $m' \leq m$,

$$\begin{aligned}E[f(\bar{X}_{i,t}(x))] &= E\left[f(X_{i,t}(x))\right] + E[f(\bar{X}_{i,t}(x)) - f(X_{i,t}(x))] \\ &= \sum_{k=0}^{m'} \frac{t^k}{k!} L_i^k f(x) + (E_t^{m'} f)(x)\end{aligned}$$

where

$$(E_t^{m'} f)(x) := (\mathrm{Err}_t^{(m')} f)(x) + E[f(\bar{X}_{i,t}(x)) - f(X_{i,t}(x))]$$

and $(\mathrm{Err}_t^{(m')} f)(x)$ is defined through the residue appearing in Proposition 2.3.6, using L_i and Q_i instead of L and P. Furthermore, using Proposition 2.6.4 (ii), we have that the error term $E_t^{m'}$ satisfies (B) in assumption $\mathcal{R}(m, t^m)$.

It remains to prove that $(\mathcal{M})$ holds for $\bar{X}_{i,t}(x)$. For the proof, we need an additional growth condition for the vector field V_i.

Proposition 2.6.5. *Assume that* $(V_i^k e_j)$ $(2 \le k \le m,\ 0 \le i \le d,\ 1 \le j \le N)$ *satisfies the linear growth condition then* $(\mathcal{M})$ *holds for* $\bar{X}_{i,t}(x)$, $i = 0, \dots, d$.

Proof. The assumption $(\mathcal{M}_0)$ follows from the smoothness and the linear growth property of $V_i^k e_j$. We only prove the moment condition (2.13) for $\bar{X}_{i,t}(x)$ $i = 1, \dots, d$. Consider the multiplication ($p \in \mathbf{N}$)

$$\Big| \sum_{k=0}^{m} \frac{(B_t^i)^k}{k!} (V_i^k e_j)(x) \Big|^{2p} = \Big| x + B_t^i V_i(x) + \sum_{k=2}^{m} \frac{(B_t^i)^k}{k!} (V_i^k e_j)(x) \Big|^{2p}.$$

Taking into account that $E[(B_t^i)^{2k+1}] = 0$, $k \in \boldsymbol{N}$. Then by the assumption, we obtain the result. □

Therefore we obtain the main result.

Theorem 2.6.6. *Assume that* $(V_i^k e_j)$ $(2 \le k \le m,\ 0 \le i \le d,\ 1 \le j \le N)$ *satisfies the linear growth condition. Let* $\bar{X}_{i,t}(x)$ *be defined by*

$$\bar{X}_{i,t}(x) = b_{2m+1}(B_t^i, V_i)x = \sum_{k=0}^{2m+1} \frac{1}{k!} (V_i^k I)(x) \int_{0<t_1<\cdots<t_k<t} 1 \circ dB_{t_1}^i \cdots \circ dB_{t_k}^i.$$

Denote by $Q_t^{\bar{X}_i}$ *the semigroup associated with* $\bar{X}_{i,t}(x)$. *Then* $Q_t^{\bar{X}_i}$ *satisfies* $(\mathcal{M})$ *and* $\mathcal{R}(m, t^m)$. *Furthermore* $J_{\le m}(Q_t^{\bar{X}_i}) = I + \sum_{j=1}^{m} \frac{t^j}{j!} L_i^j$.

c) Runge-Kutta methods: We say here that c_m is an s-stage explicit Runge-Kutta method of order m for the ODE (2.25) if it can be written in the form

$$c_m(t, V)x = x + t \sum_{i=1}^{s} \beta_i k_i(t, V)x \tag{2.27}$$

where $k_i(t, V)x$ defined inductively by

$$\begin{aligned} k_1(t, V)x &= V(x), \\ k_i(t, V)x &= V\Big(x + t \sum_{j=1}^{i-1} \alpha_{i,j} k_j(t, V)x\Big),\ 2 \le i \le s, \end{aligned}$$

and satisfies

$$|\exp(tV)x - c_m(t, V)x| \le C_m e^{K|t|}(1 + |x|^{m+1})|t|^{m+1}$$

for some constants $((\beta_i, \alpha_{i,j})_{1 \le j < i \le s})$. Runge-Kutta formulas of order less than or equal to 7 are well known. For details, see e.g. Butcher [13].

The following proposition can be shown by the same argument as in the proof of Proposition 2.6.4.

Proposition 2.6.7 (stochastic Runge-Kutta). *(i) For every $p \geq 1$,*

$$\|X_{i,t}(x) - c_{2m+1}(B_t^i, V_i)x\|_{L^p} \leq C(p, m, T)(1 + |x|^{2(m+1)})t^{m+1} \tag{2.28}$$

(ii) Let $f \in C_p^1$. Then we have

$$E[|f(X_{i,t}(x)) - f(c_{2m+1}(B_t^i, V_i)x)|] \leq C(m, T)\|f\|_{C_p^1}(1 + |x|^{2(m+1)})t^{m+1} \tag{2.29}$$

Next we show that $(\mathcal{M})$ still holds for the Runge-Kutta schemes.

Proposition 2.6.8. *$(\mathcal{M})$ holds for $c_m(B_t^i, V_i)x$, $i = 0, \ldots, d$.*

Proof. We first note that for every $1 \leq j \leq s$, there exists a function of the form $p_j = \sum_{k=0}^{j-1} a_{jk}|t|^k$ such that

$$|k_j(t, V)x| \leq p_j(t)(1 + |x|).$$

The assumption $(\mathcal{M}_0)$ follows from the smoothness and the linear growth property of V_i. We now prove (2.13). In the case $i = 0$, this is obvious by definition and the inequality (2.6.1). In the case $1 \leq i \leq d$, observe that

$$\begin{aligned} c_m(t, V)x &= x + t\sum_{l=1}^{s} \beta_l V(x) + t\sum_{l=2}^{s} \beta_l \int_0^1 \frac{d}{d\theta} V\Big(x + \theta t \sum_{j=1}^{l-1} \alpha_{l,j} k_j(t, V)x\Big) d\theta \\ &=: x + t\sum_{l=1}^{s} \beta_l V(x) + D_m(t, V)x. \end{aligned}$$

Expanding multiplications and taking expectations, as in Proposition 2.6.5, we can show that the terms containing odd powers of B_t^i have expectation 0. Finally, we obtain from the boundedness of ∂V_i that

$$|D_m(B_t^i, V_i)x| \leq p(B_t^i)(1 + |x|)$$

where $p = p(t)$ is of the form $\sum_{k=2}^{s} a_k|t|^k$. Using this, we conclude the proof. □

Consequently, as in the Taylor scheme, $\mathcal{R}(m, t^m)$ and $(\mathcal{M})$ hold for the operators associated with $c_m(t, V_0)x$ and $c_{2m+1}(B_t^i, V_i)x$, $1 \leq i \leq d$. For more on this method, we refer the reader to [68].

d) Minor extension: In the previous approximation, the assumption that $B_t \sim N(0, I_d)$ can be weakened. In fact, we can use $\sqrt{t}Z$ instead of B_t where $(Z^i)_{i=1}^d$ are independent and

$$P(Z^i = \pm\sqrt{3}) = \frac{1}{6}, \quad P(Z^i = 0) = \frac{2}{3}$$

for each $i = 1, \ldots, d$.

Proposition 2.6.9. *Let B_t be a 1-dimensional Brownian motion and Z be a* $\mathbf{R}$*-valued random variable such that for all* $0 \leq k \leq m$*,*

$$E[(Z)^k] = E[(B_1)^k]$$

and

$$E[\exp(c|Z|)] < \infty$$

for any $c > 0$*. Then, for every* $f \in C_p^{m+1}$*,*

$$|E[f(\exp(B_t V)x)] - E[f(c_m(\sqrt{t}Z, V)x)]| \leq C(m,T)(1+|x|^{p+m+1})t^{(m+1)/2}.$$

2.6.2 Compound Poisson case

Suppose that Y_t is a compound Poisson process. That is,

$$Y_t = \sum_{i=1}^{N_t} J_i$$

where (N_t) is a Poisson process with intensity λ and (J_i) are i.i.d. $\mathbf{R}^d$-valued random variables independent of (N_t) with $J_i \in \cap_{p\geq 1} L^p$.

In this case Y_t is a Lévy process with generator of the form

$$\int_{\mathbf{R}_0^d} (f(x+y) - f(x))\nu(dy)$$

where $\tau \equiv 0$, $b = 0$, $\nu(\mathbf{R}_0^d) = \lambda < \infty$ and $\nu(dy) = \lambda P(J_1 \in dy)$.

Then in this case

$$X_t^{d+1}(x) = x + \int_0^t h(X_{s-}^{d+1}(x))dY_s, \ t \in [0,T] \tag{2.30}$$

which can be solved explicitly. Indeed, let $(G_i(x))$ be defined by recursively

$$G_0 = x$$

$$G_i = G_{i-1} + h(G_{i-1})J_i.$$

Then the solution can be written as $X_t^{d+1}(x) = G_{N_t}(x)$. Define for fixed $M \in \mathbf{N}$, the approximation process $\bar{X}_{d+1,t} = G_{N_t \wedge M}(x)$. This approximation requires the simulation of at most M jumps. In fact, the rate of convergence is fast as the following result shows.

Theorem 2.6.10. *Let $M \in \mathbf{N}$. Then the process $G_{N_t\wedge M}(x)$ satisfies $(\mathcal{M})$ and $\mathcal{R}(M, t^{M-\kappa})$ for arbitrary small $\kappa > 0$. Furthermore $J_{\leq M}(Q_t^{\bar{X}_{d+1}}) = I + \sum_{j=1}^m \frac{t^j}{j!} L_{d+1}^j$.*

Proof. Note that for $f \in C_p$

$$\begin{aligned} Q_t^{\bar{X}_{d+1}} f(x) - Q_t^{d+1} f(x) &= E[f(G_{N_t \wedge M}(x))] - E[f(G_{N_t}(x))] \\ &= E[(f(G_{N_t \wedge M}(x)) - f(G_{N_t}(x)))\ 1_{\{T_{M+1} \le t\}}] \end{aligned}$$

where $T_M := \inf\{t > 0 : N_t = M\}$. By the Hölder inequality,

$$\begin{aligned} |Q_t^{\bar{X}_{d+1}} f(x) - Q_t^{d+1} f(x)| &\le 2E[\sup_{0 \le t \le T} |f(G_{N_t}(x))|^{\frac{\gamma}{\gamma-1}}]^{\frac{\gamma-1}{\gamma}} P(T_{M+1} \le t)^{\frac{1}{\gamma}} \\ &= 2E[\sup_{0 \le t \le T} |f(G_{N_t}(x))|^{\frac{\gamma}{\gamma-1}}]^{\frac{\gamma-1}{\gamma}} \Big(\int_0^t \frac{(\lambda s)^M}{M!} \lambda e^{-\lambda s} ds \Big)^{\frac{1}{\gamma}} \\ &\le C(\gamma, T) \|f\|_{C_p} (1 + |x|^p) \left(t\lambda^{-1}\right)^{(M+1)/\gamma} \end{aligned}$$

Take sufficiently small $\gamma > 1$, then $\mathcal{R}(M, t^{M-\kappa})$ holds for $Q_t^{\bar{X}_{d+1}}$ where $\kappa := (1 - 1/\gamma)(M+1) > 0$. Finally, we show $(\mathcal{M})$. Let $f_p(x) = |x|^{2p}$ $(p \in \mathbf{N})$ and $\gamma < M$. Then using the above calculation and Corollary 2.3.7, we have

$$\begin{aligned} Q_t^{\bar{X}_{d+1}} f_p(x) &= Q_t^{d+1} f_p(x) + (Q_t^{\bar{X}_{d+1}} f_p(x) - Q_t^{d+1} f_p(x)) \\ &\le (1 + K_1 t) f_p(x) + K_2 t + |Q_t^{\bar{X}_{d+1}} f_p(x) - Q_t^{d+1} f_p(x)| \\ &\le (1 + K_3 t) f_p(x) + K_4 t. \end{aligned}$$

□

2.6.3 Infinite activity case

In this subsection, we consider the SDE (2.4) under the conditions $\nu(\mathbf{R}_0^d) = \infty$. Without loss of generality, we assume that $c \equiv 0$.

a) Ignoring small jumps: Define for $\varepsilon > 0$ the finite activity (i.e. drift + compound Poisson) Lévy process (Y_t^ε) with Lévy triplet $(b, 0, \nu^\varepsilon)$ where the Lévy measure is defined by

$$\nu^\varepsilon(E) := \nu(E \cap \{y : |y| > \varepsilon\}),\ E \in \mathcal{B}(\mathbf{R}_0^d). \tag{2.31}$$

Consider the approximate coordinate SDE

$$\bar{X}_{d+1,t}(x) = x + \int_0^t h(\bar{X}_{d+1,s-}(x)) dY_s^\varepsilon,$$

whose generator is

$$L_{d+1}^{1,\varepsilon} f(x) = \nabla f(x) h(x) b + \int (f(x + h(x)y) - f(x) - \nabla f(x) h(x) \tau(y)) \nu^\varepsilon(dy).$$

Now we derive the error estimate for $\bar{X}_{d+1,t}$ by the distance between two generators using a kind of parametrix method.

Theorem 2.6.11. *Assume that $0 < \varepsilon \equiv \varepsilon(t) \leq 1$ is chosen as to satisfy that $\sigma^2(\varepsilon) := \int_{|y|\leq\varepsilon} |y|^2\nu(dy) = O(t^M)$. Then we have that $Q_t^{\bar{X}_{d+1}}$ satisfies $(\mathcal{M})$ and $\mathcal{R}(M, t^M)$. Furthermore $J_{\leq M}(Q_t^{\bar{X}_{d+1}}) = I + \sum_{j=1}^{M} \frac{t^j}{j!} L_{d+1}^j$.*

Proof. First, we remark that condition $(\mathcal{M}_0)$ follows from Proposition 5.2 in [39]. We start by noting that from Proposition 2.3.5, we have (see e.g. Kohatsu-Higa and Tankov[47])

$$\begin{aligned} Q_t^{d+1} f(x) - Q_t^{\bar{X}_{d+1}} f(x) &= \int_0^t \frac{d}{ds} (Q_{t-s}^{\bar{X}_{d+1}} Q_s^{d+1} f)(x) ds \\ &= \int_0^t (Q_{t-s}^{\bar{X}_{d+1}} (L_{d+1} - L_{d+1}^{1,\epsilon}) Q_s^{d+1} f)(x) ds \end{aligned}$$

Therefore the proof is achieved if we prove that

$$|(L_{d+1} - L_{d+1}^{1,\varepsilon}) f(x)| \leq C \|f\|_{C_p^2} (1 + |x|^{p+2}) t^M.$$

For the proof, we change here the representation of the Lévy triplets of Y_t and Y_t^ε as follows:

$$\begin{aligned} (b, 0, \nu), \tau &\Rightarrow (b_\varepsilon, 0, \nu), \tau_\varepsilon \\ (b, 0, \nu^\varepsilon), \tau &\Rightarrow (b_\varepsilon, 0, \nu^\varepsilon), \tau_\varepsilon \end{aligned}$$

where $\tau_\varepsilon(y) = y 1_{\{|y|\leq\varepsilon\}}$. Then

$$\begin{aligned} |(L_{d+1} - L_{d+1}^{1,\varepsilon}) f(x)| \leq & \Big| \int \nabla f(x) h(x) (y - \tau_\varepsilon(y)) (\nu(dy) - \nu^\varepsilon(dy)) \Big| \\ & + \Big| \int \int_0^1 (1-\theta) \frac{d^2}{d\theta^2} f(x + \theta h(x) y) d\theta (\nu(dy) - \nu^\varepsilon(dy)) \Big|. \end{aligned} \tag{2.32}$$

We first obtain that for $\varepsilon > 0$,

$$\int (y - \tau_\varepsilon(y)) (\nu(dy) - \nu^\varepsilon(dy)) = 0$$

since the support of the measure $(\nu - \nu^\varepsilon)$ is $\{|y| \leq \varepsilon\}$. Now we consider the second term of (2.32). We can immediately show that due to the polynomial growth property for f,

$$\Big| \int \int_0^1 \frac{d^2}{d\theta^2} f(x + \theta h(x) y) d\theta (\nu(dy) - \nu^\varepsilon(dy)) \Big| \leq C \|f\|_{C_p^2} (1 + |x|^{p+2}) \sigma^2(\varepsilon)$$

and hence as $\sigma^2(\varepsilon) = O(t^M)$, one obtains that $J_{\leq M}(Q_t^{\bar{X}_{d+1}}) = I + \sum_{j=1}^{m} \frac{t^j}{j!} L_{d+1}^j$ and that $Q_t^{\bar{X}_{d+1}}$ satisfies $(\mathcal{M})$ and $\mathcal{R}(M, t^M)$ follows as in the proof of Proposition 2.6.10. □

Using Theorem 2.5.1, we can incorporate the above approximating process $\bar{X}_{d+1,t}$ to the whole approximation method. This will require to first simulate the jump times of the approximating Lévy process Y^ε and then solving ode's between these times. If the task is time consuming one can also separate the jump component from the drift component as indicated by Theorem 2.5.1. The right size of ε is determined by the condition $\sigma^2(\varepsilon) \leq Ct^M$.

b) Approximation of small jumps: We apply here the Asmussen-Rosiński's approximation for small jumps of Lévy processes. The idea is that the small jumps ignored in (2.31) are close to a Brownian motion with small variance $\sigma^2(\varepsilon)$ (see details in [2]).

Consider the new approximate SDE

$$\bar{X}_{d+1,t}(x) = x + \int_0^t h(\bar{X}_{d+1,s}(x))\Sigma_\varepsilon^{1/2} dW_s + \int_0^t h(\bar{X}_{d+1,s-}(x))dY_s^\varepsilon \tag{2.33}$$

where W_t is a new d-dimensional Brownian motion independent of B_t and Y_t^ε, and Σ_ε is the symmetric and semi-positive definite $d \times d$ matrix defined as

$$\Sigma_\varepsilon = \int_{|y|\leq\varepsilon} yy^*\nu(dy). \tag{2.34}$$

We remark that Σ_ε is of the form $A\Lambda A^*$, where A is an orthogonal matrix and Λ is the diagonal matrix with entries $\lambda_1, \ldots, \lambda_d \geq 0$ (eigenvalues). Thus we use the notation $\Sigma_\varepsilon^{1/2} = A\Lambda^{1/2}$. Since the above SDE is also driven by a jump-diffusion process, we can also simulate it using the second order discretization schemes in Theorem 2.5.1.

Now we prove that rate of convergence in this case is faster than in the case that we ignore completely the small jumps (see Theorem 2.6.11).

Theorem 2.6.12. *Assume that $0 < \varepsilon \equiv \varepsilon(t) \leq 1$ is chosen as to satisfy that $\int_{|y|\leq\varepsilon} |y|^3\nu(dy) = O(t^M)$. Then we have that $Q_t^{\bar{X}_{d+1}}$ satisfies $(\mathcal{M})$ and $\mathcal{R}(M, t^M)$. Furthermore $J_{\leq M}(Q_t^{\bar{X}_{d+1}}) = I + \sum_{j=1}^M \frac{t^j}{j!} L_{d+1}^j$.*

Proof. As before, condition $(\mathcal{M}_0)$ follows from Proposition 5.2 in [39]. The SDE $\bar{X}_{d+1,t}$ corresponds to the generator

$$\begin{aligned} L_{d+1}^{2,\varepsilon} f(x) :=& \nabla f(x)h(x)b + \frac{1}{2}\sum_{k,l} \partial_{k,l} f(x)(h(x)\Sigma_\varepsilon h^*(x))_{k,l} \\ &+ \int (f(x + h(x)y) - f(x) - \nabla f(x)h(x)\tau(y))\nu^\varepsilon(dy). \end{aligned}$$

Using this representation, we have for $f \in C_p^3$,

$$\begin{aligned}(L_{d+1} - L_{d+1}^{2,\varepsilon})f(x) &= \int\int_0^1 (1-\theta)\frac{d^2}{d\theta^2} f(x+\theta h(x)y)d\theta(\nu(dy) - \nu^\varepsilon(dy)) \\ &\quad - \frac{1}{2}\sum_{k,l} \partial_{k,l} f(x)(h(x)\Sigma_\varepsilon h^*(x))_{k,l} \\ &= \int\int_0^1 \frac{(1-\theta)^2}{2}\frac{d^3}{d\theta^3} f(x+\theta h(x)y)d\theta(\nu(dy) - \nu^\varepsilon(dy)).\end{aligned}$$

Hence we finish the proof as in the proof of Theorems 2.6.10 and 2.6.11. □

If we put all the pieces together, we have the following final result. Here B_t^{ij} denote $i = 1, ..., d$, $j = 1, ..., 2n$ denote $2nd$ independent standard Brownian motions and $B_t^{0j} \equiv t$.

Theorem 2.6.13. *Assume that V_0, V and h are infinitely differentiable functions with bounded derivatives with $\int_{|y|>1} |y|^p \nu(dy) < \infty$ for all $p \in \mathbf{N}$. Define $\varepsilon \equiv \varepsilon(T, n)$ so that $\int_{|y|\le\varepsilon} |y|^3 \nu(dy) = O((\frac{T}{n})^2)$. Let $\bar{X}_{i,t}^j(x) = c_5(B_t^{ij}, V_i)x$, $i = 0, ..., d$, $j = 1, ..., 2n$, $2n$ copies of the Runge-Kutta method of order 2 as defined in (2.27) and $\bar{X}_{d+1,t}^j(x)$ $j = 1, ..., 2n$ independent copies of the approximation defined in (2.33). Then the following schemes, $X_T^{(n)} = Y_n^n \circ Y_n^{n-1} \circ ... \circ Y_n^1(x)$, are second order discretization schemes:*

N-V(a) $Y_n^j(x) = U_j \bar{X}_{0,T/(2n)}^j \circ \bar{X}_{1,T/n}^j \circ ... \circ \bar{X}_{d+1,T/n}^j \circ \bar{X}_{0,T/(2n)}^j(x) + (1-U_j)\bar{X}_{0,T/(2n)}^j \circ \bar{X}_{d+1,T/n}^j \circ ... \circ \bar{X}_{1,T/n}^j \circ \bar{X}_{0,T/(2n)}^j(x)$ *where U_j is a Bernoulli r.v. with $P(U_j = 1) = 1/2$, independent of everything else.*

N-V(b) $Y_n^j(x) = U_j \bar{X}_{d+1,T/n}^j \circ ... \circ \bar{X}_{0,T/n}^j(x) + (1-U_j)\ \bar{X}_{0,T/n}^j \circ ... \circ \bar{X}_{d+1,T/n}^j(x)$ *where U_j is a Bernoulli r.v. with $P(U_j = 1) = 1/2$, independent of everything else.*

Splitting $Y_n^j(x) = \bar{X}_{0,T/(2n)}^j \circ ... \circ \bar{X}_{d,T/(2n)}^j \circ \bar{X}_{d+1,T/n}^j \circ \bar{X}_{d,T/(2n)}^{n+j} \circ ... \circ \bar{X}_{0,T/(2n)}^{n+j}(x)$.

One can also write a similar result for higher order schemes using Theorem 2.5.3.

2.6.4 Limiting the number of jumps per interval for approximations of infinite activity Lévy driven SDE's

In the previous two approximations although $\varepsilon \in (0, 1)$ may be relatively large compared with the interval size T/n, one still faces the possibility of having many jumps in the interval $[0, T]$. Therefore we introduce the idea used in Section 2.6.2. That is, we propose another approximation that restricts the numbers of possible jumps to at most n. Throughout this section we assume that $\int_{|y|<1} |y|\nu(dy) < \infty$ and without loss of generality, we assume that $\tau(y) = y1_{|y|<1}$.

Then we decompose the operator

$$L_{d+1} = L^1_{d+1} + L^2_{d+1} + L^3_{d+1}$$

$$L^1_{d+1} f(x) := \nabla f(x) h(x) \left(b - \int_{\varepsilon < |y| \leq 1} \tau(y) \nu(dy) \right)$$

$$L^2_{d+1} f(x) := \int_{|y| \leq \varepsilon} (f(x + h(x)y) - f(x) - \nabla f(x) h(x) \tau(y)) \nu(dy)$$

$$L^3_{d+1} f(x) := \int_{\varepsilon < |y|} f(x + h(x)y) - f(x) \nu(dy).$$

The operator L^1_{d+1} can be easily approximated using any Runge-Kutta method for the ordinary differential equation $X^1_{d+1,t} = x + \left(b - \int_{\varepsilon<|y|\leq 1} \tau(y)\nu(dy) \right) \int_0^t h\left(X^1_{d+1,s}\right) ds$. We denote by $\bar{X}^1_{d+1,t}$, the Euler scheme associated with this ordinary differential equation. Therefore we only need to approximate L^2_{d+1} and L^3_{d+1}.

Let $l : \mathbf{R}^d \to \mathbf{R}_+$ be a localization function that may be used for importance sampling of the Lévy measure. Let $F^l_\varepsilon(dy) = \lambda_\varepsilon^{-1} l(y) 1_{|y| \leq \varepsilon} \nu(dy)$ with $\lambda_\varepsilon = \int_{|y| \leq \varepsilon} l(y) \nu(dy)$. Let $Y_\varepsilon \sim F_\varepsilon$. Define $\bar{X}^{2,\varepsilon}_t(x) = x + h(x) W_t \sqrt{\lambda_\varepsilon}$, where W is a d-dimensional Wiener process with covariance matrix given by $\Sigma_{ij} = l(Y^\varepsilon)^{-1} Y^\varepsilon_i Y^\varepsilon_j$ which is independent of everything else.

First we prove that $\bar{X}^{2,\varepsilon}_t(x)$ satisfies assumption $(\mathcal{M})$.

Lemma 2.6.14. *Assume that for $p \geq 2$, $\sup_{\varepsilon \in (0,1]} \int_{|y| \leq \varepsilon} |y|^p l(y)^{-\frac{p-2}{2}} \nu(dy) < \infty$, then assumption $(\mathcal{M})$ is satisfied with*

$$E\left[\left| \bar{X}^{2,\varepsilon}_{d+1}(x) \right|^p \right] \leq (1 + Kt)|x|^p + K't.$$

Proof. Let $f(x) = |x|^p$, $p \geq 2$. Using Ito's formula for $p \neq 3$ and an approximative argument in the case $p = 3$ (as in the proof of the Meyer-Ito formula) one obtains that

$$E\left[f\left(\bar{X}^{2,\varepsilon}_t(x) \right) \right] - f(x) \tag{2.35}$$

$$= \frac{p}{2} \lambda_\varepsilon E\left[l(Y^\varepsilon)^{-1} \int_0^t \left(\frac{p}{2} - 1 \right) \left| \bar{X}^{2,\varepsilon}_s(x) \right|^{p-4} \left\langle h(x) Y^\varepsilon, \bar{X}^{2,\varepsilon}_s(x) \right\rangle^2 + \left| \bar{X}^{2,\varepsilon}_s(x) \right|^{p-2} \left| h(x) Y^\varepsilon \right|^2 ds \right]$$

We use the Lipschitz property of h to obtain that

$$\begin{aligned} \left| \bar{X}^{2,\varepsilon}_s(x) \right| &= \left| x + h(x) W_s \sqrt{\lambda_\varepsilon} \right| \\ &\leq \left(1 + C \left| W_s \right| \sqrt{\lambda_\varepsilon} \right) (1 + |x|). \end{aligned}$$

Then, we have

$$\begin{aligned} &\left| E\left[f\left(\bar{X}^{2,\varepsilon}_t(x) \right) \right] - f(x) \right| \\ &\leq C_p t \left(1 + |x|^p\right) \int_{|y| < \varepsilon} |y|^2 \left(1 + \left(|y|^2 l(y)^{-1} \lambda_\varepsilon t \right)^{\frac{p-2}{2}} \right) \nu(dy). \end{aligned}$$

□

Lemma 2.6.15. *Assume that for $p \geq 2$,*

$$M_p = \sup_{\varepsilon \in (0,1]} \int_{|y| \leq \varepsilon} |y|^4 l(y)^{-1} \left(1 + \left(|y|^2 l(y)^{-1} \lambda_\varepsilon t\right)^{\frac{p-2}{2}}\right) \nu(dy) < \infty$$

and $\int_{|y| \leq \varepsilon} |y|^3 \nu(dy) \leq Ct$ then

$$\left|E\left[f(\bar{X}_t^{2,\varepsilon})\right] - f(x) - tL_{d+1}^2 f(x)\right| \leq C(p) \left\|f\right\|_{C_p^4} (1 + |x|^{p+4}) t^2.$$

That is, $\bar{X}_t^{2,\varepsilon}(x)$ satisfies assumption $\mathcal{R}(2, t^2)$.

Proof. Let $f \in C_p^4$ then applying Ito's formula, one gets

$$\begin{aligned} E\left[f(\bar{X}_t^{2,\varepsilon})\right] &= f(x) + \frac{\lambda_\varepsilon}{2} E\left[\int_0^t \sum_{i,j,k,l} \partial_{ij} f(\bar{X}_s^{2,\varepsilon}) h_{ik} h_{il}(x) l(Y^\varepsilon)^{-1} Y_k^\varepsilon Y_l^\varepsilon ds\right] \\ &= f(x) + \frac{t}{2} \int_{|y| \leq \varepsilon} \sum_{i,j,k,l} \partial_{ij} f(x) h_{ik} h_{il}(x) y_k y_l \nu(dy) + R_\varepsilon(x) \end{aligned}$$

where by Lemma 2.6.14, we have

$$|R_\varepsilon(x)| \leq C \left\|f\right\|_{C_p^4} (1 + |x|^{p+4}) t^2 \int_{|y| \leq \varepsilon} |y|^4 l(y)^{-1} \left(1 + \left(|y|^2 l(y)^{-1} \lambda_\varepsilon t\right)^{\frac{p-2}{2}}\right) \nu(dy).$$

Furthermore

$$\begin{aligned} &L_{d+1}^{2,\varepsilon} f(x) - \frac{1}{2} \int_{|y| \leq \varepsilon} \sum_{i,j,k,l} \partial_{ij} f(x) h_{ik} h_{il}(x) y_k y_l \nu(dy) \\ &= \sum_{i,j,k,l} \int_{|y| \leq \varepsilon} \int_0^1 \left(\partial_{ij} f(x + \alpha h(x) y) - \partial_{ij} f(x)\right) \alpha d\alpha h_{ik} h_{il}(x) y_k y_l \nu(dy). \end{aligned}$$

Therefore

$$\begin{aligned} &\left|L_{d+1}^{2,\varepsilon} f(x) - \frac{1}{2} \int_{|y| \leq \varepsilon} \sum_{i,j,k,l} \partial_{ij} f(x) h_{ik} h_{il}(x) y_k y_l \nu(dy)\right| \\ &\leq C \left\|f\right\|_{C_p^4} (1 + |x|^{p+3}) \int_{|y| \leq \varepsilon} |y|^3 \nu(dy). \end{aligned}$$

This finishes the proof. □

In the particular case that $l(y) = y^r$, $r = 2$, the above scheme corresponds to a Asmussen-Rosiński type approach.

The approximation for L_{d+1}^3 is defined as follows. Let $G_{\varepsilon,l}(dy) = C_{\varepsilon,l}^{-1} l(y) 1_{|y| > \varepsilon} \nu(dy)$, $C_{\varepsilon,l} = \int_{|y| > \varepsilon} l(y) \nu(dy)$ and let $Z^{\varepsilon,l} \sim G_{\varepsilon,l}$ and let $S^{\varepsilon,l}$ be a Bernoulli random variable independent of $Z^{\varepsilon,l}$. Then consider the following two cases. If $S^{\varepsilon,l} = 0$ define $\bar{X}_t^{3,\varepsilon}(x) = x$, otherwise $\bar{X}_t^{3,\varepsilon}(x) = x + h(x) l(Z^{\varepsilon,l})^{-1} Z^{\varepsilon,l}$. Then we have the following results.

Lemma 2.6.16. *Assume that for $p \geq 2$, $\sup_{\varepsilon\in(0,1]} \int_{|y|>\varepsilon} l(y)^{-p} |y|^{p+1} \nu(dy) < \infty$ and $C_{\varepsilon,l}^{-1} P[S^\varepsilon = 1] \leq Ct$ then assumption $(\mathcal{M})$ is satisfied with*

$$E\left[\left|\bar{X}_{d+1}^{3,\varepsilon}(x)\right|^p\right] \leq (1+Kt)|x|^p + K't.$$

Proof. The result follows clearly from $(f(x) = |x|^p)$

$$\begin{aligned}
&P[S^\varepsilon = 1]\left|E\left[f\left(x + h(x)l(Z^{\varepsilon,l})^{-1}Z^{\varepsilon,l}\right) - f(x)\right]\right| \\
&= C_{\varepsilon,l}^{-1} P[S^\varepsilon = 1] \int_{|y|>\varepsilon} \left(f(x + h(x)l(y)^{-1}y) - f(x)\right) l(y)\nu(dy) \\
&\leq Ct(1+|x|^p)\left(1 + \int_{|y|>\varepsilon} l(y)^{-p}|y|^{p+1}\nu(dy)\right).
\end{aligned}$$

□

Lemma 2.6.17. *Assume that for $f \in C_p^2$, we have that $\int_{|y|>\varepsilon} |y|^2(l(y)^{-1}-1) + |y|^{p+2}|l(y)^{-1}-1|^{p+1}\nu(dy) \leq C$ and $\left|C_{\varepsilon,l}^{-1} P\left[S^{\varepsilon,l}=1\right] - t\right| \leq Ct^2$ then*

$$\left|E\left[f(\bar{X}_t^{3,\varepsilon})\right] - f(x) - tL_{d+1}^3 f(x)\right| \leq Ct^2 \|f\|_{C_p^2}(1+|x|^{p+2}).$$

Proof. As before let $f \in C_p^2$ then

$$\begin{aligned}
E\left[f(\bar{X}_t^{3,\varepsilon})\right] &= f(x) + E\left[f\left(x + h(x)l(Z^{\varepsilon,l})^{-1}Z^{\varepsilon,l}\right) - f(x); S^{\varepsilon,l}=1\right] \\
&= f(x) + \int_{|y|>\varepsilon} \left(f(x+h(x)l(y)^{-1}y) - f(x)\right) l(y)\nu(dy) C_{\varepsilon,l}^{-1} P\left[S^{\varepsilon,l}=1\right].
\end{aligned}$$

Then we clearly have that

$$\begin{aligned}
&\left|E\left[f(\bar{X}_t^{3,\varepsilon})\right] - f(x) - tL_{d+1}^3 f(x)\right| \\
&\leq t\left|\int_{|y|>\varepsilon}\int_0^1 \sum_i \left(\partial_i f(x+\alpha h(x)l(y)^{-1}y) - \partial_i f(x+\alpha h(x)y)\right) d\alpha h(x)y\nu(dy)\right| \\
&\quad \times C_{\varepsilon,l}^{-1} P\left[S^{\varepsilon,l}=1\right] \\
&+ \quad \left|\int_{|y|>\varepsilon} f(x+h(x)y) - f(x)\nu(dy)\right| \left|C_{\varepsilon,l}^{-1} P\left[S^{\varepsilon,l}=1\right] - t\right| \\
&\leq C\|f\|_{C_p^2}(1+|x|^{p+2})t^2
\end{aligned}$$

This finishes the proof. □

Using the previous results we can propose various schemes of approximation of order 1 as in Theorem 2.6.13. We state the simplest type of approximation.

Theorem 2.6.18. *Assume that V_0, V and h are infinitely differentiable functions with bounded derivatives with $\int_{|y|>1}|y|^p\nu(dy)<\infty$ for all $p\in\mathbf{N}$. Define $\varepsilon\equiv\varepsilon(T,n)$ so that the conditions on Lemmas 2.6.14, 2.6.15, 2.6.16 and 2.6.17 are satisfied for $t=T/n$ and for appropriate localization functions. Let $\bar{X}^j_{i,t}(x), i=0,...,d$, $j=1,...,n$, n copies of the Euler-Maruyama method for $X_{i,t}(x)$.*

Also, let $\bar{X}^{i,\varepsilon,j}_{d+1,T/n}$, $i=1,2,3$, $j=1,...,n$ be n independent copies of the schemes defined above. Then the following scheme, $X^{(n)}_T=Y^n_n\circ Y^{n-1}_n\circ...\circ Y^1_n(x)$, $Y^j_n(x)=\bar{X}^j_{0,T/n}\circ...\circ\bar{X}^j_{d,T/n}\circ\bar{X}^{1,\varepsilon,j}_{d+1,T/n}\circ\bar{X}^{2,\varepsilon,j}_{d+1,T/n}\circ\bar{X}^{3,\varepsilon,j}_{d+1,T/n}(x)$.is a first order discretization scheme.

Achieving higher order schemes for the approximation of L^2_{d+1} can be easily obtained from the proof of Lemma 2.6.15. In fact, the required conditions are as follows. Assume that for $p\geq 2$,

$$\int_{|y|\leq\varepsilon}|y|^4l(y)^{-1}\left(1+\left(|y|^2\,l(y)^{-1}\lambda_\varepsilon t\right)^{\frac{p-2}{2}}\right)\nu(dy)\leq Ct \tag{2.36}$$

$$\int_{|y|\leq\varepsilon}|y|^3\nu(dy)\leq Ct^2. \tag{2.37}$$

For L^3_{d+1}, the idea used in the previous scheme is that the probability of having more than one jump in an interval of size T/n is of order $(T/n)^2$ and therefore they can be neglected if the goal is to achieve a scheme of order 1. Obviously, in order to obtain a higher order scheme, one has to consider the possibility of more jumps per interval. As an example, we consider the case of at most two jumps per interval with localization $l\equiv 1$.

For L^3_{d+1} one can do the following: Let $G_\varepsilon(dy)=C_\varepsilon^{-1}1_{|y|>\varepsilon}\nu(dy)$, $C_\varepsilon=\int_{|y|>\varepsilon}\nu(dy)$ and let Z^ε_1, $Z^\varepsilon_2\sim G_\varepsilon$ independent between themselves and let S^ε_1 and S^ε_2 be two independent Bernoulli random variable independent of Z^ε_1, Z^ε_2. Then consider the following cases. If $S^\varepsilon_1=0$ define $\bar{X}^{3,\varepsilon}_t(x)=x$, if $S^\varepsilon_1=1$ and $S^\varepsilon_2=0$ then $\bar{X}^{3,\varepsilon}_t(x)=x+h(x)Z^\varepsilon_1$ and finally if $S^\varepsilon_1=1$ and $S^\varepsilon_2=1$ then $\hat{X}^{3,\varepsilon}_t(x)=x+h(x)Z^\varepsilon_1+h(x+h(x)Z^\varepsilon_1)Z^\varepsilon_2$.

Define

$$\begin{aligned}p_\varepsilon&=P\left[S^\varepsilon_1=1\right]\left(1+P\left[S^\varepsilon_2=1\right]\right),\\ q_\varepsilon&=P\left[S^\varepsilon_1=1\right]P\left[S^\varepsilon_2=1\right].\end{aligned}$$

In this case we have

Lemma 2.6.19. *If $C_\varepsilon^{-1}P\left[S^\varepsilon_1=1,S^\varepsilon_2=0\right]\leq Ct$ and $C_\varepsilon^{-2}P\left[S^\varepsilon_1=1,S^\varepsilon_2=1\right]\leq Ct$ then assumption $(\mathcal{M})$ is satisfied with*

$$E\left[\left|\hat{X}^{3,\varepsilon}_{d+1}(x)\right|^p\right]\leq(1+Kt)|x|^p+K't$$

for all $p\geq 2$.

Proof. The result follows clearly from $(f(x) = |x|^p)$

$$P\left[S_1^\varepsilon = 1, S_2^\varepsilon = 0\right] \left|E\left[f\left(x + h(x)Z^\varepsilon\right) - f(x)\right]\right| \leq Ct(1 + |x|^p)\left(1 + \int_{|y|>\varepsilon} |y|^p \, \nu(dy)\right)$$
$$P\left[S_1^\varepsilon = 1, S_2^\varepsilon = 1\right] \left|E\left[f\left(x + h(x)Z_1^\varepsilon + h(x + h(x)Z_1^\varepsilon)Z_2^\varepsilon\right) - f(x)\right]\right|$$
$$\leq Ct(1 + |x|^p)\left(1 + \left(\int_{|y|>\varepsilon} |y|^p \, \nu(dy)\right)^2\right).$$

□

Lemma 2.6.20. *Assume that* $|C_\varepsilon^{-1} p_\varepsilon - t| \leq Ct^3$ *and* $|2C_\varepsilon^{-2} q_\varepsilon - t^2| \leq Ct^3$ *then*

$$\left|E\left[f(\hat{X}_t^{3,\varepsilon})\right] - f(x) - tL_{d+1}^3 f(x) - \frac{t^2}{2}\left(L_{d+1}^3\right)^2 f(x)\right|$$
$$\leq Ct^3 \left\|f\right\|_{C_p^2} (1 + |x|^{p+2})\left(1 + \left(\int_{|y|>\varepsilon} |y|\nu(dy)\right)^2\right).$$

Proof. As before let $f \in C_p^2$ then

$$E\left[f(\hat{X}_t^{3,\varepsilon})\right]$$
$$= f(x) + \int_{|y|>\varepsilon} \left(f(x + h(x)y) - f(x)\right) \nu(dy) C_\varepsilon^{-1} P\left[S_1^\varepsilon = 1, S_2^\varepsilon = 0\right]$$
$$+ E\left[\int_{|y|>\varepsilon} f(x + h(x)y + h(x + h(x)y)Z_2^\varepsilon) - f(x)\nu(dy)\right] C_\varepsilon^{-1} P\left[S_1^\varepsilon = 1, S_2^\varepsilon = 1\right]$$
$$= f(x) + L_{d+1}^3 f(x) C_\varepsilon^{-1} P\left[S_1^\varepsilon = 1, S_2^\varepsilon = 0\right]$$
$$+ \int_{|y|>\varepsilon}\int_{|y|>\varepsilon} f(x + h(x)y + h(x + h(x)y)y_1) - f(x)\nu(dy)\nu(dy_1) C_\varepsilon^{-2} P\left[S_1^\varepsilon = 1, S_2^\varepsilon = 1\right]$$

Hence,

$$E\left[f(\hat{X}_t^{3,\varepsilon})\right] = f(x) + L_{d+1}^3 f(x) C_\varepsilon^{-1}\left(P\left[S_1^\varepsilon = 1\right] + P\left[S_1^\varepsilon = 1, S_2^\varepsilon = 1\right]\right)$$
$$+ \left(L_{d+1}^3\right)^2 f(x) C_\varepsilon^{-2} P\left[S_1^\varepsilon = 1, S_2^\varepsilon = 1\right].$$

Therefore

$$\left|E\left[f(\hat{X}_t^{3,\varepsilon})\right] - f(x) - tL_{d+1}^3 f(x) - \frac{t^2}{2}\left(L_{d+1}^3\right)^2 f(x)\right|$$
$$\leq \left|L_{d+1}^3 f(x)\right| \left|C_\varepsilon^{-1} p_\varepsilon - t\right| + \left|\left(L_{d+1}^3\right)^2 f(x)\right| \left|C_\varepsilon^{-2} q_\varepsilon - \frac{t^2}{2}\right|.$$

Finally note that

$$
\begin{aligned}
&\left(L^3_{d+1}\right)^2 f(x) \\
&= \int_{\varepsilon<|y|}\int_{\varepsilon<|y_1|} \left(f(x+h(x)y+h(x+h(x)y)y_1)-2f(x+h(x)y)+f(x)\right)\nu(dy_1)\nu(dy) \\
&= \int_{\varepsilon<|y|}\int_{\varepsilon<|y_1|}\int_0^1 \nabla f(x+h(x)y+\alpha h(x+h(x)y)y_1)h(x+h(x)y)y_1 \\
&\quad - \nabla f(x+\alpha h(x)y)h(x)y d\alpha\nu(dy_1)\nu(dy)
\end{aligned}
$$

Thus $(L^3_{d+1})^2 f(x)$ is equal to

$$
\begin{aligned}
&\int_{\varepsilon<|y|}\int_{\varepsilon<|y_1|}\int_0^1 \nabla f(x+h(x)y_1+\alpha h(x+h(x)y_1)y) \\
&\quad\times\int_0^1 \nabla h(x+\beta h(x)y_1)h(x)y_1 d\beta y d\alpha\nu(dy_1)\nu(dy) \\
&\quad+\int_{\varepsilon<|y|}\int_{\varepsilon<|y_1|}\int_0^1\int_0^1 D^2 f(x+\alpha h(x)y+\beta(h(x)y_1+\alpha\left(h(x+h(x)y_1)-h(x)\right)y)) \\
&\qquad\times\left[h(x)y_1+\alpha\left(\int_0^1 \nabla h(x+\gamma h(x)y_1)d\gamma h(x)y_1\right)y, h(x)y\right] d\beta d\alpha\nu(dy_1)\nu(dy).
\end{aligned}
$$

This finishes the proof. □

A similar statement can be achieved if we limit the number of jumps in any interval. The parallel of Theorem 2.6.18 can also be stated in this case.

2.6.5 Example: Tempered stable Lévy measure

Now we consider the previous approximation in the case that the Lévy measure ν defined on $\mathbf{R}_0$ is given by

$$\nu(dy) = \frac{1}{|y|^{1+\alpha}}\Big(c_+ e^{-\lambda_+|y|}1_{y>0} + c_- e^{-\lambda_-|y|}1_{y<0}\Big)dy$$

The Lévy process associated with no Brownian term and the above Lévy measure ν is called by

- Gamma: $\lambda_+, c_+ > 0$, $c_- = 0$, $\alpha = 0$.
- Variance gamma: $\lambda_+, \lambda_-, c_+, c_- > 0$, $\alpha = 0$.
- Tempered stable: $\lambda_+, \lambda_-, c_+, c_- > 0$, $0 < \alpha < 2$.

Then, we have that for $\alpha \in [0,1)$

$$\int_{|y|\leq\varepsilon} |y|^k \nu(dy) \sim \varepsilon^{k-\alpha}, \quad k \geq 1.$$

Then $\sup_{\varepsilon\in(0,1]} \int_{|y|\leq\varepsilon} |y|\nu(dy) < \infty$. For L^2_{d+1}, we consider as localization function $l(y) = |y|^r$, then the conditions of Lemma 2.6.15 are satisfied if $\alpha < r \leq 2$ and $\varepsilon = t^{\frac{1}{3-\alpha}}$.

For L^3_{d+1}, we consider as localization $l(y) \equiv 1$, then Lemma 2.6.17 is satisfied for example in the following case. Let $P[S^\varepsilon = 1] = e^{-C_\varepsilon a(\varepsilon,t)}$ where $C_\varepsilon \sim \varepsilon^{-\alpha}$, $a(\varepsilon,t) = -\varepsilon^\alpha \log\left((t^2+t)\,\varepsilon^{-\alpha}\right)$ as $\varepsilon = t^{\frac{1}{3-\alpha}}$ then we have that

$$a = -t^{\frac{\alpha}{3-\alpha}} \log\left((t+1)t^{\frac{3-2\alpha}{3-\alpha}}\right).$$

In the case of Lemma 2.6.20, one choice of parameters is

$$P[S_1^\varepsilon = 1] = t^{\frac{6-3\alpha}{3-\alpha}}(t+1)(1+t^{\frac{\alpha}{3-\alpha}})$$
$$P[S_2^\varepsilon = 1] = \frac{1}{2(1+t^{\frac{\alpha}{3-\alpha}})}.$$

The choice of r in the above scheme is related with variance/importance sampling issues.

Remark 2.6.21. In this chapter we have presented a general set-up to handle what maybe called operator splitting methods. In particular, the method is useful when considering approximations of expectations of functionals of diffusions. The approximation problem is divided in components, each one driven by a single process. This single process, called the coordinate process can be approximated to a high order using an appropriate (stochastic) Runge-Kutta scheme if the driving process is the Brownian motion. In the case that the driving process is a Lévy process one can decompose the Lévy measure in various pieces to facilitate the analysis. Note that sometimes is not needed to know how to simulate Y but only the functional form of the Lévy measure. In comparison with the proposal presented in [45], where high order multiple integrals driven by different Wiener processes have to be simulated at each step, we believe that the present methodology is a better scheme.

The issue that local approximations of high order are interesting to study in comparison with Romberg extrapolations as introduced in [90] is similar to the discussion of using Runge-Kutta approximations in comparison with Romberg extrapolations to approximate solutions of ordinary differential equations. We believe that this article helps to open the path in this direction. In fact, it is somewhat clear from Theorem 2.4.4 that the leading constants in a Euler+Romberg method and a Runge Kutta method do not coincide.

Finally, we used the structure of this construction to easily introduce and analyze the asymptotic error of an approximating scheme for solutions of stochastic differential equations driven by Lévy processes with possibly infinite activity.

Chapter 3

A review of Cubature on Wiener space

This chapter is based on the article by Tanaka [93] published in *RIMS Kôkyûroku*.

3.1 Introduction

In this chapter, we consider the discretization of stochastic differential equations driven by Brownian motions using cubature formulas on Wiener space and study their connection to operator splitting methods discussed in Chapter 2.

A cubature formula for a finite measure ν on $\mathbf{R}^d$ is defined as follows: If there exist positive weights λ_i and points $x_i \in \mathbf{R}^d$ $(1 \leq i \leq k)$ such that for any polynomial $\mathcal{P}$ with degree less than or equal to m

$$\int_{\mathbf{R}^d} \mathcal{P}(x)\nu(dx) = \sum_{i=1}^{k} \lambda_i \mathcal{P}(x_i),$$

then we say that the pair $(\lambda_i, x_i)_{1\leq i\leq k}$ defines a cubature formula with degree m. The existence and construction of the above finite d-dimensional cubature formulas has been well-studied (see e.g. Stroud [84]). One important application of the formula $(\lambda_i, x_i)_{1\leq i\leq k}$ is the numerical integration formula

$$\int_{\mathbf{R}^d} f(x)\nu(dx) \approx \sum_{i=1}^{k} \lambda_i f(x_i)$$

for smooth functions f. The reason why this method works is based on the Taylor expansion or polynomial approximation of f. Therefore the regularity of f is a sufficient condition for the method to work well.

The main objective of this chapter is to review how to construct cubature formulas on Wiener space using splitting methods which have been applied to many research fields such as numerical partial differential equations (e.g. [30]). Cubature

formulas on Wiener space play a similar role to that in finite dimensional space in the calculation of infinite dimensional integrals.

Wiener space is defined as the space of continuous functions $C([0,1];\mathbf{R}^d)$ equipped with the so-called Wiener measure, under which the mapping $B_t := \omega(t)$ for $\omega \in C([0,1];\mathbf{R}^d)$ is a standard Brownian motion. On this space, multiple integrals with respect to the time variable $t \mapsto B_t$ have a similar importance to polynomials in finite dimension. For example, the Itô-Wiener chaos expansion theorem ([37]) shows that L^2-random variables on Wiener space can be expanded by series of the multiple integrals. As seen later, cubature formulas on Wiener space can be applied to numerical approximations of stochastic differential equations, which appear in finance, physics, filtering etc. In this case, the stochastic Taylor expansion gives the error estimation.

This chapter is organized as follows. In Section 3.2, we formulate the cubature formula and prepare some basic tools to discuss algebraic properties of the formula. In Section 3.3, we review the idea of splitting methods for exponential maps and also give some results applicable to the construction of cubature formulas.

3.2 Cubature on Wiener space

3.2.1 Definitions

Let $(B_t^1,\dots,B_t^d)_{t\in[0,1]}$ be a d-dimensional standard Brownian motion on a complete probability space $(\Omega,\mathcal{F},P)$, and set the $(d+1)$-dimensional path $B=(B_t^0,B_t^1,\dots,B_t^d)_{t\in[0,1]}$ with $B_t^0=t$. We use the following notation.

- Let $\alpha\in\mathcal{I}:=\{\emptyset\cup(\cup_{k\in\mathbf{N}}\{0,1,\dots,d\}^k)\}$ be an multi-index and then define the degree of α by

$$\|\alpha\| := \begin{cases} k+\#\{\alpha_j=0\}, & \alpha=(\alpha_1,\dots,\alpha_k)\in\{0,1,\dots,d\}^k, k\geq 1 \\ 0, & \alpha=\emptyset. \end{cases}$$

- $C_{0,BV}([0,t];\mathbf{R}^{d+1})$: the set of all $\mathbf{R}^{d+1}$-valued continuous functions $g=(g_s^0,\dots,g_s^d)_{s\in[0,t]}$ of bounded variation in $[0,t]$ and which start at zero.

- For $\alpha=(\alpha_1,\dots,\alpha_k)\in\{0,1,\dots,d\}^k$, $k\geq 1$, we define the multiple Fisk-Stratonovich integral as follows.

$$I(t,\alpha,\circ dB) := \int_{0<t_k<\cdots<t_1<t} \circ dB_{t_k}^{\alpha_k}\cdots\circ dB_{t_1}^{\alpha_1}.$$

- Similarly, for $g=(g_t^0,\dots,g_t^d)_{t\in[0,1]}\in C_{0,BV}([0,1];\mathbf{R}^{d+1})$, we define

$$I(t,\alpha,dg) := \int_{0<t_k<\cdots<t_1<t} dg_{t_k}^{\alpha_k}\cdots dg_{t_1}^{\alpha_1}.$$

- Throughout the present chapter, we say that a measurable application

$$\omega = (\omega_t)_{t\in[0,1]} : \Omega \to C_{0,BV}([0,1];\mathbf{R}^{d+1})$$

is a "random path" if it satisfies the moment condition

$$I(1,\alpha,d|\omega|) \in L^1(\Omega,\mathcal{F},P)$$

for any $\alpha \in \mathcal{I}$ and almost all $\omega \in \Omega$. Here $|\cdot|$ denotes the total variation path, i.e., $|\omega|_t^i := \sup_{0=t_0<t_1<\cdots<t_k=t} \sum_{j=1}^k |\omega_{t_j}^i - \omega_{t_{j-1}}^i|$.

- A random path ω has finite mass if there exist finite functions $(g_i)_{1\le i\le L} \subset C_{0,BV}([0,1];\mathbf{R}^{d+1})$ and positive weights $(p_i)_{1\le i\le L}$ such that $P(\omega = g_i) = p_i$ and $\sum_{i=1}^L p_i = 1$.

Definition 3.2.1. A random path ω defines a cubature formula with degree m if ω has finite mass and satisfies for every $\|\alpha\| \le m$

$$E[I(1,\alpha,\circ dB)] = E[I(1,\alpha,d\omega)] \quad \Big(= \sum_{i=1}^L p_i I(1,\alpha,dg_i)\Big). \tag{3.1}$$

We denote the space of all random paths which define cubature formulas with degree m by $(\mathrm{Cub})_m$.

Remark 3.2.2. The original paper by Lyons and Victoir [63] assumes $\omega^0(t) = t$. However the above generalization is straightforward.

We extend $(\mathrm{Cub})_m$ to a more general class which includes Ninomiya-Victoir and Ninomiya-Ninomiya schemes (degree 5 formulas). The random path ω for these two schemes already appeared in Kusuoka's papers [54, 55].

Definition 3.2.3. A random path ω defines a moment matching formula with degree m if for every $\|\alpha\| \le m$

$$E[I(1,\alpha,\circ dB)] = E[I(1,\alpha,d\omega)]. \tag{3.2}$$

We denote the space of all random paths satisfying (3.2) by $(\mathrm{M})_m$. Clearly, $(\mathrm{Cub})_m \subseteq (\mathrm{M})_m$.

Example 3.2.4. Here, we give two examples of random paths with moment matching of degrees 3 and 5. That is, elements of $(\mathrm{M})_m$ with $m = 3,\ 5$.

(Degree 3) For each $0 \le i \le d$, define

$$d\omega_t^i := B_1^i dt.$$

Then this ω defines a degree 3 formula. Indeed, due to the symmetry of the Gaussian law of B_1, if $\|\alpha\| = 1$ or 3, $E[I(1,\alpha,\circ dB)] = E[I(1,\alpha,d\omega)] = 0$. Thus it is enough

to check the case $\|\alpha\| = 2$. If $\alpha_1 \neq \alpha_2$, clearly $E[I(1, \alpha, \circ dB)] = E[I(1, \alpha, d\omega)] = 0$. If $\alpha_1 = \alpha_2 = i \geq 1$, we have by Itô's formula, $I(1, \alpha, \circ dB) = (B_1^i)^2/2 = I(1, \alpha, d\omega)$.

(Degree 5) (Ninomiya-Victoir scheme) Let Λ be a random variable with probability

$$P(\Lambda = 1) = P(\Lambda = -1) = 1/2$$

and which is independent of (B_t). Then we define a piecewise smooth path ω by

$$d\omega_t^i := \begin{cases} (d+2)dt, & \text{if } i = 0, t \in [0, \frac{1}{d+2}) \cup [\frac{d+1}{d+2}, 1) \\ (d+2)B_1^i dt, & \text{if } 1 \leq i \leq d, \Lambda = 1, t \in [\frac{i}{d+2}, \frac{i+1}{d+2}) \\ (d+2)B_1^i dt, & \text{if } 1 \leq i \leq d, \Lambda = -1, t \in [\frac{d+1-i}{d+2}, \frac{d+2-i}{d+2}) \\ 0, & \text{otherwise.} \end{cases}$$

We remark $\omega_t^0 \neq t$.

Let ω be a random path and then define the time-scaled path $(\omega_s[t])_{s \in [0,t]}$ by

$$\omega_s^i[t] := \begin{cases} t\omega_{s/t}^0, & i = 0 \\ \sqrt{t}\omega_{s/t}^i, & 1 \leq i \leq d. \end{cases}$$

Under $(\mathrm{Cub})_m$ or $(\mathrm{M})_m$, the scaling property for the Brownian motion (i.e. $B_t \overset{d}{=} \sqrt{t}B_1$ where $\overset{d}{=}$ denotes equality in law, also called equality in distribution) implies

$$E[I(t, \alpha, \circ dB)] = t^{\|\alpha\|} E[I(1, \alpha, \circ dB)] = t^{\|\alpha\|} E[I(1, \alpha, d\omega)] = E[I(t, \alpha, d\omega[t])]$$

for every $t > 0$. Therefore, it is enough to reduce our attention to the case $t = 1$ for the construction of cubature formulas.

3.2.2 Application: random ODE and stochastic Taylor expansion

Let $X_t = X_t^x$ be the unique solution to the stochastic differential equation (SDE)

$$X_t^x = x + \sum_{i=0}^{d} \int_0^t V_i(X_s^x) \circ dB_s^i \tag{3.3}$$

where $V_i \in C_b^\infty(\mathbf{R}^N; \mathbf{R}^N)$. We also define a random ordinary differential equation (ODE)

$$\bar{X}_t^x = x + \sum_{i=0}^{d} \int_0^t V_i(\bar{X}_s^x) d\omega_s^i.$$

We denote by $\bar{X}_t^x(d\omega)$ the solution.

The well-known Itô's formula in stochastic calculus is a fundamental theorem of calculus (or change of variable formula) as follows:

$$f(X_t^x) = f(x) + \sum_{i=0}^{d} \int_0^t (V_i f)(X_s^x) \circ dB_s^i$$

for a smooth function f, where V_i acts on f as a vector field $\sum_{j=1}^{N} V_i^j \frac{\partial}{\partial x_j}$ on $\mathbf{R}^N$. We can apply this formula to the integrands of the (stochastic) integrals. Then we get the so-called stochastic Taylor expansion

$$f(X_t^x) = \sum_{\|\alpha\| \le m} I(t, \alpha, \circ dB)\Big(V_{\alpha_k} \cdots V_{\alpha_1} f\Big)(x) + (\text{remainder}) \ .$$

Of course, we can also apply the fundamental of calculus to the bounded variation function ω, and then we have a similar formula

$$f(\bar{X}_t^x(d\omega[t])) = \sum_{\|\alpha\| \le m} I(t, \alpha, d\omega[t])\Big(V_{\alpha_k} \cdots V_{\alpha_1} f\Big)(x) + (\text{remainder}) \ .$$

We can show the following error estimates by using stochastic Taylor expansions.

Theorem 3.2.5. *Let a random path ω satisfy the condition $(\mathrm{M})_m$. Then for any $f \in C_b^\infty(\mathbf{R}^N; \mathbf{R})$, there exists a constant $C = C(m, f)$ such that*

$$|E[f(X_t^x)] - E[f(\bar{X}_t^x(d\omega[t]))]| \le C t^{(m+1)/2}.$$

Sketch of proof. The fundamental theorem of stochastic calculus (i.e. Itô's formula) can be applied to ω and B;

$$f(X_t^x) - \sum_{\|\alpha\| \le m} I(t, \alpha, \circ dB)\Big(V_{\alpha_k} \cdots V_{\alpha_1} f\Big)(x) =: R_m^X(t, x),$$

$$f(\bar{X}_t^x(d\omega[t])) - \sum_{\|\alpha\| \le m} I(t, \alpha, d\omega[t])\Big(V_{\alpha_k} \cdots V_{\alpha_1} f\Big)(x) =: R_m^{\bar{X}}(t, x).$$

We obtain from the assumption $(\mathrm{M})_m$

$$E\Big[\sum_{\|\alpha\| \le m} I(t, \alpha, \circ dB)\Big(V_{\alpha_k} \cdots V_{\alpha_1} f\Big)(x)\Big] = E\Big[\sum_{\|\alpha\| \le m} I(t, \alpha, d\omega[t])\Big(V_{\alpha_k} \cdots V_{\alpha_1} f\Big)(x)\Big].$$

One can easily check that the remainders R_m^X and $R_m^{\bar{X}}$ consist of the multiple integrals of B and ω with degree $m+1$ and $m+2$. Therefore the result follows from the time-scaling property for B_t and $\omega[t]$. □

If t is not small, we can use a Markov chain type approximation as follows. The precise proof can be obtained by the semigroup approach in Chapter 2.

Theorem 3.2.6. *Let $\omega(i)$ $(1 \le i \le n)$ be i.i.d. random paths satisfying the condition defining* $(\mathrm{M})_m$. *Let us define a new random path $\bar{\omega}$ in $[0,1]$ by*

$$\bar{\omega}_t := (\omega(i))_{t-(i-1)/n}[n^{-1}]$$

for $t \in [\frac{i-1}{n}, \frac{i}{n})$. Then for any $f \in C_b^\infty(\mathbf{R}^N;\mathbf{R})$, there exists a constant $C = C(m,f)$ such that

$$|E[f(X_1^x)] - E[f(\bar{X}_1^x(d\bar{\omega}))]| \le \frac{C}{n^{(m-1)/2}}.$$

3.2.3 Formal series and expansion of SDEs

When we investigate the short time asymptotics of the map $t \mapsto E[f(\bar{X}_t^x(d\omega[t]))]$, the vector fields $V_0, \dots, V_d$ are considered just as the coefficients of the series. On the other hand, we notice that the coefficient of t^k is spanned by $\{V_{\alpha_k} \cdots V_{\alpha_1} f; \|\alpha\| = k\}$. Hence it is natural to regard it as formal power series with respect to the variables $V_0, \dots, V_d$, and to forget the time parameter t with $t = 1$.

To discuss formal power series with variables $V_0, \dots, V_d$, we use the following notation.

- $A = \{v_0, v_1, \cdots, v_d\}$: Alphabets.
- Powers (words) of $v = (v_0, \dots, v_d)$:

$$v^\alpha := \begin{cases} v_{\alpha_k} \cdots v_{\alpha_1}, & \alpha = (\alpha_1, \dots, \alpha_k) \in \{0,1,\dots,d\}^k, k \ge 1 \\ 1, & \alpha = \emptyset. \end{cases}$$

- $\mathbf{R}\langle A \rangle$: $\mathbf{R}$-algebra of noncommutative polynomials on A.
- $\mathbf{R}\langle\langle A \rangle\rangle$: $\mathbf{R}$-algebra of noncommutative formal power series on A with product topology. We regard $\mathbf{R}\langle\langle A \rangle\rangle$ as the space of $\mathbf{R}$-valued functions defined on all powers of A.
- J_m : The projection from $\mathbf{R}\langle\langle A \rangle\rangle$ to polynomials of degree less than or equal to m, i.e.

$$J_m(x) := \sum_{\|\alpha\| \le m} a_\alpha v^\alpha \quad \text{for} \quad x = \sum_\alpha a_\alpha v^\alpha \in \mathbf{R}\langle\langle A \rangle\rangle,\ (a_\alpha) \subset \mathbf{R}.$$

- $\exp(x) := 1 + \sum_{k=1}^\infty \frac{x^k}{k!}$ for $x = \sum_{\|\alpha\|>0} a_\alpha v^\alpha$ and $(a_\alpha) \subset \mathbf{R}$. We note that this mapping is well-defined since $a_\emptyset = 0$ and so $J_{k-1}(x^k) = 0$ for every k.
- $\Gamma\langle\cdot\rangle$: The linear map from $\mathbf{R}\langle A \rangle$ to differential operators defined by $\Gamma\langle v^\alpha \rangle := V_{\alpha_k} \cdots V_{\alpha_1}$.

- $\Gamma\langle t, \cdot\rangle$: The linear map from $\mathbf{R}\langle A\rangle$ to differential operators defined by $\Gamma\langle t, v^\alpha\rangle := t^{\|\alpha\|} V_{\alpha_k} \cdots V_{\alpha_1}$.

Remark 3.2.7. In Lyons-Victoir [63], instead of polynomials on A, they consider the expansion with respect to Lie polynomials generated by

$$[v_{i_1}, [v_{i_2}, [\cdots [v_{i_{k-1}}, v_{i_k}] \cdots]]]$$

($[v_i, v_j] := v_i v_j - v_j v_i$). It is shown that the existence of the function in $C_{0,BV}([0,t]; \mathbf{R}^{d+1})$ corresponding to the exponential map $\exp(\mathcal{L})$ with arbitrary Lie polynomial $\mathcal{L}$ (Chen's theorem on Wiener space). Their approach for constructing cubature formulas consists of two parts: The first is to find a pair of weights (p_i) and Lie polynomials $(\mathcal{L}_i)$ such that $\sum_{i=1}^k p_i J_m(\exp(\mathcal{L}_i)) = J_m(\exp(v_0 + \frac{1}{2}\sum_{i=1}^d v_i^2))$. The second is to construct bounded variation functions that come from $\exp(\mathcal{L}_i)$.

Let us define the $\mathbf{R}\langle\langle A\rangle\rangle$-valued SDE:

$$\mathbf{X}_t = 1 + \sum_{i=0}^d \int_0^t \mathbf{X}_s v_i \circ dB_s^i$$

which has the unique solution given by

$$\mathbf{X}_t = 1 + \sum_{\|\alpha\|>0} I(t, \alpha, \circ dB) v^\alpha.$$

The following result is well-known (e.g. [63]).

Proposition 3.2.8.

$$E[\mathbf{X}_1] = \exp\Big(v_0 + \frac{1}{2}\sum_{i=1}^d v_i^2\Big).$$

Remark 3.2.9. The above result corresponds to the expansion (for $f \in C_b^\infty$)

$$\begin{aligned} E[f(X_t^x)] &= f(x) + \sum_{j=1}^k \frac{t^j}{j!}\Big(V_0 + \frac{1}{2}\sum_{i=1}^d V_i^2\Big)^j f(x) + O(t^{k+1}) \\ &= \Gamma\Big\langle t, J_k\Big(\exp\Big(v_i + \frac{1}{2}\sum_{i=1}^d v_i^2\Big)\Big)\Big\rangle f(x) + O(t^{k+1}). \end{aligned}$$

That is to say, the operation $f(\cdot) \mapsto E[f(X_t^\cdot)]$ has the structure of the form $\exp(t\mathcal{L})$ where $\mathcal{L}$ is the generator of the Markov process X_t and is given by $\mathcal{L} = V_0 + \frac{1}{2}\sum_{i=1}^d V_i^2$.

We next consider the $\mathbf{R}\langle\langle A\rangle\rangle$-valued random ODE defined as

$$\bar{\mathbf{X}}_t = 1 + \sum_{i=0}^{d} \int_0^t \bar{\mathbf{X}}_s v_i d\omega_s^i. \tag{3.4}$$

The solution is denoted by by $\bar{\mathbf{X}}_t(d\omega)$. As in the case of $\mathbf{X}_t$, we can solve $\bar{\mathbf{X}}_t(d\omega)$ as

$$\bar{\mathbf{X}}_t(d\omega) = 1 + \sum_{\|\alpha\|>0} I(t, \alpha, d\omega) v^\alpha.$$

To be more precise, we define the solution of (3.4) rigorously. Let $g \in C_{0,BV}([0,1]; \mathbf{R}^{d+1})$ and $a \in \mathbf{R}\langle\langle A\rangle\rangle$ be fixed and consider the $\mathbf{R}\langle\langle A\rangle\rangle$-valued linear ODE

$$\bar{\mathbf{Y}}_t = a + \sum_{i=0}^{d} \int_0^t \bar{\mathbf{Y}}_s v_i dg_s^i. \tag{3.5}$$

We say that $(\bar{\mathbf{Y}}_t)_{t\geq 0}$ is a solution of (3.5) if the coefficients of $\bar{\mathbf{Y}}_t$ (as a formal series) are continuous function with respect to t and $\bar{\mathbf{Y}}_t$ satisfies the equation (3.5).

Lemma 3.2.10. *The equation (3.5) has the unique solution given by*

$$\bar{\mathbf{Y}}_t = a \left(1 + \sum_{\|\alpha\|>0} I(t, \alpha, dg) v^\alpha \right).$$

Proof. We can check that the function $t \mapsto a(1 + \sum_{\|\alpha\|>0} I(t, \alpha, dg) v^\alpha)$ is a solution of (3.5). If (3.5) has another solution $\tilde{\mathbf{Y}}_t$, then using the Taylor expansion we can derive that $J_m(\bar{\mathbf{Y}}_t - \tilde{\mathbf{Y}}_t) = 0$ for every m. Therefore the uniqueness of solutions holds. □

We can define the solution of (3.4) pathwisely by means of the above lemma. By Proposition 3.2.8, we obtain the equivalent condition for cubature formulas.

Theorem 3.2.11. *Let ω be a random path. Then we have the followings.*
(i) For each $m \in \mathbf{N}$, ω satisfies $(\mathrm{M})_m$ if and only if

$$E[J_m(\bar{\mathbf{X}}_1(d\omega))] = J_m\Big(\exp\Big(v_0 + \frac{1}{2} \sum_{i=1}^{d} v_i^2 \Big) \Big).$$

(ii) Assume that ω has finite mass. Then for each $m \in \mathbf{N}$, ω satisfies $(\mathrm{Cub})_m$ if and only if the above equality holds.

Proof. Notice that $(\mathrm{M})_m$ holds if and only if

$$E[J_m(\bar{\mathbf{X}}_1(d\omega))] - E[J_m(\mathbf{X}_1)] = \sum_{\|\alpha\|\leq m} \Big(E[I(t, \alpha, d\omega)] - E[I(t, \alpha, \circ dB)] \Big) v^\alpha = 0.$$

By using Proposition 3.2.8, this condition holds if and only if

$$E[J_m(\bar{\mathbf{X}}_1(d\omega))] - J_m\Big(\exp\Big(v_0 + \frac{1}{2}\sum_{i=1}^{d} v_i^2\Big)\Big) = 0.$$

□

3.3 Splitting methods and construction of cubature formulas

In Chapter 2, we study the idea of splitting methods (or exponential product formulas) that have been applied to approximations of ODEs, PDEs and more general exponential maps (e.g. [81], [82], [85], [86], [30], [94]). For simplicity, we consider two matrices $A, B \in \mathbf{R}^{k\times k}$ such that $AB \neq BA$. We can easily show by the Taylor expansion

$$\begin{aligned}
\exp(tA)\exp(tB) &= \exp(t(A+B)) + O(t^2),\\
\exp(\frac{t}{2}A)\exp(tB)\exp(\frac{t}{2}A) &= \exp(t(A+B)) + O(t^3),\\
\frac{1}{2}\exp(tA)\exp(tB) + \frac{1}{2}\exp(tB)\exp(tA) &= \exp(t(A+B)) + O(t^3).
\end{aligned}$$

The above computation and basic ideas are applicable to more general (unbounded) operator A, B. As mentioned in Remark 3.2.9, our interest is the case where $A, B(, C, D\dots)$ are generators of some Markov processes.

3.3.1 Splitting method for $\mathbf{R}\langle\langle A\rangle\rangle$-valued SDEs

Let us define $(d+1)$ SDEs considered as the splitting of $\mathbf{X}_t$ in each direction of $(B_t^0, B_t^1, \dots, B_t^d)$. For $i = 0, 1, \dots, d$, define the $\mathbf{R}\langle\langle A\rangle\rangle$-valued SDE

$$\mathbf{X}_t^{(i)} = 1 + \int_0^t \mathbf{X}_s^{(i)} v_i \circ dB_s^i.$$

We can immediately solve the above equations.

Lemma 3.3.1. (a)

$$\mathbf{X}_1^{(0)} = \exp(v_0).$$

(b) *For* $i \geq 1$,

$$\begin{aligned}
\mathbf{X}_1^{(i)} &= \exp(B_1^i v_i),\\
E[\mathbf{X}_1^{(i)}] &= \exp\Big(\frac{v_i^2}{2}\Big).
\end{aligned}$$

Remark 3.3.2. The lemma is an algebraic version of the following probabilistic consideration: Let B_t be a one dimensional Brownian motion, W be a $C_b^\infty(\mathbf{R}^N;\mathbf{R}^N)$-vector field, and $\exp(sW)x$ be the solution to the ODE

$$Y_s^x = x + \int_0^s W(Y_r^x)dr.$$

Then using Itô's formula, we can show that $\exp(B_tW)x$ is the solution to the SDE

$$X_t^x = x + \int_0^t W(X_r^x) \circ dB_r.$$

This equation has much better analytical tractability than the original SDE (3.3) which is driven by a multidimensional Brownian motion. In other words, the splitting methods help us to avoid simulations of "Lévy areas" defined as

$$I^{ij}(t) := \int_0^t \int_0^s \circ dB_r^i \circ dB_s^j - \int_0^t \int_0^s \circ dB_r^j \circ dB_s^i$$

for $i \neq j$. It is important to point out that (i) Levy areas naturally appear in the stochastic Taylor expansion via $\int_0^t \int_0^s \circ dB_r^i \circ dB_s^j = \frac{1}{2}(B_t^i B_t^j + I^{ij}(t))$. (ii) The exact distribution of Lévy areas is not known and it is even difficult to know its moments.

We introduce some formulas of splitting methods with degree 5.

Theorem 3.3.3. [Ninomiya-Victoir scheme]: *For* $Z = (Z^1, \dots, Z^d) \sim N(0, I_d)$,

$$\begin{aligned}
&E\Big[J_5\Big(\frac{1}{2}\big(\exp(v_0/2)\exp(Z^1v_1)\cdots\exp(Z^dv_d)\exp(v_0/2)\\
&\quad + \exp(v_0/2)\exp(Z^dv_d)\cdots\exp(Z^1v_1)\exp(v_0/2)\big)\Big)\Big]\\
&= E\Big[J_5\Big(\frac{1}{2}\big(\exp(v_0)\exp(Z^1v_1)\cdots\exp(Z^dv_d) + \exp(Z^dv_d)\cdots\exp(Z^1v_1)\exp(v_0)\big)\Big)\Big]\\
&= J_5\Big(\exp\Big(v_0 + \frac{1}{2}\sum_{i=1}^d v_i^2\Big)\Big).
\end{aligned}$$

[Strang's splitting type scheme]: *For* $Z = (Z^1, \dots, Z^{2d-1}) \sim N(0, I_{2d-1})$,

$$\begin{aligned}
&E\Big[J_5\Big(\exp(v_0/2)\exp(Z^1v_1/\sqrt{2})\cdots\exp(Z^{d-1}v_{d-1}/\sqrt{2})\exp(Z^dv_d)\\
&\qquad \exp(Z^{d+(d-1)}v_{d-1}/\sqrt{2})\cdots\exp(Z^{d+1}v_1/\sqrt{2})\exp(v_0/2)\Big)\Big]\\
&= J_5\Big(\exp\Big(v_0 + \frac{1}{2}\sum_{i=1}^d v_i^2\Big)\Big).
\end{aligned}$$

Proof. Using the independence of (Z^i), we can derive

$$E[\exp(Z^i v_i)\exp(Z^j v_j)] = E[\exp(Z^i v_i)]E[\exp(Z^j v_j)] = \exp\Big(\frac{v_i^2}{2}\Big)\exp\Big(\frac{v_j^2}{2}\Big)$$

for $i, j \geq 1$, $i \neq j$. Therefore we can obtain the desired results from formal computation of the Taylor series for exponential maps such as we have seen in previous for the matrices A, B. □

Remark 3.3.4. The weight "$\frac{1}{2}$" corresponds to the probability weight of a Bernoulli random variable independent of Z (recall Example 3.2.4).

Another formula is given by Ninomiya and Ninomiya ([68]). They focus on the number of solving or approximating ODEs. The proof differs from Theorem 3.3.3 due to the lack of independence.

Theorem 3.3.5. [Ninomiya-Ninomiya scheme]: *For* $Z = (Z^1, \ldots, Z^{2d}) \sim N(0, I_{2d})$,

$$E\Big[J_5\Big(\exp\Big(\frac{v_0}{2} + \sum_{i=1}^{d}\Big(\frac{1}{2}Z^i + \frac{1}{\sqrt{2}}Z^{d+i}\Big)v_i\Big)\exp\Big(\frac{v_0}{2} + \sum_{i=1}^{d}\Big(\frac{1}{2}Z^i - \frac{1}{\sqrt{2}}Z^{d+i}\Big)v_i\Big)\Big)\Big]$$
$$= J_5\Big(\exp\Big(v_0 + \frac{1}{2}\sum_{i=1}^{d} v_i^2\Big)\Big).$$

Proof. As in the proof of Theorem 3.3.3, it follows from the computation of the moments of correlated Gaussian random variables with degree 2 and 4. □

3.3.2 Construction of paths of bounded variation

We give here a hint for construction of moment matching or cubature formulas of degree 5.

Lemma 3.3.6. *Let* $Z = (Z^0, \ldots, Z^d)$ *be a random variable and for* $0 \leq i \leq d$,

$$d\omega_t^i := Z^i dt.$$

Then

$$\bar{\mathbf{X}}_1(d\omega) = \exp\Big(\sum_{i=0}^{d} Z^i v_i\Big).$$

Proof. The result follows from

$$\bar{\mathbf{X}}_1(d\omega) = 1 + \sum_{j=1}^{\infty}\Big(\sum_{i=0}^{d} Z^i v_i\Big)^j \int_{0<t_j<\cdots<t_1<1} dt_j \cdots dt_1 = 1 + \sum_{j=1}^{\infty}\Big(\sum_{i=0}^{d} Z^i v_i\Big)^j \frac{1}{j!}.$$

□

The above lemma shows the relationship between the exponential $\exp(\sum_{i=0}^d Z^i v_i)$ and the random ODE $\bar{\mathbf{X}}_1(d\omega)$. We now extend this lemma to discuss more general compositions of the exponential maps including Ninomiya-Ninomiya type schemes.

Theorem 3.3.7. (1) *Let $\ell \in \mathbf{N}$ be fixed and $Z = (Z^{ij})_{0\le i\le d, 1\le j\le \ell}$ be a $\mathbf{R}^{(d+1)\ell}$-valued random variable. Let us define for each $1 \le j \le \ell$*

$$d\omega_t^i := \ell Z^{ij} dt, \quad t \in \Big[\frac{j-1}{\ell}, \frac{j}{\ell}\Big). \tag{3.6}$$

Then we have

$$\bar{\mathbf{X}}_1(d\omega) = \exp\Big(\sum_{i=0}^d Z^{i1} v_i\Big) \cdots \exp\Big(\sum_{i=0}^d Z^{i\ell} v_i\Big).$$

(2) *Let $(Z^{0j})_{1\le j\le \ell}$ be non-negative constants such that $\sum_{j=1}^\ell Z^{0j} = 1$, and $(Z^{ij})_{1\le i\le d, 1\le j\le \ell}$ be Gaussian random variables (which need not to be independent). Assume that ω defined in (3.6) satisfies $(\mathrm{M})_m$, and $\bar{Z} = (\bar{Z}^{ij})_{1\le i\le d, 1\le j\le \ell}$ is a discrete $\mathbf{R}^{d\ell}$-valued random variable with probabilities $(p_l)_{1\le l\le L}$ so that*

$$E[\mathcal{P}((Z^{ij})_{1\le i\le d, 1\le j\le \ell})] = \sum_{l=1}^L p_l \mathcal{P}((\bar{Z}^{ij}(l))_{1\le i\le d, 1\le j\le \ell})$$

for any polynomial $\mathcal{P}$ on $\mathbf{R}^{d\ell}$ with degree less than or equal to m. Then a random path $\bar{\omega}$ given by

$$\begin{aligned} d\bar{\omega}_t^0 &:= \ell Z^{0j} dt, \quad t \in \Big[\frac{j-1}{\ell}, \frac{j}{\ell}\Big), \\ d\bar{\omega}_t^i &:= \ell \bar{Z}^{ij} dt, \quad t \in \Big[\frac{j-1}{\ell}, \frac{j}{\ell}\Big), i \ge 1 \end{aligned}$$

defines a cubature formula with degree m.

Proof. The result (1) is obtained from Lemma 3.3.6 and the uniqueness of solutions of $\bar{\mathbf{X}}_t(d\omega)$ (Lemma 3.2.10). Indeed, for $t \in [1/\ell, 2/\ell)$,

$$\begin{aligned} &\exp\Big(\sum_{i=0}^d Z^{i1} v_i\Big) \exp\Big((t-1/\ell)\sum_{i=0}^d Z^{i2} v_i\Big) \\ &= \bar{\mathbf{X}}_{1/\ell}\Big(1 + \sum_{i=0}^d \int_{1/\ell}^t \exp\Big((s-1/\ell)\sum_{i=0}^d Z^{i2} v_i\Big) v_i d\omega_s^i\Big) \\ &= 1 + \sum_{i=0}^d \Big(\int_0^{1/\ell} \exp\Big(s\sum_{i=0}^d Z^{i1} v_i\Big) v_i d\omega_s^i \\ &\quad + \int_{1/\ell}^t \exp\Big(\sum_{i=0}^d Z^{i1} v_i\Big) \exp\Big((s-1/\ell)\sum_{i=0}^d Z^{i2} v_i\Big) v_i d\omega_s^i\Big). \end{aligned}$$

This implies $\bar{\mathbf{X}}_{2/\ell}(d\omega) = \exp\left(\sum_{i=0}^{d} Z^{i1} v_i\right) \exp\left(\sum_{i=0}^{d} Z^{i2} v_i\right)$. We obtain the result for $\bar{\mathbf{X}}_1(d\omega)$ by induction.

Through the representation via exponential maps, we notice that the conditions $(\mathrm{M})_m$ and $(\mathrm{Cub})_m$ depend only on the polynomials of Z with degree less than or equal to m. Therefore the assertion (2) immediately follows. □

Remark 3.3.8. Theorem 3.3.7 lifts the original problem of cubature formula essentially in infinite dimension down the finite dimensional problem of Gaussian measure.

Remark 3.3.9. Let a path of ω in Theorem 3.3.7 be fixed. Then the random ODE $\bar{X}_t(d\omega)$ becomes an ODE which has piecewise random coefficients. We can apply the Runge-Kutta method for the ODE in each interval (See [68], [94]).

Chapter 4

Implementation using interpolated-lattice

This chapter is based on the paper by Tanaka [92] published in *Journal of Computational and Applied Mathematics*.

4.1 Introduction

The aim of this chapter is to construct higher-order discretization schemes based on Markov chain weak approximations and multidimensional interpolation methods. The target problem is to speed up the computation of option prices under commonly used N-dimensional stochastic processes with $N \leq 5$. We will emphasize through computational experiments that the scheme is more effective than the Monte Carlo-based approach *in terms of CPU times* in the three dimensional case.

Let us consider a Stratonovich stochastic differential equation defined as

$$X_t(x) = x + \sum_{i=0}^{d} \int_0^t V_i(X_s(x)) \circ dW_s^i \tag{4.1}$$

with smooth coefficients $V_i : \mathbf{R}^N \to \mathbf{R}^N$ whose derivatives of any order (≥ 0) are bounded. Here $\mathbf{W}_t = (W_t^1, \ldots, W_t^d)$ is a d-dimensional standard Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ and $W_t^0 = t$ for convention. We assume that an underlying asset dynamics (stock, interest rate, etc) follows the above SDE, and attempt to evaluate the following values.

- European option price $P_T f$:

$$P_t f(x) = E[f(X_t(x))] \quad (\text{at time } t = T\text{: maturity}) \tag{4.2}$$

- Bermudan option price H_T: For expiry dates $0 < T_1 < \cdots < T_l = T$,

$$\begin{aligned} H_{T-T_k}(x) &= \max(g_k(x), P_{T_{k+1}-T_k} H_{T-T_{k+1}}(x)) \quad \text{for } k \geq 1 \\ H_T(x) &= P_{T_1} H_{T-T_1}(x) \end{aligned} \tag{4.3}$$

where f and $(g_k)_{1\leq k\leq l}$ are Lipschitz continuous functions and set $H_0 = \max(0, g_l(x))$. The functions (g_k) are sometimes represented as an expectation of SDEs such as (4.2). Several approaches in numerical discretizations have been known. However, as the dimension of the underlying asset process increases, we face the problem of exponentially increasing computational cost, called *the curse of dimensionality*.

In derivative pricing problems, the dimensions N and d become higher (≥ 2) when we evaluate some complex instruments or models such as basket options ($N, d \geq 2$), cross currency derivatives ($N, d \geq 3$), equity/credit hybrid products ($N, d \geq 2$), Asian or other path-dependent options ($N + 1$), multi-factor Markov-functional interest rate models ($N, d \geq 2$), LIBOR market models (N:large, $d \geq 2$), stochastic volatility models ($N + 1$, $d + 1$) and so on. In the following, we review basic methods to evaluate them.

The well-known tree/lattice methods for some Markovian models (e.g. Brownian motion, geometric Brownian motion) are very simple, and work well when the dimension of the state space N is low. Similarly, the PDE approach such as finite difference methods, finite element methods and the method of lines (FDM,FEM,MOL for short, see details in e.g. [78, 96, 34]) has advantages in low dimensional settings ($N = 1, 2$). However if $N \geq 3$, it turns out to be too hard to solve numerically PDEs or Bellman equations that stem from (4.2), (4.3) since we need to solve unrealistically large dimensional linear equations for them. For a similar reason, fast Fourier transforms (FFTs, see e.g. [14]) that are very efficient approach for models whose distributions are known cannot be used for higher dimensions.

Over the past two decade, the sparse grid method has been applied to FDMs, FEMs, MOL and FFTs (e.g. [12], [76]) to overcome the curse of dimensionality. The method decreases the computational cost with respect to "space" discretization from exponential order m^N to $m(\log m)^{N-1}$ where m denotes here the number of nodes for each direction. The essence of this reduction is the representation through an orthogonal hierarchical basis on full grids, and the restriction of its support from $\{l_1, \ldots, l_N \leq l\} \subset \mathbf{N}^N$ to $\{\sum_{j=1}^N l_j \leq l + N - 1\}$ where l_j stands for the level of the associated hierarchical basis of j-th direction. We will use the sparse grid techniques in a part of our scheme explained later.

On the other hand, higher-order "time" discretization problems in the PDE approach have not been used except special situations (Crank-Nicolson for $N = 1$, the alternating direction implicit (ADI) method for $N = 2$ etc, [96]), because of difficulty in performing stability analysis. Though it depends on the situation, time discretization error often accounts for a majority in total error, and thus the second order accuracy is desired.

Some higher-order time discretization algorithms arising from weak approximation problems have been introduced by several authors (cf. [63], [69], [68]; Section 4.2 for details). The weak approximation approach is based on (quasi-)Monte Carlo simulations, therefore, it does not seem easy to use in order to evaluate Bermudan or American options (see related works in [5], [10], [15], [62]). Besides, when we use Monte Carlo simulations, we cannot avoid instability of sensitivity calculation.

Hence, nontrivial variance reduction techniques will be required.

This chapter provides higher-order backward numerical schemes for pricing European and Bermudan-style derivatives *without* Monte Carlo simulations. The goal of this chapter is to find higher-order tree/lattice methods which rely on a semigroup type discretization in order to generate trees. The semigroups can be approximated using small size grids together with the interpolation methods. The sparse grid method plays an important role in avoiding the curse of dimensionality to some extent.

Roughly speaking, the implementation consists of the following steps:

1. Generate grid points G for each time step t_k $(0 = t_0 < \cdots < t_n = T)$, and begin recursion step $2 \to 3$ from $k = n-1$ to 0.

2. If $k = n-1$, then set $u_0 := f$ on $\mathbf{R}^N$. If $k < n-1$, then extend (interpolate or extrapolate) the values $u_{n-(k+1)}$ from the grid points G to whole space $\mathbf{R}^N$ (or sufficiently large domain which contains G).

3. On each grid point $x \in G$, calculate $u_{n-k}(x)$ by

$$\begin{aligned} u_{n-k}(x) &:= E[u_{n-(k+1)}(M_k(x))] \\ &\approx E[u_{n-(k+1)}(X_{t_{k+1}})|X_{t_k} = x] = P_{t_{k+1}-t_k} u_{n-(k+1)}(x) \end{aligned}$$

   for some discrete-valued random variable $M_k(x)$.

Then we expect that $u_n \approx P_T f$. We refer to this two-step backward scheme as the *interpolated lattice* (I-L) scheme. A pioneering work is Chevance [16] who introduces a one-dimensional case of the above type scheme with the projection (choice of the closest grid point) instead of interpolations. More recently, a projection type technique called quantization has been developed in Bally et.al. [5]. Similar ideas via interpolation methods (but, based on Monte Carlo grid sketching) can be found in Berridge [7, chapter 5] and Kargin [43]. We will define an operator $u_j \mapsto u_{j+1}$ (see the definition (4.14)) which corresponds to the recursion step $2 \to 3$, and will show that the operator can be represented by just multiplications of sparse matrices (see the equation (4.21)). In other words, there exists the sparse matrix $\mathbb{M}$ such that

$$u_{j+1} = \mathbb{M} \times u_j \text{ on } G. \tag{4.4}$$

This explicit form of the linear equation is similar to that of explicit FDMs. However, they quite differ in the rate of convergence and the stability (Theorem 4.2.8 for smooth payoffs and Theorem 4.2.10 for Lipschitz payoffs).

This chapter is organized as follows. In Section 4.2, we first explain the details of the I-L scheme with some examples and formulate the convergence results. In Section 4.3, we introduce a space interpolation technique known as the sparse grid method to overcome the curse of dimensionality, and also explain how to implement (4.4) effectively. Section 4.4 gives the precise error estimates for convergence of the

schemes. In Section 4.7 we also discuss the stability property for the approximations in more general cases. Finally, computational tests are provided for some interest rate and cross currency derivatives in Section 4.6. The discussion includes the comparison of sparse grids and Monte Carlo simulations.

4.2 The algorithm

This section establishes numerical schemes for pricing European and Bermudan options. First, we describe the idea of the I-L scheme and discuss one-dimensional examples in computational finance. After that, we state the I-L scheme for general multidimensional diffusions and give our main result about the rate of convergence. Finally, we will apply the scheme to pricing Bermudan-style derivatives.

4.2.1 The methodology and one-dimensional examples

Let $u(t,x) := P_t f(x) = E[f(X_t(x))]$ for the stochastic differential equation (X_t) defined in (4.1). Then under appropriate conditions for $f(\in C^2)$, it is well-known that $u(t,x)$ is a solution of a second order parabolic PDE

$$\begin{cases} \frac{d}{dt}u(t,x) = \mathcal{L}u(t,x) \\ u(0,x) = f(x). \end{cases} \tag{4.5}$$

where $\mathcal{L}$ is a second order differential operator which is given by

$$\mathcal{L} = V_0 + \frac{1}{2}\sum_{i=1}^{d} V_i^2.$$

Here we identify a smooth function $V : \mathbf{R}^N \to \mathbf{R}^N$ with a smooth vector field $\sum_{i=1}^N V^{(i)}\frac{\partial}{\partial x_i}$ on $\mathbf{R}^N$. Considering the Taylor expansion of $t \mapsto u(t,x)$ around $t = 0$, we can prove $u(t,x) = \sum_{j=0}^m \frac{t^j}{j!}(\mathcal{L}^j f)(x) + O(t^{m+1})$ if f is regular (See Lemma 4.4.2). Therefore $u(t,x) = P_t f(x)$ can be considered as a formal exponential map $e^{t\mathcal{L}}$, and as we will see in the following, this fact is useful for constructing higher-order time discretization schemes.

We briefly review how to construct a numerical scheme satisfying that $P_t f(x) \approx \sum_i p_i f(y_i)$ where $y_i = y_i(t,x)$ are nodes in which the approximation process takes values, and $p_i = p_i(t,x)$ are their corresponding probabilities. Let $M_t(x)$ be a random variable which is measurable on $\Omega \times [0,T] \times \mathbf{R}^N$ and define the operator Q_t by $Q_t f(x) = E[f(M_t(x))]$. Then we have the following representation (Markov chain representation):

$$(Q_{T/n})^n f(x) = E[f(M^1_{T/n} \circ \cdots \circ M^n_{T/n}(x))]$$

where $M^i_{T/n}$ are independent copies of $M_{T/n}$ and $\circ$ is defined as $(M^i_t \circ M^j_t)(x) := M^i_t(M^j_t(x))$. Using the semigroup property of P_t, we notice that

$$P_T f(x) - (Q_{T/n})^n f(x) = \sum_{k=0}^{n-1} (Q_{T/n})^k (P_{T/n} - Q_{T/n}) P_{T-\frac{k+1}{n}T} f(x).$$

From this expression, if the process M_t approximates X_t in the sense that it satisfies the local error estimate $(P_t - Q_t)f(x) = O(t^{m+1})$, then we can expect that the Markov chain $M^1_{T/n} \circ \cdots \circ M^n_{T/n}$ converges with m-th order, i.e. $E[f(X_T(x))] - E[f(M^1_{T/n} \circ \cdots \circ M^n_{T/n}(x))] = O(n^{-m})$. Therefore, once we find an algorithm with high acurracy on small time interval $[0, t]$, we can extend it to whole time interval $[0, T]$.

Let us define the notation of ODEs, which are closely related to approximations of Stratonovich SDEs. We denote by V a smooth vector field on $\mathbf{R}^N$.

Definition 4.2.1. $\exp(V)x$ denotes $z_1(x)$ which is a solution at time 1 of the ordinary differential equation

$$\frac{dz_t(x)}{dt} = V(z_t(x)), \ z_0(x) = x. \tag{4.6}$$

Remark 4.2.2. $\exp(W^i_t V)x$ is a solution of the SDE $dX_t = V(X_t) \circ dW^i_t$. More general relations are studied in Kunita [48].

Our interest in approximating P_t by Q_t is to find a stochastic process $\hat{X}_t$ satisfying that

$$E[f(\hat{X}_t(x))] = \sum_{j=0}^{m} \frac{t^j}{j!} (\mathcal{L}^j f)(x) + O(t^{m+1}).$$

We define $\mathcal{L}_0 := V_0$ and $\mathcal{L}_i := \frac{1}{2} V_i^2$ $(1 \le i \le d)$, and consider approximations of $e^{t\mathcal{L}}$ by combining the semigroup operations $e^{t\mathcal{L}_0}, \ldots, e^{t\mathcal{L}_d}$. For example, when $d = 1$, Strang's splitting method [81] is defined as the operation $e^{\frac{t}{2}\mathcal{L}_0} e^{t\mathcal{L}_1} e^{\frac{t}{2}\mathcal{L}_0} = e^{t(\mathcal{L}_0 + \mathcal{L}_1)} + O(t^3)$. Notice that $e^{t\mathcal{L}_i}$ corresponds to the stochastic process $\exp(W^i_t V_i)x$ by Remark 4.2.2, and moreover compositions of operators of the type $(e^{t\mathcal{L}_i})_{0 \le i \le d}$ correspond to combinations of Markov chains determined by $(\exp(W^i_t V_i)x)_{0 \le i \le d}$. This is the key idea of construction of algorithms with respect to time discretization (cf. [53], [63], [69], [68], [94]).

We now consider the case $N = d = 1$. Under the one-dimensional setting, we can define a splitting tree algorithm with 3 successor nodes:

$$\hat{X}_t(x) = \exp(t/2 V_0) \exp(\sqrt{t} Z V_1) \exp(t/2 V_0) x \tag{4.7}$$

where Z takes values in $\pm\sqrt{3}, 0$ with probability 1/6, 1/6 and 2/3 respectively. This scheme corresponds to an approximation of Strang's splitting $e^{\frac{t}{2}\mathcal{L}_0} e^{t\mathcal{L}_1} e^{\frac{t}{2}\mathcal{L}_0}$ and is known to be a second order method (see e.g. [94]).

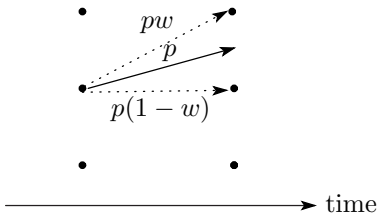

Figure 4.1: Interpolated lattice ($N = 1$): The dots represent grid points on G in each time. p stands for the probability of a node of the lattice and w stands for the weight induced by interpolations. $0 \leq p, w \leq 1$.

Let us demonstrate an example of approximating X_t, given by Hull-White short rate model. The Hull-White model is the solution of the one-dimensional SDE defined by

$$r_t = \theta_t + X_t, \quad dX_t = -\alpha X_t dt + \sigma dW_t, \quad X_0 = 0$$

where θ_t is a deterministic function which depends on the initial yield curve and the parameters α, σ, and also depends on the choice of numeraire (i.e. the choice of forward measure). Then the tree for X_t corresponding to (4.7) is $\hat{X}_t(x) = xe^{-\alpha t} + \sigma e^{-\alpha t/2}\sqrt{t}Z$. This is a non-recombining tree, whose computational cost grows exponentially as the number n increases (compare with Hull-White [35]). To avoid this difficulty, we shall develop in this chapter numerical tree/lattice schemes which decrease grid points in which the conditional expectations are calculated even in the multidimensional case.

We now consider approximating values of conditional expectations on given finite grid points $x_i \in G$. For each time step t_k, let us calculate the following values at $x \in G$, (g represents the conditional expectation which has been calculated in backward recursion)

$$E[g(X_{t_{k+1}})|X_{t_k} = x] \approx E[g(\hat{X}_{t_{k+1}-t_k})|\hat{X}_0 = x] =: Q_{t_{k+1}-t_k}g(x). \tag{4.8}$$

However, we do not know the corresponding values of $g(\cdot)$ since the random variable $\hat{X}_t$ does not always take its values in G. Therefore in order to obtain approximate values of $g(x)$ ($x \notin G$), we should interpolate its value by using the values $g(x_i)$ on grids $G = \{x_i\}$. The idea is viewed in Figure 4.1.

Let $G = \{-K = x_0 < \cdots < x_m = K\}$, $\delta := \max_{1 \leq i \leq m} |x_i - x_{i-1}|$, and assume that the values $g(x_i)$ are known. Then we shall obtain the approximate values of $g(x)$, $x \in [-K, K]$ by interpolations. Here for $g \in C^k([-K, K])$, $g^{(k)}$ denotes its k-th derivative.

Example 4.2.3 (Linear)**.** Define $\tilde{g}(x) = \frac{x_{i+1}-x}{x_{i+1}-x_i}g(x_i) + \frac{x-x_i}{x_{i+1}-x_i}g(x_{i+1})$ for $x \in [x_i, x_{i+1}]$. Then for $g \in C^2([-K, K])$,

$$\|g - \tilde{g}\|_\infty \leq C\|g^{(2)}\|_\infty \delta^2.$$

The piecewise linear interpolation has a lower rate of convergence than piecewise polynomial interpolations, and in contrast, has the advantage of the monotonicity of the L^∞-norm $\|\tilde{g}\|_\infty \leq \|g\|_\infty$. This property allows us to avoid the numerical oscillation from multiple iterations.

Example 4.2.4 (Natural cubic spline)**.** Let $\tilde{g}$ be a piecewise polynomial of degree 3 such that $\tilde{g} \in C^1([-K, K])$ and $\tilde{g}^{(2)}(x_0) = g^{(2)}(x_m) = 0$. Then for $g \in C^4([-K, K])$,

$$\|g - \tilde{g}\|_\infty \leq C\|g^{(4)}\|_\infty \delta^4.$$

This approximate function can be obtained by solving a tridiagonal linear equation.

These procedures provide to approximate the values of $E[f(X_T(x))]$ by backward recursion with respect to the time (t_k) of the underlying processes as follows.

$$\begin{aligned}\{Q_{t_{k+1}-t_k} \cdots Q_{t_n - t_{n-1}} f(x)\}_{x \in \mathbf{R}^N} &\overset{tree/lattice}{\longrightarrow} \{Q_{t_k - t_{k-1}} Q_{t_{k+1}-t_k} \cdots Q_{t_n - t_{n-1}} f(x_i)\}_{x_i \in G} \\ &\overset{interpolation}{\longrightarrow} \{Q_{t_k - t_{k-1}} Q_{t_{k+1}-t_k} \cdots Q_{t_n - t_{n-1}} f(x)\}_{x \in \mathbf{R}^N}.\end{aligned}$$

Note that the lattice can take the values out of $[-K, K]$ near the boundary. Therefore, in that case, we need an appropriate extrapolation technique.

In summary, the total error of the above scheme is represented by the following three terms:

1. time discretization error
2. space discretization (interpolation) error
3. localization (extrapolation) error with respect to boundary parameter K

In the next subsection we shall formulate the problems in general multidimensional settings.

4.2.2 The general algorithm and main result

We now turn to generalize the discussion in Section 4.2.1 to multidimensional stochastic differential equation of the form (4.1). The generalization consists of two parts; (i) multidimensional version of the formula (4.7) and (ii) general interpolation method in dimension N.

We first consider a generalization of the numerical schemes with respect to the formula (4.7), which is interpreted as cubature formulas for multidimensional Gaussian measures and some computation of noncommutative exponential maps. The following formula is a multidimensional generalization for Z that we used in the formula (4.7).

Assumption 4.1 (Gaussian cubature formula with degree 5)**.** Let $Z = (Z^1, \ldots, Z^d)$ be a finite discrete-valued random variable on $\mathbf{R}^d$ satisfying that

$$E[Z^\alpha] = E[\mathbf{W}_1^\alpha]$$

for $|\alpha| \leq 5$. Here we denote $X^\alpha = (X^1)^{\alpha_1} \cdots (X^d)^{\alpha_d}$ for a $\mathbf{R}^d$-vector X.

For details of cubature formulas with degree 5, we refer the reader to Stroud [84] and Victoir [98], in which the cubature formulas with $O(d^2)$ or $O(d^3)$ points can be found. We now introduce a three dimensional formula (14 points, [98]):

$$P\Big(Z = \Big(\eta_1\sqrt{\frac{5}{2}}, 0, 0\Big)\Big) = P\Big(Z = \Big(0, \eta_2\sqrt{\frac{5}{2}}, 0\Big)\Big) = P\Big(Z = \Big(0, 0, \eta_3\sqrt{\frac{5}{2}}\Big)\Big) = \frac{4}{25},$$

$$P\Big(Z = \Big(\eta_4\sqrt{5}, \eta_5\sqrt{5}, \eta_6\sqrt{5}\Big)\Big) = \frac{1}{200}$$

where $\eta_j = \pm 1$, $j = 1, \ldots, 6$.

Before proceeding to the construction of general lattice schemes, we additionally introduce the explicit Runge-Kutta methods for approximating the ODEs since sometimes it is not easy to obtain the closed-form solution.

Definition 4.2.5. We say here that $\exp_{RK}^{(m)}(tV)$ is an s-stage explicit Runge-Kutta method of order m for the ODE (4.6) if it can be written in the form

$$\exp_{RK}^{(m)}(tV)x = x + t\sum_{i=1}^{s} \beta_i k_i(t, V)x \tag{4.9}$$

where $k_i(t, V)x$ defined inductively by

$$k_1(t, V)x = V(x), \qquad k_i(t, V)x = V\Big(x + t\sum_{j=1}^{i-1} \alpha_{i,j} k_j(t, V)x\Big),\ 2 \leq i \leq s,$$

and satisfies

$$|\exp(tV)x - \exp_{RK}^{(m)}(tV)x| \leq C_m |t|^{m+1}. \tag{4.10}$$

The Runge-Kutta formulas $((\beta_i, \alpha_{i,j})_{1 \leq j < i \leq s})$ of order less than or equal to 7 are well known (cf. Butcher [13] for more details).

We are now ready to state a general lattice scheme, inspired by Strang [81] and Ninomiya-Victoir [69], as a multidimensional generalization of the formula (4.7) in Section 4.2.1.

Define a time approximation operator Q_t on $L^\infty(\mathbf{R}^N)$ by

$$Q_t f(x) = E[f(\hat{X}_t(x))] \tag{4.11}$$

where

$$\hat{X}_t(x) :=$$
$$\begin{cases} \exp^{(2)}_{RK}(t/2V_0)\exp^{(5)}_{RK}(\sqrt{t}Z^1V_1)\cdots\exp^{(5)}_{RK}(\sqrt{t}Z^dV_d)\exp^{(2)}_{RK}(t/2V_0)x & \text{if } \Lambda = 1 \\ \exp^{(2)}_{RK}(t/2V_0)\exp^{(5)}_{RK}(\sqrt{t}Z^dV_d)\cdots\exp^{(5)}_{RK}(\sqrt{t}Z^1V_1)\exp^{(2)}_{RK}(t/2V_0)x & \text{if } \Lambda = -1 \end{cases}$$

with a Bernoulli random variable independent of Z and $P(\Lambda = 1) = P(\Lambda = -1) = 1/2$. This approximation comes from (an approximation of) the operation of exponential maps

$$\frac{1}{2}e^{\frac{t}{2}\mathcal{L}_0}e^{t\mathcal{L}_d}\cdots e^{t\mathcal{L}_1}e^{\frac{t}{2}\mathcal{L}_0} + \frac{1}{2}e^{\frac{t}{2}\mathcal{L}_0}e^{t\mathcal{L}_1}\cdots e^{t\mathcal{L}_d}e^{\frac{t}{2}\mathcal{L}_0},$$

which is formally equal to $e^{t\mathcal{L}} + O(t^3)$.

We next begin with the extension of interpolations from one to N dimension. We denote by $C^k_b(\mathbf{R}^N)$ the set of all C^k functions with bounded derivatives of any order (less than or equal to k). Define a seminorm on $C^k_b(\mathbf{R}^N)$ by $\|f\|_{\infty,k} := \sum_{1\le|\alpha|\le k}\|\partial^\alpha f\|_\infty = \sum_{1\le|\alpha|\le k}\sup_{x\in\mathbf{R}^N}|\partial^\alpha f(x)|$. In particular, we often write $C_b(\mathbf{R}^N)$ as $C^0_b(\mathbf{R}^N)$. Let $C_{b,Lip}(\mathbf{R}^N)$ be the set of all bounded Lipschitz functions. The Lipschitz constant for $f \in C_{b,Lip}(\mathbf{R}^N)$ is denoted by $\|f\|_{Lip}$.

We denote $\mathcal{D} := [L_1, R_1]\times\cdots\times[L_N, R_N]$ a domain for interpolations, and $\mathcal{D}^\delta = \{x_j\}_{j=1}^M$ its space discretization (i.e. finite subset in $\mathcal{D}$) with meshwidth parameter δ. Let $C(\mathcal{D})$ be the set of all continuous functions on $\mathcal{D}$, and a norm $\|\cdot\|_\mathcal{D}$ is defined by $\|f\|_\mathcal{D} := \sup_{x\in\mathcal{D}}|f(x)|$. Similarly let $C^k(\mathcal{D})$ be the set of all C^k functions with a seminorm $\|f\|_{\mathcal{D},k} := \sum_{1\le|\alpha|\le k}\|\partial^\alpha f\|_\mathcal{D}$. In general, $\mathcal{D}$ should be included by $\mathrm{supp}(P_{X_T(x)})$ which defines the support of the law of $X_T(x)$ for a fixed x. For example if (X_t) moves only on $[0,\infty)^N$, $\mathcal{D}$ should be a subset of $[0,\infty)^N$. We note that, if an approximation $\tilde{X}^{(n)}_T(x)$ converges $X_T(x)$ in law, then $P(\tilde{X}^{(n)}_T(x)\notin O) \approx P(X_T(x)\notin O) = 0$ for any open set that includes $\mathrm{supp}(P_{X_T(x)})$.

Let $I^\mathcal{D}_\delta$ be an interpolation and extrapolation operator defined for any measurable fuction f on $\mathbf{R}^N$ by

$$I^\mathcal{D}_\delta f(x) := \sum_{j=1}^M b_j(x)f(x_j) \tag{4.12}$$

for $x\in\mathbf{R}^N$ with coefficients $(b_j)_{1\le j\le M}$ determined independently of f. We consider the following technical assumption:

Assumption 4.2. (1) If $f\in C_b(\mathbf{R}^N)$, then $I^\mathcal{D}_\delta f\in C_{b,Lip}(\mathbf{R}^N)$.
(2) Order of accuracy: There exist some $\beta>0$, $\beta^*\in\mathbf{N}$ and $C>0$ (independent of $\mathcal{D}$) such that for any $f\in C(\mathbf{R}^N)$ with $f|_\mathcal{D}\in C^{\beta^*}(\mathcal{D})$,

$$\|I^\mathcal{D}_\delta f - f\|_\mathcal{D} \le C\delta^\beta\|f\|_{\mathcal{D},\beta^*}.$$

(3) Probabilistic representation: $b_j(x)\ge 0$ for every j and $x\in\mathbf{R}^N$, and

$$\sum_{j=1}^M b_j(x)\equiv 1.$$

We note that the assumption 4.2-(3) implies $\|I_\delta^{\mathcal{D}} f\|_\infty \leq \|f\|_\infty$, and $I_\delta^{\mathcal{D}} f \leq I_\delta^{\mathcal{D}} g$ if $f \leq g$.

Remark 4.2.6. The index β^* might depend on the dimension N.

For mathematical simplicity, we also assume

Assumption 4.3 (extrapolation by boundary values)**.** We define for $x \in \mathcal{D}^c$

$$I_\delta^{\mathcal{D}} f(x) = I_\delta^{\mathcal{D}} f(x_*)$$

where x_* is defined by

$$x_*^j = \begin{cases} x^j & x^j \in [L_j, R_j] \\ R_j & x^j > R_j \\ L_j & x^j < L_j. \end{cases}$$

On the whole space $\mathbf{R}^N$, we also define $I^{\mathcal{D}} f := f$ on $\mathcal{D}$, and $I^{\mathcal{D}} f(x) := f(x_*)$ on $\mathcal{D}^c$. Clearly, we have $\|I^{\mathcal{D}} g\|_\infty \leq \|g\|_\infty$. It is important to point out that under Assumption 4.2 (3) and 4.3, there exist transition probability measures $P_x^{\mathcal{D}}(dy)$, $P_x^{\mathcal{D},\delta}(dy)$ such that

$$(I^{\mathcal{D}} f)(x) = \int_{\mathbf{R}^N} f(y) P_x^{\mathcal{D}}(dy) \quad \text{and} \quad (I_\delta^{\mathcal{D}} f)(x) = \int_{\mathbf{R}^N} f(y) P_x^{\mathcal{D},\delta}(dy) \tag{4.13}$$

for $f \in C_b(\mathbf{R}^N)$. In fact, using the Dirac measure $\delta_x(dy)$ concentrated at $x \in \mathbf{R}^N$, we can derive that $P_x^{\mathcal{D}}(dy) = \delta_{x_*}(dy)$ and $P_x^{\mathcal{D},\delta}(dy) = \int_{\mathbf{R}^N} \sum_{1 \leq j \leq M} b_j(z) \delta_{x_j}(dy) P_x^{\mathcal{D}}(dz)$.

The concrete algorithms can be seen in Example 4.2.3-4.2.4, and in Section 4.3. The calculation process $I_\delta^{\mathcal{D}}$ usually consists of two procedures. The first procedure is called only once to determine its coefficients, and the approximate values of $f(x)$ are obtained by calls to fast separate routine (the second procedure).

Remark 4.2.7. The assumption 4.2-(3) is not satisfied in the case of general spline functions. That is, there exist constants $M, m(\delta) > 1$ such that

$$m(\delta) \leq \sup_{f \in C(\mathcal{D}), f \neq 0} (\|I_\delta^{\mathcal{D}} f\|_{\mathcal{D}} / \|f\|_{\mathcal{D}}) \leq M.$$

The detailed explanation can be found in de Boor [21, chapter XIII].

We now define a space-time approximation operator $Q_t^{\delta,\mathcal{D}}$ by

$$Q_t^{\delta,\mathcal{D}} = Q_t \circ I_\delta^{\mathcal{D}}. \tag{4.14}$$

Let $0 = t_0 < t_1 < \cdots < t_n = T$ be a partition of $[0, T]$. We will use the backward algorithm $Q_{t_1 - t_0}^{\delta,\mathcal{D}} \cdots Q_{t_{n-1} - t_{n-2}}^{\delta,\mathcal{D}} Q_{t_n - t_{n-1}} f(x)$ for the approximation of the final value $P_T f(x)$.

Through the probabilistic representations (4.11) and (4.13), we can define three $\mathbf{R}^N$-valued Markov chains as follows:

- Under Assumption 4.1, let $\bar{X}^{(n)}_{t_k}(x)$ be a Markov chain determined by $\bar{X}^{(n)}_0(x) = x$ and $Q_{t_1-t_0}\cdots Q_{t_k-t_{k-1}}f(x) = E[f(\bar{X}^{(n)}_{t_k}(x))]$ for every $1 \le k \le n$, $f \in C_b(\mathbf{R}^N)$.

- Under Assumption 4.1 and 4.3, let $\bar{X}^{(n,\mathcal{D})}_{t_k}(x)$ be a Markov chain determined by $\bar{X}^{(n,\mathcal{D})}_0(x) = x$, $\bar{X}^{(n,\mathcal{D})}_{t_1}(x) = \bar{X}^{(n)}_{t_1}(x)$ and $Q_{t_1-t_0}\cdots(I^{\mathcal{D}}Q_{t_k-t_{k-1}})f(x) = E[f(\bar{X}^{(n,\mathcal{D})}_{t_k}(x))]$ for every $2 \le k \le n$, $f \in C_b(\mathbf{R}^N)$.

- Under Assumption 4.1, 4.2 and 4.3, let $\bar{X}^{(n,\delta,\mathcal{D})}_{t_k}(x)$ be a Markov chain determined by $\bar{X}^{(n,\delta,\mathcal{D})}_0(x) = x$, $\bar{X}^{(n,\delta,\mathcal{D})}_{t_1}(x) = \bar{X}^{(n)}_{t_1}(x)$ and $Q_{t_1-t_0}\cdots(I^{\mathcal{D}}_{\delta}Q_{t_k-t_{k-1}})f(x) = E[f(\bar{X}^{(n,\delta,\mathcal{D})}_{t_k}(x))]$ for every $2 \le k \le n$, $f \in C_b(\mathbf{R}^N)$.

These will be used to analyze the extrapolation error.

Our main result of the present chapter in terms of error estimates is the following.

Theorem 4.2.8. *Let $t_i := \frac{iT}{n}$, $(0 \le i \le n)$. Suppose Assumption 4.1, 4.2 and 4.3 hold. Then for $f \in C_b^{6\vee\beta^*}(\mathbf{R}^N)$, we have*

$$\begin{aligned}\left|P_Tf(x) - \left(Q^{\delta,\mathcal{D}}_{T/n}\right)^{n-1}Q_{T/n}f(x)\right| &\le C_1\|f\|_{\infty,6}\ n^{-2} + C_2\|f\|_{\infty,\beta^*}\ n\delta^{\beta} \\ &\quad + C_3\|\nabla f\|_{\infty}\sqrt{n}\max_{0\le k\le n-2}P(\bar{X}^{(n,\delta,\mathcal{D})}_{kT/n}(x) \notin \mathcal{D}).\end{aligned} \tag{4.15}$$

Here the constants $(C_j)_{1\le j\le 3}$ depend only on T.

In particular, if $\mathcal{D} = [-K_1, K_1]\times\cdots\times[-K_N, K_N]$, the tail probability $P(\bar{X}^{(n,\delta,\mathcal{D})}_{kT/n}(x) \notin \mathcal{D})$ is evaluated as follows: For any $p \in \mathbf{N}$, there exist constants $C_{1,j}$, $C_{2,j}$ $(1 \le j \le N)$ depending on p, T, x such that

$$P(\bar{X}^{(n,\delta,\mathcal{D})}_{kT/n}(x) \notin \mathcal{D}) \le \sum_{1\le j\le N}\frac{C_{1,j} + C_{2,j}k\delta^{\beta}}{K_j^{2p}}.$$

Remark 4.2.9. We remark some theoretical and practical points.

(i) The convergence rate of the form $n^{-\alpha} + n\delta^{\beta}$ $(\alpha, \beta > 0)$ has been investigated in earlier works by Chevance [16] and Bally et al. [5].

(ii) For a general domain $\mathcal{D} = [L_1, R_1]\times\cdots\times[L_N, R_N]$, we should consider the estimate for the tail probabilities $P(c_j(\bar{X}^{(n,\delta,\mathcal{D}),j}_{t_i}(x)) \notin [c_j(L_j), c_j(R_j)])$ with some monotone increasing functions $(c_j)_{1\le j\le N}$.

(iii) Note that, in explicit FDMs, the choice of n (i.e. Δt) strongly depends on δ due to the stability condition in order to avoid numerical oscillation. In contrast, the scheme introduced above does not cause high numerical oscillation in terms of the choice of large Δt. In addition, considering the fact that the time discretization error converges to zero with second order accuracy, we expect

that the error is sufficiently small even if the mesh size $\Delta t = T/n$ is not so small (for example, $n = 10$ with $T = 1$). Conversely, as the number of partition n increases, the space discretization error grows linearly. However, since the number n need not be so large, the error is acceptable to some extent in practice (see numerical tests given later). Incidentally, compared with the time and space discretization error, the error of localization (the third term on the right hand side of (4.15)) is practically negligible.

(iv) Let us consider general SDEs with time-dependent coefficients $V_i(t, x)$. In general we can add an extra process $X_t^0 = t$ to the original process $(X_t^i)_{1 \le i \le N}$, and consider the $(N + 1)$-dimensional SDE $(X_t^0, (X_t^i)_{1 \le i \le N})$, which is of the form (4.1). Therefore the space-time approximation $Q_{T/n}^{\delta, \mathcal{D}}$ has to be defined on the extended state space $[0, \infty) \times \mathbf{R}^N$. In the particular case of interest rate products we can proceed as follows: Let $(T_j)_{j=1}^k$ be maturity dates at which plain options (swaptions) are tradable in a financial market. Then we often use financial models with $V_i(t, x) = V_i(T_j, x), t \in [T_j, T_{j+1})$, so that the approximation $Q_{(T_{j+1}-T_j)/n_j}^{\delta, \mathcal{D}}$ with the state space $\mathbf{R}^N$ can be applied for each partition $[T_j, T_{j+1})$.

4.2.3 Alternative methods for time discretization

The other efficient numerical schemes which give second order accuracy with respect to time discretization can be seen in Ninomiya and Ninomiya [68] and Lyons and Victoir [63]. Ninomiya-Ninomiya scheme uses $2d$ (correlated) Gaussian random variables, so that the number of nodes become more than the above one. On the other hand, Lyons-Victoir scheme uses cubature formulas for multiple Wiener integrals with degree 5, which the number of nodes also tends to increase more than that for Gaussian measures. However, in spite of such limitations, we should not conclude that the scheme described above is best since the total computational cost is also composed of the other aspects such as approximating ODEs, the size of the constants C_i, and so on. At any rate, we shall use the above scheme based on Ninomiya-Victoir throughout this chapter.

When $d = 1$, using the negative time solution of the ODE for V_0, we can find a scheme with more than third order convergence. An example of such consideration is found in Alfonsi [1]. Another very high order scheme with $d = 1$ or 2 (based on a cubature formula for multiple Wiener integrals) is given by Gyurkó and Lyons [32].

4.2.4 Bermudan-style derivatives

The approximation result in previous section can be applied to pricing Bermudan-style derivatives. Let $0 < T_0 < \cdots < T_M$ be expiry dates and define

$$H_{T-T_k}(x) = \max(g_k(x), P_{T_{k+1}-T_k} H_{T-T_{k+1}}(x))$$

where

$$g_k(x) = \begin{cases} f(x) & \text{(option)} \\ \tilde{g}_k(x) & \text{(plain vanilla swaption)} \\ \bar{g}_k(x) + P_{T_{k+1}-T_k} g_{k+1}(x) & \text{(exotic swaption)} \end{cases}$$

as its price at $t = T_k$.

$\tilde{g}_k$ is, for example, of the form $\tilde{g}_k(x) = \sum_{j>k} c_j D(T_k, T_j, x)$ with closed-form bond price (discount function) formulas $D(T_k, T_j, x)$.

The reason why we call the third case "exotic" swaption is that the swap value may not be obtained by closed-form depending on the pricing model or products. In this case, the parallel backward calculation for $P_{T_{k+1}-T_k} H_{T-T_{k+1}}$ and $P_{T_{k+1}-T_k} g_{k+1}$ will be required.

If the partition $T_{k+1} - T_k$ is not so small (e.g. half year), we can identify the numerical procedure on $[T_k, T_{k+1}]$ as the one explained in European option case on $[0, T_{k+1} - T_k]$. Since the function $\max(\cdot, \cdot)$ is not smooth, we cannot apply directly the result in Theorem 4.2.8 to pricing Bermudan-style derivatives even if f is smooth. For this, we shall discuss later some estimates derived from Malliavin calculus. Specifically we review a property of the derivatives of $P_t f$ under the uniformly elliptic condition in the next subsection.

4.2.5 The case of irregular functionals

This subsection describes the results of the rate of convergence when f is not smooth but Lipschitz continuous.

Theorem 4.2.10. *Suppose Assumption 4.1, 4.2, 4.3 and the uniformly elliptic condition, i.e. there exists $\epsilon > 0$ such that $\inf_{x \in \mathbf{R}^N} \sum_{i=1}^d \langle V_i(x), \xi \rangle^2 \geq \epsilon |\xi|^2$ for all $\xi \in \mathbf{R}^N$. Let $s_k = t_k - t_{k-1} > 0$ and $\delta_k > 0$ for each time step. Then we have for any Lipschitz continuous function f,*

$$\begin{aligned} &\left| P_T f(x) - Q_{s_1}^{\delta_1, \mathcal{D}} \cdots Q_{s_{n-1}}^{\delta_{n-1}, \mathcal{D}} Q_{s_n} f(x) \right| \\ &\quad \leq \|f\|_{Lip} \Big(C_1 \Big(\sum_{k=1}^{n-1} \frac{s_k^3}{(T - t_k)^{5/2}} + \sqrt{s_n} \Big) + C_2 \sum_{k=1}^{n-1} \frac{\delta_k^\beta}{(T - t_k)^{(\beta^* - 1)/2}} \\ &\qquad + C_3 \sum_{k=0}^{n-2} \sqrt{s_{k+1}} P(\bar{X}_{t_k}^{(n, \delta, \mathcal{D})}(x) \notin \mathcal{D}) \Big). \end{aligned}$$

Here the constants $(C_j)_{1 \leq j \leq 3}$ depend only on T.

Remark 4.2.11. This theorem asserts that if f has no smooth property, the mesh-width s_k and δ_k should be refined smaller near the final time (maturity date) T.

The algorithm given in this section for Bermudan options is justified by Theorem 4.2.10. Indeed, let us consider time interval $[T_j, T_{j+1}]$ as $[0, T]$ in the case

of pricing European options. By the inequality $\|\max(g_k, P_{T_{k+1}-T_k}H_{T-T_{k+1}})\|_{Lip} \leq \max(\|g_k\|_{Lip}, \|P_{T_{k+1}-T_k}H_{T-T_{k+1}}\|_{Lip})$, we can use the convergence result in Theorem 4.2.10 as $f = \max(g_k, P_{T_{k+1}-T_k}H_{T-T_{k+1}})$ for each expiry date.

More generally, let us consider a bounded measurable function f and compute

$$E[f(X_T + \sqrt{c}\tilde{W})]$$

where $\tilde{W}$ is a standard normal random variable independent of X_t. Then the discussion given in previous can be applied to this expectation since its k-th derivatives grow like $c^{-k/2}$ as $c \downarrow 0$ near the maturity date T. For example, this formulation can be applied to pricing discretely monitored barrier options, if the convolution $y \mapsto E[f(y + \sqrt{c}\tilde{W})]$ is obtained by closed-form (or easy to be approximated).

4.3 Grid sketching and implementation issues

As stated in previous section, we have a motivation to approximate functions by space discretizations. In this section we introduce interpolation techniques on full (rectangular) grids, and on so-called sparse grids.

For notational convenience we use the bijective map $b : \mathcal{D} \to [0,1]^N$ by $b(x) = (\frac{x_1-L_1}{R_1-L_1}, \cdots, \frac{x_N-L_N}{R_N-L_N})$, $x = (x^1, \ldots, x^N) \in \mathbf{R}^N$ and consider $[0,1]$ as the region $[L_i, R_i]$ for each $j = 1, \ldots, N$. Of course, considering the importance of accuracy in each node, we may use the other (nonlinear) coordinate transforms instead of b defined above.

4.3.1 Uniform full grids

This subsection summarizes basic results on the interpolation theory. We denote by $\mathcal{D}^\delta = \left\{(x_{i_1}^{(1)}, \ldots, x_{i_N}^{(N)})\right\}_{0\leq i_j \leq m}^{1\leq j\leq N}$ rectangular grids in $\mathcal{D} = [0,1]^N$ with equidistant meshwidth $\delta = 1/m$ (i.e. $x_i^{(j)} = i/m$).

Piecewise linear interpolation

One simple algorithm is the linear interpolation, which is defined by

$$I_\delta^{\mathcal{D}} f(x) = \sum_{\mathbf{k}=(k_1,\ldots,k_N)\in\{0,1\}^N} b_{\mathbf{i},\mathbf{k}}(x) f(x_{i_1+k_1}^{(1)}, \ldots, x_{i_N+k_N}^{(N)}) \tag{4.16}$$

if $x^j \in [x_{i_j}^{(j)}, x_{i_j+1}^{(j)}]$, where $b_{\mathbf{i},\mathbf{k}}(x) = \delta^{-N} \prod_{1\leq j\leq N}((1-k_j)(x_{i_j+1}^{(j)} - x^j) + k_j(x^j - x_{i_j}^{(j)}))$, $\mathbf{i} = (i_1, \ldots, i_N)$. Note that $I_\delta^{\mathcal{D}}$ is well-defined when $x^j = x_i^{(j)}$ for some i and j. This method satisfies Assumption 4.2 with $\beta = \beta^* = 2$. The computational cost at one point x is very small (N times binary searches and some arithmetic operations). However, the number of nodes needed for the total calculation procedure $\left(Q_{T/n}^{\delta,\mathcal{D}}\right)^{n-1} Q_{T/n} f$ is $O(m^N) \times (n-1)$, so that it does not work well for large N.

Spline interpolation

More generally, we can use piecewise polynomials $(\Phi_i)_{1\le i\le m}$ as follows.

$$I_\delta^{\mathcal{D}} f(x) = \sum_{i=1}^{m} a_i \Phi_i(x) \tag{4.17}$$

with $f(x_j) = I_\delta^{\mathcal{D}} f(x_j)$ for every $x_j \in \mathcal{D}^\delta$. This approximation is also written in the form (4.12). As seen in Example 4.2.4, we expect that the rate of convergence is $\sup_{x\in\mathcal{D}} |f(x) - I_\delta^{\mathcal{D}} f(x)| \le C\delta^\beta$ with $\beta \ge 3$. Therefore, higher-order convergence may decrease the number m needed for required accuracy. However even so, the order $O(m^N)$ still seems to be too large for $N \ge 4$. We remark that polynomial interpolations generally do not satisfy the condition (3) in Assumption 4.2.

Remark 4.3.1 (Least-Square Monte Carlo)**.** We are interested in constructing approximations of $f(x) := E[g(X_{t_{i+1}})|X_{t_i} = x]$ for the SDE (X_t), such as the form of (4.17) with some basis functions (Φ_i). Let $(x_j, y_j)_{1\le j\le M}$ be i.i.d. copies of $(X_{t_i}, g(X_{t_{i+1}}))$. Then the Least-Square Monte Carlo method ([15], [62]) is formulated as

$$\min_{\{a_i\}} \sum_{j=1}^{M} \Big(y_j - \sum_{i=1}^{m} a_i \Phi_i(x_j) \Big)^2$$

in order to get an approximation $\sum_{i=1}^m a_i\Phi_i \approx f$. In this method, what basis functions we should use for general diffusions has not been clarified.

4.3.2 Sparse grids for higher dimensions

In this subsection we explain the sparse grid interpolation techniques to break the curse of dimensionality. It is expected that the method is more effective for the dimension $3 \le N \le 10$, although it may not work like Monte Carlo simulations for several hundred dimensions. For detailed information, see e.g. Bungartz and Griebel [12], Reisinger and Wittum [76], Leentvaar and Oosterlee [59] and the references therein.

A sparse grid interpolation for a function $f(x)$ in dimension N ($x \in \mathbf{R}^N$) is an operator $I_\delta^{\mathcal{D}}$ of the type (4.12) such that

$$\sup_{x\in\mathcal{D}} |f(x) - I_\delta^{\mathcal{D}} f(x)| \le C\delta^2 \log_2(\delta^{-1})^{N-1} \|f\|_{\infty,2} \tag{4.18}$$

and its computational time is $O(\log_2(\delta^{-1})^{N-1})$, and the number of grid points is $O(\delta^{-1}\log_2(\delta^{-1})^{N-1})$ as $\delta \downarrow 0$. This means that the scheme breaks the curse of dimensionality to some extent. Unfortunately, the assumption 4.2-(3) is not satisfied in the case of sparse grid interpolations explained below. Therefore, we need to derive some fine properties of $Q_t^{\delta,\mathcal{D}}$ with respect to the stability. We will discuss it in Section 4.5.

We now introduce the details of the methodology. Let $\phi_{l,i} : [0,1] \to [0,1]$ be a shape function

$$\phi_{l,i}(x) = \begin{cases} 1 - |x/h_l - i|, & x \in [(i-1)h_l, (i+1)h_l] \cap [0,1] \\ 0, & otherwise \end{cases}$$

with $h_l = 2^{-l}$ and $i = 0, 1, \ldots, 2^l$. We also define a product of N piecewise linear functions by

$$\phi_{\mathbf{l},\mathbf{i}}(x) = \prod_{j=1}^{N} \phi_{l_j,i_j}(x^j)$$

with $\mathbf{l} = (l_1, \ldots, l_N) \in (\mathbf{N} \cup \{0\})^N$ and $\mathbf{i} = (i_1, \ldots, i_N)$. Let $V_\mathbf{l}$ be an associated space spanned by $\phi_{\mathbf{l},\mathbf{i}},\ i_j = 0, 1, \ldots, 2^{l_j},\ 1 \le j \le N$. Namely, $V_\mathbf{l}$ corresponds to the space which is spanned by piecewise linear functions on full grids with meshwidth 2^{-l_j} for each direction.

Starting from the above basis functions, we reformulate the linear interpolation (4.16) on full grids. First we consider the case $N = 1$. Let $\alpha_{l,i}$ be given by

$$\alpha_{l,i} = f(x_{l,i}) - \frac{f(x_{l,i-1}) + f(x_{l,i+1})}{2}$$

where $x_{l,j} = j2^{-l}$ and $f : [0,1] \to \mathbf{R}$. In particular, let $\alpha_{l,0} = f(x_{l,0})$ and $\alpha_{l,2^l} = f(x_{l,2^l})$. Let us denote this operation $\alpha_{l,i}$ by $\left[-\frac{1}{2}, 1, -\frac{1}{2}\right]_{l,i} f$. For general N-dimensional spaces, define the coefficients $\alpha_{\mathbf{l},\mathbf{i}}$ by tensor products, that is, for each function $f : [0,1]^N \to \mathbf{R}$,

$$\alpha_{\mathbf{l},\mathbf{i}} := \Big(\prod_{j=1}^{N} \left[-\frac{1}{2}, 1, -\frac{1}{2}\right]_{l_j,i_j} f \Big) := g_0 \in \mathbf{R},$$

where $g_N := f$, $g_j : [0,1]^j \to \mathbf{R}$ is defined as

$$g_j(x^1, \ldots, x^j) := \left[-\frac{1}{2}, 1, -\frac{1}{2}\right]_{l_{j+1},i_{j+1}} g_{j+1}(x^1, \ldots, x^j, \cdot)$$

for $1 \le j \le N-1$ inductively, and $g_0 := [-\frac{1}{2}, 1, -\frac{1}{2}]_{l_1,i_1} g_1$. Then the piecewise linear interpolation denoted by $f_\mathbf{l}$ has another representation

$$f_\mathbf{l} = \sum_{i_1=0}^{l_1} \cdots \sum_{i_N=0}^{l_N} \alpha_{\mathbf{l},\mathbf{i}} \phi_{\mathbf{l},\mathbf{i}}.$$

From now on, we explain how to decrease the number of summations without serious deterioration of the rate of convergence.

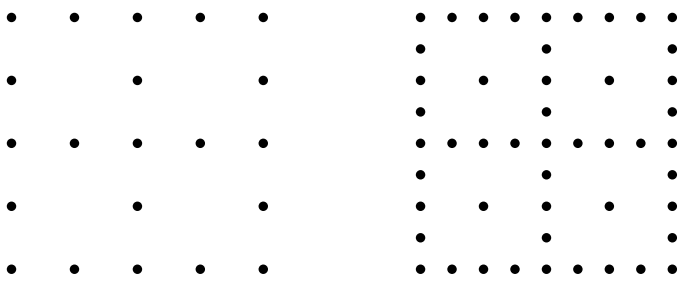

Figure 4.2: Two dimensional sparse grids with level $m_s = 2$ (left) and $m_s = 3$ (right).

Hierarchical subspace: For $\mathbf{l} \in (\mathbf{N} \cup \{0\})^N$, define the hierarchical subspace of $V_\mathbf{l}$ by

$$W_\mathbf{l} = \mathrm{span}\{\phi_{\mathbf{l},\mathbf{i}} \mid \mathbf{i} \in B_\mathbf{l}\}$$

where the index set $B_\mathbf{l}$ is given by

$$B_\mathbf{l} = \left\{ \mathbf{i} \in (\mathbf{N} \cup \{0\})^N \ ; \ \begin{array}{ll} i_j = 1, 3, \ldots, 2^{l_j} - 1, & \text{if } l_j > 0, \\ i_j = 0, 1, & \text{if } l_j = 0, \end{array} \ 1 \le j \le N. \right\}$$

We define the sparse grids of level m_s (≥ 1) by

$$V_{m_s}^s := \bigoplus_{\|\mathbf{l}\|_1 \le m_s + N - 1} W_\mathbf{l}$$

where $\|\mathbf{l}\|_1 = \sum_{j=1}^N \max(l_j, 1)$. Figure 4.2 represents the sparse grids in the case $N = 2$.

We now describe the dimension of the sparse grids (see e.g. [12, Lemma 3.6]).

Lemma 4.3.2. *The number of inner grid points in the subspace $V_{m_s}^s$ is given by*

$$|V_{m_s}^s| = O(2^{m_s} m_s^{N-1}).$$

The comparison of the full grids and the sparse grids can be viewed in Table 4.1. While the number of grid points of the full grids grows exponentially, the one of the sparse grids grows asymptotically linearly.

Sparse grid interpolation: The linear interpolation on the sparse grids is given by

$$f_{m_s}^s = \sum_{\|\mathbf{l}\|_1 \le m_s + N - 1} \sum_{\mathbf{i} \in B_\mathbf{l}} \alpha_{\mathbf{l},\mathbf{i}} \phi_{\mathbf{l},\mathbf{i}}. \tag{4.19}$$

It is known that the above approximation for sufficiently smooth functions fulfills the convergence property (4.18) with $\delta = 2^{-m_s}$, that is, $\sup_{x \in \mathcal{D}} |f(x) - I_\delta^\mathcal{D} f(x)| \le C 2^{-2m_s} m_s^{N-1} \|f\|_{\infty,2}$ (cf. Bungartz and Griebel [12, Theorem 3.8]).

level	2^{m_s}	full grid	sparse grid $V^s_{m_s}$
5	32	35,937	1,505
6	64	274,625	3,713
7	128	2,146,689	8,961
8	256	16,974,593	21,249
9	512	135,005,697	49,665
10	1024	1,076,890,625	114,689

Table 4.1: Number of grid points ($N = 3$)

Combination techniques (Smolyak's algorithm): An alternative approach to formulate sparse grid interpolations is the so-called combination techniques. Using the relationship between hierarchical basis and nodal basis on full grids, one obtains

$$f^s_{m_s} = \sum_{p=0}^{N-1} (-1)^p \begin{pmatrix} N-1 \\ p \end{pmatrix} \sum_{\substack{\|\mathbf{l}\|_1 = m_s+N-1-p \\ l_j>0}} f_{\mathbf{l}},$$

that is, a weighted sum of full grid interpolations on few hierarchical subspaces. The procedure of interpolations on full grids is clear as we have seen in (4.16), and requires some arithmetic operations (whose cost depends only on N) and binary searches. One notices that binary searches are needed only N-times for computing $(f_\ell)_{\|\ell\|\le m_s+N-1}$, and therefore the computational time of the binary searches is of $O(Nm_s)$ and ignorable compared with the other costs. Hence, we conclude that the total computational time to obtain a value of $f^s_{m_s}$ is proportional to the number of summation, $O(m_s^{N-1})$.

An extension of sparse grid interpolation is *non-equidistant* sparse grids. Let γ be a non-negative vector (usually, only one element is positive) and define a γ-non-equidistant sparse grid interpolation as

$$f^s_{m_s,\gamma} := \sum_{p=0}^{N-1} (-1)^p \begin{pmatrix} N-1 \\ p \end{pmatrix} \sum_{\substack{\|\mathbf{l}\|_1 = m_s+N-1-p \\ l_j>0}} f_{\mathbf{l}+\gamma}. \tag{4.20}$$

Denote the corresponding grid points by $V^s_{m_s,\gamma}$. Since currency (or stock) options with stochastic interest rates are mainly dominated by the currency dynamics due to its high volatility compared with interest rate dynamics, it is more effective to refine the mesh of the currency dynamics (see also Leentvaar and Oosterlee [59]). The same argument is also valid for stochastic volatility models.

Data structure: From the above mentioned schemes, the exact computation is due to the correspondence between the pairs of levels $\mathbf{l}$ and coordinate vectors $\mathbf{i}$, and the addresses of data (grid points, function values) which queues up irregularly as follows.

$$\text{key } (\mathbf{l}, \mathbf{i}) \xrightarrow{hash} \text{index (address) } p \longrightarrow \text{grid points } x_p, \text{ values } f \circ b^{-1}(x_p).$$

The so-called hash table allows us to run the mapping with the cost of $O(1)$. Several libraries, such as Boost C++ libraries(unordered_map), provide hash tables for arbitrary data type. For more efficient implementations, see Feuersänger [23].

4.3.3 Sparse matrix formulation: sparse grid interpolated lattice

It may be natural to compute (4.20) for each node of lattice, and each time step. However it could be very expensive in total computational time, since tensor product operations become too many even if we use sparse grids. Let us denote the inner grid points in $V^s_{m_s,\gamma}$ by $\{x_i\}_{1\le i\le n_p}$. To reduce the computational time, we would like to investigate properties of a matrix $\mathbb{M} = \mathbb{M}(T/n, \delta, \mathcal{D}) \in \mathbf{R}^{n_p} \otimes \mathbf{R}^{n_p}$ such that

$$Q^{\delta,\mathcal{D}}_{T/n} f(x_k) = \Big(\mathbb{M} \times \mathbf{f}\Big)^{(k)}$$

where $\mathbf{f} := (f(x_1), \dots, f(x_{n_p}))^T \in \mathbf{R}^{n_p}$ and $(v)^{(k)}$ denotes k-th elemnt of the vector $v \in \mathbf{R}^{n_p}$. We can see that the rows of the matrix $\mathbb{M}$ are sparse vectors through the representation

$$\begin{aligned} Q^{\delta,\mathcal{D}}_{T/n} f(x_k) &= \sum_{i=1}^{l} p_i f^s_{m_s,\gamma}(\hat{X}_{T/n}(x_k, i)) = \sum_{i=1}^{l} p_i \Big\langle v(\hat{X}_{T/n}(x_k, i), \delta, \mathcal{D}), \mathbf{f} \Big\rangle_{\mathbf{R}^{n_p}} \\ &= \Big\langle \sum_{i=1}^{l} p_i v(\hat{X}_{T/n}(x_k, i), \delta, \mathcal{D}), \mathbf{f} \Big\rangle_{\mathbf{R}^{n_p}} \end{aligned} \tag{4.21}$$

where $(\hat{X}_{T/n}(x_k, i))_{i=1}^{l}$ are the finite lattice points of $\hat{X}_{T/n}(x_k)$ defined in (4.11). Here $v(y, \delta, \mathcal{D})$ are the sparse vectors determined by tensor calculation for $f^s_{m_s,\gamma}(y)$ on hierarchical subspaces.

Without loss of generality, we may assume $\gamma = (0, \dots, 0)$. From the definition of tensor calculation on full grids and hierarchical structure of sparse grids, one obtains the following estimates.

Proposition 4.3.3. *The number of nonzero elements in $\mathbb{M}$ is bounded by*

$$N_{cub} \times N_{sp} \times |V^s_{m_s}|$$

where N_{cub} is the number of lattice points depending only on the dimension d, and N_{sp} is the number of nodes used in hierarchical tensor products with $O(m_s^{N-1})$.

Remark 4.3.4.

(i) Clearly, the above upper bound is not optimal because the nodes called by the interpolations are full of redundancies for each lattice point.

(ii) A similar formulation of the matrix form can be used in full grid interpolations (4.16) with the number of nonzero elements $\approx N_{cub} \times 2^N \times (2^{m_s}+1)^N$.

(iii) The computational time mainly consists of generating the matrix $\mathbb{M}$. Furthermore, the bottleneck could be the hash function lookup. We note that the matrix can be used without recalculation to tackle a class of instruments with the same underlying asset dynamics. This property is very convenient for pricing and risk managing large derivative portfolios.

Coarse-grained parallelization: One notices that the operation (4.21) can be run in parallel for each row. Indeed, through the careful consideration to avoid storage confilict, we can use parallel computing techniques for sparse grid interpolated lattice schemes. The detailed procedure is the following.

1. Construct a hash table (on the global memory).
2. For each processor, set (small) local memories used in lattice construction, binary search, tensor calculation, etc.
3. For each processor, execute the operation (4.21) in parallel by using the corresponding local memory.

4.4 Local and global error analysis

The goal of this section is to give the precise error estimates of the rates of convergence, which have been partially indicated in Theorem 4.2.8. The crucial task is to investigate the short time asymptotics of the operators P_t and Q_t.

4.4.1 Basic lemmas for SDEs with smooth coefficients

Lemma 4.4.1 (e.g. [74])**.** *There exists a version of $X_t(x)$ such that a map $x \mapsto X_t(x)$ is infinite times continuous differentiable almost surely and in L^p-convergence sense. Moreover, we have*

$$E[\sup_{0\le t\le T}|X_t(x)|^p] \le C(T,p)(1+|x|^p) < \infty,$$
$$\sup_{x\in\mathbf{R}^N} E[\sup_{0\le t\le T}|\partial_x^\alpha X_t(x)|^p] \le C(T,p,\alpha) < \infty$$

for any $p \ge 2$ and multi-index α with $|\alpha| \ge 1$.

Lemma 4.4.2. *If $f \in C_b^{2(m+1)}(\mathbf{R}^N)$, then $P_t f \in C_b^{2(m+1)}(\mathbf{R}^N)$ and $\|\partial^\alpha P_t f\|_\infty \le C_1(T,m)\|f\|_{\infty,|\alpha|}$ for $|\alpha| \le 2(m+1)$. Moreover, we have*

$$P_t f(x) = \sum_{j=0}^{m} \frac{t^j}{j!}\mathcal{L}^j f(x) + \int_0^t \frac{(t-s)^m}{m!} P_s(\mathcal{L}^{m+1} f)(x) ds$$

and $\|P_t f - \sum_{j=0}^m \frac{t^j}{j!}\mathcal{L}^j f\|_\infty \le C_2(T,m)\ t^{m+1}\|f\|_{\infty,2(m+1)}$.

4.4.2 Proof of Theorem 4.2.8

The following theorem proves the second order convergence (thus locally third order) with respect to the time discretization. The proof is similar to that of [1], [68], and [94].

Theorem 4.4.3. *Let $f \in \mathbf{C}_b^6(\mathbf{R}^N)$. Then we have*

$$\|P_t f - Q_t f\|_\infty \leq C t^3 \|f\|_{\infty,6}.$$

Proof. We first define an approximate operator without Runge-Kutta scheme:

$$\begin{aligned} Q_t^* f(x) =& \frac{1}{2} E[f(\exp(t/2V_0)\exp(\sqrt{t}Z^1V_1)\cdots\exp(\sqrt{t}Z^dV_d)\exp(t/2V_0)x)] \\ &+ \frac{1}{2} E[f(\exp(t/2V_0)\exp(\sqrt{t}Z^dV_d)\cdots\exp(\sqrt{t}Z^1V_1)\exp(t/2V_0)x)]. \end{aligned}$$

In the following we prove that both $P_t - Q_t^*$ and $Q_t^* - Q_t$ converge to zero with third order accuracy.

Step 1: To show the third order convergence of $P_t - Q_t^*$, let us prepare the Taylor formula for ODEs. Expanding $t \mapsto f(\exp(tV_i)x)$ around $t = 0$, we have for any $0 \leq m \leq 6$,

$$f\left(\exp(tV_i)x\right) \;=\; \sum_{j=0}^{m} \frac{t^j}{j!} V_i^j f(x) + \int_0^t \frac{(t-u)^m}{m!} V_i^{m+1} f(\exp(uV_i)x) du.$$

We apply this expansion to the functions inside the expectation (4.11). Then the representation

$$\begin{aligned} f(\exp(s_0V_0)\exp(s_1V_1)\cdots\exp(s_dV_d)\exp(s_{d+1}V_0)x) = \sum_{\|\alpha\|\leq 5} p(\alpha, s, V, f)(x) \\ + \text{ remainder}, \end{aligned}$$

is obtained where $\|\alpha\| := 2(\alpha_0 + \alpha_{d+1}) + \sum_{i=1}^d \alpha_i$ and the functions $p(\alpha, s, V, f) \in C_b(\mathbf{R}^N)$ ($s = (s_0, \ldots, s_{d+1})$, $\alpha = (\alpha_0, \ldots, \alpha_{d+1}) \in (\mathbf{N} \cup \{0\})^{d+2}$) have the form

$$p(\alpha, s, V, f) = a_\alpha s_0^{\alpha_0} \cdots s_{d+1}^{\alpha_{d+1}} (V_{\alpha_0} \cdots V_{\alpha_{d+1}} f)$$

for some $a_\alpha \in \mathbf{R}$. For obtaining the constants a_α of $p(\alpha, s, V, f)$, we refer the reader to e.g. [68], [94]. We can observe here that $\sum_{\|\alpha\|=0} p(\alpha, s, V, f) = f$,

$$\sum_{\|\alpha\|=2} p(\alpha, s, V, f) = (s_0 + s_{d+1})V_0 f + \sum_{i=1}^d \frac{s_i^2}{2} V_i^2 f + \sum_{1\leq i<j\leq d} s_i s_j V_j V_i f,$$

$$\begin{aligned}\sum_{\|\alpha\|=4} p(\alpha,s,V,f) &= (\frac{s_0^2+s_{d+1}^2}{2}+s_0s_{d+1})V_0^2f+\sum_{i=1}^d(s_{d+1}s_i^2V_0V_i^2+s_0s_i^2V_i^2V_0)f\\ &\quad+\sum_{i=1}^d\frac{s_i^4}{24}V_i^4f+\sum_{1\le i<j\le d}\frac{s_i^2s_j^2}{4}V_j^2V_i^2f\\ &\quad+\sum_{\|\alpha\|=4,1\le\exists i\le d\ s.t.\ \alpha_i=1} p(\alpha,s,V,f),\end{aligned}$$

and the remainder term is bounded by

$$C'(s_0^3+s_1^6+\cdots+s_d^6+s_{d+1}^3)\|f\|_{\infty,6}.$$

On the other hand, we obtain by 4.4.1

$$\|P_tf-(f+t\mathcal{L}f+\frac{t^2}{2}\mathcal{L}^2f)\|_\infty\le C_1t^3\|f\|_{\infty,6}.$$

Accordingly, the moment condition for Z up to the order 5 shows that

$$\|P_tf-Q_t^*f\|_\infty\le C_2t^3\|f\|_{\infty,6}.$$

Step 2: Let now derive the error estimate of $Q_t^*-Q_t$. It holds that

$$\begin{aligned}&\Big|\prod_{i=0}^{d+1}\exp(s_iV_i)x-\prod_{i=0}^{d+1}\exp_{RK}^{(k_i)}(s_iV_i)x\Big|\\ &\le\sum_{i=0}^{d+1}\Big|\Big(\prod_{j=0}^{i-1}\exp(s_jV_j)\Big)\exp(s_iV_i)\Big(\prod_{j=i+1}^{d+1}\exp_{RK}^{(k_j)}(s_jV_j)\Big)x\\ &\quad-\Big(\prod_{j=0}^{i-1}\exp(s_jV_j)\Big)\exp_{RK}^{(k_i)}(s_iV_i)\Big(\prod_{j=i+1}^{d+1}\exp_{RK}^{(k_j)}(s_jV_j)\Big)x\Big|\end{aligned}$$

where we denote $V_{d+1}:=V_0$, $k_0=k_{d+1}=2$ and $k_i=5$ $(1\le i\le d)$. Using the Lipschitz continuity $|\exp(sV_i)x-\exp(sV_i)y|\le C'e^{C''|s|}|x-y|$ (that follows from Gronwall's inequality), we can show that

$$\begin{aligned}\text{RHS} &\le \sum_{i=0}^{d+1}(C')^ie^{C''\sum_{j=0}^{i-1}|s_j|}\Big|\Big(\exp(s_iV_i)-\exp_{RK}^{(k_i)}(s_iV_i)\Big)\Big(\prod_{j=i+1}^{d+1}\exp_{RK}^{(k_j)}(s_jV_j)\Big)x\Big|\\ &\le \sum_{i=0}^{d+1}\tilde{C}'e^{C''\sum_{j=0}^{i-1}|s_j|}|s_i|^{k_i+1}.\end{aligned}$$

Consequently, since f is Lipschitz continuous, we have

$$\|Q_t^*f-Q_tf\|_\infty\le C_3t^3\|\nabla f\|_\infty.$$

This is the end of the proof. □

Now we prove the result in global order of convergence when f is smooth.

Proof of Theorem 4.2.8. First of all, we consider the decomposition

$$P_t f - Q_t^{\delta,\mathcal{D}} f = (P_t f - Q_t f) + Q_t(f - I_\delta^{\mathcal{D}} f) \tag{4.22}$$

for $f \in C_b^{6\vee\beta^*}(\mathbf{R}^N)$. By Theorem 4.4.3, the first term is bounded by $C_1 t^3 \|f\|_{\infty,6}$. Assumption 4.2 (2) for the space discretization gives that

$$|Q_t(f - I_\delta^{\mathcal{D}} f)(y)| \le C_2 \delta^\beta \|f\|_{\infty,\beta^*} + |Q_t\big((f - I_\delta^{\mathcal{D}} f)1_{\mathcal{D}^c}\big)(y)|.$$

Note that if $y \in \mathcal{D}$, we have

$$\begin{aligned} |Q_t\big((f - I_\delta^{\mathcal{D}} f)1_{\mathcal{D}^c}\big)(y)| &\le C\|\nabla f\|_\infty \big(\sup_{y\in\mathcal{D}} \|\hat{X}_t(y) - y\|_{L^\infty(\Omega;\mathbf{R}^N)}\big) P(\hat{X}_t(y) \notin \mathcal{D}) \\ &\le C'\|\nabla f\|_\infty \sqrt{t} P(\hat{X}_t(y) \notin \mathcal{D}). \end{aligned} \tag{4.23}$$

Here we used Assumption 4.3 for the first inequality and Assumption 4.1 for the second inequality.

By the semigroup property, we have

$$\begin{aligned} P_T f - \big(Q_{T/n}^{\delta,\mathcal{D}}\big)^{n-1} Q_{T/n} f = (Q_{T/n}^{\delta,\mathcal{D}})^{n-1}(P_{T/n} - Q_{T/n}) f \\ + \sum_{k=0}^{n-2} (Q_{T/n}^{\delta,\mathcal{D}})^k (P_{T/n} - Q_{T/n}^{\delta,\mathcal{D}}) P_{T-\frac{k+1}{n}T} f. \end{aligned}$$

We also have by the assumptions $\|Q_t^{\delta,\mathcal{D}} f\|_\infty \le \|I_\delta^{\mathcal{D}} f\|_\infty \le \|f\|_\infty$. (This property implies the stability of the scheme for multiple iterations.) We now recall that $P_s f \in C_b^{6\vee\beta^*}$ (see 4.4.1). Accordingly, we have

$$\begin{aligned} &\Big|P_T f(x) - \big(Q_{T/n}^{\delta,\mathcal{D}}\big)^{n-1} Q_{T/n} f(x)\Big| \\ &\le C_1 \|f\|_{\infty,6} n^{-2} + C_2 \|f\|_{\infty,\beta^*} n \delta^\beta \\ &+ \sum_{k=0}^{n-2} |(Q_{T/n}^{\delta,\mathcal{D}})^k Q_{T/n}\big((P_{T-\frac{k+1}{n}T} f - I_\delta^{\mathcal{D}} P_{T-\frac{k+1}{n}T} f)1_{\mathcal{D}^c}\big)(x)|. \end{aligned}$$

Since $(Q_{T/n}^{\delta,\mathcal{D}})^k$ acts on the values in $\mathcal{D}^\delta$, we obtain from (4.23)

$$\begin{aligned} &|(Q_{T/n}^{\delta,\mathcal{D}})^k Q_{T/n}\big((P_{T-\frac{k+1}{n}T} f - I_\delta^{\mathcal{D}} P_{T-\frac{k+1}{n}T} f)1_{\mathcal{D}^c}\big)(x)| \\ &\le C_T \|\nabla f\|_\infty n^{-1/2} P(\bar{X}_{kT/n}^{(n,\delta,\mathcal{D})}(x) \notin \mathcal{D}). \end{aligned}$$

This finishes the proof of the first claim (4.15).

Finally, we turn to prove the second claim. We notice that $P(\bar{X}_{kT/n}^{(n,\delta,\mathcal{D})}(x) \notin \mathcal{D}) \le \sum_{1\le j\le N} P(\bar{X}_{kT/n}^{(n,\delta,\mathcal{D}),j}(x) \notin [-K_j, K_j])$, and for any $p \in \mathbf{N}$, $K_j^{2p} P(\bar{X}_{kT/n}^{(n,\delta,\mathcal{D}),j}(x) \notin [-K_j, K_j]) \le E[|\bar{X}_{kT/n}^{(n,\delta,\mathcal{D}),j}(x)|^{2p}]$.

We can give the following estimate for $\hat{X}_t$: Let $f_p(x) := |x|^p$, $x \in \mathbf{R}^N$. Then

$$Q_t f_{2p}(x) = E[f_{2p}(\hat{X}_t(x))] \le (1 + C_{p,T}t) f_{2p}(x) + C'_{p,T}t.$$

(See also the above estimate in [94, Lemma 4.1, Proposition 6.7] under more general settings.) Combining the above estimate and Assumption 4.2, we have

$$I_\delta^{\mathcal{D}} Q_t f_{2p}(x) \le (1 + C_{p,T}t)(I_\delta^{\mathcal{D}} f_{2p})(x) + C'_{p,T}t \le (1 + C_{p,T}t) f_{2p}(x) + C'_{p,T}t + C''_{p,T}\delta^\beta.$$

The second inequality is obtained from $f_{2p} \in C^\infty$. This implies $E[|\bar{X}^{(n,\delta,\mathcal{D}),j}_{kT/n}(x)|^{2p}] \le C_{1,j} + C_{2,j}k\delta^\beta$ for some $C_{1,j}, C_{2,j}$, and therefore $K_j^{2p} P(\bar{X}^{(n,\delta,\mathcal{D}),j}_{kT/n} \notin \mathcal{D}) \le C_{1,j} + C_{2,j}k\delta^\beta < \infty$. □

4.4.3 Proof of Theorem 4.2.10

As already seen in the previous proof, the smoothness of $P_t f$ plays an important role for the error estimates, and thus some regularity results for $\partial^\alpha P_t f$ are required. The following theorem can be found in [57, 83].

Theorem 4.4.4. *Under the uniformly elliptic condition, we have*

$$\|\partial^\alpha P_t f\|_\infty \le C t^{-\frac{|\alpha|-1}{2}} \|\nabla f\|_\infty \tag{4.24}$$

for multi-index α and $f \in C_b^{|\alpha|}(\mathbf{R}^N)$.

Remark 4.4.5. Similar extended results in [57], [52] show that $P_t f$ has same estimates (at least) to the direction of the vector fields instead of ∂^α.

The above result in Malliavin calculus allows us to prove

Proof. By simple calculation, we can show that the inequality

$$\|P_t f - Q_t f\|_\infty \le \|P_t f - f\|_\infty + \|f - Q_t f\|_\infty \le C_1 \sqrt{t} \|\nabla f\|_\infty$$

holds. As in the proof of Theorem 4.4.3, we have for $f \in C_b^{6 \vee \beta^*}$ and $s > 0$,

$$\begin{aligned} |P_t P_s f(y) - Q_t^{\delta,\mathcal{D}} P_s f(y)| \le{}& C_1 t^3 \|P_s f\|_{\infty,6} + C_2 \delta^\beta \|P_s f\|_{\infty,\beta^*} \\ &+ |Q_t\big((P_s f - I_\delta^{\mathcal{D}} P_s f) 1_{\mathcal{D}^c}\big)(y)|. \end{aligned}$$

Therefore Theorem 4.4.4 implies

$$\text{RHS} \le C_1 \frac{t^3}{s^{5/2}} \|\nabla f\|_\infty + C_2 \frac{\delta^\beta}{s^{(\beta^*-1)/2}} \|\nabla f\|_\infty + |Q_t\big((P_s f - I_\delta^{\mathcal{D}} P_s f) 1_{\mathcal{D}^c}\big)(y)|.$$

Consequently, we obtain by approximations of f the above result with bounds $\|f\|_{Lip}$ instead of $\|\nabla f\|_\infty$. □

4.5 Discussion: stability analysis

As a further research, it is quite important to provide the error estimates for interpolated lattice schemes inducing general interpolations; for example, cubic spline interpolations ($N = 1$), sparse grid interpolations ($N \geq 2$). Because of lack of stability by means of Assumption 4.2-(3), we have only obtained the local error estimates (i.e. consistency) in Theorem 4.4.3 and (4.22) for such general interpolation algorithms.

In the present chapter, we cannot indicate a precise answer (i.e. the exact rate of convergence) to this problem. But we shall give some hints in order to deal with them.

Throughout this section, we use the notation of a norm $\|\cdot\|$, which is defined on a subspace of $C_b(\mathbf{R}^N)$, and is not only $\|\cdot\|_\infty$. The norm $\|\cdot\|$ satisfies at least that $\|Q_t\| \leq q(t) = 1 + O(t)$.

4.5.1 Von Neumann stability analysis

First, through the matrix representation of the scheme, we can analyze the stability of matrix multiplications.

Consider Fourier series on $\mathcal{D} = [L_1, R_1] \times \cdots \times [L_N, R_N]$ for f:

$$f(x_j) = \sum_{\mathbf{k} \in \mathbf{Z}^N} C_{\mathbf{k}} \exp(\sqrt{-1}\langle \mathbf{k}, x_j \rangle)$$

and also for $\mathbb{M}f = Q^{\delta,\mathcal{D}}_{T/n} f$ (on $\mathcal{D}^\delta$):

$$(\mathbb{M}f)(x_j) = \sum_{\mathbf{k} \in \mathbf{Z}^N} C_{\mathbf{k}} (\mathbb{M} \exp(\sqrt{-1}\langle \mathbf{k}, \cdot \rangle))(x_j)$$

where $C_{\mathbf{k}}$ is the Fourier coefficients of f. Then define an amplification factor $\rho^{(\mathbf{k})}$ by

$$\rho^{(\mathbf{k})} := \max_{x_j \in \mathcal{D}^\delta} \left| \frac{(\mathbb{M} \exp(\sqrt{-1}\langle \mathbf{k}, \cdot \rangle))(x_j)}{\exp(\sqrt{-1}\langle \mathbf{k}, x_j \rangle)} \right|$$

for each N-component $\mathbf{k}$.

By the discussion of Lax-Richtmyer theorem in [58], the upper bound estimates for $\|\mathbb{M}\| = \|Q^{\delta,\mathcal{D}}_{T/n}\|$ is given by the estimates for $\sup_{\mathbf{k}} \rho^{(\mathbf{k})} = \sup_{\mathbf{k}} \rho^{(\mathbf{k})}(T/n, \delta, \mathcal{D})$. We are concerned with when it holds that

$$\sup_{\mathbf{k}} \rho^{(\mathbf{k})} \approx 1.$$

Sometimes $\sup_{\mathbf{k}} \rho^{(\mathbf{k})}$ is analyzed theoretically, for instance, in case of some FDMs. In the case where explicit calculation of $\rho^{(\mathbf{k})}$ is difficult, computational experiments will help us to understand (empirically) whether the scheme is stable or not.

4.5.2 An analysis of operator norm

Second we give a direct approach using an operator norm defined on $C_b(\mathbf{R}^N)$. We consider an inequality

$$\|I_\delta^{\mathcal{D}} g\| \leq c(\delta)\|g\|$$

instead of Assumption 4.2-(3). The analysis of the asymptotics of $c(\delta)$ is equivalent to that of

$$\delta \mapsto \sup_{g \in \mathcal{C}, g \neq 0} \frac{\|I_\delta^{\mathcal{D}} g\|}{\|g\|}$$

where a certain subspace $\mathcal{C} \subset C_b(\mathbf{R}^N)$ includes the error functions $P_{kT/n}f - (Q_{T/n}^{\delta,\mathcal{D}})^k f$, $1 \leq k \leq n-1$ for a fixed initial function f.

A key problem is to ascertain whether it holds that

$$c(\delta) \to 1 \text{ as } \delta \downarrow 0.$$

This is known as a very complex problem in the field of interpolation theory (see e.g. [21]). If the above type estimate holds, then we rewrite Theorem 4.2.8 as follow:

$$\begin{aligned}\|P_{kT/n} - (Q_{T/n}^{\delta,\mathcal{D}})^k f\| &\leq \|(P_{T/n} - Q_{T/n}^{\delta,\mathcal{D}})P_{(k-1)T/n} f\| \\ &\quad + \|Q_{T/n}^{\delta,\mathcal{D}}(P_{(k-1)T/n} - (Q_{T/n}^{\delta,\mathcal{D}})^{k-1}) f\| \\ &\leq \text{ (local error) } + q(T/n)c(\delta)\|(P_{(k-1)T/n} - (Q_{T/n}^{\delta,\mathcal{D}})^{k-1}) f\|\end{aligned}$$

and hence

$$\text{(global error)} \leq n\ (q(T/n)c(\delta))^n \times \text{ (local error)}$$

with $c(\delta) \to 1$. In practice, only several tens of the number of partition n is needed. Thus the $c(\delta)^n$ term may have a small effect on total error.

Under such general conditions, we show a partial result of the behaviour of the error $P_{kT/n} - (Q_{T/n}^{\delta,\mathcal{D}})^k f$ with the supremum norm.

Modify the assumptions introduced in Section 4.2 as follows.

Assumption 4.4. Assumption 4.2-(3) is not satisfied, and alternatively suppose that
(1) Convergence for Lipschitz continuous functions:

$$\lim_{\delta \downarrow 0} \|I_\delta^{\mathcal{D}} g - g\|_{\mathcal{D}} = 0$$

for any Lipschitz continuous function g.
(2) Uniformly boundedness: For a fixed $\delta^* > 0$,

$$\sup_{0 < \delta \leq \delta^*} \sup_{g \in C(\mathcal{D}), g \neq 0} \frac{\|I_\delta^{\mathcal{D}} g\|_{\mathcal{D}}}{\|g\|_{\mathcal{D}}} < \infty.$$

Now we can prove the next lemma.

Lemma 4.5.1. *Suppose that* $(g_\delta), g$ *are Lipschitz continuous and* $\lim_{\delta\downarrow 0}\|g_\delta - g\|_\infty = 0$. *Then*

$$\lim_{\delta\downarrow 0}\|I_\delta^{\mathcal{D}} g_\delta - I^{\mathcal{D}} g\|_\infty = 0.$$

Proof. It immediately follows from $\|I_\delta^{\mathcal{D}} g_\delta - I^{\mathcal{D}} g\|_\infty \le \|I_\delta^{\mathcal{D}} g_\delta - I_\delta^{\mathcal{D}} g\|_\infty + \|I_\delta^{\mathcal{D}} g - I^{\mathcal{D}} g\|_\infty \le \|I_\delta^{\mathcal{D}}\|_\infty \|g_\delta - g\|_\infty + \|I_\delta^{\mathcal{D}} g - I^{\mathcal{D}} g\|_\infty$ and Assumption 4.4. □

Additionally, one can easily obtain

Lemma 4.5.2. *For any Lipschitz continuous function* g, $P_t g$ *and* $Q_t g$ *are also Lipschitz continuous.*

We can eventually obtain the following convergence result.

Theorem 4.5.3. *Under Assumption 4.1, 4.3 and 4.4, we have for* $f \in C_b^{6\vee\beta^*}(\mathbf{R}^N)$

$$\lim_{\delta\to 0}\left|P_T f(x) - \left(Q_{T/n}^{\delta,\mathcal{D}}\right)^{n-1} Q_{T/n} f(x)\right| \le C_1 \|f\|_{\infty,6} n^{-2} + C_3 \|\nabla f\|_\infty \sqrt{n} \max_{0\le k\le n-2} P(\bar{X}_{kT/n}^{(n,\mathcal{D})}(x) \notin \mathcal{D}).$$

Proof. From Lemma 4.5.1, we obtain $\left(Q_{T/n}^{\delta,\mathcal{D}}\right)^{n-1} Q_{T/n} f \to \left(Q_{T/n} I^{\mathcal{D}}\right)^{n-1} Q_{T/n} f$ as $\delta \to 0$. □

4.6 Numerical experiments

4.6.1 1-dimensional example

In this section, we provide computational tests for a widely used interest rate model of Hull and White, which has been introduced in Section 4.2.1.

Under the Hull-White model, the zero-coupon bond price of maturity S follows

$$\begin{aligned} D(t,S) &= D(t,S,r(t)) \\ &= \exp\Big(-\int_t^S f(0,s)ds + B(t,S)(f(0,t) - r(t)) - \frac{\sigma^2}{4\alpha}(1-e^{-2\alpha t})B(t,S)^2\Big) \end{aligned}$$

where $B(t,S) = (1-e^{-\alpha(S-t)})/\alpha$ and $t \mapsto f(0,t)$ is the initial forward rate curve.

Let now consider European options on zero-coupon bonds with maturity S, expiry date T and strike price L. For the computaion, let us change the numeraire from the bank account to the bond of maturity T. Under the so-called T-forward measure, we can obtain the expression

$$\begin{aligned} dX_t &= -\alpha X_t dt + \sigma dW_t^T, \quad (W_t^T : \text{Brownian motion}) \\ \theta_t &= f(0,t) + \frac{\sigma^2}{2\alpha^2}(1-e^{-\alpha t})^2 - \int_0^t \frac{\sigma^2}{\alpha}(1-e^{-\alpha(T-s)})e^{-\alpha(t-s)}ds \end{aligned}$$

(see e.g. [9]). Then the price of European put bond options follows

$$D(0,T)E^T[\max(0, L - D(T, S, \theta_T + X_T))|X_0 = 0]$$

where E^T stands for the expectation under T-forward measure. We now provide computational tests for the expectation by using an I-L scheme.

Let us take the parameters as $T = 1.0$, $S = 5.0$, $f(0,t) \equiv 0.02$, $\sigma = 0.005$, $\alpha = 1.0$, $L = e^{-0.02\times 4.0}$, and in the following, the boundary $[-K, K] = [-0.02, 0.02]$. We will describe the absolute error, that is, |approximate price − true price|, where the true price (= 0.1165 with notional amount = 100) is obtained by the analytical bond price formula. In the tests, we do not use here coordinate transform techniques (i.e. non-uniform grid sketching) which could give higher accuracy.

Figure 4.3 demonstrates the convergence results of the I-L scheme with linear interpolations. On the left, the absolute error is plotted for fixed n as the number of grid points m increases. On the right, the absolute error is plotted for fixed m in contrast to the left. For insufficient number of grid points m, we can see that the space discretization error grows like $O(n)$. (Note that linear interpolations for convex functions have positive bias.) The practical acceptable error, such as 10^{-4}, is achieved for about $(n, m) = (15, 400)$.

Figure 4.4 shows the results in the case of spline interpolations. As far as we can see the results, spline interpolations work within the interpolated lattice framework, better than linear interpolations.

In Table 4.2, we illustrate the approximate prices for various strike prices. The error seems to be larger around at-the-money since the non-differentiability of the payoff function affects the rate of convergence (Theorem 4.2.10). Even if, in the case of a deep out-of-the-money, the absolute error looks quite small.

4.6.2 3-dimensional example

We next consider pricing cross currency options as a 3-dimensional example. Cross currency options with two stochastic interest rates on some standard models fortunately have analytical formulas, so that we can evaluate an accuracy of sparse grid interpolated lattice (SGIL for short) schemes. We can apply the methods explained below to exotic cross currency derivatives such as cancelable PRDC swaps.

Let now consider a spot FX rate S and two stochastic interest rate dynamics (i.e. domestic and foreign) which are given by $r_d(t) = \theta_d(t) + X_d(t)$, $r_f(t) = \theta_f(t) + X_f(t)$,

$$\begin{aligned} dS(t)/S(t) &= (r_d(t) - r_f(t))dt + \sigma dW_t^1, \\ dX_d(t) &= -\alpha_d X_d(t)dt + \sigma_d dW_t^2, \\ dX_f(t) &= -\alpha_f X_f(t)dt + \sigma_f dW_t^3 \end{aligned}$$

under the risk-neutral measure (the spot-martingale measure). The correlation of the Brownian motion (W_t^1, W_t^2, W_t^3) is given by $d\langle W^i, W^j\rangle_t = \rho_{ij}dt$.

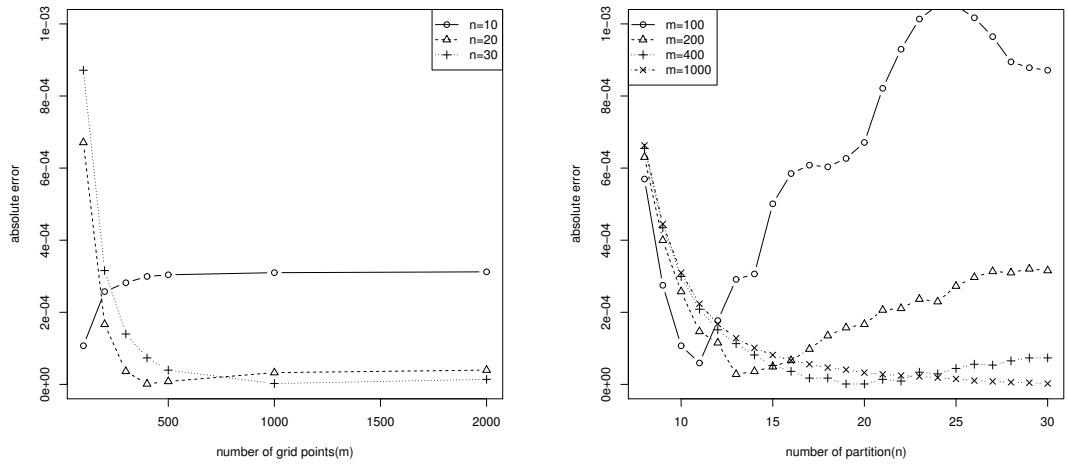

Figure 4.3: Convergence of I-L scheme (linear interpolation).

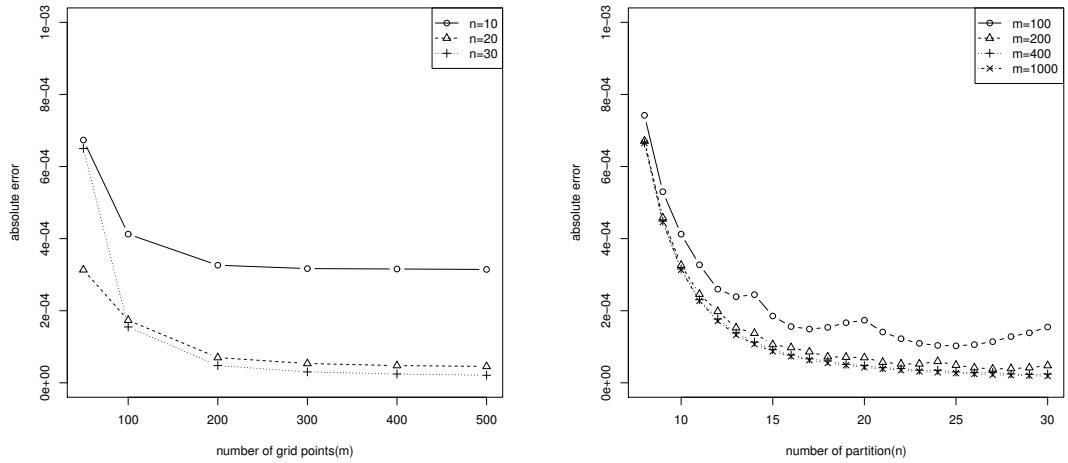

Figure 4.4: Convergence of I-L scheme (cubic spline interpolation).

price									
	linear				cubic spline				analytical
n	10		20		10		20		
m	200	400	200	400	200	400	200	400	
$L = 0.915$	0.00027	0.00026	0.00028	0.00027	0.00026	0.00026	0.00027	0.00027	0.00027
$L = 0.920$	0.02219	0.02216	0.02229	0.02218	0.02214	0.02214	0.02214	0.02215	0.02216
$L = 0.925$	0.23145	0.23139	0.23165	0.23149	0.23138	0.23137	0.23142	0.23143	0.23146
$L = 0.930$	0.67580	0.67580	0.67582	0.67581	0.67579	0.67579	0.67580	0.67580	0.67580
$L = 0.935$	1.16484	1.16484	1.16484	1.16484	1.16484	1.16484	1.16484	1.16484	1.16484

Table 4.2: Bond option price (notional amount = 100): convergence results for various strike prices.

We start approximating the values of European put option with maturity T and strike price L. The price is determined by the expression $D(0,T)E^T[\max(0, L-S_T)]$ under domestic T-forward measure. Through the change of measure, one obtains

$$E^T[\max(0, L - S_T)] = E^T[\max(0, L - S(0)\exp(V_T + X_T^1))] \tag{4.25}$$

where we have used $V_T = \int_0^T (\theta_d^T(t) - \theta_f^T(t) - \theta_S^T(t))dt - \frac{\sigma^2}{2}T$,

$$\begin{aligned}
\theta_d^T(t) &= f_d(0,t) + \frac{\sigma_d^2}{2\alpha_d^2}(1 - e^{-\alpha_d t})^2 - \int_0^t \frac{\sigma_d^2}{\alpha_d}(1 - e^{-\alpha_d(T-s)})e^{-\alpha_d(t-s)}ds, \\
\theta_f^T(t) &= f_f(0,t) + \frac{\sigma_f^2}{2\alpha_f^2}(1 - e^{-\alpha_f t})^2 - \int_0^t (\sigma\sigma_f\rho_{13} + \frac{\rho_{23}\sigma_d\sigma_f}{\alpha_d}(1 - e^{-\alpha_d(T-s)}))e^{-\alpha_f(t-s)}ds, \\
\theta_S^T(t) &= \rho_{12}\frac{\sigma\sigma_d}{\alpha_d}(1 - e^{-\alpha_d(T-t)}),
\end{aligned}$$

and X_t follows

$$\begin{aligned}
dX^1(t) &= (X^2(t) - X^3(t))dt + \sigma dW_t^{T,1}, \\
dX^2(t) &= -\alpha_d X^2(t)dt + \sigma_d dW_t^{T,2}, \\
dX^3(t) &= -\alpha_f X^3(t)dt + \sigma_f dW_t^{T,3}
\end{aligned}$$

where $(W_t^{T,j})$ stands for a 3-dimensional Brownian motion with the same correlation structure as (W_t^j). Noticing that the corresponding vector fields V_1, V_2, V_3 are commutative, we obtain a second order scheme by the simple expression

$$\exp(t/2V_0)\exp(\sqrt{t}(Z^1V_1 + Z^2V_2 + Z^3V_3))\exp(t/2V_0)x \tag{4.26}$$

with 14 successor nodes.

We take the parameters as follows: $T = 1.0$, $S_0 = 100$, $f_d(0,t) \equiv 0.02$, $f_f(0,t) \equiv 0.06$, $\sigma = 0.15$, $\sigma_d = 0.005$, $\sigma_f = 0.015$, $\alpha_d = \alpha_f = 1.0$, $\rho_{21} = \rho_{31} = \rho_{32} = 0.1$. We now evaluate an at-the-money(ATM) forward put option, that is, $L = S_0 e^{(0.02-0.06)T}$ in (4.25).

We set the boundary $[-K_1, K_1]\times[-K_2, K_2]\times[-K_3, K_3] = [-0.6, 0.6]\times[-0.02, 0.02]\times[-0.06, 0.06]$, and use a coordinate transform for each direction by

$$b^j(x) = \begin{cases}
(x + K_j)/8K_j & x \in [-K_j, -K_j/2) \\
1/16 + (x + K_j/2)/2K_j & x \in [-K_j/2, -K_j/8) \\
1/4 + 2(x + K_j/8)/K_j & x \in [-K_j/8, K_j/8) \\
3/4 + (x - K_j/8)/2K_j & x \in [K_j/8, K_j/2) \\
15/16 + (x - K_j/2)/8K_j & x \in [K_j/2, K_j].
\end{cases}$$

This transform has an effect of reducing the constant C_2 in Theorem 4.2.8 and 4.2.10.

In Table 4.3 we show the results when non-equidistant sparse grid interpolation techniques are used. Here we take the number $n = 20$, that is, $\Delta t = T/n = 0.05$.

m_s	price	absolute error	relative error	time [sec, 1 thread]	time [sec, 2 thread]
		$\gamma = (0,0,0)$			
5	5.65937	0.03786	0.674%	<0.1	<0.1
6	5.63067	0.00916	0.163%	0.2	0.1
7	5.62320	0.00169	0.030%	0.8	0.5
8	5.62130	0.00021	0.004%	2.7	1.7
9	5.62126	0.00025	0.004%	9.7	5.8
		$\gamma = (1,0,0)$			
4	5.65954	0.03803	0.677%	<0.1	<0.1
5	5.63072	0.00921	0.164%	0.1	<0.1
6	5.62318	0.00167	0.030%	0.5	0.3
7	5.62130	0.00021	0.004%	1.7	1.2
8	5.62126	0.00025	0.004%	6.1	3.7
		$\gamma = (2,0,0)$			
4	5.63154	0.01003	0.178%	0.1	<0.1
5	5.62333	0.00186	0.032%	0.3	0.2
6	5.62132	0.00019	0.003%	1.1	0.7
7	5.62126	0.00025	0.004%	4.0	2.4
		$\gamma = (3,0,0)$			
4	5.62405	0.00254	0.045%	0.2	0.1
5	5.62146	0.00005	0.001%	0.7	0.4
6	5.62129	0.00022	0.004%	2.5	1.5

Table 4.3: Results of SGIL for ATM forward currency put options, when we fix $n = 20$ ($\Delta t = 0.05$). The analytical price is 5.62151. Test on Intel(R) Core(TM) i5-661 @3.33GHz, 4GB RAM with C++ and OpenMP programming (Visual C++ compiler).

Comparing with the results which can be obtained for different γ, we confirm that the space discretization error is mainly dominated by currency dynamics. If 10^{-2} absolute error is admissible in practical views, the computation takes less than 1 second.

Comparison of sparse grids and plain Monte Carlo: The Monte Carlo method is a simple and very powerful tool for computing expectations. Using the equation (4.26) with standard normal random variables Z, we test the accuracy of Monte Carlo simulations from a viewpoint of statistical error.

num. of simulation	price	1 StDev	± 1 StDev	2 StDev / Price	time [sec, 1 thread]
100,000	5.60012	0.024	[5.576, 5.624]	0.850%	0.4
500,000	5.63425	0.011	[5.624, 5.645]	0.379%	1.8
1,000,000	5.63602	0.008	[5.628, 5.644]	0.268%	3.6
10,000,000	5.62296	0.002	[5.621, 5.625]	0.085%	35.0

Table 4.4: Results of Monte Carlo simulations with $n = 20$ in the equation (4.26).

Table 4.4 shows the results of plain Monte Carlo simulations. If the target error (in the sense of confidence interval) is about 0.01-0.05, there is no significant difference of sparse grids and plain Monte Carlo. When we require more precision, for e.g. calibration and computing sensitivities, the sparse grid method has the superiority, which will be more remarkable in computing Bermudan or American options.

4.7 Some remarks

Higher-order convergence of space-time discretization allows us to reduce dramatically computational time through few time steps and efficient grid sketching. Sparse grid interpolated lattice schemes are well-suited for several dimensional models, and are simple to be implemented as long as cubature formulas and sparse grid interpolations are provided. We do not have to be bothered by solving linear equations arising from discretized PDEs.

The notable features of (sparse grid) interpolated lattice schemes are summarized below.

- **Second order accuracy**: From the practical viewpoint, one may argue that there are advantages to converge with second order accuracy with respect to time discretization.

- **Breaking the curse of dimensionality**: By virtue of celebrated works on cubature formulas and sparse grid methods, we can compute the approximate values of expectations within the interpolated lattice framework for higher dimensions ($2 \leq N \leq 4$ or 5, probably depending on PC performance).

- **Stability**: There is no need to choose smaller and smaller Δt depending on our choice of δ. Probabilistic constructions of time discretization schemes automatically allow us to avoid instability due to the positiveness of probabilities. Conversely, a weak point is the fact that the space discretization error may grow linearly in proportion to the number of partition n. However, it does definitely not cause high oscillation, and the method has practically acceptable accuracy in the field of option pricing.

- **Flexibility**: Since the method is based on the weak approximations of SDEs, a broad class of models with unknown distributions is covered.

- **Parallelization**: The operations of sparse grid interpolated lattice schemes are compatible with coarse-grained parallel computing. In addition, once we obtain the sparse matrix $\mathbb{M}$, it can be used in computing a derivative portfolio with the same pricing model and parameters.

The reader might hope to find and use a scheme which has a higher accuracy to deal with, for example, long-term swaptions. A way to obtain such a scheme is

to develop higher order (more than third order) space-time discretization schemes. However, recall that higher order methods take in general more time per unit operation, therefore, this does not always speed up the computation in practice. Another way is to find a case-by-case methodology which reduces the constants appeared in the error estimates (cf. Theorem 4.2.8); see Chapter 5 in this thesis. Of course from a viewpoint of implementations, some developments in computer science, such as parallel computing, more effective algorithms for sparse grid tensor products, faster hash table storage, and sparse BLAS for multiplications of the matrix $\mathbb{M}$, help us to improve the computational efficiency.

Chapter 5

Strong approximation with asymptotic method

This chapter is based on the paper by Tanaka and Yamada [95] accepted for publication in *International Journal of Theoretical and Applied Finance*.

5.1 Introduction

In the present chapter, we study an asymptotic method that accelerates numerical schemes for perturbed random variables. The general concept is as follows. Suppose that F^ϵ is a random variable depending on a small parameter ϵ. Let us consider an approximation $\bar{F}^\epsilon$ for F^ϵ independently with respect to ϵ. Then the bias $F^\epsilon - \bar{F}^\epsilon$ may be close to the bias $F^0 - \bar{F}^0$, since ϵ has a small effect on the value of $F^\epsilon - \bar{F}^\epsilon$. Therefore, we expect that

$$\bar{F}^\epsilon - \bar{F}^0 + F^0 \text{ is a better approximation than } \bar{F}^\epsilon. \tag{5.1}$$

In particular, our interest is to study the above property when F^ϵ is a functional of a stochastic process and $\bar{F}^\epsilon$ comes from time discretization for it. In many cases, F^0 is a simpler model than F^ϵ and its exact distribution is well-known (e.g. Gaussian random variables or functionals of Gaussian processes). Even if the exact distribution of F^0 is unknown, it seems to be possible to provide a new scheme $\bar{F}^\epsilon - \bar{F}^0 + \tilde{F}^0$ with another more efficient scheme $\tilde{F}^0$ for F^0.

A further development of the above acceleration can be considered through the Taylor expansion of $\epsilon \mapsto F^\epsilon - \bar{F}^\epsilon$, that is,

$$F^\epsilon - \bar{F}^\epsilon \approx F^0 - \bar{F}^\epsilon + \frac{d}{d\epsilon}(F^\epsilon - \bar{F}^\epsilon)|_{\epsilon=0} + \cdots . \tag{5.2}$$

The idea (5.1) comes from the 0-th order term in (5.2). Therefore if it is possible to compute $\frac{d}{d\epsilon}(F^\epsilon - \bar{F}^\epsilon)|_{\epsilon=0}$ easily, we can expect to construct a higher-order scheme with respect to ϵ. This new idea has not been taken into account in this thesis, and should be studied with concrete examples in future work.

We now turn to consider the error estimates for the type of acceleration (5.1). Here, we are interested in the following three error structures:

- Strong error:

$$E[|F^\epsilon - (\bar{F}^\epsilon - \bar{F}^0 + F^0)|^p]^{1/p}. \tag{5.3}$$

- Weak error:

$$|E[F^\epsilon] - (E[\bar{F}^\epsilon] - E[\bar{F}^0] + E[F^0])|. \tag{5.4}$$

- Monte Carlo bias estimator for $\frac{1}{M}\sum_{j=1}^M (F^{\epsilon,j} - F^{0,j}) + E[F^0]$ where $(F^{\epsilon,j})_j$ be an i.i.d. sampling of F^ϵ:

$$\mathrm{Var}\Big(\frac{1}{M}\sum_{j=1}^M (F^{\epsilon,j} - F^{0,j})\Big). \tag{5.5}$$

Notice that in the case of Monte Carlo bias (5.5), the term $\frac{1}{M}\sum_{j=1}^M (F^{0,j}) - E[F^0]$ works as a *control variates* method. For applications in strong error (5.3), we need an exact or accurate numerical simulation method for F^0. On the other hand, in the cases of weak error (5.4) and Monte Carlo bias (5.5), we have to know the value of $E[F^0]$, and therefore we need a closed formula or an accurate numerical scheme for $E[F^0]$ such as the fast Fourier transform in one dimension.

When F^ϵ denotes a functional of a stochastic differential equation X_t^ϵ, $\bar{F}^\epsilon$ corresponds to a certain time discretization scheme $\bar{X}_t^{\epsilon,(n)}$ (n: number of partition). Takahashi-Yoshida [88] derived the following results in weak error sense (5.4) and Monte Carlo bias sense (5.5) for the Euler-Maruyama scheme $\bar{X}_t^{\epsilon,(n)}$:

$$E[f(X_T^\epsilon)] - (E[f(\bar{X}_T^{\epsilon,(n)})] - E[f(\bar{X}_T^{0,(n)})] + E[f(X_T^0)]) = O\Big(\frac{\epsilon}{n}\Big),$$
$$\mathrm{Var}^{1/2}\Big(\frac{1}{M}\sum_{j=1}^M (f(\bar{X}_T^{\epsilon,(n),j}) - f(\bar{X}_T^{0,(n),j}))\Big) = O\Big(\frac{\epsilon}{M^{1/2}}\Big),$$

and hence the total error (the root-mean-squared error; RMSE) is equal to

$$\mathrm{Var}^{1/2}\Big(E[f(X_T^\epsilon)] - \frac{1}{M}\sum_{j=1}^M (f(\bar{X}_T^{\epsilon,(n),j}) - f(\bar{X}_T^{0,(n),j})) - E[f(X_T^0)]\Big)$$
$$= O\Big(\frac{\epsilon}{n} + \frac{\epsilon}{M^{1/2}}\Big).$$

Here they assumed some appropriate conditions for f and the coefficients of X_t^ϵ. This is the case where $F^\epsilon = f(X_T^\epsilon)$ and $\bar{F}^\epsilon = f(\bar{X}_T^{\epsilon,(n)})$ in (5.4), and $F^\epsilon = f(\bar{X}_T^{\epsilon,(n)})$ in (5.5). In order to make the total error $O(\gamma)$ with weak and Monte Carlo bias, the

standard Euler-Maruyama scheme with i.i.d. sampling requires the computational cost $n \cdot M = O(\gamma^{-3})$, and in contrast, the accelerated Euler-Maruyama scheme with i.i.d. sampling requires the cost $O(\epsilon^3\gamma^{-3})$. That is, the asymptotic method (5.1) for the Euler-Maruyama scheme is $O(\epsilon^3)$-times faster than the standard method. Moreover, we can construct a sampling scheme whose computational cost turns out to be $O(\epsilon^{1-\delta}\gamma^{-2}(\log\gamma/\epsilon)^2)$ for any $\delta > 0$ and Lipschitz continuous function f via the multi-level Monte Carlo method. See Theorem 5.4.5.

In this chapter, we develop the error analysis for the Euler-Maruyama and Milstein schemes with the asymptotic method in strong sense (5.3). Under suitable conditions, we will show that for any $p \geq 2$,

$$E\Big[\sup_{0\leq t\leq T}|X_t^\epsilon - (\bar{X}_t^{\epsilon,(n)} - \bar{X}_t^{0,(n)} + X_t^0)|^p\Big]^{1/p} = O\Big(\frac{\epsilon}{n^\alpha}\Big) \tag{5.6}$$

with $\alpha = 1/2$ $(= 1)$ for the Euler-Maruyama (Milstein, resp.) scheme $\bar{X}_t^{\epsilon,(n)}$. Although strong convergence is usually very slow, the asymptotic method (5.1) helps to improve the speed of convergence.

A simplest example of (5.6) is for the case where the SDE becomes the ODE when $\epsilon = 0$, namely,

$$dX_t^\epsilon = b(X_t^\epsilon)dt + \epsilon\sigma(X_t^\epsilon)dB_t.$$

However, from the viewpoint of applications, we can also consider the (0th-order) ϵ-expansion around linear models like Black-Scholes (See an analytical expansion in Kunitomo-Takahashi [51] and Takahashi-Yamada [87]). Indeed, we can treat a perturbed stochastic differential equations such as

$$\begin{aligned} dX_t^\epsilon &= b_t^\epsilon X_t^\epsilon dt + \sqrt{\sigma_t^\epsilon} X_t^\epsilon dB_t, \\ db_t^\epsilon &= h_b(b_t^\epsilon)dt + \epsilon V_b(b_t^\epsilon)dB_t, \\ d\sigma_t^\epsilon &= h_\sigma(\sigma_t^\epsilon)dt + \epsilon V_\sigma(\sigma_t^\epsilon)dB_t. \end{aligned}$$

Notice that X_t^0 becomes the Black-Scholes model with time-dependent coefficients. Therefore there are many applications in the models of dynamic assets with stochastic volatility and/or stochastic interest rate. In particular, we will discuss more general stochastic differential equations so-called local-stochastic volatility type models.

This chapter is organized as follows: Section 5.2 is devoted to state theoretical results for strong convergence (5.6). In Section 5.3 we discuss pathwise simulation of stochastic volatility models. In Section 5.4, we introduce the multi-level Monte Carlo method and its acceleration by the asymptotic method. In Section 5.5 some numerical experiments for the SABR stochastic volatility model are given.

5.2 Strong convergence results

As seen in the previous intruduction, the asymptotic method (5.1) for discretizing stochastic processes is very natural to speed up the discretization procedure. We

state here the basic setting to discuss the approximation schemes. Let us consider a stochastic differential equation (SDE) of the form

$$dX_t^\epsilon = b(X_t^\epsilon, \epsilon)dt + \sigma(X_t^\epsilon, \epsilon)dB_t, \quad X_0^\epsilon = x_0, \tag{5.7}$$

where $b \in C(\mathbf{R}^N \times [0,1]; \mathbf{R}^N)$, $\sigma \in C(\mathbf{R}^N \times [0,1]; \mathbf{R}^N \times \mathbf{R}^d)$, and B_t is a d-dimensional standard Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying usual conditions. Throughout the present chapter, we use the equidistant partition $t_i = \frac{i}{n}T$, $0 \leq i \leq n$. The Euler-Maruyama and Milstein schemes will be considered with some smoothness conditions for the coefficients of the SDE.

5.2.1 The Euler-Maruyama scheme with asymptotic method

Let $\bar{X}_t^{\epsilon,(n)}$ be the Euler-Maruyama scheme for the SDE X_t^ϵ (Maruyama [64]): For $t \in [t_i, t_{i+1}]$,

$$\bar{X}_t^{\epsilon,(n)} := \bar{X}_{t_i}^{\epsilon,(n)} + b(\bar{X}_{t_i}^{\epsilon,(n)}, \epsilon)(t - t_i) + \sigma(\bar{X}_{t_i}^{\epsilon,(n)}, \epsilon)(B_t - B_{t_i}). \tag{5.8}$$

The implementation of (5.8) is very simple. Indeed, practitioners only need to know how to simulate normal random variables. The error of the scheme has been analyzed deeply by many researchers (see e.g. [90], [45], [6]). Roughly speaking, the strong order of convergence is equal to $1/2$, and the weak order is equal to 1.

We now prepare the assumptions for $\bar{X}$.

(H_1): $|b(x,\epsilon)| + |\sigma(x,\epsilon)| \leq C(1 + |x|)$.

(H_2): $|b(x,\epsilon) - b(y,\epsilon)| + |\sigma(x,\epsilon) - \sigma(y,\epsilon)| \leq C|x - y|$.

(H_3): $|b(x,\epsilon) - b(x,0)| + |\sigma(x,\epsilon) - \sigma(x,0)| \leq C\epsilon(1 + |x|)$.

(H_4): For every ϵ, $b(\cdot,\epsilon), \sigma(\cdot,\epsilon) \in C^1$ and $|\partial b(x,\epsilon) - \partial b(y,0)| + |\partial\sigma(x,\epsilon) - \partial\sigma(y,0)| \leq C(\epsilon + |x - y|)$.

The above constant C is independent of $(x, y, \epsilon) \in \mathbf{R}^N \times \mathbf{R}^N \times [0,1]$.

Let us define the accelerated Euler-Maruyama scheme as

$$\bar{Y}_t^{\epsilon,(n)} := \bar{X}_t^{\epsilon,(n)} - \bar{X}_t^{0,(n)} + X_t^0.$$

The property (5.1) for strong convergence is formulated rigorously as follows.

Theorem 5.2.1. *Suppose that (H_1)-(H_4) hold. Then for any $p \geq 2$, there exists a constant $C = C(T, x_0, p)$ such that*

$$E\Big[\sup_{0\leq t\leq T} |X_t^\epsilon - \bar{Y}_t^{\epsilon,(n)}|^p\Big]^{1/p} \leq C\frac{\epsilon}{n^{1/2}}.$$

In particular, if we consider the small volatility model $dX_t^\epsilon = b(X_t^\epsilon)dt + \epsilon dB_t$ and the ODE $dX_t^0 = b(X_t^0)dt$, then intuitively speaking, $(X_t^\epsilon - \bar{X}_t^\epsilon) - (\bar{X}_t^0 - \bar{X}_t^0)$ cancels out the error from the drift term (except the effect of ϵ), and the error from small volatility ϵ only remains. Hence the total error is proportional to ϵ.

Of course, more general situations can be considered, for example, if b and σ depends on time t, then some smoothness assumptions with respect to (t, ϵ) are needed in addition to (H_1)-(H_4). We will not attempt to prove this, but basically the asymptotic method works as well.

Remark 5.2.2. The rate of convergence of the Euler-Maruyama scheme basically relies on the smoothness (or the Lipschitz continuity) of coefficients of SDEs. If the coefficients are not smooth but Hölder continuous, the speed of convergence may be slow, as seen in the paper by Yan [97]. For obtaining the strong rate of convergence $O(n^{-1/2})$ with $\sigma(x) = x^\alpha$ $(1/2 \leq \alpha < 1)$, a modified Euler-type scheme (called a symmetrized Euler scheme) was developed by Berkaoui et al. ([8]). We should mention that the Euler-Maruyama scheme may not converge strongly when the coefficients are non-globally Lipschitz continuous. For example, in [36] a sufficient condition that the scheme explodes is given.

5.2.2 The Milstein scheme with asymptotic method

We next discuss the Milstein scheme which has a higher order rate of convergence than the Euler-Maruyama scheme in strong sense. Just for notational convenience, we only consider the case $d = 1$. Of course, in general dimensional setting with commutative vector fields $(\sigma^j)_{1 \leq j \leq d}$, we can use the (accelerated) Milstein scheme as well.

Throughout this section, we assume the following smoothness.

- For every ϵ, $\sigma(\cdot, \epsilon) \in C^2$.

The Milstein scheme $\hat{X}_t^{\epsilon,(n)}$ for the SDE X_t^ϵ is defined by

$$\begin{aligned}
\hat{X}_t^{\epsilon,(n)} &:= \hat{X}_{t_i}^{\epsilon,(n)} + b(\hat{X}_{t_i}^{\epsilon,(n)}, \epsilon)(t - t_i) + \sigma(\hat{X}_{t_i}^{\epsilon,(n)}, \epsilon)(B_t - B_{t_i}) \\
&\quad + \sigma\sigma'(\hat{X}_{t_i}^{\epsilon,(n)}, \epsilon) \int_{t_i}^t \int_{t_i}^s dB_r dB_s \\
&= \hat{X}_{t_i}^{\epsilon,(n)} + b(\hat{X}_{t_i}^{\epsilon,(n)}, \epsilon)(t - t_i) + \sigma(\hat{X}_{t_i}^{\epsilon,(n)}, \epsilon)(B_t - B_{t_i}) \\
&\quad + \frac{1}{2}\sigma\sigma'(\hat{X}_{t_i}^{\epsilon,(n)}, \epsilon)((B_t - B_{t_i})^2 - (t - t_i))
\end{aligned}$$

for $t \in [t_i, t_{i+1}]$.

We use the (stronger) assumptions for $\hat{X}$.

(H_1'): (H_1) & $|\sigma\sigma'(x, \epsilon)| + |b\sigma'(x, \epsilon)| + |\sigma^2\sigma''(x, \epsilon)| \leq C(1 + |x|)$.

(H_2'): (H_2) & $|\sigma\sigma'(x,\epsilon)-\sigma\sigma'(y,\epsilon)|+|b\sigma'(x,\epsilon)-b\sigma'(y,\epsilon)|+|\sigma^2\sigma''(x,\epsilon)-\sigma^2\sigma''(y,\epsilon)| \le C|x-y|$.

(H_3'): (H_3) & $|\sigma\sigma'(x,\epsilon)-\sigma\sigma'(x,0)|+|b\sigma'(x,\epsilon)-b\sigma'(x,0)|+|\sigma^2\sigma''(x,\epsilon)-\sigma^2\sigma''(x,0)| \le C\epsilon(1+|x|)$.

(H_4'): (H_4) & $|(\sigma\sigma')'(x,\epsilon)-(\sigma\sigma')'(y,0)| \le C(\epsilon+|x-y|)$.

Let us define the accelerated Milstein scheme as

$$\hat{Y}_t^{\epsilon,(n)} := \hat{X}_t^{\epsilon,(n)} - \hat{X}_t^{0,(n)} + X_t^0.$$

Then we can get the higher order convergence rate.

Theorem 5.2.3. *Suppose that (H_1')-(H_4') hold. Then for any $p \ge 2$, there exists a constant $C = C(T, x_0, p)$ such that*

$$E\Big[\sup_{0\le t\le T}|X_t^\epsilon - \hat{Y}_t^{\epsilon,(n)}|^p\Big]^{1/p} \le C\frac{\epsilon}{n}.$$

5.2.3 Proof of Theorem 5.2.1 and 5.2.3

We use the following notations.

- $\eta(s) := t_i$ if $s \in [t_i, t_{i+1})$.
- $\bar{X}_t^\epsilon \equiv \bar{X}_t^{\epsilon,(n)}$, $\hat{X}_t^\epsilon \equiv \hat{X}_t^{\epsilon,(n)}$.

We will apply the Burkholder-Davis-Gundy (BDG) inequality

$$c_p E[\langle M\rangle_T^{p/2}] \le E[\sup_{0\le t\le T}|M_t|^p] \le C_p E[\langle M\rangle_T^{p/2}]$$

to the proofs below: Here $p > 0$ and M_t is a continuous local martingale.

Using the BDG inequality and Gronwall inequality, we can show the following moment estimates. (See [45] for the proof in the case of L^2-norm.)

Lemma 5.2.4. *(i) Suppose that the assumptions (H_1)-(H_2) hold. Then for any $p \ge 2$, we have*

$$\sup_{\epsilon\in[0,1]} E[\sup_{0\le t\le T}|X_t^\epsilon|^p] + \sup_{\epsilon\in[0,1]} E[\sup_{0\le t\le T}|\bar{X}_t^\epsilon|^p] < \infty,$$

$$\sup_{\epsilon\in[0,1]} \max_{0\le i\le n-1} E[\sup_{t_i\le t\le t_{i+1}}|X_t^\epsilon - X_{t_i}^\epsilon|^p] \le C(T,x_0,p)/n^{1/2}.$$

$$\sup_{\epsilon\in[0,1]} E[\sup_{0\le t\le T}|X_t^\epsilon - \bar{X}_t^\epsilon|^p] \le C(T,x_0,p)/n^{1/2}.$$

(ii) Suppose that the assumptions (H_1')-(H_2') hold. Then for any $p \ge 2$, we have

$$\sup_{\epsilon\in[0,1]} E[\sup_{0\le t\le T}|\hat{X}_t^\epsilon|^p] < \infty,$$

$$\sup_{\epsilon\in[0,1]} E[\sup_{0\le t\le T}|X_t^\epsilon - \hat{X}_t^\epsilon|^p] \le C(T,x_0,p)/n.$$

We now give an important lemma for the proof of the main theorems.

Lemma 5.2.5. *(i) Under (H_1)-(H_3), we have for any $p \geq 2$,*

$$E\Big[\sup_{0\leq t\leq T}|X_t^\epsilon - X_t^0|^p\Big]^{1/p} \leq C(T,x_0,p)\epsilon, \tag{5.9}$$

and

$$\max_{0\leq i\leq n-1} E\Big[\sup_{t_i\leq t\leq t_{i+1}}|X_t^\epsilon - (X_{t_i}^\epsilon - X_{t_i}^0 + X_t^0)|^p\Big]^{1/p} \leq C(T,x_0,p)\frac{\epsilon}{n^{1/2}}. \tag{5.10}$$

(ii) Under (H_1)-(H_3), for any $p \geq 2$,

$$E\Big[\sup_{0\leq t\leq T}|\bar{X}_t^\epsilon - \bar{X}_t^0|^p\Big]^{1/p} \leq C(T,x_0,p)\epsilon.$$

(iii) Under (H_1')-(H_3'), for any $p \geq 2$,

$$E\Big[\sup_{0\leq t\leq T}|\hat{X}_t^\epsilon - \hat{X}_t^0|^p\Big]^{1/p} \leq C(T,x_0,p)\epsilon.$$

Proof. (i): We first note that

$$X_t^\epsilon - X_t^0 = \int_0^t (b(X_s^\epsilon,\epsilon) - b(X_s^0,0))ds + \int_0^t (\sigma(X_s^\epsilon,\epsilon) - \sigma(X_s^0,0))dBs,$$

and by the BDG inequality for the stochastic integral term,

$$\begin{aligned} E[\sup_{0\leq s\leq t}|X_s^\epsilon - X_s^0|^p] \leq C_p\Big(&E\Big[\Big(\int_0^t |b(X_s^\epsilon,\epsilon) - b(X_s^0,0)|ds\Big)^p\Big] \\ &+ E\Big[\Big(\int_0^t (\sigma(X_s^\epsilon,\epsilon) - \sigma(X_s^0,0))^2 ds\Big)^{p/2}\Big]\Big). \end{aligned}$$

Using the conditions (H_1)-(H_3) for the above, we have immediately

$$G(t) := E[\sup_{0\leq s\leq t}|X_s^\epsilon - X_s^0|^p] \leq C_1\epsilon^p + C_2\int_0^t G(s)ds.$$

Here the constants C_1 and C_2 do not depend on ϵ. Thus from the Gronwall inequality we obtain (5.9).

We next consider the second result (5.10). Since

$$X_t^\epsilon - (X_{\eta(t)}^\epsilon - X_{\eta(t)}^0 + X_t^0) = \int_{\eta(t)}^t (b(X_s^\epsilon,\epsilon) - b(X_s^0,0))ds + \int_{\eta(t)}^t (\sigma(X_s^\epsilon,\epsilon) - \sigma(X_s^0,0))dB_s,$$

the inequality (5.10) follows from (H_2)-(H_3) and (5.9).

The proofs for (ii) and (iii) are straightforward as in (5.9). □

The following lemma will be used such as the Lipschitz continuous property.

Lemma 5.2.6. *(i) Assume that* (H_1)*-*(H_4) *hold. Then*

$$\begin{aligned}&|b(x_1,\epsilon)-b(y_1,\epsilon)+b(y_2,0)-b(x_2,0)|\\&\le C((\epsilon+|x_1-x_2|+|y_1-y_2|)(x_1-y_1)+|x_1-y_1+y_2-x_2|).\\&|\sigma(x_1,\epsilon)-\sigma(y_1,\epsilon)+\sigma(y_2,0)-\sigma(x_2,0)|\\&\le C((\epsilon+|x_1-x_2|+|y_1-y_2|)(x_1-y_1)+|x_1-y_1+y_2-x_2|).\end{aligned}$$

(ii) Assume that (H_1')*-*(H_4') *hold. Then*

$$\begin{aligned}&|\sigma\sigma'(x_1,\epsilon)-\sigma\sigma'(y_1,\epsilon)+\sigma\sigma'(y_2,0)-\sigma\sigma'(x_2,0)|\\&\le C((\epsilon+|x_1-x_2|+|y_1-y_2|)(x_1-y_1)+|x_1-y_1+y_2-x_2|).\end{aligned}$$

Proof. We only prove for b. By the mean value theorem,

$$b(x_1,\epsilon)-b(y_1,\epsilon)+b(y_2,0)-b(x_2,0)=\xi^{\epsilon}_{x_1,y_1}(x_1-y_1)+\xi^{0}_{y_2,x_2}(y_2-x_2)$$

where $\xi^{\epsilon}_{x,y}:=\int_0^1\partial b(\rho x+(1-\rho)y,\epsilon)d\rho=\xi^{\epsilon}_{y,x}$. Taking the difference again in the right hand side, we have

$$\xi^{\epsilon}_{x_1,y_1}(x_1-y_1)=(\xi^{\epsilon}_{x_1,y_1}-\xi^{0}_{x_2,y_2})(x_1-y_1)+\xi^{0}_{y_2,x_2}(x_1-y_1).$$

Finally, using the assumption (H_4) for $(\xi^{\epsilon}_{x_1,y_1}-\xi^{0}_{x_2,y_2})$, we obtain the result. □

Now we shall prove the theorems.

Proof of Theorem 5.2.1. Let us define

$$G_1(t):=E[\sup_{0\le s\le t}|X^{\epsilon}_s-\bar{Y}^{\epsilon,(n)}_s|^p].$$

By using the Gronwall inequality, our goal becomes to show the following:

$$G_1(t)\le C_1\frac{\epsilon^p}{n^{p/2}}+C_2\int_0^t G_1(s)ds$$

where C_1 and C_2 depend only on T,x_0,p.

We now compute

$$X^{\epsilon}_t-\bar{Y}^{\epsilon,(n)}_t=e^{\epsilon}(t)+\bar{e}^{\epsilon}(t)$$

where

$$\begin{aligned}e^{\epsilon}(t)=&\int_0^t(b(X^{\epsilon}_{\eta(s)},\epsilon)-b(\bar{X}^{\epsilon}_{\eta(s)},\epsilon)+b(\bar{X}^{0}_{\eta(s)},0)-b(X^{0}_{\eta(s)},0))ds\\&+\int_0^t(\sigma(X^{\epsilon}_{\eta(s)},\epsilon)-\sigma(\bar{X}^{\epsilon}_{\eta(s)},\epsilon)+\sigma(\bar{X}^{0}_{\eta(s)},0)-\sigma(X^{0}_{\eta(s)},0))dB_s\end{aligned}$$

and

$$\bar{e}^{\epsilon}(t) = \int_0^t (b(X_s^{\epsilon},\epsilon) - b(X_{\eta(s)}^{\epsilon},\epsilon) + b(X_{\eta(s)}^0,0) - b(X_s^0,0))ds \\ + \int_0^t (\sigma(X_s^{\epsilon},\epsilon) - \sigma(X_{\eta(s)}^{\epsilon},\epsilon) + \sigma(X_{\eta(s)}^0,0) - \sigma(X_s^0,0))dB_s.$$

For $\bar{e}^{\epsilon}(t)$, we obtain from Lemma 5.2.6,

$$\begin{aligned} &|b(X_s^{\epsilon},\epsilon) - b(X_{\eta(s)}^{\epsilon},\epsilon) + b(X_{\eta(s)}^0,0) - b(X_s^0,0)| \\ &\leq\ C((\epsilon + |X_s^{\epsilon} - X_s^0| + |X_{\eta(s)}^{\epsilon} - X_{\eta(s)}^0|)|X_s^{\epsilon} - X_{\eta(s)}^{\epsilon}| \\ &\quad + |X_s^{\epsilon} - (X_{\eta(s)}^{\epsilon} - X_{\eta(s)}^0 + X_s^0)|), \end{aligned}$$

and

$$\begin{aligned} &|\sigma(X_s^{\epsilon},\epsilon) - \sigma(X_{\eta(s)}^{\epsilon},\epsilon) + \sigma(X_{\eta(s)}^0,0) - \sigma(X_s^0,0)| \\ &\leq\ C((\epsilon + |X_s^{\epsilon} - X_s^0| + |X_{\eta(s)}^{\epsilon} - X_{\eta(s)}^0|)|X_s^{\epsilon} - X_{\eta(s)}^{\epsilon}| \\ &\quad + |X_s^{\epsilon} - (X_{\eta(s)}^{\epsilon} - X_{\eta(s)}^0 + X_s^0)|). \end{aligned}$$

Hence the integral term in $\bar{e}^{\epsilon}(t)$ is evaluated by

$$\begin{aligned} &E[\sup_{0\leq r\leq t} |\int_0^r (b(X_s^{\epsilon},\epsilon) - b(X_{\eta(s)}^{\epsilon},\epsilon) + b(X_{\eta(s)}^0,0) - b(X_s^0,0))ds|^p] \\ &\leq\ C_3\Big((\epsilon^p + 2\sup_{0\leq s\leq t}\|X_s^{\epsilon} - X_s^0\|_{2p}^p)\sup_{0\leq s\leq t}\|X_s^{\epsilon} - X_{\eta(s)}^{\epsilon}\|_{2p}^p \\ &\quad + \sup_{0\leq s\leq t}\|X_s^{\epsilon} - (X_{\eta(s)}^{\epsilon} - X_{\eta(s)}^0 + X_s^0)\|_p^p\Big) \end{aligned}$$

By using the BDG inequality, the stochastic integral term in $\bar{e}^{\epsilon}(t)$ also has the same bound (except the size of constant C_3). Consequently, we have by Lemma 5.2.4, 5.2.5,

$$E[\sup_{0\leq s\leq t} |\bar{e}^{\epsilon}(s)|^p] \leq C_4\frac{\epsilon^p}{n^{p/2}}.$$

Applying a similar calculus to $e^{\epsilon}(t)$, we also get

$$\begin{aligned} E[\sup_{0\leq s\leq t} |e^{\epsilon}(s)|^p] &\leq C_5\frac{\epsilon^p}{n^{p/2}} + C_6\int_0^t E[|X_{\eta(s)}^{\epsilon} - (\bar{X}_{\eta(s)}^{\epsilon} - \bar{X}_{\eta(s)}^0 + X_{\eta(s)}^0)|^p]ds \\ &\leq C_5\frac{\epsilon^p}{n^{p/2}} + C_6\int_0^t G_1(s)ds. \end{aligned}$$

This finishes the proof of Theorem 5.2.1. □

Proof of Theorem 5.2.3. Similarly to the proof of Theorem 5.2.1, we shall show that for $G_2(t) := E[\sup_{0\le s\le t}|X_s^\epsilon - \hat{Y}_t^{\epsilon,(n)}|^p]$,

$$G_2(t) \le C_1\frac{\epsilon^p}{n^p} + C_2\int_0^t G_2(s)ds.$$

Now we consider the decomposition

$$X_t^\epsilon - \hat{Y}_t^{\epsilon,(n)} = \tilde{e}^\epsilon(t) + \sum_{i=1}^3 \hat{e}_i^\epsilon(t),$$

where

$$\begin{aligned}
\tilde{e}^\epsilon(t) = &\int_0^t (b(X_{\eta(s)}^\epsilon,\epsilon) - b(\hat{X}_{\eta(s)}^\epsilon,\epsilon) + b(\hat{X}_{\eta(s)}^0,0) - b(X_{\eta(s)}^0,0))ds \\
&+ \int_0^t (\sigma(X_{\eta(s)}^\epsilon,\epsilon) - \sigma(\hat{X}_{\eta(s)}^\epsilon,\epsilon) + \sigma(\hat{X}_{\eta(s)}^0,0) - \sigma(X_{\eta(s)}^0,0))dB_s \\
&+ \int_0^t\int_{\eta(s)}^s (\sigma\sigma'(X_{\eta(r)}^\epsilon,\epsilon) - \sigma\sigma'(\hat{X}_{\eta(r)}^\epsilon,\epsilon) \\
&\quad + \sigma\sigma'(\hat{X}_{\eta(r)}^0,0) - \sigma\sigma'(X_{\eta(r)}^0,0))dB_r dB_s,
\end{aligned}$$

and

$$\begin{aligned}
\hat{e}_1^\epsilon(t) &= \int_0^t\int_{\eta(s)}^s (b\sigma'(X_r^\epsilon,\epsilon) - b\sigma'(X_r^0,0))drdB_s, \\
\hat{e}_2^\epsilon(t) &= \int_0^t\int_{\eta(s)}^s (\frac{1}{2}\sigma^2\sigma''(X_r^\epsilon,\epsilon) - \frac{1}{2}\sigma^2\sigma''(X_r^0,0))drdB_s, \\
\hat{e}_3^\epsilon(t) &= \int_0^t\int_{\eta(s)}^s (\sigma\sigma'(X_r^\epsilon,\epsilon) - \sigma\sigma'(X_{\eta(r)}^\epsilon,\epsilon) + \sigma\sigma'(X_{\eta(r)}^0,0) - \sigma\sigma'(X_r^0,0))dB_r dB_s.
\end{aligned}$$

By a similar manner as in the proof of Theorem 5.2.1, we can also obtain

$$E[\sup_{0\le s\le t}(|\tilde{e}^\epsilon(s)|^p + \sum_{i=1}^3 |\hat{e}_i^\epsilon(s)|^p)] \le C_1\frac{\epsilon^p}{n^p} + C_2\int_0^t G_2(s)ds.$$

Indeed, compared with Theorem 5.2.1, the reason why we can get the rate n^{-p} above is due to the additional integrals $\int_{\eta(s)}^s \cdot\, dr$ or $\int_{\eta(s)}^s \cdot\, dB_r$ inside the error terms. □

5.3 Application to pathwise simulation of stochastic volatility models

Our goal in this section is to construct a faster pathwise approximation for perturbed stochastic differential equations which appear in financial modeling of volatility.

5.3.1 An accelerated scheme for SABR model

In financial modeling, the SABR model plays a role to fit the implied volatility especially in short time. The model is given by the SDE (Hagan et al. [33])

$$\begin{aligned} dS_t &= \sqrt{\alpha_t} S_t^\beta dB_t^1 \\ d\alpha_t &= \nu \alpha_t (\rho dB_t^1 + \sqrt{1-\rho^2} dB_t^2). \end{aligned}$$

The volatility is not a mean-reversion process, hence this model does not suit for pricing long-dated options. If $\beta < 1$, as far as the authors know, there is no exact pathwise simulation method for the above SDE. In weak sense, several accurate simulation methods via Bessel processes are known.

To avoid that the volatility process α_t becomes negative in approximation procedures, we use a logarithmic transform for α_t.

$$\begin{aligned} dS_t &= \sqrt{\alpha_0 \exp(\tilde{\alpha}_t)} S_t^\beta dB_t^1 \\ d\tilde{\alpha}_t &= -\frac{\nu^2}{2} dt + \nu (\rho dB_t^1 + \sqrt{1-\rho^2} dB_t^2). \end{aligned}$$

Consider $\epsilon = \nu$. Since we do not know exact pathwise simulation methods for S_t^0, we substitute the Milstein scheme $\hat{S}_t^0$ for S_t^0. Therefore, we can use an $O(\frac{\epsilon}{\sqrt{n}} + \frac{1}{n})$-scheme defined by

$$\tilde{Y}_t^\epsilon := \bar{S}_t^\epsilon - \bar{S}_t^0 + \hat{S}_t^0.$$

When ν is small enough, a typical sample path is like Figure 5.1. Here we use $n = 16$ for the standard Euler-Maruyama scheme (Standard E-M) and the above accelerated scheme (Accelerated).

We next turn to consider another formal approximation scheme. Formally, $x^\beta \approx x$ when $\beta \approx 1$ and especially $x \approx 1$. Thus consider the scaling $L_t := S_t/S_0$, $\beta = \beta(\epsilon)$.

$$\begin{aligned} dL_t^\epsilon &= \sqrt{\alpha_0 \exp(\tilde{\alpha}_t)} S_0^{\beta-1} (L_t^\epsilon)^\beta dB_t^1, \\ dL_t^0 &= \sqrt{\alpha_0} S_0^{\beta-1} L_t^0 dB_t^1. \end{aligned}$$

Here $S_0^{\beta-1}$ is just a constant coming from the scaling, thus we do not change the constant $S_0^{\beta-1}$ even when $\epsilon = 0$. The accelerated scheme that we want to use is

$$\check{Y}_t^\epsilon := \bar{S}_t^\epsilon - S_0(\bar{L}_t^0 - L_t^0).$$

Since L_t^0 is a log-normal process, it is useful to compute the path $t \mapsto X_t^0$ and $E[f(X_T^0)]$. We will check the efficiency of $\check{Y}$ through a numerical test later.

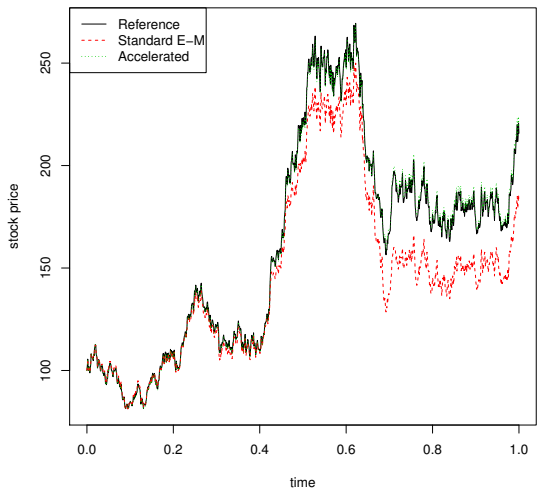

Figure 5.1: A sample path of discretized SABR model when ν is small.

5.3.2 General stochastic volatility models

The following model is an extension of local-stochastic volatility models applicable to both short and long term contingent claims in financial markets.

$$dS_t = \mu S_t dt + \sqrt{\alpha_t} S_t^{\beta} dB_t^1 + S_{t-} dJ_t$$
$$d\alpha_t = \lambda(\theta - \alpha_t)dt + \nu \alpha_t^{\gamma}(\rho dB_t^1 + \sqrt{1-\rho^2} dB_t^2)$$

where J_t is a compound Poisson process, which is often used to adapt especially short-dated large volatility smile/skew.

We remark that it is difficult to fit short-dated volatility smile/skew under the Heston model ($\beta = 1$, $\gamma = 1/2$, $J_t \equiv 0$), and then ν can take very large value. On the other hand, under general models with β and J_t, the parameter ν need not to be so large.

Let $\{\tau_j\}$ be the random jump times associated to J_t and consider a new time partition $\{\tilde{t}_k\} := \{t_i\} \cup \{\tau_j\}$. On the time interval $[\tilde{t}_k, \tilde{t}_{k+1})$ we can regard the approximation problem for S_t as the one for a continuous SDE. In particular by taking $\epsilon = \nu = 0$, the model becomes the CEV model with time-dependent coefficients. For a technical reason, we should consider some carefull treatments around zero of the function $(\cdot)^{\gamma}$ (See [11, 8, 61]).

5.4 Application to multi-level Monte Carlo method

The theoretical results we obtained in previous can be applied to the multi-level Monte Carlo method (MLMC in short). We propose an accelerated Monte Carlo

sampling for Takahashi-Yoshida's weak convergence method.

5.4.1 A brief review of MLMC with Euler-Maruyama scheme

We forget the parameter ϵ for the time being, and denote by X_t the continuous SDE X_t^ϵ defined by (5.7). Let us define $P := f(X_T)$ and $\bar{P}_l := f(\bar{X}_T^{(n_l)})$, and consider the time-step size $T/n_l = T/k^l$ for a fixed $k \in \mathbf{N}$. Let $L \in \mathbf{N}$ and the sampling of multi-level Monte Carlo is defined by

$$Y = \sum_{l=0}^{L} Y_l$$

where each Y_l is independently distributed and is given by

$$Y_l = \frac{1}{N_l} \sum_{j=1}^{N_l} \begin{cases} \bar{P}_0^{(j)}, & \text{if } l = 0, \\ (\bar{P}_l - \bar{P}_{l-1})^{(j)}, & \text{if } l \geq 1, \end{cases}$$

with i.i.d. sampling $\bar{P}_0^{(j)}$ or $(\bar{P}_l - \bar{P}_{l-1})^{(j)}$, $j = 1, \ldots, N_l$. The most important point is to use the same Brownian motion path $(B_t)_{t \geq 0}$ for simulating $\bar{P}_l$ and $\bar{P}_{l-1}$, and so the concept of the multi-level Monte Carlo method concerns the strong (pathwise) convergence rate.

Clearly we show that

$$E[P] - E[Y] = E[P] - E[\bar{P}_L],$$

therefore the weak rate of convergence depends only on the last number L. Moreover, we obtain from the independence of (Y_l),

$$\mathrm{Var}(Y) = \sum_{l=0}^{L} \frac{1}{N_l} \mathrm{Var}(Y_l).$$

and by definition $\mathrm{Var}(Y_l) = \mathrm{Var}(\bar{P}_l - \bar{P}_{l-1})$. Suppose some suitable conditions for f and X_t. Then one can obtain

$$\begin{aligned} E[\bar{P}_l - P] &= O(1/n_l), \\ \mathrm{Var}^{1/2}(\bar{P}_l - \bar{P}_{l-1}) &\leq \|\bar{P}_l - P\|_2 + \|\bar{P}_{l-1} - P\|_2 \\ &= O(1/n_l^{1/2}). \end{aligned}$$

The last estimate is the strong convergence result in L^2.

The total computational cost C is determined by the level L, the number of sampling $(N_l)_{l=0}^L$, and the number of partition $n_l (= k^l)$ so that

$$C = \sum_{l=0}^{L} N_l n_l.$$

Suppose that the required RMSE is $O(\gamma)$. Then by choosing $N_l = O(\gamma^{-2} L n_l^{-1})$, the total variance $\mathrm{Var}(Y)$ is of $O(\gamma^2)$. Now if we set $L = \log(\gamma^{-1})/\log(k) + O(1)$, then the total time discretization error $E[\bar{P}_L - P] = O(\gamma)$. Consequently $C = O(\gamma^{-2}(\log\gamma)^2)$ for the required accuracy $O(\gamma)$.

5.4.2 Accelerated MLMC sampling (with smooth payoffs)

From now on, we reconsider the sampling of the accelerated Euler-Maruyama scheme introduced by Takahashi and Yoshida from the standard Monte Carlo method

$$\frac{1}{M}\sum_{j=1}^{M}(f(\bar{X}_T^{\epsilon,(n),j}) - f(\bar{X}_T^{0,(n),j})) + E[f(X_T^0)]$$

to the multi-level Monte Carlo method via

$$\bar{P}_l^{\mathrm{new}} := f(\bar{X}_T^{\epsilon,(n_l)}) - f(\bar{X}_T^{0,(n_l)}) + E[f(X_T^0)].$$

Remark 5.4.1. We can also consider another MLMC sampling method via

$$\bar{P}_l^{\mathrm{another}} := f(\bar{X}_T^{\epsilon,(n_l)} - \bar{X}_T^{0,(n_l)} + X_T^0),$$

whose computational cost is $O(\epsilon^2\gamma^{-2}(\log\gamma/\epsilon)^2)$ for Lipschitz functions f. However, in this case we cannot take advantage of the explicit formula for the term $E[f(X_T^0)]$.

Giles [27] assumed that f is Lipschitz continuous to analyze the variance of estimator. On the other hand, we need $f \in C_b^2 := \{g \in C^2(\mathbf{R}^N;\mathbf{R}) \mid \partial_i g \text{ and } \partial_{ij} g \text{ are bounded}, 1 \le i,j \le N\}$ in order to use the asymptotics with respect to ϵ (Notice that $|f(x_1) - f(y_1) + f(y_2) - f(x_2)| \not\le C|x_1 - y_1 + y_2 - x_2|$ in general). Our analysis follows from the lemma below.

Lemma 5.4.2. *For $f \in C_b^2$,*

$$\begin{aligned}|f(x_1) - f(y_1) + f(y_2) - f(x_2)| \le& \frac{\|\nabla^2 f\|_\infty}{2}(|x_1 - x_2| + |y_1 - y_2|)|x_1 - y_1| \\ &+ \|\nabla f\|_\infty |x_1 - y_1 + y_2 - x_2|.\end{aligned}$$

Proof. This can be proved immediately by using the mean value theorem twice (See also Lemma 5.2.6). □

Then we have the following variance estimate.

Proposition 5.4.3. *Assume that (H_1)-(H_4) hold. For $f \in C_b^2$, we have*

$$\mathrm{Var}^{1/2}(\bar{P}_l^{\mathrm{new}} - \bar{P}_{l-1}^{\mathrm{new}}) \le C\epsilon n_l^{-1/2}.$$

Proof. The variance of the difference $\bar{P}_l^{\text{new}} - \bar{P}_{l-1}^{\text{new}}$ is estimated as

$$\begin{aligned}\text{Var}^{1/2}(\bar{P}_l^{\text{new}} - \bar{P}_{l-1}^{\text{new}}) &\le \|f(\bar{X}_T^{\epsilon,(n_l)}) - f(\bar{X}_T^{0,(n_l)}) - (f(\bar{X}_T^{\epsilon,(n_{l-1})}) - f(\bar{X}_T^{0,(n_{l-1})}))\|_2 \\ &\le \|f(X_T^\epsilon) - f(\bar{X}_T^{\epsilon,(n_l)}) + f(\bar{X}_T^{0,(n_l)}) - f(X_T^0)\|_2 \\ &\quad + \|f(X_T^\epsilon) - f(\bar{X}_T^{\epsilon,(n_{l-1})}) + f(\bar{X}_T^{0,(n_{l-1})}) - f(X_T^0)\|_2.\end{aligned}$$

Thus by Lemma 5.4.2,

$$\begin{aligned}&\|f(X_T^\epsilon) - f(\bar{X}_T^{\epsilon,(n_l)}) + f(\bar{X}_T^{0,(n_l)}) - f(X_T^0)\|_2 \\ &\le C((\|X_T^\epsilon - X_T^0\|_4 + \|\bar{X}_T^{\epsilon,(n_l)} - \bar{X}_T^{0,(n_l)}\|_4)\|X_T^\epsilon - \bar{X}_T^{\epsilon,(n_l)}\|_4 \\ &\quad + \|X_T^\epsilon - \bar{X}_T^{\epsilon,(n_l)} + \bar{X}_T^{0,(n_l)} - X_T^0\|_2).\end{aligned}$$

Hence using Lemma 5.2.4-5.2.5 and Theorem 5.2.1, we get the rate of convergence. □

For the use of the multi-level Monte Carlo method, we have obtained the results as follows.

$$\begin{aligned}E[\bar{P}_l^{\text{new}} - P] &= O(\epsilon/n_l), \\ \text{Var}^{1/2}(\bar{P}_l^{\text{new}} - \bar{P}_{l-1}^{\text{new}}) &= O(\epsilon/n_l^{1/2}).\end{aligned}$$

So the estimator for $(\bar{P}_l^{\text{new}})_{l\ge 0}$ has an equivalent effect to the one for $(\bar{P}_l)_{l\ge 0}$ with the required error $O(\gamma/\epsilon)$. Consequently we get the order of computational cost $O(\epsilon^2\gamma^{-2}(\log(\gamma/\epsilon))^{-2})$. Both the asymptotic method and multi-level Monte Carlo method are very easily computable, so that practitioners will get large benefit only with small additional implementation cost.

Remark 5.4.4. Clearly, we can also check the variance estimate for the accelerated Milstein scheme. Let $\hat{P}_l^{\text{new}} := f(\hat{X}_T^{\epsilon,(n_l)}) - f(\hat{X}_T^{0,(n_l)}) + E[f(X_T^0)]$. By a similar argument, we derive that under (H_1')-(H_4') and $f \in C_b^2$,

$$\text{Var}^{1/2}(\hat{P}_l^{\text{new}} - \hat{P}_{l-1}^{\text{new}}) \le C\epsilon n_l^{-1}. \tag{5.11}$$

We have not obtained weak convergence results for the accelerated Milstein scheme yet. However, we guess that from the basic proof of Takahashi-Yoshida [88], it holds that

$$E[f(X_T^\epsilon)] - (E[f(\hat{X}_T^{\epsilon,(n)})] - E[f(\hat{X}_T^{0,(n)})] + E[f(X_T^0)]) = O\Big(\frac{\epsilon}{n}\Big) \tag{5.12}$$

under some smoothness conditions for f and the coefficients of X_t^ϵ. Thus combining the results (5.11), (5.12) and the discussion in Giles [27, 26], we finally conclude that the total computational cost is $O(\epsilon^2\gamma^{-2})$.

5.4.3 Lipschitz payoffs

Let us consider the first component $(X_T^\epsilon)^{(1)}$ as an asset dynamics. Our interest is pricing an option $f((X_T^\epsilon)^{(1)})$ with Lipschitz payoffs $f : \mathbf{R} \to \mathbf{R}$. Set $\bar{P}_l^{\text{new}} = f((\bar{X}_T^{\epsilon,(n_l)})^{(1)}) - f((\bar{X}_T^{0,(n_l)})^{(1)}) + E[f((X_T^0)^{(1)})]$. Then we can obtain an upper bound estimate as follows.

Theorem 5.4.5. *Assume* (H_1)*-*(H_4) *and* $f : \mathbf{R} \to \mathbf{R}$ *is a Lipschitz continuous function whose weak derivative has bounded variation in* $\mathbf{R}$*. In addition, suppose* $(X_T^0)^{(1)}$ *has a bounded density, and* $(\bar{X}_T^0)^{(1)}$ *also has a bounded density uniformly with respect to* n*. Then we have for any small* $\delta > 0$*,*

$$\text{Var}^{1/2}(\bar{P}_l^{\text{new}} - \bar{P}_{l-1}^{\text{new}}) \leq C\epsilon^{(1-\delta)/2} n_l^{-1/2}.$$

Proof. See 5.4.5. □

This theorem implies that the required computational cost turns out to be $O(\epsilon^{1-\delta}\gamma^{-2}(\log\gamma/\epsilon)^2)$, with $L = \log(\epsilon\gamma^{-1})/\log(k) + O(1)$ and $N_l = O(\epsilon^{1-\delta}\gamma^{-2}Ln_l^{-1})$.

We now summarize strong rate of convergence for $\bar{P}_l - \bar{P}_{l-1}$ and $\bar{P}_l^{\text{new}} - \bar{P}_{l-1}^{\text{new}}$ in Table 5.1.

Payoff	Standard E-M	Accelerated E-M
C_b^2	$O(n_l^{-1/2})$	$O(\epsilon n_l^{-1/2})$
Lipschitz	$O(n_l^{-1/2})$	$O(\epsilon^{(1-\delta)/2} n_l^{-1/2})$
Digital	$O(n_l^{-(1-\delta)/4})$, ([3], [28])	-

Table 5.1: Strong rate of convergence of $\bar{P}_l - \bar{P}_{l-1}$ and $\bar{P}_l^{\text{new}} - \bar{P}_{l-1}^{\text{new}}$.

5.4.4 Localization for irregular payoffs

The regularity of f seems to be essential for the accelerated MLMC method introduced in previous. For example, we will see through computational experiments that the acceleration with discontinuous functions f does not work so well.

We now propose a localization technique for this problem. Let us define a decomposition

$$f = f_s + f_{ir}$$

where f_s is a smooth (at least Lipschitz continuous) function with $f \approx f_s$. Then we apply the accelerated MLMC to the smooth part f_s and the standard MLMC to the irregular part f_{ir}. In other words, we consider the MLMC method for

$$\bar{P}_l^{\text{new}}(f_s) := f(\bar{X}_T^\epsilon) - f_s(\bar{X}_T^0) + E[f_s(X_T^0)].$$

The standard MLMC for discontinuous functions was studied in Avikainen [3] and Giles et al. [28].

5.4.5 Proof of Theorem 5.4.5

Throughout this section, we use the following notations without confusion:

- $f(X_T^\epsilon) \equiv f((X_T^\epsilon)^{(1)})$, $f(\bar{X}_T^\epsilon) \equiv f((\bar{X}_T^\epsilon)^{(1)})$.
- $\|f\|_{\mathrm{Lip}} := \inf\{K \geq 0 : |f(x) - f(y)| \leq K|x-y|, \text{ for all } x, y \in \mathbf{R}\}$.
- $\|f\|_{\mathrm{TV}} := \sup_{-\infty<x_0<\cdots<x_m<\infty} \sum_{j=1}^m |f(x_j) - f(x_{j-1})|$.

We say f has bounded variation in $\mathbf{R}$ if $\|f\|_{\mathrm{TV}} < \infty$.

The following lemma plays a crucial role in the proof of the theorem.

Lemma 5.4.6 (Avikainen [3], Theorem 2.4)**.** *Let X and $\hat{X}$ be real valued random variables with $X, \hat{X} \in L^p$ $(p \geq 1)$. In addition, suppose X has a bounded density. Then for any function f of bounded variation in $\mathbf{R}$ and $q \geq 1$, there exists a constant $C > 0$ depending on p, q, and the essential supremum for a density of X such that*

$$\|f(X) - f(\hat{X})\|_q \leq C\|f\|_{\mathrm{TV}}\|X - \hat{X}\|_p^{\frac{p}{q(p+1)}}.$$

By the next lemma, we can obtain an approximation sequence of the payoff f.

Lemma 5.4.7. *Let f be a bounded Lipschitz continuous function whose weak derivative has bounded variation in $\mathbf{R}$. Then there exists a sequence $(f_j)_{j\geq 1} \subset C^1(\mathbf{R})$ such that*

$$\begin{aligned}
&\|f - f_j\|_\infty \to 0, \quad \textit{as } j \to \infty, \\
&\|f_j'\|_\infty \leq \|f\|_{\mathrm{Lip}} \quad \textit{for all } j \geq 1, \\
&\|f_j'\|_{\mathrm{TV}} \leq \|f'\|_{\mathrm{TV}} \quad \textit{for all } j \geq 1.
\end{aligned}$$

Proof. The approximate sequence can be constructed by mollifier convolutions $f_h := (f * \phi_h)$, that is, $\phi_h := \frac{1}{h}\phi(\frac{x}{h})$ with the conditions (i) $\phi \in C^\infty$, (ii) $\mathrm{supp}(\phi) \subset \{|x| \leq 1\}$, (iii) $\phi \geq 0$, and (iv) $\int_{\mathbf{R}} \phi(x)dx = 1$. □

Proof of Theorem 5.4.5. Assume that f is a bounded C^1 function whose derivative has bounded variation. As seen in Proposition 5.4.3, note that

$$\begin{aligned}
&\|f(X_T^\epsilon) - f(\bar{X}_T^\epsilon) + f(\bar{X}_T^0) - f(X_T^0)\|_2 \\
&\leq \|\int_0^1 (f'(\rho X_T^\epsilon + (1-\rho)\bar{X}_T^\epsilon) - f'(\rho X_T^0 + (1-\rho)\bar{X}_T^0))d\rho \cdot ((X_T^\epsilon)^{(1)} - (\bar{X}_T^\epsilon)^{(1)})\|_2 \\
&\quad + \|f\|_{\mathrm{Lip}}\|X_T^\epsilon - \bar{X}_T^\epsilon + \bar{X}_T^0 - X_T^0\|_2.
\end{aligned}$$

The final line is bounded by $\|f\|_{\mathrm{Lip}} \times O(\epsilon/n^{1/2})$, and thus we turn to focus on the estimate for the second line. The second line is bounded by

$$\begin{aligned}
&\|\int_0^1 (f'(\rho X_T^\epsilon + (1-\rho)\bar{X}_T^\epsilon) - f'(\rho X_T^0 + (1-\rho)\bar{X}_T^0))d\rho\|_{2p}\|X_T^\epsilon - \bar{X}_T^\epsilon\|_{2q} \\
&\leq \frac{C_{p,T,x_0}}{n^{1/2}} \int_0^1 \|(f'(\rho X_T^\epsilon + (1-\rho)\bar{X}_T^\epsilon) - f'(\rho X_T^0 + (1-\rho)\bar{X}_T^0))\|_{2p}d\rho
\end{aligned}$$

for any $p, q > 1$ such that $1/p + 1/q = 1$. Now using Lemma 5.4.6, we have

$$\begin{aligned}
&\int_0^1 \|(f'(\rho X_T^\epsilon + (1-\rho)\bar{X}_T^\epsilon) - f'(\rho X_T^0 + (1-\rho)\bar{X}_T^0))\|_{2p} d\rho \\
&\le C_{p,r} \|f'\|_{\mathrm{TV}} \int_0^1 \|(\rho X_T^\epsilon + (1-\rho)\bar{X}_T^\epsilon) - (\rho X_T^0 + (1-\rho)\bar{X}_T^0)\|_{2p}^{\frac{r}{2p(r+1)}} d\rho \\
&\le C_{p,r,T,x_0} \|f'\|_{\mathrm{TV}}\, \epsilon^{\frac{r}{2p(r+1)}}.
\end{aligned}$$

To obtain the result, we choose small $p > 1$ and large $r \ge 1$ such that $\frac{r}{2p(r+1)} > \frac{1}{2} - \delta$.

Finally, for general f, consider $f_K := (f \wedge K) \vee (-K)$ for $K > 0$ as a first approximation, and apply Lemma 5.4.7 to f_K. Then we obtain the desired result by taking the limit. □

5.5 Simulations

5.5.1 Numerical experiments for SABR model

In this section, we want to study an estimator of L^2-norm $(\frac{1}{M}\sum_{j=1}^M (X_T^{\epsilon,j} - \tilde{Y}_T^{\epsilon,(n),j})^2)^{1/2}$ for the SABR model. As a reference path, we use $\bar{X}_T^{\epsilon,(n_{\mathrm{ref}})}$ instead of X_T^ϵ.

We set the parameters as follows.

- $S_0 = 100$, $\beta = 0.9$, $\alpha_0 = 0.16 \times S_0^{2(1-\beta)}$, $\nu = 0.1$, $\rho = -0.6$, $T = 1$
- $n_{\mathrm{ref}} = 2^{14}$, $n = 8, 16, 32, 64, 128, 256$.

Here we considered a scaling for α_0 (via $S_t \approx S_0^{1-\beta} S_t^\beta$). The number of simulation M for the test is 10^5. The results are given in Figure 5.2. The accelerated scheme is faster than the standard method in both cases of L^2-error.

We next study the case with several ν. Let us compute the L^2-error ratio for a random variable Z which is defined as

$$\frac{E[|\bar{S}_T^{(n_{\mathrm{ref}})} - Z|^2]^{1/2}}{E[|\bar{S}_T^{(n_{\mathrm{ref}})} - \bar{S}_T^{(n)}|^2]^{1/2}} \times 100\ (\%).$$

We fix the other parameters in the previous. In Figure 5.3, we can check the efficiency of the asymptotic method (only) when ν is small enough.

Finally we compare $\tilde{Y}$ and $\check{Y}$ with different β. Figure 5.4 shows that the efficiency of $\check{Y}$ is very close to that of $\tilde{Y}$ as $\beta \approx 1$. Therefore if $\beta \approx 1$, we can apply the analytical tractability of $\check{Y}$ to pathwise simulation, computing expectations, or so on.

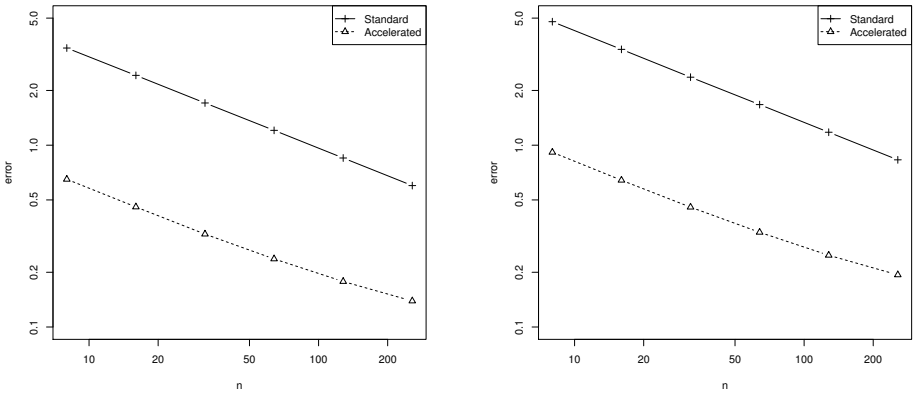

Figure 5.2: L^2-error : $E[|\bar{S}_T^{(n_{\mathrm{ref}})} - \tilde{Y}_T^{(n)}|^2]^{1/2}$ for the left and $E[\max_{0\le i\le n_{\mathrm{ref}}} |\bar{S}_{t_i}^{(n_{\mathrm{ref}})} - \tilde{Y}_{t_i}^{(n)}|^2]^{1/2}$ for the right.

5.5.2 Numerical tests for MLMC

To show that the accelerated method is more efficient than the standard method with MLMC, we take a numerical test for $E[\bar{P}_l^{\mathrm{new}} - P] = O(\epsilon/n_l)$, and $\mathrm{Var}^{1/2}(\bar{P}_l^{\mathrm{new}} - \bar{P}_{l-1}^{\mathrm{new}}) = O(\epsilon/n_l^{1/2})$ under the SABR model with small parameter ν. Let us consider payoff functions (European and digital options)

$$f(x) = \max(0, x - 100) \text{ or } f(x) = 1_{\{x-100\ge 0\}}$$

and the parameters

- $S_0 = 100$, $\beta = 1$, $\alpha_0 = 0.16$, $\nu = 0.1$, $\rho = -0.6$, $T = 1$

The level structure of MLMC is given by $k = 4$, i.e., $n_l = 4^l$. As a localization for digital option, we use

$$f_s(x) = (\max(x - 100 + h, 0) - \max(x - 100 - h, 0))/2h.$$

Here we set $h = 1.0$.

Figure 5.5 and 5.6 show the numerical results. We used the number of simulation $M = 10^7$ for the left, and $M = 10^5$ for the right. The results basically imply that the accelerated method works better than the standard one as in preceding numerical experiments. Remarkably the accelerated method performs worse in the case of variance estimates for digital option, likely due to discontinuity of the payoff function. In contrast, the localized scheme (Accelerated_loc) performs better than the others to some extent. We note that for general $1/2 \le \beta < 1$, the (semi-)analytical formula for CEV option pricing model can be used in order to compute $E[f(S_T^0)]$ (See [79]).

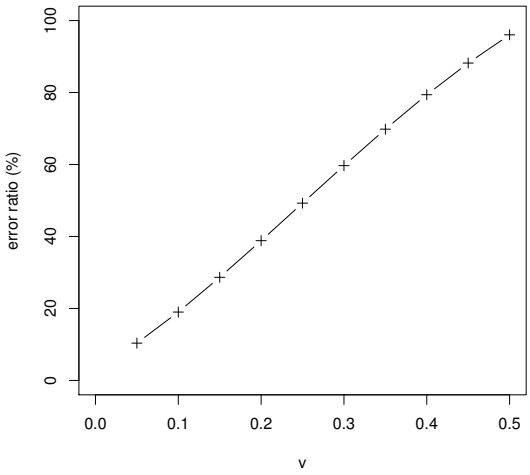

Figure 5.3: L^2-error ratio for $\tilde{Y}$ with different ν.

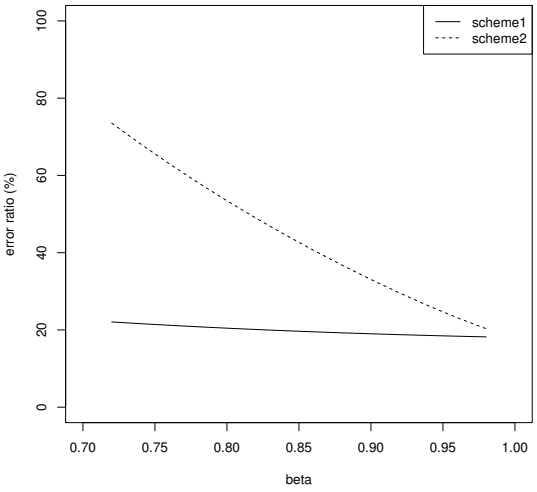

Figure 5.4: L^2-error ratio for $\tilde{Y}$ (scheme1) and $\check{Y}$ (scheme2) with different β.

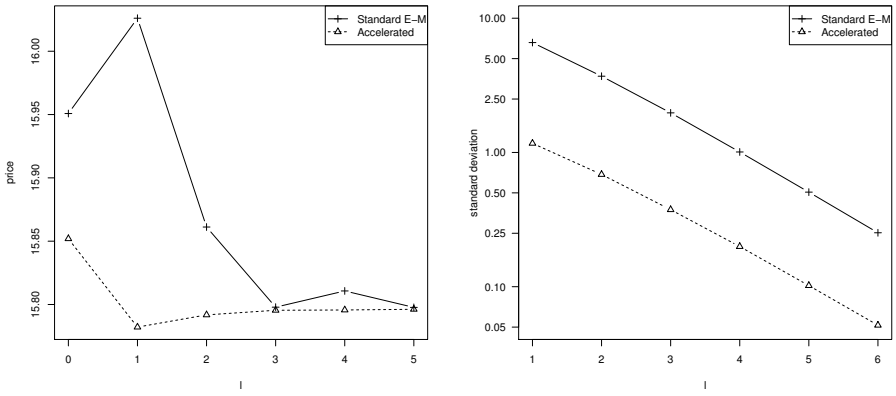

Figure 5.5: European option: (Left) A comparison of weak convergence between $\bar{P}_l$ and $\bar{P}_l^{\text{new}}$. (Right) A comparison of standard deviation between $\bar{P}_l - \bar{P}_{l-1}$ and $\bar{P}_l^{\text{new}} - \bar{P}_{l-1}^{\text{new}}$.

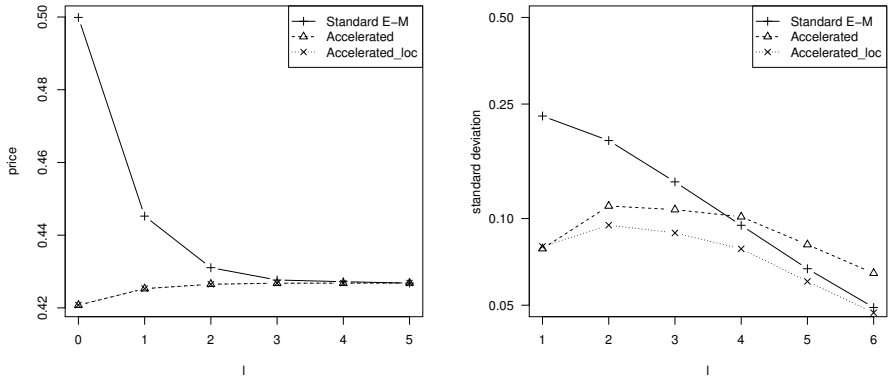

Figure 5.6: Digital option: (Left) A comparison of weak convergence between $\bar{P}_l$ and $\bar{P}_l^{\text{new}}$. (Right) A comparison of standard deviation between $\bar{P}_l - \bar{P}_{l-1}$ and $\bar{P}_l^{\text{new}} - \bar{P}_{l-1}^{\text{new}}$.

Chapter 6

Discrete approximation for nonlinear filtering

This chapter is based on the unpublished paper by Tanaka [93] (arXiv preprint: 1311.6090).

6.1 Introduction

The aim of this chapter is to determine the convergence rate of Picard's filter for nonlinear filtering in a more general condition than that of Picard ([73]), and to understand deeply why the scheme can perform with the rate. Although Picard's filter is based on an Euler-type approximation of stochastic differential equations, the error estimate does not rely on the standard argument of strong and weak convergence of the Euler-type scheme. As seen in the following, the properties of stochastic integrals under a conditional probability make the proof of convergence much more complicated.

Let us first formulate the nonlinear filtering problem with continuous time observations. Consider a stochastic process $(X_t)_{t\geq 0}$ (often called the signal process) defined as the solution of an N-dimensional stochastic differential equation

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s \tag{6.1}$$

with $x \in \mathbf{R}^N$ and an N-dimensional standard Brownian motion $B = (B_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions. We observe another d-dimensional process $(Y_t)_{t\geq 0}$ (called the observation process) defined by

$$Y_t = \int_0^t h(X_s)ds + W_t$$

where $W = (W_t)_{t\geq 0}$ is a d-dimensional standard Brownian motion independent of B. We denote the filtrations associated to B and Y by $(\mathcal{F}_t^B)$ and $(\mathcal{F}_t^Y)$ respectively.

The primary goal of nonlinear filtering problem is to investigate the evolution of the conditional distribution of X_T under the observation $(Y_t)_{0\le t\le T}$. In other words, we are interested in computing the value

$$E^P[g(X_T)|\mathcal{F}_T^Y]. \tag{6.2}$$

For this purpose, we consider the new probability measure Q on $\mathcal{F}_\infty = \sigma(\cup_{t\ge 0}\mathcal{F}_t)$ under which (Y_t) is a standard Brownian motion independent of (X_t), and (X_t) has the same law under P and Q. Throughout this chapter, we denote the expectation under Q by $E[\,\cdot\,]$. Then the conditional expectation (6.2) has the expression

$$E^P[g(X_T)|\mathcal{F}_T^Y] = \frac{E[g(X_T)\Phi_T|\mathcal{F}_T^Y]}{E[\Phi_T|\mathcal{F}_T^Y]}$$

with the Radon-Nikodym derivative

$$\Phi_t = \exp\Big(\sum_{j=1}^d \Big(\int_0^t h^j(X_s)dY_s^j - \frac{1}{2}\int_0^t (h^j)^2(X_s)ds\Big)\Big).$$

This is called the Kallianpur-Striebel formula (cf. [42], [4]). We need time discretization methods in order to compute $E[g(X_T)\Phi_T|\mathcal{F}_T^Y]$ since the stochastic integral term cannot be computed exactly.

Fix $T > 0$ and $\eta(t) = t_i := iT/n$ if $t \in [iT/n, (i+1)T/n)$. Let us use the notations: $\|\cdot\|_p \equiv \|\cdot\|_{p,Q} := E[|\cdot|^p]^{1/p}$ and $\|\cdot\|_{p,P} := E^P[|\cdot|^p]^{1/p}$. We now consider an approximation by a Riemann sum for Φ_T. Jean Picard showed the following surprizing result of L^2-convergence.

Theorem 6.1.1 ([73])**.** *Assume that g, b and σ are Lipshitz continuous and $h \in C_b^2(\mathbf{R}^N;\mathbf{R}^d)$. Then*

$$\Big\|E[g(X_T)\Phi_T|\mathcal{F}_T^Y] - E[g(X_T)\tilde{\Phi}_T|\mathcal{F}_T^Y]\Big\|_2 \le \frac{C_T}{n} \tag{6.3}$$

where

$$\tilde{\Phi}_t \equiv \tilde{\Phi}_t(X) := \exp\Big(\sum_{j=1}^d \Big(\int_0^t h^j(X_{\eta(s)})dY_s^j - \frac{1}{2}\int_0^t (h^j)^2(X_{\eta(s)})ds\Big)\Big).$$

Remark 6.1.2. The assumption $\|h\|_\infty < \infty$ can be weakened (see [73], [18]). For example, Picard ([73]) discusses the condition

$$E\Big[\exp\Big((1+\varepsilon)TH\Big(\sup_{0\le t\le T}|X_t|\Big)\Big)\Big] < \infty, \quad \text{for some } \varepsilon > 0$$

where

$$H(y) := \sup\Big\{\sum_{j=1}^d (h^j)^2(x); |x| \le y\Big\}.$$

The convergence error (6.3) is related to both of weak convergence of $\mathcal{F}_T^B$- measurable random variables and strong convergence of $\mathcal{F}_T^Y$-measurable random variables. Very roughly speaking, the order of convergence of the error is mainly from $\int_0^T (h(X_s) - h(X_{\eta(s)}))dY_s$. We notice that the difference $h(X_s) - h(X_{\eta(s)})$ has the weak error of $O(1/n)$, but this is averaged over the trajectory of (Y_s). That is why the rate of convergence is not so obvious. The proof given by Picard is quite complicated since we have to deal carefully with $\int_0^T \cdot \, dY_s$ under the conditional expectation $E[\, \cdot \, |\mathcal{F}_T^Y]$. In this work, we generalize the result (6.3) in terms of the regularity of g and L^p-estimates with $p > 2$ using several techniques in Malliavin calculus, and however, h is basically assumed to be bounded because of the difficulty in L^p-moment estimates for Φ_T and $\tilde{\Phi}_T$. See Theorem 6.2.1 and its proof.

In general situation, the distribution of X_t is unknown, and therefore some approximation schemes need to be applied to X_t. Let $\tilde{X}$ be a time discretization scheme of X, such as the Euler-Maruyama approximation or the stochastic ODE approximations via cubature formulas on Wiener space ([63], [69]). Then we have to analyze the error

$$\left\| E[g(X_T)\tilde{\Phi}_T(X)|\mathcal{F}_T^Y] - E[g(\tilde{X}_T)\tilde{\Phi}_T(\tilde{X})|\mathcal{F}_T^Y] \right\|_p \tag{6.4}$$

and this type of problem is discussed in e.g. [19], [20]. In the case where $\tilde{X}$ is Euler-Maruyama scheme, several researchers give error estimates for (6.3) and (6.4) simultaneously (e.g. [89], [65]).

It remains to discuss simulations of $E[g(\tilde{X}_T)\tilde{\Phi}_T(\tilde{X})|\mathcal{F}_T^Y]$ by Monte Carlo. In practice, the procedure of estimation for this is performed step-by-step for each observation time $T = t_1, \cdots, t_n, \cdots$. Hence it is important to construct special simulation methods, recursively in time T, to avoid the recalculation of the conditional expectation and explosion of time series data. For the reason, particle filters (or sequencial Monte Carlo methods) were originally developed by [31] and [44] for discrete time filtering. Recent developments of particle filters can be found in [22] and references therein.

Another approach to the computational problem for (6.2) is known as the stochastic partial differential equation (SPDE) approach. We can derive the equation of the dynamics of $t \mapsto E[g(X_t)\Phi_t|\mathcal{F}_t^Y]$ ($g \in C^2$) which is called the Zakai equation (cf. [4], [50]). The Zakai equation follows a SPDE with the finite dimensional noise Y. In that case, we have to consider time discretizations for the SPDE and give some error estimates for strong convergence (see e.g. [29]). We point out the relationship between the Zakai equation and Picard's filter $E[g(X_t)\tilde{\Phi}_t|\mathcal{F}_t^Y]$ in Remark 6.2.4.

The present chapter is organized as follows. In Section 6.2, we state the main result which is an extension of Picard's theorem, and shall give only the outline of the proof. In Section 6.3, we show the main part of the proof using infinite dimensional analysis on Wiener space.

6.2 L^p-convergence result

6.2.1 An extension of Picard's theorem

Let us fix $T > 0$. Throughout this chapter, the condition

$$E^P[\Phi_T^{-1}] = 1 \tag{6.5}$$

is always assumed to define the probability measure Q on $\mathcal{F}_T$, i.e. $Q(A) := E^P[1_A \Phi_T^{-1}]$ for $A \in \mathcal{F}_T$. The assumptions (A2)-(A3) introduced below imply the condition (6.5). See Kallianpur [41], Section 11.3.

We shall extend Picard's theorem as follows.

Theorem 6.2.1. *Assume that the following conditions hold:*

(A1) *The function $g : \mathbf{R}^N \to \mathbf{R}$ is a measurable function such that $g(X_T) \in \cap_{p \geq 1} L^p(\Omega, \mathcal{F}_T, Q)$.*

(A2) *The coefficients b and σ are Lipshitz continuous.*

(A3) *The function $h : \mathbf{R}^N \to \mathbf{R}^d$ is a C^2-function of polynomial growth with all derivatives.*

(A4) *For every $p \geq 1$,*

$$\|\Phi_T\|_p + \sup_n \|\tilde{\Phi}_T\|_p \leq K(p, T) < \infty.$$

Then for every $p \geq 1$, there exists a constant $C = C(p, T) > 0$ such that

$$\left\| E[g(X_T)\Phi_T | \mathcal{F}_T^Y] - E[g(X_T)\tilde{\Phi}_T | \mathcal{F}_T^Y] \right\|_p \leq \frac{C}{n}. \tag{6.6}$$

A typical example of (A4) is that h is bounded. The following corollary is an immediate consequence of Theorem 6.2.1.

Corollary 6.2.2. *Suppose the assumptions* (A1)-(A3) hold, and moreover h is assumed to be bounded. Then for every $p \geq 1$, there exists a constant $C = C(p, T) > 0$ such that

$$\left\| E^P[g(X_T)|\mathcal{F}_T^Y] - \frac{E[g(X_T)\tilde{\Phi}_T|\mathcal{F}_T^Y]}{E[\tilde{\Phi}_T|\mathcal{F}_T^Y]} \right\|_{p,P} \leq \frac{C}{n}.$$

Proof. Let $\rho_T(g) := E[g(X_T)\Phi_T|\mathcal{F}_T^Y]$ and $\tilde{\rho}_T(g) := E[g(X_T)\tilde{\Phi}_T|\mathcal{F}_T^Y]$. The error is expressed as

$$\frac{\rho_T(g)}{\rho_T(\mathbf{1})} - \frac{\tilde{\rho}_T(g)}{\tilde{\rho}_T(\mathbf{1})} = \frac{\rho_T(g) - \tilde{\rho}_T(g)}{\rho_T(\mathbf{1})} + \frac{\tilde{\rho}_T(g)}{\rho_T(\mathbf{1})\tilde{\rho}_T(\mathbf{1})}(\tilde{\rho}_T(\mathbf{1}) - \rho_T(\mathbf{1})).$$

It is possible to show from the boundedness of h that the $L^p(\Omega, \mathcal{F}_T, Q)$-norms of Φ_T, $\tilde{\Phi}_T$, $\rho_T(\mathbf{1})^{-1}$ and $\tilde{\rho}_T(\mathbf{1})^{-1}$ are bounded for every $p \geq 1$. Hence we obtain from Cauchy-Schwartz's inequality

$$\begin{aligned}\Big\| \frac{\rho_T(g)}{\rho_T(\mathbf{1})} - \frac{\tilde{\rho}_T(g)}{\tilde{\rho}_T(\mathbf{1})} \Big\|_{p,P} &= E\Big[\Big| \frac{\rho_T(g)}{\rho_T(\mathbf{1})} - \frac{\tilde{\rho}_T(g)}{\tilde{\rho}_T(\mathbf{1})} \Big|^p \Phi_T \Big]^{1/p} \\ &\leq C_1(p,T) \|\rho_T(g) - \tilde{\rho}_T(g)\|_{2p} - C_2(p,T) \|\rho_T(\mathbf{1}) - \tilde{\rho}_T(\mathbf{1})\|_{2p},\end{aligned}$$

which proves the desired result. □

Remark 6.2.3. For the proof of Theorem 6.2.1, the probability space $(\Omega, \mathcal{F}_T, Q)$ can be replaced by any other probability space on which $(X_t, Y_t)_{0 \leq t \leq T}$ has the same law. In the following, we fix the probability space so that $(B_t)_{0 \leq t \leq T}$ and $(Y)_{0 \leq t \leq T}$ are independent Brownian motions, and $(X_t)_{0 \leq t \leq T}$ is the solution of (6.1). The probability space will be assumed to be the Wiener space in Section 6.3.

Remark 6.2.4. As mentioned in the introduction, the time evolution

$$\rho_t(g) : t \mapsto E[g(X_t)\Phi_t | \mathcal{F}_t^Y], (g \in C_b^2)$$

solves the Zakai equation

$$\rho_t(g) = \rho_0(g) + \int_0^t \rho_s(\mathcal{L}g) ds + \int_0^t \rho_s(gh^{\mathsf{T}}) dY_s$$

where $\rho_0(g) = E[g(X_0)] = g(x)$ and $\mathcal{L}$ is the generator of X, i.e.

$$(\mathcal{L}g)(x) = \sum_{i=1}^N b^i(x) \frac{\partial g}{\partial x_i}(x) + \frac{1}{2} \sum_{i,j=1}^N (\sigma^i \sigma^j)(x) \frac{\partial^2 g}{\partial x_i \partial x_j}(x).$$

Picard's filter $\tilde{\rho}_t(g) : t \mapsto E[g(X_t)\tilde{\Phi}_T | \mathcal{F}_t^Y]$ can be understood as a semigroup-type approximation (or Markov chain approximation) in the following sense. Let X_t^x be a stochastic flow of the SDE (6.1) and $(P_t g)(x) := E[g(X_t^x)]$. Define a parameterized operator $\tilde{P}_t^y$, $y \in \mathbf{R}^d$ by

$$(\tilde{P}_t^y g)(x) := (P_t g)(x) \exp\Big(\sum_{j=1}^d \Big(h^j(x) y^j - \frac{1}{2} (h^j)^2(x) t \Big) \Big).$$

Then we can deduce that for $t_i \leq t < t_{i+1}$,

$$\tilde{\rho}_t(g) = \tilde{P}_{t_1 - t_0}^{Y_{t_1} - Y_{t_0}} \circ \cdots \circ \tilde{P}_{t_i - t_{i-1}}^{Y_{t_i} - Y_{t_{i-1}}} \circ \tilde{P}_{t - t_i}^{Y_t - Y_{t_i}}(g),$$

and $\tilde{P}_{t-t_i}^{Y_t - Y_{t_i}}(g)(x)$ is a solution of the evolution equation

$$\tilde{P}_{t-t_i}^{Y_t - Y_{t_i}}(g) = g(x) + \int_{t_i}^t \tilde{P}_{s-t_i}^{Y_s - Y_{t_i}}(\mathcal{L}g) ds + \int_{t_i}^t \tilde{P}_{s-t_i}^{Y_s - Y_{t_i}}(g) h^{\mathsf{T}}(x) dY_s,$$

which can be considered as the Zakai equation with the freezing coefficient $h(x)$.

6.2.2 Outline of proof

The proof of Theorem 6.2.1 is entirely different from the original one in [73]. Let us compute

$$g(X_T)\Phi_T - g(X_T)\tilde{\Phi}_T$$
$$= g(X_T)\Gamma_T \sum_{j=1}^{d} \Big(\int_0^T (h^j(X_s) - h^j(X_{\eta(s)}))dY_s^j - \frac{1}{2}\int_0^T ((h^j)^2(X_s) - (h^j)^2(X_{\eta(s)}))ds \Big)$$

where $\Gamma_T = \int_0^1 \Gamma_T(\rho)d\rho$, $\Gamma_T(\rho) = \exp(\rho \log(\Phi_T) + (1-\rho)\log(\tilde{\Phi}_T))$. Applying Itô's formula to $\zeta(X_s)$ with $\zeta = h^j$ or $(h^j)^2 \in C^2$, we have

$$\zeta(X_s) - \zeta(X_{\eta(s)}) = \int_{\eta(s)}^{s} \nabla\zeta(X_r)\sigma(X_r)dB_r + \int_{\eta(s)}^{s} (\mathcal{L}\zeta)(X_r)dr.$$

So the error $E[g(X_T)\Phi_T|\mathcal{F}_T^Y] - E[g(X_T)\tilde{\Phi}_T|\mathcal{F}_T^Y]$ can be decomposed into four parts $(E_i)_{1\le i\le 4}$:

$$E_1 = E\Big[g(X_T)\Gamma_T \sum_{j=1}^{d} \int_0^T \Big(\int_{\eta(s)}^{s} \nabla(h^j)(X_r)\sigma(X_r)dB_r \Big) dY_s^j \Big| \mathcal{F}_T^Y \Big]$$
$$E_2 = E\Big[g(X_T)\Gamma_T \sum_{j=1}^{d} \int_0^T \Big(\int_{\eta(s)}^{s} \mathcal{L}h^j(X_r)dr \Big) dY_s^j \Big| \mathcal{F}_T^Y \Big]$$
$$E_3 = -\frac{1}{2}E\Big[g(X_T)\Gamma_T \sum_{j=1}^{d} \int_0^T \Big(\int_{\eta(s)}^{s} \nabla((h^j)^2)(X_r)\sigma(X_r)dB_r \Big) ds \Big| \mathcal{F}_T^Y \Big]$$
$$E_4 = -\frac{1}{2}E\Big[g(X_T)\Gamma_T \sum_{j=1}^{d} \int_0^T \Big(\int_{\eta(s)}^{s} \mathcal{L}(h^j)^2(X_r)dr \Big) ds \Big| \mathcal{F}_T^Y \Big].$$

We are going to prove that

$$\|E_i\|_p \le \frac{C(i,p,T)}{n}$$

for $p \ge 2$ and $1 \le i \le 4$. The estimation for E_1 is the most difficult task since E_1 includes both dB and dY parts. First, we give the estimates for E_2 and E_4.

Proposition 6.2.5. *Under the assumption* (A1)-(A4), for every $p \ge 1$, there exists a constant $C = C(p,T) > 0$ such that

$$\|E_2\|_p + \|E_4\|_p \le \frac{C}{n}.$$

Proof. We only prove $\|E_2\|_p \le C/n$. By the assumption (A4), it holds that

$$\|\Gamma_T\|_q \le \|\Phi_T\|_q + \|\tilde{\Phi}_T\|_q \le K(q,T) < \infty$$

for every $q \geq 1$. Using the Cauchy-Schwartz inequality and Burkholder-Davis-Gundy inequality, we have

$$\|E_2\|_p \leq \|g(X_T)\Gamma_T\|_{2p} E\Big[\Big(\sum_{j=1}^{d}\int_0^T \Big(\int_{\eta(s)}^{s} \mathcal{L}h^j(X_r)dr\Big)dY_s^j\Big)^{2p}\Big]^{1/2p}$$
$$\leq C_1(p,T)\sum_{j=1}^{d} E\Big[\Big(\int_0^T \Big(\int_{\eta(s)}^{s} \mathcal{L}h^j(X_r)dr\Big)^2 ds\Big)^{p}\Big]^{1/2p}.$$

We can finally get the estimate

$$E\Big[\Big(\int_0^T \Big(\int_{\eta(s)}^{s} \mathcal{L}h^j(X_r)dr\Big)^2 ds\Big)^{p}\Big]^{1/2p} \leq E\Big[\sup_{0\leq r\leq T}|(\mathcal{L}h^j)(X_r)|^{2p}\Big(\int_0^T (s-\eta(s))^2 ds\Big)^{p}\Big]^{1/2p}$$
$$\leq \frac{C_2(p,T)}{n}.$$

□

6.3 The estimation via infinite dimensional analysis

6.3.1 A brief review of Malliavin calculus and Hilbert space valued martingales

Let $(\Omega, \mathcal{F}, Q)$ be a d-dimensional Wiener space and $(B_t)_{0\leq t\leq T}$ be the d-dimensional canonical Brownian motion on $(\Omega, \mathcal{F}, Q)$. More precisely, $\Omega = C([0,T];\mathbf{R}^d)$, $\mathcal{F}$ is the Borel σ-field on Ω, and Q is the Wiener measure under which the coodinate map $t \mapsto B_t, B \in \Omega$ becomes a standard Brownian motion.

The Malliavin derivative $D : L^2(\Omega) \supset \mathrm{Dom}(D) \to L^2(\Omega; L^2([0,T];\mathbf{R}^d))$ is defined as the extension of the following closable operator for smooth Wiener functional F:

$$F = f\Big(\int_0^T h_1(s)dB_s, \ldots, \int_0^T h_m(s)dB_s\Big)$$

where $f : \mathbf{R}^m \to \mathbf{R}$ is a polynomial function and $(h_i) \subset L^2([0,T];\mathbf{R}^d)$. Then

$$DF := \sum_{i=1}^{m}(\partial_i f)\Big(\int_0^T h_1(s)dB_s, \ldots, \int_0^T h_m(s)dB_s\Big)h_i.$$

The Skorohod integral $\delta : L^2(\Omega; L^2([0,T];\mathbf{R}^d)) \supset \mathrm{Dom}(\delta) \to L^2(\Omega)$ is the adjoint operator of D. Let K be a real separable Hilbert space. We can similarly define D and δ for K-valued Wiener functionals. The spaces $\mathbb{D}^{1,p}(K) \subset L^p(\Omega; K)$ are defined as the Sobolev spaces induced by the derivative operator D for K-valued Wiener

functionals. For the details of the precise formulation of Malliavin calculus, we refer to [80] and [70].

We prepare some results for the Skorohod integral δ (cf. [70]).

Lemma 6.3.1. *For* $u(\cdot) = \sum_{i=1}^n F_i 1_{[t_i,t_{i+1})}(\cdot) \in L^2([0,T];\mathbf{R}^d)$ *with* $F_i \in \mathbb{D}^{1,2}(\mathbf{R}^d)$, *we have*

$$\delta(u) = \sum_{i=1}^n F_i \cdot (B_{t_{i+1}} - B_{t_i}) - \sum_{i=1}^n \int_{t_i}^{t_{i+1}} \sum_{j=1}^d D_r^{(j)} F_i^{(j)} dr.$$

Lemma 6.3.2 (Continuity of δ)**.** *Let* $p > 1$. *There exists* $C > 0$ *such that*

$$\|\delta(u)\|_p \le C \|u\|_{\mathbb{D}^{1,p}(L^2([0,T];\mathbf{R}^d))}$$

for every $u \in \mathbb{D}^{1,p}(L^2([0,T];\mathbf{R}^d))$

We will use a kind of Fubini's theorem below.

Lemma 6.3.3. *Let* $(u_s)_{0\le s\le T} \in L^2([0,T];\mathbb{D}^{1,2}(L^2([0,T];\mathbf{R}^d)))$, *then*

$$\int_0^T \delta(u_s(\cdot))ds = \delta\Big(\int_0^T u_s(\cdot)ds\Big) \quad a.s. \tag{6.7}$$

Proof. Let $u_s^k = \sum_{j=1}^{m_k} a_j^k 1_{B_j^k}(s)$ with $a_j^k \in \mathbb{D}^{1,2}(L^2([0,T];\mathbf{R}^d))$ and $B_j^k \in \mathcal{B}([0,T])$ such that $u^k \to u$ in the norm of $L^2([0,T];\mathbb{D}^{1,2}(L^2([0,T];\mathbf{R}^d)))$ as $k \to \infty$. Clearly we have

$$\int_0^T \delta(u_s^k(\cdot))ds = \delta\Big(\int_0^T u_s^k(\cdot)ds\Big).$$

It suffices to check the limit of both sides. By taking L^2-norm,

$$\begin{aligned}\Big\| \int_0^T \delta(u_s^k(\cdot))ds - \int_0^T \delta(u_s(\cdot))ds\Big\|_2^2 &\le C_1 \int_0^T \|\delta(u_s^k(\cdot) - u_s(\cdot))\|_2^2 ds \\ &\le C_2 \int_0^T \|u_s^k(\cdot) - u_s(\cdot)\|^2_{\mathbb{D}^{1,2}(L^2([0,T];\mathbf{R}^d))} ds\end{aligned}$$

and

$$\begin{aligned}\Big\|\delta\Big(\int_0^T u_s^k(\cdot)ds\Big) - \delta\Big(\int_0^T u_s(\cdot)ds\Big)\Big\|_2^2 &\le C_3 \Big\| \int_0^T (u_s^k(\cdot) - u_s(\cdot))ds\Big\|^2_{\mathbb{D}^{1,2}(L^2([0,T];\mathbf{R}^d))} \\ &\le C_4 \int_0^T \|u_s^k(\cdot) - u_s(\cdot)\|^2_{\mathbb{D}^{1,2}(L^2([0,T];\mathbf{R}^d))} ds.\end{aligned}$$

Thus we obtain the result (6.7) as $k \to \infty$. □

We can derive the following fundamental inequalities for Hilbert space valued martingales.

Lemma 6.3.4. *Let M_t be a continuous K-valued martingale with respect to a filtration $(\mathcal{F}_t)$ which satisfies the usual conditions. Then for every $p > 0$, there exists positive constants K_p, $c_p < C_p$ such that*

Doob's inequality:

$$E\Big[\sup_{0\le t\le T}|M_t|_K^p\Big] \le K_p E\Big[|M_T|_K^p\Big].$$

Burkholder-Davis-Gundy's inequality:

$$c_p E\Big[\langle M\rangle_T^{p/2}\Big] \le E\Big[\sup_{0\le t\le T}|M_t|_K^p\Big] \le C_p E\Big[\langle M\rangle_T^{p/2}\Big].$$

Proof. See e.g. [80, Theorem 3.1]. □

Lemma 6.3.5. *If $F \in L^p(\mathcal{F}_T^B; K)$ for some $p \ge 2$, then there exists an unique process $f_s = (f_s^1, \dots, f_s^d)$ such that f_s^i are K-valued progressively measurable processes satisfying*

$$F = E[F] + \int_0^T f_s dB_s,$$

and

$$E\Big[\Big(\int_0^T \sum_{i=1}^d |f_s^i|_K^2 ds\Big)^{p/2}\Big] \le C_p E[|F|_K^p]. \tag{6.8}$$

In particular, if $F \in \mathbb{D}^{1,2}(\mathcal{F}_T^B; K)$, then we have the so-called Clark-Ocone formula

$$f_s(\omega) = E[D_s F|\mathcal{F}_s^B](\omega) \text{ a.e. } (s,\omega) \in [0,T]\times\Omega.$$

Proof. We check only (6.8) using the inequalities in Lemma 6.3.4:

$$\begin{aligned} E\Big[\Big(\int_0^T \sum_{i=1}^d |f_s^i|_K^2 ds\Big)^{p/2}\Big] &\le C_1(p) E\Big[\Big|\int_0^T f_s dB_s\Big|_K^{p/2}\Big] \\ &= C_1(p) E[|F - E[F]|_K^p] \\ &\le C_2(p) E[|F|_K^p]. \end{aligned}$$

□

6.3.2 Infinite dimensional Itô calculus for E_3

Let us define two Wiener spaces $(\mathcal{W}_B, \mathcal{B}(\mathcal{W}_B), P^{\mathcal{W}_B})$ and $(\mathcal{W}_Y, \mathcal{B}(\mathcal{W}_Y), P^{\mathcal{W}_Y})$ on which $(B_t)_{0\le t\le T}$ and $(Y_t)_{0\le t\le T}$ are canonical Brownian motions respectively. From now on we specify

$$(\Omega, \mathcal{F}, Q) = (\mathcal{W}_B, \mathcal{B}(\mathcal{W}_B), P^{\mathcal{W}_B}) \times (\mathcal{W}_Y, \mathcal{B}(\mathcal{W}_Y), P^{\mathcal{W}_Y}).$$

We denote by $E^{\mathcal{W}_B}$ and $E^{\mathcal{W}_Y}$ the expectations under $P^{\mathcal{W}_B}$ and $P^{\mathcal{W}_Y}$ respectively. Since B and Y are independent, we notice that $E[\,\cdot\,|\mathcal{F}_T^Y] = E^{\mathcal{W}_B}[\,\cdot\,]$.

We now return to prove $\|E_3\|_p = O(1/n)$. The fundamental idea to get the order of convergence is as follows (see also [17]): Let $F \in L^2(\mathcal{W}_B \times \mathcal{W}_Y; \mathbf{R})$ and θ_s be a $\mathcal{F}_s^B$-adapted process with finite moments. We are going to give the error estimates for the type of $E^{\mathcal{W}_B}[F\int_{t_i}^{t_{i+1}} \theta_s dB_s]$. Let us consider

$$L^2(\mathcal{W}_B \times \mathcal{W}_Y; \mathbf{R}) \cong L^2(\mathcal{W}_B; L^2(\mathcal{W}_Y; \mathbf{R})).$$

By Lemma 6.3.5, we obtain the representation $F = E^{\mathcal{W}_B}[F] + \int_0^T f_s dB_s$, which has already appeared in Picard's paper [73, Proposition 1]. Applying this representation to $E^{\mathcal{W}_B}[F\int_{t_i}^{t_{i+1}} \theta_s dB_s]$, we obtain a conditional duality formula

$$E^{\mathcal{W}_B}\Big[F\int_{t_i}^{t_{i+1}} \theta_s dB_s\Big] = E^{\mathcal{W}_B}\Big[\int_{t_i}^{t_{i+1}} f_s\theta_s ds\Big] \in L^2(\mathcal{W}_Y; \mathbf{R}).$$

This means that it is possible to prove the convergence of $O(1/n)$ from the term $\int_{t_i}^{t_{i+1}} \cdot\, ds$ if (f_s) has good moment estimates.

Lemma 6.3.6. *Let $p \geq 2$ and suppose $F \in L^p(\mathcal{W}_B \times \mathcal{W}_Y; \mathbf{R})$ has the representation $F = E^{W_B}[F] + \int_0^T f_s dB_s$ (in Lemma 6.3.5), then there exists a constant $C = C(p) > 0$ such that*

$$E\Big[\Big(\int_0^T |f_s|^2 ds\Big)^{p/2}\Big] \leq CE[|F|^p]. \tag{6.9}$$

Proof. Recall that $|\cdot|$ is the norm on $\mathbf{R}^d$. Let us consider the $L^2(\mathcal{W}_Y; \mathbf{R})$-valued martingale $\int_0^t f_s dB_s$ as the $\mathbf{R}$-valued stochastic integral for the $\mathbf{R}^d$-valued $\mathcal{F}_s^B \vee \mathcal{F}_T^Y$-progressively measurable process f_s on $(\Omega, \mathcal{F}, Q)$ (see e.g. [18, Lemma 21.2]). We can apply Lemma 6.3.4 with $K = \mathbf{R}$ to it. □

Proposition 6.3.7. *Under the assumption* (A1)-(A4), for every $p \geq 1$, there exists a constant $C = C(p, T) > 0$ such that

$$\|E_3\|_p \leq \frac{C}{n}.$$

Proof. We prove only the one dimensional case. Let $\theta_r = \frac{1}{2}(h^2)'(X_r)\sigma(X_r)$. Using Itô's formula for stochastic integrals with respect to B_t, we can deduce that

$$E^{\mathcal{W}_B}\Big[E^{\mathcal{W}_B}[g(X_T)\Gamma_T]\int_0^T\int_{\eta(s)}^s \theta_r dB_r ds\Big] = 0$$

and

$$\begin{aligned}
E^{\mathcal{W}_B}\Big[\int_0^T f_s dB_s \int_0^T\int_{\eta(s)}^s \theta_r dB_r ds\Big] &= \int_0^T E^{\mathcal{W}_B}\Big[\int_0^T f_r dB_r \int_{\eta(s)}^s \theta_r dB_r\Big] ds \\
&= \int_0^T E^{\mathcal{W}_B}\Big[\int_{\eta(s)}^s f_r\theta_r dr\Big] ds \\
&= \int_0^T\int_{\eta(s)}^s E^{\mathcal{W}_B}[f_r\theta_r] dr ds.
\end{aligned}$$

We notice that

$$|E^{\mathcal{W}_B}[f_r\theta_r]| \le E^{\mathcal{W}_B}[|f_r|^2]^{1/2} \sup_{0\le r\le T} E^{\mathcal{W}_B}[|\theta_r|^2]^{1/2}.$$

Therefore the estimate (6.9) in Lemma 6.3.6 implies

$$\begin{aligned}\Big\| E^{\mathcal{W}_B}\Big[\int_0^T \int_{\eta(s)}^s f_r\theta_r dr ds\Big]\Big\|_p^p &\le \sup_{0\le r\le T} E^{\mathcal{W}_B}[|\theta_r|^2]^{p/2} \\ &\qquad E^{\mathcal{W}_Y}\Big[\Big(\int_0^T \int_{\eta(s)}^{\eta(s)+T/n} E^{\mathcal{W}_B}[|f_r|^2]^{1/2} dr ds\Big)^p\Big] \\ &\le C_1\Big(\frac{T}{n}\Big)^p E\Big[\Big(\int_0^T |f_r|^2 dr\Big)^{p/2}\Big] \\ &\le \frac{C_2}{n^p}\|g(X_T)\Gamma_T\|_p^p\end{aligned}$$

for some constant $C_2 = C_2(p,T)$. □

6.3.3 Partial Malliavin calculus for E_1

For analyzing E_1, we study partial Malliavin calculus introduced in [71]. Consider Malliavin calculus for each space of $(\mathcal{W}_B, \mathcal{B}(\mathcal{W}_B), P^{\mathcal{W}_B})$ and $(\mathcal{W}_Y, \mathcal{B}(\mathcal{W}_Y), P^{\mathcal{W}_Y})$. Let us denote the Sobolev spaces, the Malliavin derivative, and the Skorohod integral on $(\mathcal{W}_B, \mathcal{B}(\mathcal{W}_B), P^{\mathcal{W}_B})$ by $\mathbb{D}_B^{k,p}$, D_t^B, δ_B, and on $(\mathcal{W}_Y, \mathcal{B}(\mathcal{W}_Y), P^{\mathcal{W}_Y})$ by $\mathbb{D}_Y^{k,p}$, D_t^Y, δ_Y. We note that D^B and D^Y are naturally extended to $(N+d)$-dimensional Wiener space $(\Omega, \mathcal{F}, Q)$, and the pair (D^B, D^Y) coincides with the standard Malliavin derivative $D : \Omega \to L^2([0,T]; \mathbf{R}^{N+d})$ in the following sense: Let us consider an orthogonal decomposition

$$L^2([0,T]; \mathbf{R}^{N+d}) = L_B^2 \oplus L_Y^2$$

with

$$\begin{aligned}L_B^2 &= \{f \in L^2([0,T]; \mathbf{R}^{N+d}) : f^{(j)} \equiv 0 \text{ for } N < j \le N+d\} \cong L^2([0,T]; \mathbf{R}^N), \\ L_Y^2 &= \{f \in L^2([0,T]; \mathbf{R}^{N+d}) : f^{(j)} \equiv 0 \text{ for } 1 \le j \le N\} \cong L^2([0,T]; \mathbf{R}^d).\end{aligned}$$

Let Π_B and Π_Y be the projections from $L^2([0,T]; \mathbf{R}^{N+d})$ to L_B^2 and L_Y^2 respectively. Then we can define $D^B := \Pi_B \circ D$ and $D^Y := \Pi_Y \circ D$ on the $(N+d)$-dimensional Wiener space $(\Omega, \mathcal{F}, Q)$. This formulation is called the "partial" Malliavin calculus ([56], [71]).

Let K be a real separable Hilbert space and $G \in L^2(\mathcal{W}_B; K)$. We define by J_t^B the projection so that $G = E^{\mathcal{W}_B}[G] + \int_0^T J_s^B(G) dB_s$. In particular, if $G \in \mathbb{D}_B^{1,2}(\mathbb{D}_Y^{1,2}(\mathbf{R})) \subset L^2(\mathcal{W}_B; \mathbb{D}_Y^{1,2}(\mathbf{R}))$, we have by the Clark-Ocone formlua

$$J_s^B(G) = E^{\mathcal{W}_B}[D_s^B G | \mathcal{F}_s^B] \in \mathbb{D}_Y^{1,2}(\mathbf{R}). \tag{6.10}$$

We note that $\mathbb{D}_B^{1,2}(\mathbb{D}_Y^{1,2}(\mathbf{R})) \neq \mathbb{D}_{(B,Y)}^{2,2}(\mathbf{R})$.

Let us first present auxiliary lemma which will be used in later computations.

Lemma 6.3.8. (i): *For $G \in L^2(\mathcal{W}_B; \mathbb{D}^{1,2}_Y(K))$,*

$$D^Y E^{\mathcal{W}_B}[G] = E^{\mathcal{W}_B}[D^Y G] \quad a.s.$$

(ii): *If $\xi \in \mathbb{D}^{1,p}_B(L^2([0,T];\mathbf{R}^d))$ with some $p \geq 2$, then $\int_0^T \xi_s dY_s \in \mathbb{D}^{1,p}_B(\mathbb{D}^{1,2}_Y(\mathbf{R}))$ and*

$$\begin{aligned} D^B\Big(\int_0^T \xi_s dY_s\Big) &= \int_0^T (D^B\xi_s) dY_s, \\ D^Y\Big(\int_0^T \xi_s dY_s\Big) &= \xi, \\ D^Y D^B\Big(\int_0^T \xi_s dY_s\Big) &= D^B D^Y\Big(\int_0^T \xi_s dY_s\Big) = D^B\xi. \end{aligned}$$

Proof. (i): We choose an approximation sequence (G_k) of the form $G_k = \sum_{i=1}^m S_i 1_{A_i}$, $S_i \in \mathbb{D}^{1,2}_Y(K)$ and $A_i \in \mathcal{B}(\mathcal{W}_B)$. For each k, G_k clearly satisfies the desired equality. Thus we obtain the result using the continuity of D. (ii): This is a version of the proof of [70, Proposition 1.3.8], recall that $D^B(Y_t) = 0$. □

We will take an approximation sequence $(Z_\ell)_\ell \subset \mathbb{D}^{1,2p}_B(\mathbf{R})$ such that $Z_\ell \to g(X_T)$ in $L^{2p}(\mathcal{W}_B)$ as $\ell \to \infty$. The following lemma plays a key role for the estimate of E_1.

Lemma 6.3.9. *Let $p \geq 2$ and $Z \in \mathbb{D}^{1,2p}_B(\mathbf{R})$. Then under the assumptions* (A2)-(A4), *$Z\Gamma_T(\rho) \in \mathbb{D}^{1,p}_B(\mathbb{D}^{1,2}_Y(\mathbf{R}))$. Moreover, let (θ_s) be a $\mathbf{R}^d$-valued continuous $\mathcal{F}^B_s$-progressively measurable process with $E[\sup_{0\leq s\leq T}|\theta_s|^4]^{1/4} \leq M$, then there exists a constant $C = C(p,T)$ such that*

$$E^{\mathcal{W}_Y}\Big[\Big(\int_0^T \operatorname*{ess\,sup}_{0\leq r\leq T} |E^{\mathcal{W}_B}[D^Y_r J^B_s(Z\Gamma_T(\rho))\cdot\theta_s]|^2 ds\Big)^{p/2}\Big] \leq M^p C \|Z\Gamma_T(\rho)\|^p_p.$$

Proof. We can check that $X_t \in \cap_{p\geq 1}\mathbb{D}^{1,p}_B$ under Assumption (A2). Using the chain rule of Malliavin derivative, we obtain from Lemma 6.3.8 and Assumption (A4)

$$\Gamma_T(\rho,k) := \sum_{l=0}^k \frac{(\log(\Gamma_T(\rho)))^l}{l!} \in \bigcap_{p\geq 1} \mathbb{D}^{1,p}_B(\mathbb{D}^{1,2}_Y(\mathbf{R})).$$

Thus taking the limit $k \to \infty$, we can show that

$$\Gamma_T(\rho) \in \bigcap_{p\geq 1} \mathbb{D}^{1,p}_B(\mathbb{D}^{1,2}_Y(\mathbf{R})),$$

which implies $Z\Gamma_T(\rho) \in \mathbb{D}^{1,p}_B(\mathbb{D}^{1,2}_Y(\mathbf{R}))$.

Applying the Clark-Ocone formula (6.10) to $Z\Gamma_T(\rho)$, we deduce that

$$D^Y_r J^B_s(Z\Gamma_T(\rho)) = D^Y_r E^{\mathcal{W}_B}[D^B_s(Z\Gamma_T(\rho))|\mathcal{F}^B_s] = E^{\mathcal{W}_B}[D^B_s(ZD^Y_r\Gamma_T(\rho))|\mathcal{F}^B_s].$$

We notice that

$$D_r^Y \Gamma_T(\rho) = \exp(\rho \log(\Phi_T) + (1-\rho)\log(\tilde{\Phi}_T))(\rho h(X_r) + (1-\rho)h(X_{\eta(r)}))$$

and then

$$\begin{aligned} D_s^B(ZD_r^Y\Gamma_T(\rho)) =& D_s^B(Z\Gamma_T(\rho))(\rho h(X_r) + (1-\rho)h(X_{\eta(r)})) \\ &+ Z\Gamma_T(\rho) D_s^B(\rho h(X_r) + (1-\rho)h(X_{\eta(r)})). \end{aligned}$$

This formula and the Cauchy-Schwartz inequality for the conditional expectation $E[\cdot|\mathcal{F}_s^B]$ imply

$$\begin{aligned} &|E^{\mathcal{W}_B}[D_r^Y J_s^B(Z\Gamma_T(\rho)) \cdot \theta_s]|^2 \\ &\le 2E^{\mathcal{W}_B}[|J_s^B(Z\Gamma_T(\rho))|^2]E^{\mathcal{W}_B}[|(\rho h(X_r) + (1-\rho)h(X_{\eta(r)})) \cdot \theta_s|^2] \\ &\quad + 2E^{\mathcal{W}_B}[|Z\Gamma_T(\rho)|^2]E^{\mathcal{W}_B}[|D_s^B(\rho h(X_r) + (1-\rho)h(X_{\eta(r)})) \cdot \theta_s|^2]. \end{aligned}$$

We refer for the reader to the basic estimate ([70]): for any $q \ge 1$,

$$E^{\mathcal{W}_B}\Big[\sup_{0\le t\le T}|X_t|^q\Big] + \sup_{0\le s\le T} E^{\mathcal{W}_B}\Big[\sup_{0\le t\le T}|D_s^B X_t|^q\Big] \le C_1(q,T) < \infty. \tag{6.11}$$

The above inequality allows us to show that

$$|E^{\mathcal{W}_B}[D_r^Y J_s^B(Z\Gamma_T(\rho)) \cdot \theta_s]|^2 \le C_2(p,T)(E^{\mathcal{W}_B}[|J_s^B(Z\Gamma_T(\rho))|^2] + E^{\mathcal{W}_B}[|Z\Gamma_T(\rho)|^2]).$$

We can show by Jensen's inequality and Lemma 6.3.6 that

$$\begin{aligned} E^{\mathcal{W}_Y}\Big[\Big(\int_0^T E^{\mathcal{W}_B}[|J_s^B(Z\Gamma_T(\rho))|^2]ds\Big)^{p/2}\Big] &\le E\Big[\Big(\int_0^T |J_s^B(Z\Gamma_T(\rho))|^2 ds\Big)^{p/2}\Big] \\ &\le C_3(p)E[|Z\Gamma_T(\rho)|^p]. \end{aligned}$$

Using these inequalities, we obtain the constant C in the assertion. □

We now finish the proof of the main theorem.

Proposition 6.3.10. *Let the assumptions* (A1)-(A4) hold. Then for every $p \ge 2$, there exists a constant $C = C(p,T) > 0$ such that

$$\|E_1\|_p \le \frac{C}{n}.$$

Proof. We first define

$$E_1(\rho) := E\Big[g(X_T)\Gamma_T(\rho)\sum_{j=1}^d \int_0^T \Big(\int_{\eta(s)}^s \nabla(h^j)(X_r)\sigma(X_r)dB_r\Big)dY_s^j \Big| \mathcal{F}_T^Y\Big]$$

and then

$$\|E_1\|_p \le \int_0^1 \|E_1(\rho)\|_p d\rho \le \sup_{0\le\rho\le 1} \|E_1(\rho)\|_p.$$

So it suffices to give an estimate for $\|E_1(\rho)\|_p$.

Let us define for $Z \in \mathbb{D}_B^{1,2p}(\mathbf{R})$

$$E_1(\rho, Z) := E\Big[Z\Gamma_T(\rho)\sum_{j=1}^d \int_0^T \Big(\int_{\eta(s)}^s \nabla(h^j)(X_r)\sigma(X_r)dB_r\Big)dY_s^j \Big| \mathcal{F}_T^Y\Big].$$

We shall show that

$$\|E_1(\rho, Z)\|_p \le \frac{C}{n}\|Z\Gamma_T(\rho)\|_p, \tag{6.12}$$

and then taking an approximation sequence $(Z_\ell)_\ell \subset \mathbb{D}_B^{1,2p}(\mathbf{R})$ such that $Z_\ell \to g(X_T)$ in L^{2p}, we have

$$\|E_1(\rho)\|_p \le \frac{C}{n}\|g(X_T)\Gamma_T(\rho)\|_p \le \frac{\tilde{C}(p,T)}{n},$$

which is what we want to prove.

For notational simplicity, we prove (6.12) only the case where B and Y are one dimensional Brownian motions. Let $\theta_r = (h)'(X_r)\sigma(X_r)$. By Itô's formula,

$$\int_0^T \int_{\eta(s)}^s \theta_r dB_r dY_s = \sum_{i=0}^{n-1} \Big(\Big(\int_{t_i}^{t_{i+1}} \theta_s dB_s\Big)(Y_{t_{i+1}} - Y_{t_i}) - \int_{t_i}^{t_{i+1}} (Y_s - Y_{t_i})\theta_s dB_s\Big).$$

Let us define $f_s = f_s(\rho, Z) := J_s^B(Z\Gamma_T(\rho))$. We can deduce that

$$E^{\mathcal{W}_B}\Big[E^{\mathcal{W}_B}[Z\Gamma_T(\rho)]\int_0^T \int_{\eta(s)}^s \theta_r dB_r dY_s\Big] = 0$$

and

$$\begin{aligned}
&E^{\mathcal{W}_B}\Big[\int_0^T f_s dB_s \int_0^T \int_{\eta(s)}^s \theta_r dB_r dY_s\Big] \\
&= E^{\mathcal{W}_B}\Big[\int_0^T f_s dB_s \sum_{i=0}^{n-1} \Big(\Big(\int_{t_i}^{t_{i+1}} \theta_s dB_s\Big)(Y_{t_{i+1}} - Y_{t_i}) - \int_{t_i}^{t_{i+1}} (Y_s - Y_{t_i})\theta_s dB_s\Big)\Big] \\
&= E^{\mathcal{W}_B}\Big[\sum_{i=0}^{n-1} \Big(\Big(\int_{t_i}^{t_{i+1}} f_s\theta_s ds\Big)(Y_{t_{i+1}} - Y_{t_i}) - \int_{t_i}^{t_{i+1}} (Y_s - Y_{t_i})f_s\theta_s ds\Big)\Big] \\
&= \sum_{i=0}^{n-1} \Big(\Big(\int_{t_i}^{t_{i+1}} E^{\mathcal{W}_B}[f_s\theta_s]ds\Big)(Y_{t_{i+1}} - Y_{t_i}) - \int_{t_i}^{t_{i+1}} (Y_s - Y_{t_i})E^{\mathcal{W}_B}[f_s\theta_s]ds\Big).
\end{aligned}$$

By using Lemma 6.3.1 and the fact that $D^Y E^{\mathcal{W}_B}[\cdot] = E^{\mathcal{W}_B}[D^Y \cdot]$ in Lemma 6.3.8, it holds that

$$\begin{aligned}
&\sum_{i=0}^{n-1} \Big(\int_{t_i}^{t_{i+1}} E^{\mathcal{W}_B}[f_s\theta_s]ds \Big)(Y_{t_{i+1}} - Y_{t_i}) \\
&= \delta_Y\Big(\sum_{i=0}^{n-1} \Big(\int_{t_i}^{t_{i+1}} E^{\mathcal{W}_B}[f_s\theta_s]ds \Big) 1_{[t_i,t_{i+1})}(\cdot) \Big) + \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_{i+1}} E^{\mathcal{W}_B}[(D_r^Y f_s)\theta_s]dsdr,
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} (Y_s - Y_{t_i}) E^{\mathcal{W}_B}[f_s\theta_s]ds \\
&= \int_0^T \Big(\delta_Y(E^{\mathcal{W}_B}[f_s\theta_s]1_{[\eta(s),s)}(\cdot)) + \int_{\eta(s)}^s E^{\mathcal{W}_B}[(D_r^Y f_s)\theta_s]dr \Big) ds \\
&= \delta_Y\Big(\sum_{i=0}^{n-1} \Big(\int_{\cdot}^{t_{i+1}} E^{\mathcal{W}_B}[f_s\theta_s]ds \Big) 1_{[t_i,t_{i+1})}(\cdot) \Big) + \int_0^T \int_{\eta(s)}^s E^{\mathcal{W}_B}[(D_r^Y f_s)\theta_s]drds.
\end{aligned}$$

Here we used Lemma 6.3.3 in the second equality. Consequently we derive the formula

$$\begin{aligned}
&E^{\mathcal{W}_B}\Big[\int_0^T f_s dB_s \int_0^T \int_{\eta(s)}^s \theta_r dB_r dY_s \Big] \\
&= \delta_Y\Big(\sum_{i=0}^{n-1} \Big(\int_{t_i}^{\cdot} E^{\mathcal{W}_B}[f_s\theta_s]ds \Big) 1_{[t_i,t_{i+1})}(\cdot) \Big) + \int_0^T \int_{\eta(r)}^r E^{\mathcal{W}_B}[(D_r^Y f_s)\theta_s]dsdr.
\end{aligned}$$

Using the above formula and Lemma 6.3.2, we finally get the estimate

$$\begin{aligned}
\|E_1(\rho, Z)\|_p^p &\le \frac{C_1}{n^p} E\Big[\Big(\int_0^T |f_s|^2 ds \Big)^{p/2} \Big] \\
&\quad + \frac{C_2}{n^{p/2}} E^{\mathcal{W}_Y}\Big[\Big(\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_{i+1}} |E^{\mathcal{W}_B}[(D_r^Y f_s)\theta_s]|^2 dsdr \Big)^{p/2} \Big] \\
&\le \frac{C_3}{n^p} \|Z\Gamma_T(\rho)\|_p^p + \frac{C_4}{n^p} E^{\mathcal{W}_Y}\Big[\Big(\int_0^T \operatorname*{ess\,sup}_{0\le r\le T} |E^{\mathcal{W}_B}[(D_r^Y f_s)\theta_s]|^2 ds \Big)^{p/2} \Big]
\end{aligned}$$

Applying Lemma 6.3.9 to the last term, we obtain the result (6.12). This finishes the proof. □

Bibliography

[1] Alfonsi, A., High-order discretization schemes for the CIR process: Application to affine term structure and Heston models, *Math. Comp.* **79** (2010) 209-237.

[2] Asmussen, S., Rosiński, J., Approximations of small jumps of Lévy processes with a view towards simulation, *J. Appl. Probab.* **38** (2001) 482-493.

[3] Avikainen, R., On irregular functionals of SDEs and the Euler scheme, *Finance Stoch.* **13** (2009) 381-401.

[4] Bain, A., Crisan, D., *Fundamentals of Stochastic Filtering*, Springer, 2009.

[5] Bally, V., Pages, G., Printems, J., A quantization tree method for pricing and hedging multidimensional American options, *Math. Finance* **15** (2005) 119-168.

[6] Bally, V., Talay, D., The law of the Euler scheme for stochastic differential equations (I):convergence rate of the distribution function, *Probab. Theory Related Fields* **104** (1995) 43-60.

[7] Berridge, S.J., *Irregular grid methods for pricing high-dimensional American options*, Proefschrift, 2004.

[8] Berkaoui, A., Bossy, M., Diop, A., Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence *ESAIM Probab. Stat.* **12** (2008) 1-11.

[9] Brigo, D., Mercurio, F., *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit*, Springer, 2006.

[10] Broadie, M., Glasserman, P., A stochastic mesh method for pricing high-dimensional American options, *Journal of Computational Finance* **7** (2004) 35-72.

[11] Bosy, M., Diop, A., An efficient discretisation scheme for one dimensional SDEs with a diffusion coefficient function of the form $|x|^a$, $a \in [1/2, 1)$ (*INRIA Working Paper*, 2004).

[12] Bungartz, H.J., Griebel, M., Sparse grids, *Acta Numerica* **13** (2004) 147-269.

[13] Butcher, J.C., *Numerical Methods for Ordinary Differential Equations*, Willey, 2003.

[14] Carr, P, Madan, D., Option valuation using the fast fourier transform, *Journal of Computational Finance* **2** (1998) 61-73.

[15] Carriere, J.F., Valuation of the early-exercise price for derivative securities using simulations and splines, *Insurance: Math. Econom.* **19** (1996) 19-30.

[16] Chevance, D., Numerical methods for backward stochastic differential equations, in: L. Rogers, D. Talay (Eds.), Numerical Methods in Finance, 232-244, Cambridge University Press, 1997.

[17] Clément, D., Kohatsu-Higa, A., Lamberton, D., A duality approach for the weak approximation of stochastic differential equations, *Ann. Appl. Probab.* **16** (2006) 1124-1154.

[18] Crisan, D., Discretizing the continuous-time filtering problem: order of convergence, The Oxford Handbook of Nonlinear Filtering, 572-597, Oxford University Press, 2011.

[19] Crisan, D., Ghazali, S., On the convergence rates of a general class of weak approximations of SDEs, Stochastic differential equations: theory and applications, 221-248, Interdiscip. Math. Sci., 2, World Sci. Publ., Hackensack, NJ, 2007.

[20] Crisan, D., Ortiz-Latorre, S., A KLV Filter, preprint, 2012.

[21] de Boor, C., A Practical Guide to Splines, Springer, 1978.

[22] Douset, A., Johansen, A.M., A tutorial on particle filtering and smoothing: fifteen years later, The Oxford Handbook of Nonlinear Filtering, 656-704, Oxford University Press, 2011.

[23] Feuersänger, C., *Sparse Grid Methods for Higher Dimensional Approximation*, Dissertation, Institut für Numerische Simulation, Universität Bonn, 2010.

[24] Fujiwara, T., Approximation of expectations of jump-diffusion processes (in Japanese), Master thesis, Univ. Tokyo, 2006.

[25] Fujiwara, T., Sixth order methods of Kusuoka approximation, *UTMS preprint*, 2006.

[26] Giles, M.B., Improved multilevel Monte Carlo convergence using the Milstein scheme, *Monte Carlo and Quasi-Monte Carlo Methods 2006* (2007) 343-358.

[27] Giles, M.B., Multilevel Monte Carlo path simulation, *Oper. Res.* **56** (2008) 607-617.

[28] Giles, M.B., Higham, D.J., Mao, X., Analyzing multi-level Monte Carlo for options with non-globally Lipschitz payoff, *Finance Stoch.* **13** (2009) 403-414.

[29] Gobet, E., Pagès, G., Pham, H., Printemps, J., Discretization and simulation of Zakai equation, *SIAM J. Numer. Anal.* **44** (2006) 2505-2538.

[30] Gyöngy, I., Krylov, N.V., Expansion of solutions of parameterized equations and acceleration of numerical methods, *Illinois J. Math.*, **50**, No. 2, Special Volume in Memory of Joseph Doob (2006) 473-514.

[31] Gordon, N.J., Salmond, D.J., Smith, A.F.M., Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proceedings F (Radar and Signal Processing), 140, 1993, 107-113.

[32] Gyurkó, L.G., Lyons, T., Efficient and practical implementations of cubature on Wiener space, D. Crisan(Eds.), Stochastic Analysis 2010, 73-111, Springer, 2010.

[33] Hagan, P., Kumar, D., Lesniewski, A., Woodward, D., Managing smile risk, *Wilmott magazine* **1** (2002) 84-108.

[34] Hamdi, S., Schiesser, W.E., Griffiths, G.W., Method of lines, Scholarpedia, 2(7):2859, 2007.

[35] Hull, J.C., White, A., Using Hull-White interest rate trees, *J. Derivatives* **4** (1996) 26-36.

[36] Hutzenthaler, M., Jentzen, A., Kloeden, P., Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz coefficients, *Proc. Roy. Soc. London A* **467** (2011) 1563-1576.

[37] Itô, K., Multiple Wiener integral, *Journal of the Mathematical Society of Japan*, **3** (1951) 157-169.

[38] Ikeda, N., Watanabe, S., *Stochastic Differential Equations and Diffusion Processes, Second edition*, North-Holland, Kodansha, 1989.

[39] Jacod, J., Kurtz, T.G., Méléard, S., Protter, P., The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincare Probab. Statist.* **41** (2005) 523-558.

[40] Jacod, J., Protter, P., Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** (1998) 267-307.

[41] Kallianpur, G., *Stochastic Filtering Theory*, Springer, 1980.

[42] Kallianpur, G., Striebel, C., Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors, *Ann. Math. Statist.* **39** (1968) 785-801.

[43] Kargin, V., Lattice option pricing by multidimensional interpolation, *Math. Finance* **15** (2005) 635-647.

[44] Kitagawa, G., Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, *Journal of Computational and Graphical Statistics* **5** (1996) 1-25.

[45] Kloeden, P.E., Platen, E., *Numerical Solution of Stochastic Differential Equations* Springer, 1992.

[46] Kohatsu-Higa, A., Ortiz-Latorre, S., Tankov, P., Optimal simulation schemes for Lévy driven stochastic differential equations, *Math. Comp.* (2013).

[47] Kohatsu-Higa, A., Tankov, P., Jump-adapted discretization schemes for Lévy-driven SDEs, *Stoch. Proc. Appl.* **120** (2010) 2258-2285.

[48] Kunita, H., On the representation of solutions of stochastic differential equations, J. Azéma, M. Yor (Eds), Seminar on Probability XIV, 118-141, Springer, 1980.

[49] Kunita, H., Stochastic differential equations and stochastic flows of diffeomorphisms, Lecture Notes in Math., 1097, Springer, Berlin, 1984.

[50] Kunita, H., Nonlinear filtering problems II. associated equations, The Oxford Handbook of Nonlinear Filtering, 55-94, Oxford University Press, 2011.

[51] Kunitomo, N., Takahashi, A., On Validity of the Asymptotic Expansion Approach in Contingent Claim Analysis, *Ann. Appl. Probab.* **13** (2003) 914-952.

[52] Kusuoka, S., Malliavin calculus revisited, *J. Math. Sci. Univ. Tokyo.* **10** (2003) 261-277.

[53] Kusuoka, S., Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus, *Adv. Math. Econ.* **6** (2004) 69-83.

[54] Kusuoka, S., Malliavin calculus and numerical analysis (in Japanese), lecture note, Probability Summer School, 2005.

[55] Kusuoka, S., Gaussian K-scheme, *UTMS Preprint Series*, 2009.

[56] Kusuoka, S., Stroock, D., The partial Malliavin calculus and its application to non-linear filtering, *Stochastics* **12** (1984) 83-142.

[57] Kusuoka, S., Stroock, D.W., Applications of Malliavin calculus, Part III, *J. Fac. Sci. Univ. Tokyo Sect. IA, Math.* **34** (1987) 391-442.

[58] Lax, P.D., Richtmyer, R.D., Survey of the stability of linear finite difference equations. *Comm. Pure Appl. Math.* **9** (1956) 267-293.

[59] Leentvaar, C.C.W., Oosterlee, C.W., On coordinate transformation and grid stretching for sparse grid pricing of basket options, *J. Comput. Appl. Math.* **222** (2008) 193-209.

[60] Litterer, C., Lyons, T., High order recombination and an application to cubature on Wiener space, *Ann. Appl. Probab.* **22** (2012) 1301-1327.

[61] Load, R., Koekkoek R., Van Dijk, D., A comparison of biased simulation schemes for stochastic volatility models, *Quantitative Finance* **10** (2010) 177-194.

[62] Longstaff, F.A., Schwartz, E.S., Valuing American options by simulation: A simple least-squares approach, *Rev. Financ. Stud.* **14** (2001) 113-147.

[63] Lyons, T., Victoir, N., Cubature on Wiener space, *Proc. R. Soc. Lond. Ser. A* **460** (2004) 169-198.

[64] Maruyama, G., Continuous Markov processes and stochastic equations, *Rendiconti del Circolo Matematico di Palerm* **4** (1955) 48-90.

[65] Milstein, G.N., Tretyakov, M.V., Monte Carlo methods for backward equations in nonlinear filtering, *Adv. in Appl. Probab.* **41** (2009) 63-100.

[66] Mordecki, E., Szepessy, A., Tempone, R., Zouraris, G.E., Adaptive wak approximation of diffusions with jumps, *SIAM J. Numer. Anal.* **46** (2008) 1732-1768.

[67] Ngo, H-L., Kohatsu-Higa, A., Weak approximations for SDE's driven by Lévy processes, Seminar on Stochastic Analysis, Random Fields and Application VII, 131-169, 2013.

[68] Ninomiya, M., Ninomiya, S., A new weak approximation scheme of stochastic differential equations by using the Runge-Kutta method, *Finance Stoch.* **13** (2009) 415-443.

[69] Ninomiya, S., Victoir, N., Weak approximation of stochastic differential equations and application to derivative pricing, *Appl. Math. Finance* **15** (2008) 107-121.

[70] Nualart, D., *Malliavin calculus and related topics*, Springer, Berlin, 2006.

[71] Nualart, D., Zakai, M., The partial Malliavin calculus, *Séminaire de Probabilités XXIII.* Lecture Notes in Math. 1372, 362-381, Springer, Berlin, 1989.

[72] Oshima, K., Teichmann, J., Veluscek, D., A new extrapolation method for weak approximation schemes with applications, *Ann. Appl. Probab.* **22** (2012) 1008-1045.

[73] Picard, J., Approximation of nonlinear filtering problems and order of convergence, Filtering and Control of Random Processes (Lecture Notes Control Inform. Sci. 61), Springer, Berlin, 1984, 219-236.

[74] Protter, P., *Stochastic Integration and Differential Equations, Second Edition*, Springer, 2005.

[75] Protter, P., Talay, D., The Euler scheme for Lévy driven stochastic differential equations, *Ann. Probab.* **25** (1997) 393-423.

[76] Reisinger, C., Wittum, G., Efficient hierarchical approximation of high-dimensional option pricing problems, *SIAM J. Sci. Comp.* **29** (2006) 440-458.

[77] Sato, K., *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, 1999.

[78] Seydel, R., *Tools for Computational Finance, 3rd edition*, Springer-Verlag, Berlin, 2002.

[79] Schroder, M., Computing the constant elasticity of variance option pricing formula, *J. of Finance* **211** (1989) 211-219.

[80] Shigekawa, I., *Stochastic Analysis*, Translations of Mathematical Monographs Vol. 224, American Mathematical Society, Providence, RI, 2004.

[81] Strang, G., Accurate partial difference methods I: linear Cauchy problems, *Arch. Rat. Mech. Anal.* **12** (1963) 392-402.

[82] Strang, G., On the construction and comparison of difference schemes, *SIAM J. Numer. Anal.* **5** (1968) 506-517.

[83] Stroock, D.W., Partial Differential Equations for Probabilists, Springer, 2008.

[84] Stroud, A.H., *Approximate calculation of multiple integrals*, Prentice-Hall, 1971.

[85] Suzuki, M., Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations, *Phys. Lett. A* **146** (1990) 319-323.

[86] Suzuki, M., General theory of higher-order decomposition of exponential operators and symplectic integrators, *Phys. Lett. A* **165** (1992) 387-395.

[87] Takahashi, A., Yamada, T., An asymptotic expansion with push-down of Malliavin weights, *SIAM J. Finan. Math.* **3** (2012) 95-136.

[88] Takahashi, A., Yoshida, N., Monte Carlo simulation with asymptotic method, *J. Japan Statist. Soc.* **35** (2005) 171-203.

[89] Talay, D., Efficient numerical schemes for the approximation of expectations of functionals of the solution of a SDE and applications, Filtering and Control of Random Processes (Lecture Notes Control Inform. Sci. 61), Springer, Berlin, 1984, 294-313.

[90] Talay, D., Tubaro, L., Expansion of the global error for numerical schemes solving stochastic differential equations, *Stochastic Analysis and Applications* **8** (1990) 94-120.

[91] Tanaka, H., Cubature formula on Wiener space from the viewpoint of splitting methods, *RIMS Kôkyûroku*, Vol. 1844 (2013) 50-59.

[92] Tanaka, H., Higher-order interpolated lattice schemes for multidimensional option pricing problems, *J. Comput. Appl. Math.* **255** (2014) 313-333.

[93] Tanaka, H., A new proof for the convergence of Picard's filter using partial Malliavin calculus, preprint, available at arXiv:1311.6090.

[94] Tanaka, H., Kohatsu-Higa, A., An operator approach for Markov chain weak approximations with an application to infinite activity Lévy driven SDEs, *Ann. Appl. Probab.* **19** (2009) 1026-1062.

[95] Tanaka, H., Yamada, T., Strong convergence for Euler-Maruyama and Milstein schemes with asymptotic method, accepted for publication in *International Journal of Theoretical and Applied Finance*.

[96] Thomas, J.W., *Numerical Partial Differential Equations: Finite Difference Methods*, Springer, 1995.

[97] Yan, B.L., The Euler scheme with irregular coefficients, *Ann. Probab.* **30** (2002) 1172-1194.

[98] Victoir, N., Asymmetric cubature formulae with few points in high dimension for symmetric measures, *SIAM J. Numer. Anal.* **42** (2005) 209-227.

[99] Victoir, N., *From Cubature to Rough Paths*, Ph.D thesis, University of Oxford, 2003.