

# **Towards High-order Methods for Stochastic Differential Equations with White Noise: A Spectral Approach**

by

Zhongqiang Zhang

A dissertation submitted in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy  
in the Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

May 2014

Abstract of “Towards High-order Methods for Stochastic Differential Equations with White Noise: A Spectral Approach” by Zhongqiang Zhang, Ph.D., Brown University, May 2014

We develop a recursive multistage Wiener chaos expansion method (WCE) and a recursive multistage stochastic collocation method (SCM) for numerical integration of linear stochastic advection-diffusion-reaction equations with multiplicative white noise. We show that both methods are comparably efficient in computing the first two moments of solutions for long time intervals, compared to a direct application of WCE and SCM while both methods are more efficient than the standard Monte Carlo method if high accuracy is required.

Both methods belong to Wong-Zakai approximation, where Brownian motion is truncated using its spectral expansion before any discretization in time and space. For computational convenience, WCE is associated with the Ito formulation of underlying equations and SCM is associated with the Stratonovich formulation.

We apply SCM using Smolyak’s sparse grid construction to obtain the shock location of the one-dimensional piston problem, which is modeled by stochastic Euler equations with multiplicative white noise. We show numerically that SCM is efficient for short time simulations and for small magnitudes of noises and quasi-Monte Carlo methods are efficient for moderate large-time simulations. We also illustrate the efficiency of SCM through error estimates for a linear model problem.

We further investigate the effect of a spectral approximation of Brownian motion, rather than a piecewise linear approximation, for both spatial and temporal noise. For spatial noise, we consider semilinear elliptic equations with additive noise and show that when the solution is smooth enough, the spectral approximation is superior to the piecewise linear approximation while both approximations are comparable when the solution is not smooth. For temporal noise, we use this spectral approach to design numerical schemes for stochastic delay differential equations under the Stratonovich formulation. We show that the spectral approach admits higher-order accuracy only for higher-order schemes.

Besides equations with coefficients of linear growth, we also consider stochastic ordinary differential equations with coefficients of polynomial growth. We formulate a basic relationship between local truncation error and global error of numerical methods for these equations, and apply this relationship for our explicit balanced scheme to obtain the convergence order.

© Copyright 2014 by Zhongqiang Zhang

This dissertation by Zhongqiang Zhang is accepted in its present form  
by the Division of Applied Mathematics as satisfying the  
dissertation requirement for the degree of Doctor of Philosophy.

Date \_\_\_\_\_  
George Em Karniadakis, Advisor

Date \_\_\_\_\_  
Boris Rozovskii, Co-advisor

Recommended to the Graduate Council

Date \_\_\_\_\_  
Michael V. Tretyakov, Reader

Date \_\_\_\_\_  
Marcus Sarkis, Reader

Approved by the Graduate Council

Date \_\_\_\_\_  
Peter M. Weber, Dean of the Graduate School

## Acknowledgments

I would like to thank my advisor Professor George Em Karniadakis and co-advisor Professor Boris L. Rozovskii for their guidance and support during my study at Brown University. Their passion and great vision on scientific research in applied mathematics have inspired me to explore my research field in depth. Professor George Em Karniadakis always reminds of the essence of applied mathematics and shares with me his rich knowledge and research experience, which have helped me appreciate applications beyond mathematics itself. Professor Boris L. Rozovskii introduced his works to me and has encouraged me to be more rigorous in my work. My gratitude to them is actually beyond any words.

It is my great fortune to collaborate with Professor Michael V. Tretyakov of University of Nottingham at England. Without valuable discussions with him and invaluable advice from him, I would not have this dissertation in current form, several chapters of which are papers we published together in the last few years. He carefully checked every point leaving nothing to chance, which greatly helped me improve my presentation in clarity and accuracy.

I express my sincere thanks to my committee members and readers, Professor Marcus Sarkis and Professor Michael V. Tretyakov, for their precious time in carefully reading my thesis and serving as committee members.

I would like to thank Professor Johnny Guzmán for discussing finite element methods for elliptic equations with me and Professor Hongjie Dong for his help on energy estimates of some parabolic equations with minimal regularity. I also thank Professor Chau-Hsing Su for his help on the stochastic piston problem.

My heart is full of gratitude to people at the Crunch group: Heyrim Cho, Minseok Choi, Mingge Deng, Huan Lei, Xuejin Li, Zhen Li, Yuhang Tang, Daniele Venturi, Alix Witthoft, Xiu Yang, Yue Yu, Summer Zheng and many others for their help and support in life and in research. I would also like to thank Xingjie Li at Brown University, Wanrong Cao at Southeast University, Guang Lin at

Pacific Northwest National Laboratory, and Xiaoliang Wan at Louisiana State University for many helpful discussions.

Many thanks are also due to Ms. Madeline Brewster, Ms. Camille O. Dickson, Ms. Lisa Eklund, Ms. Stephanie Han, Ms. Jean Radican and Ms. Laura Leddy for their timely help.

Finally, I have to express my heartfelt thanks to my parents and my sisters for their love and encouragement. I remain indebted to my wife, Fei Yu, for her emotional support and love. I am also grateful to my father- and mother-in-law for their love and support, especially during my hard times. This dissertation is dedicated to my deceased mother.

This work was supported by OSD/MURI grant FA9550-09-1-0613, NSF/DMS grant DMS-0915077, NSF/DMS grant DMS-1216437 and also by the Collaboratory on Mathematics for Mesoscopic Modeling of Materials (CM4) which is sponsored by DOE.

# Contents

<b>Acknowledgments</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Review of numerical methods for SPDEs . . . . .	2
1.1.1 Direct semi-discretization methods for parabolic SPDEs . . . . .	3
1.1.2 Wong-Zakai approximation for parabolic SPDEs . . . . .	6
1.1.3 Preprocessing methods for parabolic SPDEs . . . . .	8
1.1.4 Stochastic Burgers and Navier-Stokes equations . . . . .	11
1.1.5 Beyond parabolic SPDEs . . . . .	13
1.1.6 Stability and convergence of existing numerical methods . . . . .	14
1.1.7 Conclusion . . . . .	18
1.2 Approximation of Brownian motion . . . . .	18
1.2.1 Piecewise linear approximation . . . . .	18
1.2.2 Other approximations . . . . .	21
1.3 Integration methods in random space . . . . .	22
1.3.1 Monte Carlo method and its variants . . . . .	22
1.3.2 Wiener chaos expansion method . . . . .	23
1.3.3 Stochastic collocation method . . . . .	24
1.4 Objectives of this work . . . . .	25

<b>Part I: Temporal White Noise</b>	<b>29</b>
<b>2 Wiener chaos methods for linear stochastic advection-diffusion-reaction equations</b>	<b>29</b>
2.1 Introduction . . . . .	30
2.2 WCE of the SPDE solution . . . . .	31
2.3 Multistage WCE method . . . . .	33
2.4 Algorithm for computing moments . . . . .	39
2.5 Numerical tests in one dimension . . . . .	43
2.5.1 Test problems . . . . .	43
2.5.2 Application of WCE algorithms to the model problem . . . . .	44
2.5.3 Numerical results . . . . .	48
2.5.4 Comparison of the WCE algorithm and Monte Carlo-type algorithms . . . . .	50
2.6 Numerical tests with passive scalar equation . . . . .	55
2.7 Summary . . . . .	61
<b>3 Stochastic collocation methods for differential equations with white noise</b>	<b>63</b>
3.1 Introduction . . . . .	64
3.2 Sparse grid for weak integration of SDE . . . . .	67
3.2.1 Smolyak's sparse grid . . . . .	67
3.2.2 Weak-sense integration of SDE . . . . .	70
3.2.3 Illustrative examples . . . . .	75
3.3 Recursive collocation algorithm for linear SPDE . . . . .	84
3.4 Numerical experiments . . . . .	88
<b>4 Comparison between Wiener chaos methods and stochastic collocation methods</b>	<b>96</b>
4.1 Introduction . . . . .	97

4.2	Review of Wiener chaos and stochastic collocation . . . . .	99
4.2.1	Wiener chaos expansion (WCE) . . . . .	100
4.2.2	Stochastic collocation method (SCM) . . . . .	101
4.3	Error estimates . . . . .	103
4.4	Numerical results . . . . .	106
4.5	Proofs . . . . .	115
4.5.1	Proof of Theorem 4.3.1 . . . . .	115
4.5.2	Proof of Theorem 4.3.3 . . . . .	120
4.6	Appendix: derivation of Algorithm 4.2.1. . . . .	123
<b>5</b>	<b>Application example: stochastic collocation methods for 1D stochastic piston problem</b>	<b>125</b>
5.1	Introduction . . . . .	126
5.2	Theoretical Background . . . . .	128
5.2.1	Stochastic Euler equations . . . . .	130
5.3	Verification of the Stratonovich- and Ito-Euler equations . . . . .	132
5.3.1	A splitting method for stochastic Euler equations . . . . .	133
5.3.2	Stratonovich-Euler equations versus first-order perturbation analysis . . . . .	134
5.3.3	Stratonovich-Euler equations versus Ito-Euler equations . . . . .	135
5.4	The Stochastic Collocation Method . . . . .	137
5.5	Summary . . . . .	142
5.6	Appendix: a first-order model driven by different Gaussian processes . . . . .	145
	<b>Part II: Spatial White Noise</b>	<b>150</b>
<b>6</b>	<b>Semilinear Elliptic Equations with Additive Noise</b>	<b>150</b>
6.1	Introduction . . . . .	150

6.2	Main results . . . . .	153
6.2.1	Semidiscrete and fully discrete schemes . . . . .	154
6.2.2	Strong and weak convergence order . . . . .	155
6.2.3	Examples of other PDEs . . . . .	158
6.3	Numerical results . . . . .	160
6.4	Proofs . . . . .	163
6.4.1	Regularity of the solution to (6.1.1). . . . .	164
6.4.2	Proofs for strong convergence order . . . . .	166
6.4.3	Proofs for weak convergence order . . . . .	167
6.4.4	Proof of Theorem 6.2.9 . . . . .	172
6.4.5	Proof of Theorem 6.2.5 . . . . .	174
6.4.6	Proof of Theorem 6.2.7 . . . . .	175
<b>Part III: Stochastic Ordinary Differential Equations</b>		<b>180</b>
<b>7</b>	<b>A Fundamental Limit Theorem for SDEs with non-Lipschitz coefficients</b>	<b>180</b>
7.1	Introduction . . . . .	180
7.2	Fundamental theorem . . . . .	183
7.2.1	Discussion . . . . .	186
7.3	A balanced method . . . . .	188
7.4	Numerical examples . . . . .	194
7.5	Proof of the fundamental theorem . . . . .	201
7.6	Proof of Lemma 7.2.3 . . . . .	204
7.7	Proof of Lemma 7.3.2 . . . . .	206
<b>8</b>	<b>Numerical schemes for SDEs with time delay via Wong-Zakai approximation</b>	<b>210</b>
8.1	Introduction . . . . .	211

8.2	Numerical schemes for SDDEs . . . . .	213
8.2.1	Derivation of numerical schemes . . . . .	216
8.3	Numerical results . . . . .	222
8.4	Proofs . . . . .	229
<b>9</b>	<b>Conclusion</b>	<b>239</b>
9.1	Summary . . . . .	239
9.2	Future work . . . . .	242

# List of Tables

2.1	Performance of Algorithm 2.5.1 for Model (2.5.1). The parameters of the model (2.5.1) are $\sigma = 1$ , $\epsilon = 0$ , and the time $t = 10$ . In Algorithm 2.5.1 we take $M = 20$ . . .	49
2.2	<i>Model (2.5.1): performance of Algorithm 2.5.1.</i> The parameters of the model (2.5.1) are $\sigma = 1$ , $\epsilon = 0.01$ , and the time $t = 10$ . In Algorithm 2.5.1 we take $M = 20$ . . . .	49
2.3	Performance of Algorithm 2.5.1 for Model (2.5.3). The parameters of the model (2.5.3) are $\sigma = 1$ , $a = 0.5$ , and the time $t = 10$ . In Algorithm 2.5.1 we take $M = 20$ . .	50
2.4	Performance of the method (2.5.14) for Model (2.5.1). The parameters of the model (2.5.1) are $\sigma = 1$ , $\epsilon = 0$ , and the time $t = 10$ . The statistical error is computed according to (2.5.15). . . . .	53
2.5	Performance of Algorithm 2.5.1 for Model (2.5.1). The parameters of the model (2.5.1) are $\sigma = 1$ , $\epsilon = 0$ , and the time $t = 10$ . The parameters of Algorithm 2.5.1 are $\Delta = 0.1$ , $M = 20$ , $\delta t = 0.001$ . . . . .	53
2.6	<i>Model (2.5.1): performance of the method (2.5.21).</i> The parameters of the model (2.5.1) are $\sigma = 1$ , $\epsilon = 0$ , and the time $t = 10$ . . . . .	55
2.7	<i>Performance of Algorithm 2.4.1 for passive scalar equation (2.6.4).</i> The parameters of Algorithm 2.4.1 are $N = 2$ , $n = 1$ , $M = 900$ , $\delta t = 1 \times 10^{-4}$ . . . . .	59
2.8	Performance of Algorithm 2.4.1 passive scalar equation (2.6.4). The parameters of Algorithm 2.4.1 are $N = 2$ , $n = 1$ , $M = 900$ , $\delta t = 1 \times 10^{-4}$ . . . . .	61

2.9	Performance of Algorithm 2.6.11 for passive scalar equation (2.6.4). The parameter is $M = 100$ . . . . .	61
3.1	The number of sparse grid points for the sparse grid quadrature (3.2.4) using the one-dimensional Gauss-Hermite quadrature rule (3.2.2), when the sparse grid level $L \leq d$ . . . . .	69
3.2	Comparison of the SGC algorithms based on the Euler scheme (left) and on the second-order scheme (3.2.20) (right). The parameters of the model (3.4.1) are $x_0 = 0.1$ , $\theta_1 = 1$ , $\theta_2 = 0.3$ , and $T = 1$ . . . . .	90
3.3	Comparison of the SGC algorithms based on the Euler scheme (left) and on the second-order scheme (3.2.20) (right). The parameters of the model (3.4.1) are $x_0 = 0.08$ , $\theta_1 = -1$ , $\theta_2 = 2$ , and $T = 1$ . The sparse grid level $L = 4$ . . . . .	91
3.4	Errors of the SGC algorithm to the stochastic Burgers equation (3.4.3) with parameters $T = 0.5$ , $\nu = 0.1$ and $\sigma = 1$ . . . . .	93
3.5	Errors of the SGC algorithm applied to the stochastic Burgers equation (3.4.3) with parameters $\nu = 1$ , $\sigma = 0.5$ , and $T = 0.5$ . . . . .	94
3.6	Errors in computing the second moment of the solution to the stochastic advection-diffusion equation (3.4.6) with $\sigma = 0.5$ , $\beta = 0.1$ , $\epsilon = 0.2$ at $T = 5$ by Algorithm 3.3.1 with $l_* = 20$ and the one-dimensional Gauss-Hermite quadrature of order $n = 2$ (left) and $n = 3$ (right). . . . .	95
4.1	Algorithm 2.4.1: recursive multi-stage Wiener chaos method for (4.4.1) at $T = 5$ : $\sigma = 0.5$ , $\beta = 0.1$ , $\epsilon = 0.02$ , and $M = 20$ , $n = 1$ . . . . .	107
4.2	Algorithm 4.2.1: recursive multi-stage stochastic collocation method for (4.4.1) at $T = 5$ : $\sigma = 0.5$ , $\beta = 0.1$ , $\epsilon = 0.02$ , and $M = 20$ , $n = 1$ . . . . .	108

4.3	Algorithm 2.4.1: recursive multi-stage Wiener chaos expansion for commutative noises (4.4.4) at $T = 1$ : $\sigma_1 = 0.5$ , $\sigma_2 = 0.2$ , $\beta = 0.1$ , $\epsilon = 0.02$ , and $M = 30$ , $n = 1$ . . . . .	110
4.4	Algorithm 4.2.1: recursive multi-stage stochastic collocation method for commutative noises (4.4.4) at $T = 1$ : $\sigma_1 = 0.5$ , $\sigma_2 = 0.2$ , $\beta = 0.1$ , $\epsilon = 0.02$ , and $M = 30$ , $n = 1$ . . . . .	111
4.5	Algorithm 2.4.1 (recursive multi-stage Wiener chaos expansion, left) and Algorithm 4.2.1 (recursive multi-stage stochastic collocation method, right) for (4.4.8) at $T = 1$ : $\sigma_1 = 0.5$ , $\sigma_2 = 0.2$ , $\beta = 0.1$ , $\epsilon = 0.02$ , and $M = 20$ , $n = 1$ . The time step size $\delta t$ is $\Delta/10$ . The reported CPU time is in seconds. . . . .	112
4.6	Algorithm 2.4.1: recursive multi-stage Wiener chaos expansion for (4.4.8) at $T = 1$ : $\sigma_1 = 0.5$ , $\sigma_2 = 0.2$ , $\beta = 0.1$ , $\epsilon = 0.02$ . The parameters are $M = 20$ , $N = 2$ , and $n = 10$ . The time step size $\delta t$ is $\Delta/10$ . . . . .	114
5.1	The difference of variances of shock location by Stratonovich-Euler and Ito-Euler equations for $\epsilon = 0.5$ . . . . .	137
6.1	Convergence of piecewise linear finite element methods for the one-dimensional semilinear problem (6.3.1) with a spectral approximation of white noise (6.2.7). . . . .	162
6.2	Convergence of piecewise linear finite element approximation of the two-dimensional semilinear problem (6.3.3) with a spectral approximation of white noise (6.2.7). . . . .	163
7.1	<i>Example 7.4.1.</i> Mean-square errors of the selected schemes. See further details in the text. . . . .	198
7.2	<i>Example 7.4.2.</i> Mean-square errors of the selected schemes. See further details in the text. . . . .	200

7.3	<i>Example 7.4.2.</i> Computational times for the selected schemes. See further details in the text. . . . .	201
8.1	Convergence rate of Predictor-Corrector scheme (left) and Midpoint scheme (right) for Example 8.3.1 at $T = 20$ with ten white noises using $n_p = 4000$ sample paths. . .	224
8.2	Convergence rate of the Milstein-like scheme (left) for Equation (8.3.4) at $T = 1$ and comparison with the convergence rate of the Predictor-Corrector scheme (middle) and the Midpoint scheme (right) using $n_p = 4000$ sample paths. The upper rows are with $\tau = 1/16$ and the lower are with $\tau = 1/4$ . . . . .	227
8.3	Convergence rate of the Milstein-like scheme (left) for Equation (8.3.5) (single white noise) at $T = 1$ and comparison with the convergence rate of the Predictor-Corrector scheme (middle) and Midpoint scheme (right) using $n_p = 4000$ sample paths. The delay $\tau$ is taken as $1/4$ . . . . .	228

# List of Figures

1.1	Conceptual and chapter overview of this thesis . . . . .	2
2.1	Illustration of the idea of multistage WCE. The dotted line denotes the “offline” computation, where we solve the propagator up to time $\Delta$ . The dash line implies that one solves only the solution on certain time levels instead of on the entire time interval. . . . .	37
2.2	Dependence of the relative numerical error $\rho_2^r(t)$ on integration time. Model (2.5.1) is simulated by Algorithm 2.5.1 with $M = 20$ and $\delta t = \Delta/100$ and various $\Delta$ and $N$ . The parameters of (2.5.1) are $\sigma = 1$ and $\epsilon = 0$ . . . . .	51
5.1	A sketch of piston-driven shock tube with random piston motion. . . . .	128
5.2	Comparison between the results from first-order perturbation analysis (5.2.2) and solving the Stratonovich-Euler equations (5.2.8) by the splitting method (5.3.1)-(5.3.3). 136	
5.3	Comparison between solving Stratonovich-Euler equations (5.2.8) and Ito-Euler equations (5.2.9) by the splitting method (5.3.1)-(5.3.2). . . . .	138
5.4	Comparison between numerical results from Stratonovich-Euler equations (5.2.8) using the direct Monte Carlo method (5.3.1) and (5.3.3) and the stochastic collocation method (5.4.1). The sparse grid level is 2 and $\Delta$ is the size of element in time in the stochastic collocation method. . . . .	141

5.5	Comparison between numerical results from Stratonovich-Euler equations (5.2.8) using direct Monte Carlo method (5.3.1)-(5.3.3) and the QMC method for (5.4.1) with a large noise: $\epsilon = 0.5$ . . . . .	143
5.6	Comparison among variances of shock positions induced by three different Gaussian noises: Brownian motion, random process with zero mean and exponential covariance kernel $\exp(- t_1 - t_2 )$ and standard Gaussian random variable. The noise amplitude is $\epsilon = 0.1$ . . . . .	144
5.7	Comparison of variances of the solutions for four models of $k(t)$ up to time $t = 1$ : $A = 1$ . . . . .	147
5.8	Comparison of variances of the solutions for four models of $k(t)$ at large time: $A = 1$ . . . . .	148
7.1	Example 7.4.2. Trajectories of the fully-tamed scheme (7.4.2) and the balanced scheme (7.3.1) for $h = 0.1$ . The reference trajectory is simulated by the mid-point scheme (see (7.4.4) with $\lambda = 1/2$ ) using a small time step. . . . .	199
8.1	Mean-square convergence test of the Predictor-Corrector (left column) and Midpoint schemes (right column) on Example (8.3.2) at $T=5$ with different $\tau$ using $n_p = 10000$ sample paths. (a): multi white noises with non-delayed diffusion term; (b): single white noise with non-delayed diffusion term; (c): single white noise with delayed diffusion term. . . . .	226

# Chapter 1

## Introduction

Stochastic mathematical models have attracted increasing attention for their capacity of representing intrinsic uncertainty in complex systems, e.g., capturing various scales as in particle simulations at mesoscopic scale, and extrinsic uncertainty, e.g., stochastic external forces, stochastic initial condition or stochastic boundary conditions.

One important class of stochastic mathematical models is stochastic partial differential equations (SPDEs), which can be seen as deterministic partial differential equations (PDEs) with finite or infinite dimensional random processes—either with color noise or white noise. Though white noise is a purely mathematical construction, it can be a good model for rapid random fluctuations and also it is a limit of color noise when the correlation length goes to zero.

SPDEs with white noise have been derived from various applications, such as nonlinear filtering (see e.g. [415]), turbulent flows (see e.g. [40, 292]), fluid flows in random media (see e.g. [190]), particle systems (see e.g. [228]), population biology (see e.g. [86]), neuroscience (see e.g. [389]), etc.

Since analytic solutions to SPDEs can rarely be obtained, numerical methods are adopted to solve SPDEs. One of the motivations for numerical SPDEs in early literature was to solve the Zakai equation of nonlinear filtering, see e.g. [29, 68, 112, 127, 128, 129]. In the next section,

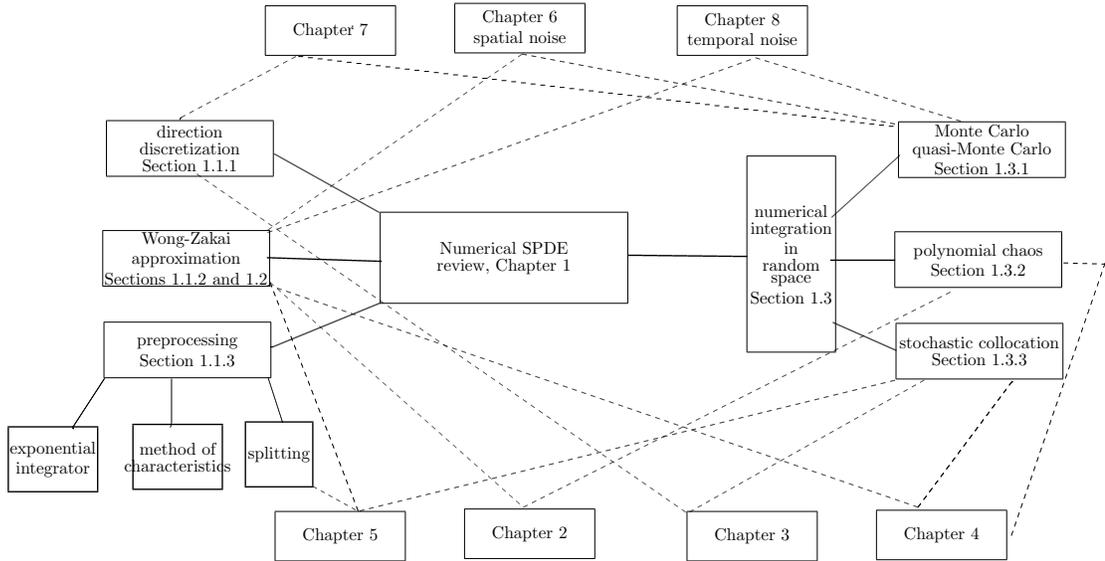


Figure 1.1: Conceptual and chapter overview of this thesis

we review some numerical methods for semilinear equation (1.1.1), advection-diffusion-reaction equation of nonlinear filtering (1.1.6), stochastic Burgers equation (1.1.13) and stochastic Navier-Stokes equation (1.1.16).

The rest of this chapter is organized as follows. In Section 1.1, we review several prevalent numerical methods for SPDEs and address the convergence and stability of these methods. We then survey one of these numerical methods, the Wong-Zakai approximation in Section 1.2. Before we present the objectives of this work, we briefly address the numerical integration methods in random space for obtaining statistics of solutions in Section 1.3. In Figure 1.1, we sketch how we organize this chapter and the overall organization of this thesis.

## 1.1 Review of numerical methods for SPDEs

In this section, we briefly review numerical methods for SPDEs and classify the numerical methods in literature into three types: *direct semi-discretization methods*, *Wong-Zakai approximation*, and *preprocessing methods*. In the first approach, we usually discretize the underlying SPDEs in

time and/or in space, applying classical techniques from time-discretization methods of stochastic ordinary differential equations (SODEs) and/or from spatial discretization methods of partial differential equations (PDEs). In the second approach, we first discretize the space-time noise before any discretization in time and space and thus we need further spatial-temporal discretizations. In the third approach, we first transform the underlying SPDE into some equivalent form before we discretize the SPDEs.

We start from considering the following SPDE over the physical domain  $\mathcal{D} \subseteq \mathbb{R}^d$ ,

$$dX = [\mathcal{A}X + f(X)] dt + g(X) dW^Q, \quad (1.1.1)$$

where  $W^Q$  is a  $Q$ -Wiener process:

$$W^Q(t, x) = \sum_{i \in \mathbb{N}^d} \sqrt{q_i} e_i(x) w_i(t), \quad (1.1.2)$$

Here  $q_i \geq 0$ ,  $i \in \mathbb{N}^d$  and  $\{e_i(x)\}$  is an orthonormal basis in  $L^2(\mathcal{D})$ . When  $q_i = 1$  for all  $i$ , we have the space-time white noise. When  $\sum_{i=1}^{\infty} q_i < \infty$ , we call it the space-time colored noise. We call the noise finite-dimensional when  $q_i = 0$  for all sufficient large  $i$ . The physical space is one-dimensional, i.e.  $d = 1$ , unless otherwise stated.

The leading operator  $\mathcal{A}$  can be second-order or fourth-order differential operators, which are usually generators of analytic semigroups. The nonlinear function  $f, g$  are usually Lipschitz continuous. The problem (1.1.1) is endowed either with only initial conditions in the whole space ( $\mathcal{D} = \mathbb{R}^d$ ) or with initial and boundary conditions in a bounded domain ( $\mathcal{D} \subsetneq \mathbb{R}^d$ ).

### 1.1.1 Direct semi-discretization methods for parabolic SPDEs

In this approach, we first review the *time-discretization* methods for (1.1.1), which can be seen as a straightforward application of numerical methods for SODEs, where increments of Brownian

motions are used. After truncating in physical space, we will have a system of finite dimensional SODEs and then we can apply standard numerical methods for SODEs, e.g. those from [218, 297, 301].

### Second-order equations

For finite dimensional noise, we can directly apply those time-discretization methods for SODEs to SPDEs as solutions are usually smooth in space. Ref. [141] considered Euler and other explicit schemes for scalar Wiener process and Ref. [220] further considered linear-implicit schemes in time under the same problem setting. Ref. [118] proposed the Milstein scheme for KPP equations with multiplicative noise using finite difference scheme in space. See [136, 309, 343, 355, 359] for more numerical results.

For infinite dimensional noise but with fast decaying  $q_i$ , Ref. [180] considered the mean-square convergence of linear-implicit and explicit Euler scheme, and Crank-Nicolson scheme in time for (1.1.1) with certain smooth  $f$  and  $g$  and proved half-order convergence for these schemes. The author remarked that for Crank-Nicolson scheme the convergence order can be improved to one for linear equations with additive noise, as in the case of SODEs. Ref. [181] proved the first-order weak convergence of these numerical schemes for (1.1.1) with additive noise. Ref. [293] considered (1.1.1) with space-time noise where  $q_i$  and  $e_i$  are eigenvalues and eigenfunctions of a specific isotropic kernel. Hausenblas [179] considered a slightly different equation

$$dX = [AX + f(t, X)] dt + \sum_j g_j(t, X) dw_j(t), \quad (1.1.3)$$

where  $\sum_j \|g_j(t)\|_2^2 < \infty$  and some boundedness of  $f$  and  $g$  is imposed. Here  $\|\cdot\|_2$  is the Sobolev-Hilbert second-order norm. The author proved half-order convergence in time for the linear-implicit and explicit Euler schemes and the Crank-Nicolson scheme.

However, if space-time white noise is considered, the convergence order in time is expected to be

less than  $1/4$ . In fact, the sample paths of the solution to heat equation with additive noise is Hölder continuous with exponent  $1/4 - \epsilon$  in time ( $\epsilon > 0$  is arbitrarily small), and thus the optimal order of convergence in time is  $1/4 - \epsilon$  if only increments of Brownian motion (with equi-spaced time steps) are used, see e.g. [5, 84] for the case of linear equations. Gyongy and Nualart introduced an implicit numerical scheme in time for the SPDE (1.1.1) with additive noise and proved convergence in probability in time without order in [167] and for (1.1.1) with mean-square order  $1/8 - \epsilon$  in time [168]. Gyongy [157, 159] also applied finite differences in space to the SPDE (1.1.1) and then used several temporal implicit and explicit schemes, including the linear-implicit Euler scheme. The author showed that these schemes converge with order  $1/2 - \epsilon$  in the space and with order  $1/4 - \epsilon$  in time for multiplicative noise with Lipschitz nonlinear terms similar to the linear equations in [5, 84].

Refs. [315, 316] proposed an implicit Euler scheme on non-uniform time grid for (1.1.1) with  $f = 0$  to reduce the computational cost where they provided upper bound [315] and lower bound [316] of the mean-square errors in terms of the computational cost.

As we mentioned before, *the solution to (1.1.1) is of low regularity and thus it is not possible to derive high-order schemes with direct time discretization methods*. See e.g. [390] for discussion on first-order schemes (Milstein type schemes) for (1.1.1) and also [209] for a review of numerical approximation of (1.1.1) along this line.

For *spatial semi-discretization* methods for solving SPDEs (including but not limited to (1.1.1)), see finite difference methods, see e.g. [5, 353, 412, 263]; finite element methods, see e.g. [5, 20, 129, 390, 393, 408, 411]; finite volume methods for hyperbolic problems, see e.g. [234, 311]; spectral methods, see e.g. [56, 68, 203, 254]). See also [164, 176, 177, 126] for acceleration schemes in space using Richardson's extrapolation method.

## Fourth-order equations

Now we consider fourth-order equations, i.e.  $\mathcal{A}$  is a fourth-order differential operator, which have been investigated in [225, 226, 227, 231, 248] etc. As the kernels associated with fourth-order operators can have more smoothing effects than those associated with second-order differential operator, we can expect better convergence in space and also in time.

Ref. [225] considered fully-discrete finite element approximations for a fourth-order linear stochastic parabolic equation with additive space-time white noise in one space dimension where strong convergence with order  $3/8$  in time and  $3/2 - \epsilon$  in space was proved.

Ref. [231] proved the convergence of finite element approximation of the nonlinear stochastic Cahn-Hilliard-Cook equation by additive space-time color noise

$$dX = \Delta^2 X + \Delta f(X) + dW^Q. \quad (1.1.4)$$

Ref. [62] presented some numerical results of a semi-implicit backward differentiation formula in time for nonlinear Cahn-Hilliard equation while no convergence analysis is given.

For the linearized Cahn-Hilliard-Cook equation ( $f = 0$ ) with additive space-time colored noise, Ref. [248] applied a standard finite element method and an implicit Euler scheme in time and obtained quasi-optimal convergence order in space. Kossioris and Zouris considered an implicit Euler scheme in time and finite elements in space for the linear Cahn-Hilliard equation with additive space-time white noise in [227] and the same equation but with even rougher noise which is the first-order spatial derivative of the space-time white noise in [226]. In [227], they proved that the strong convergence order is  $(4 - d)/8$  in time and  $(4 - d)/2 - \epsilon$  in space for  $d = 2, 3$ .

### 1.1.2 Wong-Zakai approximation for parabolic SPDEs

In this approach, we first truncate the Brownian motion with a smooth process of bounded variation yielding a PDE with finite dimensional noise. Thus, after truncating Brownian motion, we have to

discretize both in time and in space to obtain fully discrete schemes.

The most popular approximation Brownian motion in this approach is piecewise linear approximation of Brownian motion (also known as polygonal approximation [399])

$$W^{(n)}(t) = W(\mathbf{t}_i) + (W(\mathbf{t}_{i+1}) - W(\mathbf{t}_i)) \frac{t - \mathbf{t}_i}{\mathbf{t}_{i+1} - \mathbf{t}_i}, \quad t \in [\mathbf{t}_i, \mathbf{t}_{i+1}). \quad (1.1.5)$$

Piecewise linear approximation for SPDEs has been well studied in theory, see e.g. [151, 194, 371, 382, 383, 385, 121] (for mean-square convergence), [42, 152, 171, 172] (for pathwise convergence), [16, 28, 63, 69, 153, 154, 155, 170, 156, 294, 384] (for support theorem, the relation between the support of distribution of the solution and that of its Wong-Zakai approximation). For mean-square convergence of (1.1.6) with  $\mathcal{M}_k$  having no differential operator, Ref. [194] proved a half-order convergence, see also [49, 50]. For pathwise convergence, Ref. [171] proved a  $1/4 - \epsilon$ -order convergence and Ref. [172] proved a  $1/2 - \epsilon$ -order convergence when  $\mathcal{M}_k$  is a first-order differential operator.

All the aforementioned papers were on the convergence of the Wong-Zakai approximation itself, i.e., without any further discretization of the resulting PDEs. Numerical simulations haven't been well explored for SPDEs. Even for SODEs, Ref. [258] seems to be the first attempt to obtain numerical solutions from Wong-Zakai approximation, where the authors considered a stiff ODE (abbreviated for ordinary differential equations) solver instead of presenting new discretization schemes.

In this work, we will derive fully discrete schemes based on Wong-Zakai approximations and show the relationships between the derived schemes and the classical schemes (e.g. those in [218, 297, 301]); see Chapter 8 for details.

We will briefly review the literature on Wong-Zakai approximation in Section 1.2, especially on different types of approximation of Brownian motion and their applications to SPDEs and also SODEs.

### 1.1.3 Preprocessing methods for parabolic SPDEs

In this type of methods, the underlying equation is first transformed into an equivalent form, which may bring some benefits in computation, and then is dealt with time discretization techniques. For example, splitting techniques split the underlying equation into stochastic part and deterministic part and save computational cost if either part can be easily solved or even explicitly solved. In the splitting methods, we also have the freedom to use different schemes for different parts.

We will only review two methods in this class: splitting techniques and exponential integrator methods. In addition to these two methods, there are other preprocessing methods such as methods of averaging-over-characteristics, e.g. [304, 336, 360]; particle methods, e.g. [76, 77, 78, 79, 80, 243]; algebraic method, e.g. [344]; filtering on space-time noise [260]; etc.

#### Splitting methods

Splitting methods are also known as fractional step methods, see e.g. [138], and sometimes as predictor-corrector methods see e.g. [113]. They have been widely used for their computational convenience, e.g. [31, 74, 75, 160, 162, 163, 205, 249, 256, 255]. Mostly, the splitting is formulated by the following Lie-Trotter splitting, which splits the underlying problem, say (1.1.6), into two parts: ‘stochastic part’ (1.1.7a) and ‘deterministic part’ (1.1.7b). Consider the following Cauchy problem (see e.g. [113, 161, 162])

$$du(t, x) = \mathcal{L}u(t, x) dt + \sum_{k=1}^{d_1} \mathcal{M}_k u(t, x) \circ dw_k, \quad (t, x) \in (0, T] \times \mathcal{D}, \quad (1.1.6)$$

where  $\mathcal{L}$  is linear second-order differential operator,  $\mathcal{M}_k$  is linear differential operator up to first order, and  $\mathcal{D}$  is the whole space  $\mathbb{R}^d$ . The typical Lie-Trotter splitting scheme for (1.1.6) reads, over the time interval  $(t_n, t_{n+1}]$ , in integral form

$$\tilde{u}_n(t, x) = u_n(t_n, x) + \int_{t_n}^t \sum_{k=1}^{d_1} \mathcal{M}_k \tilde{u}_n(s, x) \circ dw_k(s), \quad t \in (t_n, t_{n+1}], \quad (1.1.7a)$$

$$u_n(t, x) = \tilde{u}_n(t_{n+1}) + \int_{t_n}^t \mathcal{L}u_n(s, x) ds, \quad t \in (t_n, t_{n+1}]. \quad (1.1.7b)$$

When  $\mathcal{M}_k$  is a zeroth order differential operator, Ref. [363] presented for pathwise convergence with half-order in time under  $L^2$ -norm in space when  $d_1 = 1$ . Under similar settings in [363], Ref. [205] proved that a normalization of numerical density in Zakai equation in a splitting scheme is equivalent to solving the Kushner equation (nonlinear SPDE for the normalized density, see e.g. [244]) by a similar splitting scheme (first order in the mean-square sense).

When  $\mathcal{M}_k$  is a first-order differential operator, Ref. [113] proved half-order mean-square convergence in time under the  $L^2$ -norm in space. Gyöngy and Krylov managed to provide the first-order mean-square convergence in time under higher-order Sobolev-Hilbert norms [162], and under even stronger norm in space [161].

Other than finite dimensional noise, Refs. [30, 31] considered semilinear parabolic equations (1.1.1) with multiplicative space-time color noises. With the Lie-Trotter splitting, they established strong convergence of the splitting scheme and proved half-order mean-square convergence in time. [74] obtained mean-square and pathwise convergence order of Lie-Trotter splitting methods for Cauchy problems of linear stochastic parabolic equations with additive space-time noise.

Other than the problems (1.1.1) and (1.1.6), the Lie-Trotter splitting techniques have been applied to different problems, such as stochastic hyperbolic equations (e.g. [6, 25, 349]), rough partial differential equations (e.g. [117]), stochastic Schrödinger equation (e.g. [43, 142, 255, 256, 285]), etc.

### **Integration-factor (exponential integrator) techniques**

In this approach, we first write the underlying SPDE in mild form (integration-factor) and then combine different time-discretization methods to derive fully discrete schemes. It was first proposed in [259, 313], under the name of exponential Euler scheme and was further developed to derive higher-order scheme, see e.g. [27, 206, 207, 208, 209, 210, 211].

In this approach, it is possible to derive high-order schemes in the strong sense since we may incorporate dynamics the underlying problems as shown for ODEs with smooth random inputs in [216]. By formulating Equation (1.1.1) with additive noise in mild form, we have

$$X(t) = e^{At} X_0 + \int_0^t e^{A(t-s)} f(X(s)) ds + \int_0^t e^{A(t-s)} dW^Q(s), \quad (1.1.8)$$

then we can derive an exponential Euler scheme [259, 313]:

$$X_{k+1} = e^{Ah} [X_k + hf(X_k) + W^Q(t_{k+1}) - W^Q(t_k)]. \quad (1.1.9)$$

or as in [210, 313]

$$X_{k+1} = e^{Ah} X_k + \mathcal{A}^{-1}(e^{Ah} - I)f(X_k) + \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-s)} dW^Q(s), \quad (1.1.10)$$

where  $t_k = kh$ ,  $k = 0, \dots, N$ ,  $Nh = T$ .

In certain cases, the total computational cost for the exponential Euler scheme can be reduced when  $\eta_k = \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-s)} dW^Q(s)$  is simulated as a whole instead of using increments of Brownian motion. For example, when  $\mathcal{A}e_i = -\lambda_i e_i$ , noticing that  $\eta_k$  solves

$$Y = \sum_{i=1}^{\infty} \int_{t_k}^{t_{k+1}} \mathcal{A}Y ds + \sum_{i=1}^{\infty} \int_{t_k}^{t_{k+1}} \sqrt{q_i} e_i dw_i(s), \quad (1.1.11)$$

and thus  $\eta_k$  can be represented by

$$\eta_k = \sum_{i=1}^{\infty} \sqrt{\gamma_i} e_i(x) \xi_{k,i}, \quad \xi_{k,i} = \frac{1}{\sqrt{\gamma_i}} \int_{t_k}^{t_{k+1}} e^{\lambda_i(t_{k+1}-s)} dw_i(s), \quad \gamma_i = \frac{q_i}{2\lambda_i} (1 - \exp(2\lambda_i h)). \quad (1.1.12)$$

In this way, we incorporate the interaction between the dynamics and the noise and thus we can have first-order mean-square convergence [209, 210]. See [208, 217, 261, 305] for further discussion

on additive noise.

For multiplicative noise, a first-order scheme (Milstein scheme) has been derived under this approach [211], where commutative conditions on diffusion coefficients for equations with infinite dimension noises were identified and a one-and-a-half order scheme in the mean-square sense has been derived in [27]. See also [3, 19, 21, 235, 262, 395] for further discussion on exponential integration schemes for SPDEs with multiplicative noises.

#### 1.1.4 Stochastic Burgers and Navier-Stokes equations

As a special class of parabolic SPDEs, stochastic Burgers and Navier-Stokes equations require more attention for their strong interactions between the strong nonlinearity and the noises. Similar to linear heat equation with additive noise, the convergence for time-discretization of one-dimensional Burgers equations is no more than 1/4, see [339] for multiplicative space-time noise with convergence in probability, and [36] for additive space-time noise with pathwise convergence. The convergence in space is less than 1/4, see [4] for additive space-time white noise with pathwise convergence, and [35] for additive space-time color noise with pathwise convergence.

Because of the strong nonlinearity, the discretization in space and in time may cause some effects, such as “a spatial version of the Ito-Stratonovich correction” [174, 175]. Hairer et al considered finite difference schemes for the Burgers equation with additive space-time noise in [175]:

$$\partial_t u = \nu \partial_x^2 u + (\nabla G(u)) \partial_x u + \sigma \dot{W}^Q, \quad x \in [0, 2\pi]. \quad (1.1.13)$$

If we only consider the discretization of the first-order differential operator, e.g.,

$$\partial_t u^\varepsilon = \nu \partial_x^2 u^\varepsilon + (\nabla G(u^\varepsilon)) \partial_x u^\varepsilon + \sigma \dot{W}^Q, \quad \partial_x u^\varepsilon = \frac{u(x + a\varepsilon) - u(x - b\varepsilon)}{(a + b)\varepsilon}, \quad (1.1.14)$$

then we can prove that this equation converges to (see [174])

$$\partial_t v = \nu \partial_x^2 v + (\nabla G(v)) \partial_x v - \frac{\sigma^2}{4\nu} \frac{a-b}{a+b} \Delta G(v) + \sigma \dot{W}^Q, \quad x \in [0, 2\pi] \quad (1.1.15)$$

if  $\dot{W}^Q$  is space-time white noise, and no correction term if  $\dot{W}^Q$  is more regular than space-time white noise, e.g. white in time but correlated in space. Effects of some other standard discretizations in space, e.g. Galerkin methods, and fully discretizations were also discussed in [174].

The stochastic incompressible Navier-Stokes is

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} - \nu \Delta \mathbf{u} + \nabla p = \sigma(\mathbf{u}) \dot{W}^Q, \quad \operatorname{div} \mathbf{u} = 0, \quad (1.1.16)$$

where  $\sigma$  is Lipschitz continuous. When  $\mathbb{E}[W^Q(x, t)W^Q(y, s)] = q(x, y) \min(s, t)$  and  $q(x, x)$  is square-integrable over the physical domain, Ref. [44] showed the existence and strong convergence of the solutions for the full discrete schemes in two-dimensional case. Ref. [60] considered three semi-implicit Euler schemes in time and standard finite elements methods for two-dimensional (1.1.16) with periodic boundary conditions. They presented the solution convergence in probability with order 1/4 in time similar to one-dimensional stochastic Burgers equation with additive noise. They also showed that for the corresponding Stokes problem, the fully discrete scheme converges in the strong sense with order half in time and order one in physical space.

For (1.1.16) in the bounded domain with Dirichlet boundary condition, Ref. [418] considered the backward Euler scheme and proved half-order strong convergence when the multiplicative noise is space-time color noise. Ref. [410] considered an implicit-explicit scheme and proved a convergence order depending on the regularity index of initial condition. Ref. [106] considered finite elements methods and a semi-implicit Euler for stochastic Navier-Stokes equation (1.1.16) and Ref. [107] considered similar fully discrete schemes for stochastic Navier-Stokes introduced in [292]. Ref. [409] provided *a posteriori* error estimates for stochastic Navier-Stokes equation.

See [38] ( recursive approximation), [111] (implicit scheme), [143] (Wong-Zakai approximation), [351, 413] (Galerkin approximation), [108] (Wiener chaos expansion) for more discussion on numerical methods and e.g. [95] for existence and uniqueness of (1.1.16). See also [147] for strong convergence of Fourier Galerkin methods for the hyperviscous Burgers equation and some numerical results for stochastic Burgers equation equipped with Wick product [374].

### 1.1.5 Beyond parabolic SPDEs

Compared to parabolic equations, *stochastic wave equations* of second order can have better smoothing in time: the solutions are Hölder continuous with exponent  $1/2 - \epsilon$  in time, and thus the optimal order of convergence in time is half if only increments of Brownian motion is used, see [391] for the one-dimensional wave equation with multiplicative noise.

Ref. [8] considered linear wave equation with additive single white noise in time using integration factor techniques, where the convergence of two-step finite difference schemes in time is of first-order. Ref. [397] applied exponential integration with (1.1.12) for the semilinear wave equation with additive space-time noise and obtained first-order mean-square convergence in time and half-order in space.

Ref. [342] considered finite difference schemes in space for stochastic semilinear wave equation with multiplicative space-time white noise and obtained optimal mean-square convergence with order less than  $1/3$  in space given smooth initial conditions. Finite element methods were investigated in [232] and their convergence order was identified with the regularity of the solution. Ref. [56] considered semi-discretization using spectral Galerkin methods in physical space.

Other than strong approximation of stochastic wave equations, Ref. [183] obtained second-order weak convergence both in space and in time for leap-frog scheme in both space and time solving the one-dimensional semilinear wave equation driven by additive spatial-time white noise. Ref. [341] considered weak convergence of full discrete finite element methods for the linear stochastic elastic equation driven by additive space-time noise and showed that the weak order is twice the strong

order both in time and in space.

Among *stochastic hyperbolic problems*, stochastic conservation laws have also attracted increasing interest, see e.g. [67, 100, 109, 191, 350] for some theoretical results and e.g. [25, 349, 234, 311] for some numerical studies.

Many other evolution equations have also been explored, such as *stochastic KdV equations* (see e.g. [90, 93, 94, 97, 98, 185]), *Ginzburg-Landau equation* (see e.g. [254]), *stochastic Schrödinger equations* (see e.g. [24, 87, 88, 89, 91, 92, 313]), *stochastic age-dependent population* (see e.g. [184]), etc.

For steady stochastic partial differential equations, especially for stochastic elliptic equation, see e.g. [5, 10, 32, 55, 105, 119, 165, 393]. See further discussion in Chapter 6.

### 1.1.6 Stability and convergence of existing numerical methods

There are various aspects to be considered for numerical methods for SPDEs, e.g. the sense of existence of solutions, the sense of convergence, the sense of stability, etc. Here the existence of solutions and numerical solutions to SPDEs are usually interpreted as mild solutions or as variational solutions. Convergence and stability are usually understood in the following sense: mean-square sense (or  $L^p$  in random space), pathwise (almost sure convergence), weak sense (convergence in moments and expectations of functionals of solutions). Here we use ‘strong convergence’ for convergence in the mean-square sense and ‘weak convergence’ for convergence in moments or expectations of the functional of solution.

We focus on weak convergence in this subsection. For strong convergence, we refer to [236] for an optimal convergence order of finite element methods and linear-implicit Euler scheme in time for (1.1.1); see also the aforementioned papers for strong convergence in different problem settings.

## Weak convergence

Similar to the weak convergence of numerical methods for SODEs, the main tool for the weak convergence is the Kolmogorov equation associated with the functional and the underlying SPDE [81, 83].

For linear equations, the Kolmogorov equation for SPDEs is sufficient to obtain optimal weak convergence, see e.g. [99, 124, 354]. Ref. [354] considered weak convergence of  $\theta$ -method in time and spectral methods in physical space for heat equation with additive space-time noise and showed that the weak convergence order is twice that of strong convergence for a finite dimensional functional. Ref. [124] obtained similar conclusion for more general functional, the restriction on which was further removed in [99]. More recently, there have been more works following this approach [229, 230, 252, 341] for linear equations.

For the linear Cahn-Hilliard equation with additive noise, Ref. [229] obtained the weak error for the semidiscrete schemes by linear finite elements with order  $h^{2\beta}|\log(h)|$ , where  $h^\beta$  is strong convergence order and  $\beta$  is determined by  $q_i$  and the smoothness of the initial condition. Ref. [230] provided weak convergence order for the same problem but with further time discretization and proved that the weak convergence order is twice the strong convergence order.

For nonlinear equations, Malliavian calculus for SPDEs has also been used for optimal weak convergence, see e.g. [96, 181, 396]. Ref. [181] applied Malliavin calculus to parabolic SPDE to obtain the weak convergence of linear-implicit Euler and Crank-Nicolson schemes in time for additive noise, where the first-order weak convergence (with certain condition on the functional) is obtained. Ref. [183] showed that the order of weak convergence of leap-frog both in space and time is twice that of strong convergence for wave equation with additive noise as shown for heat equations, see e.g. [96, 99, 124]. Ref. [96] established weak convergence order for the semilinear heat equations with multiplicative space-time noise and showed that the weak convergence order is twice the strong convergence order in time. Ref. [396] obtained weak convergence order of the

linear-implicit Euler scheme in time for (1.1.1) with additive noise and obtained similar conclusions.

For exponential Euler schemes for SODEs, it was proved that the weak convergence order is one (see e.g. [313]), which is the same as the mean-square convergence order, .

For weak convergence of numerical methods for elliptic equations, we can use multivariate calculus to compute the derivatives with respect to (random) parameters and Taylor's expansion, see e.g. [64, 65] and also Chapter 6.

### **Pathwise convergence**

There are two approaches to obtain pathwise convergence. The first is via mean-square convergence. By the Borel-Cantelli lemma (see e.g. [157]), it can be shown that pathwise convergence order is the same as mean-square convergence order (up to an arbitrarily small constant  $\epsilon > 0$ ). For example, Ref. [84] first deduced a pathwise convergence on schemes from the mean-square convergence order established in [159]. Refs. [19, 21, 75, 247, 246] first obtained the mean-square convergence order and then documented the pathwise convergence.

The second approach is without knowing the mean-square convergence. In [363], the authors required pathwise boundedness (uniformly boundedness in time step sizes) to have a pathwise convergence with order  $1/2 - \epsilon$ . In [208], it was shown that it is crucial to establish the pathwise regularity of the solution to obtain pathwise convergence order.

Finally, we note that there are some other senses of convergence, see e.g. [17] for convergence in probability using several approximations of white noise, see also Section 1.2.

### **Stability**

Here we will not review the stability of numerical methods for SPDEs but refer to [390] for the stability of the fully discrete schemes for (1.1.1). We also refer to the following two papers for some general framework on stability and convergence. Ref. [245] proposed a version of Lax equivalence theorem for (1.1.1) with additive and multiplicative noise while  $W^Q$  is replaced with a càdlàg (right

continuous with a left limit) square-integrable martingale. Ref. [235] suggested a general framework for Galerkin methods for (1.1.1) and applied them to Milstein schemes.

We now summarize a recent work on the mean-square stability of Milstein scheme for one-dimensional advection-diffusion equation with multiplicative scalar noise [136, 343]. Ref. [343] analyzed the linear stability (proposed in [46]) of the first-order  $\sigma$ - $\theta$ -scheme and Ref. [187] for SODEs. For a specific equation of the form (1.1.6)

$$dv = -\mu\partial_x v dt + \frac{1}{2}\partial_x^2 v dt - \sqrt{\rho}\partial_x v dW_t, \quad 0 \leq \rho < 1, \quad (1.1.17)$$

the  $\sigma$ - $\theta$  scheme reads, with time step size  $\delta t$  and space step size  $\delta x$ ,

$$\begin{aligned} V_{n+1} = & V_n - \frac{\theta}{2}\left(\frac{\delta t}{\delta x}\mu D_1 - \frac{\delta t}{\delta x^2}D_2\right)V_{n+1} - \frac{1-\theta}{2}\left(\frac{\delta t}{\delta x}\mu D_1 - \frac{\delta t}{\delta x^2}D_2\right)V_n \\ & - \frac{\delta t}{\delta x^2}\rho[\sigma D_2 V_{n+1} + (1-\sigma)D_2 V_n] \\ & - \frac{\sqrt{\rho}}{2}\frac{\sqrt{\delta t}}{\delta x}D_1 V_n \xi_n + \frac{\rho}{2}\frac{\delta t}{\delta x^2}D_2 V_n \xi_n^2, \end{aligned}$$

where  $\xi_n$  are i.i.d. independent standard Gaussian random variables,  $\theta \in [0, 1]$  and  $D_1$  and  $D_2$  are the first and second central difference operators. It was shown that when  $\sigma = -1$ ,  $\theta > 1/2$  the scheme is unconditionally stable as we have, by Fourier stability analysis,

$$\frac{\delta t}{\delta x^2}[1 - 2(\theta - \rho\sigma - \rho^2)] < 1. \quad (1.1.18)$$

When  $\sigma = 0$ ,  $\theta = 0$ , the scheme becomes the Milstein discretization in time in conjunction with finite difference schemes in physical space introduced in [136] where it requires that  $\mu^2 dt \leq 1 - \rho$  in addition to (1.1.18).

It is common that the stability region of a numerical scheme in time for SPDEs with multiplicative noise is smaller than that of the scheme for PDEs, e.g., Crank-Nicolson scheme for

(1.1.1) with multiplicative noise [390], or alternating direction explicit scheme for heat equation with multiplicative noise [359].

### 1.1.7 Conclusion

As SPDEs driven by space-time noise are usually of low regularity, especially when the noise is space-time white noise, it is difficult to obtain efficient high-order schemes in general. Therefore, it is helpful to make full use of specific properties of the underlying SPDEs and preprocessing techniques to derive higher order schemes while keeping the computational cost low. For example, we can use the exponential Euler scheme (1.1.10) with (1.1.12) when the underlying SPDEs are driven by additive noise and their leading differential operators are independent of randomness and time. When SPDEs (with multiplicative noises) have commutative noises, we can use the Milstein scheme (first order strong convergence, see e.g. [211, 236, 309]) while only sampling the increment of Brownian motions.

Another issue for numerical methods of SPDEs is to reduce the computational cost in high-dimensional random space as there are usually infinite dimensional stochastic processes whose truncations converge very slowly. This is the case even when (1.1.12) can be used. Thus, efficient infinite-dimensional integration methods should be employed to obtain the desired statistics. See Section 1.3 for a brief review of numerical integration methods in random space.

## 1.2 Approximation of Brownian motion

### 1.2.1 Piecewise linear approximation

Let us first illustrate the Wong-Zakai approximation by considering the piecewise linear approximation (1.1.5) of the one-dimensional Brownian motion  $W(t)$  for the following Ito SODEs, see e.g. [399, 400]

$$dX = b(t, X)dt + \sigma(t, X)dW(t), \quad X(0) = X_0, \quad (1.2.1)$$

and obtain the following ODE with smooth random inputs

$$dX^{(n)} = b(t, X^{(n)})dt + \sigma(t, X^{(n)})dW^{(n)}(t), \quad X(0) = X_0. \quad (1.2.2)$$

It is proved in [399, 400] that (1.2.2) converges in the mean-square sense to

$$dX = \left( b(t, X) + \frac{1}{2}\sigma(t, X)\sigma_x(t, X) \right) dt + \sigma(t, X)dW(t), \quad X(0) = X_0, \quad (1.2.3)$$

under mild assumptions, which can be written in Stratonovich form [361]

$$dX = b(t, X)dt + \sigma(t, X) \circ dW(t), \quad X(0) = X_0, \quad (1.2.4)$$

where ‘ $\circ$ ’ indicates the Stratonovich product. The term  $\frac{1}{2}\sigma(t, X)\sigma_x(t, X)$  in (1.2.3) is called the Wong-Zakai correction term.

As in the case of scalar SODEs, it is essential to identify the Wong-Zakai correction term (or the equation that the resulting equation from Wong-Zakai approximation converges to) in various cases. For SODEs with scalar noise, say (1.2.1), when Brownian motion is approximated by a process of bounded variation (rather than by piecewise linear approximation), Ref. [364] proved that the convergence to (1.2.3) holds in the pathwise sense (almost surely) if the drift  $b$  is locally Lipschitz continuous and is of linear growth and diffusion  $\sigma$  is continuous with bounded first-order derivatives. However, this conclusion does not hold if  $\sigma$  does not have bounded first-order derivative in  $x$  [364] or the approximation of Brownian motion is not differentiable [289].

For SODEs with multiple noises, Sussmann [365] derived a generic Wong-Zakai correction term for multiple noises. Refs. [242, 241] provided a practical criterion to verify whether a general approximation of Brownian motions (even general semimartingales) leads to a standard Wong-Zakai correction term (e.g.  $1/2\sigma_x\sigma$  for (1.2.4)) or other Wong-Zakai correction terms. To have a

standard Wong-Zakai correction, the bottom line for the approximation of Brownian motion is

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \int_0^T W^{(n)} dW^{(n)} - \int_0^T W(t) \circ dW(t) \right] = 0. \quad (1.2.5)$$

The convergence of Wong-Zakai approximation for SODEs has been established in different senses, e.g. pathwise convergence (e.g. [364, 365]), support theorem (the relation between the support of distribution of the solution and that of its Wong-Zakai approximation, e.g. [15, 294, 362, 385]), mean-square convergence (e.g. [13, 202, 385, 166]), convergence in probability (e.g. [14]).

The Wong-Zakai approximation has been extended in various aspects:

- from single white noise to multiple white noise, see e.g. [166, 365].
- from SODEs to SPDEs, hyperbolic equations (e.g. [295, 345, 346, 422]), parabolic equations (e.g. [2, 101]) including Burgers equation (e.g. [325, 326]) and Navier-Stokes equations (e.g. [69, 385, 384]), and equations on manifold (e.g. [41, 169]), etc.
- from piecewise linear approximation to general approximation: mollifier type (e.g. [202, 270, 272]), Ikeda-Nakao-Yamato-type approximations (e.g. [122, 200, 270]) or their extensions (e.g. [166]), (Lévy-Ciesielski) spectral type (e.g. [50, 265, 422]), general colored noises (e.g. [2, 364, 365]), etc.
- from SODEs driven by Gaussian white noise to those with general processes: general semi-martingales (e.g. [110, 121, 182, 223, 241, 242, 340]), fractional Brownian motion (e.g. [18, 378]), rough path (e.g. [116]), etc.

## 1.2.2 Other approximations

Besides the piecewise linear approximation of Brownian motion (1.1.5), there are several other approximations, e.g. *mollification of Brownian motion*, see e.g. [104, 149, 200, 272, 270, 337, 424, 423]

$$\tilde{W}(t) = \int_{t_n}^t \int_{\mathbb{R}} K(\theta, s) dW(s) d\theta, t \in [t_n, t_{n+1}), \quad (1.2.6)$$

where  $K$  is symmetric. This type of approximation was proposed for a method of lines for SODEs in [329], where no numerical results were presented. When this approximation is applied in SODEs, consistency (convergence without order) has been proved in [104, 149, 150, 273] etc. In [202], the approaches of piecewise linear approximation and mollification have been unified with proved convergence order, known as Ikeda-Nakao-Yamato-type approximations, see also [166].

Another way to approximate Brownian motion is by an *orthogonal expansion*,: (also known as Levy-Ciesielski approximation [71, 214, 237], Ito-Niso approximation [204] or Fourier approximation [333]):

$$W^{(n)}(t) = \sum_{i=1}^n \left( \int_{t_n}^t m_i(s) ds \right) \int_{t_n}^{t_{n+1}} m_i(s) dW(s), \quad t \in [t_n, t_{n+1}), \quad (1.2.7)$$

where  $\{m_i(t)\}_{i=1}^{\infty}$  is a complete orthonormal system (CONS) in  $L^2([0, T])$ ; see [201] for a history review on this approximation. With this approximation, Ogawa [328] defined the following so-called Ogawa integral, which is proved in [327] to coincide with the Stratonovich integral if the integrand  $f(t)$  is a continuous semi-martingale on the natural filtration of Brownian motion.

In fact, taking a piecewise constant basis (normalized) in (1.2.7), we then have exactly the piecewise linear approximation (1.1.5). Also, if we take any orthonormal basis starting from a constant, then taking  $n = 1$  also leads to (1.1.5). With a piecewise constant basis, the use of multiple Ito integrals (Wiener chaos expansion) and multiple Stratonovich integrals was addressed in [49, 50].

The approximation with trigonometric orthonormal basis has been used in Wiener chaos meth-

ods (see e.g. [51, 265, 268, 266, 192, 420]) and will be the approximation for our Wong-Zakai approximation throughout this work. See Chapter 8 for the Wong-Zakai approximation using (1.2.7) for SODEs with time delay.

## 1.3 Integration methods in random space

### 1.3.1 Monte Carlo method and its variants

Numerical SODEs and SPDEs are usually dependent on the Monte Carlo method and its variants to obtain the desired statistics of the solutions. The standard (brute force) Monte Carlo method is known for its slow convergence since its error is usually dominated by its statistical error,  $C/\sqrt{N}$  where  $C$  is the variance of the random process associated with the desired statistics and  $N$  is the number of sampling paths (trajectories), see e.g. [320, Chapter 1].

To accelerate the standard method, some variance reduction methods have been proposed to reduce the number  $C$ , see e.g. [134, 218, 301]. One of the recently developed variance reduction methods, the so-called multilevel Monte Carlo method has attracted more attention for numerical SODEs and SPDEs. The idea of multilevel Monte Carlo methods is to write the desired statistics in a telescoping sum and then to sample the difference terms (between terms defined on two different mesh sizes) in the telescoping sum with a small number of sampling paths, where the corresponding  $C$  is small. In this way, the computational cost is reduced since the difference terms defined on finer meshes would admit smaller variances and thus require a smaller number of sampling paths.

For the multilevel Monte Carlo method for numerical SPDEs, see e.g. [1, 23, 66, 72, 369, 370] for elliptic equations with random coefficients, [306, 307, 308] for stochastic hyperbolic equations, and [22, 136] for stochastic parabolic equations. Ref. [1] proposed time discretization schemes with large stability regions to further reduce the cost of the multilevel Monte Carlo method. However, it has been shown that the multilevel Monte Carlo methods are not robust, see e.g. [224, Chapter 4].

*Quasi-Monte Carlo methods* have also been investigated for numerical SPDEs. Quasi-Monte Carlo methods were originally designed as deterministic integration methods in random space and allowed only moderately high dimensional integrations, see e.g. [320, 356]. However, some randomized quasi-Monte Carlo methods have been successfully applied to solve stochastic elliptic equations with random coefficients, see e.g. [139, 140, 240, 239] where the solution is analytic in random space (parameter space). For the latest review on quasi-Monte Carlo methods, see [102].

Compared to the Monte Carlo type method, the following two methods have no statistical errors and allow efficient short-time integration of SPDEs.

### 1.3.2 Wiener chaos expansion method

By Cameron-Martin theorem [52], any square-integrable stochastic processes (with respect to Wiener measure) can be represented by a series of Hermite polynomials of Gaussian random variables, named Wiener Chaos expansion. This approach has been applied to study practical problems, see e.g. [132] and extended to a general class of chaos expansion (polynomial chaos) for general measures, see e.g. [405, 392]. See e.g. [402] for a review on this topic.

For linear problems driven by white noise in space or in time, the Wiener chaos expansion method has been investigated both theoretically (see e.g. [265, 266, 267, 268]) and numerically (see e.g. [393, 420]). The advantage of Wiener chaos expansion method is that the resulting system of PDEs is linear, lower triangular and deterministic. Also, the Wiener chaos expansion method can be of high accuracy.

However, there are two main difficulties for the Wiener chaos expansion as a numerical method. The first is the efficiency of long-time integration. Usually, this method is only efficient for short-time integration, see e.g. [49, 265]. This limitation can be somewhat removed when a recursive procedure is adopted for computing certain statistics, e.g. first two moments of the solution, see e.g. [420].

The second is nonlinearity. When SPDEs are nonlinear, Wiener chaos expansion methods result in fully coupled systems of deterministic PDEs while the interactions between different Wiener chaos expansion terms necessitate exhaustive computation. This effect has been shown numerically through the stochastic Burgers equation and the Navier-Stokes equations [192].

One remedy for nonlinear problems is to introduce the Wick-Malliavin approximation for nonlinear terms. Wick-Malliavin approximations can be seen as a perturbation of a Wick product formulation by adding high-order Malliavin derivatives of the nonlinear terms to the Wick product formulation, see [388] for details. Basically, lower level Wick-Malliavin approximation (with lower-order Malliavin derivatives) allows weaker nonlinear interactions between the Wiener chaos expansion terms. Let us consider Burgers equation with additive noise for example. When only the Wick product is used (zeroth-order Malliavin derivatives only), the resulting system is lower triangular and contains only one nonlinear equation. When Malliavin derivatives of up to first-order are used, the resulting system of PDEs is only weakly coupled and contains only two nonlinear equations. This approach has been shown very efficient for short-time integration of equations with quadratic nonlinearity and small noise, see e.g. [388].

The Wick product had been formulated in [190] for various SPDEs before the Wick-Malliavin approximation was introduced. The Wick product formulation has been explored with finite element methods in physical space, see e.g. [274, 275, 276, 277, 278, 279, 280, 372, 373, 374, 414, 213] and also [393] for a brief review on SPDEs equipped with Wick product.

### 1.3.3 Stochastic collocation method

In the framework of deterministic integration methods for SPDEs in random space, another solution for nonlinear SPDEs, or linear SPDEs with random coefficient is employing collocation techniques in random space. Here by stochastic collocation methods, we mean the sampling strategies using high-dimensional deterministic quadratures (with certain polynomial exactness) to evaluate desired expectations of solutions to SPDEs.

For SODEs driven by white noise in time, the stochastic collocation method has been known as cubature on Wiener space (e.g. [173, 253, 269, 317, 321]), optimal quantization (e.g. [331, 332]) to solve SODEs in random space, sparse grid of Smolyak type (e.g.[131, 130, 145, 338]), or particle simulation (e.g. [114]). For stochastic collocation methods for equations with color noise, see e.g. [9, 403].

The stochastic collocation methods result in decoupled systems of equations as Monte Carlo method and its variants do, which can be of great advantage in computation. High accuracy and fast convergence can be also observed for stochastic evolution equations, e.g. [130, 145, 338] where sparse grid of Smolyak type was used.

However, the fundamental limitation of these collocation methods is the exponential growth of sampling points with an increasing number of random parameters, see e.g. [145], and thus a failure for longer time integration, see error estimates for cubature on Wiener space (e.g. [26, 61, 103]) and conclusions for optimal quantization (e.g. [331, 332]).

One remedy of cubature methods for longer time integration of SODEs was proposed in [173, 253], where at each time step an empirical measure was reconstructed by regression such that the measure can produce moments close to the first few moments of numerical solutions computed at cubature points. However, for SPDEs, this regression approach has not been documented yet.

## 1.4 Objectives of this work

We will focus on two issues in numerical methods for SPDEs with white noise: one is deterministic integration methods in random space; the other is the effect of truncation of Brownian motion using spectral approximation.

For deterministic integration methods of SPDEs in random space, we aim at longer time numerical integration of time-dependent equations, especially of linear stochastic advection-reaction-diffusion equations. We will study Wiener chaos expansion methods (WCE) and stochastic collo-

cation methods (SCM) and compare their performance and prove their convergence order.

To achieve longer time integration, we adopt the recursive WCE proposed in [265] for the Zakai equation for nonlinear filtering and develop algorithms for the first two moments of solutions. Numerical results show that when high accuracy is required WCE is superior to Monte Carlo methods while WCE is not as efficient if only low accuracy is required, see Chapter 2.

We will show that the recursive approach for SCM for linear advection-reaction-diffusion equations is efficient for longer time integration in Chapter 3. We first analyze the error of SCM (sparse grid collocation of Smolyak type) with Euler scheme in time for linear SODEs, and show the error is small only when the noise magnitude is small and/or the integration time is small.

We will compare WCE and SCM using the recursive procedure in Chapter 4, where we derive error estimates of WCE and SCM for linear advection-reaction-diffusion equations and show that WCE and SCM are competitive in practice by careful numerical comparisons, even though WCE can be of one order higher than SCM.

Among almost all approximations for WCE and SCM, we use the Wong-Zakai approximation with spectral approximation of Brownian motion. The convergence order with respect to the number of truncation modes is half order, see Chapter 4. However, WCE can be of higher order convergence since it can preserve the orthogonality over the Wiener space (infinite dimensional) while SCM can not as the orthogonality is only valid on discrete spaces (finite dimensional), see Chapter 4.

To further investigate the effect of truncation of Brownian motions, we will study the elliptic equation with additive white noise in Chapter 6. We show that the convergence of numerical solutions with truncation of Brownian motion depends on the smoothing effects of the resolvent of the elliptic operator. We also show similar convergence when finite element methods are used. In Chapter 5, we will test the Wong-Zakai approximation in conjunction with stochastic collocation method of stochastic Euler equations modeling a stochastic piston problem and show the effectiveness of this approximation.

In Chapter 7, we will derive a relationship between local truncation error and global truncation error of one-step schemes for SODEs with non-global Lipschitz conditions.

Before we summarize our results in Chapter 9, we will explore the Wong-Zakai approximation for stochastic differential equations with time delay in Chapter 8 and show that Wong-Zakai approximation can facilitate the derivation of various numerical schemes.

As shown in Figure 1.1, we focus on deterministic integration methods in random space, i.e., polynomial chaos (WCE) and stochastic collocation (SCM), in Chapters 2-5 and compare their performance with Monte Carlo methods and/or quasi-Monte Carlo methods. In Chapter 2, we compare WCE and Monte Carlo methods and show that WCE is superior to Monte Carlo methods if high accuracy is needed. In Chapters 3 and 5, we show theoretically and numerically the efficiency of SCM for short time integration and for small magnitudes of noises. In Chapter 4, we compare WCE and SCM in conjuncture with a recursive multistage procedure and show that they are comparable in performance.

We use Monte Carlo methods in Chapters 6-8 as the dimensionality in random space is beyond deterministic integration methods.

In all chapters except Chapters 3 and 7, we apply the Wong-Zakai approximation with the Brownian motion approximated by its spectral truncation. We show that the convergence of numerical schemes based on the Wong-Zakai approximation is determined by further discretization in space (Chapter 6) or in time (Chapter 8).

## Part I: Temporal White Noise

## Chapter 2

# Wiener chaos methods for linear stochastic advection-diffusion-reaction equations

In this chapter, we develop numerical algorithms using Wiener chaos expansion (WCE) for solving second-order linear parabolic stochastic partial differential equations (SPDEs). The algorithm we propose for computing moments of the SPDE solutions is deterministic, i.e., without any statistical errors. We also compare the proposed deterministic algorithm with two other numerical methods based on the Monte Carlo technique and demonstrate that the new method is more efficient for highly accurate solutions. Numerical tests of some examples show that the scheme is of mean-square order  $\mathcal{O}(\Delta^{N/2})$  for advection-diffusion and for diffusion-reaction SPDEs with constant or variable coefficients, where  $\Delta$  is the time-step, and  $N$  is the Wiener chaos order.

## 2.1 Introduction

In this chapter we develop a new numerical method based on nonlinear filtering ideas and spectral expansions for advection-diffusion-reaction equations perturbed by random fluctuations, which form a broad class of second-order linear parabolic stochastic differential equations (SPDEs). The standard approach to constructing SPDE solvers starts with a space discretization of a SPDE, for which spectral methods (see, e.g. [68, 141, 209]), finite element methods (see, e.g. [5, 129, 408]) or spatial finite differences (see, e.g., [5, 160, 353, 412]) can be used. The result of such a space discretization is a large system of ordinary stochastic differential equations (SDEs) which requires time discretization to complete a numerical algorithm. In [89, 96] a SPDE is first discretized in time and then to this semi-discretization a finite-element or finite-difference method can be applied. Other numerical approaches include those making use of splitting techniques [31, 249, 162], quantization [137], or an approach based on the averaging-over-characteristic formula [304, 336]. In [265, 290] numerical algorithms based on the Wiener chaos expansion (WCE) were introduced for solving the nonlinear filtering problem for hidden Markov models. Since then the WCE-based numerical methods have been successfully developed in a number of directions (see e.g. [192, 406]).

In computing moments of SPDE solutions, the existing approaches to solving SPDEs are usually complemented by the Monte Carlo technique. Consequently, in these approaches numerical approximations of SPDE moments have two errors: numerical integration error and Monte Carlo (statistical) error. To reach a high accuracy, we have to run a very large number of independent simulations of the SPDE to reduce the Monte Carlo error. Instead, here we exploit WCE numerical methods to construct a *deterministic* algorithm for computing moments of the SPDE solutions without any use of the Monte Carlo technique.

The rest of this chapter is organized as follows. In Section 2.2 we introduce the linear SPDE considered in this chapter and recall the definition of its Wiener chaos solution. In Section 2.3 we revisit the method employed in [265] and apply it to a more general linear SPDE than the one

treated in [265]. Based on this, the algorithm for computing moments of the SPDE solutions is introduced in Section 2.4. To demonstrate the effectiveness of the proposed algorithm, we perform a number of numerical tests. In Section 2.5 we test the algorithm on two one-dimensional SPDEs and confirm its theoretical order of convergence. In Section 2.6 we apply this algorithm to the passive scalar equation in the periodic case in two dimensions. In both Section 2.5 and Section 2.6 we also compare the WCE-based algorithm with algorithms exploiting the Monte Carlo technique and demonstrate that while the proposed WCE-based algorithm is slower than Monte Carlo-type methods in getting results of low accuracy, in reaching higher accuracy the WCE-based algorithm can be more efficient. A summary and discussion on possible extensions are given in Section 2.7.

## 2.2 WCE of the SPDE solution

Let  $(\Omega, \mathcal{F}, P)$  be a complete probability space,  $\mathcal{F}_t$ ,  $0 \leq t \leq T$ , be a filtration satisfying the usual hypotheses, and  $(w(t), \mathcal{F}_t) = (\{w_k(t), k \geq 1\}, \mathcal{F}_t)$  be a system of one-dimensional independent standard Wiener processes. Let  $\mathcal{D}$  be an open domain in  $\mathbb{R}^d$ . Consider the following SPDE written in the form of Itô:

$$du(t, x) = [\mathcal{L}u(t, x) + f(x)] dt + \sum_{k \geq 1} [\mathcal{M}_k u(t, x) + g_k(x)] dw_k(t), \quad (t, x) \in (0, T] \times \mathcal{D}, \quad (2.2.1)$$

$$u(0, x) = u_0(x), \quad x \in \mathcal{D},$$

where

$$\mathcal{L}u(t, x) = \sum_{i,j=1}^d a_{ij}(x) D_i D_j u(t, x) + \sum_{i=1}^d b_i(x) D_i u(t, x) + c(x) u(t, x), \quad (2.2.2)$$

$$\mathcal{M}_k u(t, x) = \sum_{i=1}^d b_i^k(x) D_i u(t, x) + h^k(x) u(t, x),$$

and  $D_i := \partial_{x_i}$ . We assume that  $\mathcal{D}$  is either bounded with a regular boundary or that  $\mathcal{D} = \mathbb{R}^d$ . In the former case we will consider periodic boundary conditions and in the latter the Cauchy problem. We also assume that the coefficients of operators  $\mathcal{L}$  and  $\mathcal{M}$  are uniformly bounded and  $\mathcal{L} - \frac{1}{2} \sum_{k \geq 1} \mathcal{M}_k \mathcal{M}_k$  is nonnegative definite. When the coefficients of  $\mathcal{L}$  and  $\mathcal{M}$  are sufficiently smooth, existence and uniqueness results for the solution of (2.2.1)–(2.2.2) are available, e.g. in [347] and under weaker assumptions, see e.g. [291, 268].

Now let us recall (see details in [291, 266, 268]) the definition of a Wiener chaos solution to the linear SPDE (2.2.1)–(2.2.2). Denote by  $\mathcal{J}$  the set of multi-indices  $\alpha = (\alpha_{k,l})_{k,l \geq 1}$  of finite length  $|\alpha| = \sum_{i,k=1}^{\infty} \alpha_{k,l}$ , i.e.,

$$\mathcal{J} = \{ \alpha = (\alpha_{k,l}, k, l \geq 1), \alpha_{k,l} \in \{0, 1, 2, \dots\}, |\alpha| < \infty \}.$$

Here  $k$  denotes the number of Wiener processes and  $l$  the number of Gaussian random variables approximating each Wiener process as will be shown shortly. We represent the solution of (2.2.1)–(2.2.2) as

$$u(t, x) = \sum_{\alpha \in \mathcal{J}} \frac{1}{\sqrt{\alpha!}} \varphi_{\alpha}(t, x) \xi_{\alpha}, \quad (2.2.3)$$

where  $\{\xi_{\alpha}\}$  is a complete orthonormal system (CONS) in  $L^2(\Omega, \mathcal{F}_t, P)$ ,  $\alpha! = \prod_{k,l} (\alpha_{k,l}!)$ , and  $\varphi_{\alpha}$  satisfies the following system of equations (the propagator):

$$\begin{aligned} \frac{\partial \varphi_{\alpha}(s, x)}{\partial s} &= \mathcal{L} \varphi_{\alpha}(s, x) + f(x) \mathbf{1}_{\{|\alpha|=0\}} \\ &+ \sum_{k,l} \alpha_{k,l} m_l(s) [\mathcal{M}_k \varphi_{\alpha - (k,l)}(s, x) + g_k(x) \mathbf{1}_{\{|\alpha|=1\}}], \quad 0 < s \leq t, x \in \mathcal{D}, \\ \varphi_{\alpha}(0, x) &= u_0(x) \mathbf{1}_{\{|\alpha|=0\}}, \quad x \in \mathcal{D}, \end{aligned} \quad (2.2.4)$$

where  $\alpha^-(k, l)$  is the multi-index with components

$$(\alpha^-(k, l))_{i,j} = \begin{cases} \max(0, \alpha_{i,j} - 1), & \text{if } i = k \text{ and } j = l, \\ \alpha_{i,j}, & \text{otherwise.} \end{cases} \quad (2.2.5)$$

The random variables  $\xi_\alpha$  in (2.2.3) are constructed according to the Cameron-Martin theorem [52]:

$$\xi_\alpha := \prod_{\alpha} \left( \frac{H_{\alpha_{k,l}}(\xi_{k,l})}{\sqrt{\alpha_{k,l}!}} \right), \quad \alpha \in \mathcal{J}, \quad (2.2.6)$$

where  $\{m_l\} = \{m_l(s)\}_{l \geq 1}$  is a CONS in  $L^2([0, t])$ ,  $\xi_{k,l} = \int_0^t m_l(s) dw_k(s)$ , and  $H_n$  is the  $n$ -th Hermite polynomial:

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}. \quad (2.2.7)$$

The representation (2.2.3)-(2.2.7) is called a WCE of the SPDE solution. It is clear that a truncation of the WCE (2.2.3) presents a possibility for constructing numerical methods for SPDEs. This is considered in the next section.

## 2.3 Multistage WCE method

In addition to the multi-index length  $|\alpha| = \sum_{i,k=1}^{\infty} \alpha_{k,l}$ , we define the order of multi-index  $\alpha$  :  $d(\alpha) = \max \{l \geq 1 : \alpha_{k,l} > 0 \text{ for some } k \geq 1\}$  and the truncated set of multi-indices:

$$\mathcal{J}_{\mathbf{N}, \mathbf{n}} = \{\alpha \in \mathcal{J} : |\alpha| \leq \mathbf{N}, \quad d(\alpha) \leq \mathbf{n}\}.$$

Here  $\mathbf{N}$  is the highest Hermite polynomial order and  $\mathbf{n}$  is the maximum number of Gaussian random variables for each Wiener process. Using (2.2.3), we introduce the truncated Wiener chaos solution

$$u_{\mathbf{N}, \mathbf{n}}(t, x) = \sum_{\alpha \in \mathcal{J}_{\mathbf{N}, \mathbf{n}}} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(t, x) \xi_\alpha. \quad (2.3.1)$$

We choose the basis  $\{m_l(s)\}_{l \geq 1}$  as

$$m_1(s) = \frac{1}{\sqrt{t}}, \quad m_l(s) = \sqrt{\frac{2}{t}} \cos\left(\frac{\pi(l-1)s}{t}\right), \quad l \geq 2, \quad 0 \leq s \leq t. \quad (2.3.2)$$

See a discussion on selection of basis in [265].

The truncated expansion (2.3.1) together with (2.2.4), (2.2.6), and (2.3.2) gives us a constructive approximation of the solution to (2.2.1), where implementation requires that we numerically solve the propagator (2.2.4).

It is proved in [265, Theorem 2.2] that when  $b_i^k(t, x) = 0$ ,  $c = 0$ ,  $g_k = 0$  (reaction-diffusion equation) and the number of noises is finite there is a constant  $C > 0$  such that for any  $t \in (0, T]$

$$\mathbb{E}[\|u_{N,n}(t, \cdot) - u(t, \cdot)\|_{L^2}^2] \leq C e^{Ct} \left( \frac{(Ct)^{N+1}}{(N+1)!} + \frac{t^3}{n} \right). \quad (2.3.3)$$

Our preliminary analysis shows that in the *general case* of the equation (2.2.1), the error estimate (2.3.3) is expected to be

$$\mathbb{E}[\|u_{N,n}(t, \cdot) - u(t, \cdot)\|_{L^2}^2] \leq C e^{Ct} \left( \frac{(Ct)^{N+1}}{(N+1)!} + \frac{t^2}{n} \right). \quad (2.3.4)$$

It follows from the error estimates (2.3.3) and (2.3.4) that the error of the approximation  $u_{N,n}(t, \cdot)$  grows exponentially in time  $t$  which severely limits its practical use. To overcome this difficulty, it was proposed in [265] to introduce a time discretization with step  $\Delta > 0$  and view (2.3.1), (2.2.4), (2.2.6), (2.3.2) as the one-step approximation of the SPDE solution based on which an effective numerical method applicable to longer time intervals was constructed.

To this end, let us introduce the multi-step basis for the WCE and its corresponding propagator. Let  $0 = t_0 < t_1 < \dots < t_K = T$  be a uniform partition of the time interval  $[0, T]$  with time step

size  $\Delta$ , see Figure 2.1. Let  $\{m_k^{(i)}\} = \{m_k^{(i)}(s)\}_{k \geq 1}$  be the following CONS in  $L^2([t_i, t_{i-1}])$ :

$$\begin{aligned} m_l^{(i)} &= m_l(s - t_i), \quad t_i \leq s \leq t_{i-1}, \\ m_l(s) &= \frac{1}{\sqrt{\Delta}}, \quad m_l(s) = \sqrt{\frac{2}{\Delta}} \cos\left(\frac{\pi(l-1)s}{\Delta}\right), \quad l \geq 2, \quad 0 \leq s \leq \Delta, \\ m_l(s) &= 0, \quad l \geq 1, \quad s \notin [0, \Delta]. \end{aligned} \tag{2.3.5}$$

Define the random variables  $\xi_\alpha^{(i)}$ ,  $i = 1, \dots, K$ , as

$$\xi_\alpha^{(i)} := \prod_{\alpha}^{(i)} \left( \frac{H_{\alpha_{k,l}}(\xi_{k,l}^{(i)})}{\sqrt{\alpha_{k,l}!}} \right), \quad \alpha \in \mathcal{J}, \tag{2.3.6}$$

where  $\xi_{k,l}^{(i)} = \int_{t_i}^{t_{i-1}} m_l^{(i)}(s) dw_k(s)$ , and  $H_n$  are Hermite polynomials (2.2.7).

Let

$$u_{\Delta, N, n}(0, x) = u_0(x) \tag{2.3.7}$$

and by induction for  $i = 1, \dots, K$ :

$$u_{\Delta, N, n}(t_{i-1}, x) = \sum_{\alpha \in \mathcal{J}_{N, n}} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha^{(i)}(\Delta, x) \xi_\alpha^{(i)}, \tag{2.3.8}$$

where  $\varphi_\alpha^{(i)}(\Delta, x)$  solves the system

$$\begin{aligned} \frac{\partial \varphi_\alpha^{(i)}(s, x)}{\partial s} &= \mathcal{L} \varphi_\alpha^{(i)}(s, x) + f(x) \mathbf{1}_{\{|\alpha|=0\}} \\ &+ \sum_{k,l} \alpha_{k,l} m_l^{(i)}(s) \left[ \mathcal{M}_k \varphi_{\alpha - (l,k)}^{(i)}(s, x) + g_k(x) \mathbf{1}_{\{|\alpha|=1\}} \right], \quad s \in (0, \Delta], \\ \varphi_\alpha^{(i)}(0, x) &= u_{\Delta, N, n}(t_i, x) \mathbf{1}_{\{|\alpha|=0\}}. \end{aligned} \tag{2.3.9}$$

Thus, (2.3.7)-(2.3.9) together with (2.3.5) and (2.3.6) gives us a *recursive* method (called the RWCE method) for solving the SPDE (2.2.1), where implementation requires to numerically solve the

propagator (2.3.9) at every time step.

Based on the one-step error (2.3.3), the following global error estimate for the RWCE method is proved in [265, Theorem 2.4] (the case of  $b_i^k(t, x) = 0$ ,  $c = 0$ ,  $g_k = 0$  and finite number of noises):

$$\mathbb{E}[\|u_{\Delta, N, n}(\mathbf{t}_{i-1}, \cdot) - u(\mathbf{t}_{i-1}, \cdot)\|_{L^2}^2] \leq C e^{CT} \left( \frac{(C\Delta)^N}{(N+1)!} + \frac{\Delta^2}{n} \right), \quad i = 1, \dots, K, \quad (2.3.10)$$

for some  $C > 0$  independent of  $\Delta$ ,  $N$ , and  $n$ , i.e., this method is of global mean-square order  $\mathcal{O}\left(\frac{\Delta^{N/2}}{\sqrt{(N+1)!}} + \frac{\Delta}{\sqrt{n}}\right)$ . Moreover, based on (2.3.4), one can prove that in the general case of (2.2.1) (advection-diffusion-reaction equations) the error estimate for the RWCE method will have the form

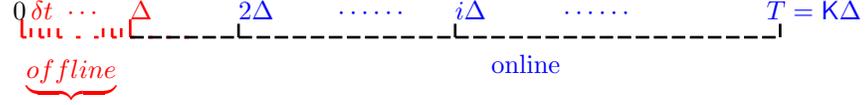
$$\mathbb{E}[\|u_{\Delta, N, n}(\mathbf{t}_{i-1}, \cdot) - u(\mathbf{t}_{i-1}, \cdot)\|_{L^2}^2] \leq C e^{CT} \left( \frac{(C\Delta)^N}{(N+1)!} + \frac{\Delta}{n} \right), \quad i = 1, \dots, K, \quad (2.3.11)$$

i.e., this method is of mean-square order  $\mathcal{O}\left(\frac{\Delta^{N/2}}{\sqrt{(N+1)!}} + \sqrt{\frac{\Delta}{n}}\right)$ .

As we already mentioned, the RWCE method requires to solve the propagator (2.3.9) at every time step, which is computationally rather expensive. To reduce the cost, we introduce a modification of this method following [265]. The idea is to expand the initial condition  $u_0(x)$  in a basis  $\{e_m\}$ , present  $u_{\Delta, N, n}(\mathbf{t}_i, x)$  as  $u_{\Delta, N, n}(\mathbf{t}_i, x) = \sum_m c_m e_m(x)$  and note that  $\varphi_\alpha(\Delta, x; u_{\Delta, N, n}(\mathbf{t}_i, \cdot)) = \sum_m c_m \varphi_\alpha(\Delta, x; e_m)$ , where  $\varphi_\alpha(s, x; \phi)$  is the solution of the propagator (2.3.9) with the initial condition  $\phi(x)$ .

The idea is illustrated in Figure 2.1 with the help of a sketch. We can first compute the propagator (2.3.12) (see below) on  $(0, \Delta]$  and obtain a problem-dependent basis  $q_{\alpha, l, m}$  (2.3.13). This step is called “offline” as in [265]. Thus, one recursively computes the solution “online” by (2.3.14) and (2.3.15) only at time  $i\Delta$  ( $i = 2, \dots, K$ ) using the obtained basis  $q_{\alpha, l, m}$ . Specifically, we proceed as follows. Let  $\{e_m\} = \{e_m(x)\}_{m \geq 1}$  be a CONS in  $L^2(\mathcal{D})$  with boundary conditions satisfied and

Figure 2.1: Illustration of the idea of multistage WCE. The dotted line denotes the “offline” computation, where we solve the propagator up to time  $\Delta$ . The dash line implies that one solves only the solution on certain time levels instead of on the entire time interval.



$(\cdot, \cdot)$  be the inner product in that space. Let  $\varphi_\alpha(s, x; \phi)$  solve the following propagator:

$$\begin{aligned} \frac{\partial \varphi_\alpha(s, x; \phi)}{\partial s} &= \mathcal{L}\varphi_\alpha(s, x; \phi) + f(x)\mathbf{1}_{\{|\alpha|=0\}} \\ &+ \sum_{k,l} \alpha_{k,l} m_l(s) [\mathcal{M}_k \varphi_{\alpha-(l,k)}(s, x; \phi) + g_k(x)\mathbf{1}_{\{|\alpha|=1\}}], \quad s \in (0, \Delta], \\ \varphi_\alpha^{(i)}(0, x) &= \phi(x)\mathbf{1}_{\{|\alpha|=0\}}, \end{aligned} \tag{2.3.12}$$

where  $m_l(s)$  are as in (2.3.2). Define

$$q_{\alpha,l,m} = (\varphi_\alpha(\Delta, \cdot; e_l), e_m), \quad l, m \geq 1, \tag{2.3.13}$$

and then find by induction the coefficients

$$\begin{aligned} \psi_m(0; \mathbf{N}, \mathbf{n}) &:= (u_0, e_m), \\ \psi_m(i; \mathbf{N}, \mathbf{n}) &:= \sum_{\alpha \in \mathcal{J}_{\mathbf{N}, \mathbf{n}}} \sum_l \frac{1}{\sqrt{\alpha!}} \psi_l(i-1; \mathbf{N}, \mathbf{n}) q_{\alpha,l,m} \xi_\alpha^{(i)}, \quad i = 1, \dots, K. \end{aligned} \tag{2.3.14}$$

It is proved in [265, Theorem 2.5] that

$$u_{\Delta, \mathbf{N}, \mathbf{n}}(t_{i-1}, x) = \sum_m \psi_m(i; \mathbf{N}, \mathbf{n}) e_m(x), \quad i = 0, \dots, K, \quad P\text{-a.s.} \tag{2.3.15}$$

We refer to the numerical method (2.3.15), (2.3.12)-(2.3.14) together with (2.3.5)-(2.3.6) as the *multistage WCE method* for the SPDE (2.2.1).

In practice, if the equation (2.2.1) has an infinite number of Wiener processes, we truncate them to a finite number  $r \geq 1$  of noises. We introduce the correspondingly truncated set  $\mathcal{J}_{\mathbf{N},\mathbf{n},r}$  so that

$$\mathcal{J}_{\mathbf{N},\mathbf{n},r} = \{\alpha \in \mathcal{J} : |\alpha| \leq \mathbf{N}, \quad d_r(\alpha) \leq \mathbf{n}\},$$

where  $d_r(\alpha) = \max\{l \geq 1 : \alpha_{k,l} > 0 \text{ for some } 1 \leq k \leq r\}$ .

**Algorithm 2.3.1.** Choose a truncation of the number of noises  $r \geq 1$  and the algorithm's parameters: a CONS  $\{e_m(x)\}_{m \geq 1}$  and its truncation  $\{e_m(x)\}_{m=1}^{\mathbf{M}}$ ; a time step  $\Delta$ ;  $\mathbf{N}$  and  $\mathbf{n}$  which together with  $r$  determine the size of the multi-index set  $\mathcal{J}_{\mathbf{N},\mathbf{n},r}$ .

*Step 1.* For each  $m = 1, \dots, \mathbf{M}$ , solve the propagator (2.3.12) for  $\alpha \in \mathcal{J}_{\mathbf{N},\mathbf{n},r}$  on the time interval  $[0, \Delta]$  with the initial condition  $e_m(x)$  and denote the obtained solution as  $\varphi_\alpha(\Delta, x; e_m)$ ,  $\alpha \in \mathcal{J}_{\mathbf{N},\mathbf{n},r}$ ,  $m = 1, \dots, \mathbf{M}$ . Note in this step, we need to choose also a time step size  $\delta t$  to solve the equations in the propagator numerically.

*Step 2.* Evaluate  $\psi_m(0; \mathbf{N}, \mathbf{n}, \mathbf{M}) = (u_0, e_m)$ ,  $m = 1, \dots, \mathbf{M}$ , where  $u_0(x)$  is the initial condition for (2.2.1), and  $q_{\alpha,l,m} = (\varphi_\alpha(\Delta, \cdot; e_l), e_m(\cdot))$ ,  $l, m = 1, \dots, \mathbf{M}$ .

*Step 3.* On the  $i$ -th time step (at time  $t = i\Delta$ ), generate the Gaussian random variables  $\xi_\alpha^{(i)}$ ,  $\alpha \in \mathcal{J}_{\mathbf{N},\mathbf{n},r}$ , according to (2.3.6), compute the coefficients

$$\psi_m(i; \mathbf{N}, \mathbf{n}, \mathbf{M}) = \sum_{\alpha \in \mathcal{J}_{\mathbf{N},\mathbf{n},r}} \sum_{l=1}^{\mathbf{M}} \frac{1}{\sqrt{\alpha!}} \psi_l(i-1; \mathbf{N}, \mathbf{n}, \mathbf{M}) q_{\alpha,l,m} \xi_\alpha^{(i)}, \quad m = 1, \dots, \mathbf{M},$$

and obtain the approximate solution of (2.2.1)

$$u_{\Delta, \mathbf{N}, \mathbf{n}}^{\mathbf{M}}(\mathbf{t}_{i-1}, x) = \sum_{m=1}^{\mathbf{M}} \psi_m(i; \mathbf{N}, \mathbf{n}, \mathbf{M}) e_m(x).$$

Algorithm 2.3.1 coincides with the algorithm proposed in [265] for (2.2.1) in the case of  $b_i^k(t, x) = 0$ ,  $c = 0$ ,  $g_k = 0$ , and finite number of noises but generalizes it to a wider class of linear SPDEs of the form (2.2.1). In particular, the algorithm from [265] was applied to the nonlinear filtering

problem for hidden Markov models in the case of independent noises in signal and observation, while Algorithm 2.3.1 is also applicable when noises in signal and observation are dependent.

Algorithm 2.3.1 allows us to simulate mean-square approximations of the solution to the SPDE (2.2.1). It can also be used together with the Monte Carlo technique for computing expectations of functionals of the solution to (2.2.1). In the next section we propose an algorithm based on Algorithm 2.3.1, which allows us to compute moments of the solution to (2.2.1) without using the Monte Carlo technique.

**Remark 2.3.2.** *We note that the cost of simulation of the random field  $u(t_i, x)$  by Algorithm 2.3.1 over  $K$  timesteps is proportional to  $KM^2 \frac{(N+nr)!}{N!(nr)!}$ .*

## 2.4 Algorithm for computing moments

Implementation of Algorithm 2.3.1 requires the generation of the random variables  $\xi_\alpha^{(i)}$  (see (2.3.6)). Then, for computing moments of the solution of the SPDE problem (2.2.1), we also need to make use of the Monte Carlo technique. As is well known, Monte Carlo methods have a low rate of convergence. In this section we present a deterministic algorithm (Algorithm 2.4.1) for computing moments, i.e., an algorithm which does not require any random numbers and does not have a statistical error. In Sections 2.5 and 2.6 we compare Algorithm 2.4.1 with some Monte Carlo-type methods and demonstrate that Algorithm 2.4.1 can be more computationally efficient when higher accuracy is required.

First, it is not difficult to see that the mean solution  $\mathbb{E}[u(t, x)]$  is equal to the solution  $\varphi_{(0)}(t, x)$  of the propagator (2.3.12) with  $\alpha = (0)$ :

$$\mathbb{E}[u(t, x)] = \varphi_{(0)}(t, x).$$

Thus evaluating the mean  $Eu(t, x)$  is reduced to numerical solution of the linear deterministic PDE

for  $\varphi_{(0)}(t, x)$ .

We limit ourselves here to presenting an algorithm for computing the second moment of the solution,  $\mathbb{E}[u^2(t, x)]$ . Other moments of the solution  $u(t, x)$  can be considered analogously.

According to Algorithm 2.3.1, we approximate the solution  $u(\mathbf{t}_{i-1}, x)$  of (2.2.1) by  $u_{\Delta, \mathbf{N}, \mathbf{n}}^M(\mathbf{t}_{i-1}, x)$  (when  $f = g_k = 0$ ) as follows:

$$\begin{aligned}\psi_m(0; \mathbf{N}, \mathbf{n}, M) &= (u_0, e_m), \quad m = 1, \dots, M, \\ \psi_m(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{n}, M) &= \sum_{\alpha \in \mathcal{J}_{\mathbf{N}, \mathbf{n}, r}} \sum_{l=1}^M \frac{1}{\sqrt{\alpha!}} \psi_l(\mathbf{t}_i; \mathbf{N}, \mathbf{n}, M) q_{\alpha, l, m} \xi_{\alpha}^{(i)}, \quad m = 1, \dots, M, \\ u_{\Delta, \mathbf{N}, \mathbf{n}}^M(\mathbf{t}_{i-1}, x) &= \sum_{m=1}^M \psi_m(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{n}, M) e_m(x), \quad i = 1, \dots, K,\end{aligned}$$

where  $q_{\alpha, l, m}$  are from (2.3.13) and  $\xi_{\alpha}^{(i)}$  are from (2.3.6). Then, we can evaluate the covariance matrices

$$Q_{lm}(0; \mathbf{N}, \mathbf{n}, M) := \psi_l(0; \mathbf{N}, \mathbf{n}, M) \psi_m(0; \mathbf{N}, \mathbf{n}, M), \quad l, m = 1, \dots, M, \quad (2.4.1)$$

$$\begin{aligned}Q_{lm}(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{n}, M) &:= E[\psi_l(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{n}, M) \psi_m(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{n}, M)] \\ &= \sum_{j, k=1}^M Q_{jk}(\mathbf{t}_i; \mathbf{N}, \mathbf{n}, M) \sum_{\alpha \in \mathcal{J}_{\mathbf{N}, \mathbf{n}, r}} \frac{1}{\alpha!} q_{\alpha, j, l} q_{\alpha, k, m}, \\ & \quad l, m = 1, \dots, M, \quad i = 1, \dots, K,\end{aligned}$$

and, consequently, the second moment of the approximate solution

$$\mathbb{E}[(u_{\Delta, \mathbf{N}, \mathbf{n}}^M(\mathbf{t}_{i-1}, x))^2] = \sum_{l, m=1}^M Q_{lm}(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{n}, M) e_l(x) e_m(x), \quad i = 1, \dots, K. \quad (2.4.2)$$

We note that implementation of (2.4.1)–(2.4.2) does not require generation of the random variables  $\xi_{\alpha}^{(i)}$ . Hence we have constructed a deterministic algorithm for computing the second moments of the solution to the SPDE (2.2.1) when  $f = g_k = 0$ , which we formulate below.

**Algorithm 2.4.1** (Recursive multistage Wiener chaos expansion). *Choose a truncation of the number of noises  $r \geq 1$  in (2.2.1) and the algorithm's parameters: a CONS  $\{e_m(x)\}_{m \geq 1}$  and its truncation  $\{e_m(x)\}_{m=1}^M$ ; a time step  $\Delta$ ;  $\mathbf{N}$  and  $\mathbf{n}$  which together with  $r$  determine the size of the multi-index set  $\mathcal{J}_{\mathbf{N},\mathbf{n},r}$ .*

*STEP 1 For each  $m = 1, \dots, M$ , solve the propagator (2.3.12) for  $\alpha \in \mathcal{J}_{\mathbf{N},\mathbf{n},r}$  on the time interval  $[0, \Delta]$  with the initial condition  $\phi(x) = e_m(x)$  and denote the obtained solution as  $\varphi_\alpha(\Delta, x; e_m)$ ,  $\alpha \in \mathcal{J}_{\mathbf{N},\mathbf{n},r}$ ,  $m = 1, \dots, M$ . Also, choose a time step size  $\delta t$  to solve the equations in the propagator numerically.*

*STEP 2 Evaluate  $\psi_m(0; \mathbf{N}, \mathbf{n}, M) = (u_0, e_m)$ ,  $m = 1, \dots, M$ , where  $u_0(x)$  is the initial condition for (2.2.1), and  $q_{\alpha,l,m} = (\varphi_\alpha(\Delta, \cdot; e_l), e_m(\cdot))$ ,  $l, m = 1, \dots, M$ .*

*STEP 3 Recursively compute the covariance matrices  $Q_{lm}(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{n}, M)$  according to (2.4.1) and obtain the second moment  $\mathbb{E}[(u_{\Delta, \mathbf{N}, \mathbf{n}}^M(\mathbf{t}_{i-1}, x))^2]$  of the approximate solution to (2.2.1) by (2.4.2).*

We emphasize again that Algorithm 2.4.1 for computing moments does not have a statistical error.

Let us discuss the error of Algorithm 2.4.1. One can show (see, e.g. [265]) that due to the orthogonality of the random variables  $\xi_\alpha^{(i)}$  in the sense that  $\mathbb{E}[\xi_\alpha^{(i)} \xi_\beta^{(j)}] = 0$  unless  $i = j$  and  $\alpha = \beta$ , the following equality holds:

$$\mathbb{E}[u^2(t, x)] - \mathbb{E}[u_{\mathbf{N}, \mathbf{n}}^2(t, x)] = \mathbb{E}[(u(t, x) - u_{\mathbf{N}, \mathbf{n}}(t, x))^2]. \quad (2.4.3)$$

Hence the error estimates for approximation of the second moment  $\mathbb{E}[u^2(\mathbf{t}_{i-1}, x)]$  by  $\mathbb{E}[u_{\Delta, \mathbf{N}, \mathbf{n}}^2(\mathbf{t}_{i-1}, x)]$  is equal to the errors given in (2.3.10) and (2.3.11).

We do not discuss here errors arising from noise truncation and from truncation of the basis  $\{e_m(x)\}_{m \geq 1}$ .

**Remark 2.4.2.** *It is not difficult to show that the computational costs of Steps 1 and 2 of Algorithm 2.4.1 are proportional to  $M^2 \frac{(N+nr)!}{N!(nr)!}$ . In general the computational cost of Step 3 over  $K$  timesteps is proportional to  $KM^4 \frac{(N+nr)!}{N!(nr)!}$ . Taking this into account together with the error estimates (2.3.10) and (2.3.11), it is usually computationally beneficial to choose  $n = 1$  and  $N = 2$  or  $1$ . The main computational cost of Algorithm 2.4.1 is due to the total number of basis functions  $M$  (in physical space) required for reaching a satisfactory accuracy. As is well known, for a fixed accuracy the number  $M$  of basis functions  $\{e_m\}_{m=1}^M$  is proportional to  $C^d$ , where  $C$  depends on a choice of the basis and on the problem. If the variance of  $u^2(t, x_i)$  is relatively large and the problem considered does not require a very large number of basis functions  $M$ , then one expects Algorithm 2.4.1 to be computationally more efficient in evaluating second moments than the combination of Algorithm 2.3.1 with the Monte Carlo technique.*

Algorithm 2.4.1's efficiency can often be improved by choosing an appropriate basis  $\{e_m\}$  so that the majority of functions  $q_{\alpha,l,m}$  are identically zero or negligible and hence can be dropped from computing the covariance matrix  $\{Q_{lm}(\mathbf{t}_{i-1}; \mathbf{N}, M)\}_{l,m=1}^M$ , significantly decreasing the computational cost of Step 3. For instance, for the periodic passive scalar equation considered in Section 2.6 we choose the Fourier basis  $\{e_m\}$ . In this case the number of zero  $q_{\alpha,l,m}$  is proportional just to  $M$  (the total number of  $q_{\alpha,l,m}$  is proportional to  $M^2$ ) and, consequently, the computational cost of Step 3 (and hence that of Algorithm 2.4.1) becomes proportional to  $M^2$  instead of the original  $M^4$ . Moreover, computation of the covariance matrix according to (2.4.1) can be done in parallel. Clearly, the use of reduced-order methods with offline/online strategies [348] can greatly reduce the value of  $M$  and hence will make the proposed method very efficient.

**Remark 2.4.3.** *It is more expensive to compute higher-order moments by a deterministic algorithm analogous to Algorithm 2.4.1. Since second moments give us such important, from the physical point of view, characteristics as energy and correlation functions, Algorithm 2.4.1 can be a competitive alternative to Monte Carlo-type methods in practical situations.*

## 2.5 Numerical tests in one dimension

We start (Section 2.5.1) with a description of two one-dimensional test problems used in the numerical tests. Then, for clarity of exposition, we illustrate application of Algorithm 2.4.1 to these problems (Section 2.5.2). We present results of numerical tests of Algorithm 2.4.1 in Section 2.5.3 and its comparison with some Monte Carlo-type algorithms in Section 2.5.4. In the next section (Section 2.6) we also perform numerical tests with a two-dimensional passive scalar equation.

### 2.5.1 Test problems

We consider the following two model problems. The first one is the *stochastic advection-diffusion equation* with periodic boundary condition, written in the Stratonovich form as

$$\begin{aligned} du(t, x) &= \epsilon u_{xx}(t, x) dt + \sigma u_x(t, x) \circ dw(t), \quad t > 0, \quad x \in (0, 2\pi), \\ u(0, x) &= \sin(x), \end{aligned} \tag{2.5.1}$$

or in the Itô form as

$$du(t, x) = a u_{xx}(t, x) dt + \sigma u_x(t, x) dw(t), \quad u(0, x) = \sin(x).$$

Here  $w(t)$  is a standard one-dimensional Wiener process,  $\sigma > 0$ ,  $\epsilon \geq 0$  are constants, and  $a = \epsilon + \sigma^2/2$ . The solution of (2.5.1) is

$$u(t, x) = e^{-\epsilon t} \sin(x + \sigma w(t)), \tag{2.5.2}$$

and its first and second moments are

$$\mathbb{E}[u(t, x)] = e^{-\epsilon t} \sin(x), \quad \mathbb{E}[u^2(t, x)] = e^{-2\epsilon t} \left( \frac{1}{2} - \frac{1}{2} e^{-2\sigma^2 t} \cos(2x) \right).$$

We note that for  $\epsilon = 0$  the equation (2.5.1) becomes degenerate.

The second model problem is the following Ito *reaction-diffusion equation* with periodic boundary condition:

$$\begin{aligned} du(t, x) &= au_{xx}(t, x) dt + \sigma u(t, x) dw(t), \quad t > 0, x \in (0, 2\pi), \\ u(0, x) &= \sin(x), \end{aligned} \tag{2.5.3}$$

where  $\sigma > 0$  and  $a \geq 0$  are constants. Its solution is

$$u(t, x) = \exp\left(-\left(a + \frac{\sigma^2}{2}\right)t + \sigma w(t)\right) \sin(x); \tag{2.5.4}$$

and its first and second moments are

$$\mathbb{E}[u(t, x)] = e^{-at} \sin(x), \quad \mathbb{E}[u^2(t, x)] = \exp(-2a - \sigma^2)t \sin^2(x).$$

In Sections 2.5.3 and 2.5.4 we will test Algorithm 2.4.1 by evaluating the second moments  $\mathbb{E}u^2(t, x)$  of the solutions to (2.5.1) and (2.5.3).

## 2.5.2 Application of WCE algorithms to the model problem

The problems (2.5.1) and (2.5.3) are simpler than the general linear SPDE (2.2.1) and, consequently, Algorithm 2.4.1 applied to them takes a simpler form (see Algorithm 2.5.1 below).

We note that when an SPDE has a single Wiener process only, the multi-index  $\alpha$  takes the form  $\alpha = (\alpha_1, \alpha_2, \dots)$ , where  $\alpha_i$  are non-negative integers. For instance, if  $|\alpha| = 0$  (i.e.,  $\alpha = (0, 0, \dots)$ ) then the corresponding  $\xi_\alpha = 1$  (cf. (2.2.6)). If  $|\alpha| = 1$ , then the multi-index  $\alpha = (0, \dots, 0, 1, 0, \dots)$  with  $\alpha_i = 1$  and the other  $\alpha_k = 0$ , and the corresponding  $\xi_\alpha = H_1(\xi_i) = \xi_i = \int_0^t m_i(s) dw(s)$ . If  $|\alpha| = 2$ , then the multi-index is either of the type  $\alpha = (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots)$  with  $\alpha_i = \alpha_j = 1$  and the other  $\alpha_k = 0$ , and consequently,  $\xi_\alpha = H_1(\xi_i)H_1(\xi_j) = \int_0^t m_i(s) dw(s) \int_0^t m_j(s) dw(s)$ ; or

$\alpha = (0, \dots, 0, 2, 0, \dots)$  with  $\alpha_i = 2$  and the other  $\alpha_k = 0$ , and consequently,  $\xi_\alpha = H_2(\xi_i)/\sqrt{2} = \frac{1}{\sqrt{2}} \left[ \left( \int_0^t m_i(s) dw(s) \right)^2 - 1 \right]$ , and so on.

The model problems (2.5.1) and (2.5.3) have a single Wiener process and they possess the following interesting feature. We observe that their solutions (2.5.2) and (2.5.4) have the form  $u(t, x) = f(t, x, w(t))$ , where  $f(t, x, y)$  is a smooth function. Consequently, the solutions are expandable in the basis consisting just of  $\xi_\alpha = H_k(w(t)/\sqrt{t})/\sqrt{k!} = H_k(\xi_1)/\sqrt{k!}$ ,  $\alpha = (k, 0, \dots, 0)$ ,  $k = 0, 1, \dots$ , i.e., we have

$$u(t, x) = \sum_{\alpha \in \mathcal{J}} \frac{\varphi_\alpha(t, x)}{\sqrt{\alpha!}} \xi_\alpha = \sum_{N=0}^{\infty} \sum_{\alpha \in \mathcal{J}_{N,1}} \frac{\varphi_\alpha(t, x)}{\sqrt{\alpha!}} \xi_\alpha = \sum_{k=0}^{\infty} \frac{\varphi_k(t, x)}{\sqrt{k!}} \eta_k, \quad (2.5.5)$$

where  $\eta_k = \xi_\alpha$  with  $\alpha = (k, 0, \dots, 0)$ ,  $k = 0, 1, \dots$ . Hence

$$u_{N,1}(t, x) = :u_N(t, x) = \sum_{k=0}^N \frac{\varphi_k(t, x)}{\sqrt{k!}} \eta_k, \quad (2.5.6)$$

which corresponds to setting  $n = 1$  in (2.3.1). It is not difficult to show (see also the discussion on error estimates after Algorithm 2.4.1 in Section 2.4) that applying Algorithm 2.4.1 to the model problems (2.5.1) and (2.5.3) is more accurate than in general cases of (2.2.1) (cf. (2.3.10) and (2.3.11) and also (2.4.3)):

$$\|\mathbb{E}[u^2(t, \cdot)] - \mathbb{E}[u_{\Delta, N}^2(t, \cdot)]\|_{L^2} \leq C \frac{(C\Delta)^N}{(N+1)!} \quad (2.5.7)$$

for all sufficiently small  $\Delta > 0$  and a constant  $C > 0$  independent of  $\Delta$  and  $N$  (as before, here we neglected errors arising from truncation of the basis  $\{e_m\}$ ).

For the problems (2.5.1) and (2.5.3), the propagator (2.3.12) takes the form (recall that here the multi-index  $\alpha$  degenerates to  $\alpha = (k, 0, \dots, 0)$ ,  $k = 0, 1, \dots$ ):

$$\partial_t \varphi_0 = a \partial_{xx}^2 \varphi_0, \quad \varphi_0(0, x; \phi) = \phi(x), \quad (2.5.8)$$

$$\partial_t \varphi_k = a \partial_{xx}^2 \varphi_k + \frac{1}{\sqrt{\Delta}} \sigma k \partial_x \varphi_{k-1}, \quad \varphi_k(0, x; 0) = 0, \quad k > 0,$$

and

$$\partial_t \varphi_0 = a \partial_{xx}^2 \varphi_0, \quad \varphi_0(0, x; \phi) = \phi(x), \quad (2.5.9)$$

$$\partial_t \varphi_k = a \partial_{xx}^2 \varphi_k + \frac{1}{\sqrt{\Delta}} \sigma k \varphi_{k-1}, \quad \varphi_k(0, x; 0) = 0, \quad k > 0,$$

respectively. We solve these propagators numerically using the Fourier collocation method with  $M$  nodes in physical space and the Crank–Nicolson time discretization with step  $\delta t$  in time. Denote by  $L_m(x)$ ,  $m = 1, \dots, M$ , the  $m$ -th Lagrangian trigonometric polynomials using  $M$  Fourier collocation nodes, i.e.,  $L_m(x)$  are  $m$ -th order trigonometric polynomials satisfying  $L_m(x_l) = \delta_{m,l}$  and  $x_l = \frac{2\pi}{M}(l-1)$ ,  $l = 1, \dots, M$ . Now, for completeness, we formulate the realization of Algorithm 2.4.1 in the case of the model problems.

**Algorithm 2.5.1.** *For given values of the model parameters  $a$  and  $\sigma$ , choose the algorithm parameters: a number of Fourier collocation nodes  $M$ , a time step  $\delta t$  for solving the propagator (2.5.8) (or (2.5.9)), and a time step  $\Delta$  and the number of Hermite polynomials  $N$ .*

*STEP 1 Solve the propagator (2.5.8) (or (2.5.9)) on the time interval  $[0, \Delta]$  with the initial condition  $\phi(x) = L_k(x)$  using the Fourier collocation method with  $M$  nodes in physical space and the Crank–Nicolson scheme with step  $\delta t$  in time and denote the obtained numerical approximation of  $\varphi_k(\Delta, x_l; L_m)$  as  $\varphi_k^{M, \delta t}(\Delta, x_l; L_m)$ ,  $l, m = 1, \dots, M$ ,  $k = 0, 1, \dots, N$ .*

*STEP 2 Recursively compute the covariance matrices*

$$Q_{lm}(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{M}) := \mathbb{E}[u_{\Delta, \mathbf{N}}^{M, \delta t}(\mathbf{t}_{i-1}, x_l) u_{\Delta, \mathbf{N}}^{M, \delta t}(\mathbf{t}_{i-1}, x_m)], \quad \mathbf{t}_{i-1} = i\Delta, \quad i = 0, \dots, K,$$

of the approximate solution to (2.5.1) (or (2.5.3)):

$$\begin{aligned}
Q_{lm}(0; \mathbf{N}, \mathbf{M}) &= u_0(x_l)u_0(x_m), \quad l, m = 1, \dots, \mathbf{M}, \\
Q_{lm}(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{M}) &= \sum_{k=1}^{\mathbf{N}} \sum_{q=1}^{\mathbf{M}} \sum_{r=1}^{\mathbf{M}} \frac{1}{k!} Q_{qr}(\mathbf{t}_i; \mathbf{N}, \mathbf{M}) \varphi_k^{M, \delta t}(\Delta, x_l; L_q) \varphi_k^{M, \delta t}(\Delta, x_m; L_r), \\
l, m &= 1, \dots, \mathbf{M}, \quad i = 1, \dots, \mathbf{K},
\end{aligned}$$

where  $u_0(x)$  is the initial condition of (2.5.1) (or (2.5.3)).

In particular, we obtain the second moment of the approximate solution to (2.5.1) (or (2.5.3)):

$$E[u_{\Delta, \mathbf{N}}^{M, \delta t}(\mathbf{t}_{i-1}, x_j)]^2 = Q_{jj}(\mathbf{t}_{i-1}; \mathbf{N}, \mathbf{M}), \quad j = 1, \dots, \mathbf{M}, \quad i = 1, \dots, \mathbf{K}.$$

We note that Algorithm 2.5.1 has four errors: (1) an error due to time discretization of the SPDE, which is controlled by  $\Delta$ ; (2) the truncation error of the one-step WCE, which is controlled by  $\mathbf{N}$ ; (3) an error due to the truncation of the spatial basis  $\{e_m\}$ , which is controlled by  $\mathbf{M}$ ; and (4) the numerical integration error in solving the propagator. The last one, in its turn, consists of the error due to space discretization, which is controlled by  $\mathbf{M}$ , and of the error due to time discretization, which is controlled by  $\delta t$ .

**Remark 2.5.2.** To approximate the solution of (2.5.1) (or (2.5.3)), one can use the truncated WCE  $u_{\mathbf{N}}(t, x)$  from (2.5.6) and, in particular, evaluate the second moment  $\mathbb{E}[u^2(t, x)]$  as

$$\mathbb{E}[u^2(t, x)] \approx \mathbb{E}[u_{\mathbf{N}}^2(t, x)] = \sum_{k=0}^{\mathbf{N}} \frac{\varphi_k^2(t, x)}{k!} \approx \sum_{k=0}^{\mathbf{N}} \frac{[\varphi_k^{M, \delta t}(t, x)]^2}{k!}, \quad (2.5.10)$$

where  $\varphi_0(t, x) = \varphi_0(t, x; u_0(x))$  and  $\varphi_k(t, x) = \varphi_k(t, x; 0)$ ,  $k > 0$ , are solutions of the propagator (2.5.8) (or (2.5.9)) and  $\varphi_k^{M, \delta t}(t, x)$  are their numerical approximations obtained, e.g., using the Fourier collocation method with  $\mathbf{M}$  nodes in physical space and the Crank–Nicolson scheme with step

$\delta t$  in time. The approximation (2.5.10) can be viewed as a one-step approximation corresponding to Algorithm 2.5.1, i.e., the first step of Algorithm 2.5.1 with  $\Delta = t$ , and its error is estimated as

$$|\mathbb{E}[u^2(t, \cdot)] - \mathbb{E}[u_{\mathbb{N}}^2(t, \cdot)]|_{L^2} \leq C e^{Ct} \frac{(Ct)^{\mathbb{N}+1}}{(\mathbb{N}+1)!}.$$

We see that this error grows exponentially with  $t$ , which was confirmed by our numerical tests with (2.5.1) (not presented here). To reach a satisfactory accuracy of the approximation (2.5.10) for a fixed  $t$ , one has to take a sufficiently large  $\mathbb{N}$  which is computationally expensive (see also Remark 2.4.2) even in the case of moderate values of  $t$ . In contrast, we demonstrate (see next section) that the error of Algorithm 2.5.1 grows linearly with time and it is relatively small even for  $\mathbb{N} = 1$ .

### 2.5.3 Numerical results

In this section we present some results of our numerical tests of Algorithm 2.5.1 on the two model problems (2.5.1) and (2.5.3).

In approximating the propagators (2.5.8) and (2.5.9) we choose a sufficiently large number of Fourier collocation nodes  $M$  and a sufficiently small time step  $\delta t$  so that errors of numerical solutions to the propagators have a negligible influence on overall accuracy of Algorithm 2.5.1 in our simulations. In all the numerical tests it was sufficient to take  $M = 20$ ; this choice of  $M$  was tested by running control tests with  $M = 80$ .

We measure numerical errors using the norms:

$$\rho_2(t) = \left( \frac{2\pi}{M} \sum_{m=1}^M (\mathbb{E}[u_{\Delta, \mathbb{N}}^{M, \delta t}(t, x_m)]^2) - \mathbb{E}[u^2(t, x_m)] \right)^{1/2},$$

and

$$\rho_\infty(t) = \max_{1 \leq m \leq M} \left| \mathbb{E} \left[ u_{\Delta, \mathbb{N}}^{M, \delta t}(t, x_m) \right]^2 - \mathbb{E} u^2(t, x_m) \right|.$$

The results of our tests on the model problem (2.5.1) in the degenerate case (i.e.,  $\epsilon = 0$ ) and in the non-degenerate case (i.e.,  $\epsilon > 0$ ) are presented in Tables 2.1 and 2.2, respectively. Table 2.3 corresponds to the tests with the second model problem (2.5.3). Numerical tests with values of the parameters other than those used for Tables 2.1-2.3 were also performed and they gave similar results.

Table 2.1: Performance of Algorithm 2.5.1 for Model (2.5.1). The parameters of the model (2.5.1) are  $\sigma = 1$ ,  $\epsilon = 0$ , and the time  $t = 10$ . In Algorithm 2.5.1 we take  $M = 20$ .

N	$\Delta$	$\delta t$	$\rho_2(10)$	$\rho_\infty(10)$
1	0.1	$1 \times 10^{-3}$	$4.69 \times 10^{-1}$	$1.87 \times 10^{-1}$
	0.01	$1 \times 10^{-4}$	$6.07 \times 10^{-2}$	$2.42 \times 10^{-2}$
	0.001	$1 \times 10^{-5}$	$6.25 \times 10^{-3}$	$2.49 \times 10^{-3}$
2	0.1	$1 \times 10^{-3}$	$1.92 \times 10^{-2}$	$7.67 \times 10^{-3}$
	0.01	$1 \times 10^{-4}$	$2.07 \times 10^{-4}$	$8.27 \times 10^{-5}$
	0.001	$1 \times 10^{-5}$	$2.09 \times 10^{-6}$	$8.33 \times 10^{-7}$
3	0.1	$1 \times 10^{-3}$	$4.82 \times 10^{-4}$	$1.99 \times 10^{-4}$
	0.01	$1 \times 10^{-4}$	$5.16 \times 10^{-7}$	$2.06 \times 10^{-7}$
	0.001	$1 \times 10^{-5}$	$3.37 \times 10^{-10}$	$1.81 \times 10^{-10}$
4	0.1	$1 \times 10^{-3}$	$9.36 \times 10^{-6}$	$3.73 \times 10^{-6}$
	0.01	$1 \times 10^{-5}$	$9.35 \times 10^{-10}$	$4.17 \times 10^{-10}$

Table 2.2: *Model (2.5.1): performance of Algorithm 2.5.1.* The parameters of the model (2.5.1) are  $\sigma = 1$ ,  $\epsilon = 0.01$ , and the time  $t = 10$ . In Algorithm 2.5.1 we take  $M = 20$ .

N	$\Delta$	$\delta t$	$\rho_2(10)$	$\rho_\infty(10)$
1	0.1	$1 \times 10^{-3}$	$3.84 \times 10^{-1}$	$1.53 \times 10^{-1}$
	0.01	$1 \times 10^{-4}$	$4.97 \times 10^{-2}$	$1.98 \times 10^{-2}$
	0.001	$1 \times 10^{-4}$	$5.11 \times 10^{-3}$	$2.04 \times 10^{-3}$
2	0.1	$1 \times 10^{-3}$	$1.58 \times 10^{-2}$	$6.28 \times 10^{-3}$
	0.01	$1 \times 10^{-4}$	$1.70 \times 10^{-4}$	$6.77 \times 10^{-5}$
	0.001	$1 \times 10^{-4}$	$1.72 \times 10^{-6}$	$6.88 \times 10^{-7}$
3	0.1	$1 \times 10^{-3}$	$3.95 \times 10^{-4}$	$1.57 \times 10^{-4}$
	0.01	$1 \times 10^{-4}$	$4.22 \times 10^{-7}$	$1.68 \times 10^{-7}$
	0.001	$1 \times 10^{-5}$	$3.65 \times 10^{-10}$	$2.01 \times 10^{-10}$
4	0.1	$1 \times 10^{-3}$	$7.67 \times 10^{-6}$	$3.06 \times 10^{-6}$
	0.01	$1 \times 10^{-5}$	$8.39 \times 10^{-10}$	$3.90 \times 10^{-10}$

Analyzing the results in Tables 2.1, 2.2 and 2.3, we observe the convergence order of  $\Delta^N$  for a fixed  $N$  in all the tests which confirms our theoretical prediction (2.5.7). We also run other cases (not presented here) to confirm the conclusion from Section 2.5.2 that the number  $n$  of random

Table 2.3: Performance of Algorithm 2.5.1 for Model (2.5.3). The parameters of the model (2.5.3) are  $\sigma = 1$ ,  $a = 0.5$ , and the time  $t = 10$ . In Algorithm 2.5.1 we take  $M = 20$ .

N	$\Delta$	$\delta t$	$\rho_2(10)$	$\rho_\infty(10)$
1	0.1	$1 \times 10^{-3}$	$5.75 \times 10^{-1}$	$3.74 \times 10^{-1}$
	0.01	$1 \times 10^{-4}$	$7.44 \times 10^{-2}$	$4.85 \times 10^{-2}$
	0.001	$1 \times 10^{-4}$	$7.65 \times 10^{-3}$	$4.98 \times 10^{-3}$
2	0.1	$1 \times 10^{-3}$	$2.36 \times 10^{-2}$	$1.53 \times 10^{-2}$
	0.01	$1 \times 10^{-4}$	$2.54 \times 10^{-4}$	$1.65 \times 10^{-4}$
	0.001	$1 \times 10^{-4}$	$2.58 \times 10^{-6}$	$1.68 \times 10^{-6}$
3	0.1	$1 \times 10^{-3}$	$5.90 \times 10^{-4}$	$3.85 \times 10^{-4}$
	0.01	$1 \times 10^{-4}$	$6.32 \times 10^{-7}$	$4.12 \times 10^{-7}$

variables  $\xi_k$  used per step does not influence the accuracy of Algorithm 2.4.1 in the case of the model problems (2.5.1) and (2.5.3).

In Figure 2.2 we demonstrate dependence of the relative numerical error

$$\rho_2^r(t) = \frac{\rho_2(t)}{\|\mathbb{E}[u^2(t, \cdot)]\|_{L^2}}$$

on integration time. These results were obtained in the degenerate case of the problem (2.5.1), but similar behavior of errors was observed in our tests with other parameters as well. One can conclude from Figure 2.2 that (after an initial fast growth) the error grows *linearly* with integration time. This is a remarkable feature of the proposed WCE-based algorithm since it implies that the algorithm can be used for long time integration of SPDEs.

#### 2.5.4 Comparison of the WCE algorithm and Monte Carlo-type algorithms

As discussed in Introduction, there are other approaches to solving SPDEs, which are usually complimented by the Monte Carlo technique when one is interested in computing moments of SPDE solutions. In this section, using the problem (2.5.1), we compare the performance of Algorithm 2.5.1 and two Monte Carlo-type algorithms, one of which is based on the method of characteristics [304] and another on the Fourier transform of the linear SPDE with subsequent simulation of SDEs and

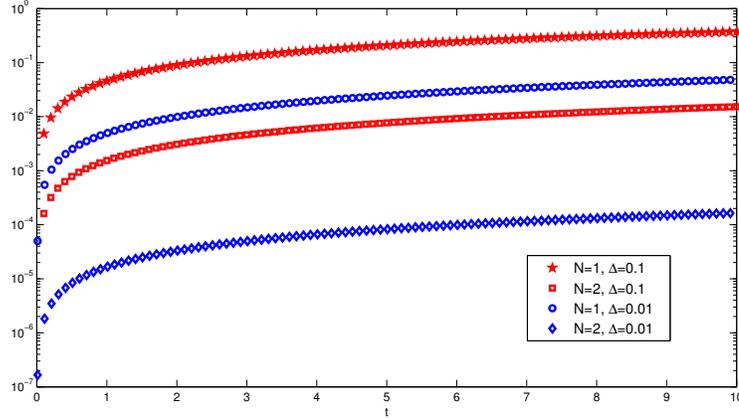


Figure 2.2: Dependence of the relative numerical error  $\rho_2^r(t)$  on integration time. Model (2.5.1) is simulated by Algorithm 2.5.1 with  $M = 20$  and  $\delta t = \Delta/100$  and various  $\Delta$  and  $N$ . The parameters of (2.5.1) are  $\sigma = 1$  and  $\epsilon = 0$ .

application of the Monte Carlo technique.

The solution of (2.5.1) with  $\epsilon = 0$  (the degenerate case) can be represented using the method of characteristics [347]:

$$u(t, x) = \sin(X_{t,x}(0)), \quad (2.5.11)$$

where  $X_{t,x}(s)$ ,  $0 \leq s \leq t$ , is the solution of the system of backward characteristics

$$dX_{t,x}(s) = \sigma \overleftarrow{dw}(s), \quad X_{t,x}(t) = x. \quad (2.5.12)$$

The notation “ $\overleftarrow{dw}(s)$ ” means backward Itô integral (see, e.g. [347]). It follows from (2.5.12) that  $X_{t,x}(0)$  has the same probability distribution as  $x + \sigma\sqrt{t}\zeta$ , where  $\zeta$  is a standard Gaussian random variable (i.e.,  $\zeta \sim \mathcal{N}(0,1)$ ). Since we are interested only in computing statistical moments, it is assumed, without loss of generality, that

$$X_{t,x}(0) = x + \sigma\sqrt{t}\zeta. \quad (2.5.13)$$

Then we can estimate the second moment  $m_2(t, x) := \mathbb{E}[u^2(t, x)]$  as

$$m_2(t, x) \doteq \hat{m}_2(t, x) = \frac{1}{L} \sum_{l=1}^L \sin^2(x + \sigma\sqrt{t}\zeta^{(l)}), \quad (2.5.14)$$

where  $\zeta^{(l)}$ ,  $l = 1, \dots, L$ , are i.i.d. standard Gaussian random variables. The estimate  $\hat{m}_2$  for  $m_2$  is unbiased, and, hence, the numerical procedure for finding  $m_2$  based on (2.5.14) has only the Monte Carlo (i.e., statistical) error which, as usual, can be quantified via half of the length of the 95% confidence interval:

$$\rho_{MC}(t, x) = 2 \frac{\sqrt{\text{Var}(\sin^2(x + \sigma\sqrt{t}\zeta))}}{\sqrt{L}}.$$

Table 2.4 gives the statistical error for  $\hat{m}_2(t, x)$  from (2.5.14) (recall that there is no space or time discretization error in this algorithm), which is computed as

$$2 \cdot \max_j \frac{\sqrt{\frac{1}{L} \sum_{l=1}^L \sin^4(x_j + \sigma\sqrt{t}\zeta^{(l)}) - [\hat{m}_2(t, x_j)]^2}}{\sqrt{L}}, \quad (2.5.15)$$

where the set of  $x_j$  is the same as the one used for producing the results of Table 2.5 by Algorithm 2.5.1 and  $\zeta^{(l)}$  are as in (2.5.14). All the tests were run using Matlab R2007b, on a Macintosh desktop computer with Intel Xeon CPU E5462 (quad-core, 2.80 GHz). Every effort was made to program and execute the different algorithms as much as possible in an identical way. The cost of simulation due to (2.5.14) is directly proportional to  $L$ . The slower time increase for smaller  $L$  in Table 2.4 is due to inclusion of the initialization time of the computer program in the time measurement.

In Table 2.5 we repeat some of the results already presented in Table 2.1, which are now also accompanied by CPU time for comparison.

Comparing the results in Tables 2.4 and 2.5, we conclude that when one sets a relatively large error tolerance level the estimate  $\hat{m}_2(t, x)$  from (2.5.14) is computationally more efficient than Algorithm 2.5.1; however, Algorithm 2.5.1 has lower costs in reaching a higher accuracy (errors

Table 2.4: Performance of the method (2.5.14) for Model (2.5.1). The parameters of the model (2.5.1) are  $\sigma = 1$ ,  $\epsilon = 0$ , and the time  $t = 10$ . The statistical error is computed according to (2.5.15).

$L$	statistical error	CPU time (sec.)
$10^2$	$8.87 \times 10^{-2}$	$6 \times 10^{-3}$
$10^4$	$7.40 \times 10^{-3}$	$6.7 \times 10^{-2}$
$10^6$	$7.09 \times 10^{-4}$	$7.4 \times 10^0$
$10^8$	$7.07 \times 10^{-5}$	$7.4 \times 10^2$
$10^{10}$	$7.07 \times 10^{-6}$	$7.3 \times 10^4$

Table 2.5: Performance of Algorithm 2.5.1 for Model (2.5.1). The parameters of the model (2.5.1) are  $\sigma = 1$ ,  $\epsilon = 0$ , and the time  $t = 10$ . The parameters of Algorithm 2.5.1 are  $\Delta = 0.1$ ,  $M = 20$ ,  $\delta t = 0.001$ .

N	$\rho_\infty(10)$	CPU time (sec.)
1	$1.87 \times 10^{-1}$	$5.7 \times 10^0$
2	$7.67 \times 10^{-3}$	$8.1 \times 10^0$
3	$1.99 \times 10^{-4}$	$1.1 \times 10^1$
4	$3.73 \times 10^{-6}$	$1.3 \times 10^1$

of order equal to or smaller than  $10^{-6}$ ). We note that variance reduction techniques (see, e.g. [301, 304] and the references therein) can be used in order to reduce the Monte Carlo error. But the aim here is to give a comparison of computational costs for the WCE-based algorithm and direct Monte Carlo methods having in mind that for complex stochastic problems it is usually rather difficult to reduce variance efficiently.

Let us now use the problem (2.5.1) with  $\epsilon = 0$  for comparison of Algorithm 2.5.1 with another approach exploiting the Monte Carlo technique. One can represent the solution of this periodic problem via the Fourier transform:

$$u(t, x) = \sum_{k \in \mathbb{Z}} e^{ikx} u_k(t) \quad (2.5.16)$$

with  $u_k(t)$ ,  $t \geq 0$ ,  $k \in \mathbb{Z}$ , satisfying the system of SDEs:

$$du_k(t) = -k^2 \frac{1}{2} \sigma^2 u_k(t) dt + ik \sigma u_k(t) dw(t), \quad \operatorname{Re} u_k(0) = 0, \quad \operatorname{Im} u_k(0) = \frac{1}{2} (\delta_{1k} - \delta_{-1k}). \quad (2.5.17)$$

Noting that here  $u_k(t) \equiv 0$  for all  $|k| \neq 1$  and re-writing (2.5.16)-(2.5.17) in the trigonometric form, we get

$$u(t, x) = u^c(t) \cos x + u^s(t) \sin x, \quad (2.5.18)$$

where

$$\begin{aligned} du^c(t) &= -\frac{1}{2}\sigma^2 u^c(t)dt + \sigma u^s(t)dw(t), \quad u^c(0) = 0, \\ du^s(t) &= -\frac{1}{2}\sigma^2 u^s(t)dt - \sigma u^c(t)dw(t), \quad u^s(0) = 1. \end{aligned} \quad (2.5.19)$$

The system (2.5.19) is a Hamiltonian system with multiplicative noise (see, e.g. [299, 301]). It is known [299, 301] that symplectic integrators have advantages in comparison with usual numerical methods in long time simulations of stochastic Hamiltonian systems. An example of a symplectic method is the midpoint scheme, which in application to (2.5.19) takes the following form:

$$\begin{aligned} \bar{u}^c(t_{k+1}) &= \bar{u}^c(t_k) + \frac{\sigma}{2}(\bar{u}^s(t_k) + \bar{u}^s(t_{k+1}))\sqrt{\Delta t}\zeta_{k+1}, \quad u^c(0) = 0, \\ \bar{u}^s(t_{k+1}) &= \bar{u}^s(t_k) - \frac{\sigma}{2}(\bar{u}^c(t_k) + \bar{u}^c(t_{k+1}))\sqrt{\Delta t}\zeta_{k+1}, \quad u^s(0) = 1, \end{aligned} \quad (2.5.20)$$

where  $\zeta_k$  are i.i.d. standard Gaussian random variables and  $\Delta t > 0$  is a time step. The scheme (2.5.20) converges with the mean-square order 1/2 and weak order 1 [301]. It is implicit but can be resolved analytically since we are dealing with the linear system here. One can recognize that (2.5.19) is a Kubo oscillator. A number of numerical tests with symplectic and non-symplectic integrators are done on a Kubo oscillator in [299, 301].

Using (2.5.18) and (2.5.20), we evaluate the second moment of the solution to (2.5.1) with  $\epsilon = 0$  as

$$m_2(t_k, x) := \mathbb{E}[u^2(t_k, x)] \doteq \mathbb{E}[(\bar{u}^c(t_k) \cos x + \bar{u}^s(t_k) \sin x)^2] \quad (2.5.21)$$

$$\doteq \hat{m}_2(t_k, x) = \frac{1}{L} \sum_{l=1}^L \left[ \bar{u}^{c,(l)}(t_k) \cos x + \bar{u}^{s,(l)}(t_k) \sin x \right]^2,$$

where  $\bar{u}^{c,(l)}(t_k)$ ,  $\bar{u}^{s,(l)}(t_k)$  are independent realizations of the random variables  $\bar{u}^c(t_k)$ ,  $\bar{u}^s(t_k)$ .

The estimate  $\hat{m}_2(t_k, x)$  from (2.5.21) has two errors: the time discretization error due to the approximation of (2.5.19) by (2.5.20) and the Monte Carlo error. The errors presented in Table 2.6 are computed as  $\max_j [\bar{m}_2(t_k, x_j) - \mathbb{E}[u^2(t_k, x_j)]]$  and are given together with the 95% confidence interval.

Table 2.6: *Model (2.5.1): performance of the method (2.5.21)*. The parameters of the model (2.5.1) are  $\sigma = 1$ ,  $\epsilon = 0$ , and the time  $t = 10$ .

$\Delta t$	$L$	Error	CPU time (sec.)
0.1	$10^4$	$8.06 \times 10^{-3} \pm 7.09 \times 10^{-3}$	$4.72 \times 10^{-1}$
0.01	$10^4$	$6.55 \times 10^{-4} \pm 7.08 \times 10^{-4}$	$3.90 \times 10^2$
0.001	$10^6$	$8.81 \times 10^{-5} \pm 7.07 \times 10^{-5}$	$3.81 \times 10^5$

Comparing the results in Tables 2.6 and 2.5, we come to the same conclusion as in our first comparison test that Algorithm 2.5.1 is computationally more efficient than the Monte Carlo-based algorithms in reaching a higher accuracy.

## 2.6 Numerical tests with passive scalar equation

A prominent example of the stochastic advection-diffusion equation (2.2.1)-(2.2.2) is a passive scalar equation, which is motivated by the study of the turbulent transport problem (see [123, 233, 264] and the references therein). Here we perform numerical tests on the *two-dimensional* ( $d = 2$ ) passive scalar equation with periodic boundary conditions:

$$du(t, x) + \sum_{k=1}^{\infty} \sum_{i=1}^d \sigma_k^i(x) D_i u \circ dw_k(t) = 0, \quad (2.6.1)$$

$$u(t, x^1 + \ell, x^2) = u(t, x^1, x^2 + \ell) = u(t, x), \quad t > 0, \quad x \in (0, \ell)^2,$$

$$u(0, x) = u_0(x), \quad x \in (0, \ell)^2,$$

where  $\circ$  indicates the Stratonovich version of stochastic integration,  $\ell > 0$ , the initial condition  $u_0(x)$  is a periodic function with the period  $(0, \ell)^2$ , and  $\sigma_k^i(x)$  are divergence-free periodic functions with the period  $(0, \ell)^2$ :

$$\operatorname{div} \sigma_k = 0. \quad (2.6.2)$$

In (2.6.1) we take a combination of such  $\sigma_k(x)$  so that the corresponding spatial covariance  $C$  is symmetric and stationary:  $C(x-y) = \sum_{k=1}^{\infty} \lambda_k \sigma_k(x) \sigma_k^T(y)$ , where  $\lambda_k$  are some non-negative numbers.

Namely, we consider

$$C(x-y) = \sum_{k=1}^{\infty} \lambda_k C(x-y; n_k, m_k), \quad (2.6.3)$$

where  $n_k, m_k$  is a sequence of positive integers, and

$$C(x-y; n, m) = \cos(2\pi (n[x^1 - y^1] + m[x^2 - y^2]) / \ell) \begin{bmatrix} m^2 & -nm \\ -nm & n^2 \end{bmatrix},$$

which can be decomposed as

$$\begin{aligned} C(x-y; n, m) &= \cos(2\pi[nx^1 + mx^2]/\ell) \begin{bmatrix} -m \\ n \end{bmatrix} \cos(2\pi [ny^1 + my^2]/\ell) \begin{bmatrix} -m & n \end{bmatrix} \\ &+ \sin(2\pi [nx^1 + mx^2]/\ell) \begin{bmatrix} -m \\ n \end{bmatrix} \sin(2\pi [ny^1 + my^2]/\ell) \begin{bmatrix} -m & n \end{bmatrix}. \end{aligned}$$

Hence,  $\{\sigma_k(x)\}_{k \geq 1}$  in (2.6.1) is an appropriate combination of vector functions of the form

$$\cos(2\pi[nx^1 + mx^2]/\ell) \begin{bmatrix} -m \\ n \end{bmatrix} \text{ and } \sin(2\pi [nx^1 + mx^2]/\ell) \begin{bmatrix} -m \\ n \end{bmatrix}.$$

We rewrite (2.6.1) in the Ito's form:

$$\begin{aligned}
du(t) + \frac{1}{2} \sum_{i,j=1}^d C_{ij}(0) D_i D_j u dt + \sum_{k=1}^{\infty} \sum_{i=1}^d \sigma_k^i(x) D_i u dw_k(t) &= 0, \\
u(t, x^1 + \ell, x^2) = u(t, x^1, x^2 + \ell) = u(t, x), \quad t > 0, \quad x \in (0, \ell)^2, \\
u(0, x) = u_0(x), \quad x \in (0, \ell)^2.
\end{aligned} \tag{2.6.4}$$

Below we present results of numerical tests of Algorithm 2.4.1 applied to (2.6.4) and its comparison with the Monte Carlo-type algorithm based on the method of characteristics from [304]. In the tests we simulated the  $L^2$ -norm of the second moment of the SPDE solution

$$\|\mathbb{E}[u^2(T, \cdot)]\|_{L^2} = \left[ \int_{[0, \ell]^2} (\mathbb{E}[u^2(T, x)])^2 dx \right]^{1/2}. \tag{2.6.5}$$

We considered the particular case of (2.6.1), (2.6.3) with  $\ell = 2\pi$ , the initial condition

$$u_0(x) = \sin(2x^1) \sin(x^2). \tag{2.6.6}$$

and with two noise terms:

$$\begin{aligned}
\sigma_1(x) = \cos(x_1 + x_2) \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \sigma_2(x) = \sin(x_1 + x_2) \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \\
\sigma_k(x) = 0 \quad \text{for } k > 2.
\end{aligned} \tag{2.6.7}$$

This example satisfies the so-called commutativity condition, see e.g. [211, 301]. The error estimate (2.3.10) holds in this case, which is confirmed in the tests, (see Chapter 4 for error estimates for single noise, one special case of commutative noises).

In Algorithm 2.4.1 we solve the propagator (2.3.12) corresponding to the SPDE (2.6.4) using fourth-order explicit Runge–Kutta with step size  $\delta t$  in time and the Fourier spectral method with

$M$  modes in physical space. We noted in Remark 2.4.2 that in general the computational cost of Algorithm 2.4.1 is proportional to  $M^4$  but with an appropriate choice of basis functions, this cost can be considerably reduced. Indeed, the Fourier basis is natural for the problem (2.6.4) and the use of this basis reduces the computational cost to being proportional to  $M^2$ . This significant reduction is based on the following observation. Since we consider finite number of noises with periodic  $\sigma_k(x)$  and  $\varphi_\alpha(\Delta, x; e_l)$  to be the solution of the propagator (2.3.12) with the initial condition equal to a single basis function  $e_l(x)$ ,  $\varphi_\alpha(\Delta, x; e_l)$  is expandable in a finite number of periodic functions  $e_k(x)$  and this number does not depend on  $M$ . Hence for fixed  $\alpha$  and  $l$  the number of nonzero  $q_{\alpha,l,m} = (\varphi_\alpha(\Delta, \cdot; e_l), e_m(\cdot))$  is finite. Therefore, the overall number of nonzero  $q_{\alpha,l,m}$  is proportional to  $M$  instead of  $M^2$ . This was tested and confirmed in our tests. We use the above fact in our computer realization of Algorithm 2.4.1 and reduce the computational cost of obtaining a single entry of the matrix  $Q_{l,m}$  from the order of  $\mathcal{O}(M^2)$  to order  $\mathcal{O}(1)$ . Hence, computational costs of Step 3 (and hence that of Algorithm 2.4.1) becomes proportional to  $M^2$  instead of the original  $M^4$ .

We do not have an exact solution of the problem (2.6.1) and hence we need a reference solution. To this end, the  $L_2$ -norm of the second moment of the SPDE solution at  $T = 1$  was computed by Algorithm 2.4.1 with parameters  $N = 2$ ,  $n = 1$ ,  $M = 900$  (i.e., 30 basis functions in each space direction),  $\delta t = 1 \times 10^{-5}$ ,  $\Delta = 1 \times 10^{-4}$  and which is equal to 1.57976 (5 d.p.). This result was also verified by the Monte Carlo-type method described below with  $\Delta t = 1 \times 10^{-3}$ ,  $M_s = 10$  and  $L = 8 \times 10^7$ , which gave  $1.579777 \pm 7.6 \times 10^{-5}$  where  $\pm$  reflects the Monte Carlo error only.

For Algorithm 2.4.1, we measure the error of computing the  $L^2$ -norm of the second moment of the SPDE solution as follows

$$\rho(T) = \left\| \mathbb{E}[u_{\text{ref}}^2(T, \cdot)] \right\|_{l^2} - \left\| \mathbb{E}[(u_{\Delta, N}^{M, \delta t}(T, \cdot))^2] \right\|_{l^2},$$

where  $\|v(\cdot)\|_{l^2} = \frac{\ell}{M_s} \left( \sum_{i,j=1}^{M_s} v^2(x_i^1, x_j^2) \right)^{1/2}$ ,  $x_i^1 = x_i^2 = (i-1)\ell/M_s$ ,  $i = 1, \dots, M_s$ , and  $\mathbb{E}[u_{\text{ref}}^2(T, \cdot)]$

is the reference solution computed as explained above. The results demonstrating second-order convergence (see (2.3.10) and the discussion after Algorithm 2.4.1 ) are given in Table 2.7. We note that we also did some control tests with  $\delta t = 1 \times 10^{-5}$  and  $M = 1600$  which showed that the errors presented in this table are not essentially influenced by the errors caused by the choice of  $\delta t = 1 \times 10^{-4}$  and cut-off of the basis at  $M = 900$ .

Table 2.7: *Performance of Algorithm 2.4.1 for passive scalar equation (2.6.4).* The parameters of Algorithm 2.4.1 are  $N = 2$ ,  $n = 1$ ,  $M = 900$ ,  $\delta t = 1 \times 10^{-4}$ .

$\Delta$	$\rho(1)$
0.05	0.1539
0.02	0.0326
0.01	0.0089
0.005	0.0023
0.0025	0.0006

Let us now describe the Monte Carlo-type algorithm based on the method of characteristics (see further details in [304]) with which we compare here the performance of Algorithm 2.4.1. The solution  $u(t, x)$  of (2.6.1) has the following (conditional) probabilistic representation (see [347, 264]):

$$u(t, x) = u_0(X_{t,x}(0)) \quad \text{a.s.}, \quad (2.6.8)$$

where  $X_{t,x}(s)$ ,  $0 \leq s \leq t$ , is the solution of the system of (backward) characteristics

$$-dX = \sum_k \sigma_k(X) \overleftarrow{dw}_k(s), \quad X(t) = x. \quad (2.6.9)$$

Due to (2.6.2) and (see [264])

$$\sum_k \frac{\partial \sigma_k}{\partial x} \sigma_k = 0, \quad (2.6.10)$$

the phase flow of (2.6.9) preserves phase volume (see e.g. [301, p. 247, Eq. (5.5)]). We also recall that the Ito and Stratonovich forms of (2.6.9) coincide. As it is known [301], it is beneficial to approximate (2.6.9) using phase volume preserving schemes, e.g., by the midpoint method [301,

Chapter 4], which for (2.6.9) takes the form (here we exploited that the Ito and Stratonovich forms of (2.6.9) coincide): for an integer  $m \geq 1$ ,

$$\begin{aligned} X_m &= x, \\ X_l &= X_{l+1} + \sum_k \sigma_k \left( \frac{X_l + X_{l+1}}{2} \right) (\zeta_k^{\Delta t})_l \sqrt{\Delta t}, \quad l = n-1, \dots, 0, \end{aligned} \tag{2.6.11}$$

where  $(\zeta_k^{\Delta t})_l$  are, e.g., i.i.d. random variables with the law

$$\zeta_k^{\Delta t} = \begin{cases} \xi_k, & |\xi_k| \leq A_{\Delta t}, \\ A_{\Delta t}, & \xi_k > A_{\Delta t}, \\ -A_{\Delta t}, & \xi_k < -A_{\Delta t}, \end{cases} \tag{2.6.12}$$

and  $\xi_k$  are i.i.d standard Gaussian random variables, and  $A_{\Delta t} = \sqrt{2c|\ln \Delta t|}$ ,  $c \geq 1$ . Its weak order is equal to one [301]. This scheme requires solving the two-dimensional nonlinear equation at each step. To solve it, we used the fixed-point method with the level of tolerance  $10^{-13}$ , and in our example two fixed-point iterations were sufficient to reach this accuracy. Using  $\bar{X}_{t,x}(0) = X_0$  obtained by (2.6.11) with  $\Delta t = T/m$ , we simulate the  $L^2$ -norm of the second moment of the SPDE solution as follows:

$$\begin{aligned} \|\mathbb{E}[u^2(T, \cdot)]\|_{L^2} &= \left( \int_{[0, \ell]^2} (\mathbb{E}[u^2(T, x)])^2 dx \right)^{1/2} \approx \|\mathbb{E}[u^2(T, \cdot)]\|_{l^2} \\ &= \frac{\ell}{M_s} \left[ \sum_{i,j=1}^{M_s} \left( \mathbb{E}[u_0^2(X_{T,x_i^1, x_j^2}(0))] \right)^2 \right]^{1/2} \approx \frac{\ell}{M_s} \left[ \sum_{i,j=1}^{M_s} (\mathbb{E}[u_0^2(\bar{X}_{T,x_i^1, x_j^2}(0))]^2 \right]^{1/2} \\ &\approx \frac{\ell}{M_s} \left[ \sum_{i,j=1}^{M_s} \left[ \frac{1}{L} \sum_{l=1}^L u_0^2(\bar{X}_{T,x_i^1, x_j^2}^{(l)}(0)) \right]^2 \right]^{1/2}, \end{aligned} \tag{2.6.13}$$

where  $x_i^1 = x_i^2 = (i-1)\ell/M_s$ ,  $i = 1, \dots, M_s$ ;  $\bar{X}_{t,x_i^1, x_j^2}^{(l)}(0)$  are independent realizations of the random variables  $\bar{X}_{t,x_i^1, x_j^2}(0)$ . The approximation in (2.6.13) has three errors: (i) the error of discretization

of the integral of the space domain  $[0, \ell]^2$  which is negligible in our example even for  $M_s = 10$ ; (ii) the error of numerical integration due to replacement of  $X_{t,x_i^1,x_j^2}(0)$  by  $\bar{X}_{t,x_i^1,x_j^2}(0)$ ; (iii) the Monte Carlo error which is measured analogously to how it was done in Section 2.5.4. We note that it is possible to reduce the variance of the estimator on the right-hand side of (2.6.13) but we do not consider it here. It is interesting that the mid-point scheme used to simulate  $\bar{X}_{t,x_i^1,x_j^2}(0)$  gave very accurate results even with relatively large time steps.

We compare Algorithm 2.4.1 and the Monte Carlo algorithm (2.6.13) by simulating the example (2.6.1), (2.6.6), (2.6.7) at  $T = 1$ . In these comparison tests, Matlab R2010b was used for each test on a single core of two Intel Xeon 5540 (2.53 GHz) quad-core Nehalem processors. From Tables 2.8 and 2.9, we can draw the same conclusion as in one dimension that for lower accuracy the Monte Carlo algorithm (2.6.13) outperforms Algorithm 2.4.1; but that Algorithm 2.4.1 is more efficient for obtaining higher accuracy.

Table 2.8: Performance of Algorithm 2.4.1 passive scalar equation (2.6.4). The parameters of Algorithm 2.4.1 are  $N = 2$ ,  $n = 1$ ,  $M = 900$ ,  $\delta t = 1 \times 10^{-4}$ .

$\Delta$	$\rho(1)$	CPU time
$1 \times 10^{-2}$	$8.89 \times 10^{-3}$	$3.7 \times 10^4$ (sec.)
$1 \times 10^{-3}$	$1.20 \times 10^{-4}$	$3.2 \times 10^5$ (sec.)
$5 \times 10^{-4}$	$3.73 \times 10^{-5}$	$1.8 \times 10^2$ (hours)

Table 2.9: Performance of Algorithm 2.6.11 for passive scalar equation (2.6.4). The parameter is  $M = 100$ .

$\Delta t$	$L$	Error	CPU time
$2 \times 10^{-1}$	$2.5 \times 10^4$	$4.68 \times 10^{-3} \pm 4.38 \times 10^{-3}$	$1.2 \times 10^1$ (sec.)
$1 \times 10^{-2}$	$4 \times 10^7$	$1.46 \times 10^{-4} \pm 1.08 \times 10^{-4}$	$3.5 \times 10^5$ (sec.)
$1 \times 10^{-3}$	$4 \times 10^8$	$\sim \times 10^{-5} \pm 3.03 \times 10^{-5}$	$9.7 \times 10^3$ (hours) <sup>1</sup>

## 2.7 Summary

We have developed a multistage Wiener chaos expansion (WCE) method for advection-diffusion-reaction equations with multiplicative noise, which form a wide class of linear parabolic SPDEs.

<sup>1</sup>This is an estimated time according to the tests with smaller  $\Delta t$ ,  $L$  and with  $M = 100$ .

We complemented this method by a deterministic algorithm for computing second moments of the SPDE solutions without any use of the Monte Carlo technique. Our numerical tests demonstrated that the proposed WCE-based deterministic algorithm can be more efficient than Monte Carlo-type methods in obtaining results of higher accuracy, scaling as  $\Delta^N$ , where  $\Delta$  is the time-step of the “online” integration and  $N$  is the order of Wiener chaos. We have also found that for obtaining results of lower accuracy, Monte Carlo-type methods outperform the deterministic algorithm for computing moments even in the one-dimensional case. The proposed WCE-based algorithm is conceptually different from Monte Carlo-type methods and thus it can be used for independent verification of results obtained by Monte Carlo solvers. The efficiency of the algorithm can be greatly improved if it is combined with reduced-order methods so that only a handful of modes will be required to represent the solution accurately in physical space, i.e., a case with small  $M$ .

Further work is required to extend the theoretical analysis of [265] to the stochastic advection-diffusion-reaction equations we have considered here as well as to weak convergence for WCE-based algorithms. The numerical experiments in Section 2.6 with periodic passive scalar equation were motivated by a non-viscous transport equation with Kraichnan’s velocity, which corresponds to an SPDE with less a regular solution than the one we simulated in Section 2.6. Though we obtained promising results in our numerical tests, simulation of the passive scalar equation with Kraichnan’s velocity requires special consideration. These aspects will be addressed in the future.

## Chapter 3

# Stochastic collocation methods for differential equations with white noise

In this chapter, we consider a sparse grid collocation method in conjunction with a time discretization of the differential equations for computing expectations of functionals of solutions to differential equations perturbed by time-dependent white noise. We first analyze the error of Smolyak's sparse grid collocation used to evaluate expectations of functionals of solutions to stochastic differential equations discretized by the Euler scheme. We show theoretically and numerically that this algorithm can have satisfactory accuracy for small magnitude of noise or small integration time, however it does not converge neither with decrease of the Euler scheme's time step size nor with increase of Smolyak's sparse grid level. Subsequently, we use this method as a building block for proposing a new algorithm by combining sparse grid collocation with a recursive procedure. This approach allows us to numerically integrate linear stochastic partial differential equations over longer times, which is illustrated in numerical tests on a stochastic advection-diffusion equation.

## 3.1 Introduction

In a number of applications from physics, financial engineering, biology and chemistry it is of interest to compute expectations of some functionals of solutions of ordinary stochastic differential equations (SDE) and stochastic partial differential equations (SPDE) driven by white noise. Usually, evaluation of such expectations requires to approximate solutions of stochastic equations and then to compute the corresponding averages with respect to the approximate trajectories. We will not consider the former in this chapter (see, e.g. [301] and references therein) and will concentrate on the latter. The most commonly used approach for computing the averages is the Monte Carlo technique, which is known for its slow rate of convergence and hence limiting computational efficiency of stochastic simulations. To speed up computation of the averages, variance reduction techniques (see, e.g. [301, 304] and the references therein), quasi-Monte Carlo algorithms [320, 356], and the multi-level Monte Carlo method [134, 135] have been proposed and used.

An alternative approach to computing the averages is (stochastic) collocation methods in random space, which are deterministic methods in comparison with the Monte Carlo-type methods that are based on a statistical estimator of a mean. The expectation can be viewed as an integral with respect to the measure corresponding to approximate trajectories. In stochastic collocation methods, one uses (deterministic) high-dimensional quadratures to evaluate these integrals. In the context of uncertainty quantification where moments of stochastic solutions are sought, collocation methods and their close counterparts (e.g., Wiener chaos expansion-based methods) have been very effective in reducing the overall computational cost in engineering problems, see e.g. [132, 368, 404].

Stochastic equations or differential equations with randomness can be split into differential equations perturbed by time-independent noise and by time-dependent noise. It has been demonstrated in a number of works (see e.g. [10, 34, 9, 403, 314, 322, 416] and references therein) that stochastic collocation methods can be a competitive alternative to the Monte Carlo technique and its variants in the case of differential equations perturbed by time-independent noise. The success of these

methods relies on smoothness in the random space and can usually be achieved when it is sufficient to consider only a limited number of random variables (i.e., in the case of a low dimensional random space). The small number of random variables significantly limits the applicability of stochastic collocation methods to differential equations perturbed by time-dependent noise as, in particular, it will be demonstrated in this chapter.

The class of stochastic collocation methods for SDE with time-dependent white noise includes cubatures on Wiener space [269], derandomization [317], optimal quantization [331, 332] and sparse grids of Smolyak type [130, 131, 145]. While derandomization and optimal quantization aim at finding quadrature rules which are in some sense optimal for computing a particular expectation under consideration, cubatures on Wiener space and a stochastic collocation method using Smolyak sparse grid quadratures (a sparse grid collocation method, SGC) use pre-determined quadrature rules in a universal way without being tailed towards a specific expectation. Since SGC is endowed with negative weights, it is, in practice, different from cubatures on Wiener space, where only quadrature rules with positive weights are used. Among quadrature rules, SGC is of particular interest due to its computational convenience. It has been considered in computational finance [130, 145], where high accuracy was observed. We note that the use of SGC in [130, 145] relies on exact sampling of geometric Brownian motion and of solutions of other simple SDE models, i.e., SGC in these works was not studied in conjunction with SDE approximations.

In this chapter, we consider a SGC method accompanied by time discretization of differential equations perturbed by time-dependent noise. Our objective is twofold. *First*, using both analytical and numerical results, we warn that straightforward carrying over stochastic collocation methods and, in particular, SGC to the case of differential equations perturbed by time-dependent noise (SDE or SPDE) usually leads to a failure. The main reason for this failure is that when integration time increases and/or time discretization step decreases, the number of random variables in approximation of SDE and SPDE grows quickly. The number of collocation points required for sufficient accuracy of collocation methods grows exponentially with the number of random variables. This

results in non-convergence of algorithms based on SGC and SDE time discretizations. Further, due to empirical evidence (see e.g. [335]), the use of SGC is limited to problems with random space dimensionality of up to 40. Consequently, SGC algorithms for differential equations perturbed by time-dependent noise can be used only over small time intervals unless a cure for its fundamental limitation is found.

In Section 2 (after brief introduction to the sparse grid of Smolyak [358] (see also [398, 131, 403]) and to the weak-sense numerical integration for SDE (see, e.g. [301])), we obtain an error estimate for a SGC method accompanied by the Euler scheme for evaluating expectations of smooth functionals of solutions of a scalar linear SDE with additive noise. In particular, we conclude that the SGC can successfully work for a small magnitude of noise and relatively short integration time while it does not converge neither with decrease of the time discretization step used for SDE approximation nor with increase of the level of Smolyak’s sparse grid. Numerical tests in Section 3.4 confirm our theoretical conclusions and we also observe first-order convergence in time step size of the algorithm using the SGC method as long as the SGC error is small relative to the error of time discretization of SDE. We note that our conclusion is, to some extent, similar to that for cubatures on Wiener space [61], for Wiener chaos method [192, 265, 266, 420] and some other functional expansion approaches [49, 50].

The *second objective* of the chapter is to suggest a possible cure for the aforementioned deficiencies, which prevent SGC to be used over longer time intervals. For longer time simulation, deterministic replacements (such as stochastic collocation methods and functional expansion methods) of the Monte Carlo technique in simulation of differential equations perturbed by time-dependent noise do not work effectively unless some restarting strategies allowing to ‘forget’ random variables from earlier time steps are employed. Examples of such strategies are the recursive approach for Wiener chaos expansion methods to compute moments of solutions to linear SPDE [265, 420] and an approach for cubatures on Wiener space based on compressing the history data via a regression at each time step [253].

Here we exploit the idea of the recursive approach to achieve accurate longer time integration by numerical algorithms using the SGC. For linear SPDE with *time-independent* coefficients, the recursive approach works as follows. We first find an approximate solution of an SPDE at a relatively small time  $t = h$ , and subsequently take the approximation at  $t = h$  as the initial value in order to compute the approximate solution at  $t = 2h$ , and so on, until we reach the final integration time  $T = Nh$ . To find second moments of the SPDE solution, we store a covariance matrix of the approximate solution at each time step  $kh$  and recursively compute the first two moments. Such an algorithm is proposed in Section 3.3; in Section 3.4 we demonstrate numerically that this algorithm converges in time step  $h$  and that it can work well on longer time intervals. At the same time, a major challenge remains: how to effectively use restarting strategies for SGC in the case of nonlinear SDE and SPDE and further work is needed in this direction.

## 3.2 Sparse grid for weak integration of SDE

### 3.2.1 Smolyak's sparse grid

Sparse grid quadrature is a certain reduction of product quadrature rules which decreases the number of quadrature nodes and allows effective integration in moderately high dimensions [358] (see also [398, 324, 131]). Here we introduce it in the form suitable for our purposes.

We will be interested in evaluating  $d$ -dimensional integrals of a function  $\varphi(y)$ ,  $y \in \mathbb{R}^d$ , with respect to a Gaussian measure:

$$I_d \varphi := \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \varphi(y) \exp\left(-\frac{1}{2} \sum_{i=1}^d y_i^2\right) dy_1 \cdots dy_d. \quad (3.2.1)$$

Consider a sequence of one-dimensional Gauss–Hermite quadrature rules  $Q_n$  with number of nodes

$n \in \mathbb{N}$  for univariate functions  $\psi(\mathbf{y})$ ,  $\mathbf{y} \in \mathbb{R}$ :

$$Q_n \psi(\mathbf{y}) = \sum_{\alpha=1}^n \psi(y_{n,\alpha}) w_{n,\alpha}, \quad (3.2.2)$$

where  $y_{n,1} < y_{n,2} < \dots < y_{n,n}$  are the roots of the Hermite polynomial  $H_n(\mathbf{y}) = (-1)^n e^{y^2/2} \frac{d^n}{dy^n} e^{-y^2/2}$  and  $w_{n,\alpha} = n! / (n^2 [H_{n-1}(y_{n,\alpha})]^2)$  are the associated weights. It is known that  $Q_n \psi$  is exactly equal to the integral  $I_1 \psi$  when  $\psi$  is a polynomial of degree less than or equal to  $2n-1$ , i.e., the polynomial degree of exactness of Gauss–Hermite quadrature rules  $Q_n$  is equal to  $2n-1$ .

We can approximate the multidimensional integral  $I_d \varphi$  by a quadrature expressed as the tensor product rule

$$\begin{aligned} I_d \varphi &\approx \bar{I}_d \varphi := Q_n \otimes Q_n \cdots \otimes Q_n \varphi(y_1, y_2, \dots, y_d) = Q_n^{\otimes d} \varphi(y_1, y_2, \dots, y_d) \quad (3.2.3) \\ &= \sum_{\alpha_1=1}^n \cdots \sum_{\alpha_d=1}^n \varphi(y_{n,\alpha_1}, \dots, y_{n,\alpha_d}) w_{n,\alpha_1} \cdots w_{n,\alpha_d}, \end{aligned}$$

where for simplicity we use the same amount on nodes in all the directions. The quadrature  $\bar{I}_d \varphi$  is exact for all polynomials from the space  $\mathcal{P}_{k_1} \otimes \dots \otimes \mathcal{P}_{k_d}$  with  $\max k_i = 2n-1$ , where  $\mathcal{P}_k$  is the space of one-dimensional polynomials of degree less than or equal to  $k$  (we note in passing that this fact is easy to prove using probabilistic representations of  $I_d \varphi$  and  $\bar{I}_d \varphi$ ). Computational costs of quadrature rules are measured in terms of a number of function evaluations which is equal to  $n^d$  in the case of the tensor product (3.2.3), i.e., the computational cost of (3.2.3) grows exponentially fast with dimension.

The sparse grid of Smolyak [358] reduces computational complexity of the tensor product rule (3.2.3) via exploiting the difference quadrature formulas:

$$A(L, d) \varphi := \sum_{d \leq |\mathbf{i}| \leq L+d-1} (Q_{i_1} - Q_{i_1-1}) \otimes \cdots \otimes (Q_{i_d} - Q_{i_d-1}) \varphi,$$

where  $Q_0 = 0$  and  $\mathbf{i} = (i_1, i_2, \dots, i_d)$  is a multi-index with  $i_k \geq 1$  and  $|\mathbf{i}| = i_1 + i_2 + \dots + i_d$ . The

number  $L$  is usually referred to as the level of the sparse grid. The sparse grid rule (3.2.4) can also be written in the following form [398]:

$$A(L, d)\varphi = \sum_{L \leq |\mathbf{i}| \leq L+d-1} (-1)^{L+d-1-|\mathbf{i}|} \binom{d-1}{|\mathbf{i}|-L} Q_{i_1} \otimes \cdots \otimes Q_{i_d} \varphi. \quad (3.2.4)$$

The quadrature  $A(L, d)\varphi$  is exact for polynomials from the space  $\mathcal{P}_{k_1} \otimes \cdots \otimes \mathcal{P}_{k_d}$  with  $|\mathbf{k}| = 2L - 1$ , i.e., for polynomials of total degree up to  $2L - 1$  [324, Corollary 1]. Due to (3.2.4), the total number of nodes used by this sparse grid rule is estimated by

$$\#S \leq \sum_{L \leq |\mathbf{i}| \leq L+d-1} i_1 \times \cdots \times i_d.$$

Table 3.1 lists the number of sparse grid points,  $\#S$ , up to level 5 when the level is not greater than  $d$ .

Table 3.1: The number of sparse grid points for the sparse grid quadrature (3.2.4) using the one-dimensional Gauss-Hermite quadrature rule (3.2.2), when the sparse grid level  $L \leq d$ .

	$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$
$\#S$	1	$2d + 1$	$2d^2 + 2d + 1$	$\frac{4}{3}d^3 + 2d^2 + \frac{14}{3}d + 1$	$\frac{2}{3}d^4 + \frac{4}{3}d^3 + \frac{22}{3}d^2 + \frac{8}{3}d + 1$

The quadrature  $\bar{I}_d\varphi$  from (3.2.3) is exact for polynomials of total degree  $2L - 1$  when  $n = L$ . It is not difficult to see that if the required polynomial exactness (in terms of total degree of polynomials) is relatively small then the sparse grid rule (3.2.4) substantially reduces the number of function evaluations compared with the tensor-product rule (3.2.3). For instance, suppose that the dimension  $d = 40$  and the required polynomial exactness is equal to 3. Then the cost of the tensor product rule (3.2.3) is  $3^{40} \doteq 1.2158 \times 10^{19}$  while the cost of the sparse grid rule (3.2.4) based on one-dimensional rule (3.2.2) is 3281.

**Remark 3.2.1.** *In this work we consider the isotropic SGC. More efficient algorithms might be built using anisotropic SGC methods [145, 323], which employ more quadrature points along the “most*

important" direction. Goal-oriented quadrature rules, e.g. [317, 331, 332], can also be exploited instead of pre-determined quadrature rules used here.

### 3.2.2 Weak-sense integration of SDE

Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $(w(t), \mathcal{F}_t^w) = ((w_1(t), \dots, w_r(t))^\top, \mathcal{F}_t^w)$  be an  $r$ -dimensional standard Wiener process, where  $\mathcal{F}_t^w$ ,  $0 \leq t \leq T$ , is an increasing family of  $\sigma$ -subalgebras of  $\mathcal{F}$  induced by  $w(t)$ .

Consider the system of Ito SDE

$$dX = a(t, X)dt + \sum_{l=1}^r \sigma_l(t, X)dw_l(t), \quad t \in (t_0, T], \quad X(t_0) = x_0, \quad (3.2.5)$$

where  $X$ ,  $a$ ,  $\sigma_r$  are  $m$ -dimensional column-vectors and  $x_0$  is independent of  $w$ . We assume that  $a(t, x)$  and  $\sigma(t, x)$  are sufficiently smooth and globally Lipschitz. We are interested in computing the expectation

$$u(x_0) = \mathbb{E}f(X_{t_0, x_0}(T)), \quad (3.2.6)$$

where  $f(x)$  is a sufficiently smooth function with growth at infinity not faster than a polynomial:

$$|f(x)| \leq K(1 + |x|^\varkappa) \quad (3.2.7)$$

for some  $K > 0$  and  $\varkappa \geq 1$ .

To find  $u(x_0)$ , we first discretize the solution of (3.2.5). Let

$$h = (T - t_0)/N, \quad t_k = t_0 + kh, \quad k = 0, \dots, N.$$

In application to (3.2.5) the Euler scheme has the form

$$X_{k+1} = X_k + a(t_k, X_k)h + \sum_{l=1}^r \sigma_l(t_k, X_k) \Delta_k w_l, \quad (3.2.8)$$

where  $X_0 = x_0$  and  $\Delta_k w_l = w_l(t_{k+1}) - w_l(t_k)$ . The Euler scheme can be realized in practice by replacing the increments  $\Delta_k w_l$  with Gaussian random variables:

$$X_{k+1} = X_k + a(t_k, X_k)h + \sum_{l=1}^r \sigma_l(t_k, X_k) \sqrt{h} \xi_{l,k+1}, \quad (3.2.9)$$

where  $\xi_{r,k+1}$  are i.i.d.  $\mathcal{N}(0,1)$  random variables. Due to our assumptions, the following error estimate holds for (3.2.9) (see e.g. [301, Chapter 2]):

$$|\mathbb{E}f(X_N) - \mathbb{E}f(X(T))| \leq Kh, \quad (3.2.10)$$

where  $K > 0$  is a constant independent of  $h$ . This first-order weak convergence can also be achieved by replacing  $\xi_{l,k+1}$  with discrete random variables [301], e.g., the weak Euler scheme has the form

$$\tilde{X}_{k+1} = \tilde{X}_k + ha(t_k, \tilde{X}_k) + \sqrt{h} \sum_{l=1}^r \sigma_l(t_k, \tilde{X}_k) \zeta_{l,k+1}, \quad k = 0, \dots, N-1, \quad (3.2.11)$$

where  $\tilde{X}_0 = x_0$  and  $\zeta_{l,k+1}$  are i.i.d. random variables with the law

$$P(\zeta = \pm 1) = 1/2. \quad (3.2.12)$$

The following error estimate holds for (3.2.11)-(3.2.12) (see e.g. [301, Chapter 2]):

$$|\mathbb{E}f(\tilde{X}_N) - \mathbb{E}f(X(T))| \leq Kh, \quad (3.2.13)$$

where  $K > 0$  can be a different constant than in (3.2.10).

Introducing the function  $\varphi(y)$ ,  $y \in \mathbb{R}^{rN}$ , so that

$$\varphi(\xi_{1,1}, \dots, \xi_{r,1}, \dots, \xi_{1,N}, \dots, \xi_{r,N}) = f(X_N), \quad (3.2.14)$$

we have

$$\begin{aligned} u(x_0) &\approx \bar{u}(x_0) := \mathbb{E}f(X_N) = \mathbb{E}\varphi(\xi_{1,1}, \dots, \xi_{r,1}, \dots, \xi_{1,N}, \dots, \xi_{r,N}) \\ &= \frac{1}{(2\pi)^{rN/2}} \int_{\mathbb{R}^{rN}} \varphi(y_{1,1}, \dots, y_{r,1}, \dots, y_{1,N}, \dots, y_{r,N}) \exp\left(-\frac{1}{2} \sum_{i=1}^{rN} y_i^2\right) dy. \end{aligned} \quad (3.2.15)$$

Further, it is not difficult to see from (3.2.11)-(3.2.12) and (3.2.3) that

$$\begin{aligned} u(x_0) &\approx \tilde{u}(x_0) := \mathbb{E}f(\tilde{X}_N) = \mathbb{E}\varphi(\zeta_{1,1}, \dots, \zeta_{r,1}, \dots, \zeta_{1,N}, \dots, \zeta_{r,N}) \\ &= Q_2^{\otimes rN} \varphi(y_{1,1}, \dots, y_{r,1}, \dots, y_{1,N}, \dots, y_{r,N}), \end{aligned} \quad (3.2.16)$$

where  $Q_2$  is the Gauss-Hermite quadrature rule with nodes  $\pm 1$  and equal weights  $1/2$ . We note that  $\tilde{u}(x_0)$  can be viewed as an approximation of  $\bar{u}(x_0)$  and that (cf. (3.2.10) and (3.2.13))  $\bar{u}(x_0) - \tilde{u}(x_0) = O(h)$ .

**Remark 3.2.2.** Let  $\zeta_{l,k+1}$  in (3.2.11) be *i.i.d.* random variables with the law

$$P(\zeta = y_{n,j}) = w_{n,j}, \quad j = 1, \dots, n, \quad (3.2.17)$$

where  $y_{n,j}$  are nodes of the Gauss-Hermite quadrature  $Q_n$  and  $w_{n,j}$  are the corresponding quadrature weights (see (3.2.2)). Then

$$\mathbb{E}f(\tilde{X}_N) = \mathbb{E}\varphi(\zeta_{1,N}, \dots, \zeta_{r,N}) = Q_n^{\otimes rN} \varphi(y_{1,1}, \dots, y_{r,N}),$$

which can be a more accurate approximation of  $\bar{u}(x_0)$  than  $\tilde{u}(x_0)$  from (3.2.16) but the weak-sense

error for the SDEs approximation  $\mathbb{E}f(\tilde{X}_N) - \mathbb{E}f(X(T))$  remains of order  $O(h)$ .

Practical implementation of  $\bar{u}(x_0)$  and  $\tilde{u}(x_0)$  usually requires the use of the Monte Carlo technique since the computational cost of, e.g. the tensor product rule in (3.2.16) is prohibitively high (cf. Section 3.2.1). In this work, we consider application of the sparse grid rule (3.2.4) to the integral in (3.2.15) motivated by lower computational cost of (3.2.4).

### Probabilistic interpretation of SGC

It is not difficult to show that SGC admits a probabilistic interpretation, e.g. in the case of level  $L = 2$  we have

$$\begin{aligned}
& A(2, N)\varphi(y_{1,1}, \dots, y_{r,1}, \dots, y_{1,N}, \dots, y_{r,N}) \\
&= (Q_2 \otimes Q_1 \otimes \dots \otimes Q_1)\varphi + (Q_1 \otimes Q_2 \otimes Q_1 \otimes \dots \otimes Q_1)\varphi \\
&\quad + \dots + (Q_1 \otimes Q_1 \otimes \dots \otimes Q_2)\varphi - (Nr - 1)(Q_1 \otimes Q_1 \otimes \dots \otimes Q_1)\varphi \\
&= \sum_{i=1}^N \sum_{j=1}^r \mathbb{E}\varphi(0, \dots, 0, \zeta_{j,i}, 0, \dots, 0) - (Nr - 1)\varphi(0, 0, \dots, 0),
\end{aligned} \tag{3.2.18}$$

where  $\zeta_{j,i}$  are i.i.d. random variables with the law (3.2.12). Using (3.2.16), (3.2.18), Taylor's expansion and symmetry of  $\zeta_{j,i}$ , we obtain the relationship between the weak Euler scheme (3.2.11) and the SGC (3.2.4):

$$\begin{aligned}
\mathbb{E}f(\tilde{X}_N) - A(2, N)\varphi &= \mathbb{E}\varphi(\zeta_{1,1}, \dots, \zeta_{r,1}, \dots, \zeta_{1,N}, \dots, \zeta_{r,N}) \\
&\quad - \sum_{i=1}^N \sum_{j=1}^r \mathbb{E}\varphi(0, \dots, 0, \zeta_{j,i}, 0, \dots, 0) - (Nr - 1)\varphi(0, 0, \dots, 0) \\
&= \sum_{|\alpha|=4} \frac{4}{\alpha!} \mathbb{E} \left[ \prod_{i=1}^N \prod_{j=1}^r (\zeta_{j,i})^{\alpha_{j,i}} \int_0^1 (1-z)^3 D^\alpha \varphi(z\zeta_{1,1}, \dots, z\zeta_{r,N}) dz \right] \\
&\quad - \frac{1}{3!} \sum_{i=1}^N \sum_{j=1}^r \mathbb{E} \left[ \zeta_{j,i}^4 \int_0^1 (1-z)^3 \frac{\partial^4}{(\partial y_{j,i})^4} \varphi(0, \dots, 0, z\zeta_{j,i}, 0, \dots, 0) dz \right],
\end{aligned} \tag{3.2.19}$$

where the multi-index  $\alpha = (\alpha_{1,1}, \dots, \alpha_{r,N}) \in \mathbb{N}_0^{rN}$ ,  $|\alpha| = \sum_{i=1}^N \sum_{j=1}^r \alpha_{j,i}$ ,  $\alpha! = \prod_{i=1}^N \prod_{j=1}^r \alpha_{j,i}!$  and  $D^\alpha = \frac{\partial^{|\alpha|}}{(\partial y_{1,1})^{\alpha_{1,1}} \dots (\partial y_{r,N})^{\alpha_{r,N}}}$ . The error of the SGC applied to weak-sense approximation of SDE is further studied in Section 3.2.3.

### Second-order schemes

In the SGC context, it is beneficial to exploit higher-order or higher-accuracy schemes for approximating the SDE (3.2.5) because they can allow us to reach a desired accuracy using larger time step sizes and therefore less random variables than the first-order Euler scheme (3.2.9) or (3.2.11). For instance, we can use the second-order weak scheme for (3.2.5) (see, e.g. [301, Chapter 2]):

$$\begin{aligned} X_{k+1} &= X_k + ha(t_k, X_k) + \sqrt{h} \sum_{i=1}^r \sigma_i(t_k, X_k) \xi_{i,k+1} + \frac{h^2}{2} \mathfrak{L}a(t_k, X_k) \\ &\quad + h \sum_{i=1}^r \sum_{j=1}^r \Lambda_i \sigma_j(t_k, X_k) \eta_{i,j,k+1} + \frac{h^{3/2}}{2} \sum_{i=1}^r (\Lambda_i a(t_k, X_k) + \mathfrak{L}\sigma_i(t_k, X_k)) \xi_{i,k+1}, \\ k &= 0, \dots, N-1, \end{aligned} \tag{3.2.20}$$

where  $X_0 = x_0$ ;  $\eta_{i,j} = \frac{1}{2} \xi_i \xi_j - \gamma_{i,j} \zeta_i \zeta_j / 2$  with  $\gamma_{i,j} = -1$  if  $i < j$  and  $\gamma_{i,j} = 1$  otherwise;

$$\Lambda_l = \sum_{i=1}^m \sigma_l^i \frac{\partial}{\partial x_i}, \quad \mathfrak{L} = \frac{\partial}{\partial t} + \sum_{i=1}^m a^i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{l=1}^r \sum_{i,j=1}^m \sigma_l^i \sigma_l^j \frac{\partial^2}{\partial x_i \partial x_j};$$

and  $\xi_{i,k+1}$  and  $\zeta_{i,k+1}$  are mutually independent random variables with Gaussian distribution or with the laws  $P(\xi = 0) = 2/3$ ,  $P(\xi = \pm\sqrt{3}) = 1/6$  and  $P(\zeta = \pm 1) = 1/2$ . The following error estimate holds for (3.2.20) (see e.g. [301, Chapter 2]):

$$|\mathbb{E}f(X(T)) - Ef(X_N)| \leq Kh^2.$$

Roughly speaking, to achieve  $O(h)$  accuracy using (3.2.20), we need only  $\sqrt{2rN}$  ( $\sqrt{rN}$  in the case of additive noise) random variables, while we need  $rN$  random variables for the Euler scheme

(3.2.9). This reduces the dimension of the random space and hence can increase efficiency and widen applicability of SGC methods (see, in particular Example 4.1 in Section 3.4 for a numerical illustration). We note that when noise intensity is relatively small, we can use high-accuracy low-order schemes designed for SDEs with small noise [300] (see also [301, Chapter 3]) in order to achieve a desired accuracy using less number of random variables than the Euler scheme (3.2.9).

### 3.2.3 Illustrative examples

In this section we show limitations of the use of SGC in weak approximation of SDEs. To this end, it is convenient and sufficient to consider the scalar linear SDE

$$dX = \lambda X dt + \varepsilon dw(t), \quad X_0 = 1, \quad (3.2.21)$$

where  $\lambda$  and  $\varepsilon$  are some constants.

We will compute expectations  $\mathbb{E}f(X(T))$  for some  $f(x)$  and  $X(t)$  from (3.2.21) by applying the Euler scheme (3.2.9) and the SGC (3.2.4). This simple example provides us with a clear insight when algorithms of this type are able to produce accurate results and when they are likely to fail. Using direct calculations, we first (see Examples 3.2.3–3.2.4 below) derive an estimate for the error  $|\mathbb{E}f(X_N) - A(2, N)\varphi|$  with  $X_N$  from (3.2.9) applied to (3.2.21) and for some particular  $f(x)$ . Then (Proposition 3.2.5) we obtain an estimate for the error  $|\mathbb{E}f(X_N) - A(L, N)\varphi|$  for a smooth  $f(x)$  which grows not faster than a polynomial function at infinity. We will observe that the considered algorithm is not convergent in time step  $h$  and not convergent in level  $L$  but it can be sufficiently accurate when noise intensity and integration time are small.

It follows from (3.2.10) and (3.2.13) that

$$\begin{aligned} |\mathbb{E}f(X_N) - A(L, N)\varphi| &\leq \left| \mathbb{E}f(\tilde{X}_N) - A(L, N)\varphi \right| + |\mathbb{E}f(X_N) - \mathbb{E}f(\tilde{X}_N)| & (3.2.22) \\ &\leq \left| \mathbb{E}f(\tilde{X}_N) - A(L, N)\varphi \right| + Kh, \end{aligned}$$

where  $\tilde{X}_N$  is from the weak Euler scheme (3.2.11) applied to (3.2.21), which can be written as  $\tilde{X}_N = (1 + \lambda h)^N + \sum_{j=1}^N (1 + \lambda h)^{N-j} \varepsilon \sqrt{h} \zeta_j$ . Introducing the function

$$\bar{X}(N; y) = (1 + \lambda h)^N + \sum_{j=1}^N (1 + \lambda h)^{N-j} \varepsilon \sqrt{h} y_j, \quad (3.2.23)$$

we see that  $\tilde{X}_N = \bar{X}(N; \zeta_1, \dots, \zeta_N)$ . We have

$$\frac{\partial}{\partial y_i} \bar{X}(N; y) = (1 + \lambda h)^{N-i} \varepsilon \sqrt{h} \quad \text{and} \quad \frac{\partial^2}{\partial y_i \partial y_j} \bar{X}(N; y) = 0. \quad (3.2.24)$$

Then we obtain from (3.2.19):

$$\begin{aligned} R & : = \mathbb{E}f(\tilde{X}_N) - A(2, N)\varphi & (3.2.25) \\ & = \varepsilon^4 h^2 \sum_{|\alpha|=4} \frac{4}{\alpha!} \mathbb{E} \left[ \prod_{i=1}^N (\zeta_i (1 + \lambda h)^{N-i})^{\alpha_i} \int_0^1 (1-z)^3 z^4 \frac{d^4}{dx^4} f(\bar{X}(N, z\zeta_1, \dots, z\zeta_N)) dz \right] \\ & \quad - \frac{1}{3!} \varepsilon^4 h^2 \sum_{i=1}^N \mathbb{E} \left[ \zeta_i^4 \int_0^1 (1-z)^3 z^4 \frac{d^4}{dx^4} f(\bar{X}(0, \dots, 0, z\zeta_i, 0, \dots, 0)) (1 + \lambda h)^{4N-4i} dz \right]. \end{aligned}$$

### Non-Convergence in time step $h$

We will illustrate the non-convergence in  $h$  through two examples.

**Example 3.2.3.** For  $f(x) = x^p$  with  $p = 1, 2, 3$ , it follows from (3.2.25) that  $R = 0$ , i.e., SGC does not introduce any additional error, and hence by (3.2.22)

$$|\mathbb{E}f(X_N) - A(2, N)\varphi| \leq Kh, \quad f(x) = x^p, \quad p = 1, 2, 3.$$

For  $f(x) = x^4$ , we get from (3.2.25):

$$R = \frac{6}{35} \varepsilon^4 h^2 \sum_{i=1}^N \sum_{j=i+1}^N (1 + \lambda h)^{4N-2i-2j}$$

$$= \frac{6}{35}\varepsilon^4 \times \begin{cases} \frac{(1+\lambda h)^{2N}-1}{\lambda^2(2+\lambda h)^2} \left[ \frac{(1+\lambda h)^{2N}+1}{1+(1+\lambda h)^2} - 1 \right], & \lambda \neq 0, \quad 1 + \lambda h \neq 0, \\ \frac{T^2}{2} - \frac{Th}{2}, & \lambda = 0. \end{cases}$$

We see that  $R$  does not go to zero when  $h \rightarrow 0$  and that for sufficiently small  $h > 0$

$$|\mathbb{E}f(X_N) - A(2, N)\varphi| \leq Kh + \frac{6}{35}\varepsilon^4 \times \begin{cases} \frac{1}{\lambda^2}(1 + e^{4T\lambda}), & \lambda \neq 0, \\ \frac{T^2}{2}, & \lambda = 0. \end{cases}$$

We observe that the SGC algorithm does not converge with  $h \rightarrow 0$  for higher moments. In the considered case of linear SDE, increasing the level  $L$  of SGC leads to the SGC error  $R$  being 0 for higher moments, e.g., for  $L = 3$  the error  $R = 0$  for up to 5th moment but the algorithm will not converge in  $h$  for 6th moment and so on (see Proposition 3.2.5 below). Further (see the continuation of the illustration below), in the case of, e.g.  $f(x) = \cos x$  for any  $L$  this error  $R$  is not zero, which is also the case for nonlinear SDE. We also note that one can expect that this error  $R$  is small when noise intensity is relatively small and either time  $T$  is small or SDE has, in some sense, stable behavior (in the linear case it corresponds to  $\lambda < 0$ ).

**Example 3.2.4.** Now consider  $f(x) = \cos(x)$ . It follows from (3.2.25) that

$$\begin{aligned} R &= \varepsilon^4 h^2 \sum_{|\alpha|=4} \frac{4}{\alpha!} \mathbb{E} \left[ \prod_{i=1}^N (\zeta_i (1 + \lambda h)^{N-i})^{\alpha_i} \int_0^1 (1-z)^3 z^4 \cos((1 + \lambda h)^N \right. \\ &\quad \left. + z \sum_{j=1}^N (1 + \lambda h)^{N-j} \varepsilon \sqrt{h} \zeta_j) dz \right] \\ &\quad - \frac{1}{3!} \varepsilon^4 h^2 \sum_{i=1}^N (1 + \lambda h)^{4N-4i} \int_0^1 (1-z)^3 z^4 \mathbb{E}[\zeta_i^4 \cos((1 + \lambda h)^N + z(1 + \lambda h)^{N-i} \varepsilon \sqrt{h} \zeta_i)] dz \end{aligned}$$

and after routine calculations we obtain

$$R = \varepsilon^4 h^2 \cos((1 + \lambda h)^N) \left[ \left( \frac{1}{6} \sum_{i=1}^N (1 + \lambda h)^{4N-4i} + 2 \sum_{i=1}^N \sum_{j=i+1}^N (1 + \lambda h)^{4N-2i-2j} \right) \right]$$

$$\begin{aligned}
& \times \int_0^1 (1-z)^3 z^4 \prod_{l=1}^N \cos(z(1+\lambda h)^{N-l} \varepsilon \sqrt{h}) dz \\
& + \left( \frac{2}{3} \sum_{i,j=1; i \neq j}^N (1+\lambda h)^{4N-3i-j} + 2 \sum_{\substack{k,i,j=1 \\ i \neq j, i \neq k, k \neq j}}^N (1+\lambda h)^{4N-2k-i-j} \right) \\
& \times \int_0^1 (1-z)^3 z^4 \prod_{l=i,j} \sin(z(1+\lambda h)^{N-l} \varepsilon \sqrt{h}) \prod_{\substack{l=1 \\ l \neq i, l \neq j}}^N \cos(z(1+\lambda h)^{N-l} \varepsilon \sqrt{h}) dz \\
& + 4 \sum_{\substack{i,j,k,m=1 \\ i \neq j, i \neq k, i \neq m, j \neq k, j \neq m, k \neq m}}^N (1+\lambda h)^{4N-i-j-k-m} \\
& \times \int_0^1 (1-z)^3 z^4 \prod_{l=i,j,m} \sin(z(1+\lambda h)^{N-l} \varepsilon \sqrt{h}) \prod_{\substack{l=1 \\ l \neq i, l \neq j, l \neq k, l \neq m}}^N \cos(z(1+\lambda h)^{N-l} \varepsilon \sqrt{h}) dz \\
& - \frac{1}{6} \sum_{i=1}^N (1+\lambda h)^{4N-4i} \int_0^1 (1-z)^3 z^4 \cos(z(1+\lambda h)^{N-i} \varepsilon \sqrt{h}) dz \Big].
\end{aligned}$$

It is not difficult to see that  $R$  does not go to zero when  $h \rightarrow 0$ . Further, taking into account that  $|\sin(z(1+\lambda h)^{N-j} \varepsilon \sqrt{h})| \leq z(1+\lambda h)^{N-j} \varepsilon \sqrt{h}$ , we get for sufficiently small  $h > 0$

$$|R| \leq C\varepsilon^4(1 + e^{4T\lambda}),$$

where  $C > 0$  is independent of  $\varepsilon$  and  $h$ . Hence

$$|\mathbb{E}f(X_N) - A(2, N)\varphi| \leq C\varepsilon^4(1 + e^{4T\lambda}) + Kh, \quad (3.2.26)$$

and we have arrived at a similar conclusion for  $f(x) = \cos x$  as for  $f(x) = x^4$ .

### Non-convergence in SGC level

Now we will address the question what the effect of increase of the level  $L$  on error estimates can be. To this end, we will need the following error estimate of a Gauss-Hermite quadrature. Let

$\psi(y)$ ,  $y \in \mathbb{R}$ , be a sufficiently smooth function which itself and its derivatives are growing not faster than a polynomial at infinity. Using the Peano kernel theorem (see e.g. [85]) and that a Gauss-Hermite quadrature with  $n$ -nodes has the order of polynomial exactness  $2n - 1$ , we obtain for the approximation error  $R_{n,\gamma}\psi$  of the Gauss-Hermite quadrature  $Q_n\psi$ :

$$R_{n,\gamma}(\psi) := Q_n\psi - I_1\psi = \int_{\mathbb{R}} \frac{d^\gamma}{dy^\gamma} \varphi(y) R_{n,\gamma}(\Gamma_{y,\gamma}) dy, \quad 1 \leq \gamma \leq 2n, \quad (3.2.27)$$

where  $\Gamma_{y,\gamma}(z) = (z-y)^{\gamma-1}/(\gamma-1)!$  if  $z \geq y$  and 0 otherwise. One can show (see, e.g. [286, Theorem 2]) that there is a constant  $c > 0$  independent of  $n$  and  $y$  such that for any  $0 < \beta < 1$

$$|R_{n,\gamma}(\Gamma_{y,\gamma})| \leq \frac{c}{\sqrt{2\pi}} n^{-\gamma/2} \exp\left(-\frac{\beta y^2}{2}\right), \quad 1 \leq \gamma \leq 2n. \quad (3.2.28)$$

We also note that (3.2.28) and the triangle inequality imply, for  $1 \leq \gamma \leq 2(n-1)$ :

$$|R_{n,\gamma}(\Gamma_{y,\gamma}) - R_{n-1,\gamma}(\Gamma_{y,\gamma})| \leq \frac{c}{\sqrt{2\pi}} [n^{-\gamma/2} + (n-1)^{-\gamma/2}] \exp\left(-\frac{\beta y^2}{2}\right). \quad (3.2.29)$$

Now we consider an error of the sparse grid rule (3.2.4) accompanied by the Euler scheme (3.2.9) for computing expectations of solutions to (3.2.21).

**Proposition 3.2.5.** *Assume that a function  $f(x)$  and its derivatives up to  $2L$ -th order satisfy the polynomial growth condition (3.2.7). Let  $X_N$  be obtained by the Euler scheme (3.2.9) applied to the linear SDE (3.2.21) and  $A(L, N)\varphi$  be the sparse grid rule (3.2.4) with level  $L$  applied to the integral corresponding to  $\mathbb{E}f(X_N)$  as in (3.2.15). Then for  $L \leq N$  and sufficiently small  $h > 0$*

$$|\mathbb{E}f(X_N) - A(L, N)\varphi| \leq K\varepsilon^{2L}(1 + e^{\lambda(2L+\varkappa)T}) \left(1 + (3c/2)^L\right) \beta^{-L/2} T^L, \quad (3.2.30)$$

where  $K > 0$  is independent of  $h$ ,  $L$  and  $N$ ;  $c$  and  $\beta$  are from (3.2.28);  $\varkappa$  is from (3.2.7).

*Proof.* We recall (see (3.2.15)) that

$$\mathbb{E}f(X_N) = I_N\varphi = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} \varphi(y_1, \dots, y_N) \exp\left(-\frac{1}{2} \sum_{i=1}^N y_i^2\right) dy.$$

Introduce the integrals

$$I_1^{(k)}\varphi = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi(y_1, \dots, y_k, \dots, y_N) \exp\left(-\frac{y_k^2}{2}\right) dy_k, \quad k = 1, \dots, N, \quad (3.2.31)$$

and their approximations  $Q_n^{(k)}$  by the corresponding one-dimensional Gauss-Hermite quadratures with  $n$  nodes. Also, let  $\mathcal{U}_{i_k}^{(k)} = Q_{i_k}^{(k)} - Q_{i_k-1}^{(k)}$ .

Using (3.2.4) and the recipe from the proof of Lemma 3.4 in [323], we obtain

$$I_N\varphi - A(L, N)\varphi = \sum_{l=2}^N S(L, l) \otimes_{k=l+1}^N I_1^{(k)}\varphi + (I_1^{(1)} - Q_L^{(1)}) \otimes_{k=2}^N I_1^{(k)}\varphi, \quad (3.2.32)$$

where

$$S(L, l) = \sum_{i_1 + \dots + i_{l-1} + i_l = L + l - 1} \otimes_{k=1}^{l-1} \mathcal{U}_{i_k}^{(k)} \otimes (I_1^{(l)} - Q_{i_l}^{(l)}). \quad (3.2.33)$$

Due to (3.2.27), we have for  $n > 1$  and  $1 \leq \gamma \leq 2(n-1)$

$$\begin{aligned} \mathcal{U}_n\psi &= Q_n\psi - Q_{n-1}\psi = [Q_n\psi - I_1(\psi)] - [Q_{n-1}\psi - I_1(\psi)] \\ &= \int_{\mathbb{R}} \frac{d^\gamma}{dy^\gamma} \psi(y) [R_{n,\gamma}(\Gamma_{y,\gamma}) - R_{n-1,\gamma}(\Gamma_{y,\gamma})] dy, \end{aligned} \quad (3.2.34)$$

and for  $n = 1$

$$\mathcal{U}_1\psi = Q_1\psi - Q_0\psi = Q_1\psi = \psi(0). \quad (3.2.35)$$

By (3.2.33), (3.2.31) and (3.2.27), we obtain for the first term in the right-hand side of (3.2.32):

$$S(L, l) \otimes_{n=l+1}^N I_1^{(n)}\varphi$$

$$\begin{aligned}
&= \sum_{i_1+\dots+i_l=L+l-1} \otimes_{n=1}^{l-1} \mathcal{U}_{i_k}^{(k)} \otimes (I_1^{(l)} - Q_{i_l}^{(l)}) \otimes_{n=l+1}^N I_1^{(n)} \varphi \\
&= \sum_{i_1+\dots+i_l=L+l-1} \otimes_{n=1}^{l-1} \mathcal{U}_{i_k}^{(k)} \otimes (I_1^{(l)} - Q_{i_l}^{(l)}) \\
&\quad \otimes \int_{\mathbb{R}^{N-l}} \varphi(y) \frac{1}{(2\pi)^{(N-l)/2}} \exp\left(-\sum_{k=l+1}^N \frac{y_k^2}{2}\right) dy_{l+1} \dots dy_N \\
&= - \sum_{i_1+\dots+i_l=L+l-1} \otimes_{n=1}^{l-1} \mathcal{U}_{i_k}^{(k)} \otimes \int_{\mathbb{R}^{N-l+1}} \frac{d^{2i_l}}{dy_l^{2i_l}} \varphi(y) R_{i_l, 2i_l}(\Gamma_{y_l, 2i_l}) \\
&\quad \times \frac{1}{(2\pi)^{(N-l)/2}} \exp\left(-\sum_{k=l+1}^N \frac{y_k^2}{2}\right) dy_l \dots dy_N.
\end{aligned}$$

Now consider two cases: if  $i_{l-1} > 1$  then by (3.2.34):

$$\begin{aligned}
S(L, l) \otimes_{n=l+1}^N I_1^{(n)} \varphi &= - \sum_{i_1+\dots+i_l=L+l-1} \otimes_{n=1}^{l-2} \mathcal{U}_{i_k}^{(k)} \otimes \int_{\mathbb{R}^{N-l+2}} \frac{d^{2i_{l-1}-2}}{dy_{l-1}^{2i_{l-1}-2}} \frac{d^{2i_l}}{dy_l^{2i_l}} \varphi(y) R_{i_l, 2i_l}(\Gamma_{y_l, 2i_l}) \\
&\quad \times [R_{i_{l-1}, 2i_{l-1}-2}(\Gamma_{y_{l-1}, 2i_{l-1}-2}) - R_{i_{l-1}-1, 2i_{l-1}-2}(\Gamma_{y_{l-1}, 2i_{l-1}-2})] \\
&\quad \times \frac{1}{(2\pi)^{(N-l)/2}} \exp\left(-\sum_{k=l+1}^N \frac{y_k^2}{2}\right) dy_{l-1} \dots dy_N
\end{aligned}$$

otherwise (i.e., if  $i_{l-1} = 1$ ) by (3.2.35):

$$\begin{aligned}
S(L, l) \otimes_{n=l+1}^N I_1^{(n)} \varphi &= - \sum_{i_1+\dots+i_l=L+l-1} \otimes_{n=1}^{l-2} \mathcal{U}_{i_k}^{(k)} \otimes \int_{\mathbb{R}^{N-l+1}} Q_1^{(l-1)} \frac{d^{2i_l}}{dy_l^{2i_l}} \varphi(y) R_{i_l, 2i_l}(\Gamma_{y_l, 2i_l}) \\
&\quad \times \frac{1}{(2\pi)^{(N-l)/2}} \exp\left(-\sum_{k=l+1}^N \frac{y_k^2}{2}\right) dy_l \dots dy_N.
\end{aligned}$$

Repeating the above process for  $i_{l-2}, \dots, i_1$ , we obtain

$$\begin{aligned}
S(L, l) \otimes_{n=l+1}^N I_1^{(n)} \varphi &= \sum_{i_1+\dots+i_l=L+l-1} \int_{\mathbb{R}^{N-\#F_{l-1}}} [\otimes_{m \in F_{l-1}} Q_1^{(m)} D^{2\alpha_l} \varphi(y)] \quad (3.2.36) \\
&\quad \times \mathcal{R}_{l, \alpha_l}(y_1, \dots, y_l) \frac{1}{(2\pi)^{(N-l)/2}} \exp\left(-\sum_{k=l+1}^N \frac{y_k^2}{2}\right) \prod_{n \in G_{l-1}} dy_n \times dy_l \dots dy_N,
\end{aligned}$$

where the multi-index  $\alpha_l = (i_1 - 1, \dots, i_{l-1} - 1, i_l, 0, \dots, 0)$  with the  $m$ -th element  $\alpha_l^m$ , the sets

$F_{l-1} = F_{l-1}(\alpha_l) = \{m : \alpha_l^m = 0, m = 1, \dots, l-1\}$  and  $G_{l-1} = G_{l-1}(\alpha_l) = \{m : \alpha_l^m > 0, m = 1, \dots, l-1\}$ ,

the symbols  $\#F_{l-1}$  and  $\#G_{l-1}$  stand for the number of elements in the corresponding sets, and

$$\mathcal{R}_{l,\alpha_l}(y_1, \dots, y_l) = -R_{i_l, 2i_l}(\Gamma_{y_l, 2i_l}) \otimes_{n \in G_{l-1}} [R_{i_n, 2i_n-2}(\Gamma_{y_n, 2i_n-2}) - R_{i_n-1, 2i_n-2}(\Gamma_{y_n, 2i_n-2})].$$

Note that  $\#G_{l-1} \leq (L-1) \wedge (l-1)$  and also recall that  $i_j \geq 1, j = 1, \dots, l$ .

Using (3.2.28), (3.2.29) and the inequality

$$\prod_{n \in G_{l-1}} [i_n^{-(i_n-1)} + (i_n-1)^{-(i_n-1)}] i_l^{-i_l} \leq (3/2)^{\#G_{l-1}},$$

we get

$$\begin{aligned} |\mathcal{R}_{l,\alpha}(y_1, \dots, y_l)| &\leq \prod_{n \in G_{l-1}} [i_n^{-(i_n-1)} + (i_n-1)^{-(i_n-1)}] i_l^{-i_l} \frac{c^{\#G_{l-1}+1}}{(2\pi)^{(\#G_{l-1}+1)/2}} \quad (3.2.37) \\ &\quad \times \exp\left(-\sum_{n \in G_{l-1}} \frac{\beta y_n^2}{2} - \frac{\beta y_l^2}{2}\right) \\ &\leq \frac{(3c/2)^{\#G_{l-1}+1}}{(2\pi)^{(\#G_{l-1}+1)/2}} \exp\left(-\sum_{n \in G_{l-1}} \frac{\beta y_n^2}{2} - \frac{\beta y_l^2}{2}\right). \end{aligned}$$

Substituting (3.2.37) in (3.2.36), we arrive at

$$\begin{aligned} &\left| S(L, l) \otimes_{n=l+1}^N I_1^{(n)} \varphi \right| \quad (3.2.38) \\ &\leq \sum_{i_1+\dots+i_l=L+l-1} \frac{(3c/2)^{\#G_{l-1}+1}}{(2\pi)^{(N-\#F_{l-1})/2}} \int_{\mathbb{R}^{N-\#F_{l-1}}} \left| \otimes_{m \in F_{l-1}} Q_1^{(m)} D^{2\alpha_l} \varphi(y) \right| \\ &\quad \times \exp\left(-\sum_{n \in G_{l-1}} \frac{\beta y_n^2}{2} - \frac{\beta y_l^2}{2} - \sum_{k=l+1}^N \frac{y_k^2}{2}\right) \prod_{n \in G_{l-1}} dy_n \times dy_l \dots dy_N. \end{aligned}$$

Using (3.2.24) and the assumption that  $\left| \frac{d^{2L}}{dx^{2L}} f(x) \right| \leq K(1+|x|^\varkappa)$  for some  $K > 0$  and  $\varkappa \geq 1$ , we

get

$$|D^{2\alpha_l} \varphi(y)| = \varepsilon^{2L} h^L \left| \frac{d^{2L}}{dx^{2L}} f(\bar{X}(N, y)) \right| (1 + \lambda h)^{2LN-2\sum_{i=1}^l i\alpha_i} \quad (3.2.39)$$

$$\leq K\varepsilon^{2L}h^L(1+\lambda h)^{2LN-2\sum_{i=1}^l i\alpha_i}(1+|\bar{X}(N,y)|^\varkappa).$$

Substituting (3.2.39) and (3.2.23) in (3.2.38) and doing further calculations, we obtain

$$\begin{aligned} \left|S(L,l)\otimes_{n=l+1}^N I_1^{(n)}\varphi\right| &\leq K\varepsilon^{2L}h^L(1+e^{\lambda\varkappa T})(1+(3c/2)^{L\wedge l})\beta^{-(L\wedge l)/2} \\ &\quad \times \sum_{i_1+\dots+i_l=L+l-1} (1+\lambda h)^{2LN-2\sum_{i=1}^l i\alpha_i} \\ &\leq K\varepsilon^{2L}h^L(1+e^{\lambda(2L+\varkappa)T})(1+(3c/2)^{L\wedge l})\beta^{-(L\wedge l)/2}\binom{L+l-2}{L-1} \\ &\leq K\varepsilon^{2L}h^L(1+e^{\lambda(2L+\varkappa)T})(1+(3c/2)^{L\wedge l})\beta^{-(L\wedge l)/2}l^{L-1}. \end{aligned} \tag{3.2.40}$$

with a new  $K > 0$  which does not depend on  $h, \varepsilon, L, c, \beta$ , and  $l$ . In the last line of (3.2.40) we used

$$\binom{L+l-2}{L-1} = \prod_{i=1}^{L-1} \left(1 + \frac{l-1}{i}\right) \leq \left[\frac{1}{L-1} \sum_{i=1}^{L-1} \left(1 + \frac{l-1}{i}\right)\right]^{L-1} \leq l^{L-1}.$$

Substituting (3.2.40) in (3.2.32) and observing that  $\left|(I_1^{(1)} - Q_L^{(1)}) \otimes_{k=2}^N I_1^{(k)}\varphi\right|$  is of order  $O(h^L)$ , we arrive at (3.2.30).  $\square$

**Remark 3.2.6.** *Due to Examples 3.2.3 and 3.2.4, the error estimate (3.2.30) proved in Proposition 3.2.5 is quite sharp and we conclude that in general the SGC algorithm for weak approximation of SDE does not converge with neither decrease of time step  $h$  nor with increase of the level  $L$ . At the same time, the algorithm can be sufficiently accurate when noise intensity  $\varepsilon$  and integration time  $T$  are relatively small.*

**Remark 3.2.7.** *It follows from the proof (see (3.2.39)) that if  $\frac{d^{2L}}{dx^{2L}}f(x) = 0$  then the error  $I_N(\varphi) - A(L,N)\varphi = 0$ . We emphasize that this is a feature of the linear SDE (3.2.21) thanks to (3.2.24), while in the case of nonlinear SDE this error remains of the form (3.2.30) even if the  $2L$ th derivative of  $f$  is zero. See also the discussion at the end of Example 3.2.3 and numerical tests in Example 3.4.1.*

**Remark 3.2.8.** *We note that it is possible to prove a proposition analogous to Proposition 3.2.5 for a more general SDE, e.g. for SDE with additive noise. Since such a proposition does not add further information to our discussion of the use of SGC and its proof is more complex than in the case of (3.2.21), we do not consider such a proposition here.*

### 3.3 Recursive collocation algorithm for linear SPDE

In the previous section we have demonstrated the limitations of SGC algorithms in application to SDEs that, in general, such an algorithm will not work unless integration time  $T$  and magnitude of noise are small. It is not difficult to understand that SGC algorithms have the same limitations in the case of SPDE as well, which, in particular, is demonstrated in Example 4.2, where a stochastic Burgers equation is considered. To cure this deficiency and achieve longer time integration in the case of linear SPDE, we will exploit the idea of the recursive approach proposed in [265, 420] in the case of a Wiener chaos expansion method. To this end, we apply the algorithm of SGC accompanied by a time discretization of SPDE over a small interval  $[(k-1)h, kh]$  instead of the whole interval  $[0, T]$  as we did in the previous section and build a recursive scheme to compute the second-order moments of the solutions to linear SPDE.

Consider the linear SPDE (2.2.1) with finite dimensional noises. We will continue to use the notation from the previous section:  $h$  is a step of uniform discretization of the interval  $[0, T]$ ,  $N = T/h$  and  $t_k = kh$ ,  $k = 0, \dots, N$ . We apply the trapezoidal rule in time to the SPDE (2.2.1):

$$\begin{aligned}
 u^{k+1}(x) &= u^k(x) + h[\tilde{\mathcal{L}}u^{k+1/2}(x) - \frac{1}{2} \sum_{l=1}^r \mathcal{M}_l g_l(x) + f(x)] \\
 &+ \sum_{l=1}^r \left[ \mathcal{M}_l u^{k+1/2}(x) + g_l(x) \right] \sqrt{h} (\xi_{lh})_{k+1}, \quad x \in \mathcal{D}, \\
 u^0(x) &= u_0(x),
 \end{aligned} \tag{3.3.1}$$

where  $u^k(x)$  approximates  $u(t_k, x)$ ,  $u^{k+1/2} = (u^{k+1} + u^k)/2$ , and  $(\xi_{lh})_k$  are i.i.d. random variables

so that

$$\xi_h = \begin{cases} \xi, & |\xi| \leq A_h, \\ A_h, & \xi > A_h, \\ -A_h, & \xi < -A_h, \end{cases} \quad (3.3.2)$$

with  $\xi \sim \mathcal{N}(0, 1)$  and  $A_h = \sqrt{2p|\ln h|}$  with  $p \geq 1$ . We note that the cut-off of the Gaussian random variables is needed in order to ensure that the implicitness of (3.3.1) does not lead to non-existence of the second moment of  $u^k(x)$  [299, 301]. Based on the standard results of numerics for SDEs [301], it is natural to expect that under some regularity assumptions on the coefficients and the initial condition of (2.2.1), the approximation  $u^k(x)$  from (3.3.1) converges with order 1/2 in the mean-square sense and with order 1 in the weak sense and in the latter case one can use discrete random variables  $\zeta_{l,k+1}$  from (3.2.12) instead of  $(\xi_{lh})_{k+1}$  (see also e.g. [96, 141] but we are not proving such a result here).

In what follows it will be convenient to also use the notation:  $u_H^k(x; \phi(\cdot)) = u_H^k(x; \phi(\cdot); (\xi_{lh})_k, l = 1, \dots, r)$  for the approximation (3.3.1) of the solution  $u(t_k, x)$  to the SPDE (2.2.1) with  $f(x) = 0$  and  $g_l(x) = 0$  for all  $l$  (homogeneous SPDE) and with the initial condition  $\phi(\cdot)$  prescribed at time  $t = t_{k-1}$ ;  $u_O^k(x) = u_O^k(x; (\xi_{lh})_k, l = 1, \dots, r)$  for the approximation (3.3.1) of the solution  $u(t_k, x)$  to the SPDE (2.2.1) with the initial condition  $\phi(x) = 0$  prescribed at time  $t = t_{k-1}$ . Note that  $u_O^k(x) = 0$  if  $f(x) = 0$  and  $g_l(x) = 0$  for all  $l$ .

Let  $\{e_i\} = \{e_i(x)\}_{i \geq 1}$  be a complete orthonormal system (CONS) in  $L^2(\mathcal{D})$  with boundary conditions satisfied and  $(\cdot, \cdot)$  be the inner product in that space. Then we can write

$$u^{k-1}(x) = \sum_{i=1}^{\infty} c_i^{k-1} e_i(x) \quad (3.3.3)$$

with  $c_i^{k-1} = (u^{k-1}, e_i)$  and, due to the SPDE's linearity:

$$u^k(x) = u_O^k(x) + \sum_{i=1}^{\infty} c_i^{k-1} u_H^k(x; e_i(\cdot)).$$

We have

$$c_l^0 = (u_0, e_l), \quad c_l^k = q_{Ol}^k + \sum_{i=1}^{\infty} c_i^{k-1} q_{Hli}^k, \quad l = 1, 2, \dots, \quad k = 1, \dots, N,$$

where  $q_{Ol}^k = (u_O^k, e_l)$  and  $q_{Hli}^k = (u_H^k(\cdot; e_i), e_l(\cdot))$ .

Using (3.3.3), we represent the second moment of the approximation  $u^k(x)$  from (3.3.1) of the solution  $u(t_k, x)$  to the SPDE (2.2.1) as follows

$$\mathbb{E}[u^k(x)]^2 = \sum_{i,j=1}^{\infty} C_{ij}^k e_i(x) e_j(x), \quad (3.3.4)$$

where the covariance matrix  $C_{ij}^k = \mathbb{E}[c_i^k c_j^k]$ . Introducing also the means  $M_i^k$ , one can obtain the recurrent relations in  $k$  :

$$\begin{aligned} M_i^0 &= c_i^0 = (u_0, e_i), \quad C_{ij}^0 = c_i^0 c_j^0, \\ M_i^k &= \mathbb{E}[q_{Oi}^k] + \sum_{l=1}^{\infty} M_l^{k-1} \mathbb{E}[q_{Hil}^k], \\ C_{ij}^k &= \mathbb{E}[q_{Oi}^k q_{Oj}^k] + \sum_{l=1}^{\infty} M_l^{k-1} (\mathbb{E}[q_{O_i}^k q_{Hjl}^k] + \mathbb{E}[q_{O_j}^k q_{Hil}^k]) + \sum_{l,p=1}^{\infty} C_{lp}^{k-1} \mathbb{E}[q_{Hil}^k q_{Hjp}^k], \\ i, j &= 1, 2, \dots, \quad k = 1, \dots, N. \end{aligned} \quad (3.3.5)$$

Since the coefficients of the SPDE (2.2.1) are time independent, all the expectations involving the quantities  $q_{Oi}^k$  and  $q_{Hil}^k$  in (3.3.5) do not depend on  $k$  and hence it is sufficient to compute them just once, on a single step  $k = 1$ , and we get

$$\begin{aligned} M_i^0 &= c_i^0 = (u_0, e_i), \quad C_{ij}^0 = c_i^0 c_j^0, \\ M_i^k &= \mathbb{E}[q_{Oi}^1] + \sum_{l=1}^{\infty} M_l^{k-1} \mathbb{E}[q_{Hil}^1], \\ C_{ij}^k &= \mathbb{E}[q_{O_i}^1 q_{O_j}^1] + \sum_{l=1}^{\infty} M_l^{k-1} (\mathbb{E}[q_{O_i}^1 q_{Hjl}^1] + \mathbb{E}[q_{O_j}^1 q_{Hil}^1]) + \sum_{l,p=1}^{\infty} C_{lp}^{k-1} \mathbb{E}[q_{Hil}^1 q_{Hjp}^1], \\ i, j &= 1, 2, \dots, \quad k = 1, \dots, N. \end{aligned} \quad (3.3.6)$$

These expectations can be approximated by quadrature rules from Section 2.1. If the number of noises  $r$  is small, then it is natural to use the tensor product rule (3.2.3) with one-dimensional Gauss–Hermite quadratures of order  $n = 2$  or  $3$  (note that when  $r = 1$ , we can use just a one-dimensional Gauss–Hermite quadrature of order  $n = 2$  or  $3$ ). If the number of noises  $r$  is large then it might be beneficial to use the sparse grid quadrature (3.2.4) of level  $L = 2$  or  $3$ . More specifically,

$$\begin{aligned}
\mathbb{E}[q_{O_i}^1] &\doteq \sum_{p=1}^{\eta} (u_O^1(\cdot; y_p), e_i(\cdot)) W_p, & \mathbb{E}[q_{H_{il}}^1] &\doteq \sum_{p=1}^{\eta} (u_H^1(\cdot; e_l; y_p), e_i(\cdot)) W_p, & (3.3.7) \\
\mathbb{E}[q_{O_i}^1 q_{O_j}^1] &\doteq \sum_{p=1}^{\eta} (u_O^1(\cdot; y_p), e_i(\cdot)) (u_O^1(\cdot; y_p), e_j(\cdot)) W_p, \\
\mathbb{E}[q_{O_i}^1 q_{H_{jl}}^1] &\doteq \sum_{p=1}^{\eta} (u_O^1(\cdot; y_p), e_i(\cdot)) (u_H^1(\cdot; e_l; y_p), e_j(\cdot)) W_p, \\
\mathbb{E}[q_{H_{il}}^1 q_{H_{jk}}^1] &\doteq \sum_{p=1}^{\eta} (u_H^1(\cdot; e_l; y_p), e_i(\cdot)) (u_H^1(\cdot; e_k; y_p), e_j(\cdot)) W_p,
\end{aligned}$$

where  $y_p \in \mathbb{R}^r$  are nodes of the quadrature,  $W_p$  are the corresponding quadrature weights, and  $\eta = n^r$  in the case of the tensor product rule (3.2.3) with one-dimensional Gauss–Hermite quadratures of order  $n$  or  $\eta$  is the total number of nodes  $\#S$  used by the sparse-grid quadrature (3.2.4) of level  $L$ . To find  $u_O^1(x; y_p)$  and  $u_H^1(x; e_l; y_p)$ , we need to solve the corresponding elliptic PDE problems, which we do using the spectral method in physical space, i.e., using a truncation of the CONS  $\{e_l\}_{l=1}^{l_*}$  to represent the numerical solution.

To summarize, we formulate the following deterministic recursive algorithm for the second-order moments of the solution to the SPDE problem (2.2.1).

**Algorithm 3.3.1.** *Choose the algorithm’s parameters: a complete orthonormal basis  $\{e_l(x)\}_{l \geq 1}$  in  $L^2(\mathcal{D})$  and its truncation  $\{e_l(x)\}_{l=1}^{l_*}$ ; a time step size  $h$ ; and a quadrature rule (i.e., nodes  $y_p$  and the quadrature weights  $W_p$ ,  $p = 1, \dots, \eta$ ).*

*Step 1. For each  $p = 1, \dots, \eta$  and  $l = 1, \dots, l_*$ , find approximations  $\bar{u}_O^1(x; y_p) \approx u_O^1(x; y_p)$  and  $\bar{u}_H^1(x; e_l; y_p) \approx u_H^1(x; e_l; y_p)$  using the spectral method in physical space.*

*Step 2. Using the quadrature rule, approximately find the expectations as in (3.3.7) but with the approximate  $\bar{u}_O^1(x; y_p)$  and  $\bar{u}_H^1(x; e_l; y_p)$  instead of  $u_O^1(x; y_p)$  and  $u_H^1(x; e_l; y_p)$ , respectively.*

*Step 3. Recursively compute the approximations of the means  $M_i^k$ ,  $i = 1, \dots, l_*$ , and covariance matrices  $\{C_{ij}^k, i, j = 1, \dots, l_*\}$  for  $k = 1, \dots, N$  according to (3.3.6) with the approximate expectations found in Step 2 instead of the exact ones.*

*Step 4. Compute the approximation of the second-order moment  $\mathbb{E}[u^k(x)]^2$  using (3.3.4) with the approximate covariance matrix found in Step 3 instead of the exact one  $\{C_{ij}^k\}$ .*

We emphasize that Algorithm 3.3.1 for computing moments does not have a statistical error. Error analysis of this algorithm will be considered elsewhere.

**Remark 3.3.2.** *Algorithms analogous to Algorithm 3.3.1 can also be constructed based on other time-discretizations methods than the trapezoidal rule used here or based on other types of SPDE approximations, e.g. one can exploit the Wong-Zakai approximation.*

**Remark 3.3.3.** *The cost of this algorithm is, similar to the algorithm in [420],  $\frac{T}{\Delta} \eta l_*^4$  and the storage is  $\eta l_*^2$ . The total cost can be reduced by employing some reduced order methods in physical space and be proportional to  $l_*^2$  instead of  $l_*^4$ . The discussion on computational efficiency of the recursive Wiener chaos method is also valid here, see [420, Remark 4.1].*

## 3.4 Numerical experiments

In this section we illustrate via three examples how the SGC algorithms can be used for the weak-sense approximation of SDEs and SPDEs. The first example is a scalar SDE with multiplicative noise, where we show that the SGC algorithm's error is small when the noise magnitude is small. We also observe that when the noise magnitude is large, the SGC algorithm does not work well. In the second example we demonstrate that the SGC can be successfully used for simulating Burgers equation with additive noise when the integration time is relatively small. In the last example we

show that the recursive algorithm from Section 3.3 works effectively for computing moments of the solution to an advection-diffusion equation with multiplicative noise over a longer integration time.

In all the tests we limit the dimension of random spaces by 40, which is an empirical limitation of the SGC of Smolyak on the dimensionality [335]. Also, we take the sparse grid level less than or equal to five in order to avoid an excessive number of sparse grid points. All the tests were run using Matlab R2012b on a Macintosh desktop computer with Intel Xeon CPU E5462 (quad-core, 2.80 GHz).

**Example 3.4.1** (modified Cox-Ingersoll-Ross (mCIR), see e.g. [73]). Consider the Ito SDE

$$dX = -\theta_1 X dt + \theta_2 \sqrt{1 + X^2} dw(t), \quad X(0) = x_0. \quad (3.4.1)$$

For  $\theta_2^2 - 2\theta_1 \neq 0$ , the first two moments of  $X(t)$  are equal to

$$\mathbb{E}X(t) = x_0 \exp(-\theta_1 t), \quad \mathbb{E}X^2(t) = -\frac{\theta_2^2}{\theta_2^2 - 2\theta_1} + \left(x_0^2 + \frac{\theta_2^2}{\theta_2^2 - 2\theta_1}\right) \exp((\theta_2^2 - 2\theta_1)t).$$

In this example we test the SGC algorithms based on the Euler scheme (3.2.8) and on the second-order weak scheme (3.2.20). We compute the first two moments of the SDE's solution and measure the errors of the algorithms as

$$\rho_1^r(T) = \frac{|\mathbb{E}X(T) - \mathbb{E}X_N|}{|\mathbb{E}X(T)|}, \quad \rho_2^r(T) = \frac{|\mathbb{E}X^2(T) - \mathbb{E}X_N^2|}{\mathbb{E}X^2(T)}. \quad (3.4.2)$$

Table 3.2 presents the errors for the SGC algorithms based on the Euler scheme (left) and on the second-order scheme (3.2.20) (right), when the noise magnitude is small. For the parameters given in the table's description, the exact values (up to 4 d.p.) of the first and second moments are  $3.679 \times 10^{-2}$  and  $4.162 \times 10^{-2}$ , respectively. We see that increase of the SGC level  $L$  above 2 in the Euler scheme case and above 3 in the case of the second-order scheme does not improve accuracy. When the SGC error is relatively small in comparison with the error due to time discretization, we

observe decrease of the overall error of the algorithms in  $h$ : proportional to  $h$  for the Euler scheme and to  $h^2$  for the second-order scheme. We underline that in this experiment the noise magnitude is small.

Table 3.2: Comparison of the SGC algorithms based on the Euler scheme (left) and on the second-order scheme (3.2.20) (right). The parameters of the model (3.4.1) are  $x_0 = 0.1$ ,  $\theta_1 = 1$ ,  $\theta_2 = 0.3$ , and  $T = 1$ .

$h$	$L$	$\rho_1^r(1)$	order	$\rho_2^r(1)$	order	$L$	$\rho_1^r(1)$	order	$\rho_2^r(1)$	order
$5 \times 10^{-1}$	2	$3.20 \times 10^{-1}$	–	$3.72 \times 10^{-1}$	–	3	<b><math>6.05 \times 10^{-2}</math></b>	–	$8.52 \times 10^{-2}$	–
$2.5 \times 10^{-1}$	2	$1.40 \times 10^{-1}$	1.2	$1.40 \times 10^{-1}$	1.4	3	$1.14 \times 10^{-2}$	2.4	$2.10 \times 10^{-2}$	2.0
$1.25 \times 10^{-1}$	2	<b><math>6.60 \times 10^{-2}</math></b>	1.1	$4.87 \times 10^{-2}$	1.5	3	$1.75 \times 10^{-3}$	2.7	$6.73 \times 10^{-3}$	1.6
$6.25 \times 10^{-2}$	2	$3.21 \times 10^{-2}$	1.0	$8.08 \times 10^{-3}$	2.6	4	$3.64 \times 10^{-4}$	2.3	$1.21 \times 10^{-3}$	2.5
$3.125 \times 10^{-2}$	2	$1.58 \times 10^{-2}$	1.0	$1.12 \times 10^{-2}$	-0.5	4	$8.48 \times 10^{-4}$	-1.2	$3.75 \times 10^{-4}$	1.7
$2.5 \times 10^{-2}$	2	$1.26 \times 10^{-2}$		$1.49 \times 10^{-2}$		2	$9.02 \times 10^{-4}$		$5.72 \times 10^{-2}$	
$2.5 \times 10^{-2}$	3	$1.26 \times 10^{-2}$		$1.48 \times 10^{-2}$		3	$9.15 \times 10^{-5}$		$2.84 \times 10^{-3}$	
$2.5 \times 10^{-2}$	4	$1.26 \times 10^{-2}$		$1.55 \times 10^{-2}$		4	$1.06 \times 10^{-4}$		$2.77 \times 10^{-4}$	
$2.5 \times 10^{-2}$	5	$1.26 \times 10^{-2}$		$1.56 \times 10^{-2}$		5	$1.06 \times 10^{-4}$		$1.81 \times 10^{-4}$	

In Table 3.3 we give results of the numerical experiment when the noise magnitude is not small. For the parameters given in the table’s description, the exact values (up to 4 d.p.) of the first and second moments are 0.2718 and 272.3202, respectively. Though for the Euler scheme there is a proportional to  $h$  decrease of the error in computing the mean, there is almost no decrease of the error in the rest of this experiment. The large value of the second moment apparently affects efficiency of the SGC here. For the Euler scheme, increasing  $L$  and decreasing  $h$  can slightly improve accuracy in computing the second moment, e.g. the smallest relative error for the second moment is 56.88% when  $h = 0.03125$  and  $L = 5$  (this level requires 750337 sparse grid points) out of the considered cases of  $h = 0.5, 0.25, 0.125, 0.0625$ , and  $0.03125$  and  $L \leq 5$ . For the mean, increase of the level  $L$  from 2 to 3, 4 or 5 does not improve accuracy. For the second-order scheme (3.2.20), relative errors for the mean can be decreased by increasing  $L$  for a fixed  $h$ : e.g., for  $h = 0.25$ , the relative errors are 0.5121 0.1753, 0.0316 and 0.0086 when  $L = 2, 3, 4$ , and 5, respectively.

We also see in Table 3.3 that the SGC algorithm based on the second-order scheme may not admit higher accuracy than the one based on the Euler scheme, e.g. for  $h = 0.5, 0.25, 0.125$  the second-order scheme yields higher accuracy while the Euler scheme demonstrates higher accuracy

for smaller  $h = 0.0625$  and  $0.03125$ . Further decrease in  $h$  was not considered because this would lead to increase of the dimension of the random space beyond 40 when the sparse grid of Smolyak (3.2.4) may fail and the SGC algorithm may also lose its competitive edge with Monte Carlo-type techniques.

Table 3.3: Comparison of the SGC algorithms based on the Euler scheme (left) and on the second-order scheme (3.2.20) (right). The parameters of the model (3.4.1) are  $x_0 = 0.08$ ,  $\theta_1 = -1$ ,  $\theta_2 = 2$ , and  $T = 1$ . The sparse grid level  $L = 4$ .

$h$	$\rho_1^r(1)$	order	$\rho_2^r(1)$	$\rho_1^r(1)$	$\rho_2^r(1)$
$5 \times 10^{-1}$	$1.72 \times 10^{-1}$	–	$9.61 \times 10^{-1}$	$2.86 \times 10^{-2}$	$7.69 \times 10^{-1}$
$2.5 \times 10^{-1}$	$1.02 \times 10^{-1}$	0.8	$8.99 \times 10^{-1}$	$8.62 \times 10^{-3}$	$6.04 \times 10^{-1}$
$1.25 \times 10^{-1}$	$5.61 \times 10^{-2}$	0.9	$7.87 \times 10^{-1}$	$1.83 \times 10^{-2}$	$7.30 \times 10^{-1}$
$6.25 \times 10^{-2}$	$2.96 \times 10^{-2}$	0.9	$6.62 \times 10^{-1}$	$3.26 \times 10^{-2}$	$8.06 \times 10^{-1}$
$3.125 \times 10^{-2}$	$1.52 \times 10^{-2}$	1.0	$5.64 \times 10^{-1}$	$4.20 \times 10^{-2}$	$8.40 \times 10^{-1}$

Via this example we have shown that the SGC algorithms based on first- and second-order schemes can produce sufficiently accurate results when noise magnitude is small and that the second-order scheme is preferable since for the same accuracy it uses random spaces of lower dimension than the first-order Euler scheme, compare e.g. the error values highlighted by bold font in Table 3.2 and see also the discussion at the end of Section 2.2. When the noise magnitude is large (see Table 3.3), the SGC algorithms do not work well as it was predicted in Section 2.3.

**Example 3.4.2** (Burgers equation with additive noise). Consider the stochastic Burgers equation [82, 192]:

$$du + u \frac{\partial u}{\partial x} dt = \nu \frac{\partial^2 u}{\partial x^2} dt + \sigma \cos(x) dw, \quad 0 \leq x \leq \ell, \quad \nu > 0 \quad (3.4.3)$$

with the initial condition  $u_0(x) = 2\nu \frac{2\pi}{\ell} \frac{\sin(\frac{2\pi}{\ell}x)}{a + \cos(\frac{2\pi}{\ell}x)}$ ,  $a > 1$ , and periodic boundary conditions. In the numerical tests the used values of the parameters are  $\ell = 2\pi$  and  $a = 2$ .

Apply the Fourier collocation method in physical space and the trapezoidal rule in time to (3.4.3):

$$\frac{\vec{u}_{j+1} - \vec{u}_j}{h} - \nu D^2 \frac{\vec{u}_{j+1} + \vec{u}_j}{2} = -\frac{1}{2} D \left( \frac{\vec{u}_{j+1} + \vec{u}_j}{2} \right)^2 + \sigma \Gamma \sqrt{h} \xi_j, \quad (3.4.4)$$

where  $\vec{u}_j = (u(t_j, x_1), \dots, u(t_j, x_M))^\top$ ,  $t_j = jh$ ,  $D$  is the Fourier spectral differential matrix,  $\xi_j$  are i.i.d  $\mathcal{N}(0, 1)$  random variables, and  $\Gamma = (\cos(x_1), \dots, \cos(x_M))^\top$ . The Fourier collocation points are  $x_m = m\frac{\ell}{M}$  ( $m = 1, \dots, M$ ) in physical space and in the experiment we used  $M = 100$ . We aim at computing moments of  $\vec{u}_j$ , which are integrals with respect to the Gaussian measure corresponding to the collection of  $\xi_j$ , and we approximate these integrals using the SGC from Section 2. The use of the SGC amounts to substituting  $\xi_j$  in (3.4.4) by sparse-grid nodes, which results in a system of (deterministic) nonlinear equations of the form (3.4.4). To solve the nonlinear equations, we used the fixed-point iteration method with tolerance  $h^2/100$ .

The errors in computing the first and second moments are measured as follows

$$\begin{aligned} \rho_1^{r,2}(T) &= \frac{\|\mathbb{E}u_{\text{ref}}(T, \cdot) - \mathbb{E}u_{\text{num}}(T, \cdot)\|}{\|\mathbb{E}u_{\text{ref}}(T, \cdot)\|}, & \rho_2^{r,2}(T) &= \frac{\|\mathbb{E}u_{\text{ref}}^2(T, \cdot) - \mathbb{E}u_{\text{num}}^2(T, \cdot)\|}{\|\mathbb{E}u_{\text{ref}}^2(T, \cdot)\|}, \\ \rho_1^{r,\infty}(T) &= \frac{\|\mathbb{E}u_{\text{ref}}(T, \cdot) - \mathbb{E}u_{\text{num}}(T, \cdot)\|_\infty}{\|\mathbb{E}u_{\text{ref}}(T, \cdot)\|_\infty}, & \rho_2^{r,\infty}(T) &= \frac{\|\mathbb{E}u_{\text{ref}}^2(T, \cdot) - \mathbb{E}u_{\text{num}}^2(T, \cdot)\|_\infty}{\|\mathbb{E}u_{\text{ref}}^2(T, \cdot)\|_\infty}, \end{aligned} \quad (3.4.5)$$

where  $\|v(\cdot)\| = \left(\frac{2\pi}{M} \sum_{m=1}^M v^2(x_m)\right)^{1/2}$ ,  $\|v(\cdot)\|_\infty = \max_{1 \leq m \leq M} |v(x_m)|$ ,  $x_m$  are the Fourier collocation points, and  $u_{\text{num}}$  and  $u_{\text{ref}}$  are the numerical solution obtained by the SGC algorithm and the reference solution, respectively. The first and second moments of the reference solution  $u_{\text{ref}}$  were computed by the same solver in space and time (3.4.4) but accompanied by the Monte Carlo method with a large number of realizations ensuring that the statistical errors were negligible.

First, we choose  $\nu = 0.1$  and  $\sigma = 1$ . We obtain the reference solution with  $h = 10^{-4}$  and  $1.92 \times 10^6$  Monte Carlo realizations. The corresponding statistical error is  $1.004 \times 10^{-3}$  for the mean (maximum of the statistical error for  $\mathbb{E}u_{\text{ref}}(0.5, x_j)$ ) and  $9.49 \times 10^{-4}$  for the second moment (maximum of the statistical error for  $\mathbb{E}u_{\text{ref}}^2(0.5, x_j)$ ) with 95% confidence interval, and the corresponding estimates of  $L^2$ -norm of the moments are  $\|\mathbb{E}u_{\text{ref}}(0.5, \cdot)\| \doteq 0.18653$  and  $\|\mathbb{E}u_{\text{ref}}^2(0.5, \cdot)\| \doteq 0.72817$ . We see from the results of the experiment presented in Table 3.4 that for  $L = 2$  the error in computing the mean decreases when  $h$  decreases up to  $h = 0.05$  but the accuracy does not improve

with further decrease of  $h$ . For the second moment, we observe no improvement in accuracy with decrease of  $h$ . For  $L = 4$ , the error in computing the second moment decreases with  $h$ . When  $h = 0.0125$ , increasing the sparse grid level improves the accuracy for the mean:  $L = 3$  yields  $\rho_1^{r,2}(0.5) \doteq 9.45 \times 10^{-3}$  and  $L = 4$  yields  $\rho_1^{r,2}(0.5) \doteq 8.34 \times 10^{-3}$ . As seen in Table 3.4, increase of the level  $L$  also improves accuracy for the second moment when  $h = 0.05, 0.25$ , and  $0.125$ .

Table 3.4: Errors of the SGC algorithm to the stochastic Burgers equation (3.4.3) with parameters  $T = 0.5$ ,  $\nu = 0.1$  and  $\sigma = 1$ .

$h$	$\rho_1^{r,2}(0.5), L = 2$	$\rho_1^{r,2}(0.5), L = 3$	$\rho_2^{r,2}(0.5), L = 2$	$\rho_2^{r,2}(0.5), L = 3$	$\rho_2^{r,2}(0.5), L = 4$
$2.5 \times 10^{-1}$	$1.28 \times 10^{-1}$	$1.3661 \times 10^{-1}$	$4.01 \times 10^{-2}$	$1.05 \times 10^{-2}$	$1.25 \times 10^{-2}$
$1.00 \times 10^{-1}$	$4.70 \times 10^{-2}$	$5.3874 \times 10^{-2}$	$4.48 \times 10^{-2}$	$4.82 \times 10^{-3}$	$4.69 \times 10^{-3}$
$5.00 \times 10^{-2}$	$2.75 \times 10^{-2}$	$2.7273 \times 10^{-2}$	$4.73 \times 10^{-2}$	$5.89 \times 10^{-3}$	$2.82 \times 10^{-3}$
$2.50 \times 10^{-2}$	$2.51 \times 10^{-2}$	$1.4751 \times 10^{-2}$	$4.87 \times 10^{-2}$	$6.92 \times 10^{-3}$	$2.34 \times 10^{-3}$
$1.25 \times 10^{-2}$	$2.67 \times 10^{-2}$	$9.4528 \times 10^{-3}$	$4.95 \times 10^{-2}$	$7.51 \times 10^{-3}$	$2.29 \times 10^{-3}$

Second, we choose  $\nu = 1$  and  $\sigma = 0.5$ . We obtain the first two moments of the reference  $u_{\text{ref}}$  using  $h = 10^{-4}$  and the Monte Carlo method with  $3.84 \times 10^6$  realizations. The corresponding statistical error is  $3.2578 \times 10^{-4}$  for the mean and  $2.2871 \times 10^{-4}$  for the second moment with 95% confidence interval, and the corresponding estimates of  $L^2$ -norm of the moments are  $\|\mathbb{E}u_{\text{ref}}(0.5, \cdot)\| \doteq 1.11198$  and  $\|\mathbb{E}u_{\text{ref}}^2(0.5, \cdot)\| \doteq 0.66199$ .

The results of the experiment are presented in Table 3.5. We see that accuracy is sufficiently high and there is some decrease of errors with decrease of time step  $h$ . However, as expected, no convergence in  $h$  is observed and further numerical tests (not presented here) showed that taking  $h$  smaller than  $1.25 \times 10^{-2}$  and level  $L = 2$  or  $3$  does not improve accuracy. In additional experiments we also noticed that there was no improvement of accuracy for the mean when we increased the level  $L$  up to 5. For the second moment, we observe some improvement in accuracy when  $L$  increases from 2 to 3 (see Table 3.5) but additional experiments (not presented here) showed that further increase of  $L$  (up to 5) does not reduce the errors.

For the errors measured in  $L^\infty$ -norm (3.4.5) we had similar observations (not presented here) as in the case of  $L^2$ -norm.

Table 3.5: Errors of the SGC algorithm applied to the stochastic Burgers equation (3.4.3) with parameters  $\nu = 1$ ,  $\sigma = 0.5$ , and  $T = 0.5$ .

$h$	$\rho_1^{r,2}(0.5), L = 2$	$\rho_2^{r,2}(0.5), L = 2$	$\rho_2^{r,2}(0.5), L = 3$
$2.5 \times 10^{-1}$	$4.94 \times 10^{-3}$	$8.75 \times 10^{-3}$	$8.48 \times 10^{-3}$
$1 \times 10^{-1}$	$8.20 \times 10^{-4}$	$1.65 \times 10^{-3}$	$1.13 \times 10^{-3}$
$5 \times 10^{-2}$	$4.88 \times 10^{-4}$	$1.18 \times 10^{-3}$	$6.47 \times 10^{-4}$
$2.5 \times 10^{-2}$	$3.83 \times 10^{-4}$	$1.08 \times 10^{-3}$	$5.01 \times 10^{-4}$
$1.25 \times 10^{-2}$	$3.45 \times 10^{-4}$	$1.07 \times 10^{-3}$	$4.26 \times 10^{-4}$

In summary, this example has illustrated that SGC algorithms can produce accurate results in finding moments of solutions of nonlinear SPDE when the integration time is relatively small. Comparing Tables 3.4 and 3.5, we observe better accuracy for the first two moments when the magnitude of noise is smaller. In some situations higher sparse grid levels  $L$  improve accuracy but dependence of errors on  $L$  is not monotone. No convergence in time step  $h$  and in level  $L$  was observed which is consistent with our theoretical prediction in Section 2.

**Example 3.4.3** (Stochastic advection-diffusion equation). Consider the stochastic advection-diffusion equation in the Ito sense:

$$du = \left( \frac{\epsilon^2 + \sigma^2}{2} \frac{\partial^2 u}{\partial x^2} + \beta \sin(x) \frac{\partial u}{\partial x} \right) dt + \sigma \frac{\partial u}{\partial x} dw(s), \quad (t, x) \in (0, T] \times (0, 2\pi), \quad (3.4.6)$$

$$u(0, x) = \phi(x), \quad x \in (0, 2\pi),$$

where  $w(s)$  is a standard scalar Wiener process and  $\epsilon \geq 0$ ,  $\beta$ , and  $\sigma$  are constants. In the tests we took  $\phi(x) = \cos(x)$ ,  $\beta = 0.1$ ,  $\sigma = 0.5$ , and  $\epsilon = 0.2$ .

We apply Algorithm 3.3.1 to (3.4.6) to compute the first two moments at a relatively large time  $T = 5$ . The Fourier basis was taken as CONS. Since (3.4.6) has a single noise only, we used one-dimensional Gauss–Hermite quadratures of order  $n$ . The implicitness due to the use of the trapezoidal rule was resolved by the fixed-point iteration with stopping criterion  $h^2/100$ .

As we have no exact solution of (3.4.6), we chose to find the reference solution by Algorithm 4.2 from [420] (a recursive Wiener chaos method accompanied by the trapezoidal rule in time and

Fourier collocation method in physical space) with the parameters: the number of Fourier collocation points  $M = 30$ , the length of time subintervals for the recursion procedure  $h = 10^{-4}$ , the highest order of Hermite polynomials  $P = 4$ , the number of modes approximating the Wiener process  $n = 4$ , and the time step in the trapezoidal rule  $h = 10^{-5}$ . It gives the second moment in the  $L^2$ -norm  $\|\mathbb{E}u_{\text{ref}}^2(1, \cdot)\| \doteq 1.065195$ . The errors are computed as follows

$$\varrho_2^2(T) = \left| \|\mathbb{E}u_{\text{ref}}^2(T, \cdot)\| - \|\mathbb{E}u_{\text{numer}}^2(T, \cdot)\| \right|, \quad \varrho_2^{r,2}(T) = \frac{\varrho_2^2(T)}{\|\mathbb{E}u_{\text{ref}}^2(T, \cdot)\|}, \quad (3.4.7)$$

where the norm is defined as in (3.4.5).

Table 3.6: Errors in computing the second moment of the solution to the stochastic advection-diffusion equation (3.4.6) with  $\sigma = 0.5$ ,  $\beta = 0.1$ ,  $\epsilon = 0.2$  at  $T = 5$  by Algorithm 3.3.1 with  $l_* = 20$  and the one-dimensional Gauss–Hermite quadrature of order  $n = 2$  (left) and  $n = 3$  (right).

$h$	$\varrho_2^{r,2}(5)$	order	CPU time (sec.)	$\varrho_2^{r,2}(5)$	order	CPU time (sec.)
$5 \times 10^{-2}$	$1.01 \times 10^{-3}$	–	7.41	$1.06 \times 10^{-3}$	–	$1.10 \times 10$
$2 \times 10^{-2}$	$4.07 \times 10^{-4}$	1.0	$1.65 \times 10$	$4.25 \times 10^{-4}$	1.0	$2.43 \times 10$
$1 \times 10^{-2}$	$2.04 \times 10^{-4}$	1.0	$3.43 \times 10$	$2.12 \times 10^{-4}$	1.0	$5.10 \times 10$
$5 \times 10^{-3}$	$1.02 \times 10^{-4}$	1.0	$6.81 \times 10$	$1.06 \times 10^{-4}$	1.0	$1.00 \times 10^2$
$2 \times 10^{-3}$	$4.08 \times 10^{-5}$	1.0	$1.70 \times 10^2$	$4.25 \times 10^{-5}$	1.0	$2.56 \times 10^2$
$1 \times 10^{-3}$	$2.04 \times 10^{-5}$	1.0	$3.37 \times 10^2$	$2.12 \times 10^{-5}$	1.0	$5.12 \times 10^2$

The results of the numerical experiment are given in Table 3.6. We observe first-order convergence in  $h$  for the second moments. We notice that increasing the quadrature order  $n$  from 2 to 3 does not improve accuracy which is expected. Indeed, the used trapezoidal rule is of weak order one in  $h$  in the case of multiplicative noise and more accurate quadrature rule cannot improve the order of convergence. We note in passing that in the additive noise case we expect to see the second order convergence in  $h$  when  $n = 3$  due to the properties of the trapezoidal rule.

In conclusion, we showed that recursive Algorithm 3.3.1 can work effectively for accurate computing of second moments of solutions to linear stochastic advection-diffusion equations at relatively large time. We observed convergence of order one in  $h$ .

## Chapter 4

# Comparison between Wiener chaos methods and stochastic collocation methods

In this chapter, we compare Wiener chaos expansion and stochastic collocation methods for linear advection-reaction-diffusion equations with multiplicative white noise. Both methods are constructed based on a recursive multi-stage algorithm for long-time integration. We derive error estimates for both methods and compare their numerical performance. Numerical results confirm that the recursive multi-stage stochastic collocation method is of order  $\Delta$  (time step size) in the second-order moments while the recursive multi-stage Wiener chaos method is of order  $\Delta^N + \Delta^2$  ( $N$  is the order of Wiener chaos) for advection-diffusion-reaction equations with commutative noises, in agreement with the theoretical error estimates. However, for non-commutative noises, both methods are of order one in the second-order moments.

## 4.1 Introduction

Partial differential equations (PDEs) driven by white noise have different interpretation of stochastic products and lead to different numerical approximations, unlike the PDEs driven by color noise. Specifically, stochastic products for white noise are usually interpreted with two different products: the Ito-Wick product and the Stratonovich product, see e.g. [7]. Different products lead to different performance of numerical solvers for PDEs driven by white noise, especially when Wiener chaos expansion (WCE) and stochastic collocation methods (SCM) in random space are used. In this chapter, we will show theoretically and through numerical examples that for white noise driven PDEs, WCE and SCM have quite different performance when the noises are commutative. This is different from how WCE and SCM behave for PDEs driven by color noise. For elliptic equations with color noise, it is demonstrated in [11] that there are only small differences in the numerical performance of generalized polynomial chaos expansion and SCM.

To apply WCE and SCM, we first discretize the Brownian motion with its truncated spectral expansion, see e.g. [333, Chapter IX] and [265], and subsequently we employ the corresponding functional expansion (WCE and SCM) to represent the solution in random space. In principle, we can employ any functional expansion, however, different expansions are preferred for different stochastic products because of computational efficiency. In practice, WCE is associated with the Ito-Wick product, see (4.2.2), as the product is defined with Wiener chaos modes yielding a weakly coupled system (lower-triangular system) of PDEs for linear equations. On the other hand, SCM is associated with the Stratonovich product, see (4.2.6), yielding a decoupled system of PDEs. These different formulations lead to different numerical performance as we demonstrate in Section 4.4; in particular, WCE can be of second-order convergence in time while SCM is only of first-order in time in the second-order moments for commutative noises. Further, when the noises serve as the advection coefficients, SCM can be more accurate than WCE when both methods are of first order convergence as the SCM (Stratonovich formulation) can lead to smaller diffusion coefficient than

those for WCE (Ito-Wick formulation).

Both methods are actually Wong-Zakai type approximations [399, 400], according to which we discretize only the Brownian motion, hence resulting in PDEs with finite dimensional random inputs. The latter can be solved numerically using a variety of space-time discretization methods. Here we will employ functional expansion methods in random space, e.g., WCE [49, 265] and SCM [403]. Compared to the Monte Carlo method, these functional expansion methods have no statistical errors and allow efficient short-time integration of SPDEs [49, 50, 192, 265, 266, 420].

However, a fundamental limitation of these expansion methods is the exponential growth of error with time and the increasing complexity as the number of random variables is increasing, generated by the discretization of the Brownian motion. To deal with this complexity, a recursive WCE method was proposed in [265] for the Zakai equation of nonlinear filtering with uncorrelated observations. More recently, a recursive multi-stage approach was developed to efficiently solve linear stochastic advection-diffusion-reaction equations using either WCE [420].

Some numerical results of WCE for SPDEs have been presented in [420] for linear advection-diffusion-reaction equations and in [192] for nonlinear SPDEs including the stochastic Burgers equation and the Navier-Stokes equations. These numerical results have demonstrated that WCE in conjunction with the recursive multi-stage approach are efficient for long-time integration of linear advection-diffusion-reaction equations. Although the number of operations for the recursive multi-stage WCE is of order  $M^4$ , where  $M$  is the number of nodes employed in the discretization of physical space, this computational complexity can be reduced to the order of  $M^2$  using sparse representations (see e.g. [352]), as demonstrated in [420].

The main point of this chapter is the derivation of theoretical error estimates for both WCE and SCM methods and subsequent comparison of the numerical performance of the two methods for commutative and non-commutative noises. In addition, we will develop a recursive multi-stage SCM using a spectral truncation of Brownian motion. Specifically, in this chapter we will derive the error estimate of WCE for linear advection-diffusion-reaction equations with white noise in

the advection velocity and that of SCM with white noise in the reaction rate. We note that the convergence rate of WCE is known only for linear advection-diffusion-reaction equations with white noise in the reaction rate although the convergence of WCE for linear advection-diffusion-reaction equations has been studied for some time [264, 265, 266, 268].

This chapter is organized as follows. After the Introduction section, in Section 4.2, we review the WCE method and SCM for linear parabolic SPDEs and develop a new recursive SCM using a spectral truncation of Brownian motion, following the same recursive procedure as WCE in [265, 420]. In Section 4.3 we present the error estimates for both methods for linear advection-diffusion-reaction equations, with the proofs presented in Section 4.5. In Section 4, we present numerical results of WCE and SCM for linear SPDEs with both commutative and non-commutative noises and verify the error estimates of WCE and SCM for commutative noises.

## 4.2 Review of Wiener chaos and stochastic collocation

In this section, we briefly review WCE and SCM for the linear SPDE (2.2.1).

In both WCE and SCM, we discretize the Brownian motion using the following spectral representations (see e.g. [265, 420]):

$$\lim_{n \rightarrow \infty} \mathbb{E}[(w(t) - w^{(n)}(t))^2] = 0, \quad w^{(n)}(t) = \sum_{i=1}^n \int_0^t m_i(s) ds \xi_i, \quad t \in [0, T], \quad (4.2.1)$$

where  $\{m_i\}_{i=1}^{\infty}$  is a CONS (complete orthonormal system) in  $L^2([0, T])$ , and  $\xi_i$  are mutually independent standard Gaussian random variables. The expansion (4.2.1) is an extension of Fourier expansion of Brownian motion that is the Wiener construction [333, Chapter IX] and is used in [214, 237].

### 4.2.1 Wiener chaos expansion (WCE)

The SPDE (2.2.1) with finite dimensional noises can be written in the following form using the Ito-Wick product

$$\begin{aligned} du(t, x) &= [\mathcal{L}u(t, x) + f(x)] dt + \sum_{k=1}^q [\mathcal{M}_k u(t, x) + g_k(x)] \diamond \dot{w}_k dt, \quad (t, x) \in (0, T] \times \mathcal{D}, \\ u(0, x) &= u_0(x), \quad x \in \mathcal{D}, \end{aligned} \quad (4.2.2)$$

where  $\dot{w}_k$  is formally the first-order derivative of  $w_k$  in time, i.e.,  $\dot{w} = \frac{d}{dt}w$ . To obtain the coefficients  $\varphi_\alpha(t, x; \phi)$ , we approximate  $w_k$  with the spectral truncation (4.2.1),  $w_k^{(n)}$ , and then we substitute the representation (2.2.3) into (2.2.1); by multiplying  $\xi_\alpha$  on both sides of (2.2.1), and taking expectation with the properties of the Ito-Wick product  $\xi_\alpha \diamond \xi_\beta = \sqrt{\frac{(\alpha+\beta)!}{\alpha!\beta!}} \xi_{\alpha+\beta}$  and  $\mathbb{E}[\xi_\alpha \xi_\beta] = \delta_{\alpha=\beta}$ , we then have that the coefficients  $\varphi_\alpha(t, x; \phi)$  satisfy the following propagator

$$\begin{aligned} \frac{\partial \varphi_\alpha(t, x; \phi)}{\partial t} &= \mathcal{L}\varphi_\alpha(t, x; \phi) + f(x) \mathbf{1}_{\{|\alpha|=0\}} \\ &\quad + \sum_{k=1}^q \sum_{l=1}^n \alpha_{k,l} m_l(t) [\mathcal{M}_k \varphi_{\alpha^-(k,l)}(t, x; \phi) + g_k(x) \mathbf{1}_{\{|\alpha|=1\}}], \quad t \in (0, T], \\ \varphi_\alpha(0, x) &= \phi(x) \mathbf{1}_{\{|\alpha|=0\}}, \end{aligned} \quad (4.2.3)$$

where  $\alpha^-(k, l)$  is the multi-index with components

$$(\alpha^-(k, l))_{i,j} = \begin{cases} \max(0, \alpha_{i,j} - 1), & \text{if } i = k \text{ and } j = l, \\ \alpha_{i,j}, & \text{otherwise.} \end{cases} \quad (4.2.4)$$

In practical computations, we are only interested in the truncated Wiener chaos solution (2.3.1). However, the error induced by the truncation of Wiener chaos expansion grows exponentially with time and thus WCE is not efficient for long-time integration. To control the error behavior, we can use the recursive WCE (see Algorithm 2.4.1) for computing the second moments,  $\mathbb{E}[u^2(t, x)]$ , of the solution of the SPDE (2.2.1). See Chapter 2 for more details.

Note that in Algorithm 2.4.1 we discretize the Brownian motion using the following spectral representation in a multi-element version, i.e., using  $K$  multi-elements [265, 420]:

$$w^{(n,K)}(t) = \sum_{k=1}^K \sum_{i=1}^n \int_{\mathbf{t}_{k-1} \wedge t}^{\mathbf{t}_k \wedge t} m_{i,k}(s) ds \xi_{i,k}, \quad t \in [0, T], \quad (4.2.5)$$

where  $0 = \mathbf{t}_0 < \mathbf{t}_1 < \dots < \mathbf{t}_K = T$ ,  $\mathbf{t}_k \wedge t$  is the minimum of  $\mathbf{t}_k$  and  $t$ ,  $\{m_{i,k}\}_{i=1}^\infty$  is a CONS in  $L^2([\mathbf{t}_k, \mathbf{t}_{k+1}])$ , and  $\xi_{i,k}$  are mutually independent standard Gaussian random variables. This approximation of the Brownian motion will be also used in the stochastic collocation methods presented below.

## 4.2.2 Stochastic collocation method (SCM)

This method leads to fully decoupled system instead of a weakly coupled system from the WCE.

First, we rewrite the SPDE (2.2.1) with finite dimensional noises in Stratonovich form

$$\begin{aligned} du(t, x) &= [\tilde{\mathcal{L}}u(t, x) + f(x)] dt + \sum_{k=1}^q [\mathcal{M}_k u(t, x) + g_k(x)] \circ \dot{w}_k dt, \quad (t, x) \in (0, T] \times \mathcal{D}, \\ u(0, x) &= u_0(x), \quad x \in \mathcal{D}, \end{aligned} \quad (4.2.6)$$

where  $\tilde{\mathcal{L}}u = \mathcal{L}u - \frac{1}{2} \sum_{1 \leq k \leq q} \mathcal{M}_k [\mathcal{M}_k u + g_k]$ . Second, we replace the Brownian motion with its multi-element spectral expansion (4.2.5), and obtain the following partial differential equation with smooth random inputs:

$$\begin{aligned} d\tilde{u}(t, x) &= [\tilde{\mathcal{L}}\tilde{u}(t, x) + f(x)] dt + \sum_{k=1}^q [\mathcal{M}_k \tilde{u}(t, x) + g_k(x)] dw_k^{(n,K)}(t), \quad (t, x) \in (0, T] \times \mathcal{D}, \\ \tilde{u}(0, x) &= u_0(x), \quad x \in \mathcal{D}. \end{aligned} \quad (4.2.7)$$

Now we can apply standard numerical techniques of high integration to obtain  $p$ -th moments of the solution to (2.2.1)

$$\mathbb{E}[\tilde{u}_{\mathbf{n},\mathbf{K}}^p(x, t)] = \frac{1}{(2\pi)^{nq\mathbf{K}/2}} \int_{\mathbb{R}^{nq\mathbf{K}}} F(u_0(x), x, t, \mathbf{y}) e^{-\frac{\mathbf{y}^\top \mathbf{y}}{2}} d\mathbf{y}, \quad p = 1, 2, \dots \quad (4.2.8)$$

where  $\mathbf{y} = (y_{i,k,l})$ ,  $i \leq n, k \leq \mathbf{K}, l \leq q$  and the functional  $F$  represents the solution functional for (4.2.7). Here we employ sparse grid collocation [125, 358] if the dimension  $n\mathbf{K}$  is moderately large. As pointed out in [9, 403], we are led to a fully *decoupled* system of equations as in the case of Monte Carlo methods.

In practice, we use the sparse grid quadrature rule (3.2.4). Here again, the direct application of SCM is efficient only for short-time integration. To achieve long-time integration, we apply the recursive multi-stage idea used in Algorithm 2.4.1 and similarly we have the following algorithm for the second moments of the approximate solution when  $f = g_k = 0$ , see Appendix 4.6 for its derivation.

**Algorithm 4.2.1** (Recursive multi-stage stochastic collocation method). *Choose a CONS  $\{e_m(x)\}_{m \geq 1}$  and its truncation  $\{e_m(x)\}_{m=1}^M$ ; a time step  $\Delta$ ; the sparse grid level  $\mathbf{L}$  and  $\mathbf{n}$ , which together with the number of noises  $q$  determine the sparse grid  $\mathcal{H}_{\mathbf{L}}^{nq}$  which contains  $\eta(\mathbf{L}, nq)$  sparse grid points.*

*Step 1. For each  $m = 1, \dots, M$ , solve the system of equations (4.2.7) on the sparse grid  $\mathcal{H}_{\mathbf{L}}^{nq}$  in the time interval  $[0, \Delta]$  with the initial condition  $\phi(x) = e_m(x)$  and denote the obtained solution as  $v_\kappa(\Delta, x; e_m)$ ,  $m = 1, \dots, M$ , and  $\kappa = 1, \dots, \eta(\mathbf{L}, nq)$ . Also, choose a time step size  $\delta t$  to solve (4.2.7) numerically.*

*Step 2. Evaluate  $\Phi_m(0; \mathbf{L}, \mathbf{n}, M) = (u_0, e_m)$ ,  $m = 1, \dots, M$ , where  $u_0(x)$  is the initial condition for (2.2.1), and  $h_{\kappa,l,m} = (v_\kappa(\Delta, \cdot; e_l), e_m)$ ,  $l, m = 1, \dots, M$ .*

*Step 3. Recursively compute the covariance matrices  $H_{lm}(\mathbf{t}_i; \mathbf{L}, \mathbf{n}, M)$ ,  $l, m = 1, \dots, M$ , as follows:*

$$H_{lm}(0; \mathbf{N}, \mathbf{n}, M) = (u_0, e_l)(u_0, e_m),$$

$$H_{lm}(\mathbf{t}_i; \mathbf{L}, \mathbf{n}, \mathbf{M}) = \sum_{j,k=1}^{\mathbf{M}} H_{jk}(\mathbf{t}_{i-1}; \mathbf{L}, \mathbf{n}, \mathbf{M}) \sum_{\kappa=1}^{\eta(\mathbf{L}, \mathbf{n}q)} h_{\kappa,j,l} h_{\kappa,k,m} \mathbb{W}_{\kappa}, \quad i = 1, \dots, \mathbf{K},$$

where  $\mathbb{W}_{\kappa}$  are the sparse grid quadrature weights corresponding to the sparse grid points in Step 1, and obtain the second moments  $\mathbb{E}[\tilde{u}_{\Delta, \mathbf{L}, \mathbf{n}}^{\mathbf{M}}(t_i, x)]^2$  of the approximate solution to (2.2.1) as

$$\mathbb{E}[\tilde{u}_{\Delta, \mathbf{L}, \mathbf{n}}^{\mathbf{M}}(t_i, x)]^2 = \sum_{l,m=1}^{\mathbf{M}} H_{lm}(\mathbf{t}_i; \mathbf{L}, \mathbf{n}, \mathbf{M}) e_l(x) e_m(x), \quad i = 1, \dots, \mathbf{K}. \quad (4.2.9)$$

**Remark 4.2.2.** Similar to Algorithm 2.4.1, the cost of this algorithm is  $\frac{T}{\Delta} \eta(\mathbf{L}, \mathbf{n}q) \mathbf{M}^4$  and the storage is  $\eta(\mathbf{L}, \mathbf{n}q) \mathbf{M}^2$ . The total cost can be reduced to the order of  $\mathbf{M}^2$  by adopting some reduced order methods in physical space. The discussion on computational efficiency of the recursive WCE methods, see [420, Remark 4.1], is also valid for Algorithm 4.2.1.

### 4.3 Error estimates

Though WCE and SCM use the same spectral truncation of Brownian motion, the former is associated with the Ito-Wick product while the latter is related to the Stratonovich product. Note that WCE employs orthogonal polynomials as basis and SCM does not have such orthogonality. This difference allows WCE to have better convergence rate than SCM in the second-order moments, see Corollary 4.3.2 and Remark 4.3.4.

Assume that there exist a positive constant  $\delta_{\mathcal{L}}$  and a real number  $C_{\mathcal{L}}$  such that

$$(\mathcal{L}v, v) + \delta_{\mathcal{L}} \|v\|_{H^1}^2 \leq C_{\mathcal{L}} \|v\|^2, \quad (4.3.1)$$

where  $(\cdot, \cdot)$  is the duality between  $H^{-1}(\mathcal{D})$  and  $H^1(\mathcal{D})$  associated with the inner-product over  $L^2(\mathcal{D})$ . The operator  $\mathcal{L}$  generates a semi-group  $\{\mathcal{T}_t\}_{t \geq 0}$ , which has the following properties: for  $g \in H^r(\mathcal{D})$ ,

$$\|\mathcal{T}_t g\|_{H^r}^2 \leq C(r, \mathcal{L}) e^{2C_{\mathcal{L}} t} \|g\|_{H^r}^2, \quad (4.3.2)$$

where  $C(0, \mathcal{L}) = 1$  and

$$\int_s^t e^{2C_{\mathcal{L}}(t-\theta)} \|\mathcal{T}_t g\|_{H^{r+1}}^2 d\theta \leq \delta_{\mathcal{L}}^{-1} C(r, \mathcal{L}) e^{2C_{\mathcal{L}}(t-s)} \|g\|_{H^r}^2. \quad (4.3.3)$$

Also, we assume that there exists a constant  $\tilde{C}(r, \mathcal{M})$  such that

$$\|\mathcal{M}_k g\|_{H^r}^2 \leq \tilde{C}(r, \mathcal{M}) \|g\|_{H^{r+1}}^2, \quad \text{for } g \in H^{r+1}, k = 1, \dots, q. \quad (4.3.4)$$

For the WCE for the SPDE (2.2.1) with single noise, i.e.,  $q = 1$ , we have the following result.

**Theorem 4.3.1.** *Assume that  $\sigma_{i,k}, a_{i,j}, b_i, c, \nu_k$  in (2.2.2) belong to  $C_b^{r+1}(\mathcal{D})$  and  $u_0 \in H^r(\mathcal{D})$ , where  $r \geq N + 2$  and  $N$  is the order of Wiener chaos. Let  $u$  in (2.2.3) be the Wiener chaos solution to (2.2.1) and  $u_{N,n}$  in (2.3.1) the truncated Wiener chaos solution. Then for  $\mathbf{t}_i = i\Delta$ , it holds that, when  $C_1 < \delta_{\mathcal{L}}$ ,*

$$\begin{aligned} & \mathbb{E}[\|u_{N,n}(\mathbf{t}_i, \cdot) - u(\mathbf{t}_i, \cdot)\|^2] \\ & \leq (C_{\lfloor r \rfloor} \Delta)^N e^{2C_{\mathcal{L}}T} \left[ \frac{e^{C_{\lfloor r \rfloor}T}}{(N+1)!} + \frac{(C_{\lfloor r \rfloor} \Delta)^{\lfloor r \rfloor - N - 1}}{\lfloor r \rfloor!} \frac{\delta_{\mathcal{L}}}{\delta_{\mathcal{L}} - C_1} \right] \|u_0\|_{H^r}^2 \\ & \quad + 2C^2(N+2, \mathcal{L}) e^{2C_{N+2}T + 2C_{\mathcal{L}}T} \frac{\Delta^2}{n\pi^2} \|u_0\|_{H^{N+2}}^2, \end{aligned}$$

where the constants  $\delta_{\mathcal{L}}$  and  $C_{\mathcal{L}}$  are from (4.3.1) and  $C_{\lfloor r \rfloor} = C(\lfloor r \rfloor, \mathcal{L}) \tilde{C}(\lfloor r \rfloor - 1, \mathcal{M})$ . The constants  $C(\lfloor r \rfloor, \mathcal{L})$  and  $\tilde{C}(\lfloor r \rfloor - 1, \mathcal{M})$  are from (4.3.2) and (4.3.4), respectively.

From Theorem 4.3.1, we have that the mean-square error of the recursive multi-stage WCE is  $\mathcal{O}(\Delta^{N/2}) + \mathcal{O}(\Delta)$ .

**Corollary 4.3.2.** *Under the conditions of Theorem 4.3.1, we have*

$$\left| \mathbb{E}[\|u_{N,n}(\mathbf{t}_i, \cdot)\|^2] - \mathbb{E}[\|u(\mathbf{t}_i, \cdot)\|^2] \right| = \mathbb{E}[\|u_{N,n}(\mathbf{t}_i, \cdot) - u(\mathbf{t}_i, \cdot)\|^2]$$

$$\begin{aligned} &\leq (C_{\lfloor r \rfloor} \Delta)^N e^{2C_{\mathcal{L}} T} \left[ \frac{e^{C_{\lfloor r \rfloor} T}}{(N+1)!} + \frac{(C_{\lfloor r \rfloor} \Delta)^{\lfloor r \rfloor - N - 1}}{\lfloor r \rfloor!} \frac{\delta_{\mathcal{L}}}{\delta_{\mathcal{L}} - C_1} \right] \|u_0\|_{H^r}^2 \\ &\quad + 2C^2(N+2, \mathcal{L}) e^{2C_{N+2} T + 2C_{\mathcal{L}} T} \frac{\Delta^2}{n\pi^2} \|u_0\|_{H^{N+2}}^2. \end{aligned}$$

This corollary claims that the convergence rate of the error in second-order moments is twice that of the mean-square error, i.e.,  $\mathcal{O}(\Delta^N) + \mathcal{O}(\Delta^2)$ . This corollary can be proved by the orthogonality of WCE. In fact, it holds that

$$\mathbb{E}[u^2(\mathbf{t}_i, x)] - \mathbb{E}[u_{N,n}^2(\mathbf{t}_i, x)] = \mathbb{E}[(u(\mathbf{t}_i, x) - u_{N,n}(\mathbf{t}_i, x))^2], \quad (4.3.5)$$

as the different terms in the Cameron-Martin basis are mutually orthogonal. Then integrating over the physical domain and by the Fubini Theorem, we reach the conclusion by Theorem 4.3.1.

For SCM for the SPDE (2.2.1), we consider here Wong-Zakai type approximation (4.2.7), as the case of  $n = 1$  has been considered in [50, 194].

**Theorem 4.3.3.** *Suppose that  $u_0 \in L^4(\mathcal{D})$ . Let  $u(t, x)$  be the solution to (2.2.1) and  $\tilde{u}(t, x)$  the solution to (4.2.7). When  $f = g_r = \sigma_{i,r} = 0$ ,  $\nu_r$ ,  $c$  is bounded, and  $\nu_r$  is Lipschitz continuous with respect to its all variables, we have one-step error as, i.e., when  $K = 1$ ,*

$$\mathbb{E}[|u(t, x) - \tilde{u}(t, x)|^2] \leq C \exp(CT)(T^3 + T^2)n^{-1+\varepsilon}, \quad (4.3.6)$$

where  $\varepsilon > 0$  is sufficiently small and the constant  $C$  depends on the boundedness of  $\nu_r(t, x)$  and  $c(t, x)$ . When  $K > 1$  ( $K\Delta = T$ ), we have for  $\mathbf{t}_i = i\Delta$ ,  $1 \leq i \leq K$ ,

$$\mathbb{E}[|u(\mathbf{t}_i, x) - \tilde{u}(\mathbf{t}_i, x)|^2] \leq C \exp(CT)\Delta n^{-1+\varepsilon}. \quad (4.3.7)$$

**Remark 4.3.4.** *Under some smoothness assumptions on the coefficients and the initial condition of Equation (2.2.1), the error  $\left| \mathbb{E}[|\tilde{u}_{\text{SCM}}(\mathbf{t}_i, x)|^2] - \mathbb{E}[|\tilde{u}(\mathbf{t}_i, x)|^2] \right|$  is expected to be  $\Delta^{L-1}$  ( $T^L$  in*

one-step) as in multiple Stratonovich integral expansion [50].

For SCM, we do not have orthogonality as for WCE, see (4.3.5), and thus we expect that the convergence in moments is as usual for SDEs, i.e., at most first order in  $\Delta$  when  $n$  is small. This finding will be verified by numerical examples in the next section.

## 4.4 Numerical results

In this section, we compare Algorithms 2.4.1 and 4.2.1 for linear stochastic advection-diffusion-reaction equations with commutative and non-commutative noises. We will test the computational performance of these two methods in terms of accuracy and computational cost. All the tests were run using Matlab R2012b, on a Macintosh desktop computer with Intel Xeon CPU E5462 (quad-core, 2.80 GHz). Every effort was made to program and execute the different algorithms as much as possible in an identical way.

The computational complexity for Algorithm 2.4.1 is  $\binom{N+nq}{N} \frac{T}{\Delta} M^4$  [420] and that for Algorithm 4.2.1 is  $\eta(L, nq) \frac{T}{\Delta} M^4$ . The ratio of the computational cost of SCM over that of WCE is  $\eta(L, nq) / \binom{N+nq}{N}$ . For example, when  $N = 1$  and  $L = 2$ , the ratio is  $(1 + 2nq)/(1 + nq)$ , which will be used in the three numerical examples. The complexity is increasing exponentially with  $nq$  and  $L$ , see e.g. [125], or  $N$  but is increasing linearly with  $\frac{T}{\Delta}$ . Hence, we only consider low values of  $L$  and  $N$ .

**Example 4.4.1** (Single noise). *We consider a single noise in (2.2.1) in Ito's form over the domain  $(0, T] \times (0, 2\pi)$ :*

$$du = [(\epsilon + \frac{1}{2}\sigma^2)\partial_x^2 u + \beta \sin(x)\partial_x u] dt + \sigma \partial_x u dw(t), \quad (4.4.1)$$

*or in Stratonovich form*

$$du = [\epsilon \partial_x^2 u + \beta \sin(x)\partial_x u] dt + \sigma \partial_x u \circ dw(t), \quad (4.4.2)$$

with non-random initial condition  $u(0, x) = \cos(x)$ , where  $w(t)$  is a standard scalar Wiener process,  $\epsilon > 0$ ,  $\beta$ ,  $\sigma$  are constants.

In this example, we compare Algorithms 2.4.1 and 4.2.1 for (4.4.1) with the parameters  $\beta = 0.1$ ,  $\sigma = 0.5$  and  $\epsilon = 0.02$ . We will show that the recursive multi-stage WCE is at most of order  $\Delta^2$  in the second-order moments and the recursive multi-stage SCM is of order  $\Delta$ .

In Step 1, Algorithm 2.4.1, we employ Crank-Nicolson in time and Fourier collocation in physical space. We obtain the solution by Algorithm 2.4.1 with the same solver but finer resolution as a reference solution, as we have no exact solution to (4.4.1). The reference solution is obtained by  $M = 30$ ,  $\Delta = 10^{-4}$ ,  $N = 4$ ,  $n = 4$ ,  $\delta t = 10^{-5}$ . It gives the second-order moments in  $L^2$ -norm  $\|\mathbb{E}[u_{\text{ref}}^2]\| = 1.0651945500628588$  and in the  $L^\infty$ -norm  $\|\mathbb{E}[u_{\text{ref}}^2]\|_\infty = 0.51747461411047124$ .

The errors are computed with (3.4.7) and in the following sense

$$\varrho_2^\infty(T) = \left| \|\mathbb{E}[u_{\text{ref}}^2(T)]\|_\infty - \|\mathbb{E}[u_{\text{numer}}^2(T)]\|_\infty \right|, \quad \varrho_2^{r,\infty}(T) = \frac{\varrho_2^\infty(T)}{\|\mathbb{E}[u_{\text{ref}}^2(T)]\|_\infty}, \quad (4.4.3)$$

where  $\|v\| = \left( \frac{2\pi}{M} \sum_{m=1}^M v^2(x_m) \right)^{\frac{1}{2}}$ ,  $\|v\|_\infty = \max_{1 \leq m \leq M} |v(x_m)|$ ,  $x_m$  are the Fourier collocation points and  $u_{\text{numer}}$  is the numerical solution.

With the above truncation parameters, the recursive WCE is of second-order convergence in  $\Delta$  for the second-order moments when  $N = 2$  and of first-order convergence when  $N = 1$  from Table 4.1. We note that when  $N = 3$ , the method is still second-order in  $\Delta$  (not presented here). This verifies the estimate  $\mathcal{O}(\Delta^N) + \mathcal{O}(\Delta^2)$  in Corollary 4.3.2.

Table 4.1: Algorithm 2.4.1: recursive multi-stage Wiener chaos method for (4.4.1) at  $T = 5$ :  $\sigma = 0.5$ ,  $\beta = 0.1$ ,  $\epsilon = 0.02$ , and  $M = 20$ ,  $n = 1$ .

$\Delta$	$\delta t$	$N$	$\varrho_2^{r,2}(T)$	order	$\varrho_2^{r,\infty}(T)$	order	CPU time (sec.)
1.0e-1	1.0e-2	1	1.5249e-2	–	8.8177e-3	–	3.57
1.0e-2	1.0e-3	1	1.5865e-3	$\Delta^{0.98}$	8.9310e-4	$\Delta^{0.99}$	33.22
1.0e-3	1.0e-4	1	1.5934e-4	$\Delta^{1.00}$	8.9429e-5	$\Delta^{1.00}$	348.03
1.0e-1	1.0e-2	2	1.9070e-4	–	4.1855e-5	–	5.14
1.0e-2	1.0e-3	2	2.0088e-6	$\Delta^{1.98}$	4.2889e-7	$\Delta^{1.99}$	51.75
1.0e-3	1.0e-4	2	2.0386e-8	$\Delta^{1.99}$	4.8703e-9	$\Delta^{1.94}$	490.04

In Step 1, Algorithm 4.2.1, we use a third-order strong-stability-preserving Runge-Kutta scheme [212] in time and Fourier collocation method in physical space. A Crank-Nicolson scheme in time was also used but led to no changes in convergence order or in accuracy and thus the results are not presented. We observe in Table 4.2 that the convergence order for second-order moments is one in  $\Delta$  even when the sparse grid level  $L$  is 2, 3 and 4 (the latter is not presented here). The errors for  $L = 3$  are more than half in magnitude smaller than those for  $L = 2$  while the time cost for  $L = 3$  is about 1.5 times of that for  $L = 2$ .

Table 4.2: Algorithm 4.2.1: recursive multi-stage stochastic collocation method for (4.4.1) at  $T = 5$ :  $\sigma = 0.5$ ,  $\beta = 0.1$ ,  $\epsilon = 0.02$ , and  $M = 20$ ,  $n = 1$ .

$\Delta$	$\delta t$	$L$	$\varrho_2^{r,2}(T)$	order	$\varrho_2^{r,\infty}(T)$	order	CPU time (sec.)
1.0e-1	1.0e-2	2	3.7467e-4	–	3.0692e-3	–	3.67
1.0e-2	1.0e-3	2	3.7496e-5	$\Delta^{1.00}$	3.0441e-4	$\Delta^{1.00}$	34.25
1.0e-3	1.0e-4	2	3.7501e-6	$\Delta^{1.00}$	3.0416e-5	$\Delta^{1.00}$	332.06
1.0e-1	1.0e-2	3	1.4095e-4	–	3.2416e-4	–	5.18
1.0e-2	1.0e-3	3	1.3518e-5	$\Delta^{1.02}$	2.9879e-5	$\Delta^{1.04}$	50.95
1.0e-3	1.0e-4	3	1.3459e-6	$\Delta^{1.00}$	2.9623e-6	$\Delta^{1.00}$	494.07

In summary, from Tables 4.1 and 4.2, we observe that the recursive multi-stage WCE is  $\mathcal{O}(\Delta^N) + \mathcal{O}(\Delta^2)$  and the recursive multi-stage SCM is  $\mathcal{O}(\Delta)$ , as predicted by the error estimates in Section 4.3. While the SCM and the WCE are of the same order when  $N = 1$  and  $L \geq 2$ , the former can be more accurate than the latter. In fact, when  $N = 1$  and  $L = 2$ , the recursive multi-stage SCM error is almost two orders of magnitude smaller than the recursive multi-stage WCE while the computational cost for both is almost the same, as predicted ( $\binom{N+nq}{N} = \eta(L, nq) = 2$ ). The recursive multi-stage WCE with  $N = 2$  is of order  $\Delta^2$  and its errors are almost two orders of magnitude smaller than those by the recursive multi-stage SCM (with level 2 or 3) for the second-order moments.

In this example, the recursive multi-stage SCM outperforms the recursive multi-stage WCE with  $N = 1$ . The reason can be as follows. In SCM, we solve an advection-dominant equation rather than a diffusion-dominant equation in WCE, as SCM is associated with the Stratonovich product which leads to the removal of the term  $\frac{1}{2}\sigma^2\partial_x^2u$  in the resulting equation, see (4.4.2). The larger  $\sigma$  is, the more dominant the diffusion is. In fact, results for  $\sigma = 1$  and  $\sigma = 0.1$  (not presented here)

show that when  $\sigma = 1$ , the relative error of SCM with  $L = 2$  is almost three orders of magnitude smaller than WCE with  $N = 1$ ; when  $\sigma = 0.1$ , the relative error of SCM with  $L = 2$  is only less than one order of magnitude smaller than WCE with  $N = 1$ . With Crank-Nicolson in time and Fourier collocation in physical space, we cannot achieve better accuracy for WCE with  $N = 1$  and  $\Delta$  no less than 0.0005 when  $M \leq 40$ . These findings suggest the use of efficient PDE solvers for different resulting PDEs from WCE and SCM (to be precise, from Ito-Wick formulation and Stratonovich formulation), respectively.

**Example 4.4.2** (Commutative noises). *We consider two commutative noises in (2.2.1) in Ito's form over the domain  $(0, T] \times (0, 2\pi)$ :*

$$\begin{aligned} du &= [(\epsilon + \frac{1}{2}\sigma_1^2 \cos^2(x))\partial_x^2 u + (\beta \sin(x) - \frac{1}{4}\sigma_1^2 \sin(2x))\partial_x u] dt \\ &\quad + \sigma_1 \cos(x)\partial_x u dw_1(t) + \sigma_2 u dw_2(t), \end{aligned} \tag{4.4.4}$$

*or in Stratonovich form*

$$du = [\epsilon\partial_x^2 u + \beta \sin(x)\partial_x u] dt + \sigma_1 \cos(x)\partial_x u \circ dw_1(t) + \sigma_2 u \circ dw_2(t), \tag{4.4.5}$$

*with non-random initial condition  $u(0, x) = \cos(x)$ , where  $(w_1(t), w_2(t))$  is a standard Wiener process,  $\epsilon > 0$ ,  $\beta$ ,  $\sigma_1$ ,  $\sigma_2$  are constants. The noises satisfy the commutative conditions:*

$$(\sigma_1 \cos(x)\partial_x)(\sigma_2 u) = \sigma_2(\sigma_1 \cos(x)\partial_x u).$$

In this example, we take  $\sigma_1 = 0.5$ ,  $\sigma_2 = 0.2$ ,  $\beta = 0.1$ ,  $\epsilon = 0.02$ . We again observe first-order convergence for SCM and WCE with  $N = 1$ , and second-order convergence for WCE with  $N = 2$  as in the last example with single noise.

We choose the same solver for the recursive multi-stage WCE and SCM but a Crank-Nicolson

scheme in time for SCM. We compute the errors as follows: when  $\Delta_1 > \Delta_2$ ,

$$\bar{\varrho}_2^2(T) = \left| \left| \mathbb{E}[u_{\Delta_1}^2(T)] \right| - \left| \mathbb{E}[u_{\Delta_2}^2(T)] \right| \right|, \quad \bar{\varrho}_2^{r,2}(T) = \frac{\bar{\varrho}_2^2(T)}{\left| \mathbb{E}[u_{\Delta_2}^2(T)] \right|}, \quad (4.4.6)$$

$$\bar{\varrho}_2^\infty(T) = \left| \left| \mathbb{E}[u_{\Delta_1}^2(T)] \right|_\infty - \left| \mathbb{E}[u_{\Delta_2}^2(T)] \right|_\infty \right|, \quad \bar{\varrho}_2^{r,\infty}(T) = \frac{\bar{\varrho}_2^\infty(T)}{\left| \mathbb{E}[u_{\Delta_2}^2(T)] \right|_\infty}. \quad (4.4.7)$$

In Tables 4.3 and 4.4, we choose  $\Delta_2 = \Delta_1/10$ , i.e., the errors for  $\Delta$  is computed by the difference of the numerical solution obtained by  $\Delta_1 = \Delta$  ( $\delta t = \Delta_1/10$ ) and that by  $\Delta_2 = \Delta/10$ , ( $\delta t = \Delta_2/10$ ), both with  $n = 1$  and  $M = 30$ .

For WCE, we observe in Table 4.3 convergence of order  $\Delta^N$  in the second-order moments: first-order convergence when  $N = 1$ , and second-order convergence when  $N = 2$ . Numerical results for  $N = 3$  (not presented here) show that the convergence order is still two even though the accuracy is further improved when  $N$  increases from 2 to 3. This is consistent with our estimate  $\mathcal{O}(\Delta^N) + \mathcal{O}(\Delta^2)$  in Corollary 4.3.2.

We also tested the case  $n = 2$  which gives similar results and same convergence order.

Table 4.3: Algorithm 2.4.1: recursive multi-stage Wiener chaos expansion for commutative noises (4.4.4) at  $T = 1$ :  $\sigma_1 = 0.5$ ,  $\sigma_2 = 0.2$ ,  $\beta = 0.1$ ,  $\epsilon = 0.02$ , and  $M = 30$ ,  $n = 1$ .

$\Delta$	$\delta t$	$N$	$\bar{\varrho}_2^{r,2}(T)$	order	$\bar{\varrho}_2^{r,\infty}(T)$	order	CPU time (sec.)
1.0e-1	1.0e-2	1	1.5213e-3	–	1.4833e-3	–	3.19
1.0e-2	1.0e-3	1	1.6206e-4	$\Delta^{0.97}$	1.5603e-4	$\Delta^{0.98}$	32.74
1.0e-3	1.0e-4	1	1.6323e-5	$\Delta^{1.00}$	1.5694e-5	$\Delta^{1.00}$	329.15
1.0e-1	1.0e-2	2	4.0210e-5	–	2.9237e-5	–	6.53
1.0e-2	1.0e-3	2	4.4359e-7	$\Delta^{1.96}$	3.2771e-7	$\Delta^{1.95}$	65.89
1.0e-3	1.0e-4	2	4.4682e-9	$\Delta^{2.00}$	3.3484e-9	$\Delta^{1.99}$	657.55

For SCM, we observe first-order convergence in  $\Delta$  from Table 4.4 when  $L = 2, 3$ . When  $L = 4$ , the errors are at the same level as  $L = 3$ . Note that  $L = 3$  actually leads to a bit worse accuracy, compared with the case  $L = 2$ . We also tested  $n = 2$  and observed no improved accuracy for  $L = 2, 3, 4$ .

For the two commutative noises, we conclude from this example that the recursive multi-stage

Table 4.4: Algorithm 4.2.1: recursive multi-stage stochastic collocation method for commutative noises (4.4.4) at  $T = 1$ :  $\sigma_1 = 0.5$ ,  $\sigma_2 = 0.2$ ,  $\beta = 0.1$ ,  $\epsilon = 0.02$ , and  $M = 30$ ,  $n = 1$ .

$\Delta$	$\delta t$	L	$\bar{\varrho}_2^{r,2}(T)$	order	$\bar{\varrho}_2^{r,\infty}(T)$	order	CPU time (sec.)
1.0e-1	1.0e-2	2	1.2318e-4	–	1.1253e-3	–	5.18
1.0e-2	1.0e-3	2	1.1880e-5	$\Delta^{1.02}$	1.0921e-4	$\Delta^{1.01}$	54.70
1.0e-3	1.0e-4	2	1.1837e-6	$\Delta^{1.00}$	1.0889e-5	$\Delta^{1.00}$	545.20
1.0e-1	1.0e-2	3	2.3402e-4	–	1.8692e-4	–	13.26
1.0e-2	1.0e-3	3	2.3127e-5	$\Delta^{1.01}$	1.6291e-5	$\Delta^{1.06}$	142.23
1.0e-3	1.0e-4	3	2.3102e-6	$\Delta^{1.00}$	1.6062e-6	$\Delta^{1.00}$	1420.24

WCE is of order  $\Delta^N + \Delta^2$  in the second-order moments and that the recursive multi-stage SCM is of order  $\Delta$  in the second-order moments no matter what sparse grid level is. The errors of recursive multi-stage SCM is one order of magnitude smaller than those of recursive multi-stage WCE with  $N = 1$  while the time cost of SCM is about 1.6 times of that cost of WCE. For large magnitude of noises ( $\sigma_1 = \sigma_2 = 1$ , numerical results are not presented), we observed that the SCM with  $L = 2$  and WCE with  $N = 1$  have the same order-of-magnitude accuracy. In this example, the use of SCM with  $L = 2$  for small magnitude of noises is competitive with the use of WCE with  $N = 1$ .

**Example 4.4.3** (Non-commutative noises). *We consider two non-commutative noises in (2.2.1) in Ito's form over the domain  $(0, T] \times (0, 2\pi)$ :*

$$\begin{aligned}
 du &= [(\epsilon + \frac{1}{2}\sigma_1^2)\partial_x^2 u + \beta \sin(x)\partial_x u + \frac{1}{2}\sigma_2^2 \cos^2(x)u] dt \\
 &\quad + \sigma_1 \partial_x u dw_1(t) + \sigma_2 \cos(x)u dw_2(t),
 \end{aligned} \tag{4.4.8}$$

or in Stratonovich form

$$du = [\epsilon \partial_x^2 u + \beta \sin(x)\partial_x u] dt + \sigma_1 \partial_x u \circ dw_1(t) + \sigma_2 \cos(x)u \circ dw_2(t), \tag{4.4.9}$$

with non-random initial condition  $u(0, x) = \cos(x)$ , where  $(w_1(t), w_2(t))$  is a standard Wiener

process,  $\epsilon > 0$ ,  $\beta$ ,  $\sigma_1$ ,  $\sigma_2$  are constants. The noises are non-commutative since

$$\sigma_2 \cos(x)(\sigma_1 \partial_x u) - (\sigma_1 \partial_x)(\sigma_2 \cos(x)u) = \sigma_1 \sigma_2 \sin(x) \partial_x u \neq 0.$$

We take the same constants  $\epsilon > 0$ ,  $\beta$ ,  $\sigma_1$ ,  $\sigma_2$  as in the last example. We also take the same solver as in the last example. In the current example, we observe only first-order convergence for SCM (level  $L = 2, 3, 4$ ) and WCE ( $N = 1, 2, 3$ ) when  $n = 1, 2$ , see Table 4.5 for parts of the numerical results.

The errors are computed by (4.4.6) and are defined by the difference of numerical solution obtained by parameters in a current line and that obtained by parameters in the subsequent line in the tables. For example, the errors of Line 2 in Table 4.5 are computed by the difference of numerical solution with  $\Delta = 0.05$  and that with  $\Delta = 0.02$ . The errors in the last line are computed by the difference of numerical solutions with  $\Delta = 0.001$  and those with  $\Delta = 0.0005$ , while other truncation parameters are fixed except  $\delta t = \Delta/10$ .

Table 4.5: Algorithm 2.4.1 (recursive multi-stage Wiener chaos expansion, left) and Algorithm 4.2.1 (recursive multi-stage stochastic collocation method, right) for (4.4.8) at  $T = 1$ :  $\sigma_1 = 0.5$ ,  $\sigma_2 = 0.2$ ,  $\beta = 0.1$ ,  $\epsilon = 0.02$ , and  $M = 20$ ,  $n = 1$ . The time step size  $\delta t$  is  $\Delta/10$ . The reported CPU time is in seconds.

$\Delta$	N	$\bar{\varrho}_2^{r,2}(T)$	order	time (sec.)	L	$\bar{\varrho}_2^{r,2}(T)$	order	time (sec.)
1.0e-1	1	1.8648e-3	–	1.04	2	3.2616e-4	–	1.65
5.0e-2	1	1.1431e-3	$\Delta^{0.71}$	2.11	2	1.9301e-4	$\Delta^{0.76}$	3.31
2.0e-2	1	3.8525e-4	$\Delta^{1.19}$	5.12	2	6.3899e-5	$\Delta^{1.21}$	8.64
1.0e-2	1	1.9346e-4	$\Delta^{1.00}$	10.19	2	3.1868e-5	$\Delta^{1.00}$	17.12
5.0e-3	1	1.1635e-4	$\Delta^{0.73}$	20.01	2	<b>1.9095e-5</b>	$\Delta^{0.74}$	<b>33.82</b>
2.0e-3	1	3.8827e-5	$\Delta^{1.20}$	50.39	2	6.3606e-6	$\Delta^{1.20}$	86.44
1.0e-3	1	<b>1.9422e-5</b>	$\Delta^{1.00}$	<b>100.09</b>	2	3.1795e-6	$\Delta^{1.00}$	172.13
1.0e-1	2	5.7118e-5	–	2.16	3	8.1479e-5	–	4.03
5.0e-2	2	<b>2.5022e-5</b>	$\Delta^{1.19}$	<b>4.11</b>	3	4.6873e-5	$\Delta^{0.80}$	8.68
2.0e-2	2	6.7011e-6	$\Delta^{1.44}$	9.97	3	<b>1.5294e-5</b>	$\Delta^{1.22}$	<b>22.08</b>
1.0e-2	2	3.0336e-6	$\Delta^{1.14}$	20.03	3	7.5859e-6	$\Delta^{1.00}$	43.85
5.0e-3	2	1.7175e-6	$\Delta^{0.82}$	40.25	3	4.5320e-6	$\Delta^{0.74}$	88.35
2.0e-3	2	5.5527e-7	$\Delta^{1.23}$	101.34	3	1.5074e-6	$\Delta^{1.20}$	223.15
1.0e-3	2	2.7438e-7	$\Delta^{1.02}$	201.35	3	7.5310e-7	$\Delta^{1.00}$	450.13

In this example, our error estimate for recursive multi-stage WCE is not valid any more and

the numerical results suggest that the errors behave as  $\Delta^N + C\Delta/n$ . For  $N = 1$  and  $n = 10$  (not presented), the error is almost the same as  $n = 1$ . While  $N = 2$  and  $n = 10$ , the error first decreases as  $\mathcal{O}(\Delta^2)$  for large time step size and then as  $\mathcal{O}(\Delta)$  for small time step size; see Table 4.6. When  $N = 2$  and  $n = 10$ , the errors with  $\Delta = 0.005, 0.002, 0.001$  are ten percent ( $1/n$ ) of those with the same parameters but  $n = 1$  in Table 4.5. Here the constant in front of  $\Delta$ ,  $C/n$ , plays an important role: when  $\Delta$  is large and this constant is small, then the order of two can be observed; when  $\Delta$  is small,  $C\Delta/n$  is dominant so that only first-order convergence can be observed.

The recursive multi-stage SCM is of first-order convergence when  $L = 2, 3, 4$  and  $n = 1, 2, 10$  (only parts of the results presented). In contrast to Example 4.4.2, the errors from  $L = 3$  are one order of magnitude smaller those from  $L = 2$ . Recalling that the number of sparse grid points is  $\eta(2, 2) = 5$  and  $\eta(3, 2) = 13$ , we have the cost for  $L = 3$  is about 2.6 times of that for  $L = 2$ . However, it is expected that in practice, a low level sparse grid is more efficient than a high level one when  $nq$  is large as the number of sparse grid points  $\eta(L, nq)$  is increasing exponentially with  $nq$  and  $L$ . In other words,  $L = 2$  is preferred when SPDEs with many noises (large  $q$ ) are considered.

As discussed in the beginning of this section, the ratio of time cost for SCM and WCE is  $\eta(L, nq)/\binom{N+nq}{N}$ . The cost of recursive multi-stage SCM with  $L = 2$  is at most 1.8 times (1.6 predicted by the ratio above,  $q = 2$  and  $n = 1$ ) of that of recursive multi-stage WCE with  $N = 1$ . However, in this example, the accuracy of the recursive multi-stage SCM is one order of magnitude smaller than that of the recursive multi-stage WCE when  $N = 1$  and  $L = 2$ . In Table 4.5, we present in bold the errors between  $1.5 \times 10^{-5}$  and  $2.5 \times 10^{-5}$ . Among the four cases listed in the table, the most efficient, for the given accuracy above, is WCE with  $N = 2$ , which outperforms SCM with  $L = 3$  and SCM with  $L = 2$ . Also, WCE with  $N = 1$  is less efficient than the other three cases. We also observed that when  $\sigma_1 = \sigma_2 = 1$ , SCM with  $L = 2$  is one order of magnitude smaller than WCE with  $N = 1$  (results not presented here).

For non-commutative noises in this example, we show that the error for WCE is  $\Delta^2 + C\Delta/n$  and the error for SCM is  $\Delta$ . The numerical results suggest that SCM with  $L = 2$  is competitive

with WCE with  $N = 1$  for both small and large magnitude of noises if  $n = 1$ .

Table 4.6: Algorithm 2.4.1: recursive multi-stage Wiener chaos expansion for (4.4.8) at  $T = 1$ :  $\sigma_1 = 0.5$ ,  $\sigma_2 = 0.2$ ,  $\beta = 0.1$ ,  $\epsilon = 0.02$ . The parameters are  $M = 20$ ,  $N = 2$ , and  $n = 10$ . The time step size  $\delta t$  is  $\Delta/10$ .

$\Delta$	$\bar{\varrho}_2^{r,2}(T)$	order	$\bar{\varrho}_2^{r,\infty}(T)$	order	CPU time (sec.)
1.0e-1	3.5280e-5	–	1.9366e-5	–	84.00
5.0e-2	1.1122e-5	$\Delta^{1.67}$	5.9400e-6	$\Delta^{1.71}$	160.50
2.0e-2	1.9283e-6	$\Delta^{1.91}$	9.7383e-7	$\Delta^{1.97}$	391.40
1.0e-2	6.2037e-7	$\Delta^{1.64}$	2.9242e-7	$\Delta^{1.74}$	749.40
5.0e-3	2.6087e-7	$\Delta^{1.25}$	1.1245e-7	$\Delta^{1.38}$	1557.60
2.0e-3	6.8288e-8	$\Delta^{1.46}$	2.6917e-8	$\Delta^{1.56}$	3887.50
1.0e-3	3.0622e-8	$\Delta^{1.16}$	1.1452e-8	$\Delta^{1.23}$	7677.40

With these three examples, we observe that the convergence order of the recursive multi-stage SCM in the second-order moments is one for commutative and non-commutative noises. We verified that our error estimate for WCE,  $\Delta^N + \Delta^2$ , is valid for commutative noises, see Examples 4.4.1 and 4.4.2; the numerical results for non-commutative noises, see Example 4.4.3, suggest the errors are of order  $\Delta^N + C\Delta/n$  where  $C$  is a constant depending on the coefficients of the noises.

For stochastic advection-diffusion-reaction equations, different formulations of stochastic products (Ito-Wick product for WCE, Stratonovich product for SCM) lead to different numerical performance. When the white noise is in the velocity, the Ito-Wick formulation will have stronger diffusion than that in the Stratonovich formulation in the resulting PDE. As stronger diffusion requires more resolution, the recursive multi-stage WCE with  $N = 1$  may produce less accurate results than those by the recursive multi-stage SCM with  $L = 2$  with the same PDE solver under the same resolution, as shown in the first and the third examples.

In conclusion, we recommend the recursive multi-stage SCM with  $L = 2$ ,  $n = 1$  and also the recursive multi-stage WCE with  $N = 1$ ,  $n = 1$ , as both can outperform each other in certain cases. For commutative noises, recursive multi-stage WCE with  $N = 2$  may be used when the number of noises,  $q$ , is small and hence the number of WCE modes is small so that the computational cost would grow slowly.

## 4.5 Proofs

Denote by  $C_b^r(\mathcal{D})$  the set of continuous functions that have bounded up to  $r$ -th order derivatives with finite norm

$$\|f\|_{C_b^r} = \max_{0 \leq |\beta| \leq \lfloor r \rfloor} \|D^\beta f\|_{L^\infty} + \sup_{\substack{x, y \in \mathcal{D} \\ |\beta| = \lfloor r \rfloor, r > \lfloor r \rfloor}} \frac{|D^\beta f(x) - D^\beta f(y)|}{|x - y|^{r - \lfloor r \rfloor}},$$

where  $\lfloor r \rfloor$  is the integer part of the real number  $r$ .

### 4.5.1 Proof of Theorem 4.3.1

The idea of the proof is to first establish the error estimate for one-step ( $\Delta = T$ ) error and then that for the multi-step ( $\Delta = T/K$ ) error, the latter of which can be done with the same argument of the proof in [265, Theorem 2.4]. The one-step error consists of two steps: estimating  $\mathbb{E}[(u(t) - u_N(t))^2]$  and  $\mathbb{E}[(u_N(t) - u_{N,n}(t))^2]$  and the conclusion follows from the triangle inequality.

We now state the estimate of these two quantities when  $f = g_k = 0$  and  $q = 1$  in (2.2.1).

**Lemma 4.5.1.** *Assume that  $\sigma_{i,k}, a_{i,j}, b_i, c, \nu_k$  belongs to  $C_b^{r+1}(\mathcal{D})$  and  $u_0 \in H^r(\mathcal{D})$ , where  $r \geq N+1$ .*

*When  $C_1 < \delta_{\mathcal{L}}$ , we have, for any  $t \in (0, T]$ ,*

$$\mathbb{E}[\|u(t) - u_N(t)\|^2] \leq (C_{\lfloor r \rfloor} t)^{N+1} e^{2C_{\mathcal{L}} t} \left[ \frac{e^{C_{\lfloor r \rfloor} t}}{(N+1)!} + \frac{(C_{\lfloor r \rfloor} t)^{\lfloor r \rfloor - N - 1}}{\lfloor r \rfloor!} \frac{1}{1 - \frac{C_1}{\delta_{\mathcal{L}}}} \right] \|u_0\|_{H^{\lfloor r \rfloor}}^2, \quad (4.5.1)$$

*where the constants  $\delta_{\mathcal{L}}$  and  $C_{\mathcal{L}}$  are from (4.3.1) and  $C_{\lfloor r \rfloor} = C(\lfloor r \rfloor, \mathcal{L}) \tilde{C}(\lfloor r \rfloor - 1, \mathcal{M})$ . The constants  $C(\lfloor r \rfloor, \mathcal{L})$  and  $\tilde{C}(\lfloor r \rfloor - 1, \mathcal{M})$  are from (4.3.2) and (4.3.4), respectively.*

**Lemma 4.5.2.** *Under the assumptions of Lemma (4.5.1) and  $r \geq N+2$ , we have*

$$E \|u_{N,n}(t, \cdot) - u_N(t, \cdot)\|^2 \leq 2C^2(N+2, \mathcal{L}) e^{2C_{N+2} t + 2C_{\mathcal{L}} t} \frac{t^3}{n\pi^2} \|u_0\|_{H^{N+2}}^2, \quad (4.5.2)$$

*where  $C_{\mathcal{L}}$  is from (4.3.1) and  $C(N+2, \mathcal{L})$ ,  $C_{N+2}$  are from Lemma 4.5.1.*

For simplicity, we denote by  $s^k$  the ordered set  $(s_1, \dots, s_k)$ . For  $k \geq 1$ , denote  $ds^k := ds_1 \dots ds_k$ ,

and

$$\begin{aligned} F(t; s^k; x) &= \mathcal{T}_{t-s_k} \mathcal{M}_{i_k} \cdots \mathcal{T}_{s_2-s_1} \mathcal{M}_{i_1} \mathcal{T}_{s_1} u_0(x), \\ \int^{(k)} (\cdots) ds^k &= \int_0^t \int_0^{s_k} \cdots \int_0^{s_2} (\cdots) ds_1 \dots ds_k. \\ \int_{(k)} (\cdots) ds^k &= \int_0^t \int_{s_1}^t \cdots \int_{s_{k-1}}^t (\cdots) ds_k \dots ds_2 ds_1. \end{aligned}$$

*Proof of Lemma 4.5.1.* As  $\mathbb{E}[\|u(t) - u_N(t)\|^2] = \sum_{k>N} \sum_{|\alpha|=k} \frac{\|\varphi_\alpha(t)\|^2}{\alpha!}$ , one should check  $\sum_{|\alpha|=k} \frac{\varphi_\alpha^2(t, x)}{\alpha!}$  which equals to  $\int^{(k)} |F(t; s^k; x)|^2 ds^k$ , by Proposition A.1 in [265]. Then by Fubini theorem,

$$\sum_{|\alpha|=k} \frac{\|\varphi_\alpha(t)\|^2}{\alpha!} = \int^{(k)} \|F(t; s^k; \cdot)\|^2 ds^k.$$

Denote  $X_k = \mathcal{T}_{s_k-s_{k-1}} \mathcal{M}_{i_{k-1}} \cdots \mathcal{T}_{s_2-s_1} \mathcal{M}_{i_1} \mathcal{T}_{s_1} u_0$ ,  $Y_k = \mathcal{M}_{i_k} X_k$ ,  $k \geq 1$  and also  $X = \mathcal{T}_{t-s_k} Y_k$ .

Then  $X_k = \mathcal{T}_{s_k-s_{k-1}} Y_{k-1}$  and  $Y_{k-1} = \mathcal{M}_{i_{k-1}} X_{k-1}$ .

By the assumption that  $\sigma_{i,k}, a_{i,j}, b_i, c, \nu_k$  belongs to  $C_b^{r+1}(\mathcal{D})$ , it can be readily checked that (4.3.2) and (4.3.3) hold if (4.3.1) holds, and (4.3.4) holds.

Without loss of generality, assume that  $r$  is a positive integer. If  $r \geq k$ , by the definition of  $F$  and the estimate (4.3.2), we have

$$\begin{aligned} \|F(t; s^k; \cdot)\|^2 &\leq e^{2C_{\mathcal{L}}(t-s_k)} \|Y_k\|^2 = e^{2C_{\mathcal{L}}(t-s_k)} \|\mathcal{M}_{i_k} X_k\|^2 \\ &\leq \tilde{C}(0, \mathcal{M}) e^{2C_{\mathcal{L}}(t-s_k)} \|X_k\|_{H^1}^2 \\ &\leq C_1 e^{2C_{\mathcal{L}}(t-s_{k-1})} \|Y_{k-1}\|_{H^1}^2 \\ &\leq \cdots \\ &\leq C_r^k e^{2C_{\mathcal{L}}t} \|u_0\|_{H^k}^2, \end{aligned}$$

where  $C_r = C(r, \mathcal{L}) \tilde{C}(r-1, \mathcal{M})$  and  $\tilde{C}(r-1, \mathcal{M})$  is from (4.3.4), depending only on the coefficients

of  $\mathcal{M}$  with their derivatives up to order  $r - 1$ . We then have

$$\int^{(k)} \|F(t; s^k; \cdot)\|^2 ds^k \leq C_r^k e^{2C_{\mathcal{L}}t} \|u_0\|_{H^k}^2 \int^{(k)} ds^k. \quad (4.5.3)$$

If  $r < k$ , we change the integration order, apply (4.3.2), (4.3.4), and (4.3.3), and have

$$\begin{aligned} & \int^{(k)} \|F(t; s^k; \cdot)\|^2 ds^k = \int^{(k)} \|X\|^2 ds^k = \int_{(k)} \|X\|^2 ds^k \\ & \leq \int_{(k)} e^{2C_{\mathcal{L}}(t-s_k)} \|Y_k\|^2 ds^k = \int_{(k)} e^{2C_{\mathcal{L}}(t-s_k)} \|\mathcal{M}_{i_k} X_k\|^2 ds^k \\ & \leq \tilde{C}(0, \mathcal{M}) \int_{(k)} e^{2C_{\mathcal{L}}(t-s_k)} \|X_k\|_{H^1}^2 ds^k \\ & = \tilde{C}(0, \mathcal{M}) \int_{(k-1)} \int_{s_{k-1}}^t e^{2C_{\mathcal{L}}(t-s_k)} \|X_k\|_{H^1}^2 ds_k ds^{k-1} \\ & \leq \delta_{\mathcal{L}}^{-1} C_1 \int_{(k-1)} e^{2C_{\mathcal{L}}(t-s_{k-1})} \|Y_{k-1}\|^2 ds^{k-1}. \end{aligned}$$

Repeating the procedure gives

$$\int^{(k)} \|F(t; s^k; \cdot)\|^2 ds^k \leq \delta_{\mathcal{L}}^{r-k} C_1^{k-r} \int_{(r)} e^{2C_{\mathcal{L}}(t-s_r)} \|Y_r\|^2 ds^r. \quad (4.5.4)$$

By (4.5.3) and (4.5.4), and  $\int^{(k)} ds^k = \frac{t^k}{k!}$ , we conclude that, for  $r \geq \mathbf{N} + 1$ ,

$$\begin{aligned} \mathbb{E}[\|u(t) - u_{\mathbf{N}}(t)\|^2] &= \sum_{k > \mathbf{N}} \int^{(k)} \|F(t; s^k; \cdot)\|^2 ds^k \\ &= \sum_{\mathbf{N} < k \leq r} \int^{(k)} \|F(t; s^k; \cdot)\|^2 ds^k + \sum_{k > r} \int^{(k)} \|F(t; s^k; \cdot)\|^2 ds^k \\ &\leq \sum_{\mathbf{N} < k \leq r} \frac{t^k}{k!} C_r^k e^{2C_{\mathcal{L}}t} \|u_0\|_{H^k}^2 + \frac{t^r}{r!} C_r^r e^{2C_{\mathcal{L}}t} \|u_0\|_{H^r}^2 \sum_{k > r} \delta_{\mathcal{L}}^{r-k} C_1^{k-r} \\ &\leq (C_r t)^{\mathbf{N}+1} e^{2C_{\mathcal{L}}t} \left[ \frac{e^{C_r t}}{(\mathbf{N} + 1)!} + \frac{(C_r t)^{r-\mathbf{N}-1}}{r!} \frac{C_1}{\delta_{\mathcal{L}} - C_1} \right] \|u_0\|_{H^r}^2, \end{aligned}$$

where we require  $C_1 < \delta_{\mathcal{L}}$ . This ends the proof.

**Remark 4.5.3.** Lemma 4.5.1 still holds for  $r = \infty$  if  $C_\infty < \infty$ . In fact, by (4.5.3), we have

$$\mathbb{E}[\|u(t) - u_N(t)\|^2] \leq \sum_{k \geq N} \frac{t^k}{k!} C_\infty^k e^{2C_\infty t} \|u_0\|_{H^k}^2 \leq (C_\infty t)^{N+1} e^{2C_\infty t} \frac{e^{C_\infty t}}{(N+1)!} \|u_0\|_{H^\infty}^2.$$

If  $r < \infty$ , we require that  $C_1 < \delta_\mathcal{L}$ , which is actually  $\tilde{C}(0, \mathcal{M})C(1, \mathcal{L}) < \delta_\mathcal{L}$ . An example of this condition is as follows:  $\mathcal{L} = \Delta$ ,  $\mathcal{M}_k = \frac{1}{2}D_k$  and thus  $\tilde{C}(0, \mathcal{M})C(1, \mathcal{L}) = \frac{1}{2} < \delta_\mathcal{L} = 1$ .

*Proof of Lemma 4.5.2.* It can be proved as in [265, p.446-449] that

$$\sum_{|\alpha|=k, i_k^\alpha=l} \frac{\|\varphi_\alpha(t)\|^2}{\alpha!} \leq kt \|F_j(t; s^k; \cdot)\|^2 \int_0^t M_l^2(s) ds \frac{t^{k-1}}{(k-1)!}, \quad (4.5.5)$$

where  $M_l(t) = \int_0^t m_l(s) ds$  and

$$\begin{aligned} F_j(t; s^k; x) &= \frac{\partial F(t; s^k; x)}{\partial s_j} = \mathcal{T}_{t-s_k} \mathcal{M}_{i_k} \cdots \mathcal{T}_{s_{j+1}-s_j} \mathcal{M}_{i_j} \mathcal{L} \mathcal{T}_{s_j-s_{j-1}} \cdots \mathcal{T}_{s_1} u_0(x) \\ &\quad - \mathcal{T}_{t-s_k} \mathcal{M}_{i_k} \cdots \mathcal{M}_{i_{j+1}} \mathcal{L} \mathcal{T}_{s_{j+1}-s_j} \cdots \mathcal{T}_{s_1} u_0(x) =: F_j^1 + F_j^2. \end{aligned}$$

Due to the same structure of two terms in  $F_j(t; s^k; x)$ , we only need to estimate  $\|F_j^1\|$ . Repeatedly using (4.3.2) gives

$$\begin{aligned} \|F_j^1\|^2 &= \|\mathcal{T}_{t-s_k} \mathcal{M}_{i_k} \cdots \mathcal{T}_{s_{j+1}-s_j} \mathcal{M}_{i_j} \mathcal{L} \mathcal{T}_{s_j-s_{j-1}} \cdots \mathcal{T}_{s_1} u_0\|^2 \\ &\leq e^{2C_\mathcal{L} t - s_k} \|\mathcal{M}_{i_k} \cdots \mathcal{T}_{s_{j+1}-s_j} \mathcal{M}_{i_j} \mathcal{L} \mathcal{T}_{s_j-s_{j-1}} \cdots \mathcal{T}_{s_1} u_0\|^2 \\ &\leq \tilde{C}(0, \mathcal{M}) e^{2C_\mathcal{L}(t-s_k)} \|\mathcal{T}_{s_k-s_{k-1}} \cdots \mathcal{T}_{s_{j+1}-s_j} \mathcal{M}_{i_j} \mathcal{L} \mathcal{T}_{s_j-s_{j-1}} \cdots \mathcal{T}_{s_1} u_0\|_{H^1}^2 \\ &\leq C_1 e^{2C_\mathcal{L}(t-s_k)} \|\mathcal{M}_{i_{k-1}} \cdots \mathcal{T}_{s_{j+1}-s_j} \mathcal{M}_{i_j} \mathcal{L} \mathcal{T}_{s_j-s_{j-1}} \cdots \mathcal{T}_{s_1} u_0\|_{H^1}^2 \\ &\leq \cdots \\ &\leq C_{k-j}^{k-j-1} e^{2C_\mathcal{L}(t-s_j)} \|\mathcal{T}_{s_j-s_{j-1}} \cdots \mathcal{T}_{s_1} u_0\|_{H^{k-j+2}}^2. \end{aligned}$$

Similarly, we have

$$\|\mathcal{T}_{s_j - s_{j-1}} \cdots \mathcal{T}_{s_1} u_0\|_{H^{k-j+2}}^2 \leq C_{k+2}^j e^{2C_{\mathcal{L}} s_j} \|u_0\|_{H^{k+2}}^2.$$

Thus, we arrive at that

$$\|F_j^1\|^2 \leq C_{k-j}^{k-j-1} \|\mathcal{T}_{s_j - s_{j-1}} \cdots \mathcal{T}_{s_1} u_0\|_{H^{k-j+2}}^2 \leq C_{\mathbf{N}+2}^{k+1} e^{2C_{\mathcal{L}} t} \|u_0\|_{H^{\mathbf{N}+2}}^2 \quad (4.5.6)$$

Then by (4.5.5) and (4.5.6), setting  $M_l(s) = \frac{\sqrt{2t}}{(l-1)\pi} \sin(\frac{(l-1)\pi}{t}s)$ , we obtain that

$$\begin{aligned} \mathbb{E}[\|u_{\mathbf{N}}(t) - u_{\mathbf{N},n}(t)\|^2] &= \sum_{l \geq \mathbf{n}+1} \sum_{k=1}^{\mathbf{N}} \sum_{|\alpha|=k, i_k^\alpha=l} \frac{\|\varphi_\alpha(t)\|^2}{\alpha!} \\ &\leq \sum_{l \geq \mathbf{n}+1} \sum_{k=1}^{\mathbf{N}} kt \max_j \|F_j(t; s^k; \cdot)\|^2 \int_0^t M_l^2(s) ds \frac{t^{k-1}}{(k-1)!} \\ &\leq \sum_{l \geq \mathbf{n}+1} \sum_{k=1}^{\mathbf{N}} 2kt \max_j \|F_j^1\|^2 \frac{t^2}{\pi^2(l-1)^2} \frac{t^{k-1}}{(k-1)!} \\ &\leq \sum_{l \geq \mathbf{n}+1} \frac{2t^3}{\pi^2(l-1)^2} \sum_{k=1}^{\mathbf{N}} k \max_j \|F_j^1\|^2 \frac{t^{k-1}}{(k-1)!} \\ &\leq \frac{2t^3}{n\pi^2} \sum_{k=1}^{\mathbf{N}} k \max_j \|F_j^1\|^2 \frac{t^{k-1}}{(k-1)!}. \end{aligned}$$

The summation in the last inequality can be estimated by (4.5.6) as follows:

$$\begin{aligned} \sum_{k=1}^{\mathbf{N}} k \max_j \|F_j^1\|^2 \frac{t^{k-1}}{(k-1)!} &\leq \sum_{k=1}^{\mathbf{N}} C_{\mathbf{N}+2}^{k+1} e^{2C_{\mathcal{L}} t} \|u_0\|_{H^{\mathbf{N}+2}}^2 \frac{kt^{k-1}}{(k-1)!} \\ &\leq C^2(\mathbf{N}+2, \mathcal{L}) e^{2C_{\mathbf{N}+2} t + 2C_{\mathcal{L}} t} \|u_0\|_{H^{\mathbf{N}+2}}^2. \end{aligned}$$

This completes the proof.

### 4.5.2 Proof of Theorem 4.3.3

To prove Theorem 4.3.3, we need the following conditional probabilistic representation (the conditional Feynman-Kac formula) for (2.2.1), see e.g. [347],

$$u(t, x) = \mathbb{E}_Q[u_0(\hat{X}_{t,x}(T))\hat{Y}_{t,x,1}(T) + \hat{Z}_{t,x,1,0}(T)], \quad 0 \leq t \leq T. \quad (4.5.7)$$

where  $\hat{X}_{t,x}$ ,  $\hat{Y}_{t,x,y}$ , and  $\hat{Z}_{t,x,y,z}$  is the solution to the following SDE

$$\begin{aligned} d\hat{X} &= [b(s, \hat{X}) - \sum_{r=1}^q \sigma_r(s, \hat{X})\nu_r(s, \hat{X})] ds \\ &+ \sum_{r=1}^p \alpha_r(s, \hat{X}) d\hat{B}_r + \sum_{r=1}^p \sigma_r(s, \hat{X}) d\hat{w}_r, \quad \hat{X}(t) = x, \end{aligned} \quad (4.5.8)$$

$$d\hat{Y} = [c(s, \hat{X})\hat{Y} ds + \sum_{r=1}^q \nu_r(s, \hat{X})\hat{Y}] d\hat{w}_r, \quad \hat{Y}(t) = y, \quad (4.5.9)$$

$$d\hat{Z} = [f(s, \hat{X})\hat{Y} ds + \sum_{r=1}^q g_r(s, \hat{X})\hat{Y}] d\hat{w}_r, \quad \hat{Z}(t) = z, \quad (4.5.10)$$

and  $B(s) = (B_1(s), \dots, B_p(s))^\top$  is a  $p$ -dimensional standard Wiener process independent of  $w(s)$  and  $\mathbb{E}_Q[\cdot]$  is the expectation with respect to  $\mathcal{F}_t^B$ . The  $d \times p$  matrix  $\alpha(t, x)$  is defined as  $\alpha(t, x)\alpha^\top(t, x) = 2a(t, x) - \sigma(t, x)\sigma^\top(t, x)$ .

We consider the case  $f = g_k = \sigma_{ik} = 0$ , i.e., the Zakai equation of nonlinear filtering when observation is uncorrelated as in [265].

As the case of  $n = 1$  is considered in [50, 194], here we consider only the case  $K = 1$ . The proof is for time-dependent coefficients which covers the case of time-independent ones in (2.2.1).

*Proof.* The solution to (2.2.1) can be represented as

$$u(t, x) = \mathbb{E}_Q[u_0(\hat{X}_{t,x}) \exp(\sum_{r=1}^q \int_0^t \nu_r(s, \hat{X}_{t,x}) d\hat{w}_r(s) + \int_0^t \bar{c}(t, \hat{X}_{t,x}) ds)], \quad (4.5.11)$$

where  $\bar{c}(t, \hat{X}_{t,x}) = c(t, \hat{X}_{t,x}) - \frac{1}{2} \sum_{r=1}^q \nu_r^2(s, \hat{X}_{t,x})$ . The solution to (4.2.7) using spectral expansion of Brownian motion can be represented by

$$\tilde{u}(t, x) = \mathbb{E}_Q[u_0(\hat{X}_{t,x}) \exp(\sum_{r=1}^q \int_0^t \nu_r(s, \hat{X}_{t,x}) \dot{w}_r(s) ds + \int_0^t \bar{c}(t, \hat{X}_{t,x}) ds)]. \quad (4.5.12)$$

Thus, we have

$$\begin{aligned} \tilde{u}(t, x) - u(t, x) &= \mathbb{E}_Q[u_0(\hat{X}_{t,x}) \exp(\int_0^t \bar{c}(t, \hat{X}_{t,x}) ds) \\ &\quad \times \left( \exp(\sum_{r=1}^q \int_0^t \nu_r(s, \hat{X}_{t,x}) \dot{w}_r(s) ds) - \exp(\sum_{r=1}^q \int_0^t \nu_r(s, \hat{X}_{t,x}) dw_r(s)) \right)], \end{aligned}$$

By the fact that  $e^x - e^y = e^{\theta x + (1-\theta)y}(x - y)$  ( $0 \leq \theta \leq 1$ ), boundedness of  $\bar{c}(t, x)$  and Cauchy inequality, we have

$$\begin{aligned} &\mathbb{E}[|\tilde{u}(t, x) - u(t, x)|^2] \quad (4.5.13) \\ &\leq \mathbb{E}[(\mathbb{E}_Q[u_0(\hat{X}_{t,x}) \exp(\int_0^t \bar{c}(t, \hat{X}_{t,x}) ds)]) \exp(\sum_{r=1}^q \int_0^t \nu_r(s, \hat{X}_{t,x}) [\theta d\tilde{w}_r(s) + (1-\theta) dw_r(s)]) \\ &\quad \times (\sum_{r=1}^q \int_0^t \nu_r(s, \hat{X}_{t,x}) [d\tilde{w}_r(s) - dw_r(s)])]^2] \\ &\leq C \exp(Ct) \mathbb{E}_Q[u_0^4(\hat{X}_{t,x})]^{1/2} \mathbb{E}[(\mathbb{E}_Q[\exp(\sum_{r=1}^q \int_0^t 4\nu_r(s, \hat{X}_{t,x}) [\theta d\tilde{w}_r(s) + (1-\theta) dw_r(s)])]^{1/2}] \\ &\quad \times \mathbb{E}[\mathbb{E}_Q[(\sum_{r=1}^q \int_0^t \nu_r(s, \hat{X}_{t,x}) [d\tilde{w}_r(s) - dw_r(s)]^2)]]] \\ &\leq C \exp(Ct) \mathbb{E}_Q[(\mathbb{E}[\exp(\sum_{r=1}^q \int_0^t 4\nu_r(s, \hat{X}_{t,x}) [\theta d\tilde{w}_r(s) + (1-\theta) dw_r(s)])]^{1/2}] \\ &\quad \times \mathbb{E}_Q[\mathbb{E}[(\sum_{r=1}^q \int_0^t \nu_r(s, \hat{X}_{t,x}) [d\tilde{w}_r(s) - dw_r(s)]^2)]]]. \end{aligned}$$

Hence, we need to estimate  $I =: \mathbb{E}[\left| \sum_{r=1}^q \int_0^t \nu_r(s, \hat{X}_{t,x}) [d\tilde{w}_k(s) - dw_k(s)] \right|^2]$  and

$$II =: \mathbb{E}[(\mathbb{E}_Q[\exp(\sum_{r=1}^q \int_0^t 4\nu_r(s, \hat{X}_{t,x}) [\theta d\tilde{w}_r(s) + (1-\theta) dw_r(s)])]^{1/2})].$$

By the independence of  $\xi_{k,i}$  and  $\hat{X}_{t,x}$  and a standard estimate in spectral methods (e.g. [33]), we have

$$\begin{aligned}
I &= \mathbb{E}\left[\left|\sum_{r=1}^q \sum_{i=n+1}^{\infty} \int_0^t \nu_r(s, \hat{X}_{t,x}) m_{k,i}(s) ds \xi_{k,i}\right|^2\right] \\
&= \mathbb{E}\left[\sum_{r=1}^q \sum_{i=n+1}^{\infty} \left(\int_0^t \nu_r(s, \hat{X}_{t,x}) m_{k,i}(s) ds\right)^2\right] \\
&\leq CT^{1-\varepsilon} n^{-1+\varepsilon} \sum_{r=1}^q \|\nu_r\|_{\frac{1-\varepsilon}{2}, 2, [0, T]}^2,
\end{aligned} \tag{4.5.14}$$

where the Slobodeckij semi-norm  $|f|_{\theta, p, \mathcal{D}}$  is defined by  $(\int_{\mathcal{D}} \int_{\mathcal{D}} \frac{|f(x)-f(y)|^p}{|x-y|^{p\theta+d}} dx dy)^{1/2}$ . The constant  $T^{1-\varepsilon}$  appears due to Bramble-Hilbert lemma, see e.g. [70]. Since  $\nu_r$  is Lipschitz continuous in its first and second variables and  $\mathbb{E}_Q[|\hat{X}_{t,x}(s) - \hat{X}_{t,x}(u)|^2] \leq C|s-u|$ , we have

$$\begin{aligned}
\mathbb{E}_Q[\|\nu_r\|_{\frac{1-\varepsilon}{2}, 2, [0, T]}^2] &= \int_0^T \int_0^T \frac{\mathbb{E}_Q\left[\left|\nu_r(s, \hat{X}_{T,x}(s)) - \nu_r(s_1, \hat{X}_{T,x}(s_1))\right|^2\right]}{|s-s_1|^{2-\varepsilon}} ds_1 ds \\
&\leq C(T^{2+\varepsilon} + T^{1+\varepsilon}).
\end{aligned} \tag{4.5.15}$$

Thus, by (4.5.14) and (4.5.15), we have

$$I \leq C(T^3 + T^2)n^{-1+\varepsilon}. \tag{4.5.16}$$

Now we estimate  $II$ . Recall the following facts

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\sum_{r=1}^q \int_0^t 4\nu_r(s, \hat{X}_{t,x}) dw_r\right)\right] &= \exp\left(\sum_{r=1}^q 8 \int_0^t \nu_r^2(s, \hat{X}_{t,x}) ds\right), \\
\mathbb{E}\left[\exp\left(\sum_{r=1}^q \int_0^t 4\nu_r(s, \hat{X}_{t,x}) \dot{w}_r ds\right)\right] &\leq \exp\left(\sum_{r=1}^q 8 \int_0^t \nu_r^2(s, \hat{X}_{t,x}) ds\right),
\end{aligned}$$

and thus we have

$$II \leq 4 \exp(Ct). \tag{4.5.17}$$

From here, (4.5.16), and (4.5.13), we reach the conclusion for one-step error and hence we can readily obtain the multi-step error estimate.  $\square$

## 4.6 Appendix: derivation of Algorithm 4.2.1.

The idea of recursive SCM is to use SCM over small time interval  $(\mathbf{t}_{i-1}, \mathbf{t}_i]$  instead of over the whole interval  $(0, T]$  and to compute the second-order moments of the solution recursively in time. The derivation of such a recursive algorithm will make use of properties of the problem (2.2.1) and orthogonality of the basis both in physical space and random space as will be shown shortly.

Denote the solution of sparse grid collocation by

$$\tilde{u}_{\Delta, \mathbf{L}, \mathbf{n}}^{\mathbf{M}}(\mathbf{t}_i, x) = \sum_{\kappa=1}^{\eta(\mathbf{L}, nq)} v_{\kappa}(\Delta, x, \tilde{u}_{\Delta, \mathbf{L}, \mathbf{n}}^{\mathbf{M}}(\mathbf{t}_{i-1}, x)) \mathcal{I}_{\kappa}(\zeta^{(i)}),$$

where  $\zeta^{(i)}$  is a  $nq$ -dimensional Gaussian random variable used over the interval  $(\mathbf{t}_{i-1}, \mathbf{t}_i]$  and  $\mathcal{I}_{\kappa}(\zeta^{(i)})$  is the sparse grid interpolation operator which is a polynomial of order no more than  $\mathbf{L} - 1$ . It equals to 1 on the  $\kappa$ -th point of the sparse grid  $\mathcal{H}_{\mathbf{L}}^{nq}$  and is zero on the rest of sparse grid points.

We solve (4.2.7) with spectral methods in physical space, i.e., using a truncation of a CONS in physical space, e.g.,  $\{e_m\}_{m=1}^{\mathbf{M}}$  to represent the numerical solution. Define  $\Phi_m(\mathbf{t}_i, \mathbf{L}, \mathbf{n}, \mathbf{M}) = (\tilde{u}_{\Delta, \mathbf{L}, \mathbf{n}}^{\mathbf{M}}(\mathbf{t}_i, \cdot), e_m)$ . Then by the linearity of the problem (4.2.7) and  $f = g_k = 0$ , we have

$$\tilde{u}_{\Delta, \mathbf{L}, \mathbf{n}}^{\mathbf{M}}(\mathbf{t}_i, x) = \sum_{\kappa=1}^{\eta(\mathbf{L}, nq)} \sum_{l=1}^{\mathbf{M}} \Phi_l(\mathbf{t}_{i-1}, \mathbf{L}, \mathbf{n}, \mathbf{M}) v_{\kappa}(\Delta, x, e_l) \mathcal{I}_{\kappa}(\zeta^{(i)}).$$

By the orthonormality of  $e_m$ , we then have

$$\Phi_m(\mathbf{t}_i, \mathbf{L}, \mathbf{n}, \mathbf{M}) = \sum_{\kappa=1}^{\eta(\mathbf{L}, nq)} \sum_{l=1}^{\mathbf{M}} \Phi_l(\mathbf{t}_{i-1}, \mathbf{L}, \mathbf{n}, \mathbf{M}) h_{\kappa, l, m} \mathcal{I}_{\kappa}(\zeta^{(i)}), \quad (4.6.1)$$

where  $h_{\kappa,l,m} = (v_{\kappa}(\Delta; \cdot, e_l), e_m)$ . The second moments can be computed by

$$\mathbb{E}[\tilde{u}_{\Delta,L,n}^M(\mathbf{t}_i, x)]^2 = \sum_{l,m=1}^M H_{lm}(\mathbf{t}_i, L, n, M) e_l(x) e_m(x), \quad (4.6.2)$$

where  $H_{lm}(\mathbf{t}_i, L, n, M) = \mathbb{E}[\Phi_l(\mathbf{t}_i, L, n, M) \Phi_m(\mathbf{t}_i, L, n, M)]$ . With (4.6.1), we have

$$H_{lm}(\mathbf{t}_i, L, n, M) = \sum_{\kappa=1}^{\eta(L,nq)} \sum_{\alpha=1}^{\eta(L,nq)} \sum_{j=1}^M \sum_{k=1}^M H_{jk}(\mathbf{t}_{i-1}, L, n, M) h_{\kappa,j,l} h_{\kappa,k,m} \mathbb{E}[\mathcal{I}_{\kappa}(\zeta^{(i)}) \mathcal{I}_{\alpha}(\zeta^{(i)})]. \quad (4.6.3)$$

The matrix  $H_{lm}(\mathbf{t}_i, L, n, M)$  can be simplified as

$$H_{lm}(\mathbf{t}_i, L, n, M) = \sum_{j=1}^M \sum_{k=1}^M H_{jk}(\mathbf{t}_{i-1}, L, n, M) \sum_{\kappa=1}^{\eta(L,nq)} h_{\kappa,j,l} h_{\kappa,k,m} W_{\kappa},$$

as  $\mathbb{E}[\mathcal{I}_{\kappa}(\zeta^{(i)}) \mathcal{I}_{\alpha}(\zeta^{(i)})] = W_{\kappa} \delta_{\kappa,\alpha}$  since the polynomial exactness of Smolyak sparse grid is  $2L - 1$  [324, Corollary 1].

## Chapter 5

# Application example: stochastic collocation methods for 1D stochastic piston problem

In this chapter, we apply stochastic collocation method for a one-dimensional piston problem. We consider a piston with a velocity perturbed by Brownian motion moving into a straight tube filled with a perfect gas at rest. The shock generated ahead of the piston can be located by solving the one-dimensional Euler equations driven by white noise using the Stratonovich or Ito formulations. We approximate the Brownian motion with its spectral truncation and subsequently apply stochastic collocation using either sparse grid or the quasi-Monte Carlo (QMC) method. In particular, we first transform the Euler equations with an unsteady stochastic boundary into stochastic Euler equations over a fixed domain with a time-dependent stochastic source term. We then solve the transformed equations by splitting them up into two parts, i.e., a ‘deterministic part’ and a ‘stochastic part’. Numerical results verify the Stratonovich-Euler and Ito-Euler models against stochastic perturbation results, and demonstrate the efficiency of sparse grid and QMC for

small and large random piston motions, respectively. The variance of shock location of the piston grows cubically in the case of white noise in contrast to colored noise reported in [251], where the variance of shock location grows quadratically with time for short times and linearly for longer times.

## 5.1 Introduction

In recent years, the polynomial chaos method and its extensions for colored noise have been advanced significantly for computational fluid dynamics problems, see e.g. [221, 318]. Although there has been little attention paid on high-order numerical methods for white noise, white noise is nevertheless important in computational modeling, e.g. as a limit of colored noise when the correlation length goes to zero. An extremely small correlation length for colored noise will produce a high dimensional problem in random space and causes the so-called “curse-of-dimensionality” for high-order numerical methods which are prohibitively expensive. Unlike colored noise, white noise requires a fundamentally different calculus (see e.g. [330]) and therefore the development of new numerical methods.

Here, we revisit the stochastic piston problem in [251], which defines a testbed for numerical solvers in both random and physical space. The piston driven by time-varying random motions moves into a straight tube filled with a perfect gas at rest. Of interest is to quantify the perturbation of the shock position ahead of the piston corresponding to the random motion. For the perturbed shock position, Lin et al. [251] obtained analytical solutions for small amplitudes of noises and numerical solutions for large amplitudes of noises, with the method of stochastic perturbation analysis and polynomial chaos, respectively. A specific random motion of the piston was studied where the piston velocity was perturbed by a *correlated* random process with zero mean and exponential covariance kernel. It was concluded that the variance of the shock location grows quadratically with time for small time and linearly for large time by both the perturbation analysis

and numerical simulations of the corresponding Euler equations. Numerical results from the Monte Carlo method and the polynomial chaos method (e.g. [405]) for the stochastic Euler equations showed good agreement with the results from the perturbation analysis.

Here we consider the case of piston velocity perturbed by *Brownian motion*, which leads to the Euler equations subject to white noise rather than the Euler equations subject to colored noise in [251]. We will use the Monte Carlo method and the recently developed stochastic collocation method for equations driven by white noise. Note that the method of perturbation analysis in [251] is independent of the type of noises when they have *continuous paths* in the random space so that the results by the perturbation analysis can be understood in a path-wise sense. Therefore, the stochastic piston problem defined in [251] can serve as a rigorous testbed of evaluating numerical stochastic solvers. So we will use the variances from perturbation analysis as reference solutions.

Although, the Monte Carlo method is one of the popular methods for solving equations driven by white noise [218], it converges slowly as the total error of the method is dominated by the statistical error, which is proportional to  $\frac{1}{\sqrt{N}}$  with  $N$  being the number of sampling points. To avoid this slow convergence induced by the statistical error, Zhang et al. [421] proposed a new stochastic collocation method for time-dependent equations driven by white noise in time. Stochastic collocation methods are based on high-dimensional integration quadrature rules instead of statistical methods [9, 368, 403]. While the main difficulty of the stochastic collocation method comes from the large number of random variables, we proposed a spectral expansion of the Brownian motion to reduce the number of random variables up to relatively large time for time-dependent equations so that the stochastic collocation method can be applied efficiently. Here we further extend this approach to conservation laws by adopting the quasi-Monte Carlo (QMC) method to compute up to larger time and/or for large amplitudes of noises. The QMC method is efficient and converges faster than the Monte Carlo method if relatively high dimensional integration is considered, see e.g. [320, 357]; see also [139] for the application of the QMC method to elliptic equations in random porous media.

This chapter is organized as follows. In Section 2, we describe the piston problem driven by random processes and review two different approaches to obtain the shock location: the perturbation analysis and the one-dimensional Euler equations. When the piston is driven by the Brownian motion, we introduce two types of Euler equations according to different interpretations of stochastic products for white noise, i.e., the Stratonovich-Euler equations and the Ito-Euler equations. In Section 3, we describe a splitting method for the Euler equations before comparing the variances from the two stochastic Euler equations with those from first-order perturbation analysis. We demonstrate that indeed the Stratonovich-Euler equations are suitable for obtaining the variances of perturbations piston locations. We apply a stochastic collocation method to solve the Stratonovich-Euler equations in the splitting-method setting. We conclude in Section 5 with a summary and comments on computational efficiency. The Appendix includes some details of the stochastic collocation method and of a model ordinary differential equation problem.

## 5.2 Theoretical Background

Suppose that the piston velocity is perturbed by a time-dependent random process so that the piston velocity is  $u_p = U_p + v_p(t, \omega)$ , where  $\omega$  is a point in random space; see Figure 5.1 for a sketch of shock tube driven by a piston perturbed with random motion. Here we write  $v_p(t, \omega) = \epsilon U_p V(t, \omega)$

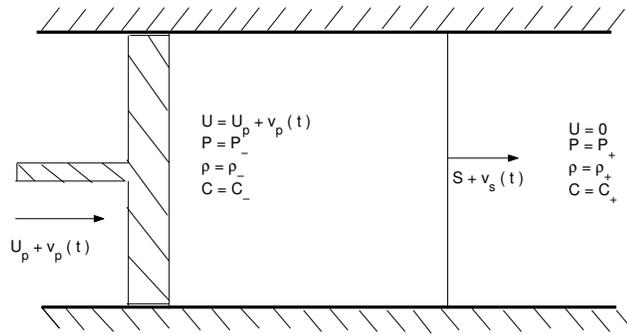


Figure 5.1: A sketch of piston-driven shock tube with random piston motion.

and denote the stochastic process  $V(t, \omega)$  as  $V(t)$  for brevity.

When  $\epsilon = 0$ , i.e., no perturbation is imposed on the piston, the piston moves into the tube with a constant velocity  $U_p$ , the shock speed  $S$  (and thus the shock location) can be determined analytically, see [251, 250]. When  $\epsilon$  is very small, one can determine the perturbation process of the shock location using the first-order perturbation analysis [251], that is:

$$z(t) = \epsilon U_p q S' \sum_{n=0}^{\infty} (-r)^n \int_0^t V(\alpha \beta^n t_1) dt_1, \quad (5.2.1)$$

where  $z(t) + tS$  is the shock location induced by the random motion of piston,

$$\begin{aligned} S' &= \frac{\gamma + 1}{4} \frac{S}{S - \frac{\gamma+1}{4} U_p}, \\ q &= \frac{2}{1+k}, \quad r = \frac{1-k}{1+k}, \quad k = C_- \frac{S + S' U_p}{1 + \gamma S U_p}, \\ \alpha &= \frac{C_- + U_p - S}{C_-}, \quad \beta = \frac{C_- + U_p - S}{C_- + S - U_p}. \end{aligned}$$

Here  $\gamma$  is the ratio of the specific heats and  $C_-$  the sound speed behind the shock when the piston is unperturbed. The first two moments of the perturbation process  $z(t)$  are

$$\begin{aligned} \mathbb{E}[z(t)] &= 0, \\ \mathbb{E}[z^2(t)] &= (\epsilon U_p q S')^2 \mathbb{E}\left[\left(\sum_{n=0}^{\infty} (-r)^n \int_0^t V(\alpha \beta^n t_1, \cdot) dt_1\right)^2\right]. \end{aligned}$$

We note that the perturbation analysis in [251] is independent of the perturbation process whenever the process is continuous such that the analysis can be understood in a path-wise way. By taking  $V(t, \omega)$  as the Brownian motion  $W(t)$  (omitting  $\omega$ ), we then have

$$\begin{aligned} \mathbb{E}[z^2(t)] &= (\epsilon U_p q S')^2 \mathbb{E}\left[\left(\sum_{n=0}^{\infty} (-r)^n \int_0^t W(\alpha \beta^n t_1) dt_1\right)^2\right] \\ &= (\epsilon U_p q S')^2 \mathbb{E}\left[\left(\sum_{n=0}^{\infty} (-r)^n \int_0^t \sqrt{\alpha \beta^n} W(t_1) dt_1\right)^2\right] \end{aligned}$$

$$\begin{aligned}
&= (\epsilon U_p q S')^2 \left( \sum_{n=0}^{\infty} (-r)^n \sqrt{\alpha \beta^n} \right)^2 \mathbb{E} \left[ \left( \int_0^t W(t_1) dt_1 \right)^2 \right] \\
&= \frac{\alpha t^3}{3} (\epsilon U_p q S')^2 \frac{1}{(1 + r \beta^{\frac{1}{2}})^2},
\end{aligned} \tag{5.2.2}$$

where we use the scaling property of Brownian motion ( $W(\alpha \beta^n t_1) = \sqrt{\alpha \beta^n} W(t_1)$ ) and  $\int_0^t W(t_1) dt_1$  is a Gaussian process with zero mean and variance  $\frac{t^3}{3}$ .

### 5.2.1 Stochastic Euler equations

The stochastic piston problem can be modeled by the following Euler equations with unsteady stochastic boundary:

$$\frac{\partial}{\partial t} \mathbf{U} + \frac{\partial}{\partial x} (f(\mathbf{U})) = 0, \tag{5.2.3}$$

where  $\mathbf{U} = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}$ ,  $f(\mathbf{U}) = \begin{pmatrix} \rho u \\ \rho u^2 + P \\ u(P + E) \end{pmatrix}$ ,  $\rho$  is density,  $u$  is velocity,  $E$  is total energy, and  $P$  is pressure given by  $(\gamma - 1)(E - \frac{1}{2}\rho u^2)$  and  $\gamma = 1.4$ . The initial and boundary conditions are given by

$$u(x, 0) = 0, \quad P(x, 0) = P_+, \quad \rho(x, 0) = \rho_+, \quad x > X_p(t),$$

$$P(X_p(t), 0) = P_-, \quad \rho(X_p(t), 0) = \rho_-,$$

and

$$u(X_p(t), t) = \frac{\partial}{\partial t} X_p(t) = u_p(t), \quad t > 0,$$

where  $X_p(t)$  is the position of the piston, and  $u_p(t)$  is the velocity of the piston.

This problem is a moving boundary problem and can be transformed to a fixed boundary

problem by defining a new coordinate  $(y, \tau)$  from  $(x, t)$  via the following transform:

$$y = x - \int_0^\tau u_p(\tau_1, \omega) d\tau_1, \quad \tau = t. \quad (5.2.4)$$

Defining  $v = u - u_p$ , we then have the following Euler equations with a source term [251]:

$$\frac{\partial}{\partial \tau} \mathbf{V} + \frac{\partial}{\partial y} (f(\mathbf{V})) = g(\mathbf{V}) \frac{\partial u_p}{\partial \tau}, \quad (5.2.5)$$

where  $\mathbf{V} = \begin{pmatrix} \rho \\ \rho v \\ \tilde{E} \end{pmatrix}$ ,  $\tilde{E} = \frac{P}{\gamma-1} + \frac{1}{2}\rho v^2$  and  $g(\mathbf{V}) = \begin{pmatrix} 0 \\ -\rho \\ -\rho v \end{pmatrix}$ . The initial and boundary conditions are given by

$$\begin{aligned} v(y, 0) = -U_p, \quad P(y, 0) &= P_+, \quad \rho(y, 0) = \rho_+, \quad y > 0, \\ P(0, 0) &= P_-, \quad \rho(0, 0) = \rho_-, \end{aligned} \quad (5.2.6)$$

and

$$v(0, \tau) = 0, \quad \tau \geq 0.$$

Our goal here is to compute the variance of the shock location perturbation  $z(\tau)$ . The perturbation of the shock location is  $z(\tau) = X_s(\tau) - \tau S = X_s(t) - tS$ , where  $X_s(\tau) = Y_s(\tau) + \int_0^\tau u_p(t_1) dt_1$  is the shock location while  $Y_s(\tau)$  is the shock location under the new coordinate  $(y, \tau)$ .

If we take  $u_p(t) = U_p(1 + \epsilon W(t))$ , where  $W(t)$  is a scalar Brownian motion, we are led to the following Euler equations

$$\frac{\partial}{\partial \tau} \mathbf{V} + \frac{\partial}{\partial y} (f(\mathbf{V})) = \epsilon U_p g(\mathbf{V}) * \dot{W}, \quad (5.2.7)$$

where ‘\*’ denotes two different products as follows:

(1) Stratonovich-Euler equations

$$\frac{\partial}{\partial \tau} \mathbf{V} + \frac{\partial}{\partial y} (f(\mathbf{V})) = \epsilon U_p g(\mathbf{V}) \circ \dot{W}, \quad (5.2.8)$$

where ‘ $\circ$ ’ is the Stratonovich product, or

(2) Ito-Euler equations

$$\frac{\partial}{\partial \tau} \mathbf{V} + \frac{\partial}{\partial y} (f(\mathbf{V})) = \epsilon U_p g(\mathbf{V}) \cdot \dot{W}, \quad (5.2.9)$$

where ‘ $\cdot$ ’ is the Ito product. The initial and boundary conditions are imposed as above. The meaning of ‘ $\circ$ ’ will be explained in Section 5.3.2 and that of ‘ $\cdot$ ’ in Section 5.3.3.

We will verify these two models (5.2.8) and (5.2.9) by solving them numerically with a splitting method in the next section.

### 5.3 Verification of the Stratonovich- and Ito-Euler equations

In the previous section, we introduced two approaches to obtain the variances of the shock location. Here, we verify the correctness of the stochastic Euler equations by comparing the variances of the shock location obtained by two approaches, i.e., the first-order perturbation analysis and the numerical solution of the stochastic Euler equations, up to time  $T = 5$ .

For numerical simulations, we consider the piston velocity  $U_p = 1.25$ , where the Mach number of the shock is  $M = 2$  and  $\gamma = 1.4$ . We normalize all velocities with  $C_+$ , the sound speed ahead of the shock, i.e.  $C_+ = 1$ . Then, the initial conditions are given through the unperturbed relations of states variables [251] as follows:

$$P_+ = 4.5, \quad P_- = 1.0, \quad \rho_+ = 3.73, \quad \rho_- = 1.4.$$

### 5.3.1 A splitting method for stochastic Euler equations

We use a source-term (noise-term) splitting method proposed in [191] for a scalar conservation law with time-dependent white noise source term. Holden and Risebro [191] considered a Cauchy problem on the whole line with multiplicative white noise in Ito's sense:  $\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}f(u) = g(u)\dot{W}(t)$  with deterministic essentially bounded initial condition where  $f, g$  are both Lipschitz, and  $g$  has bounded support. They proved the *almost-sure-convergence* of this splitting method to a weak solution of the Cauchy problem assuming initial condition having bounded support and finitely many extrema while provided no convergence rate.

Here we extend this splitting method to the system (5.2.7). Specifically, given the solution at  $\tau_n, \mathbf{V}^n$ , to obtain the solution at  $\tau_{n+1}$ , we first solve, on the small time interval  $[\tau_n, \tau_{n+1})$ ,

$$\frac{\partial}{\partial \tau} \mathbf{V}^{(1)} + \frac{\partial}{\partial y} (f(\mathbf{V}^{(1)})) = 0, \quad (5.3.1)$$

with the boundary conditions (5.2.6) and initial condition  $\mathbf{V}^{(1)}(\tau_n) = \mathbf{V}^n$ ; then we solve the following Cauchy problem, again on  $[\tau_n, \tau_{n+1})$ ,

$$\frac{\partial}{\partial \tau} \mathbf{V}^{(2)} = \epsilon U_p g(\mathbf{V}^{(2)}) * \dot{W}, \quad (5.3.2)$$

with the initial condition  $\mathbf{V}^{(2)}(\tau_n) = \mathbf{V}^{(1)}(\tau_{n+1})$ . Then the solution at time  $\tau_{n+1}$ ,  $\mathbf{V}^{n+1}$ , is set as  $\mathbf{V}^{(2)}(\tau_{n+1})$  (subject to the error from the splitting). If we denote by  $S(\tau, \tau_n)$  the operator which takes  $\mathbf{V}(\tau_n)$  as initial condition at  $\tau_n$  to the weak solution of (5.3.1) and by  $R(\tau, \tau_n)$  the operator which takes the initial condition at time  $\tau_n$  to the solution of the stochastic differential equation (5.3.2). Then the approximate solution at  $\tau_{n+1}$  is defined by  $\mathbf{V}^{n+1} = R(\tau_{n+1}, \tau_n)S(\tau_{n+1}, \tau_n)\mathbf{V}^n$ . Thus we define a sequence of approximate solution,  $\{\mathbf{V}^n\}$ , to (5.2.7) at time  $\{\tau_n\}$ .

The application of splitting technique requires numerical methods for (5.3.1) and (5.3.2). The splitting scheme allows us to deploy efficient existing methods to solve them separately. To solve

(5.3.1), we use a fifth-order WENO scheme in physical space and second-order strong-property-preserving (SPP) Runge-Kutta in time [212]. In solving (5.3.2), we will employ two different methods: the Monte Carlo method and the stochastic collocation method developed in this work. We employ 1000 points for the fifth-order WENO scheme over the interval  $[0, 5]$  and the time step size  $d\tau = 0.0005$  so that the error from time discretization is negligible. As we mentioned before, our goal is to compute the variance of the perturbed shock location. Since there is always only one shock, we obtain  $Y_s(\tau)$  by finding the biggest jump of pressure, where the error is of order  $\mathcal{O}(dx)$  ( $dx$  is the mesh size in physical space).

### 5.3.2 Stratonovich-Euler equations versus first-order perturbation analysis

We first compare the results obtained by solving the Stratonovich-Euler equations with the Monte Carlo method and those obtained from first-order perturbation analysis.

To solve the Stratonovich-Euler equations (5.2.8) with the splitting method, we need to solve (5.3.2) as follows. By the definition of the Stratonovich integral, we have that, for a square-integrable stochastic process  $h(t)$ ,

$$\int_0^T h(t) \circ dW = \lim_{n \rightarrow \infty} \sum_{i=1}^n h(t_{i+1/2}) \Delta W_i,$$

where  $t_{n+1/2} = \frac{t_{n+1} + t_n}{2}$  and  $\Delta W_n = W(t_{n+1}) - W(t_n)$ . The limit is understood in the mean-square sense [330]. Thus, we will solve Equation (5.3.2) by the following Crank-Nicolson scheme

$$\mathbf{V}^{(2)}(\tau_{n+1}) = \mathbf{V}^{(2)}(\tau_n) + \epsilon U_p g(\mathbf{V}^{(2)}(\tau_{n+1/2})) \Delta W_n. \quad (5.3.3)$$

In our simulation, the values of function  $g(\mathbf{V}^{(2)}(\tau))$  at  $\tau_{n+1/2}$  are approximated by the average values  $\frac{g(\mathbf{V}^{(2)}(\tau_n)) + g(\mathbf{V}^{(2)}(\tau_{n+1}))}{2}$ . Note that for the specific form of  $g$ , we do not have to invert the

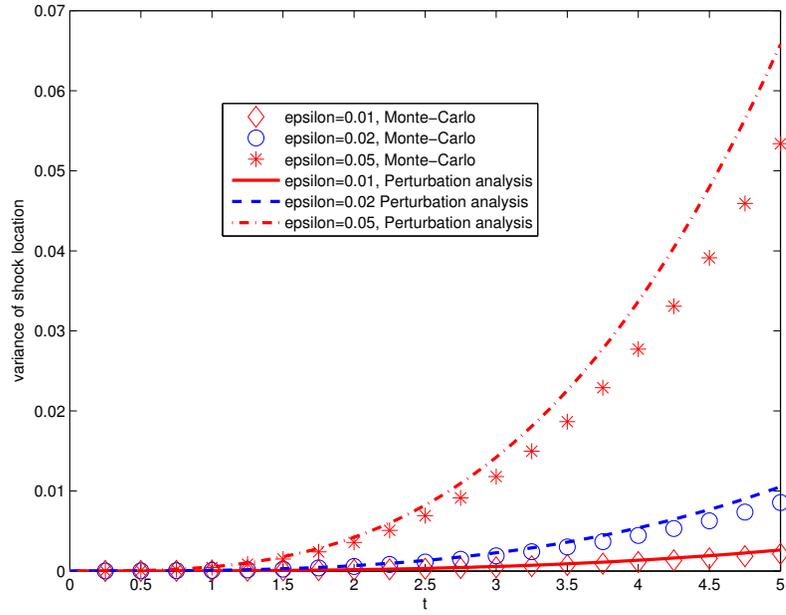
resulting matrix in (5.3.3).

Figure 5.2 verifies that the Stratonovich-Euler equations (5.2.8) can capture the variances of shock location for the stochastic piston problem driven by Brownian motion. Here we employ 10,000 realizations so that the statistical error can be neglected for noises with amplitude no less than 0.05. We also note that for noises with amplitude less than 0.05, the error of the adopted methods is dominated by the statistical error from the Monte Carlo method and also the space discretization error from WENO. Figure 5.2 presents the variances obtained by the Monte Carlo method (5.3.1)-(5.3.3) and those from variances estimates by the first-order perturbation analysis (5.2.2). We observe the agreement between the results from the Monte Carlo method and the perturbation analysis within small time and for small noises. Figure 5.2(a) shows the results for small noises, i.e.,  $\epsilon \sim O(10^{-2})$  while Figure 5.2(b) for large noises, i.e.,  $\epsilon \sim O(10^{-1})$ . The difference between the variances from the Monte Carlo method and the first-order perturbation analysis (5.2.2) is at most 12% – 13% of the variances (5.2.2), up to time  $T = 5$ , for all cases except for the case  $\epsilon = 0.5$ ; for the latter, the difference between the variances is at most 19.3% of the variance (5.2.2). However, for small time ( $t < 1$ ) the variances by Monte Carlo and perturbation analysis agree well, while they deviate much after  $t = 2$ . This effect can be explained as follows. For  $t < 1$ , the variance of the driving process (Brownian motion) has small value ( $\sqrt{t}$ ) corresponding to a weak perturbation; while at later time it has larger value increasing substantially the perturbation. (We remind the reader that the perturbation process in [251] has unit variance.)

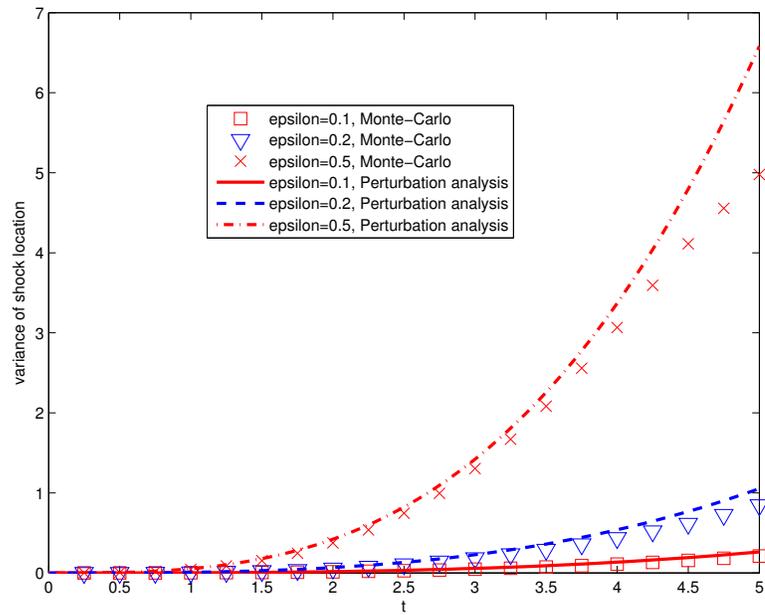
### 5.3.3 Stratonovich-Euler equations versus Ito-Euler equations

For the Ito-Euler equations (5.2.9), we solve (5.3.2) by the forward Euler scheme

$$\mathbf{V}^{(2)}(\tau_{n+1}) = \mathbf{V}^{(2)}(\tau_n) + \epsilon U_p g(\mathbf{V}^{(2)}(\tau_n)) \Delta W_n. \quad (5.3.4)$$



(a) Small noises



(b) Large noises

Figure 5.2: Comparison between the results from first-order perturbation analysis (5.2.2) and solving the Stratonovich-Euler equations (5.2.8) by the splitting method (5.3.1)-(5.3.3).

Recall that the Ito integral is defined as, see e.g. [330],

$$\int_0^T h(t) \cdot dW = \lim_{n \rightarrow \infty} \sum_{i=1}^n h(t_i) \Delta W_i.$$

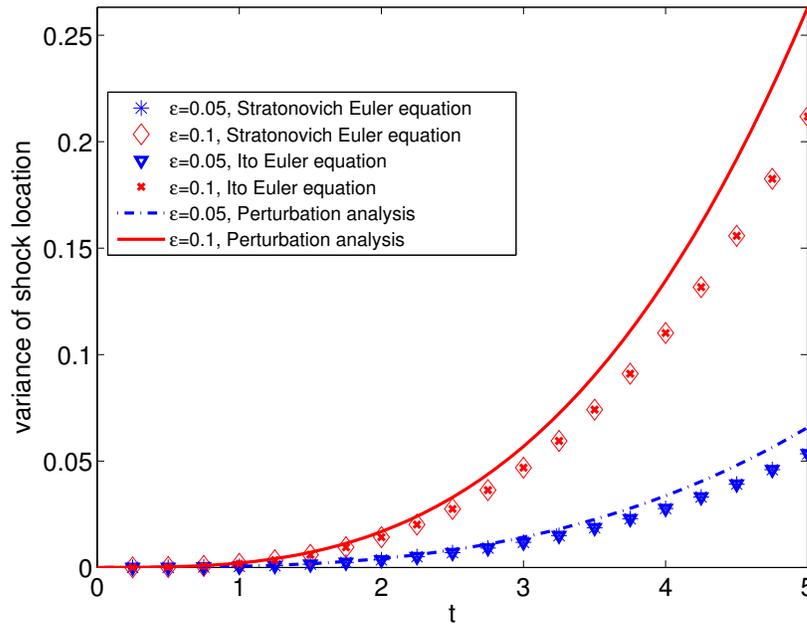
Next we compare the numerical results for the Stratonovich-Euler equations and the Ito-Euler ones using the above discretization in time. We observe from Figure 5.3 that for both small and large noises, these two types of equations have almost the same variances for the perturbed shock location  $\mathbb{E}[z^2(t)]$  up to time  $T = 5$ . Actually, the difference of variances by the Stratonovich-Euler and Ito-Euler equations for  $\epsilon \leq 0.2$  is less than  $10^{-3}$  up to time  $t = 5$  which lies within the discretization errors. For  $\epsilon = 0.5$ , we present in Table 5.1 the difference of variances for these two approaches using the same sequence of Monte Carlo points. The Stratonovich-Euler equations exhibits larger variances in large time but the difference from those by the Ito-Euler equations is less than 10% of the variances by Ito-Euler equations. We then conclude that the Stratonovich-Euler equations are a suitable model for the piston problem driven by Brownian motion and we will consider only this approach hereafter.

Table 5.1: The difference of variances of shock location by Stratonovich-Euler and Ito-Euler equations for  $\epsilon = 0.5$ .

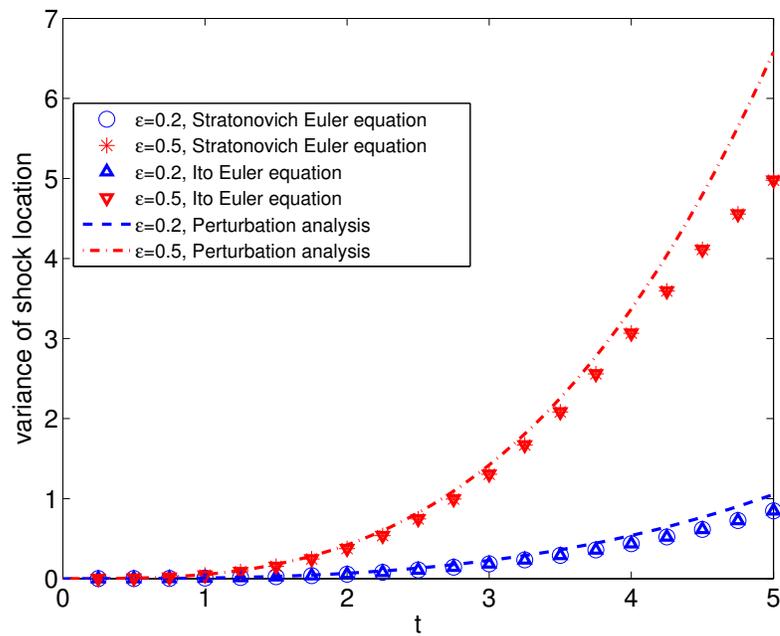
t	1.0	2.0	3.0	4.0	5.0
	0.0007	0.0129	0.0742	0.2353	0.2421

## 5.4 The Stochastic Collocation Method

Next we test the stochastic collocation method versus the Monte Carlo method for the Stratonovich-Euler equations (5.2.8). To solve the Stratonovich-Euler equations (5.2.8), we again use the splitting method (5.3.1)-(5.3.2). In (5.3.2), we adopt the stochastic collocation method, where we first introduce a spectral approximation for the Brownian motions and subsequently apply the sparse grid method. Specifically, we first approximate Brownian motion with its spectral approximation,



(a)  $\epsilon = 0.05, 0.1$



(b)  $\epsilon = 0.2, 0.5$

Figure 5.3: Comparison between solving Stratonovich-Euler equations (5.2.8) and Ito-Euler equations (5.2.9) by the splitting method (5.3.1)-(5.3.2).

using  $K$  multi-elements:

$$W^{(n,K)}(\tau) = \sum_{k=0}^{K-1} \sum_{i=1}^n \int_0^\tau \chi_{[t_k, t_{k+1})}(s) m_{k,i}(s) ds \xi_{k,i}, \quad \tau \in [0, T],$$

where  $0 = t_0 < t_1 < \dots < t_K = T$ ,  $\chi_{[t_k, t_{k+1})}(\tau)$  is the indicator function of the interval  $[t_k, t_{k+1})$ ,  $\{m_{k,i}\}_{i=1}^\infty$  is a complete orthonormal basis in  $L^2([t_k, t_{k+1}])$ , and  $\xi_{k,i}$  are mutually independent standard Gaussian random variables (with zero mean and variance one). Hence, we obtain the following partial differential equation with smooth inputs:

$$\frac{\partial}{\partial \tau} \mathbf{V}^{(2)} = \epsilon U_p g(\mathbf{V}^{(2)}) \sum_{k=0}^K \sum_{i=1}^n \chi_{[t_{k-1}, t_k)}(\tau) m_{k,i}(\tau) \xi_{k,i}. \quad (5.4.1)$$

In (5.4.1) we apply the stochastic collocation method [9, 368, 403] for smooth noises; see Appendix A for a brief review on the stochastic collocation method for white noise. The stochastic collocation method we adopt here is the sparse grid of Smolyak type based on 1D Gaussian-Hermite quadrature; we refer to [125] for implementation details.

The first issue we have for the piston problem here is the discontinuity of the solution to (5.2.8), where the condition for spectral approximation to work may be invalid [364]. In practice, we solve the problem with the WENO scheme, which smears the shock somewhat, and thus we have higher regularity than that of the original problem. A second issue is that the use of the stochastic collocation method (Smolyak sparse grid) with Gaussian quadrature may not exhibit fast convergence because of the low regularity. Thus, we use  $n = 1$  or  $2$  with large  $K$  (small time step in  $W^{(n,K)}$ ) instead of large  $n$  with small  $K$ . This choice of  $n$  is verified with control tests with  $n = 3, 4$  for different  $K$ , where the numerical results show large deviations from those of Monte Carlo method with high oscillations. We choose a low sparse grid level (i.e. two) to be consistent with the ‘available regularity’ (numerical tests with high sparse grid level show an instability). The third issue is the so-called ‘curse-of-dimensionality’. In practice, when the number of random

variables,  $Kn$ , increases, the Smolyak sparse grid method will not work well and will be replaced by the QMC method.

Here we adopt a uniform partition of the time interval  $[0, T]$ , that is  $\mathbf{t}_k = (k-1)\Delta$ ,  $k = 1, \dots, K$ . The complete orthonormal basis we employ in  $L^2([\mathbf{t}_k, \mathbf{t}_{k+1}])$  is the cosine basis

$$m_{k,1}(t) = \frac{1}{\sqrt{\Delta}}, \quad m_{k,i}(t) = \sqrt{\frac{2}{\Delta}} \cos\left(\frac{(i-1)\pi}{\Delta}(t - \mathbf{t}_k)\right), \quad i \geq 2.$$

Figure 5.4 compares the numerical results from the Monte Carlo method (5.3.1)–(5.3.3) and the stochastic collocation method for (5.3.1) and (5.4.1) with both small and large noises. For each  $\epsilon$ , we use different  $\Delta$  (the length of the uniform partition of time interval  $[0, T]$ ), i.e. different size of elements  $K$ . We note that all the numerical solutions obtained by the stochastic collocation method agree with those from the Monte Carlo method (5.3.1)–(5.3.3) within small time. Here we do not observe convergence in  $n$ , recalling that such convergence requires smoothness in random space.

We note that smaller  $\Delta$  and larger  $n$  may lead to a larger number of random variables and thus the break down of the sparse grid method [403]. So we first test the cases of small  $\Delta$  such that we can apply the sparse grid method. Figure 5.4 shows that a low level sparse grid method works well for the piston problem with small perturbations. We note that our sparse grid level is two and thus the number of collocation points is  $2n\frac{T}{\Delta} + 1$ .

When  $n = 1$ , we observe in Figure 5.4 good agreement of the results by the stochastic collocation method and the Monte Carlo method in small time ( $t \leq 2$ ). Notice that when  $n = 1$ , (5.4.1) is the classical Wong-Zakai approximation [399]

$$\frac{\partial}{\partial \tau} \mathbf{V}^{(2)} = g(\mathbf{V}^{(2)}) \frac{1}{\sqrt{\Delta}} \sum_{k=0}^{K-1} \chi_{[\mathbf{t}_k, \mathbf{t}_{k+1})}(\tau) \xi_{k,1}. \quad (5.4.2)$$

However, for  $n = 2$ , there are some disagreements between the results. In Figure 5.4(a) and 5.4(c), the results of the case  $n = 2$  and  $\Delta = 0.2$  (note that we have  $nK = 50$  random variables)

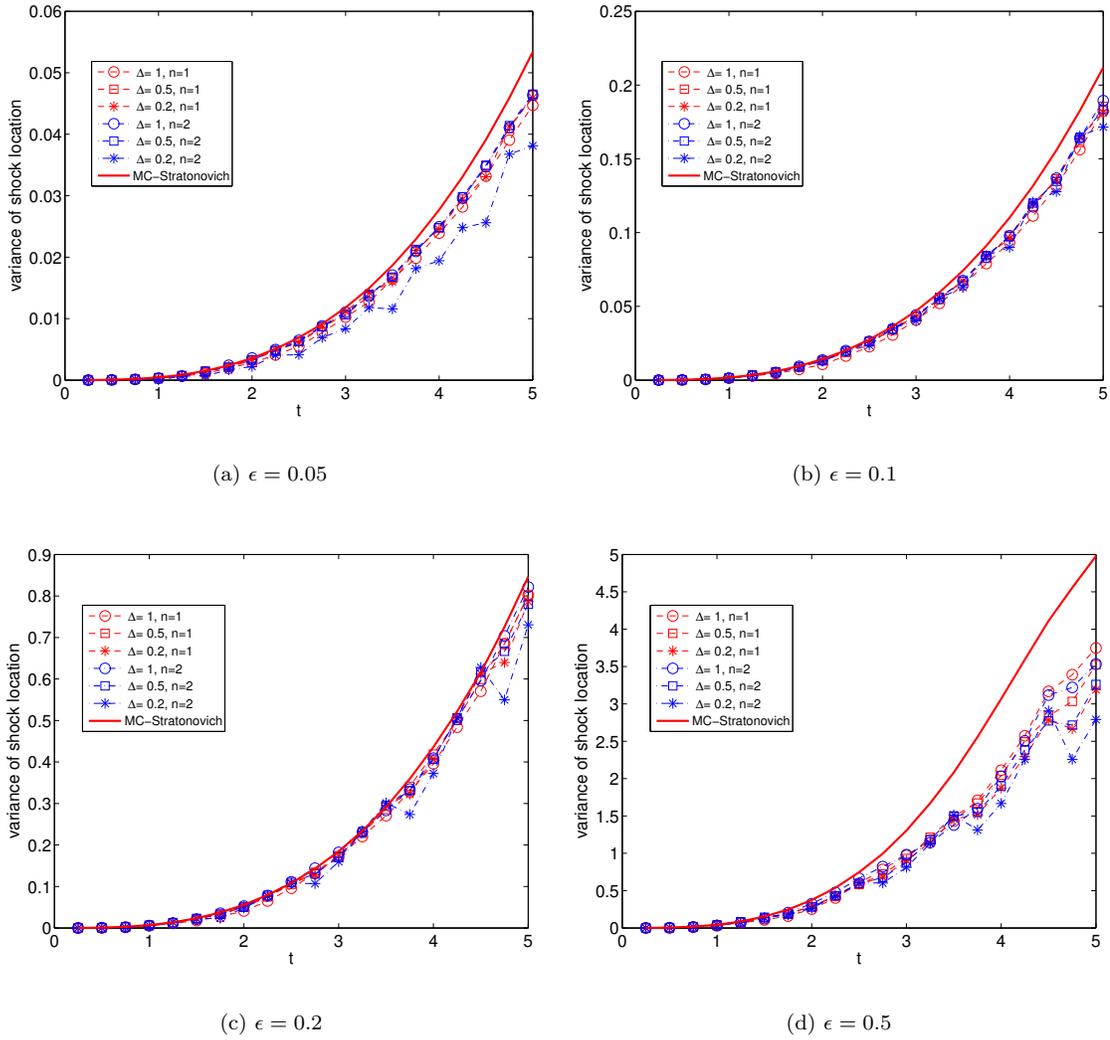


Figure 5.4: Comparison between numerical results from Stratonovich-Euler equations (5.2.8) using the direct Monte Carlo method (5.3.1) and (5.3.3) and the stochastic collocation method (5.4.1). The sparse grid level is 2 and  $\Delta$  is the size of element in time in the stochastic collocation method.

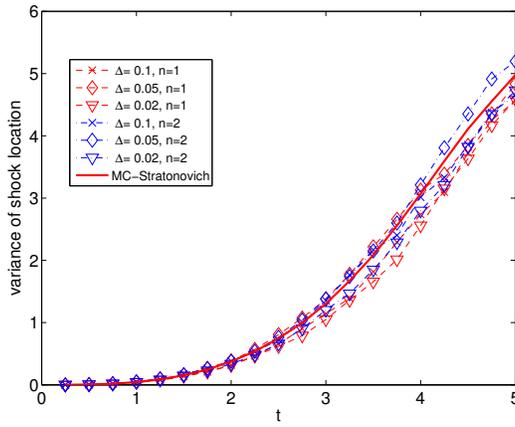
underestimate those results from the Monte Carlo method and the stochastic collocation method with a smaller number of random variables ( $n = 1$ ). The larger number of random variables ( $n = 2$  here) does not result in convergence since we do not have a smooth solution as we mention above.

For the case with large perturbation,  $\epsilon = 0.5$ , we require smaller  $\Delta$  and thus more random variables. This is why we observe the disagreement in Figure 5.4(d). For all cases in Figure 5.4, we observe a deviation of numerical results by stochastic collocation methods from those of Monte Carlo method over large time. Similar effects arise in the application of spectral methods in random space, e.g., in Wiener chaos methods. The interested reader may refer to [420] for a discussion of this effect.

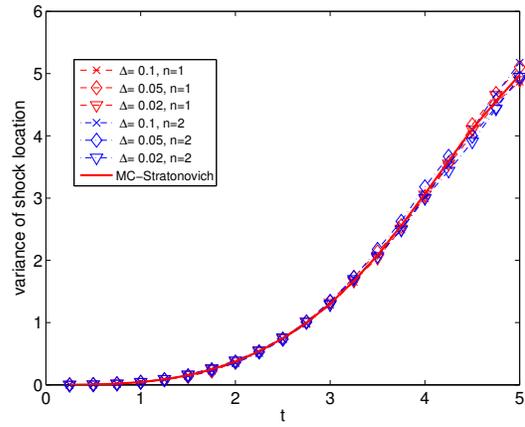
To adapt to the high dimensionality (large number of random variables), we employ the QMC method instead of sparse grid methods. We consider two popular QMC sequences: one is a scrambled Halton with the method RR2 proposed in [222]; and the other is a scrambled Sobol sequence suggested in [287]. Both sequences lie in hypercube and thus an inverse transformation is adopted to generate sequences in the entire space based on these two sequences. In Figure 5.5, we test the large noise case, i.e.  $\epsilon = 0.5$ . Both Halton and Sobol sequences work if a moderately large sample of the sequences is adopted. For 1000 sample points, variances from both sequences are closer to those from Monte Carlo method (5.3.1)-(5.3.3) than those from 500 sample points of both sequences.

## 5.5 Summary

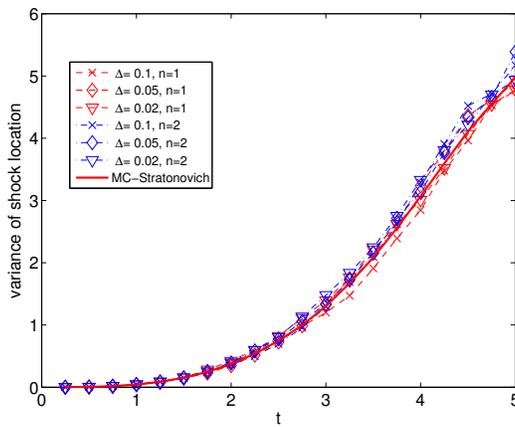
We simulated a stochastic piston problem by time-varying Brownian motions of a piston moving inside an adiabatic tube of constant area, which is governed by the Euler equations driven by white noise. By splitting the Euler equations into two parts – a ‘deterministic part’ and a ‘stochastic part’ – we solved the ‘stochastic part’ by the Monte Carlo method and the stochastic collocation method. The numerical results show that the variances of the shock location grow cubically with time, which are significantly different from those from colored noise driven piston. In Figure 5.6 we compare the



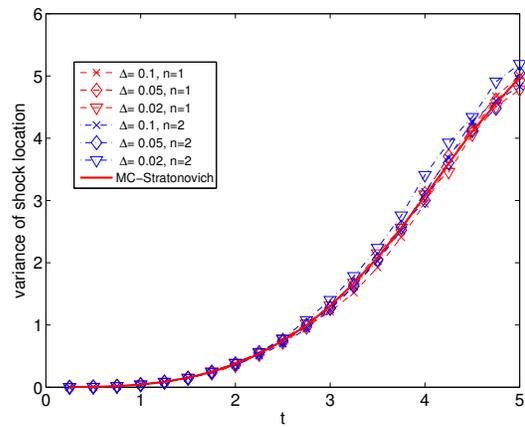
(a) Sobol sequence: 500 sample points



(b) Sobol sequence: 1000 sample points



(c) Halton sequence: 500 sample points



(d) Halton sequence: 1000 sample points

Figure 5.5: Comparison between numerical results from Stratonovich-Euler equations (5.2.8) using direct Monte Carlo method (5.3.1)-(5.3.3) and the QMC method for (5.4.1) with a large noise:  $\epsilon = 0.5$ .

variances of shock positions induced by three different Gaussian noises: Brownian motion, random process with zero mean and exponential covariance kernel  $\exp(-|t_1 - t_2|)$ , and standard Gaussian random variable, where the noise amplitude is  $\epsilon = 0.1$ . The results are obtained via the stochastic perturbation analysis. The case of Brownian motion induces smaller values of variances than the other two cases for short times and greater values of variances for longer times. We note that the

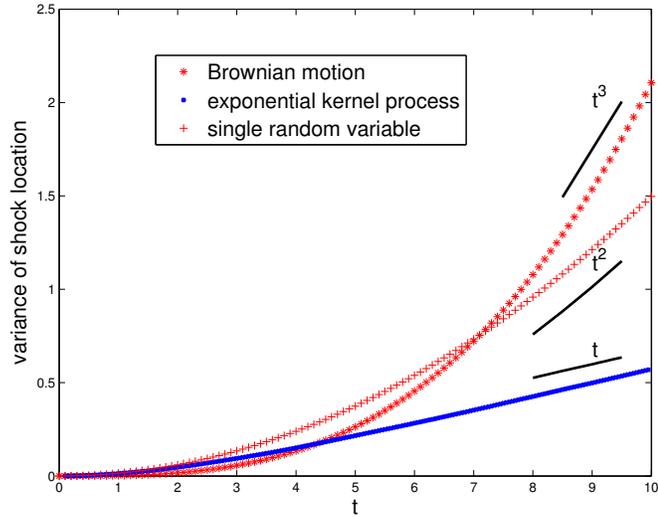


Figure 5.6: Comparison among variances of shock positions induced by three different Gaussian noises: Brownian motion, random process with zero mean and exponential covariance kernel  $\exp(-|t_1 - t_2|)$  and standard Gaussian random variable. The noise amplitude is  $\epsilon = 0.1$ .

effects of different Gaussian processes are similar to a first-order stochastic differential equation responding to different Gaussian processes. The shock location depends on the time integration of the underlying Gaussian processes as the solution to stochastic differential equation does; see Appendix B for details.

Firstly, we solved the ‘stochastic part’ using the Monte Carlo method by the definition of Stratonovich integral and verified the Stratonovich-Euler equations by the first-order perturbation analysis presented in [251]. Secondly, we solved the Stratonovich-Euler equations by solving the ‘stochastic part’ with the stochastic collocation method using a multi-element spectral approximation of the Brownian motion. Finally, we tested two types of QMC sequences for the ‘stochastic

part' using a multi-element spectral approximation of the Brownian motion when the noise is large. The stochastic collocation and QMC methods are superior to the Monte Carlo method in the sense that they can achieve faster convergence than the classic Monte Carlo method.

The low accuracy of the stochastic collocation method, especially for long times, is caused by the discontinuity of the solution. Due to the deterministic solver, we have that the accuracy for the numerical shock location is only first-order in the spatial step size, i.e.,  $\mathcal{O}(dx)$ . For small noises, we had agreement between the results from the Euler solver and those from perturbation analysis. However, for large noises, we need small time-interval  $\Delta$  for the stochastic collocation method to converge. As smaller time-interval  $\Delta$  leads to larger number of random variables, we adopted the QMC method which led to accurate solutions.

With regards to computational efficiency, the stochastic collocation method is more efficient than Monte Carlo simulation when a small number of random variables are involved, where the number of collocation points is far less than Monte Carlo sampling points. As time becomes larger, we introduce more random variables and thus we need to employ the more efficient QMC method. In other applications involving long-time integration, it may be possible to use all three different ways of sampling, i.e., starting with sparse grid for early time, continuing with the QMC for moderate time and even switching to the Monte Carlo method for long time.

## 5.6 Appendix: a first-order model driven by different Gaussian processes

Consider the following simple ordinary differential equation with multiplicative noise:

$$dy = k(t, \omega)y dt, \quad y(0) = y_0. \quad (5.6.1)$$

Here we take  $y_0 = 1$  for simplicity. Suppose  $k(t, \omega)$  is a Gaussian random variable or process with zero mean. Specifically,  $k(t, \omega)$  will take the following form:

- $k(t, \omega) =: \xi \sim \mathcal{N}(0, 1)$ ;
- $k(t, \omega) =: V(t, \omega)$  where the two-point correlation function of  $V(t)$  is  $\exp(-\frac{|t_1 - t_2|}{A})$ .
- $k(t, \omega) =: W(t, \omega)$  is the standard Brownian motion:  $\mathbb{E}[W(t)W(s)] = \min(t, s)$ .
- $k(t, \omega) =: \dot{W}(t, \omega)$  is the white noise:  $\mathbb{E}[\dot{W}(t)\dot{W}(s)] = \delta(t - s)$ .

**Remark 5.6.1.** When  $k(t, \omega) =: \dot{W}(t, \omega)$ , Equation (5.6.1) is understood in the Stratonovich sense:

$$dy = y \circ dW(t), \quad y(0) = y_0. \quad (5.6.2)$$

The exact solution to Equation (5.6.1) is  $y = y_0 \exp(K(t))$ , where  $K(t) = \int_0^t k(s) ds$  is again Gaussian with mean zero and variance  $\sigma^2$ . Then we have the moments of the solution  $y$ , for  $m = 1, 2, \dots$ ,  $\mathbb{E}[y^m(t)] = y_0^m \exp(\frac{m^2}{2}\sigma^2)$ , where  $\sigma^2 = t^2, 2At + 2A^2(\exp(-\frac{t}{A}) - 1), \frac{t^3}{3}, t$  for the listed processes, respectively. Figures 5.7 and 5.8 illustrate the different behavior of second-order moments with small time  $t \in [0, 1]$  and larger time  $t > 1$ . The amplitudes of variances are similar to those of variances of the shock location in Figure 5.6, indicating three different behaviors in three different time intervals.

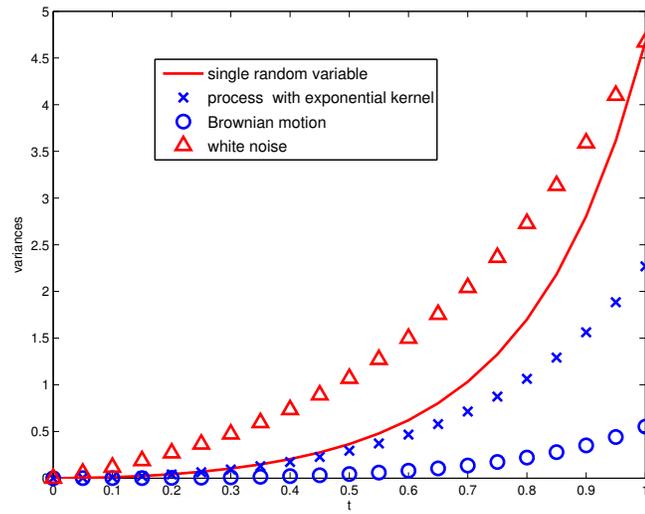
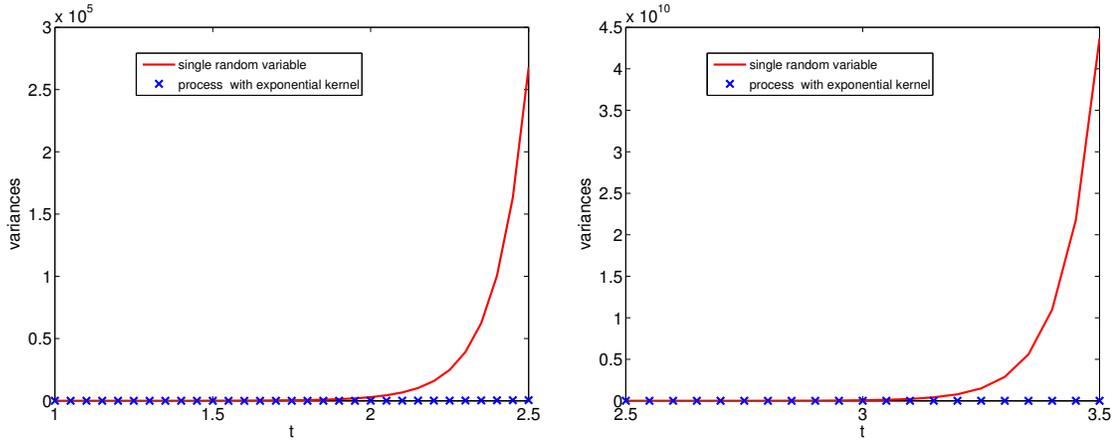
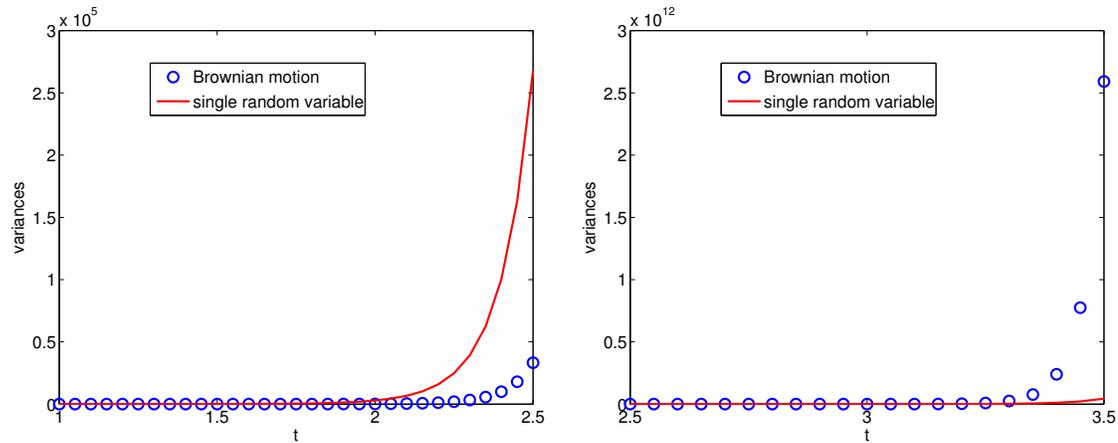


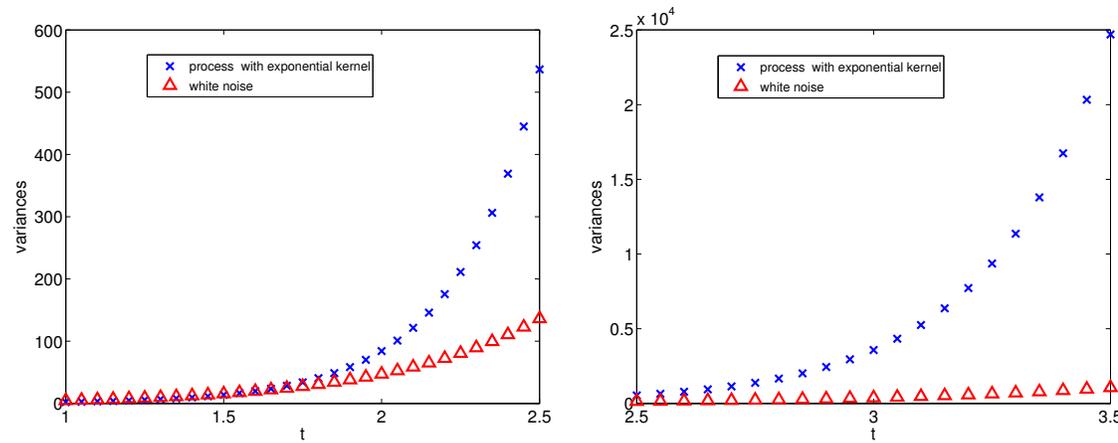
Figure 5.7: Comparison of variances of the solutions for four models of  $k(t)$  up to time  $t = 1$ :  $A = 1$ .



(a) Comparison between the cases of single random variable and a process with exponential kernel



(b) Comparison between the cases of Brownian motion and single random variable



(c) Comparison between the cases of white noise and a process with exponential kernel

Figure 5.8: Comparison of variances of the solutions for four models of  $k(t)$  at large time:  $A = 1$ .

## Part II: Spatial White Noise

## Chapter 6

# Semilinear Elliptic Equations with Additive Noise

We investigate in this chapter the strong and weak convergence order of piecewise linear finite element methods for a class of semilinear elliptic equations with additive spatial white noise using a spectral truncation of white noise. We show that the strong convergence order of the finite element approximation is  $h^{2-d/2-\epsilon}$  where  $h$  is the element size,  $d$  is the dimension and  $\epsilon > 0$  is arbitrarily small. We also show that the weak convergence order is twice the strong convergence order for one- and two-dimensional problems. Numerical results confirm our prediction for a two-dimensional semilinear elliptic problem.

### 6.1 Introduction

We study numerical approximation of the following semilinear elliptic equation with additive white noise using a spectral approximation of spatial white noise:

$$-\Delta u(x) + f(u(x)) = g(x) + \frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} W(x), \quad x = (x_1, \dots, x_d) \in \mathcal{D}, \quad (6.1.1)$$

with Dirichlet boundary condition

$$u(x) = 0, \quad x \in \partial\mathcal{D}, \quad (6.1.2)$$

where  $\mathcal{D} = (0, 1)^d$ ,  $W(x)$  is a Brownian sheet on  $\bar{\mathcal{D}} = [0, 1]^d$ ,  $f$  is Lipschitz continuous and  $g \in L^2(\mathcal{D})$  so that (6.1.1) is well-posed, see Section 6.2 for details.

Discretizing white noise (Brownian motion) is important in the numerical methods of stochastic differential equations (SDEs) driven by white noise. In the literature of numerical SDEs, several types of discretizations of Brownian motion have been proposed: piecewise linear approximation (see e.g. [400, 399] and [202, p. 396]), spectral approximation (see e.g. [301]), wavelet approximation (also known as Levy-Ciesielsky approximation, see e.g. [214]) and mollifier approximation (see e.g. [202, p. 397] and [329]), and so on.

Investigating such a benchmark problem (6.1.1) will be helpful to better understand the influence of discretizing Brownian motion/sheet as well as more complex noises in the context of approximating stochastic partial differential equations. For example, when higher dimensional white noise is considered, which is the case for space-time white noise (see e.g. [209]), we can combine one of the above approximation methods in each dimension and thus have different approximation of white noise. It is then crucial to understand the performance of different approximation methods in a simple case such as the problem (6.1.1) .

The piecewise linear approximation for Brownian motion leads to a piecewise constant approximation of white noise, which has been widely used in approximating temporal noise for solving SDEs (see e.g. [218] and [301]) as well as in approximating spatial noises (see e.g. [5, 55, 105, 165]). Using piecewise constant approximation, Gyongy and Martinez [165] considered a finite difference scheme in physical space for the problem (6.1.1) and obtained dimension-dependent convergence order in the mean-square sense:  $h$  if  $d = 1$ ; and  $h^{2-\frac{d}{2}-\epsilon}$  if  $d = 2, 3$ , where  $h$  is the finite difference step size. Here and throughout the paper,  $\epsilon > 0$  is an arbitrary small constant. For finite element methods for (6.1.1) in physical space, [5] obtained the mean-square convergence order  $h$  for one-

dimensional linear problem and [55] considered a two-dimensional problem (6.1.1) over a general bounded convex domain and established the mean-square convergence order  $h^{1-\epsilon}$ . In other words, the finite element methods basically yield the same mean-square convergence order as the finite difference methods do for  $d = 1, 2$ , when piecewise constant approximation of white noise is used.

With a spectral approximation of the spatial additive noise, we will show that the mean-square convergence order is  $h^{2-d/2-\epsilon}$ , where we use piecewise linear finite element approximation in physical space, see Theorem 6.2.5. Specifically, in the one-dimensional case, we obtain the mean-square convergence order  $h^{3/2-\epsilon}$  instead of  $h$  from the piecewise constant approximation of white noise [5]. We note that for  $d = 1$ , the solution is actually in  $H^{3/2-\epsilon}(\mathcal{D})$  and the spectral approximation benefits from the smoothness of the solution as will be shown in Section 6.2, where we also show similar effects for fourth-order equations.

We will also show that the weak convergence order of the spectral approximation of white noise is twice its mean-square convergence order when only the white noise is discretized in (6.1.1), see Theorem 6.2.3. While further discretizing (6.1.1) with a piecewise linear finite element method, we show that the weak error is  $h^{4-d-\epsilon}$  for  $d = 1, 2$ , see Theorem 6.2.7. We will also present some numerical results for one- and two-dimensional semilinear elliptic equations in Section 6.3.

Besides the problem considered here, this spectral approximation can be and has been considered for elliptic equations with multiplicative noise, see e.g. linear elliptic equation with lognormal diffusivity [64, 65, 119, 120] and with white noise diffusivity [393]. The spectral approximation of white noise can be further applied to those problems in a more complex domain and with different boundary conditions, see e.g. [5, 57, 59, 58, 105, 215, 390, 407, 408, 412] where the piecewise constant approximation is used.

## 6.2 Main results

For (6.1.1) to be well-posed, we require the following assumption as in [45, 55, 165]: the nonlinear function  $f$  can be more general than just Lipschitz continuous as in the Introduction. The following assumption can allow  $f$  to be a sum of non-decreasing bounded functions and a Lipschitz continuous function with a small Lipschitz constant.

**Assumption 6.2.1.** *The function  $f$  satisfies the following conditions:*

- *There exists a constant  $L < C_p$  such that*

$$[f(s) - f(t)](s - t) \geq -L |s - t|^2, \quad \forall s, t \in \mathbb{R}. \quad (6.2.1)$$

- *There exist constants  $M \geq 0$  and  $R \geq 0$  such that*

$$|f(s) - f(t)| \leq M + R |s - t|, \quad \forall s, t \in \mathbb{R}. \quad (6.2.2)$$

Here  $C_p$  is the constant in the Poincaré inequality:

$$\|\nabla v\|^2 \geq C_p \|v\|^2, \quad v \in H_0^1(\mathcal{D}).$$

Under Assumption 6.2.1, the solution to (6.1.1) is proved to exist and be unique in  $L^p(\Omega, L^2(\mathcal{D}))$  when  $d \leq 3$  [45]. For  $d > 4$ , for Equation (6.1.1) to be well-posed in  $L^p(\Omega, L^2(\mathcal{D}))$ , [284] considered additive color noise instead of white noise. The solution to (6.1.1) is understood in the sense of a mild solution:

$$u(x) + \int_{\mathcal{D}} \mathcal{K}(x, y) f(u(y)) dy = \int_{\mathcal{D}} \mathcal{K}(x, y) g(y) dy + \int_{\mathcal{D}} \mathcal{K}(x, y) dW(y). \quad (6.2.3)$$

where  $\mathcal{K}(x, y)$  is Green's function of the Laplace equation  $-\Delta v = c(x)$  and  $v(x) = 0$  on  $\partial\mathcal{D}$  and

$$v(x) = \int_{\mathcal{D}} \mathcal{K}(x, y) c(y) dx.$$

### 6.2.1 Semidiscrete and fully discrete schemes

Here we represent the spatial white noise  $\dot{W}(x)$  with an orthogonal series expansion

$$\frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} W(x) = \sum_{\alpha \in \mathcal{J}} e_{\alpha}(x) \xi_{\alpha}, \quad (6.2.4)$$

or for the spatial Brownian motion (Brownian sheet)

$$W(x) = \sum_{\alpha \in \mathbb{N}^d} \int_0^{x_d} \int_0^{x_{d-1}} \cdots \int_0^{x_1} e_{\alpha}(y) dy_1 \cdots dy_d \xi_{\alpha}, \quad (6.2.5)$$

where  $\{e_{\alpha}(x)\}_{|\alpha|=1}^{\infty}$  is a complete orthonormal basis in  $L^2(\mathcal{D})$ ;  $\xi_{\alpha}$ ,  $\alpha = (\alpha_1, \alpha_2, \alpha_3)$  are independent standard Gaussian random variables. In practice, we can take any orthonormal basis in  $L^2(\mathcal{D})$ .

Here we take the eigenfunctions of the elliptic equation

$$-\Delta \psi = \lambda \psi, \quad x \in \mathcal{D} \quad \psi = 0, x \in \partial \mathcal{D}, \quad (6.2.6)$$

which can form an orthonormal basis in  $L^2(\mathcal{D})$ . We denote the truncation of  $W(x)$  (6.2.5) by  $W_n$ :

$$W_n(x) = \sum_{|\alpha| \leq n, \alpha \in \mathbb{N}^d} \int_0^{x_d} \int_0^{x_{d-1}} \cdots \int_0^{x_1} e_{\alpha}(y) dy_1 \cdots dy_d \xi_{\alpha}. \quad (6.2.7)$$

A semi-discrete scheme of (6.1.1) (to be precise, (6.2.3)) is then as follows

$$u_n(x) + \int_{\mathcal{D}} \mathcal{K}(x, y) f(u_n(y)) dy = \int_{\mathcal{D}} \mathcal{K}(x, y) g(y) dy + \int_{\mathcal{D}} \mathcal{K}(x, y) dW_n(y). \quad (6.2.8)$$

which is equivalent to

$$-\Delta u_n(x) + f(u_n(x)) = g(x) + \frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} W_n(x). \quad (6.2.9)$$

The scheme (6.2.9) requires further discretization in physical space. Here we consider a finite element approximation. Let  $V_h$  be a linear finite element subspace of  $H_0^1(\mathcal{D})$  with quasi-uniform triangulation  $T_h$ . Then the linear finite element approximation of  $u_n$  in (6.2.9) is to find  $u_n^h \in V_h$  such that

$$(\nabla u_n^h, \nabla v) + (f(u_n^h), v) = (g + \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W_n, v), \quad \forall v \in V_h. \quad (6.2.10)$$

## 6.2.2 Strong and weak convergence order

In this section, we will only present our conclusions on strong and weak convergence orders while the proofs can be found in Section 6.4.

**Theorem 6.2.2** (Strong error). *Let  $u$  be the solution to (6.1.1) and  $u_n$  the solution to (6.2.9).*

*Under Assumption 6.2.1, we have*

$$\mathbb{E}[\|u - u_n\|^2] \leq C(M(\sum_{|\alpha|=n+1}^{\infty} \lambda_{\alpha}^{-2})^{1/2} + \sum_{|\alpha|=n+1}^{\infty} \lambda_{\alpha}^{-2}) \leq C\epsilon^{-1}(Mn^{-(2-d/2-\epsilon/2)} + (C_p + R)^2 n^{-(4-d-\epsilon)}),$$

*where the constant  $C$  depends only on  $d, C_p, L$  and  $M, R, \lambda_{\alpha}$  are eigenvalues of the problem (6.2.6).*

*The constants  $L, M, R$  are from Assumption 6.2.1.*

Under further smoothness of the nonlinear function  $f$ , we show that the weak convergence order is twice of the mean-square convergence order.

**Theorem 6.2.3** (Weak error). *Let  $u$  be the solution to (6.1.1) and  $u_n$  the solution to (6.2.9). In addition to Assumption 6.2.1, assume also that  $f$  and  $F$  and their derivatives up to fourth-order*

are of at most polynomial growth at infinity:

$$\left| \frac{d^k}{dx^k} G(x) \right| \leq c(1 + |x|^\varkappa), \quad \varkappa < \infty, k = 1, 2, 3, 4. \quad (6.2.11)$$

Furthermore, we assume that  $M = 0$  in Assumption 6.2.1. Then we have

$$\|\mathbb{E}[F(u)] - \mathbb{E}[F(u_n)]\| \leq C \sum_{|\alpha|=\mathfrak{n}+1}^{\infty} \lambda_\alpha^{-2} \leq C \mathfrak{n}^{-(4-d-\epsilon)}. \quad (6.2.12)$$

The constant  $C$  depends on  $d, C_p, L, M, R$  as in Theorem 6.2.2 and also the constant in (6.2.11).

Before we state the conclusion for finite element approximation (6.2.10), we illustrate the above conclusions for a linear elliptic equation with additive noise.

**Example 6.2.4** (Linear elliptic equation, see e.g. [5, 105]). We consider the following linear problem:

$$\begin{aligned} -\Delta u + bu(x) &= g(x) + \frac{\partial}{\partial x} W(x), \quad x \in (0, 1), \\ u(0) = u(1) &= 0, \end{aligned}$$

where  $b > -\pi^2$  and  $g \in L^2([0, 1])$ . The solution can be represented by  $u = \sum_{k=1}^{\infty} \frac{\xi_k + g_k}{b + k^2\pi^2} e_k$ , where  $g_k = (g, e_k)$ ,  $\frac{\partial}{\partial x} W(x) = \sum_{k=1}^{\infty} e_k \xi_k$ ,  $\xi_k$  are i.i.d. standard random variables. The basis  $\{e_k\}_{k=1}^{\infty}$  can be any orthonormal basis in  $L^2([0, 1])$ , e.g.  $e_k = \sqrt{2} \sin(k\pi x)$ , which are the corresponding eigenfunctions of  $-\Delta u = \lambda u$  over  $[0, 1]$  with  $u(0) = u(1) = 0$ . The first two moments of  $u$  are

$$\mathbb{E}[u] = \sum_{k=1}^{\infty} \frac{g_k}{b + k^2\pi^2} e_k \quad \mathbb{E}[\|u\|^2] = \sum_{k=1}^{\infty} \frac{1 + g_k^2}{(b + k^2\pi^2)^2} = \sum_{k=1}^{\infty} \frac{1}{(b + k^2\pi^2)^2} + \|\mathbb{E}[u]\|^2.$$

It can be readily checked that there exists  $C > 0$  independent of  $\mathfrak{n}$  such that

$$\mathbb{E}[\|u - u_n\|^2] = \mathbb{E}[\|u\|^2 - \|u_n\|^2] \leq C \frac{1}{\mathfrak{n}^3},$$

where  $u_n = \sum_{k=1}^n \frac{\xi_k + g_k}{b + k^2 \pi^2} e_k$ .

**Theorem 6.2.5** (Finite element approximation, strong error). *Let  $u$  be the solution to (6.1.1) and  $u_n^h$  the solution to (6.2.10). Under Assumption 6.2.1, we have the following estimate for piecewise linear finite element approximation of (6.1.1),*

$$\begin{aligned} \mathbb{E}[\|u - u_n^h\|^2] &\leq 2\mathbb{E}[\|u - u_n\|^2] + 2\mathbb{E}[\|u_n - u_n^h\|^2] \\ &\leq C\epsilon^{-1}(n^{-(4-d-\epsilon)} + Mn^{-(2-\frac{d}{2}-\epsilon/2)}) + C(h^4 n^d + Mh^2 n^{d/2}). \end{aligned}$$

When taking  $n = 1/h$ , we have

$$\mathbb{E}[\|u - u_n^h\|^2] \leq C\epsilon^{-1}(h^{4-d-\epsilon} + Mh^{2-\frac{d}{2}-\epsilon/2}). \quad (6.2.13)$$

**Remark 6.2.6.** *The convergence order in Theorem 6.2.5 is optimal as the solution  $u$  to (6.1.1) belongs to  $H^{2-d/2-\epsilon}(\mathcal{D})$  when  $M = 0$  in Assumption 6.2.1, see Theorem 6.4.3. Compared to the methods of finite difference and finite element in [5, 165, 55], the convergence order is half order higher than the convergence order presented in [5] for the one-dimensional problem and are the same for higher dimensional problems [165, 55].*

**Theorem 6.2.7** (Finite element approximation, weak error). *Let  $u$  be the solution to (6.1.1) and  $u_n^h$  the solution to (6.2.10). Under the conditions of Theorem 6.2.3, we have, for  $d = 1, 2$ ,*

$$\mathbb{E}[\|u\|^2 - \|u_n^h\|^2] \leq C\epsilon^{-1}h^{4-d-\epsilon}n^{\epsilon(\kappa+1)(d-1)}.$$

Taking  $n$  at the order of  $1/h$ , we have

$$\mathbb{E}[\|u\|^2 - \|u_n^h\|^2] \leq C\epsilon^{-1}h^{4-d-\epsilon(\kappa+1)(d-1)+1}.$$

Here the constant is similar to that in Theorem 6.2.5 and  $\kappa$  from (6.2.11).

### 6.2.3 Examples of other PDEs

It seems that the fact that the weak convergence order is twice of the strong-order convergence is quite general. In the following example, we show that it is true for an advection equation with multiplicative noises using the spectral approximation (6.2.7).

**Example 6.2.8** (Advection-reaction, see e.g. [367]).

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = \sigma(u-1) \circ \dot{W}(x), \quad x \in [0, L] \quad (6.2.14)$$

with initial condition  $u_0(x)$  and zero inflow. The stochastic product  $u \circ \dot{W}$  is the Stratonovich product. Equation (6.2.14) can be written in Ito's form

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = \frac{\sigma^2}{2}(u-1) + \sigma(u-1) \diamond W(x), \quad x \in [0, L], \quad (6.2.15)$$

where ' $\diamond$ ' represents the Ito-Wick product. The exact solution of (6.2.14) is

$$u = 1 + [u_0(x-t) - 1] \exp[\sigma W(x) - \sigma W(x-t)]. \quad (6.2.16)$$

Applying the truncated spectral expansion (6.2.7) in one-dimensional physical space, we then have the following approximation to Equation (6.2.14):

$$\frac{\partial u_n}{\partial t} + \frac{\partial u_n}{\partial x} = \sigma(u_n-1) \frac{d}{dx} W_n(x), \quad x \in [0, L] \quad (6.2.17)$$

whose solution is

$$u_n = 1 + [u_0(x-t) - 1] \exp[\sigma W_n(x) - \sigma W_n(x-t)]. \quad (6.2.18)$$

**Theorem 6.2.9.** *Let  $u$  be the solution to (6.2.14) and  $u_n$  the solution to (6.2.17). Then we have*

$$\mathbb{E}[|u - u_n|^2] \leq C \frac{1}{n}, \quad |\mathbb{E}[u^k - u_n^k]| \leq C \frac{1}{n}, \quad \forall k > 0, \quad (6.2.19)$$

where  $C$  depends only on  $t$ ,  $x$  and  $\sigma$  in the former inequality and also on  $k$  in the latter.

For one-dimensional advection equations with multiplicative noise, we have the order of  $1/\sqrt{n}$  for strong convergence and  $1/n$  for weak convergence. We do not expect better convergence order as in the case of elliptic equation, where the smoothing of the inverse of Laplacian operator is involved. The following example shows that when better smoothing effects appear for biharmonic equations, the strong convergence order can be even higher than that in the case of the elliptic operators.

**Example 6.2.10** (Linear biharmonic equations with additive noise). *Consider the following linear biharmonic equation with additive noise*

$$\Delta^2 u + bu = g(x) + \frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} W(x), \quad x = (x_1, \dots, x_d) \in \mathcal{D} = [0, 1]^d, \quad (6.2.20)$$

with  $u = 0$  and  $\Delta u = 0$  on  $\partial\mathcal{D}$ ,  $g \in L^2(\mathcal{D})$ . Suppose the operator  $\Delta^2$  has eigenvalues  $\lambda_\alpha$  and eigenfunctions  $e_\alpha$ . Then  $\lambda_\alpha \sim \frac{1}{\alpha_1^4 + \dots + \alpha_d^4}$ . We approximate (6.2.20) by truncating the white noise using the spectral representation (6.2.4):

$$\Delta^2 u_n + bu_n = g(x) + \sum_{|\alpha| \leq n} e_\alpha(x) \xi_\alpha. \quad (6.2.21)$$

Then we have

$$u = \sum_{\alpha} \frac{g_\alpha + \xi_\alpha}{b + \lambda_\alpha} e_\alpha, \quad u_n = \sum_{|\alpha| \leq n} \frac{g_\alpha + \xi_\alpha}{b + \lambda_\alpha} e_\alpha. \quad (6.2.22)$$

Similar to Theorem 6.2.2, we can conclude that

$$\mathbb{E}[\|u\|^2 - \|u_n\|^2] = \mathbb{E}[\|u - u_n\|^2] \leq Cn^{-(8-d-\epsilon)}. \quad (6.2.23)$$

Before ending this section, we remark that our approach can be further extended as follows: 1) the domain  $\mathcal{D}$  can be a bounded domain with a smooth boundary  $\partial\mathcal{D}$ ; 2) the operator  $\Delta$  can be replaced by general self-adjoint, positive-definite, linear operators, say  $\mathcal{A}$ , with compact inverse. For example, one can consider the problem in [55] with bounded convex  $\mathcal{D}$  and  $\mathcal{A} = \Delta$ . We emphasize again that any orthonormal basis in  $L^2(\mathcal{D})$  can be used in the spectral expansion (6.2.5), though it may be convenient to use the eigenfunctions of  $\mathcal{A}$  as a basis when they can be explicitly obtained.

### 6.3 Numerical results

In this section, we present some numerical results of piecewise linear finite element approximation of one- and two- dimensional semilinear elliptic equations (6.1.1) with spatial Brownian motion approximated by its spectral truncation (6.2.7).

To compute the expectations, we use quasi-Monte Carlo sampling for the one-dimensional problem as we can have relative low random dimensions (no more than 40) and Monte Carlo sampling for the two-dimensional problem as higher random dimension is adopted in the spectral approximation (6.2.5) for higher dimensional problems. The experiments were performed using Matlab R2012a on a Macintosh desktop computer with Intel Xeon CPU E5462 (quad-core, 2.80 GHz). A fixed-point iteration method with tolerance  $h^2/100$  was used to solve the nonlinear algebraic equations at each step of the implicit schemes.

**Example 6.3.1** (One-dimensional elliptic).

$$-\Delta u = \frac{1}{2} \sin(u) + \frac{\partial}{\partial x} W(x), \quad x \in \mathcal{D} = (0, 2), \quad (6.3.1)$$

with zero Dirichlet boundary condition.

In this example, we will truncate the Brownian motion as follows  $W(x) = \sum_{k=1}^n \int_{-1}^x m_k(y) dy \xi_k$ , where we use the cosine basis in  $L^2(\mathcal{D})$ :

$$m_1(x) = \frac{1}{\sqrt{|\mathcal{D}|}}, \quad m_k(x) = \sqrt{\frac{2}{|\mathcal{D}|}} \cos\left(\frac{(k-1)\pi}{|\mathcal{D}|} x\right), \quad k \geq 2.$$

Here we employ the number of elements and the number of modes for Brownian motion no more than forty so that we can use a deterministic version of quasi-Monte Carlo integration method for evaluating the expectations. See [222] on practical aspects of quasi-Monte Carlo methods.

The quasi-Monte Carlo sequence we used is a Sobol sequence generated in Matlab using the following command:

```
p=sobolset(n, 'Skip', 1e3, 'Leap', 20); X=net(p,N); X=erfinv(2*X-1)*sqrt(2);
```

where we take the number of quasi-Monte Carlo sample paths  $N = 5 \times 10^7$ .

In this example, the errors are measured in the weak sense:

$$\rho_2^r = \frac{|\mathbb{E}[\|u_{n_1}^{h_1}\|^2] - \mathbb{E}[\|u_{n_2}^{h_2}\|^2]|}{\mathbb{E}[\|u_{n_2}^{h_2}\|^2]}, \quad n_2 > n_1, \quad (6.3.2)$$

where  $\|v\|$  is the  $L^2$  norm in physical space and  $n_i$  is set to be the number of element associated with  $h_i$  ( $n_i = 2/h_i$  in this example), for  $i = 1, 2$ .

We take  $\sigma = 1$  and obtained  $\mathbb{E}[\|u_{40}^{1/20}\|^2]$  by  $N = 5 \times 10^7$  quasi-Monte Carlo sample paths which gives the value 0.2617935 (up to 7 digit).

In Table 6.1, we observe that the convergence order of finite elements error is close to 2.5.

This is also the case for linear problems (with  $u$  in place of  $\sin(u)$ ), where we observed similar convergence order (results are not presented here). We can not observe the optimal convergence  $h^3$  as in Theorem 6.2.5 as we believe that the element size  $h$  is relatively too large to see such a convergence order.

We did not test the convergence order with larger number of finite elements (smaller  $h$ ) and modes for Brownian motion (larger  $n$ ) as the evaluation of expectations is a higher dimensional integration and usually requires randomized/random sampling methods, such as randomized quasi-Monte Carlo simulation or Monte Carlo simulation. To reduce the accompanied statistical errors from the randomized/random sampling methods, a huge of number of sampling paths or an effective variance reduction method, see e.g. [301], should be applied. This issue is out of the scope of the paper and thus is not considered here.

Table 6.1: Convergence of piecewise linear finite element methods for the one-dimensional semilinear problem (6.3.1) with a spectral approximation of white noise (6.2.7).

# element	n	$\rho_2^r$	order	CPU time(s.)
8	8	$6.463 \times 10^{-4}$	–	$8.24 \times 10^2$ <sup>1</sup>
16	16	$1.737 \times 10^{-4}$	$h^{1.90}$	$2.03 \times 10^3$
24	24	$6.495 \times 10^{-5}$	$h^{2.43}$	$3.57 \times 10^3$
32	32	$3.047 \times 10^{-5}$	$h^{2.63}$	$5.07 \times 10^3$

**Example 6.3.2** (Two-dimensional elliptic equation).

$$-\Delta u + \sin(u) = \sigma \frac{\partial^2}{\partial x_1 \partial x_2} W(x), \quad x \in \mathcal{D} = (0, 1) \times (0, 1), \quad (6.3.3)$$

with zero Dirichlet boundary conditions.

In this example, we test the weak convergence of piecewise linear finite element (rectangular element) approximation of (6.3.3) with different noise magnitudes. In simulations we used the Monte Carlo methods with Mersenne twister random generator (seed 100) to compute expectations.

<sup>1</sup>The computational time includes the time of generating quasi-Monte Carlo sequences, which is much smaller than the total computational time.

The errors are measured in the following weak sense:

$$\rho_1^r = \frac{\left| \left| (\mathbb{E}[u_n^h])^2 \right| - \left| (\mathbb{E}[u_{2n}^{h/2}])^2 \right| \right|}{\left| \mathbb{E}[u_{h/2}] \right|^2}, \quad \rho_2^r = \frac{\left| \mathbb{E}[\|u_n^h\|^2] - \mathbb{E}[\mathbb{E}[u_{2n}^{h/2}]^2] \right|}{\mathbb{E}[\|u_{2n}^{h/2}\|^2]}.$$

When  $32 \times 32$  elements are used, we employ  $2 \times 10^5$  Monte Carlo sample paths and obtain

$$\sigma = 0.5, \quad \left| \left| (\mathbb{E}[u_{32}^{\sqrt{2}/32}]^2) \right| \right| = 0.22861 \pm 2.3 \times 10^{-4} \quad \text{and} \quad \mathbb{E}[\|u_{32}^{\sqrt{2}/32}\|^2] = 0.22965 \pm 4.5 \times 10^{-4};$$

$$\sigma = 1, \quad \left| \left| (\mathbb{E}[u_{32}^{\sqrt{2}/32}]^2) \right| \right| = 0.22861 \pm 4.7 \times 10^{-4}, \quad \mathbb{E}[\|u_{32}^{\sqrt{2}/32}\|^2] = 0.23278 \pm 9.0 \times 10^{-4}.$$

The numbers after ‘ $\pm$ ’ are the statistical errors with the 95% confidence interval.

In Table 6.2, we observe a second-order convergence of piecewise approximation (6.2.10) for the two-dimensional semilinear problem (6.3.3), which is consistent with our theoretical prediction in Theorem 6.2.7 for  $\sigma = 0.5$  and  $\sigma = 1$ .

Table 6.2: Convergence of piecewise linear finite element approximation of the two-dimensional semilinear problem (6.3.3) with a spectral approximation of white noise (6.2.7).

$\sigma$	# MC	# element	$\rho_1^r$	order	$\rho_2^r$	order	time(s.)
0.5	$1 \times 10^3$	$4 \times 4$	$1.831 \times 10^{-2} \pm 3.2 \times 10^{-3}$	–	$1.800 \times 10^{-2} \pm 6.5 \times 10^{-3}$	–	0.1
0.5	$4 \times 10^4$	$8 \times 8$	$4.201 \times 10^{-3} \pm 5.3 \times 10^{-4}$	$h^{2.12}$	$4.172 \times 10^{-3} \pm 1.0 \times 10^{-3}$	$h^{2.11}$	2.2
0.5	$8 \times 10^4$	$16 \times 16$	$1.113 \times 10^{-3} \pm 3.7 \times 10^{-4}$	$h^{1.92}$	$1.121 \times 10^{-3} \pm 7.1 \times 10^{-4}$	$h^{1.90}$	80.1
1.0	$2 \times 10^3$	$4 \times 4$	$1.779 \times 10^{-2} \pm 4.5 \times 10^{-3}$	–	$1.662 \times 10^{-2} \pm 9.2 \times 10^{-3}$	–	0.1
1.0	$1 \times 10^5$	$8 \times 8$	$4.281 \times 10^{-3} \pm 6.7 \times 10^{-4}$	$h^{2.05}$	$4.177 \times 10^{-3} \pm 1.3 \times 10^{-3}$	$h^{1.99}$	71.6
1.0	$2 \times 10^5$	$16 \times 16$	$1.231 \times 10^{-3} \pm 4.7 \times 10^{-4}$	$h^{1.80}$	$1.255 \times 10^{-3} \pm 9.0 \times 10^{-4}$	$h^{1.73}$	191.2

## 6.4 Proofs

The eigenvalue problem (6.2.6) admits a nondecreasing sequence  $\{\lambda_m\}_{m=1}^\infty$  of positive eigenvalues and  $\lim_{m \rightarrow \infty} \lambda_m \rightarrow \infty$ . The Green function  $\mathcal{K}(x, y)$  can be represented by

$$\mathcal{K}(x, y) = \sum_{\alpha \in \mathbb{N}^d} \frac{1}{\pi^2 |\alpha|^2} e_\alpha(x) e_\alpha(y), \quad (6.4.1)$$

where  $e_\alpha(x) = \prod_{i=1}^d \sqrt{2} \sin(\pi \alpha_i x_i)$  ( $\alpha_i = 1, 2, \dots$ ) are the orthonormal eigenfunctions of  $-\Delta$ . We will use the single-indexed  $e_k$  and/or multi-indexed  $e_\alpha$  for the eigenfunctions if no confusion arises. For any  $s \in \mathbb{R}$ , we define

$$\dot{H}^s = \dot{H}^s(\mathcal{D}) = \mathcal{D}((-\Delta)^{s/2}) = \left\{ v \mid \|v\|_s = \|(-\Delta)^{s/2} v\| = \left( \sum_{k=1}^{\infty} \lambda_k^s(v, e_k) \right)^{1/2} < \infty \right\}.$$

It is known that  $\dot{H}^s = H^s$ , see e.g. [375].

#### 6.4.1 Regularity of the solution to (6.1.1).

To prove our conclusions, we need some regularity results for (6.1.1).

**Lemma 6.4.1.** *There exists a constant  $C$  depending only on  $d$  that*

$$\int_{\mathcal{D}} |\mathcal{K}(x, y)|^2 dy \leq C, \quad \int_{\mathcal{D}} \|\mathcal{K}(\cdot, y)\|_{2-d/2-\epsilon}^2 dy \leq C. \quad (6.4.2)$$

*Proof.* By (6.4.1) and orthonormality of  $\{e_\alpha\}$ , we have

$$\int_{\mathcal{D}} |\mathcal{K}(x, y)|^2 dy = \sum_{\alpha} \frac{1}{\pi^4 |\alpha|^4} e_{\alpha}^2(x) \leq C \sum_{\alpha} \frac{1}{\pi^2 |\alpha|^4} \leq C.$$

By the fact that  $\lambda_{\alpha} \leq C |\alpha|^2$ , we then have

$$\int_{\mathcal{D}} \|\mathcal{K}(\cdot, y)\|_{2-d/2-\epsilon}^2 dy = \sum_{\alpha} \lambda_{\alpha}^{2-d/2-\epsilon} \frac{1}{\pi^4 |\alpha|^4} \leq C \sum_{\alpha} \frac{1}{|\alpha|^{d+\epsilon}} \leq C \epsilon^{-1},$$

where we use the fact that the series  $\sum_{\alpha \in \mathbb{N}^d} \frac{1}{|\alpha|^s}$  converges if and only if  $s > d + \epsilon$  with  $\epsilon > 0$ .  $\square$

**Lemma 6.4.2.** *For any  $\epsilon > 0$ , we have*

$$\mathbb{E} \left[ \left\| \frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} W_n \right\|_{\dot{H}^{-d/2-\epsilon}}^2 \right] \leq \mathbb{E} \left[ \left\| \frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} W \right\|_{\dot{H}^{-d/2-\epsilon}}^2 \right] < C(d) \epsilon^{-1}, \quad (6.4.3)$$

*Proof.* By the definition of norms in  $\dot{H}^{-\beta}$ , where  $\beta = d/2 + \epsilon$ , we have

$$\mathbb{E}\left[\left\|\frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} W_n\right\|_{\dot{H}^{-\beta}}^2\right] \leq \mathbb{E}\left[\left\|\frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} W\right\|_{\dot{H}^{-\beta}}^2\right] = \sum_{\alpha \in \mathbb{N}^d} \lambda_\alpha^{-\beta} = \frac{1}{\pi^{2\beta}} \sum_{\alpha \in \mathbb{N}^d} \frac{1}{|\alpha|^{2\beta}}.$$

Then we have for  $2\beta > d + \epsilon$ ,  $\left\|\frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} W\right\|_{\dot{H}^{-\beta}}^2 \leq C(d)\epsilon^{-1} < \infty$ .  $\square$

From Lemmas 6.4.1 and 6.4.2, we have the following regularity results for (6.1.1).

**Theorem 6.4.3.** *Under Assumption 6.2.1, we have the following regularity for the solution to (6.1.1):*

$$\mathbb{E}[\|u\|^p] < \infty, \quad \mathbb{E}[\|u\|_{L^p}^2] < \infty, \quad 2 \leq p < \infty, \quad (6.4.4)$$

and furthermore, if  $M = 0$  in Assumption 6.2.1 (to be precise, (6.2.2)),

$$\mathbb{E}[\|u\|_{2-d/2-\epsilon}^2] < C\epsilon^{-1}. \quad (6.4.5)$$

*Proof.* We first establish the  $L^p$ -stability for Equation (6.1.1) for  $p > 2$  since the  $L^2$  regularity can be found in [165]. By (6.2.3) and (6.2.2), taking  $L^p$ -norm over both sides, we then have

$$\mathbb{E}[\|u\|_{L^p}^2] \leq C\mathbb{E}\left[\left\|\int_{\mathcal{D}} |\mathcal{K}(\cdot, y)(1 + |u|) dy\right\|_{L^p}\right]^2 + C\mathbb{E}\left[\left\|\sum_{|\alpha|=1}^{\infty} \frac{1}{\pi^2 |\alpha|^2} e_\alpha(x) \xi_\alpha\right\|_{L^p}\right]^2 \quad (6.4.6)$$

By the Cauchy-Schwarz inequality and by Lemma 6.4.1, we have

$$\mathbb{E}\left[\left\|\int_{\mathcal{D}} |\mathcal{K}(\cdot, y)u| dy\right\|_{L^p}^2\right] \leq \left\|\int_{\mathcal{D}} \mathcal{K}^2(\cdot, y) dy\right\|_{L^p}^2 \mathbb{E}[\|u\|^2] \leq C\mathbb{E}[\|u\|^2].$$

Then by  $\mathbb{E}\left[\left\|\sum_{|\alpha|=1}^{\infty} \frac{1}{\pi^2 |\alpha|^2} e_\alpha(x) \xi_\alpha\right\|_{L^p}^2\right] < \infty$  and (6.4.6), we have the second inequality in (6.4.4). With Lemma 6.4.2, the estimate (6.4.5) can be proved similarly.  $\square$

## 6.4.2 Proofs for strong convergence order

**Lemma 6.4.4.** For  $\eta = \int_{\mathcal{D}} \mathcal{K}(x, y) d[W(y) - W_n(y)]$ , we have

$$\mathbb{E}[\|\eta\|^2] \leq C(d)\epsilon^{-1}(n+1)^{-(4-d-\epsilon)}. \quad (6.4.7)$$

*Proof.* By (6.4.1), (6.2.5) and (6.2.7), we have

$$\begin{aligned} \mathbb{E}[\|\eta\|^2] &= \mathbb{E}\left[\left(\int_{\mathcal{D}} \mathcal{K}(x, y) d[W(y) - W_n(y)]\right)^2\right] \\ &= \sum_{|\alpha| \geq n+1} \frac{1}{\pi^4 |\alpha|^4} \leq (n+1)^{-(4-d-\epsilon)} \sum_{|\alpha| \geq n+1} \frac{1}{\pi^4 |\alpha|^{d+\epsilon}} \\ &\leq C(d)\epsilon^{-1} n^{-(4-d-\epsilon)}, \end{aligned}$$

where we use the fact that the series  $\sum_{\alpha \in N^d} \frac{1}{|\alpha|^s}$  converges if and only if  $s > d + \epsilon$  with  $\epsilon > 0$ .  $\square$

**Lemma 6.4.5.** Under Assumption 6.2.1, then we have

$$\mathbb{E}[\|u - u_n\|^2] \leq C(M \left(\mathbb{E}[\|\eta\|^2]\right)^{\frac{1}{2}} + (C_p + R)^2 \mathbb{E}[\|\eta\|^2]).$$

where the constant  $C$  depends only on  $C_p$ ,  $L$  and  $M, R$ .

The proof of Lemma 6.4.5 is similar to that of Theorem 2.3 in [165]. We present the proof here for completeness.

*Proof.* By (6.2.3) and (6.2.8), the error equation reads

$$u(x) - u_n(x) = - \int_{\mathcal{D}} \mathcal{K}(x, y) [f(u(y)) - f(u_n(y))] dy + \eta(x). \quad (6.4.8)$$

Multiplying  $[f(u(y)) - f(u_n(y))]$  over both sides of (6.1.1), applying the inequality (6.2.1) and

integrating over the domain  $\mathcal{D}$ , we have

$$\begin{aligned}
-L \|u - u_n\|^2 &\leq \int_{\mathcal{D}} \eta(x)[f(u(x)) - f(u_n(x))] dx \\
&\quad - \int_{\mathcal{D}} \int_{\mathcal{D}} \mathcal{K}(x, y)[f(u(y)) - f(u_n(y))] dy [f(u(x)) - f(u_n(x))] dx \\
&\leq \int_{\mathcal{D}} \eta(x)[f(u(x)) - f(u_n(x))] dx - C_p \int_{\mathcal{D}} \left( \int_{\mathcal{D}} \mathcal{K}(x, y)[f(u(y)) - f(u_n(y))] dy \right)^2 dx \\
&= \int_{\mathcal{D}} \eta(x)[f(u(x)) - f(u_n(x))] dx - C_p \int_{\mathcal{D}} [u - u_n - \eta]^2 dx.
\end{aligned}$$

Then, by (6.2.2) and the Cauchy-Schwarz inequality, we have

$$(C_p - L) \|u - u_n\|^2 \leq (C_p + R) \|u - u_n\| \|\eta\| + 2M \|\eta\|.$$

By Assumption 6.2.1 ( $C_p - L > 0$ ), we have

$$\|u - u_n\|^2 \leq C(M \|\eta\| + (C_p + R)^2 \|\eta\|^2),$$

where the constant  $C$  depends only on  $C_p$ ,  $L$  and  $M$ ,  $R$ . □

Theorem 6.2.2 follows from the triangle inequality and Lemmas 6.4.4 and 6.4.5.

### 6.4.3 Proofs for weak convergence order

To prove Theorem 6.2.3, we need the following lemmas. In the proofs, we will use single-indexed eigenvalues  $\lambda_i$  and eigenfunctions  $e_i(x)$ . We introduce the following equation

$$-\Delta u(x) + f(u(x)) = g(x) + \sum_{i=1}^{\infty} e_i(x) y_i. \quad (6.4.9)$$

**Lemma 6.4.6.** *In addition to Assumption 6.2.1, assume also that  $f$  satisfies the polynomial growth*

condition (6.2.11). Then there exists a constant  $C > 0$  depending only on  $d, \kappa$  and  $\beta$  that

$$\mathbb{E}[\|D^\beta u\|_{L^q}^2] \leq C \prod_i \lambda_i^{-2\beta_i}, \text{ for } 1 \leq |\beta| \leq 4, \quad 2 \leq q \leq \infty.$$

**Lemma 6.4.7.** *Suppose that  $F$  satisfies the polynomial growth condition (6.2.11). Under the conditions of Lemma 6.4.6, we then have for some constant  $C > 0$  depending only on  $d, \kappa$  and  $\beta$ ,*

$$\mathbb{E}[\|D^\beta(F(u))\|^2] \leq C \prod_i \lambda_i^{-2\beta_i}, \text{ for } |\beta| \leq 4.$$

**Proof of Lemma 6.4.6.** To estimate the derivatives of solution with respect to parameters, we need the following auxiliary equation: for  $\tilde{g} \in L^2(\mathcal{D})$ ,

$$-\Delta v + f'(u)v = \tilde{g}(x), \quad x \in \mathcal{D}, \quad v = 0 \text{ on } \partial\mathcal{D}. \quad (6.4.10)$$

By Assumption 6.2.1, we claim the following estimate

$$\|v\|_{L^q} \leq c \left\| \int_{\mathcal{D}} \mathcal{K}(x, y) \tilde{g}(y) dy \right\|_{L^\infty}, \quad \forall 2 \leq q \leq \infty. \quad (6.4.11)$$

We first establish the case  $q = 2$ . Equation (6.4.10) can be written in the integral form as

$$v(x) + \int_{\mathcal{D}} \mathcal{K}(x, y) f'(u)v dy = \int_{\mathcal{D}} \mathcal{K}(x, y) \tilde{g}(y) dy. \quad (6.4.12)$$

Multiplying  $f'(u)v$  over both sides and by the Poincare inequality, (6.4.12) and (6.4.10), we have

$$\begin{aligned} 0 &= (f'(u)v, v) + \left( \int_{\mathcal{D}} \mathcal{K}(\cdot, y) f'(u(y))v(y) dy, f'(u)v \right) - \left( \int_{\mathcal{D}} \mathcal{K}(\cdot, y) \tilde{g}(y) dy, f'(u)v \right) \\ &\geq (f'(u)v, v) + C_p \left\| \int_{\mathcal{D}} \mathcal{K}(\cdot, y) f'(u(y))v(y) dy \right\|^2 - \left( \int_{\mathcal{D}} \mathcal{K}(\cdot, y) \tilde{g}(y) dy, f'(u)v \right) \\ &\geq -L \|v\|^2 + C_p \left\| v - \int_{\mathcal{D}} \mathcal{K}(\cdot, y) \tilde{g}(y) dy \right\|^2 - \left( \int_{\mathcal{D}} \mathcal{K}(\cdot, y) \tilde{g}(y) dy, f'(u)v \right). \end{aligned}$$

Then, by the fact that  $f' \geq -L > -C_p$  and  $|f'| \leq R$ , we have (6.4.11) when  $q = 2$ . After taking  $L^q$ -norm over both side of (6.4.12) and by  $\int_{\mathcal{D}} \mathcal{K}^2(x, y) dy \leq C$  (Lemma 6.4.1), we have

$$\|v\|_{L^q} \leq RC \|v\| + \left\| \int_{\mathcal{D}} \mathcal{K}(\cdot, y) \tilde{g}(y) dy \right\|_{L^q},$$

and thus by Lemma 6.4.1, we reach (6.4.11).

Taking the derivative with respect to  $y_i$  in Equation (6.4.9), we have

$$-\Delta D^{\epsilon_i} u + f'(u) D^{\epsilon_i} u = e_i(x).$$

Thus, by (6.4.11) and (6.4.1) , we have

$$\|D^{\epsilon_i} u\|_{L^q} \leq c \lambda_i^{-1}. \quad (6.4.13)$$

Taking the derivatives with respect to  $y_i$  and  $y_j$  in Equation (6.4.9), we have the following equation:

$$-\Delta D^{\epsilon_i + \epsilon_j} u + f'(u) D^{\epsilon_i + \epsilon_j} u = -f''(u) D^{\epsilon_i} u D^{\epsilon_j} u.$$

Then by (6.4.11) and (6.4.13), we have

$$\|D^{\epsilon_i + \epsilon_j} u\|_{L^q} \leq c \lambda_i^{-1} \lambda_j^{-1} \|f''(u)\|, \quad (6.4.14)$$

where we assume that  $f''(x)$  is of at most polynomial growth. Similarly, we have

$$\begin{aligned} \|D^{\epsilon_i + \epsilon_j + \epsilon_k} u\|_{L^q} &\leq c \lambda_i^{-1} \lambda_j^{-1} \lambda_k^{-1} (\|f^{(3)}(u)\| + \|f''(u)\|), \\ \|D^{\epsilon_i + \epsilon_j + \epsilon_k + \epsilon_l} u\|_{L^q} &\leq c \lambda_i^{-1} \lambda_j^{-1} \lambda_k^{-1} \lambda_l^{-1} (\|f^{(4)}(u)\| + \|f^{(3)}(u)\| + \|f''(u)\|). \end{aligned}$$

By the assumption of polynomial growth at infinity for  $f$  and its derivatives and the  $L^p$ -stability

(6.4.4), we reach the conclusion.  $\square$

**Proof of Lemma 6.4.7.** By the multivariate chain rule (also known as multivariate Faa di Bruno formula) and Lemma 6.4.6, we have  $D^{\epsilon_i + \epsilon_j} F(u) = F'(u) D^{\epsilon_i + \epsilon_j} u + F''(u) D^{\epsilon_i} u D^{\epsilon_j} u$ , and thus by (6.4.13)

$$\|D^{\epsilon_i + \epsilon_j} F(u)\|^2 \leq c(\|F'(u)\|^2 + \|F''(u)\|^2) \lambda_i^{-2} \lambda_j^{-2}, \quad (6.4.15)$$

and similarly,

$$\|D^{2\epsilon_i + 2\epsilon_j} F(u)\|^2 \leq c(\|F'(u)\|^2 + \|F''(u)\|^2 + \|F^{(3)}(u)\|^2 + \|F^{(4)}(u)\|^2) \lambda_i^{-4} \lambda_j^{-4}. \quad (6.4.16)$$

The conclusion then follows from the assumption of polynomial growth of  $F$  and its derivatives at infinity (6.2.11) and Lemma 6.4.6.  $\square$

**Proof of Theorem 6.2.3.** By the first-order Taylor's expansion, we have for  $m > n$

$$\begin{aligned} & \mathbb{E}[F(u_m) - F(u_n)] \\ &= \mathbb{E}[F(u_m(\xi_1, \dots, \xi_{n'}, \dots, \xi_{m'})) - F(u_n(\xi_1, \dots, \xi_n))] \\ &= \mathbb{E}\left[\sum_{i=n'+1}^{m'} D^{\epsilon_i} (F(u_m(\xi_1, \dots, \xi_{n'}, 0, \dots, 0))) \xi_i\right] \\ &+ \sum_{i,j=n'+1}^{m'} \mathbb{E}\left[\int_0^1 (1-t) D^{\epsilon_i + \epsilon_j} (F(u_m(\xi_1, \dots, \xi_{n'}, t\xi_{n'+1}, \dots, t\xi_{m'}))) \xi_i \xi_j\right] \\ &= \sum_{i,j=n'+1}^m \int_0^1 (1-t) \mathbb{E}[D^{\epsilon_i + \epsilon_j} (F(u_m(\xi_1, \dots, \xi_{n'}, t\xi_{n'+1}, \dots, t\xi_{m'}))) \xi_i \xi_j], \quad (6.4.17) \end{aligned}$$

where we used the fact  $\xi_i$  ( $i \geq n'+1$ ) is independent of  $F(u_m(\xi_1, \dots, \xi_{n'}, 0, \dots, 0))$  and  $\mathbb{E}[\xi_i] = 0$  and we defined  $n' = n!/d!/(n-d)!$ . We also denoted by  $D^{\epsilon_i} \phi(y_1, y_2, \dots, y_{m'})$  the first-order derivatives of  $\phi$  in the  $y_i$ -direction (replacing  $\xi_i$  with  $y_i$ );  $\epsilon_i$  is a vector of  $m'$  dimension taking values 1 at the  $i$ -th element and 0 otherwise. More generally,  $D^\alpha$  represents multivariate derivatives with respect to the parameters  $y$ .

To estimate (6.4.17), we split the term into two parts:

$$\begin{aligned}
I &= \sum_{i=n'+1}^{m'} \int_0^1 (1-t) \mathbb{E}[D^{2\epsilon_i}(F(u_m(\xi_1, \dots, \xi_{n'}, t\xi_{n'+1}, \dots, t\xi_{m'})))\xi_i^2], \\
II &= 2 \sum_{i < j, i, j = n'+1}^{m'} \int_0^1 (1-t) \mathbb{E}[D^{\epsilon_i + \epsilon_j}(F(u_m(\xi_1, \dots, \xi_{n'}, t\xi_{n'+1}, \dots, t\xi_{m'})))\xi_i \xi_j].
\end{aligned}$$

By Lemma 6.4.7, we have

$$\begin{aligned}
\|I\| &= \left\| \int_0^1 (1-t) \mathbb{E}[D^{2\epsilon_i}(F(u_m(\xi_1, \dots, \xi_{n'}, t\xi_{n'+1}, \dots, t\xi_{m'})))\xi_i^2] dt \right\| \\
&\leq \int_0^1 (1-t) \mathbb{E}[\|D^{2\epsilon_i}(F(u_m(\xi_1, \dots, \xi_{n'}, t\xi_{n'+1}, \dots, t\xi_{m'})))\| \xi_i^2] dt \leq c\lambda_i^{-2}. \quad (6.4.18)
\end{aligned}$$

For II, we use the recipe of the proof of Theorem 2.8 in [65]. For simplicity, we define that

$$X_{i,j}^{t,r,s} = (\xi_1, \dots, \xi_{n'}, t\xi_{n'+1}, \dots, tr\xi_i, \dots, ts\xi_j, \dots, t\xi_{m'}).$$

Noticing that  $\mathbb{E}[D^{\epsilon_i + \epsilon_j}(F(u_m(X_{i,j}^{t,0,1}))\xi_i \xi_j) = 0$  ( $i < j$ ), we have

$$\begin{aligned}
&\int_0^1 (1-t) \mathbb{E}[D^{\epsilon_i + \epsilon_j}(F(u_m(X_{i,j}^{t,1,1}))\xi_i \xi_j] dt \\
&= \int_0^1 (1-t) \mathbb{E}[D^{\epsilon_i + \epsilon_j}(F(u_m(X_{i,j}^{t,1,1}))\xi_i \xi_j] dt - \int_0^1 (1-t) \mathbb{E}[D^{\epsilon_i + \epsilon_j}(F(u_m(X_{i,j}^{t,0,1}))\xi_i \xi_j] dt \\
&= \int_0^1 \int_0^1 (1-t)t \mathbb{E}[D^{2\epsilon_i + \epsilon_j}(F(u_m(X_{i,j}^{t,r,1}))\xi_i \xi_j] dt dr.
\end{aligned}$$

With  $\mathbb{E}[D^{2\epsilon_i + \epsilon_j}(F(u_m(X_{i,j}^{t,r,0}))\xi_i \xi_j) = 0$  ( $i < j$ ), we have similarly

$$\begin{aligned}
&\int_0^1 \int_0^1 (1-t)t \mathbb{E}[D^{2\epsilon_i + \epsilon_j}(F(u_m(X_{i,j}^{t,r,1}))\xi_i \xi_j] dt dr, \\
&= \int_0^1 \int_0^1 \int_0^1 (1-t)t^2 \mathbb{E}[D^{2\epsilon_i + 2\epsilon_j}(F(u_m(X_{i,j}^{t,r,s}))\xi_i^2 \xi_j^2] dt dr ds,
\end{aligned}$$

and thus for  $i < j$

$$\int_0^1 (1-t) \mathbb{E}[D^{\epsilon_i + \epsilon_j}(F(u_m(X_{i,j}^{t,1,1}))\xi_i \xi_j] dt$$

$$= \int_0^1 \int_0^1 \int_0^1 (1-t)t^2 \mathbb{E}[D^{2\epsilon_i+2\epsilon_j}(F(u_m(X_{i,j}^{t,r,s}))\xi_i^2\xi_j^2)] dt dr ds.$$

Now we can bound  $II$  as with Lemma 6.4.7,

$$\|II\| \leq \left\| \int_0^1 (1-t) \mathbb{E}[D^{2\epsilon_i+2\epsilon_j}(F(u_m(X_{i,j}^{t,1,1}))\xi_i\xi_j)] dt \right\| \leq c\lambda_i^{-2}\lambda_j^{-2}. \quad (6.4.19)$$

Thus, we have by (6.4.17), (6.4.18), and (6.4.19),

$$\|\mathbb{E}[F(u_m) - F(u_n)]\| \leq c \sum_{|\alpha|=n+1}^m \lambda_\alpha^{-2} + c \left( \sum_{|\alpha|=n+1}^m \lambda_\alpha^{-2} \right)^2. \quad (6.4.20)$$

Then by  $\lambda_\alpha^{-1} \leq C|\alpha|^{-2}$ , we arrive at the conclusion.  $\square$

#### 6.4.4 Proof of Theorem 6.2.9

We first prove the strong convergence. Note that

$$\mathbb{E}[(u - u_n)^2] = (u_0(x-t) - 1)^2 \mathbb{E}[(\exp(\sigma W(x) - \sigma W(x-t)) - \exp(\sigma W_n(x) - \sigma W_n(x-t)))^2]. \quad (6.4.21)$$

By the fact that  $\exp(a) - \exp(b) = \exp(\theta a + (1-\theta)b)(a-b)$  where  $a \leq \theta \leq b$  and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \mathbb{E}[(u - u_n)^2] &= (u_0(x-t) - 1)^2 \mathbb{E}[(\exp(\sigma W(x) - \sigma W(x-t)) - \exp(\sigma W_n(x) - \sigma W_n(x-t)))^2] \\ &\leq (u_0(x-t) - 1)^2 (\mathbb{E}[(\exp(4\sigma\theta(W(x) - W(x-t)) + 4(1-\theta)\sigma(W_n(x) - \sigma W_n(x-t)))^{1/2} \\ &\quad \times \sigma^2 (\mathbb{E}[(W(x) - W(x-t)) - (W_n(x) - W_n(x-t))]^4)]^{1/2}). \end{aligned}$$

It requires to estimate the two expectations in the above inequality.

$$(\mathbb{E}[(\exp(4\sigma\theta(W(x) - W(x-t)) + 4(1-\theta)\sigma(W_n(x) - \sigma W_n(x-t)))^{1/2}$$

$$\begin{aligned}
&\leq (\mathbb{E}[(\exp(8\sigma\theta(W(x) - W(x-t))))^{1/4}] (\mathbb{E}[(\exp(8(1-\theta)\sigma(W_n(x) - W_n(x-t))))^{1/4}] \\
&\leq \exp(2\sigma^2\theta t) (\mathbb{E}[(\exp(4(1-\theta)\sigma(W_n(x) - W_n(x-t))))^{1/4}] \\
&\leq \exp(8\sigma^2\theta^2 t) \exp(8(1-\theta)^2\sigma^2 t) \leq \exp(8\sigma^2 t).
\end{aligned} \tag{6.4.22}$$

Now we estimate the  $\mathbb{E}[(W(x) - W(x-t) - (W_n(x) - W_n(x-t)))^4]$ .

$$\begin{aligned}
&\mathbb{E}[(W(x) - W(x-t) - (W_n(x) - W_n(x-t)))^4] \\
&= \mathbb{E}[(\sum_{k=n+1}^{\infty} [M_k(x) - M_k(x-t)]\xi_k)^4] \\
&= \mathbb{E}[(\sum_{k=n+1}^{\infty} \sum_{l=n+1}^{\infty} [M_k(x) - M_k(x-t)]^2 [M_l(x) - M_l(x-t)]^2 \xi_k^2 \xi_l^2)] \\
&\leq 3 \sum_{k=n+1}^{\infty} \sum_{l=n+1}^{\infty} [M_k(x) - M_k(x-t)]^2 [M_l(x) - M_l(x-t)]^2 \\
&= 3(\sum_{k=n+1}^{\infty} [M_k(x) - M_k(x-t)]^2)^2 \leq c \frac{1}{n},
\end{aligned} \tag{6.4.23}$$

where  $M_k = \int_0^x m_k(y) dy$  with  $m_1(x) = 1/\sqrt{L}$ ,  $m_k(x) = 2/\sqrt{L} \cos(\pi(k-1)t/L)$  and  $c$  depends only on  $t$  and  $x$ .

By (6.4.22) and (6.4.23), we have the first estimate in (6.2.19).

Now we prove the weak convergence. It suffices to check  $\mathbb{E}[(u-1)^k] - \mathbb{E}[(u_n-1)^k]$ . By (6.2.16) and (6.2.18), we have

$$\begin{aligned}
|\mathbb{E}[(u-1)^k] - \mathbb{E}[(u_n-1)^k]| &= |(u_0(x-t)-1)^k \exp(\frac{k^2}{2}\sigma^2\mathbb{E}[(W(x)-W(x-t))^2]) \\
&\quad - (u_0(x-t)-1)^k \exp(\frac{k^2}{2}\sigma^2\mathbb{E}[(W_n(x)-W_n(x-t))^2])| \\
&\leq |(u_0(x-t)-1)^k| \exp(\frac{k^2}{2}\sigma^2\mathbb{E}[(W(x)-W(x-t))^2]) \\
&\quad \times \frac{k^2}{2}\sigma^2(\mathbb{E}[(W(x)-W(x-t))^2] - \mathbb{E}[(W_n(x)-W_n(x-t))^2]),
\end{aligned}$$

where we have used the fact  $e^x - e^y = e^{\theta x + (1-\theta)y}(x-y)$  ( $0 \leq \theta \leq 1$ ) and that  $\mathbb{E}[(W_n(x) -$

$W_n(x-t)^2] \leq \mathbb{E}[(W(x) - W(x-t))^2]$ . By  $\mathbb{E}[(W(x) - W(x-t))^2] - \mathbb{E}[(W_n(x) - W_n(x-t))^2] = \sum_{k=n+1}^{\infty} (M_k(x) - M_k(x-t))^2$ , we then have

$$|\mathbb{E}[(u-1)^k] - \mathbb{E}[(u_n-1)^k]| \leq \frac{k^2}{2} \sigma^2 |(u_0(x-t)-1)^k| \exp\left(\frac{k^2}{2} \sigma^2 t\right) \frac{c}{n}.$$

Hence, the estimate of the weak convergence order follows.  $\square$

### 6.4.5 Proof of Theorem 6.2.5

Define the Ritz projection  $R_h : H_0^1(\mathcal{D}) \rightarrow V_h$  by

$$(\nabla R_h w, \nabla v) = (\nabla w, \nabla v), \quad \forall v \in V_h, \quad w \in H_0^1(\mathcal{D}).$$

Then it holds that, see e.g. [375], there is a constant  $C$  independent of  $h$  such that for  $0 \leq l < r \leq 2$

$$\|w - R_h w\|_l \leq Ch^{r-l} \|w\|_r, \quad w \in H^2(\mathcal{D}) \cap H_0^1(\mathcal{D}), \quad (6.4.24)$$

*Proof.* It can be readily checked from (6.2.9) and (6.2.10) that

$$(\nabla(R_h u_n - u_n^h), \nabla v) + (f(u_n) - f(u_n^h), v) = 0, \quad v \in V_h. \quad (6.4.25)$$

Taking  $v = R_h u_n - u_n^h$  and by (2.1) and (2.3), the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \|\nabla(R_h u_n - u_n^h)\|^2 &= -(f(u_n) - f(u_n^h), u_n - u_n^h) + (f(u_n) - f(u_n^h), R_h u_n - u_n) \\ &\leq L \|u_n - u_n^h\|^2 + c(M + R \|u_n - u_n^h\|) \|R_h u_n - u_n\| \\ &\leq \frac{L + C_p}{2} \|u_n - u_n^h\|^2 + CM \|R_h u_n - u_n\| + C \|R_h u_n - u_n\|^2. \end{aligned}$$

Then by the Poincare inequality  $\|\nabla(R_h u_n - u_n^h)\|^2 \geq C_p \|R_h u_n - u_n^h\|^2$ , the triangle inequality and

$L < C_p$ , there exists a constant  $C$  independent of  $h$  but dependent of  $C_p, R, L$ :

$$\|R_h u_n - u_n^h\|_1^2 + \|u_n - u_n^h\|^2 \leq C(M \|R_h u_n - u_n\| + \|R_h u_n - u_n\|^2). \quad (6.4.26)$$

Then by (6.4.24), we have

$$\|R_h u_n - u_n^h\|_1^2 + \|u_n - u_n^h\|^2 \leq C(Mh^2 \|u_n\|_2 + h^4 \|u_n\|_2^2). \quad (6.4.27)$$

Similar to the proof of Theorem 6.4.3, we have

$$\mathbb{E}[\|u_n\|_2^2] \leq C\mathbb{E}\left[\left\|g + \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W_n\right\|^2\right] \leq Cn^d, \quad (6.4.28)$$

where we have used the fact that

$$\mathbb{E}\left[\left\|\frac{\partial^d}{\partial x_1 \cdots \partial x_d} W_n\right\|^2\right] = \sum_{|\alpha| \leq n} \|e_\alpha\|^2 \leq Cn^d.$$

By (6.4.26) and (6.4.28), we obtain that

$$\|u_n - u_n^h\|^2 \leq C(Mh^2 \|u_n\|_2 + h^4 \|u_n\|_2^2),$$

whence we can reach the conclusion by setting  $h = n^{-1}$ . □

### 6.4.6 Proof of Theorem 6.2.7

By Theorem 6.2.3, we have

$$\mathbb{E}[\|u\|^2 - \|u_n\|^2] \leq Cn^{-(4-d-\epsilon)}.$$

By the standard estimate of the Ritz operator in negative norms, (see e.g. [375, Theorem 5.1]),

$$\|u_n - R_h u_n\|_{-r} \leq Ch^{q+r} \|u_n\|_q, \quad 1 \leq q \leq s, \quad 0 \leq r \leq s-2 \quad (6.4.29)$$

we have, by taking  $q = r = 2$ ,

$$\mathbb{E}[\|u_n\|^2 - \|R_h u_n\|^2] \leq \mathbb{E}[\|u_n - R_h u_n\|_{-2} \|u_n + R_h u_n\|_2] \leq Ch^4 (\mathbb{E}[\|u_n\|_2^2]).$$

Then by (6.4.28),  $\mathbb{E}[\|u_n\|^2 - \|R_h u_n\|^2] \leq Ch^4 n^d$ , and thus we have by Theorem 6.2.3

$$\mathbb{E}[\|u\|^2] - \mathbb{E}[\|R_h u_n\|^2] = \mathbb{E}[\|u\|^2 - \|u_n\|^2] + \mathbb{E}[\|u_n\|^2 - \|R_h u_n\|^2] \leq Cn^{-(4-d-\epsilon)} + Ch^4 n^d. \quad (6.4.30)$$

By the triangle inequality, we need to estimate  $\mathbb{E}[\|u_n^h\|^2] - \mathbb{E}[\|R_h u_n\|^2]$ . To this end, we introduce the following linear adjoint problem over the domain  $\mathcal{D}$ :

$$-\Delta\psi + f'(u_n)\psi = \phi, \quad \psi|_{\partial\mathcal{D}} = 0. \quad (6.4.31)$$

Then we have  $\|\psi\|_2 \leq C\|\phi\|$  since  $f'(u_n) \geq -L > -C_p$  is bounded. Introducing  $e = R_h u_n - u_n^h$ ,  $e_1 = u_n - u_n^h$ ,  $e_2 = R_h u_n - u_n$ , we then have,

$$\begin{aligned} (e, \phi) &= (\nabla e, \nabla\psi) + (f'(u_n)e, \psi) = (\nabla e, \nabla R_h\psi) + (f'(u_n)e, \psi) \\ &= (f'(u_n)e, \psi - R_h\psi) - (f(u_n) - f(u_n^h) - f'(u_n)e, R_h\psi), \\ &= (f'(u_n)e_1, \psi - R_h\psi) - (f(u_n) - f(u_n^h) - f'(u_n)e_1, R_h\psi - \psi) \\ &\quad - (f(u_n) - f(u_n^h) - f'(u_n)e_1, \psi) + (f'(u_n)e_2, \psi), \end{aligned}$$

where we have used the error equation (6.4.25) and the definition of the Ritz projection. Thus we

have, by (6.4.24),  $|f'(u_n)| \leq R$  and Taylor's expansion,

$$\begin{aligned} |(e, \phi)| &\leq C \|e_1\| h^2 \|\psi\|_2 - \frac{1}{2} (f''(\theta u_n + (1-\theta)u_n^h) e_1^2, \psi) + (f'(u_n) e_2, \psi) \\ &\leq C \|e_1\| h^2 \|\psi\|_2 + \|e_1\|^2 \|\psi\|_\infty (1 + \|u_n\|_{L^\infty}^\kappa + \|u_n^h\|_{L^\infty}^\kappa) + C \|e_2\|_{-2} \|\psi\|_2, \end{aligned} \quad (6.4.32)$$

where we used the polynomial growth condition (6.2.11) for  $f''$ . Then we have, by the embedding  $\|v\|_\infty \leq C \|v\|_2$ , (6.4.32), and  $\|\psi\|_2 \leq C \|\phi\|$  that

$$|(e, \phi)| \leq C h^2 (\|e_1\| + \|e_1\|^2 (1 + \|u_n\|_{L^\infty}^\kappa + \|u_n^h\|_{L^\infty}^\kappa) + \|e_2\|_{-2}) \|\phi\|.$$

Thus by the definition of negative norm, (6.4.29) and the Hölder inequality, we have

$$\begin{aligned} \mathbb{E}[\|e\|_{-1}^2] &\leq C \mathbb{E}[(\|e_1\|^2 + \|e_2\|^2) h^4 + \mathbb{E}[\|e_1\|^4 (1 + \|u_n\|_{L^\infty}^\kappa + \|u_n^h\|_{L^\infty}^\kappa)^2]] \\ &\leq C \mathbb{E}[(\|e_1\|^2 + \|e_2\|^2) h^4 + \mathbb{E}[\|e_1\|^4]] \\ &\quad + C (\mathbb{E}[\|e_1\|^{4(1+\epsilon)}])^{1/(1+\epsilon)} \left( (\mathbb{E}[\|u_n\|_{L^\infty}^{2\kappa(1+1/\epsilon)}])^{\epsilon/(1+\epsilon)} + (\mathbb{E}[\|u_n^h\|_{L^\infty}^{2\kappa(1+1/\epsilon)}])^{\epsilon/(1+\epsilon)} \right) \end{aligned} \quad (6.4.33)$$

Similar to the proofs of Lemmas 6.4.6 and 6.4.7, we have

$$\begin{aligned} (\mathbb{E}[\|u_n\|_{L^\infty}^{2\kappa(1+1/\epsilon)}])^{\epsilon/(1+\epsilon)} + (\mathbb{E}[\|u_n^h\|_{L^\infty}^{2\kappa(1+1/\epsilon)}])^{\epsilon/(1+\epsilon)} &\leq C \mathbf{n}^{\epsilon\kappa(d-1)}, \quad d = 1, 2, \\ \mathbb{E}[\|u_n\|_2^p] &\leq C \mathbf{n}^{pd/2}, \quad p \geq 2. \end{aligned} \quad (6.4.34)$$

Then by (6.4.34), (6.4.33) and (6.4.26), we have, for  $d = 1, 2$

$$\begin{aligned} (\mathbb{E}[\|e\|_{-1}^2])^{1/2} &\leq C h^2 (\mathbb{E}[\|e_1\|^2] + \mathbb{E}[\|e_2\|^2])^{1/2} + (\mathbb{E}[\|e_1\|^4])^{1/2} + C (\mathbb{E}[\|e_1\|^{4(1+\epsilon)}])^{1/(1+\epsilon)} \mathbf{n}^{\epsilon\kappa(d-1)} \\ &\leq C h^{4-d/2-\epsilon} + C h^{4-d-\epsilon} \mathbf{n}^{\epsilon\kappa(d-1)}, \end{aligned} \quad (6.4.35)$$

whence by (6.4.29), we have

$$\begin{aligned}
\left| \mathbb{E}[\|u_n^h\|^2 - \|R_h u_n\|^2] \right| &\leq \mathbb{E}[\|u_n^h - R_h u_n\|_{-1} \|u_n^h + R_h u_n\|_1] \\
&\leq (\mathbb{E}[\|e\|_{-1}^2])^{1/2} \mathbb{E}[\|u_n^h + R_h u_n\|_1^2] \\
&\leq Ch^{4-d-2\epsilon} (\mathbb{E}[\|u_n\|_1^2])^{1/2} \leq Ch^{4-d-\epsilon} \mathbf{n}^{\epsilon(\kappa+1)(d-1)}.
\end{aligned}$$

Then by the triangle inequality and (6.4.30), we reach the conclusion.

## Part III: Stochastic Ordinary Differential Equations

## Chapter 7

# A Fundamental Limit Theorem for SDEs with non-Lipschitz coefficients

In this chapter, we prove a fundamental mean-square convergence theorem for stochastic differential equations (SDEs) with coefficients of polynomial growth at infinity and satisfying a one-sided Lipschitz condition. We apply the theorem to a number of existing numerical schemes. We also propose a special balanced scheme which is explicit and of half-order mean-square convergence. Some numerical tests are presented.

### 7.1 Introduction

Let  $(\Omega, \mathcal{F}, P)$  be a complete probability space and  $\mathcal{F}_t^w$  be an increasing family of  $\sigma$ -subalgebras of  $\mathcal{F}$  induced by  $w(t)$  for  $0 \leq t \leq T$ , where  $(w(t), \mathcal{F}_t^w) = ((w_1(t), \dots, w_m(t))^\top, \mathcal{F}_t^w)$  is an  $m$ -dimensional

standard Wiener process. We consider the system of Ito stochastic differential equations (SDEs):

$$dX = a(t, X)dt + \sum_{r=1}^m \sigma_r(t, X)dw_r(t), \quad t \in (t_0, T], \quad X(t_0) = X_0, \quad (7.1.1)$$

where  $X$ ,  $a$ ,  $\sigma_r$  are  $d$ -dimensional column-vectors and  $X_0$  is independent of  $w$ . We suppose that any solution  $X_{t_0, X_0}(t)$  of (7.1.1) is regular on  $[t_0, T]$ , i.e., it is defined for all  $t_0 \leq t \leq T$  [178].

In traditional numerical analysis for SDEs [218, 297, 301], it is assumed that the SDEs coefficients are globally Lipschitz which is a significant limitation as most of the models of applicable interest have coefficients which grow faster at infinity than a linear function. If the global Lipschitz condition is violated, the convergence of many usual numerical methods can disappear, see e.g. [186, 193, 302, 366]. This has been the motivation for the recent interest in both theoretical support of existing numerical methods and developing new methods or approaches for solving SDEs under nonglobal Lipschitz assumptions on the coefficients.

In most of SDEs applications (e.g., in molecular dynamics, financial engineering and other problems of mathematical physics), one is interested in simulating averages  $\mathbb{E}\varphi(X(T))$  of the solution to SDEs – the task for which the weak-sense SDEs approximation is sufficient and effective [297, 301]. The problem with divergence of weak-sense schemes was addressed in [302] (see also [303]) for simulation of averages at finite time and also of ergodic limits when ensemble averaging is used. The concept of rejecting exploding trajectories proposed and justified in [302] allows us to use any numerical method for solving SDEs with nonglobally Lipschitz coefficients for estimating averages. Following this concept, we do not take into account the approximate trajectories  $X(t)$  which leave a sufficiently large ball  $S_R := \{x : |x| < R\}$  during the time  $T$ . See other approaches for resolving this problem in the context of computing averages, including the case of simulating ergodic limits via time averaging, e.g. in [37, 288, 366].

In this work, we deal with mean-square (strong) approximation of SDEs with nonglobal Lipschitz coefficients. Mean-square schemes have their own area of applicability (e.g. for simulating scenarios,

visualization of stochastic dynamics, filtering, etc., see further discussion on this in [197, 218, 301] and references therein). Furthermore, mean-square approximation is of theoretical interest and it also provides a guidance in constructing weak-sense schemes (see, e.g. [218, 297, 301]).

In the case of weak approximation we often have to simulate large dimensional complicated stochastic systems using the Monte Carlo technique (or time averaging), which is typical for molecular dynamics applications, or we have to perform calculations on a daily basis, which is usual, e.g., in financial applications. Hence the cost per time step of a weak numerical integrator should be low, which, in particular, essentially prohibits the use of implicit methods. In contrast, areas of applicability of mean-square schemes, as a rule, do not involve simulation of a large number of trajectories or over very long time periods and, consequently, there are more relaxed requirements on the cost per time step of mean-square schemes and efficient and reliable implicit schemes have practical interest. Strong schemes for SDEs with nonglobal Lipschitz coefficients have been considered in a number of recent works, see e.g. [186, 187, 193, 197, 198, 199, 282, 283, 319] and the references therein); see an extended literature review on this topic in [197].

In this work, we present a variant of the fundamental mean-square convergence theorem in the case of SDEs with nonglobal Lipschitz coefficients, which is analogous to Milstein's fundamental theorem for the global Lipschitz case [296] (see also [297, 301]). More precisely, we assume that the SDEs coefficients can grow polynomially at infinity and satisfy a one-sided Lipschitz condition. The theorem is stated in Section 7.2 and proved in Section 7.5. Its corollary on almost sure convergence is also given. In Section 7.2 we start discussion on applicability of the fundamental theorem, including its application to the drift-implicit Euler scheme and thus establish its order of convergence. Strong convergence (but without order) of this scheme was proved for SDEs with nonglobal Lipschitz drift and diffusion in [197, 282] and more recently its convergence with order half was proved in [283].

We propose a particular balanced method (see the class of balanced methods in [298, 301]) and prove its convergence with order half in the nonglobal Lipschitz setting in Section 7.3. Apparently,

this is the first time when mean-square convergence with an order has been proved for an explicit scheme under the conditions which allow polynomial growth of both drift and diffusion coefficients. Some numerical experiments supporting our results are presented in Section 7.4.

In [376], we also have included fully implicit (i.e., implicit both in drift and diffusion) mean-square schemes for one-sided Lipschitz drift coefficient which grows superlinearly and not faster than polynomial growth at infinity . The fully implicit schemes was proposed and motivated by symplectic integration of stochastic Hamiltonian equations in [299] (see also [301]), where their convergence was proved under globally Lipschitz coefficients. See [376] for more details.

## 7.2 Fundamental theorem

Let  $X_{t_0, X_0}(t) = X(t)$ ,  $t_0 \leq t \leq T$ , be a solution of the system (7.1.1). We will assume the following.

**Assumption 7.2.1.** (i) *The initial condition is such that*

$$\mathbb{E}|X_0|^{2p} \leq K < \infty, \quad \text{for all } p \geq 1. \quad (7.2.1)$$

(ii) *For a sufficiently large  $p_0 \geq 1$  there is a constant  $c_1 \geq 0$  such that for  $t \in [t_0, T]$ ,*

$$(x - y, a(t, x) - a(t, y)) + \frac{2p_0 - 1}{2} \sum_{r=1}^m |\sigma_r(t, x) - \sigma_r(t, y)|^2 \leq c_1 |x - y|^2, \quad x, y \in \mathbb{R}^d. \quad (7.2.2)$$

(iii) *There exist  $c_2 \geq 0$  and  $\varkappa \geq 1$  such that for  $t \in [t_0, T]$ ,*

$$|a(t, x) - a(t, y)|^2 \leq c_2(1 + |x|^{2\varkappa-2} + |y|^{2\varkappa-2})|x - y|^2, \quad x, y \in \mathbb{R}^d. \quad (7.2.3)$$

The condition (7.2.2) implies that

$$(x, a(t, x)) + \frac{2p_0 - 3}{2} \sum_{r=1}^m |\sigma_r(t, x)|^2 \leq c_0 + c'_1 |x|^2, \quad t \in [t_0, T], \quad x \in \mathbb{R}^d, \quad (7.2.4)$$

where  $c_0 = |a(t, 0)|^2/2 + \frac{(2p_0-3)(2p_0-1)}{4} \sum_{r=1}^m |\sigma_r(t, 0)|^2$  and  $c'_1 = c_1 + 1/2$ . The inequality (7.2.4) together with (7.2.1) is sufficient to ensure finiteness of moments [178]: there is  $K > 0$

$$\mathbb{E}|X_{t_0, X_0}(t)|^{2p} < K(1 + \mathbb{E}|X_0|^{2p}), \quad 1 \leq p \leq p_0 - 1, \quad t \in [t_0, T]. \quad (7.2.5)$$

Also, (7.2.3) implies that

$$|a(t, x)|^2 \leq c_3 + c'_2|x|^{2\kappa}, \quad t \in [t_0, T], \quad x \in \mathbb{R}^d, \quad (7.2.6)$$

where  $c_3 = 2|a(t, 0)|^2 + 2c_2(\kappa - 1)/\kappa$  and  $c'_2 = 2c_2(1 + \kappa)/\kappa$ .

Introduce the one-step approximation  $\bar{X}_{t,x}(t+h)$ ,  $t_0 \leq t < t+h \leq T$ , for the solution  $X_{t,x}(t+h)$  of (7.1.1), which depends on the initial point  $(t, x)$ , a time step  $h$ , and  $\{w_1(\theta) - w_1(t), \dots, w_m(\theta) - w_m(t), t \leq \theta \leq t+h\}$  and which is defined as follows:

$$\bar{X}_{t,x}(t+h) = x + A(t, x, h; w_i(\theta) - w_i(t), i = 1, \dots, m, t \leq \theta \leq t+h). \quad (7.2.7)$$

Using the one-step approximation (7.2.7), we recurrently construct the approximation  $(X_k, \mathcal{F}_{t_k})$ ,  $k = 0, \dots, N$ ,  $t_{k+1} - t_k = h_{k+1}$ ,  $T_N = T$ :

$$\begin{aligned} X_0 &= X(t_0), \quad X_{k+1} = \bar{X}_{t_k, X_k}(t_{k+1}) \\ &= X_k + A(t_k, X_k, h_{k+1}; w_i(\theta) - w_i(t_k), i = 1, \dots, m, t_k \leq \theta \leq t_{k+1}). \end{aligned} \quad (7.2.8)$$

The following theorem is a generalization of Milstein's fundamental theorem [296] (see also [297, 301, Chapter 1]) from the global to nonglobal Lipschitz case. It also has similarities with a strong convergence theorem in [186] proved for the case of nonglobal Lipschitz drift, global Lipschitz diffusion and Euler-type schemes.

For simplicity, we will consider a uniform time step size, i.e.,  $h_k = h$  for all  $k$ .

**Theorem 7.2.2.** *Suppose (i) Assumption 7.2.1 holds;*

*(ii) The one-step approximation  $\bar{X}_{t,x}(t+h)$  from (7.2.7) has the following orders of accuracy: for some  $p \geq 1$  there are  $\alpha \geq 1$ ,  $h_0 > 0$ , and  $K > 0$  such that for arbitrary  $t_0 \leq t \leq T-h$ ,  $x \in \mathbb{R}^d$ , and all  $0 < h \leq h_0$ :*

$$|\mathbb{E}[X_{t,x}(t+h) - \bar{X}_{t,x}(t+h)]| \leq K(1 + |x|^{2\alpha})^{1/2} h^{q_1}, \quad (7.2.9)$$

$$[\mathbb{E}|X_{t,x}(t+h) - \bar{X}_{t,x}(t+h)|^{2p}]^{1/(2p)} \leq K(1 + |x|^{2\alpha p})^{1/(2p)} h^{q_2} \quad (7.2.10)$$

with

$$q_2 \geq \frac{1}{2}, \quad q_1 \geq q_2 + \frac{1}{2}; \quad (7.2.11)$$

*(iii) The approximation  $X_k$  from (7.2.8) has finite moments, i.e., for some  $p \geq 1$  there are  $\beta \geq 1$ ,  $h_0 > 0$ , and  $K > 0$  such that for all  $0 < h \leq h_0$  and all  $k = 0, \dots, N$ :*

$$\mathbb{E}|X_k|^{2p} < K(1 + \mathbb{E}|X_0|^{2p\beta}). \quad (7.2.12)$$

Then for any  $N$  and  $k = 0, 1, \dots, N$  the following inequality holds:

$$[\mathbb{E}|X_{t_0, X_0}(t_k) - \bar{X}_{t_0, X_0}(t_k)|^{2p}]^{1/(2p)} \leq K(1 + \mathbb{E}|X_0|^{2\gamma p})^{1/(2p)} h^{q_2 - 1/2}, \quad (7.2.13)$$

where  $K > 0$  and  $\gamma \geq 1$  do not depend on  $h$  and  $k$ , i.e., the order of accuracy of the method (7.2.8) is  $q = q_2 - 1/2$ .

The theorem is proved in Section 7.5 and it uses the following lemma.

**Lemma 7.2.3.** *Suppose Assumption 7.2.1 holds. For the representation*

$$X_{t,x}(t+\theta) - X_{t,y}(t+\theta) = x - y + Z_{t,x,y}(t+\theta), \quad (7.2.14)$$

we have for  $1 \leq p \leq (p_0 - 1)/\varkappa$ :

$$\mathbb{E}|X_{t,x}(t+h) - X_{t,y}(t+h)|^{2p} \leq |x - y|^{2p}(1 + Kh), \quad (7.2.15)$$

$$\mathbb{E}|Z_{t,x,y}(t+h)|^{2p} \leq K(1 + |x|^{2\varkappa-2} + |y|^{2\varkappa-2})^{p/2}|x - y|^{2p}h^p. \quad (7.2.16)$$

This lemma is proved in Section 7.6. Theorem 7.2.2 has the following corollary.

**Corollary 7.2.4.** *In the setting of Theorem 7.2.2 for  $p \geq 1/(2q)$  in (7.2.13), there is  $0 < \varepsilon < q$  and an a.s. finite random variable  $C(\omega) > 0$  such that*

$$|X_{t_0, X_0}(t_k) - X_k| \leq C(\omega)h^{q-\varepsilon},$$

*i.e., the method (7.2.8) for (7.1.1) converges with order  $q - \varepsilon$  a.s.*

The corollary is proved using the Borel-Cantelli-type of arguments (see, e.g. [158, 304]).

## 7.2.1 Discussion

In this section we make a number of observations concerning Theorem 7.2.2.

**1.** As a rule, it is not difficult to check the conditions (7.2.9)-(7.2.10) following the usual routine calculations as in the global Lipschitz case [218, 297, 301]. We note that in order to achieve the optimal  $q_1$  and  $q_2$  in (7.2.9)-(7.2.10) additional assumptions on smoothness of  $a(t, x)$  and  $\sigma_r(t, x)$  are usually needed.

In contrast to the conditions (7.2.9)-(7.2.10), checking the condition (7.2.12) on moments of a method  $X_k$  is often rather difficult. In the case of global Lipschitz coefficients, boundedness of moments of  $X_k$  is just direct implication of the boundedness of moments of the SDEs solution and the one-step properties of the method (see [301, Lemma 1.1.5]). There is no result of this type in the case of nonglobal Lipschitz SDEs and each scheme requires a special consideration. For a number of strong schemes boundedness of moments in nonglobal Lipschitz cases were proved (see,

e.g. [186, 193, 198, 197, 366]). In Section 7.3 we show boundedness of moments for a balanced method. See also [376] for fully implicit methods.

Roughly speaking, Theorem 7.2.2 says that if moments of  $X_k$  are bounded and the scheme was proved to be convergent with order  $q$  in the global Lipschitz case then the scheme has the same convergence order  $q$  in the considered nonglobal Lipschitz case.

**2.** Assumptions and the statement of Theorem 7.2.2 include the famous fundamental theorem of Milstein [296] proved under the global conditions on the SDEs coefficients (of course, as discussed in the previous point, this case does not need the assumption (7.2.12)).

**3.** Consider the drift-implicit scheme [301, p. 30]:

$$X_{k+1} = X_k + a(t_{k+1}, X_{k+1})h + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h}, \quad (7.2.17)$$

where  $\xi_{rk} = (w_r(t_{k+1}) - w_r(t_k)) / \sqrt{h}$  are Gaussian  $\mathcal{N}(0, 1)$  i.i.d. random variables. Assume that the coefficients  $a(t, x)$  and  $\sigma_r(t, x)$  have continuous first-order partial derivatives in  $t$  and the coefficient  $a(t, x)$  also has continuous first-order partial derivatives in  $x^i$  and that all these derivatives and the coefficients themselves satisfy inequalities of the form (7.2.3). It is not difficult to show that the one-step approximation corresponding to (7.2.17) satisfies (7.2.9) and (7.2.10) with  $q_1 = 2$  and  $q_2 = 1$ , respectively. Its boundedness of moments, in particular, under the condition (7.2.4) for time steps  $h \leq 1/(2c_1)$ , is proved in [197]. Then, due to Theorem 7.2.2, (7.2.17) converges with mean-square order  $q = 1/2$  (note that for  $q = 1/2$ , it is sufficient to have  $q_1 = 3/2$  which can be obtained under lesser smoothness of  $a$ ). Further, in the case of additive noise (i.e.,  $\sigma_r(t, x) = \sigma_r(t)$ ,  $r = 1, \dots, m$ ),  $q_1 = 2$  and  $q_2 = 3/2$  and (7.2.17) converges with mean-square order 1 due to Theorem 7.2.2. We note that convergence of (7.2.17) with order half in the global Lipschitz case is well known [218, 297, 301]; in the case of nonglobal Lipschitz drift and global Lipschitz diffusion was proved in [186, 193] (see also related results in [158, 366]); and under Assumption 7.2.1 strong convergence of (7.2.17) without order was proved in [197, 282] and more recently its strong order

half was independently established in [283].

**5.** Due to the bound (7.2.5) on the moments of the solution  $X(t)$ , it would be natural to require that  $\beta$  in (7.2.12) should be equal to 1. Indeed, (7.2.12) with  $\beta = 1$  holds for the drift-implicit method (7.2.17) [197] and for fully implicit methods (see [376, Section 4] or (7.4.4)). However, this is not the case for tamed-type methods (see [198]) or the balanced method from Section 7.3.

**6.** The constant  $K$  in (7.2.13) depends on  $p, t_0, T$  as well as on the SDEs coefficients. The constant  $\gamma$  in (7.2.13) depends on  $\alpha, \beta$  and  $\varkappa$ .

**7.** Let us illustrate Assumption 7.2.1 (ii) on a one-dimensional SDE:  $dX = -\mu X|X|^{r_1-1}dt + \lambda X^{r_2}dw$  with  $\mu, \lambda > 0, r_1 \geq 1$ , and  $r_2 \geq 1$ . If  $r_1 + 1 > 2r_2$  or  $r_1 = r_2 = 1$ , then (7.2.2) is valid for any  $p_0 \geq 1$ . If  $r_1 + 1 = 2r_2$  and  $r_1 > 1$  then (7.2.2) is valid for  $1 \leq p_0 \leq \mu/\lambda^2 + 1/2$ .

### 7.3 A balanced method

In this section we propose a particular balanced scheme from the class of balanced methods introduced in [298] (see also [301]) and prove its mean-square convergence with order half using Theorem 7.2.2. As far as we know, this variant of balanced schemes has not been considered before. In Section 7.4 we test the balanced scheme on a model problem and demonstrate that it is more efficient than the tamed scheme (7.4.2) (see Section 7.4) from [197]. We also note that it was mentioned in [197] that a balanced scheme suitable for the nonglobal Lipschitz case could potentially be derived.

Consider the following balanced-type scheme for (7.1.1):

$$X_{k+1} = X_k + \frac{a(t_k, X_k)h + \sum_{r=1}^m \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h}}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k)\xi_{rk}|}, \quad (7.3.1)$$

where  $\xi_{rk}$  are Gaussian  $\mathcal{N}(0, 1)$  i.i.d. random variables.

We prove two lemmas which show that the scheme (7.3.1) satisfies the conditions of Theorem 7.2.2. The first lemma is on boundedness of moments, which uses a stopping time technique

(see also, e.g. [302, 197]).

**Lemma 7.3.1.** *Suppose Assumption 7.2.1 holds with sufficiently large  $p_0$ . For all natural  $N$  and all  $k = 0, \dots, N$  the following inequality holds for moments of the scheme (7.3.1):*

$$\mathbb{E}|X_k|^{2p} \leq K(1 + \mathbb{E}|X_0|^{2p\beta}), \quad 1 \leq p \leq \frac{p_0 - 1}{4(3\kappa - 2)} - \frac{1}{2}, \quad (7.3.2)$$

with some constants  $\beta \geq 1$  and  $K > 0$  independent of  $h$  and  $k$ .

**Proof.** In the proof we shall use the letter  $K$  to denote various constants which are independent of  $h$  and  $k$ . We note in passing that the case  $\kappa = 1$  (i.e., when  $a(t, x)$  is globally Lipschitz) is trivial.

The following elementary consequence of the inequalities (7.2.4) and (7.2.6) will be used in the proof: there exists a constant  $K > 0$  such that

$$\sum_{r=1}^m |\sigma_r(t, x)|^2 \leq K(1 + |x|^{2\kappa}). \quad (7.3.3)$$

We observe from (7.3.1) that

$$|X_{k+1}| \leq |X_k| + 1 \leq |X_0| + (k + 1). \quad (7.3.4)$$

Let  $R > 0$  be a sufficiently large number. Introduce the events

$$\tilde{\Omega}_{R,k} := \{\omega : |X_l| \leq R, \quad l = 0, \dots, k\}, \quad (7.3.5)$$

and their compliments  $\tilde{\Lambda}_{R,k}$ . We first prove the lemma for integer  $p \geq 1$ . We have

$$\begin{aligned} \mathbb{E}\chi_{\tilde{\Omega}_{R,k+1}}(\omega)|X_{k+1}|^{2p} &\leq \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)|X_{k+1}|^{2p} \\ &= \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)|(X_{k+1} - X_k) + X_k|^{2p} \leq \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)|X_k|^{2p} + \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)|X_k|^{2p-2} \\ &\quad \times [2p(X_k, X_{k+1} - X_k) + p(2p - 1)|X_{k+1} - X_k|^2] \end{aligned} \quad (7.3.6)$$

$$+ K \sum_{l=3}^{2p} \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-l} |X_{k+1} - X_k|^l.$$

Consider the second term in the right-hand side of (7.3.6):

$$\begin{aligned} & \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-2} [2p(X_k, X_{k+1} - X_k) + p(2p-1)|X_{k+1} - X_k|^2] \\ &= 2p \mathbb{E} \left( \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-2} \mathbb{E} \left[ \left( X_k, \frac{a(t_k, X_k)h + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h}}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \right) \right. \right. \\ & \quad \left. \left. + \frac{2p-1}{2} \left| \frac{a(t_k, X_k)h + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h}}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \right|^2 \middle| \mathcal{F}_{t_k} \right] \right). \end{aligned} \quad (7.3.7)$$

Since  $\mathbb{E} \xi_{rk} \prod_{j=1}^m |\xi_{jk}|^{\alpha_j} = 0$  for all  $r$  and any  $\alpha_j \geq 0$ ,  $j = 1, \dots, m$ , and  $\xi_{rk}$  are independent of  $\mathcal{F}_{t_k}$ , we obtain

$$\begin{aligned} & \chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ \frac{\sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h}}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \middle| \mathcal{F}_{t_k} \right] \\ &= \chi_{\tilde{\Omega}_{R,k}} \sum_{r=1}^m \mathbb{E} \left[ \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h} \sum_{i=0}^{\infty} (-1)^i \left[ h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}| \right]^i \middle| \mathcal{F}_{t_k} \right] \\ &= 0. \end{aligned} \quad (7.3.8)$$

Using that  $\mathbb{E} \xi_{rk} \xi_{lk} \prod_{j=1}^m |\xi_{jk}|^{\alpha_j} = 0$  for  $l \neq r$  and any  $\alpha_j \geq 0$ ,  $j = 1, \dots, m$ , we analogously get for  $l \neq r$ :

$$\chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ \frac{\sigma_r(t_k, X_k) \xi_{rk} \sqrt{h} \sigma_l(t_k, X_k) \xi_{lk} \sqrt{h}}{(1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|)^2} \middle| \mathcal{F}_{t_k} \right] = 0. \quad (7.3.9)$$

Then the conditional expectation in (7.3.7) becomes

$$\begin{aligned} A &:= \chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ \left( X_k, \frac{a(t_k, X_k)h + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h}}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \right) \right. \\ & \quad \left. + \frac{2p-1}{2} \left| \frac{a(t_k, X_k)h + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h}}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \right|^2 \middle| \mathcal{F}_{t_k} \right] \end{aligned} \quad (7.3.10)$$

$$\begin{aligned}
&= \chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ \frac{(X_k, a(t_k, X_k)h)}{1 + h|a(t_k, X_k)| + \sqrt{h} \left| \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \right|} \right. \\
&\quad \left. + \frac{2p-1}{2} \frac{a^2(t_k, X_k)h^2 + h \sum_{r=1}^m (\sigma_r(t_k, X_k) \xi_{rk})^2}{\left(1 + h|a(t_k, X_k)| + \sqrt{h} \left| \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \right|\right)^2} \middle| \mathcal{F}_{t_k} \right] \\
&\leq \chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ \frac{(X_k, a(t_k, X_k)h)}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \right. \\
&\quad \left. + \frac{2p-1}{2} \frac{h \sum_{r=1}^m |\sigma_r(t_k, X_k)|^2 \xi_{rk}^2}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \middle| \mathcal{F}_{t_k} \right] + \frac{2p-1}{2} \chi_{\tilde{\Omega}_{R,k}} a^2(t_k, X_k) h^2 \\
&= \chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ \frac{(X_k, a(t_k, X_k)h) + \frac{2p-1}{2} h \sum_{r=1}^m |\sigma_r(t_k, X_k)|^2}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \right. \\
&\quad \left. + \frac{2p-1}{2} \frac{h \sum_{r=1}^m |\sigma_r(t_k, X_k)|^2 (\xi_{rk}^2 - 1)}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \middle| \mathcal{F}_{t_k} \right] + \frac{2p-1}{2} \chi_{\tilde{\Omega}_{R,k}} a^2(t_k, X_k) h^2.
\end{aligned}$$

Using (7.2.4) and (7.2.6), we obtain

$$\begin{aligned}
A &\leq c_0 h + c'_1 |X_k|^2 h \chi_{\tilde{\Omega}_{R,k}} \tag{7.3.11} \\
&\quad + \frac{2p-1}{2} h \chi_{\tilde{\Omega}_{R,k}} \sum_{r=1}^m |\sigma_r(t_k, X_k)|^2 \mathbb{E} \left[ \frac{(\xi_{rk}^2 - 1)}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \middle| \mathcal{F}_{t_k} \right] \\
&\quad + K h^2 + K \chi_{\tilde{\Omega}_{R,k}} |X_k|^{2\neq} h^2.
\end{aligned}$$

Since  $\mathbb{E}(\xi_{rk}^2 - 1) = 0$ , moments of  $\xi_{rk}$  are bounded and  $\xi_{rk}$  are independent of  $\mathcal{F}_{t_k}$ , we obtain for the expectation in the second term in (7.3.11):

$$\begin{aligned}
&\chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ \frac{(\xi_{rk}^2 - 1)}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} \middle| \mathcal{F}_{t_k} \right] \tag{7.3.12} \\
&= \chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ \frac{(\xi_{rk}^2 - 1)}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k) \xi_{rk}|} - (\xi_{rk}^2 - 1) \middle| \mathcal{F}_{t_k} \right] \\
&= -\chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ (\xi_{rk}^2 - 1) \frac{h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_l(t_k, X_k) \xi_{lk}|}{1 + h|a(t_k, X_k)| + \sqrt{h} \sum_{l=1}^m |\sigma_r(t_k, X_k) \xi_{lk}|} \middle| \mathcal{F}_{t_k} \right] \\
&\leq \chi_{\tilde{\Omega}_{R,k}} \mathbb{E} \left[ |\xi_{rk}^2 - 1| \left( h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_l(t_k, X_k)| |\xi_{lk}| \right) \middle| \mathcal{F}_{t_k} \right] \\
&\leq \chi_{\tilde{\Omega}_{R,k}} K \left( h|a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k)| \right).
\end{aligned}$$

Using (7.2.6) and (7.3.3), we get from (7.3.11)-(7.3.12):

$$\begin{aligned}
A &\leq c_0 h + c'_1 \chi_{\tilde{\Omega}_{R,k}} |X_k|^2 h + Kh \chi_{\tilde{\Omega}_{R,k}} \sum_{r=1}^m |\sigma_r(t_k, X_k)|^2 \\
&\times \left[ h |a(t_k, X_k)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t_k, X_k)| \right] + Kh^2 + K \chi_{\tilde{\Omega}_{R,k}} |X_k|^{2\kappa} h^2 \\
&\leq \chi_{\tilde{\Omega}_{R,k}} Kh (1 + |X_k|^2 + |X_k|^{2\kappa} h + |X_k|^{3\kappa} h^{1/2}) \leq \chi_{\tilde{\Omega}_{R,k}} Kh (1 + |X_k|^2 + |X_k|^{3\kappa} h^{1/2}).
\end{aligned} \tag{7.3.13}$$

Now consider the last term in (7.3.6):

$$\begin{aligned}
&\mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-l} |X_{k+1} - X_k|^l \\
&\leq K \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-l} \left[ h^l |a(t_k, X_k)|^l + h^{l/2} \sum_{r=1}^m |\sigma_r(t_k, X_k)|^l |\xi_{rk}|^l \right] \\
&\leq K \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-l} h^{l/2} [1 + |X_k|^{l\kappa}],
\end{aligned} \tag{7.3.14}$$

where we used (7.2.6) and (7.3.3) again as well as the fact that  $\chi_{\tilde{\Omega}_{R,k}}(\omega)$  and  $X_k$  are  $\mathcal{F}_{t_k}$ -measurable while  $\xi_{rk}$  are independent of  $\mathcal{F}_{t_k}$ .

Combining (7.3.6), (7.3.7), (7.3.10), (7.3.13) and (7.3.14), we obtain

$$\begin{aligned}
&\mathbb{E} \chi_{\tilde{\Omega}_{R,k+1}}(\omega) |X_{k+1}|^{2p} \\
&\leq \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p} + Kh \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-2} [1 + |X_k|^2 + |X_k|^{3\kappa} h^{1/2}] \\
&+ K \sum_{l=3}^{2p} \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-l} h^{l/2} [1 + |X_k|^{l\kappa}] \\
&\leq \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p} + Kh \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p} + K \sum_{l=2}^{2p} \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-l} h^{l/2} \\
&+ Kh^{3/2} \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-2+3\kappa} + Kh \sum_{l=3}^{2p} \mathbb{E} \chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p+l(\kappa-1)} h^{l/2-1}.
\end{aligned} \tag{7.3.15}$$

Choosing

$$R = R(h) = h^{-1/(6\kappa-4)}, \tag{7.3.16}$$

we get, for  $l = 3, \dots, 2p$ ,  $\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega) |X_k|^{2p-2+3\alpha} h^{l/2-1} \leq \chi_{\tilde{\Omega}_{R(h),k}}(\omega) |X_k|^{2p}$  and also  $\chi_{\tilde{\Omega}_{R(h),k}}(\omega) |X_k|^{2p+l(\alpha-1)} h^{l/2-1} \leq \chi_{\tilde{\Omega}_{R(h),k}}(\omega) |X_k|^{2p}$ , and hence we rewrite (7.3.15) as

$$\begin{aligned} & \mathbb{E}\chi_{\tilde{\Omega}_{R(h),k+1}}(\omega) |X_{k+1}|^{2p} \\ & \leq \mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega) |X_k|^{2p} + Kh\mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega) |X_k|^{2p} + K \sum_{l=1}^p \mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega) |X_k|^{2(p-l)} h^l \\ & \leq \mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega) |X_k|^{2p} + Kh\mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega) |X_k|^{2p} + Kh, \end{aligned} \tag{7.3.17}$$

where in the last line we have used Young's inequality. From here, we get by Gronwall's inequality that

$$\mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega) |X_k|^{2p} \leq K(1 + \mathbb{E}|X_0|^{2p}), \tag{7.3.18}$$

where  $R(h)$  is from (7.3.16) and  $K$  does not depend on  $k$  and  $h$  but it depends on  $p$ .

It remains to estimate  $\mathbb{E}\chi_{\tilde{\Lambda}_{R(h),k}}(\omega) |X_k|^{2p}$ . We have

$$\begin{aligned} \chi_{\tilde{\Lambda}_{R,k}} &= 1 - \chi_{\tilde{\Omega}_{R,k}} = 1 - \chi_{\tilde{\Omega}_{R,k-1}} \chi_{|X_k| \leq R} = \chi_{\tilde{\Lambda}_{R,k-1}} + \chi_{\tilde{\Omega}_{R,k-1}} \chi_{|X_k| > R} \\ &= \dots = \sum_{l=0}^k \chi_{\tilde{\Omega}_{R,l-1}} \chi_{|X_l| > R}, \end{aligned}$$

where we put  $\chi_{\tilde{\Omega}_{R,-1}} = 1$ . Then, using (7.3.4), (7.3.18), (7.2.1), and Cauchy-Bunyakovsky-Schwarz's and Markov's inequalities, we obtain

$$\begin{aligned} \mathbb{E}\chi_{\tilde{\Lambda}_{R(h),k}}(\omega) |X_k|^{2p} &= \mathbb{E} \sum_{l=0}^k |X_k|^{2p} \chi_{\tilde{\Omega}_{R(h),l-1}} \chi_{|X_l| > R(h)} \\ &\leq (\mathbb{E}|X_0 + k|^{4p})^{1/2} \sum_{l=0}^k \left( \mathbb{E} \left[ \chi_{\tilde{\Omega}_{R(h),l-1}} \chi_{|X_l| > R(h)} \right] \right)^{1/2} \\ &= (\mathbb{E}|X_0 + k|^{4p})^{1/2} \sum_{l=0}^k \left( P(\chi_{\tilde{\Omega}_{R(h),l-1}} |X_l| > R) \right)^{1/2} \\ &\leq (\mathbb{E}|X_0 + k|^{4p})^{1/2} \sum_{l=0}^k \frac{\left( \mathbb{E}(\chi_{\tilde{\Omega}_{R(h),l-1}} |X_l|^{2(2p+1)(6\alpha-4)}) \right)^{1/2}}{R(h)^{(2p+1)(6\alpha-4)}} \end{aligned} \tag{7.3.19}$$

$$\begin{aligned} &\leq K (\mathbb{E}|X_0 + k|^{4p})^{1/2} \left( \mathbb{E}(1 + |X_0|^{2(2p+1)(6\kappa-4)}) \right)^{1/2} kh^{2p+1} \\ &\leq K(1 + \mathbb{E}|X_0|^{4p+2(2p+1)(6\kappa-4)})^{1/2}, \end{aligned}$$

which together with (7.3.18) implies (7.3.2) for integer  $p \geq 1$ . Then, by Jensen's inequality, (7.3.2) holds for non-integer  $p$  as well.  $\square$

The next lemma gives estimates for the one-step error of the balanced scheme (7.3.1).

**Lemma 7.3.2.** *Assume that (7.2.5) holds. Assume that the coefficients  $a(t, x)$  and  $\sigma_r(t, x)$  have continuous first-order partial derivatives in  $t$  and that these derivatives and the coefficients satisfy inequalities of the form (7.2.3). Then the scheme (7.3.1) satisfies the inequalities (7.2.9) and (7.2.10) with  $q_1 = 3/2$  and  $q_2 = 1$ , respectively.*

The proof of this lemma is given in Section 7.7. Lemmas 7.3.1 and 7.3.2 and Theorem 7.2.2 imply the following result.

**Proposition 7.3.3.** *Under the assumptions of Lemmas 7.3.1 and 7.3.2 the balanced scheme (7.3.1) has mean-square order half, i.e., for it the inequality (7.2.13) holds with  $q = q_2 - 1/2 = 1/2$ .*

**Remark 7.3.4.** *In the additive noise case the mean-square order of the balanced scheme (7.3.1) does not improve ( $q_1$  and  $q_2$  remain  $3/2$  and  $1$ , respectively).*

## 7.4 Numerical examples

In this section we will test the following schemes: the balanced method (7.3.1) from Section 7.3; the drift-implicit scheme (7.2.17), the fully implicit Euler scheme (7.4.4) with  $\lambda = 1$ ; the mid-point method (7.4.4) with  $\lambda = 1/2$ ; the drift-tamed Euler scheme (a modified balanced method) [198]:

$$X_{k+1} = X_k + h \frac{a(X_k)}{1 + h |a(X_k)|} + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h}; \quad (7.4.1)$$

the fully-tamed scheme [197]:

$$X_{k+1} = X_k + \frac{a(X_k)h + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h}}{\max\left(1, h \left| ha(X_k) + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h} \right| \right)}; \quad (7.4.2)$$

and the trapezoidal scheme [301, p. 30]:

$$X_{k+1} = X_k + \frac{h}{2} [a(X_{k+1}) + a(X_k)] + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h}. \quad (7.4.3)$$

As before,  $\xi_{rk} = (w_r(t_{k+1}) - w_r(t_k))/\sqrt{h}$  are Gaussian  $\mathcal{N}(0, 1)$  i.i.d. random variables. The two fully implicit scheme for (7.1.1) are from [376]:

$$\begin{aligned} X_{k+1} &= X_k + a(t_{k+\lambda}, (1-\lambda)X_k + \lambda X_{k+1})h \\ &\quad - \lambda \sum_{r=1}^m \sum_{j=1}^d \frac{\partial \sigma_r}{\partial x^j}(t_{k+\lambda}, (1-\lambda)X_k + \lambda X_{k+1}) \sigma_r^j(t_{k+\lambda}, (1-\lambda)X_k + \lambda X_{k+1})h \\ &\quad + \sum_{r=1}^m \sigma_r(t_{k+\lambda}, (1-\lambda)X_k + \lambda X_{k+1}) (\zeta_{rh})_k \sqrt{h}, \end{aligned} \quad (7.4.4)$$

where  $0 \leq \lambda \leq 1$ ,  $t_{k+\lambda} = t_k + \lambda h$  and  $(\zeta_{rh})_k$  are i.i.d. random variables so that

$$\zeta_h = \begin{cases} \xi, & |\xi| \leq A_h, \\ A_h, & \xi > A_h, \\ -A_h, & \xi < -A_h, \end{cases} \quad (7.4.5)$$

with  $\xi \sim \mathcal{N}(0, 1)$  and  $A_h = \sqrt{2l |\ln h|}$  with  $l \geq 1$ . We recall [301, Lemma 1.3.4] that

$$|\mathbb{E}[(\xi^2 - \zeta_h^2)]| \leq (1 + 2\sqrt{2l |\ln h|})h^l. \quad (7.4.6)$$

The drift-tamed Euler scheme (7.4.1) converges with strong convergence order half under Assumption 7.2.1 together with Lipschitz diffusion coefficient [198]. When  $1/2 < \lambda \leq 1$ , the fully implicit

scheme (7.4.4) is expected to converge with order half under similar conditions, see [376]. The fully-tamed Euler scheme (7.4.2) is proved to have strong convergence but without order given under Assumption 7.2.1, see [197].

In all the experiments with fully implicit schemes, where the truncated random variables  $\zeta$  are used, we took  $l = 2$  in (7.4.6). The experiments were performed using Matlab R2012a on a Macintosh desktop computer with Intel Xeon CPU E5462 (quad-core, 2.80 GHz). In simulations we used the Mersenne twister random generator with seed 100. Newton's method was used to solve the nonlinear algebraic equations at each step of the implicit schemes.

We test the methods on two model problems. The first one has nonglobal Lipschitz drift, global Lipschitz diffusion and two noncommutative noises. The second example satisfies Assumption 7.2.1 (nonglobal Lipschitz both drift and diffusion). The aim of the tests is to compare performance of the methods: their accuracy (i.e., roughly speaking, size of prefactors at a power of  $h$ ) and computational costs. We note that experiments cannot prove or disprove boundedness of moments of the schemes since experiments rely on a finite sample of trajectories run over a finite time interval while blow-up of moments in divergent methods (e.g., explicit Euler scheme) is, in general, a result of large deviations [288, 302].

**Example 7.4.1.** *Our first test model is the Stratonovich SDE of the form:*

$$dX = (1 - X^5) dt + X \circ dw_1 + dw_2, \quad X(0) = 0. \quad (7.4.7)$$

In Ito's sense, the drift of the equation becomes  $a(t, x) = 1 - x^5 + x/2$ . Here we tested the balanced method (7.3.1); the drift-tamed scheme (7.4.1); the fully implicit Euler scheme (7.4.4) with  $\lambda = 1$ ; the mid-point method (7.4.4) with  $\lambda = 1/2$ . We note that for all the methods tested on this example except the mid-point rule mean-square convergence with order half is proved either in earlier papers [198, 197, 283] or here as it was described before.

To compute the mean-square error, we run  $M$  independent trajectories  $X^{(i)}(t)$ ,  $X_k^{(i)}$ :

$$\left(E[X(T) - X_N]^2\right)^{1/2} \doteq \left(\frac{1}{M} \sum_{i=1}^M [X^{(i)}(T) - X_N^{(i)}]^2\right)^{1/2}. \quad (7.4.8)$$

We took time  $T = 50$  and  $M = 10^4$ . The reference solution was computed by the mid-point method with small time step  $h = 10^{-4}$ . It was verified that using a different implicit scheme for simulating a reference solution does not affect the outcome of the tests. We chose the mid-point scheme as a reference since in all the experiments it produced the most accurate results.

Table 7.1 gives the mean-square errors and experimentally observed convergence rates for the corresponding methods. We checked that the number of trajectories  $M = 10^4$  was sufficiently large for the statistical errors not to significantly hinder the mean-square errors (the Monte Carlo error computed with 95% confidence was at least 10 time smaller than the reported mean-square errors except values for (7.4.1) at  $h = 0.1$  and  $0.05$  where it was at least 5 time smaller than the mean-square errors). In addition to the data in the table, we evaluated errors for (7.3.1) for smaller time steps:  $h = 0.002$  – the error is  $9.27\text{e-}02$  (rate  $0.41$ ),  $0.001$  –  $6.86\text{e-}02$  ( $0.44$ ). The observed rates of convergence of all the tested methods are close to the predicted  $1/2$ . For a fixed time step  $h$ , the most accurate scheme is the mid-point one, the less accurate scheme is the new balanced method (7.3.1). To produce the result with accuracy  $\sim 0.06 - 0.07$ , in our experiment of running  $M = 10^4$  trajectories the scheme (7.4.1) required 170 sec., the mid-point (7.4.4) with  $\lambda = 1/2$  – 329 sec., (7.4.4) with  $\lambda = 1$  – 723 sec., and (7.3.1) – 1870 sec. That is, our experiments confirmed the conclusion of [198] that the drift-tamed (modified balance method) (7.4.1) from [198] is highly competitive. We note that (7.4.1) is not applicable when diffusion grows faster than a linear function and that in this case the balanced method (7.3.1) can outcompete implicit schemes as it is shown in the next example.

Table 7.1: *Example 7.4.1.* Mean-square errors of the selected schemes. See further details in the text.

$h$	(7.4.4), $\lambda = 1$ rate	(7.4.4), $\lambda = 1/2$ rate	(7.4.1) rate	(7.3.1) rate
0.1	1.712e-01 –	1.443e-01 –	3.748e-01 –	3.594e-01 –
0.05	1.234e-01 0.47	9.224e-02 0.65	2.103e-01 0.83	3.017e-01 0.25
0.02	7.692e-02 0.52	5.261e-02 0.61	9.472e-02 0.87	2.297e-01 0.30
0.01	5.478e-02 0.49	3.549e-02 0.57	6.104e-02 0.63	1.778e-01 0.37
0.005	3.935e-02 0.48	2.487e-02 0.51	3.959e-02 0.62	1.354e-01 0.39

**Example 7.4.2.** Consider the SDE in the Stratonovich sense:

$$dX = (1 - X^5) dt + X^2 \circ dw, \quad X(0) = 0. \quad (7.4.9)$$

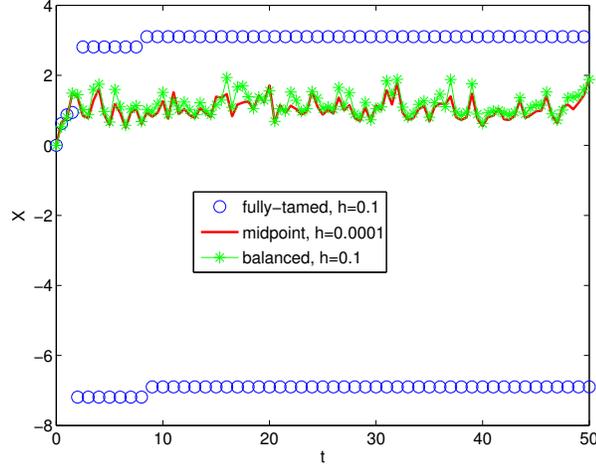
In Ito's sense, the drift of the equation becomes  $a(t, x) = 1 - x^5 + x^3$ .

Here we tested the balanced method (7.3.1); the fully-tamed Euler scheme (7.4.2); the drift-implicit scheme (7.2.17); the fully implicit Euler scheme (7.4.4) with  $\lambda = 1$ ; the mid-point method (7.4.4) with  $\lambda = 1/2$ ; and the trapezoidal scheme (7.4.3). We recall that in the case of nonglobal Lipschitz drift and diffusion, for the drift-implicit scheme (7.2.17) and the balanced method (7.3.1) mean-square convergence with order half is shown earlier in this work and for (7.2.17) also in [283]; it is not difficult to generalize the results of [282] to show boundedness of higher moments of the trapezoidal scheme (7.4.3) and then, using Theorem 7.2.2, to prove its mean-square convergence with order half (see also [283]), which is supported by the experiments. Strong convergence of (7.4.2) without order is proved in [197]. We note that it can be proved directly that implicit algebraic equations arising from application of the mid-point and fully implicit Euler schemes to (7.4.9) have unique solutions under a sufficiently small time step.

The reference solution was computed by the mid-point method with small time step  $h = 10^{-4}$ . The time  $T = 50$  and  $M = 10^4$  in (7.4.8).

The fully-tamed scheme (7.4.2) did not produce accurate results until the time step size is at least  $h = 0.005$  and we do not then report its errors here but see the remark below.

Figure 7.1: Example 7.4.2. Trajectories of the fully-tamed scheme (7.4.2) and the balanced scheme (7.3.1) for  $h = 0.1$ . The reference trajectory is simulated by the mid-point scheme (see (7.4.4) with  $\lambda = 1/2$ ) using a small time step.



**Remark 7.4.3.** *In contrast to the explicit balance scheme (7.3.1), the nature of the explicit fully-tamed scheme (7.4.2) can lead to spurious oscillations, which significantly reduces its practical usefulness. Indeed, if at a step  $k_*$ , the event  $O := \left| ha(X_k) + \sum_{r=1}^m \sigma_r(t_k, X_k) \xi_{rk} \sqrt{h} \right| > 1/h$  happens, then in the case of (7.4.9) the trajectory  $X_k$ ,  $k > k_*$ , oscillates approximately between  $X_{k_*}$  and  $X_{k_*} - \text{sgn}(X_{k_*})/h$ . Since the probability of the event  $O$  is positive for any step size  $h > 0$  and grows with integration time, it is unavoidable that in some scenarios (i.e., on some trajectories) such oscillatory behavior will appear. For instance, in this experiment for  $h = 0.1$  we observed 305 out of 1000 paths for which  $O$  happened over the time interval  $[0, 5]$ , 582 – over  $[0, 10]$  and 989 – over  $[0, 50]$ ; for  $h = 0.05$  – 866 out of 1000 paths over the time interval  $[0, 50]$ . Typical trajectories of the balance scheme (7.3.1) and the fully-tamed scheme (7.4.2) are presented in Fig. 7.1, where the reference solution is computed by the mid-point scheme with a small time step  $h = 0.0001$ . From the practical point of view, (7.4.2) works as long as the explicit Euler scheme works (cf. [288] and also [301, p. 17]). The strong convergence (without order) of (7.4.2) [197] in comparison with the explicit Euler scheme is due to the following fact. When event  $O$  happens for the Euler scheme its sequence  $X_k$  starts oscillating with growing amplitude which leads to unboundedness of its moments*

and, consequently, its divergence in the mean-square sense. For (7.4.2), the oscillations are bounded by  $\sim 1/h$  and since the probability of  $O$  over a finite time interval rapidly decreasing with decrease of  $h$ , then the moments are bounded uniformly in  $h$ . At the same time, the one-step approximation of (7.4.2) does not satisfy the conditions (7.2.9) and (7.2.10) of Theorem 7.2.2. We note that the explicit balanced-type scheme (7.3.1) does not have such drawbacks as (7.4.2).

Table 7.2 gives the mean-square errors and experimentally observed convergence rates for the corresponding methods. We checked that the number of trajectories  $M = 10^4$  was sufficiently large for the statistical errors not to significantly hinder the mean-square errors (the Monte Carlo error computed with 95% confidence was at least ten time smaller than the reported mean-square errors). In addition to the data in the table, we evaluated errors for (7.3.1) for smaller time steps:  $h = 0.002$  – the error is  $3.70\text{e-}02$  (rate 0.41),  $0.001$  –  $2.73\text{e-}02$  (0.44),  $0.0005$  –  $2.00\text{e-}02$  (0.45), i.e., for smaller  $h$  the observed convergence rate of (7.3.1) becomes closer to the theoretically predicted order  $1/2$ . Since (7.4.9) is with single noise, the mid-point scheme demonstrates the first order of convergence. The other implicit schemes show the order half as expected.

Table 7.2: *Example 7.4.2.* Mean-square errors of the selected schemes. See further details in the text.

$h$	(7.2.17)	rate	(7.4.4), $\lambda = 1$	rate	(7.4.4), $\lambda = 1/2$	rate	(7.4.3)	rate	(7.3.1)	rate
0.2	3.449e-01	–	1.816e-01	–	1.378e-01	–	4.920e-01	–	2.102e-01	–
0.1	2.441e-01	0.50	1.331e-01	0.45	8.723e-02	0.66	3.526e-01	0.48	1.637e-01	0.36
0.05	1.592e-01	0.62	9.619e-02	0.47	<b>5.344e-02</b>	0.71	2.230e-01	0.66	1.270e-01	0.37
0.02	8.360e-02	0.70	6.599e-02	0.41	2.242e-02	0.95	1.048e-01	0.82	9.170e-02	0.36
0.01	<b>5.460e-02</b>	0.61	<b>4.919e-02</b>	0.42	1.145e-02	0.97	<b>5.990e-02</b>	0.81	7.065e-02	0.38
0.005	3.682e-02	0.57	3.522e-02	0.48	5.945e-03	0.95	3.784e-02	0.66	<b>5.393e-02</b>	0.39

Table 7.3 presents the time costs in seconds. Let us fix tolerance level at  $0.05 - 0.06$ . We highlighted in bold the corresponding values in both tables. We see that in this example the mid-point scheme is the most efficient which is due to its first order convergence in the commutative case. Among methods of half-order, the balanced method (7.3.1) is the fastest and one can expect that for multi-dimensional SDEs the explicit scheme (7.3.1) can considerably outperform implicit

methods (see a similar outcome for the drift-tamed method (7.4.1) supported by experiments in [198]; note that (7.4.1), in comparison with (7.3.1), is, as a rule, divergent when diffusion is growing faster than a linear function on infinity).

Table 7.3: *Example 7.4.2.* Computational times for the selected schemes. See further details in the text.

$h$	(7.2.17)	(7.4.4), $\lambda = 1$	(7.4.4), $\lambda = 1/2$	(7.4.3)	(7.3.1)
0.2	9.25e+00	1.10e+01	9.33e+00	1.20e+01	3.98e+00
0.1	1.77e+01	2.17e+01	1.80e+01	2.30e+01	7.49e+00
0.05	3.42e+01	4.26e+01	<b>3.51e+01</b>	4.48e+01	1.41e+01
0.02	8.33e+01	1.04e+02	8.69e+01	1.10e+02	3.37e+01
0.01	<b>1.64e+02</b>	<b>2.05e+02</b>	1.73e+02	<b>2.19e+02</b>	6.62e+01
0.005	3.25e+02	4.07e+02	3.47e+02	4.37e+02	<b>1.32e+02</b>

## 7.5 Proof of the fundamental theorem

Note that in this and the two next sections we shall use the letter  $K$  to denote various constants which are independent of  $h$  and  $k$ . The proof exploits the idea of the proof of this theorem in the global Lipschitz case [296].

Consider the error of the method  $\bar{X}_{t_0, X_0}(t_{k+1})$  at the  $(k+1)$ -step:

$$\begin{aligned} \rho_{k+1} &:= X_{t_0, X_0}(t_{k+1}) - \bar{X}_{t_0, X_0}(t_{k+1}) = X_{t_k, X(t_k)}(t_{k+1}) - \bar{X}_{t_k, X_k}(t_{k+1}) \\ &= (X_{t_k, X(t_k)}(t_{k+1}) - X_{t_k, X_k}(t_{k+1})) + (X_{t_k, X_k}(t_{k+1}) - \bar{X}_{t_k, X_k}(t_{k+1})). \end{aligned} \quad (7.5.1)$$

The first difference in the right-hand side of (7.5.1) is the error of the solution arising due to the error in the initial data at time  $t_k$ , accumulated at the  $k$ -th step, which we can re-write as

$$\begin{aligned} S_{t_k, X(t_k), X_k}(t_{k+1}) &= S_{k+1} := X_{t_k, X(t_k)}(t_{k+1}) - X_{t_k, X_k}(t_{k+1}) = \rho_k + Z_{t_k, X(t_k), X_k}(t_{k+1}) \\ &= \rho_k + Z_{k+1}, \end{aligned}$$

where  $Z$  is as in (7.2.14). The second difference in (7.5.1) is the one-step error at the  $(k+1)$ -step

and we denote it as  $r_{k+1}$  :

$$r_{k+1} = X_{t_k, X_k}(t_{k+1}) - \bar{X}_{t_k, X_k}(t_{k+1}).$$

Let  $p \geq 1$  be an integer. We have

$$\begin{aligned} \mathbb{E}|\rho_{k+1}|^{2p} &= \mathbb{E}|S_{k+1} + r_{k+1}|^{2p} & (7.5.2) \\ &= \mathbb{E}[(S_{k+1}, S_{k+1}) + 2(S_{k+1}, r_{k+1}) + (r_{k+1}, r_{k+1})]^p \\ &\leq \mathbb{E}|S_{k+1}|^{2p} + 2p\mathbb{E}|S_{k+1}|^{2p-2}(\rho_k + Z_{k+1}, r_{k+1}) + K \sum_{l=2}^{2p} \mathbb{E}|S_{k+1}|^{2p-l} |r_{k+1}|^l. \end{aligned}$$

Due to (7.2.15) of Lemma 7.2.3, the first term on the right-hand side of (7.5.2) is estimated as

$$\mathbb{E}|S_{k+1}|^{2p} \leq \mathbb{E}|\rho_k|^{2p}(1 + Kh). \quad (7.5.3)$$

Consider the second term on the right-hand side of (7.5.2):

$$\begin{aligned} \mathbb{E}|S_{k+1}|^{2p-2}(\rho_k + Z_{k+1}, r_{k+1}) &= \mathbb{E}|\rho_k|^{2p-2}(\rho_k, r_{k+1}) & (7.5.4) \\ + \mathbb{E}\left(|S_{k+1}|^{2p-2} - |\rho_k|^{2p-2}\right)(\rho_k, r_{k+1}) &+ \mathbb{E}|S_{k+1}|^{2p-2}(Z_{k+1}, r_{k+1}). \end{aligned}$$

Due to  $\mathcal{F}_{t_k}$ -measurability of  $\rho_k$  and due to the conditional variant of (7.2.9), we get for the first term on the right-hand side of (7.5.4):

$$\mathbb{E}|\rho_k|^{2p-2}(\rho_k, r_{k+1}) \leq K\mathbb{E}|\rho_k|^{2p-1}(1 + |X_k|^{2\alpha})^{1/2}h^{q_1}. \quad (7.5.5)$$

Consider the second term on the right-hand side of (7.5.4) and first of all note that it is equal to zero for  $p = 1$ . We have for integer  $p \geq 2$  :

$$\mathbb{E}\left(|S_{k+1}|^{2p-2} - |\rho_k|^{2p-2}\right)(\rho_k, r_{k+1}) \leq K\mathbb{E}|Z_{k+1}||\rho_k||r_{k+1}| \sum_{l=0}^{2p-3} |S_{k+1}|^{2p-3-l} |\rho_k|^l.$$

Further, using  $\mathcal{F}_{t_k}$ -measurability of  $\rho_k$  and the conditional variants of (7.2.10), (7.2.15) and (7.2.16) and the Cauchy-Bunyakovsky-Schwarz inequality (twice), we get for  $p \geq 2$  :

$$\begin{aligned} & \mathbb{E} \left( |S_{k+1}|^{2p-2} - |\rho_k|^{2p-2} \right) (\rho_k, r_{k+1}) \\ & \leq K \mathbb{E} |\rho_k|^{2p-1} (1 + |X(t_k)|^{2\kappa-2} + |X_k|^{2\kappa-2})^{1/4} h^{q_2+1/2} (1 + |X_k|^{2\alpha})^{1/2}. \end{aligned} \quad (7.5.6)$$

Due to  $\mathcal{F}_{t_k}$ -measurability of  $\rho_k$ , the conditional variants of (7.2.10) and (7.2.16) and the Cauchy-Bunyakovsky-Schwarz inequality (twice), we obtain for the third term on the right-hand side of (7.5.4):

$$\begin{aligned} & \mathbb{E} |S_{k+1}|^{2p-2} (Z_{k+1}, r_{k+1}) \\ & \leq \mathbb{E} [\mathbb{E} (S_{k+1}^{4p-4} | \mathcal{F}_{t_k})]^{1/2} \mathbb{E} (|Z_{k+1}|^4 | \mathcal{F}_{t_k})^{1/4} \mathbb{E} (|r_{k+1}|^4 | \mathcal{F}_{t_k})^{1/4} \\ & \leq K \mathbb{E} |\rho_k|^{2p-1} (1 + |X(t_k)|^{2\kappa-2} + |X_k|^{2\kappa-2})^{1/4} h^{q_2+1/2} (1 + |X_k|^{4\alpha})^{1/4}. \end{aligned} \quad (7.5.7)$$

Due to  $\mathcal{F}_{t_k}$ -measurability of  $\rho_k$  and due to the conditional variants of (7.2.10) and (7.2.15) and the Cauchy-Bunyakovsky-Schwarz inequality, we estimate the third term on the right-hand side of (7.5.2):

$$\begin{aligned} K \sum_{l=2}^{2p} \mathbb{E} |S_{k+1}|^{2p-l} |r_{k+1}|^l & \leq K \sum_{l=2}^{2p} \mathbb{E} [\mathbb{E} (|S_{k+1}|^{4p-2l} | \mathcal{F}_{t_k})]^{1/2} \mathbb{E} (|r_{k+1}|^{2l} | \mathcal{F}_{t_k})^{1/2} \\ & \leq K \sum_{l=2}^{2p} \mathbb{E} [|\rho_k|^{2p-l} h^{lq_2} (1 + |X_k|^{2l\alpha})^{1/2}]. \end{aligned} \quad (7.5.8)$$

Substituting (7.5.3)-(7.5.8) in (7.5.2) and recalling that  $q_1 \geq q_2 + 1/2$ , we obtain

$$\begin{aligned} & \mathbb{E} |\rho_{k+1}|^{2p} \leq \mathbb{E} |\rho_k|^{2p} (1 + Kh) + K \mathbb{E} |\rho_k|^{2p-1} (1 + |X_k|^{2\alpha})^{1/2} h^{q_2+1/2} \\ & + K \mathbb{E} |\rho_k|^{2p-1} (1 + |X(t_k)|^{2\kappa-2} + |X_k|^{2\kappa-2})^{1/4} h^{q_2+1/2} (1 + |X_k|^{2\alpha})^{1/2} \\ & + K \mathbb{E} |\rho_k|^{2p-1} (1 + |X(t_k)|^{2\kappa-2} + |X_k|^{2\kappa-2})^{1/4} h^{q_2+1/2} (1 + |X_k|^{4\alpha})^{1/4} \end{aligned}$$

$$\begin{aligned}
& +K \sum_{l=2}^{2p} \mathbb{E}[|\rho_k|^{2p-l} h^{lq_2} (1 + |X_k|^{2l\alpha})^{1/2}] \\
& \leq \mathbb{E}|\rho_k|^{2p} (1 + Kh) \\
& +K \mathbb{E} |\rho_k|^{2p-1} (1 + |X(t_k)|^{2\kappa-2} + |X_k|^{2\kappa-2})^{1/4} h^{q_2+1/2} (1 + |X_k|^{2\alpha})^{1/2} \\
& +K \sum_{l=2}^{2p} \mathbb{E}[|\rho_k|^{2p-l} h^{lq_2} (1 + |X_k|^{2l\alpha})^{1/2}].
\end{aligned}$$

Then using Young's inequality and the conditions (7.2.5) and (7.2.12), we obtain

$$\mathbb{E}|\rho_{k+1}|^{2p} \leq \mathbb{E}|\rho_k|^{2p} + Kh \mathbb{E}|\rho_k|^{2p} + K(1 + \mathbb{E}|X_0|^{\beta p(\kappa-1)+2p\alpha\beta}) h^{2p(q_2-1/2)+1}$$

whence (7.2.13) with integer  $p \geq 1$  follows by application of Gronwall's inequality. Then by Jensen's inequality (7.2.13) holds for non-integer  $p$  as well.  $\square$

## 7.6 Proof of Lemma 7.2.3

Lemma 7.2.3 is an analogue of Lemma 1.1.3 in [301].

**Proof.** Introduce the process  $S_{t,x,y}(s) = S(s) := X_{t,x}(s) - X_{t,y}(s)$  and note that  $Z(s) = S(s) - (x - y)$ . We first prove (7.2.15). Using the Ito formula and the condition (7.2.2) (recall that (7.2.2) implies (7.2.5)), we obtain for  $\theta \geq 0$ :

$$\begin{aligned}
\mathbb{E}|S(t+\theta)|^{2p} & = |x - y|^{2p} + 2p \int_t^{t+\theta} \mathbb{E}|S|^{2p-2} \left[ S^\top(a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) \right. \\
& \quad \left. + \frac{1}{2} \sum_{r=1}^m |\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))|^2 \right] ds \\
& + 2p(p-1) \int_t^{t+\theta} \mathbb{E}|S|^{2p-4} \left| S^\top(s) \sum_{r=1}^m [\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))] \right|^2 ds \\
& \leq |x - y|^{2p} + 2p \int_t^{t+\theta} \mathbb{E}|S|^{2p-2} \left[ S^\top(a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) \right. \\
& \quad \left. + \frac{2p-1}{2} \sum_{r=1}^m |\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))|^2 \right] ds
\end{aligned}$$

$$\leq |x - y|^{2p} + 2pc_1 \int_t^{t+\theta} \mathbb{E}|S(s)|^{2p} ds$$

from which (7.2.15) follows after applying Gronwall's inequality.

Now we prove (7.2.16). Using the Ito formula and the condition (7.2.2), we obtain for  $\theta \geq 0$  :

$$\begin{aligned} \mathbb{E}|Z(t+\theta)|^{2p} &= 2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2} \left[ Z^\top (a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) \right. \\ &\quad \left. + \frac{1}{2} \sum_{r=1}^m |\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))|^2 \right] ds \\ + 2p(p-1) \int_t^{t+\theta} \mathbb{E}|Z|^{2p-4} &\left| Z^\top \sum_{r=1}^m [\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))] \right|^2 ds \\ &\leq 2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}(s) \left[ S^\top (a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) \right. \\ &\quad \left. + \frac{2p-1}{2} \int_t^{t+\theta} \sum_{r=1}^m |\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))|^2 \right] ds \\ &\quad - 2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}(x - y, a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) ds \\ &\leq 2pc_1 \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2} |S|^2 ds \\ &\quad - 2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}(x - y, a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) ds. \end{aligned} \tag{7.6.1}$$

Using Young's inequality, we get for the first term in the right-hand side of (7.6.1):

$$\begin{aligned} 2pc_1 \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2} |S|^2 ds &\leq 4pc_1 \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2} (|Z|^2 + |x - y|^2) ds \\ &\leq K \int_t^{t+\theta} \mathbb{E}|Z|^{2p} ds + K|x - y|^2 \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2} ds. \end{aligned} \tag{7.6.2}$$

Consider the second term in the right-hand side of (7.6.1). Using Hölder's inequality (twice), (7.2.3), (7.2.15) and (7.2.5), we obtain

$$- 2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}(x - y, a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) ds \tag{7.6.3}$$

$$\begin{aligned}
&\leq 2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2} |a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))| |x - y| ds \\
&\leq K|x - y| \int_t^{t+\theta} [\mathbb{E}|Z|^{2p}]^{1-1/p} [\mathbb{E}|a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))|^p]^{1/p} ds \\
&\leq K|x - y| \int_t^{t+\theta} [\mathbb{E}|Z|^{2p}]^{1-1/p} \\
&\quad \times (\mathbb{E}[(1 + |X_{t,x}(s)|^{2\kappa-2} + |X_{t,y}(s)|^{2\kappa-2})^{p/2} |X_{t,x}(s) - X_{t,y}(s)|^p])^{1/p} ds \\
&\leq K|x - y| \int_t^{t+\theta} [\mathbb{E}|Z|^{2p}]^{1-1/p} (\mathbb{E}[(1 + |X_{t,x}(s)|^{2\kappa-2} + |X_{t,y}(s)|^{2\kappa-2})^p])^{1/2p} \\
&\quad \times (\mathbb{E}[|X_{t,x}(s) - X_{t,y}(s)|^{2p}])^{1/2p} ds \\
&\leq K|x - y|^2 (1 + |x|^{2\kappa-2} + |y|^{2\kappa-2})^{1/2} \int_t^{t+\theta} [\mathbb{E}|Z|^{2p}]^{1-1/p} ds.
\end{aligned}$$

Substituting (7.6.2) and (7.6.3) in (7.6.1) and applying Hölder's inequality to  $\mathbb{E}|Z|^{2p-2}$ , we get

$$\begin{aligned}
\mathbb{E}|Z(t + \theta)|^{2p} &\leq K \int_t^{t+\theta} \mathbb{E}|Z|^{2p} ds \\
&\quad + K|x - y|^2 (1 + |x|^{2\kappa-2} + |y|^{2\kappa-2})^{1/2} \int_t^{t+\theta} [\mathbb{E}|Z|^{2p}]^{1-1/p} ds
\end{aligned}$$

whence we obtain (7.2.16) for integer  $p \geq 1$  using Gronwall's inequality as, e.g. in [310, p. 360], and then by Jensen's inequality for non-integer  $p > 1$  as well.  $\square$

## 7.7 Proof of Lemma 7.3.2

As in the global Lipschitz case [298, 301], the proof of Lemma 7.3.2 is routine one-step error analysis.

**Proof.** We start with proving an auxiliary result. Let a function  $\varphi(t, x)$  have continuous first-order partial derivative in  $t$  and that the derivative and the function satisfy inequalities of the form (7.2.3). For  $\alpha \geq 1$  and  $s \geq t$ , we have

$$\begin{aligned}
\mathbb{E}|\varphi(s, X_{t,x}(s)) - \varphi(t, x)|^\alpha &\leq K\mathbb{E}|\varphi(s, X_{t,x}(s)) - \varphi(s, x)|^\alpha + K|\varphi(s, x) - \varphi(t, x)|^\alpha \\
&\leq K\mathbb{E}[(1 + |X_{t,x}(s)|^{\kappa-1} + |x|^{\kappa-1})|X_{t,x}(s) - x|]^\alpha + K\left|\int_t^s \frac{\partial}{\partial s}\varphi(s', x)ds'\right|^\alpha
\end{aligned}$$

$$\begin{aligned}
&\leq K\mathbb{E}(1 + |X_{t,x}(s)|^{\varkappa-1} + |x|^{\varkappa-1})^\alpha \left| \int_t^s a(s', X_{t,x}(s')) ds' \right|^\alpha \\
&+ K \sum_{r=1}^q \mathbb{E}(1 + |X_{t,x}(s)|^{\varkappa-1} + |x|^{\varkappa-1})^\alpha \left| \int_t^s \sigma_r(s', X_{t,x}(s')) dw_r(s') \right|^\alpha \\
&\quad + K(1 + |x|^{\alpha\varkappa})(s-t)^\alpha.
\end{aligned}$$

Then, using the Cauchy-Bunyakovsky-Schwarz inequality, (7.2.5) and (7.2.3), we get

$$\begin{aligned}
\mathbb{E} |\varphi(s, X_{t,x}(s)) - \varphi(t, x)|^\alpha &\leq K(1 + |x|^{\alpha(\varkappa-1)}) \left[ \mathbb{E} \left( \int_t^s (1 + |X_{t,x}(s')|^\varkappa) ds' \right)^{2\alpha} \right]^{1/2} \\
&+ K(1 + |x|^{\alpha(\varkappa-1)}) \sum_{r=1}^q \left[ \mathbb{E} \left| \int_t^s \sigma_r(s', X_{t,x}(s')) dw_r(s') \right|^{2\alpha} \right]^{1/2} + K(1 + |x|^{\alpha\varkappa})(s-t)^\alpha.
\end{aligned} \tag{7.7.1}$$

By the inequality for powers of Ito integrals from [133, p. 26], we obtain that

$$\mathbb{E} \left| \int_t^s \sigma_r(s', X_{t,x}(s')) dw_r(s') \right|^{2\alpha} \leq K(s-t)^{\alpha-1} \int_t^s \mathbb{E} |\sigma_r(s', X_{t,x}(s'))|^{2\alpha} ds'. \tag{7.7.2}$$

And, by the same recipe as in [133, p. 26] which exploits Hölder's inequality, it is not difficult to get

$$\mathbb{E} \left[ \int_t^s |1 + X_{t,x}(s')|^\varkappa ds' \right]^{2\alpha} \leq K(s-t)^{2\alpha-1} \int_t^s \mathbb{E} |1 + X_{t,x}(s')|^{2\alpha\varkappa} ds'. \tag{7.7.3}$$

It follows from (7.7.1)-(7.7.3), the assumption that  $\sigma_r$  satisfy (7.2.3) and from (7.2.5) that

$$\mathbb{E} |\varphi(s, X_{t,x}(s)) - \varphi(t, x)|^\alpha \leq K(1 + |x|^{2\alpha\varkappa-\alpha})[(s-t)^{\alpha/2} + (s-t)^\alpha], \tag{7.7.4}$$

which, in particular, holds for the functions  $a(t, x)$  and  $\sigma_r(t, x)$  under the conditions of the lemma.

Now consider the one-step approximation of the SDE (7.1.1), which corresponds to the balanced method (7.3.1):

$$X = x + \frac{a(t, x)h + \sum_{r=1}^m \sigma_r(t, x)\xi_r\sqrt{h}}{1 + h|a(t, x)| + \sqrt{h} \sum_{r=1}^m |\sigma_r(t, x)\xi_r|} \tag{7.7.5}$$

and the one-step approximation corresponding to the explicit Euler scheme:

$$\tilde{X} = x + a(t, x)h + \sum_{r=1}^m \sigma_r(t, x)\xi_r\sqrt{h}. \quad (7.7.6)$$

We start with analysis of the one-step error of the Euler scheme:

$$\tilde{\rho}(t, x) := X_{t,x}(t+h) - \tilde{X}.$$

Using (7.7.4), we obtain

$$\begin{aligned} |\mathbb{E}\tilde{\rho}(t, x)| &= \left| \mathbb{E} \int_t^{t+h} (a(s, X_{t,x}(s)) - a(t, x)) ds \right| \leq \mathbb{E} \int_t^{t+h} |a(s, X_{t,x}(s)) - a(t, x)| ds \quad (7.7.7) \\ &\leq Kh^{3/2}(1 + |x|^{2\kappa-1}). \end{aligned}$$

(We remark that assuming additional smoothness of  $a(t, x)$ , we can get an estimate for  $\mathbb{E}\tilde{\rho}(t, x)$  of order  $O(h^2)$  but this will not improve the result of this lemma for the balanced scheme (7.3.1).)

Further,

$$\begin{aligned} \mathbb{E}\tilde{\rho}^{2p}(t, x) &\leq K \mathbb{E} \left| \int_t^{t+h} (a(s, X_{t,x}(s)) - a(t, x)) ds \right|^{2p} \quad (7.7.8) \\ &\quad + K \sum_{r=1}^q \mathbb{E} \left| \int_t^{t+h} (\sigma_r(s, X_{t,x}(s)) - \sigma_r(t, x)) dw_r(s) \right|^{2p}. \end{aligned}$$

Using the same recipe as in (7.7.3) and then applying (7.7.4), we get for the first term in (7.7.8):

$$\begin{aligned} \mathbb{E} \left| \int_t^{t+h} (a(s, X_{t,x}(s)) - a(t, x)) ds \right|^{2p} &\leq Kh^{2p-1} \int_t^{t+h} \mathbb{E} |a(s, X_{t,x}(s)) - a(t, x)|^{2p} ds \quad (7.7.9) \\ &\leq Kh^{3p}(1 + |x|^{4p\kappa-2p}). \end{aligned}$$

Using an inequality of the form (7.7.2) and then applying (7.7.4), we obtain

$$\begin{aligned} & \mathbb{E} \left| \int_t^{t+h} (\sigma_r(s, X_{t,x}(s)) - \sigma_r(t, x)) dw_r(s) \right|^{2p} \\ & \leq Kh^{p-1} \int_t^{t+h} \mathbb{E} |\sigma_r(s, X_{t,x}(s)) - \sigma_r(t, x)|^{2p} ds \leq Kh^{2p}(1 + |x|^{4p\kappa-2p}). \end{aligned} \quad (7.7.10)$$

It follows from (7.7.8)-(7.7.10) that

$$\mathbb{E}\tilde{\rho}^{2p}(t, x) \leq Kh^{2p}(1 + |x|^{4p\kappa-2p}). \quad (7.7.11)$$

Now we compare the one-step approximations (7.7.5) of the balanced scheme and (7.7.6) of the Euler scheme:

$$X = x + \frac{a(t, x)h + \sum_{r=1}^m \sigma_r(t, x)\xi_r\sqrt{h}}{1 + h|a(t, x)| + \sqrt{h}\sum_{r=1}^m |\sigma_r(t, x)\xi_r|} = \tilde{X} - \rho(t, x), \quad (7.7.12)$$

where

$$\rho(t, x) = \left( a(t, x)h + \sum_{r=1}^m \sigma_r(t, x)\xi_r\sqrt{h} \right) \frac{h|a(t, x)| + \sqrt{h}\sum_{r=1}^m |\sigma_r(t, x)\xi_r|}{1 + h|a(t, x)| + \sqrt{h}\sum_{r=1}^m |\sigma_r(t, x)\xi_r|}.$$

Using the equality (7.3.8) and the assumptions made on the coefficients (see (7.2.3)), we obtain

$$|\mathbb{E}\rho(t, x)| = \left| a(t, x)h \mathbb{E} \frac{h|a(t, x)| + \sqrt{h}\sum_{r=1}^m |\sigma_r(t, x)\xi_r|}{1 + h|a(t, x)| + \sqrt{h}\sum_{r=1}^m |\sigma_r(t, x)\xi_r|} \right| \leq Kh^{3/2}(1 + |x|^{2\kappa}),$$

which together with (7.7.12) and (7.7.7) implies that (7.7.5) satisfies (7.2.9) with  $q_1 = 3/2$ . Further,

$$\mathbb{E}\rho^{2p}(t, x) \leq h^{2p} \mathbb{E} \left[ \sqrt{h}|a(t, x)| + \sum_{r=1}^m |\sigma_r(t, x)\xi_r| \right]^{4p} \leq Kh^{2p}(1 + |x|^{4p\kappa}),$$

which together with (7.7.12) and (7.7.11) implies that (7.7.5) satisfies (7.2.10) with  $q_2 = 1$ .  $\square$

## Chapter 8

# Numerical schemes for SDEs with time delay via Wong-Zakai approximation

In this chapter, we derive three numerical schemes for stochastic delay differential equations (SDDEs) using the Wong-Zakai approximation. By approximating the Brownian motion with its truncated spectral expansion and then using different discretizations in time, we obtain three schemes: a Predictor-Corrector scheme, a Midpoint scheme and a Milstein-like scheme. We prove that the Predictor-Corrector scheme converges with order half in the mean-square sense while the Milstein-like scheme converges with order one. Numerical tests confirm the theoretical prediction and demonstrate that the Midpoint scheme is of half-order convergence. Numerical results also show that the Predictor-Corrector and Midpoint schemes can be of first-order convergence under commutative noises when there is no delay in the diffusion coefficients.

## 8.1 Introduction

Numerical solution of stochastic delay differential equations (SDDEs) has attracted increasing interest recently, as memory effects in the presence of noise are modeled with SDDEs in engineering and finance, e.g., [146, 188, 334, 377, 387]. Most of numerical methods for SDDEs have focused on the convergence and stability of different time-discretization schemes since the early work [379, 380]. Currently, several time-discretization schemes have been well-studied: the Euler-type schemes (the forward Euler scheme [12, 238] and the drift-implicit Euler scheme [196, 257, 401]), the Milstein schemes [46, 189, 195, 219], the split-step schemes [148, 394, 417], and also some multi-step schemes [47, 48, 53, 54].

These schemes are usually based on the Ito-Taylor expansion, see e.g. [219] or anticipative calculus, see e.g. [195]. Here we employ a different approach, the so-called Wong-Zakai (WZ) approximation, see e.g. [271, 381, 386]. The difference between WZ approximation and the aforementioned schemes is that in WZ we first approximate the Brownian motion with an absolute continuous process and then apply proper time-discretization schemes for the resulting equation while the aforementioned schemes are ready for simulation without any further time discretization. The WZ approximation thus can be viewed as an intermediate step for deriving numerical schemes and can provide more flexibility of discretization of Brownian motion before performing any time discretization. Moreover, with the WZ approximation we can apply Taylor's expansion rather than Ito-Taylor expansion and anticipative calculus. In this chapter, we show the flexibility of this approach and derive three numerical schemes for SDDEs using the Stratonovich formulation with a spectral truncation of Brownian motion.

In WZ, we employ the classical piecewise linear interpolation of Brownian motion in e.g. [271, 381, 386] and a Fourier approximation of Brownian motion. Specifically, we will derive three distinct schemes using different time-discretization techniques. After approximating the Brownian motion by a spectral expansion, we then use the trapezoidal rule and the predictor-corrector strategy to

obtain a Predictor-Corrector scheme and prove its convergence in the mean-square sense. We also use the midpoint rule within the WZ approximation to derive a fully implicit scheme (implicit in both drift and diffusion terms). These two schemes are convergent with strong order half for SDDEs, as shown numerically in Section 8.3.

If no delay arises, the Predictor-Corrector scheme and the Midpoint scheme coincide with those for stochastic differential equations without delay. The Predictor-Corrector scheme degenerates into a family of the Predictor-Corrector scheme in [39], which were proposed in order to overcome numerical stability introduced by the Euler scheme and other one-step explicit schemes. Without delay, our Midpoint scheme becomes one of the symplectic-preserving schemes in [299] for stochastic Hamiltonian systems. Though we will only focus on the convergence of these schemes and check their numerical performance, we expect that these schemes have larger stability regions than the Euler scheme for SDDEs as in the cases without delay.

Based on Taylor expansion of the diffusion terms, we also derive a first-order scheme (called Milstein-like), which is similar to the Milstein scheme [189, 195, 219]. The Milstein-like scheme we propose here can be readily used in routine simulation unlike the Milstein scheme [189, 195, 219] which requires additional approximation of the double integrals. Specifically, the double integrals are approximated with spectral truncation using truncation parameters reciprocal to the time step size to achieve first-order convergence, which will be shown both in theory and in computation. The spectral truncations we use are from the piecewise linear interpolation and a Fourier expansion. Comparison between these two truncations will be presented for a specific numerical example in Section 8.3, where it is shown that the Fourier approach is faster than the piecewise constant approach. It is worth noting that the approximation of double integrals in the present context is similar to those using numerical integration techniques which has been long explored, see e.g. [218, 301].

The rest of this chapter is organized as follows. In Section 8.2, we show how to derive our schemes from WZ approximation to the Stratonovich SDDEs. Numerical results will be presented

in Section 8.3 to illustrate the convergence of the three schemes and to compare their numerical performance. We will show that the Milstein-like scheme is much slower than the Predictor-Corrector and Midpoint schemes as in each step the evaluation of double integrals are expensive, no matter what approximation for the double integrals is used. Finally, we prove in Section 8.4 that the Predictor-Corrector scheme is of half-order convergence in the mean-square sense while the Milstein-like scheme is of first-order convergence.

## 8.2 Numerical schemes for SDDEs

Consider the following SDDE with constant delay in Stratonovich form:

$$\begin{aligned} dX(t) &= f(X(t), X(t-\tau))dt + \sum_{l=1}^r g_l(X(t), X(t-\tau)) \circ dW_l(t), \quad t \in (0, T], \\ X(t) &= \phi(t), \quad t \in [-\tau, 0], \end{aligned} \tag{8.2.1}$$

where  $\tau > 0$  is a constant,  $(W(t), \mathcal{F}_t) = (\{W_l(t), 1 \leq l \leq r\}, \mathcal{F}_t)$  is a system of one-dimensional independent standard Wiener process, the functions  $f : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $g_l : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $\phi(t) : [-\tau, 0] \rightarrow \mathbb{R}^d$  is continuous with  $\mathbb{E}\|\phi\|_{L^\infty}^2 < \infty$ . We also assume that  $\phi(t)$  is  $\mathcal{F}_0$ -measurable.

For the mean-square stability of Equation (8.2.1), we assume that  $f$ ,  $g_l$ ,  $\partial_x g_l g_l$  and  $\partial_{x_\tau} g_l g_l$ , ( $\partial_x$  and  $\partial_{x_\tau}$  denote the derivatives with respect to the first and second variables, respectively),  $l = 1, 2, \dots, r$  in Equation (8.2.1) satisfy the following Lipschitz conditions:

$$|v(x_1, y_1) - v(x_2, y_2)|^2 \leq L_v(|x_1 - x_2|^2 + |y_1 - y_2|^2), \tag{8.2.2}$$

and the linear growth conditions

$$|v(x_1, y_1)|^2 \leq K(1 + |x_1|^2 + |y_1|^2) \tag{8.2.3}$$

for every  $x_1, y_1, x_2, y_2 \in \mathbb{R}^d$ , where  $L_v, K$  are positive constants, which depend only on  $v$ . Under these conditions, Equation (8.2.1) has a unique sample-continuous and  $\mathcal{F}_t$ -adapted strong solution  $X(t) : [-\tau, +\infty) \rightarrow \mathbb{R}^d$ , see e.g. [281, 312].

The Wong-Zakai (WZ) approximation, see e.g. [399, 400], is a semi-discretization method where Brownian motion is approximated by finite dimensional absolute continuous stochastic processes before any discretization in time or in physical space. There are different types of WZ approximation, see e.g. [202, 328, 364, 423]. Here we use an orthogonal expansion approach for WZ approximation of Brownian motion:

$$W^{(n)}(t) = \sum_{j=1}^n \int_0^t m_j(s) ds \int_0^T m_j(t) dW, \quad t \in [0, T], \quad (8.2.4)$$

where  $\{m_j(t)\}_{j=1}^\infty$  is a complete orthonormal system (CONS) in  $L^2([0, T])$  and  $\xi_j =: \int_0^T m_j(t) dW$ , are mutually independent standard Gaussian random variables. In this chapter, we will use a piecewise version of spectral expansion (8.2.4) by taking a partition  $0 = \mathbf{t}_0 < \mathbf{t}_1 < \dots < \mathbf{t}_{N_\Delta-1} < \mathbf{t}_{N_\Delta} = T$  and choosing a truncated CONS,  $\{m_j^{(n)}(t)\}_{j=1}^{N_h}$  in  $L^2([\mathbf{t}_n, \mathbf{t}_{n+1}])$  for  $n = 0, \dots, N_\Delta - 1$ :

$$W^{(N_h, n)}(t) = \sum_{n=0}^{N_\Delta-1} \sum_{j=1}^{N_h} \int_0^t \chi_{[\mathbf{t}_n, \mathbf{t}_{n+1})} m_j^{(n)}(s) ds \xi_j^{(n)}, \quad \xi_j^{(n)} = \int_{\mathbf{t}_n}^{\mathbf{t}_{n+1}} m_j^{(n)}(s) dW \quad (8.2.5)$$

where  $\chi$  is the indicator function.

Here different choices of CONS lead to different representations. The orthonormal piecewise constant basis over time interval  $[\mathbf{t}_n, \mathbf{t}_{n+1})$ , with  $\Delta' = (\mathbf{t}_{n+1} - \mathbf{t}_n)/N_h$ ,

$$m_j^{(n)}(t) = \frac{\sqrt{N_h}}{\sqrt{\mathbf{t}_{n+1} - \mathbf{t}_n}} \chi_{[\mathbf{t}_n + (j-1)\Delta', \mathbf{t}_n + j\Delta')}, \quad j = 1, 2, \dots, N_h, \quad (8.2.6)$$

gives the classical piecewise linear interpolation (see e.g. [202, 385, 399, 400]), if  $N_h = 1$ ,

$$W^{(1, n)}(t) = W(\mathbf{t}_n) + (t - \mathbf{t}_n) \frac{W(\mathbf{t}_{n+1}) - W(\mathbf{t}_n)}{\mathbf{t}_{n+1} - \mathbf{t}_n}, \quad t \in [\mathbf{t}_n, \mathbf{t}_{n+1}]. \quad (8.2.7)$$

The orthonormal Fourier basis gives Wiener's representation (see e.g. [218, 301, 333]):

$$\begin{aligned} m_1^{(n)}(t) &= \frac{1}{\sqrt{\mathbf{t}_{n+1} - \mathbf{t}_n}}, \quad m_{2k}^{(n)}(t) = \sqrt{\frac{2}{\mathbf{t}_{n+1} - \mathbf{t}_n}} \sin\left(\frac{2k\pi}{\mathbf{t}_{n+1} - \mathbf{t}_n}(t - \mathbf{t}_n)\right), \\ m_{2k+1}^{(n)}(t) &= \sqrt{\frac{2}{\mathbf{t}_{n+1} - \mathbf{t}_n}} \cos\left(\frac{2k\pi}{\mathbf{t}_{n+1} - \mathbf{t}_n}(t - \mathbf{t}_n)\right), \quad t \in [\mathbf{t}_n, \mathbf{t}_{n+1}]. \end{aligned} \quad (8.2.8)$$

Note that taking  $N_h = 1$  in (8.2.8) leads to the piecewise linear interpolation (8.2.7). Besides, we can also use the wavelet basis, which gives the Levy-Ciesielsky representation [214].

In this chapter, we consider the spectral approximation (8.2.5) with piecewise constant basis (8.2.6) and Fourier basis (8.2.8). With these approximations, we have the following WZ approximation for Equation (8.2.1)

$$\begin{aligned} d\tilde{X}(t) &= f(\tilde{X}(t), \tilde{X}(t - \tau))dt + \sum_{l=1}^r g_l(\tilde{X}(t), \tilde{X}(t - \tau))d\tilde{W}_l(t), \quad t \in [0, T], \\ \tilde{X}(t) &= \phi(t), \quad t \in (-\tau, 0], \end{aligned} \quad (8.2.9)$$

where  $\tilde{W}_l(t)$  can be any approximation of  $W_l(t)$  described above. For the piecewise linear interpolation (8.2.7), we have the following consistency of the WZ approximation (8.2.9) to Equation (8.2.1).

**Theorem 8.2.1** (Consistency, [381]). *Suppose  $f$  and  $g_l$  in Equation (8.2.1) are Lipschitz continuous and satisfy conditions (8.2.2) and have second-order continuous and bounded partial derivatives. Suppose also the initial segment  $\phi(t), t \in [-\tau, 0]$  to be on the probability space  $(\Omega, \mathcal{F}, P)$  and  $\mathcal{F}_0$ -measurable and right continuous, and  $E\|\phi\|_{L^\infty}^2 < \infty$ . For  $\tilde{X}(t)$  in (8.2.9) with piecewise linear approximation of Brownian motion (8.2.7), we have for any  $t \in (0, T]$ ,*

$$\lim_{n \rightarrow \infty} \sup_{0 \leq s \leq t} \mathbb{E}[|X(s) - \tilde{X}(s)|^2] = 0. \quad (8.2.10)$$

The consistency of the WZ approximation with spectral approximation (8.2.5) can be established

by the argument of integration by parts as in [171, 202], under similar conditions on the drift and diffusion coefficients.

### 8.2.1 Derivation of numerical schemes

We will further discretize Equation (8.2.9) in time and derive several numerical schemes for (8.2.1). To this end, we take an uniform time step size  $h$ , which satisfies  $\tau = mh$  and  $m$  is a positive integer;  $N_T = T/h$  ( $T$  is the final time);  $t_n = nh$ ,  $n = 0, 1, \dots, N_T$ . For simplicity, we take the same partition for the WZ approximation exactly as the time discretization, i.e.,

$$\mathbf{t}_n = t_n, \quad n = 0, 1, \dots, N_T \quad \text{and} \quad \Delta =: \mathbf{t}_n - \mathbf{t}_{n-1} = t_n - t_{n-1} = h.$$

For Equation (8.2.9), we have the following integral form over  $[t_n, t_{n+1}]$ :

$$\begin{aligned} \int_{t_n}^{t_{n+1}} d\tilde{X}(t) &= \int_{t_n}^{t_{n+1}} f(\tilde{X}(t), \tilde{X}(t-\tau))dt + \sum_{l=1}^r \int_{t_n}^{t_{n+1}} g_l(\tilde{X}(t), \tilde{X}(t-\tau))d\tilde{W}_l(t) \quad (8.2.11) \\ &= \int_{t_n}^{t_{n+1}} f(\tilde{X}(t), \tilde{X}(t-\tau))dt + \sum_{l=1}^r \int_{t_n}^{t_{n+1}} g_l(\tilde{X}(t), \tilde{X}(t-\tau)) \sum_{j=1}^{N_h} m_j^{(n)}(t)\xi_{l,j}^{(n)} dt. \end{aligned}$$

Here we emphasize that the time-discretization for the diffusion term have to be at least half-order. Otherwise, the resulting scheme is not consistent, e.g., Euler-type schemes, in general, converge to the corresponding SDDEs in Ito sense instead of those in Stratonovich sense. In fact, if  $g_l(\tilde{X}(t), \tilde{X}(t-\tau))$  ( $l = 1, \dots, r$ ) is approximated by  $g_l(\tilde{X}(t_n), \tilde{X}(t_n-\tau))$  in Equation (8.2.11), then we have, for both Fourier basis (8.2.8) and piecewise constant basis (8.2.6),

$$\int_{t_n}^{t_{n+1}} d\tilde{X}(t) = \int_{t_n}^{t_{n+1}} f(\tilde{X}(t), \tilde{X}(t-\tau))dt + \sum_{l=1}^r g_l(\tilde{X}(t_n), \tilde{X}(t_n-\tau))\Delta W_{l,n},$$

where  $\Delta W_{l,n} = W_l(t_{n+1}) - W_l(t_n)$ . This will lead to Euler-type scheme which converges to the

following SDDE in the Ito sense, see e.g. [12, 257], instead of (8.2.1):

$$dX(t) = f(X(t), X(t - \tau))dt + \sum_{l=1}^r g_l(X(t), X(t - \tau))dW_l(t).$$

In the following, three numerical schemes for solving Equation (8.2.1) are derived using Taylor expansion and different discretizations in time in (8.2.11). The first scheme is a *Predictor-Corrector* scheme. Taking  $N_h = 1$ , we have that both base, (8.2.6) and (8.2.8), have only one term  $m_1^{(n)} = 1/\sqrt{h}$  over each subinterval. Using the trapezoidal rule to approximate the integrals on the right side of (8.2.11), we get

$$\begin{aligned} X_{n+1} = X_n + \frac{h}{2} [f(X_n, X_{n-m}) + f(X_{n+1}, X_{n-m+1})] \\ + \frac{1}{2} \sum_{l=1}^r [g_l(X_n, X_{n-m}) + g_l(X_{n+1}, X_{n-m+1})] \Delta W_{l,n}, \end{aligned} \quad (8.2.12)$$

where  $X_n$  is an approximation of  $\tilde{X}(t_n)$  (thus an approximation of  $X(t_n)$ ). The initial conditions are  $X_n = \phi(nh)$ , when  $n = -m, -m + 1, \dots, 0$ . Note that the scheme (8.2.12) is fully implicit and is not solvable as  $\Delta W_{l,n}$  can take any values in the real line. To resolve this issue, we further apply the left rectangle rule on the right side of (8.2.11) to obtain a predictor for  $X_{n+1}$  in (8.2.12) so that the resulting scheme is explicit. Actually, we arrive at a Predictor-Corrector scheme for SDDE (8.2.1):

$$\begin{aligned} \bar{X}_{n+1} = X_n + hf(X_n, X_{n-m}) + \sum_{l=1}^r g_l(X_n, X_{n-m}) \Delta W_{l,n}, \\ X_{n+1} = X_n + \frac{h}{2} [f(X_n, X_{n-m}) + f(\bar{X}_{n+1}, X_{n-m+1})] \\ + \frac{1}{2} \sum_{l=1}^r [g_l(X_n, X_{n-m}) + g_l(\bar{X}_{n+1}, X_{n-m+1})] \Delta W_{l,n}, \quad n = 0, 1, \dots, N_T - 1. \end{aligned} \quad (8.2.13)$$

**Remark 8.2.2.** Taking  $N_h = 1$  is sufficient for half-order schemes, such as the scheme *Predictor-Corrector* scheme (8.2.13) and the following *Midpoint* scheme. Both schemes employ

$\int_{t_n}^{t_{n+1}} \sum_{j=1}^{N_h} m_j^{(n)}(t) \xi_{l,j}^{(n)} dt$ , which is equal to  $\Delta W_{l,n}$  for any  $N_h \geq 1$ , according to (8.2.5) and our choices of orthonormal base (8.2.6) and (8.2.8).

**Theorem 8.2.3.** *Assume that  $f$ ,  $g_l$ ,  $\partial_x g_l g_q$  and  $\partial_{x_\tau} g_l g_q$  ( $l, q = 1, 2, \dots, r$ ) satisfy the Lipschitz condition (8.2.2) and also  $g_l$  have bounded second-order partial derivatives with respect to all variables. If  $\mathbb{E}[\|\phi\|_{L^\infty}^p] < \infty$ ,  $p \geq 4$ , then we have for the Predictor-Corrector scheme (8.2.13),*

$$\max_{1 \leq n \leq N_T} \mathbb{E}|X(t_n) - X_n|^2 = O(h). \quad (8.2.14)$$

The proof will be presented in Section 8.4.

**Remark 8.2.4.** *When  $\tau = 0$  both in drift and diffusion coefficients, the scheme (8.2.13) degenerates into one family of the Predictor-Corrector schemes in [39], which can have larger stability region than the explicit Euler scheme and some other one-step schemes, especially for stochastic differential equations with multiplicative noises. Moreover, we will numerically show that if the time delay only exists in the drift term in SDDE with commutative noise (for one-dimensional case, i.e.,  $d = 1$ , the commutative condition is  $g_l \partial_x g_q - g_q \partial_x g_l = 0$ ,  $1 \leq l, q \leq r$ ), the proposed Predictor-Corrector scheme can be convergent with order one in the mean-square sense.*

The second scheme is a *Midpoint scheme*. Taking  $N_h = 1$ , applying the Midpoint rule on the right side of (8.2.11) and by  $X(t + \frac{h}{2}) \approx \frac{1}{2}(X(t+h) + X(t))$ , we obtain the following Midpoint scheme

$$\begin{aligned} X_{n+1} &= X_n + hf \left( \frac{X_n + X_{n+1}}{2}, \frac{X_{n-m} + X_{n-m+1}}{2} \right) \\ &+ \sum_{l=1}^r g_l \left( \frac{X_n + X_{n+1}}{2}, \frac{X_{n-m} + X_{n-m+1}}{2} \right) \overline{\Delta W}_{l,n}, \quad n = 0, 1, \dots, N_T - 1, \end{aligned} \quad (8.2.15)$$

where we have truncated  $\Delta W_n$  with  $\overline{\Delta W}_n$  so that the solution to (8.2.15) has finite second-order moment and is solvable (see e.g. [301, Section 1.3]). Here  $\overline{\Delta W}_n = \zeta^{(n)} \sqrt{h}$  instead of  $\xi^{(n)} \sqrt{h}$ , where

$\zeta^{(n)}$  is a truncation of the standard Gaussian random variable  $\xi^{(n)}$  (see e.g. [301, pp. 39]):

$$\zeta^{(n)} = \xi^{(n)} \chi_{|\xi^{(n)}| \leq A_h} + \text{sgn}(\xi^{(n)}) A_h \chi_{|\xi^{(n)}| > A_h}, \quad A_h = \sqrt{4|\log(h)|}.$$

This fully implicit Midpoint scheme is symplectic if  $\tau = 0$  [299], which allows long-time integration for stochastic Hamiltonian systems. As in the case of no delay, the Midpoint scheme complies with the Stratonovich calculus without differentiating the diffusion coefficient. Again, it has first-order convergence for Stratonovich stochastic differential equations with commutative noise when no delay arises in the diffusion coefficients. However, it has half-order convergence once the delay appears in the diffusion coefficients, which will be shown numerically in Section 8.3.

The last scheme is a *Milstein-like scheme*. When  $s \in [t_n, t_{n+1}]$ , we approximate  $f(\tilde{X}(s), \tilde{X}(s-\tau))$  by  $f(\tilde{X}(t_n), \tilde{X}(t_n - \tau))$  and by the Taylor's expansion we have

$$\begin{aligned} g_l(\tilde{X}(s), \tilde{X}(s-\tau)) &\approx g_l(\tilde{X}(t_n), \tilde{X}(t_n - \tau)) + \partial_x g_l(\tilde{X}(t_n), \tilde{X}(t_n - \tau))[\tilde{X}(s) - \tilde{X}(t_n)] \\ &\quad + \partial_{x_\tau} g_l(\tilde{X}(t_n), \tilde{X}(t_n - \tau))[\tilde{X}(s-\tau) - \tilde{X}(t_n - \tau)]. \end{aligned} \quad (8.2.16)$$

Substituting the above approximations into (8.2.11) and omitting the terms whose order is higher than one in (8.2.11), we then obtain the following scheme:

$$\begin{aligned} X_{n+1} &= X_n + hf(X_n, X_{n-m}) + \sum_{l=1}^r g_l(X_n, X_{n-m}) \tilde{I}_0 \\ &\quad + \sum_{l=1}^r \sum_{q=1}^r \partial_x g_l(X_n, X_{n-m}) g_q(X_n, X_{n-m}) \tilde{I}_{q,l,t_n,t_{n+1},0} \\ &\quad + \sum_{l=1}^r \sum_{q=1}^r \partial_{x_\tau} g_l(X_n, X_{n-m}) g_q(X_{n-m}, X_{n-2m}) \chi_{t_n \geq \tau} \tilde{I}_{q,l,t_n,t_{n+1},\tau}, \quad n = 0, 1, \dots, N_T - 1, \end{aligned} \quad (8.2.17)$$

where

$$\tilde{I}_0 = \int_{t_n}^{t_{n+1}} d\tilde{W}_l(t), \quad \tilde{I}_{q,l,t_n,t_{n+1},0} = \int_{t_n}^{t_{n+1}} \int_{t_n}^t d\tilde{W}_q(s) d\tilde{W}_l(t), \quad t_n \geq 0;$$

$$\tilde{I}_{q,l,t_n,t_{n+1},\tau} = \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{t-\tau} d\tilde{W}_q(s) d\tilde{W}_l(t), \quad t_n \geq \tau.$$

Using the Fourier basis (8.2.8), the three stochastic integrals in (8.2.17) are computed by

$$\begin{aligned} \tilde{I}_0^F &= \int_{t_n}^{t_{n+1}} m_i^{(1)}(t) \xi_{l,1}^{(n)} dt = \Delta W_{l,n}, \\ \tilde{I}_{q,l,t_n,t_{n+1},0}^F &= \frac{h}{2} \xi_{q,1}^{(n)} \xi_{l,1}^{(n)} - \frac{\sqrt{2}h}{2\pi} \xi_{q,1}^{(n)} \sum_{p=1}^s \frac{1}{p} \xi_{l,2p}^{(n)} + \frac{h}{2\pi} \sum_{p=1}^{s_1} \frac{1}{p} [\xi_{q,2p+1}^{(n)} \xi_{l,2p}^{(n)} - \xi_{q,2p}^{(n)} \xi_{l,2p+1}^{(n)}], \\ \tilde{I}_{q,l,t_n,t_{n+1},\tau}^F &= \frac{h}{2} \xi_{q,1}^{(n-m)} \xi_{l,1}^{(n)} - \frac{\sqrt{2}h}{2\pi} \xi_{q,1}^{(n-m)} \sum_{p=1}^s \frac{1}{p} \xi_{l,2p}^{(n)} + \frac{h}{2\pi} \sum_{p=1}^{s_1} \frac{1}{p} [\xi_{q,2p+1}^{(n-m)} \xi_{l,2p}^{(n)} - \xi_{q,2p}^{(n-m)} \xi_{l,2p+1}^{(n)}], \end{aligned} \quad (8.2.18)$$

where  $s = \lceil \frac{N_h}{2} \rceil$  and  $s_1 = \lceil \frac{N_h-1}{2} \rceil$ . When piecewise constant basis (8.2.6) is used, these integrals are

$$\begin{aligned} \tilde{I}_0^L &= \sum_{j=0}^{N_h-1} \Delta W_{l,n,j} = \Delta W_{l,n}, \\ \tilde{I}_{q,l,t_n,t_{n+1},0}^L &= \sum_{j=0}^{N_h-1} \Delta W_{l,n,j} \left[ \frac{\Delta W_{q,n,j}}{2} + \sum_{i=0}^{j-1} \Delta W_{q,n,i} \right], \\ \tilde{I}_{q,l,t_n,t_{n+1},\tau}^L &= \sum_{j=0}^{N_h-1} \Delta W_{l,n,j} \left[ \frac{\Delta W_{q,n-m,j}}{2} + \sum_{i=0}^{j-1} \Delta W_{q,n-m,i} \right], \end{aligned} \quad (8.2.19)$$

where  $\Delta W_{k,n,j} = W_k(t_n + \frac{(j+1)h}{N_h}) - W_k(t_n + \frac{jh}{N_h})$ ,  $k = 1, \dots, r$ ,  $j = 0, \dots, N_h - 1$  and  $\Delta W_{k,n,-1} = 0$ .

In Example 8.3.3, Section 8.3, we will show that the piecewise linear interpolation is less efficient than the Fourier approximation for achieving the same order of accuracy.

The scheme (8.2.17) can be seen as further discretization of the Milstein scheme for Stratonovich SDDs proposed in [195]:

$$\begin{aligned} X_{n+1}^M &= X_n^M + hf(X_n^M, X_{n-m}^M) + \sum_{l=1}^r g_l(X_n^M, X_{n-m}^M) \Delta W_{l,n} \\ &+ \sum_{l=1}^r \sum_{q=1}^r \partial_x g_l(X_n^M, X_{n-m}^M) g_q(X_n^M, X_{n-m}^M) I_{q,l,t_n,t_{n+1},0} \\ &+ \sum_{l=1}^r \sum_{q=1}^r \partial_{x_\tau} g_l(X_n^M, X_{n-m}^M) g_q(X_{n-m}^M, X_{n-2m}^M) \chi_{t_n \geq \tau} I_{q,l,t_n,t_{n+1},\tau}, \quad n = 0, 1, \dots, N_T - 1. \end{aligned} \quad (8.2.20)$$

as the double integrals approximated by either the Fourier expansion or the piecewise linear interpolation:  $\tilde{I}_0$ ,  $\tilde{I}_{q,l,t_n,t_{n+1},0}$ , and  $\tilde{I}_{q,l,t_n,t_{n+1},\tau}$  are, respectively, approximation of the following integrals :

$$\begin{aligned} I_0 &= \int_{t_n}^{t_{n+1}} \circ dW_l(t), & I_{q,l,t_n,t_{n+1},0} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^t \circ dW_q(s) \circ dW_l(t), \quad t_n \geq 0 \\ I_{q,l,t_n,t_{n+1},\tau} &= \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{t-\tau} \circ dW_q(s) \circ dW_l(t), \quad t_n \geq \tau. \end{aligned}$$

Actually, we have the following relations.

**Lemma 8.2.5.** *For the Fourier basis (8.2.8), it holds that*

$$\tilde{I}_0^F = I_0, \quad (8.2.21)$$

$$\mathbb{E}[(\tilde{I}_{q,l,t_n,t_{n+1},0}^F - I_{q,l,t_n,t_{n+1},0})^2] = \varsigma(N_h) \frac{2\Delta^2}{(N_h\pi)^2} + \sum_{i=M}^{\infty} \frac{\Delta^2}{(i\pi)^2} \leq c \frac{\Delta^2}{\pi^2 M}, \quad (8.2.22)$$

$$\mathbb{E}[(\tilde{I}_{q,l,t_n,t_{n+1},\tau}^F - I_{q,l,t_n,t_{n+1},\tau})^2] = \varsigma(N_h) \frac{2\Delta^2}{(N_h\pi)^2} + \sum_{i=M}^{\infty} \frac{\Delta^2}{(i\pi)^2} \leq c \frac{\Delta^2}{\pi^2 M}, \quad (8.2.23)$$

where  $\varsigma(N_h) = 0$  if  $N_h$  is odd and 1 otherwise, and  $M$  is the integer part of  $(N_h + 1)/2$ .

The proof of this lemma can be found in Section 4. With Lemma 8.2.5, we can show that the Milstein-like scheme (8.2.17) can be of first-order convergence in the mean-square sense, see Section 8.4.

**Theorem 8.2.6.** *Assume that  $f$ ,  $g_l$ ,  $\partial_x g_l g_q$  and  $\partial_{x_\tau} g_l g_q$  ( $l, q = 1, 2, \dots, r$ ) satisfy the Lipschitz condition (8.2.2) and also  $g_l$  have bounded second-order partial derivatives with respect to all variables. If  $\mathbb{E}[\|\phi\|_{L^\infty}^p] < \infty$ ,  $p \geq 4$ , then we have for the Milstein-like scheme (8.2.17),*

$$\max_{1 \leq n \leq N_T} \mathbb{E}|X(t_n) - X_n|^2 = O(h^2), \quad (8.2.24)$$

when the double integrals  $\tilde{I}_{q,l,t_n,t_{n+1},0}$ ,  $\tilde{I}_{q,l,t_n,t_{n+1},\tau}$  are computed by (8.2.18) and  $N_h$  is at the order

of  $1/h$ .

When (8.2.19) is used in the Milstein-like scheme (8.2.17), the first-order strong convergence can be proved similarly when  $N_h$  is at the order of  $1/h$ .

### 8.3 Numerical results

In this section, we test the convergence rate of the proposed schemes and compare their numerical performance. In the first two examples, we test the Predictor-Corrector scheme (8.2.13) and Midpoint scheme (8.2.15) for multiple noises and show both methods are of half-order mean-square convergence. Further, we show that both schemes converge with order one in the mean-square sense for a SDDE with single white noise and no time delay in diffusion coefficients. In the last example, we investigate the Milstein-like scheme (8.2.17) and show that it is first-order convergent for SDDEs with multiple white noises.

Throughout this section, the strong error of numerical solutions is defined as

$$\rho_{h,T} = \left( \frac{1}{n_p} \sum_{i=1}^{n_p} |X_h(T, \omega_i) - X_{\frac{h}{2}}(T, \omega_i)|^2 \right)^{1/2},$$

where  $\omega_i$  denotes the  $i$ -th single sample path and  $n_p$  is the number of paths.

The numerical tests were performed using Matlab R2012a on a Dell Optiplex 780 computer with CPU (E8500 3.16 GHz). we used the Mersenne twister random generator with seed 1 and took a large number of paths so that the statistical error can be ignored. Newton's method with tolerance  $h^2/100$  was used to solve the nonlinear algebraic equations at each step of the implicit schemes.

We first test the convergence rate of Predictor-Corrector scheme (8.2.13) and Midpoint scheme (8.2.15) for a SDDE with several noises.

**Example 8.3.1.** Consider Equation (8.2.1) with the following coefficients

$$\begin{aligned}
 f &= -15X(t) + 2\sin(X(t - \tau)); \\
 g_1 &= \sin(X(t)) + 0.5X(t - \tau); \quad g_2 = 0.9X(t); \quad g_3 = 0.2X(t) + 0.2X(t - \tau); \\
 g_4 &= 2\sin(X(t)); \quad g_5 = 0.8X(t) + \cos(X(t - \tau)); \quad g_6 = X(t) + 0.5\sin(X(t - \tau)); \\
 g_7 &= 2\cos(X(t - \tau)); \quad g_8 = -X(t) + \cos(X(t - \tau)), \quad g_9 = 0.5X(t) - X(t - \tau); \\
 g_{10} &= 1.5\cos(X(t - \tau))
 \end{aligned}$$

and the initial function is  $\phi(t) = t + 0.2$ .

In this example, we test the convergence order of the Predictor-Corrector and Midpoint schemes at  $T = 20$  and with different time delays  $\tau = 2^{-4}, 2^{-2}, 1$ .

In Table 8.1, we observe that both schemes are convergent with order half in the mean-square sense. Different time delays do not influence the convergence rate of these schemes.

The amount of operations of the Predictor-Corrector scheme for Equation 8.2.1 is  $(5r + 6)dT/h$ . For the Midpoint scheme, the amount of operations is  $6C_I(r + 1)dT/h$ , where  $C_I$  is the maximum number of Newton's iterations in each time step. In Table 8.1, we observe that for both schemes, the computational cost doubles when step sizes reduce by half. The CPU time of the Midpoint scheme is about four times of what the Predictor-Corrector scheme costs, which is consistent with the prediction as the observed  $C_I$  is around 4.

We now test the convergence rate for the Predictor-Corrector scheme (8.2.13) and the Midpoint scheme (8.2.15) for SDDEs with different types of noises: non-commutative noise, single noise. We will show that the time delay in a diffusion coefficient keeps both methods only convergent at half-order, while for the SDDE with single noise, the two schemes can be of first-order accuracy in the mean-square sense if the time delay does not appear explicitly in the diffusion coefficients.

---

<sup>1</sup>We write ‘\*’ whenever no results from a smaller time step size are available and the convergence order is absent. This rule applies throughout this section.

Table 8.1: Convergence rate of Predictor-Corrector scheme (left) and Midpoint scheme (right) for Example 8.3.1 at  $T = 20$  with ten white noises using  $n_p = 4000$  sample paths.

$\tau$	$h$	$\rho_{h,T}$	order	time (s.)	$\rho_{h,T}$	order	time (s.)
$\frac{1}{16}$	$2^{-8}$	4.050e-02	0.51	220	6.698e-02	0.61	807
	$2^{-9}$	2.851e-02	0.56	420	4.378e-02	0.44	1591
	$2^{-10}$	1.936e-02	0.48	818	3.237e-02	0.53	3157
	$2^{-11}$	1.390e-02	*	1620	2.239e-02	*	6289
	$2^{-12}$	* 1	*	3221	*	*	12573
$\frac{1}{4}$	$2^{-8}$	3.955e-02	0.54	219	5.942e-02	0.51	805
	$2^{-9}$	2.728e-02	0.55	417	4.189e-02	0.48	1586
	$2^{-10}$	1.861e-02	0.46	816	3.000e-02	0.47	3148
	$2^{-11}$	1.359e-02	*	1615	2.164e-02	*	6284
	$2^{-12}$	*	*	3215	*	*	12557
1	$2^{-8}$	3.914e-02	0.53	221	5.725e-02	0.48	805
	$2^{-9}$	2.711e-02	0.54	422	4.092e-02	0.51	1588
	$2^{-10}$	1.873e-02	0.46	820	2.887e-02	0.47	3148
	$2^{-11}$	1.364e-02	*	1627	2.086e-02	*	6289
	$2^{-12}$	*	*	3236	*	*	12569

**Example 8.3.2.** Consider Equation (8.2.1) in one-dimension and assume the initial function

$\phi(t) = t + 0.2$ , with different diffusion terms:

- non-commutative white noises without delay in the diffusion coefficients:

$$dX = [-X(t) + \sin(X(t - \tau))] dt + \sin(X(t)) \circ dW_1(t) + 0.5X(t) \circ dW_2(t), \quad (8.3.1)$$

where the noises are non-commutative as  $\partial_x(\sin(x))0.5x - \partial_x(0.5x)\sin(x) \neq 0$ ;

- commutative (single) white noises without delay in the diffusion coefficients:

$$dX = [-X(t) + \sin(X(t - \tau))] dt + \sin(X(t)) \circ dW(t); \quad (8.3.2)$$

- commutative (single) white noises with delay in the diffusion coefficients:

$$dX = [-X(t) + \sin(X(t - \tau))] dt + \sin(X(t - \tau)) \circ dW(t). \quad (8.3.3)$$

From Figure 8.1(a) (non-commutative noises, Equation (8.3.1)) and Figure 8.1(c) (single delayed

diffusion, Equation (8.3.3)), we observe the half-order strong convergence. In contrast, for Equation (8.3.2) (single noise, non-delayed diffusion) in Figure 8.1 (b), the convergence order of these two schemes becomes one in the mean-square sense.

From this example, we conclude that for the Predictor-Corrector and Midpoint schemes, when the time delay only appears in the drift term, the convergence order is one for the equation with commutative noises and half for the one with non-commutative noises. However, when the diffusion coefficients contain time delays, these two schemes are only half-order even for equations with a single white noise.

In the last example, we test the Milstein-like scheme (8.2.17) using different bases, i.e., piecewise constant basis (8.2.6) and Fourier approximation (8.2.8), and compare its numerical performance with Predictor-Corrector and Midpoint schemes. For the Milstein-like scheme, we show that for multiple noises, the computational cost for achieving the same accuracy is much higher than the other two schemes; while for single noise, the computational cost for the same accuracy is lower.

**Example 8.3.3.** *We consider the Milstein-like scheme (8.2.17) for*

$$\begin{aligned} dX(t) &= [-9X(t) + \sin(X(t - \tau))]dt + [\sin(X(t)) + X(t - \tau)] \circ dW_1(t) \\ &\quad + [X(t) + \cos(0.5X(t - \tau))] \circ dW_2(t), \quad t \in (0, T] \\ X(t) &= t + \tau + 0.1, \quad t \in [-\tau, 0] \end{aligned} \tag{8.3.4}$$

and

$$\begin{aligned} dX(t) &= [-2X(t) + 2X(t - \tau)]dt + [\sin(X(t)) + X(t - \tau)] \circ dW(t), \quad t \in (0, T] \\ X(t) &= t + \tau, \quad t \in [-\tau, 0] \end{aligned} \tag{8.3.5}$$

To reduce the computational cost, the double integrals are computed by the Fourier expansion

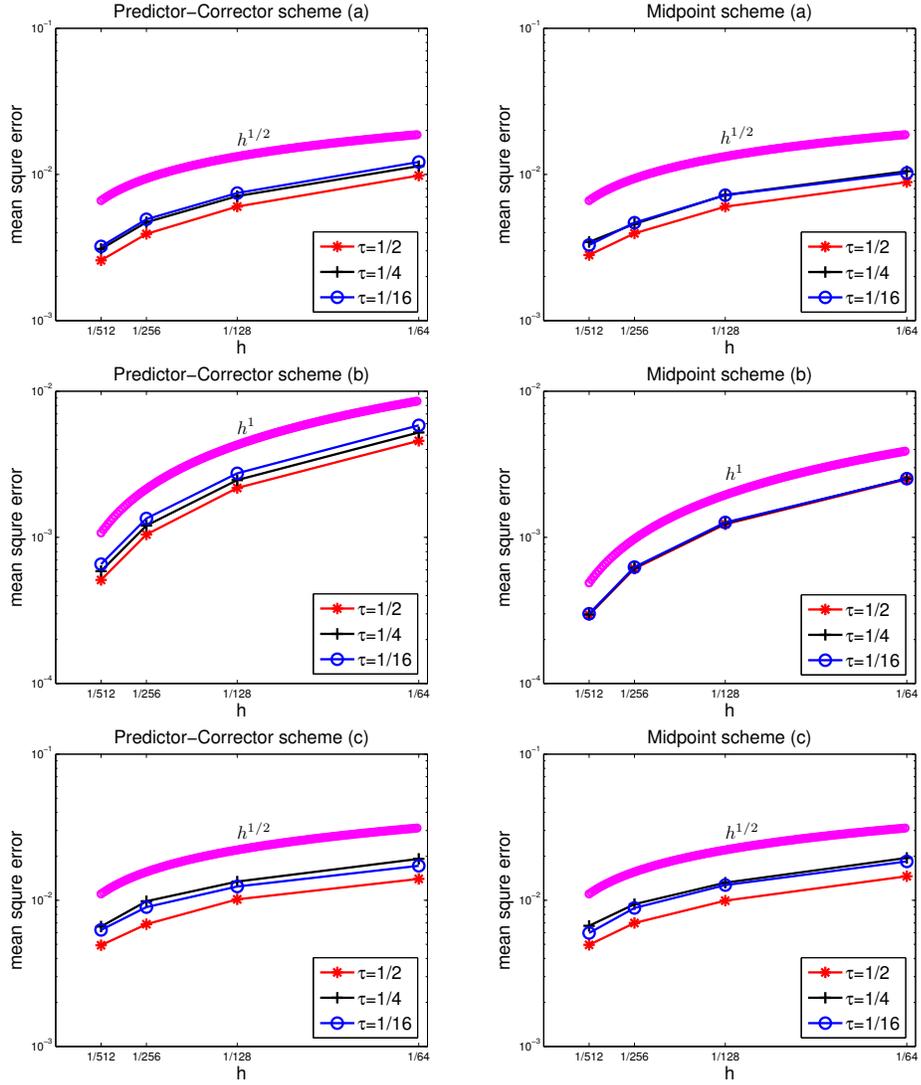


Figure 8.1: Mean-square convergence test of the Predictor-Corrector (left column) and Midpoint schemes (right column) on Example (8.3.2) at  $T=5$  with different  $\tau$  using  $n_p = 10000$  sample paths. (a): multi white noises with non-delayed diffusion term; (b): single white noise with non-delayed diffusion term; (c): single white noise with delayed diffusion term.

approximation (8.2.18) and the following relation

$$\tilde{I}_{q,l,t_n,t_{n+1},0} = \Delta W_{l,n} \Delta W_{q,n} - \tilde{I}_{l,q,t_n,t_{n+1},0}, \quad \tilde{I}_{l,l,t_n,t_{n+1},0} = \frac{(\Delta W_{l,n})^2}{2}. \quad (8.3.6)$$

We also use the following relations

$$\begin{aligned}\tilde{I}_{q,l,t_n,t_n+ph,0} &= \sum_{j=0}^{p-1} \left[ \tilde{I}_{q,l,t_n+jh,t_n+(j+1)h,0} + \Delta W_{l,n+j} \chi_{j \geq 1} \sum_{i=0}^{j-1} \Delta W_{q,n+i} \right], \\ \tilde{I}_{q,l,t_n,t_n+ph,\tau} &= \sum_{j=0}^{p-1} \left[ \tilde{I}_{q,l,t_n+jh,t_n+(j+1)h,\tau} + \Delta W_{l,n+j} \chi_{j \geq 1} \sum_{i=0}^{j-1} \Delta W_{q,n-m+i} \right].\end{aligned}$$

In Table 8.2, we show that for Equation (8.3.4), the Milstein-like scheme (8.2.17) converges with order one in the mean-square sense. Compared to the Predictor-Corrector scheme or the Midpoint scheme, when the time step sizes are the same, the computational cost for the Milstein-like scheme (8.2.17) is several times higher. In fact, in the Milstein-like scheme, the extra computational cost comes from evaluating the double integrals  $\tilde{I}_{q,l,t_n,t_{n+1},0}^F$  and  $\tilde{I}_{q,l,t_n,t_{n+1},\tau}^F$  at each time step, which requires  $7/(2h)(3r^2 - r)/2$  operations when we take the relation (8.3.6) into account.

We also test the Milstein-like scheme (8.2.17) using the piecewise constant basis (8.2.6). The computational cost is even higher than that of using the Fourier basis for the same time step size. Actually, the amount of operations for evaluating double integrals using (8.2.19) is  $(1/(2h^2) + 5/(2h) - 1)(3r^2 - r)/2$ , which is  $O(1/h^2)$ , much more than that of using the Fourier basis,  $O(1/h)$ . Our numerical tests (not presented here) confirmed the fast increase of amount of operations.

Table 8.2: Convergence rate of the Milstein-like scheme (left) for Equation (8.3.4) at  $T = 1$  and comparison with the convergence rate of the Predictor-Corrector scheme (middle) and the Midpoint scheme (right) using  $n_p = 4000$  sample paths. The upper rows are with  $\tau = 1/16$  and the lower are with  $\tau = 1/4$ .

$h$	$\rho_{h,T}$	order	time (s.)	$\rho_{h,T}$	order	time (s.)	$\rho_{h,T}$	order	time (s.)
$2^{-4}$	9.832e-02	1.27	0.72	7.164e-02	0.94	0.05	5.000e-02	0.60	0.16
$2^{-5}$	4.090e-02	1.09	1.0	3.734e-02	0.69	0.10	3.304e-02	0.55	0.29
$2^{-6}$	1.921e-02	0.99	1.7	2.308e-02	0.51	0.12	2.263e-02	0.51	0.41
$2^{-7}$	9.703e-03	*	3.3	1.616e-02	*	0.25	1.590e-02	*	0.79
$2^{-8}$	*	*	6.4	*	*	0.40	*	*	1.54
$2^{-4}$	9.307e-02	1.28	0.56	6.956e-02	0.96	0.04	5.050e-02	0.68	0.11
$2^{-5}$	3.824e-02	1.08	0.93	3.582e-02	0.70	0.10	3.155e-02	0.56	0.22
$2^{-6}$	1.804e-02	0.99	1.6	2.205e-02	0.62	0.17	2.133e-02	0.58	0.39
$2^{-7}$	9.069e-03	*	2.8	1.434e-02	*	0.26	1.425e-02	*	0.78
$2^{-8}$	*	*	5.5	*	*	0.45	*	*	1.59

However, the amount of operations of the Milstein-like scheme can be significantly reduced when there is just a single diffusion term. In Table 8.3, we observe that the Milstein-like scheme

for Equation (8.3.5) is still of first-order convergence but the Predictor-Corrector scheme and the Midpoint scheme is only of half-order convergence. For the same accuracy, the computational cost for the Milstein-like scheme using the Fourier basis is less than that for the other two schemes. In fact, for single noise, we only need to compute one double integral  $\tilde{I}_{1,1,t_n,t_{n+1},\tau}$ . Moreover, when the coefficients of the diffusion term are small, a small number of Fourier modes is required for large time step sizes, i.e.,  $N_h$  can be  $O(1)$  instead of  $O(1/h)$ . The computational cost can thus be reduced somewhat, see e.g. [301, Chapter 3] for such a discussion for equations with small noises without delay.

Table 8.3: Convergence rate of the Milstein-like scheme (left) for Equation (8.3.5) (single white noise) at  $T = 1$  and comparison with the convergence rate of the Predictor-Corrector scheme (middle) and Midpoint scheme (right) using  $n_p = 4000$  sample paths. The delay  $\tau$  is taken as  $1/4$ .

$h$	$\rho_{h,T}$	order	time (s.)	$h$	$\rho_{h,T}$	order	time (s.)	$\rho_{h,T}$	order	time (s.)
$2^{-4}$	3.164e-02	0.91	0.19	$2^{-7}$	1.252e-02	0.44	0.18	1.263e-02	0.45	0.59
$2^{-5}$	1.688e-02	0.99	0.28	$2^{-8}$	9.219e-03	0.51	0.37	9.246e-03	0.51	1.09
$2^{-6}$	8.499e-03	0.90	0.46	$2^{-9}$	6.462e-03	0.49	0.56	6.471e-03	0.48	2.05
$2^{-7}$	4.570e-03	*	0.79	$2^{-10}$	4.617e-03	*	1.03	4.627e-03	*	3.97
$2^{-8}$	*	*	1.40	$2^{-11}$	*	*	1.91	*	*	7.58

In summary, the proposed Predictor-Corrector scheme and Midpoint scheme are convergent with half-order in the mean-square sense, see Example 8.3.1. We also show that these two schemes can be of first-order in the mean-square sense if the underlying SDDEs with single noise (commutative noise) and the time delay is only in the drift coefficients, see Example 8.3.2. In Example 8.3.3 the numerical tests show that our proposed Milstein-like scheme is of first-order in the mean-square sense for SDDEs with non-commutative noise wherever the time delay appears, i.e. in the drift and/or diffusion coefficients. Compared to the other two schemes, the Milstein-like scheme is more accurate but is more expensive as it requires evaluations of double integrals, with cost inversely proportional to the time step size and proportional to the square of number of noises. However, for SDDEs with single noise, the Milstein-like scheme (with the Fourier basis) can be superior to the Predictor-Corrector scheme and the Midpoint scheme both in terms of accuracy and computational cost.

## 8.4 Proofs

In this section, we prove Theorems 8.2.3 and 8.2.6 and Lemma 8.2.5. While proofs of Theorems 8.2.3 and 8.2.6 are presented only for the one-dimensional problem (8.2.1) ( $d = 1$ ), they can be extended to multi-dimensional case  $d > 1$  without difficulty.

**Proof of Theorem 8.2.3.** We recall that for the Milstein scheme (8.2.20), see [195],

$\max_{1 \leq n \leq N_T} \mathbb{E}|X(t_n) - X_n^M|^2 = O(h^2)$ . Then by the triangle inequality, it suffices to prove

$$\max_{1 \leq n \leq N_T} \mathbb{E}|X_n^M - X_n|^2 = O(h). \quad (8.4.1)$$

We denote that  $f_n = f(X_n, X_{n-m})$  and  $g_{l,n} = g_l(X_n, X_{n-m})$  and also

$$\begin{aligned} \rho_{f_n} &= f(\bar{X}_{n+1}, X_{n-m+1}) - f_n, \\ \rho_{g_{l,n}} &= g_l(\bar{X}_{n+1}, X_{n-m+1}) - [g_{l,n} + \partial_x g_{l,n} \sum_{q=1}^r g_{q,n} \Delta W_{q,n} + \partial_{x_\tau} g_{l,n} \sum_{q=1}^r g_{q,n-m} \Delta W_{q,n-m}]. \end{aligned} \quad (8.4.2)$$

With (8.4.2), we can rewrite (8.2.13) as follows

$$\begin{aligned} X_{n+1} &= X_n + h f_n + \sum_{l=1}^r g_{l,n} \Delta W_{l,n} + \frac{1}{2} \sum_{l=1}^r \sum_{q=1}^r \partial_x g_{l,n} \Delta W_{q,n} \Delta W_{l,n} \\ &\quad + \frac{1}{2} \sum_{l=1}^r \sum_{q=1}^r \partial_{x_\tau} g_{l,n} g_{q,n-m} \Delta W_{q,n-m} \Delta W_{l,n} + \rho_n, \end{aligned} \quad (8.4.3)$$

where  $\rho_n = h \rho_{f_n} + \frac{1}{2} \sum_{l=1}^r \rho_{g_{l,n}} \Delta W_{l,n}$ .

It can be readily checked that if  $f, g_l$  satisfy the Lipschitz condition (8.2.2), and  $g_l$  has bounded second-order derivatives ( $l = 1, \dots, r$ ), then by the Predictor-Corrector scheme (8.2.13) and Taylor's expansion of  $g_l(\bar{X}_{n+1}, X_{n-m+1})$ , we have  $h^2 \mathbb{E}[\rho_{f_n}^2] \leq Ch^3$ ,  $\mathbb{E}[(\rho_{g_{l,n}} \Delta W_{l,n})^2] \leq Ch^3$ , and thus by the triangle inequality,

$$\mathbb{E}[\rho_n^2] \leq Ch^3, \quad (8.4.4)$$

where the constant  $C$  depends on  $r$  and Lipschitz constants, independent of  $h$ .

Subtracting (8.4.3) from (8.2.20) and taking expectation after squaring over both sides, we have

$$\begin{aligned}\mathbb{E}[(X_{n+1}^M - X_{n+1})^2] &= \mathbb{E}[(X_n^M - X_n)^2] + 2\mathbb{E}[(X_n^M - X_n)\left(\sum_{i=0}^4 R_i - \rho_n\right)] \\ &\quad - 2\sum_{i=0}^4 \mathbb{E}[\rho_n R_i] + \sum_{i,j=0}^4 \mathbb{E}[R_i R_j] + \mathbb{E}[\rho_n^2],\end{aligned}\tag{8.4.5}$$

where we denote  $f_n^M = f(X_n^M, X_{n-m}^M)$  and  $g_{l,n}^M = g_l(X_n^M, X_{n-m}^M)$  and

$$\begin{aligned}R_0 &= h(f_n^M - f_n) + \sum_{l=1}^r (g_{l,n}^M - g_{l,n})\Delta W_{l,n}, \\ R_1 &= \sum_{l=1}^r \sum_{q=1}^r [\partial_x g_{l,n}^M g_{q,n}^M - \partial_x g_{l,n} g_{q,n}] \frac{\Delta W_{q,n} \Delta W_{l,n}}{2}, \\ R_2 &= \sum_{l=1}^r \sum_{q=1}^r [\partial_{x_\tau} g_{l,n}^M g_{q,n-m}^M - \partial_{x_\tau} g_{l,n} g_{q,n-m}] \frac{\Delta W_{q,n-m} \Delta W_{l,n}}{2}, \\ R_3 &= \sum_{l=1}^r \sum_{q=1}^r \partial_x g_{l,n}^M g_{q,n}^M (I_{q,l,t_n,t_{n+1},0} - \frac{\Delta W_{q,n} \Delta W_{l,n}}{2}), \\ R_4 &= \sum_{l=1}^r \sum_{q=1}^r \partial_{x_\tau} g_{l,n}^M g_{q,n-m}^M (I_{q,l,t_n,t_{n+1},\tau} - \frac{\Delta W_{q,n-m} \Delta W_{l,n}}{2}).\end{aligned}$$

By the Lipschitz condition for  $f$  and  $g_l$ , and adaptedness of  $X_n, X_n^M$ , we have

$$\mathbb{E}[R_0^2] \leq C(h^2 + h)(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]).\tag{8.4.6}$$

To bound  $\mathbb{E}[R_i^2]$  ( $i = 1, 2, 3, 4$ ), we require that  $X_n$  and  $X_n^M$  have bounded (up to) fourth-order moments, which can be readily checked from the Predictor-Corrector scheme (8.2.13) and the Milstein scheme (8.2.20) under our assumptions. By the Lipschitz condition of  $g_l$  and  $\partial_{x_\tau} g_l g_q$ , we have

$$\mathbb{E}[R_2^2] \leq C \max_{1 \leq l, q \leq r} \mathbb{E}[(|X_n^M - X_n| + |X_{n-m}^M - X_{n-m}|)^2 (\Delta W_{q,n-m} \Delta W_{l,n})^2],$$

whence by Cauchy inequality and the boundedness of  $\mathbb{E}[X_n^4]$  and  $\mathbb{E}[(X_n^M)^4]$ , we have  $\mathbb{E}[R_2^2] \leq Ch^2$ . Similarly, we have  $\mathbb{E}[R_1^2] \leq Ch^2$ . By Lemma 8.2.5, and linear growth condition (8.2.3) for  $\partial_{x_\tau} g_l g_q$ , we obtain

$$\mathbb{E}[R_4^2] \leq C \max_{1 \leq l < q \leq r} \mathbb{E}[(1 + |X_n^M|^2 + |X_{n-m}^M|^2)(I_{q,l,t_n,t_{n+1},\tau} - \frac{\Delta W_{q,n-m} \Delta W_{l,n}}{2})^2] \leq Ch^2,$$

since  $X_n^M, X_{n-m}^M$  have bounded fourth-order moments and by the Burkholder-Davis-Gundy inequality, it holds that for  $l \neq q$

$$\begin{aligned} & \mathbb{E}[(I_{q,l,t_n,t_{n+1},\tau} - \frac{\Delta W_{q,n-m} \Delta W_{l,n}}{2})^4] \\ = & \mathbb{E}[(\int_{t_n}^{t_{n+1}} (W_q(t-\tau) - \frac{W_q(t_{n+1}-\tau) + W_q(t_n-\tau)}{2}) \circ dW_l)^4] \\ \leq & C(\mathbb{E}[\int_{t_n}^{t_{n+1}} (W_q(t-\tau) - \frac{W_q(t_{n+1}-\tau) + W_q(t_n-\tau)}{2})^2 ds])^2 \leq Ch^4. \end{aligned}$$

Similarly, we have  $\mathbb{E}[R_3^2] \leq Ch^2$ . Thus we have proved that

$$\mathbb{E}[R_i^2] \leq Ch^2, \quad i = 1, 2, 3, 4. \quad (8.4.7)$$

By the basic inequality  $2ab \leq a^2 + b^2$ , we have

$$2 |\mathbb{E}[(X_n^M - X_n)\rho_n]| \leq h\mathbb{E}[(X_n^M - X_n)^2] + h^{-1}\mathbb{E}[\rho_n^2]. \quad (8.4.8)$$

By the fact that  $X_n$  and  $X_n^M$  are  $\mathcal{F}_{t_n}$ -measurable and Lipschitz condition for  $f$ ,

$$\begin{aligned} 2\mathbb{E}[(X_n^M - X_n)R_0] &= 2h\mathbb{E}[(X_n^M - X_n)(f_n - f_{n-m})] \\ &\leq Ch(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]). \end{aligned} \quad (8.4.9)$$

Further, by the Lipschitz condition (8.2.2) for  $\partial_x g_l g_l$ , we have

$$\begin{aligned} 2\mathbb{E}[(X_n^M - X_n)R_1] &= \sum_{l=1}^r \mathbb{E}[(X_n^M - X_n)\partial_x g_{l,n}^M g_{l,n}^M - \partial_x g_{l,n} g_{l,n}] \mathbb{E}[(\Delta W_{l,n})^2] \\ &\leq Ch(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]). \end{aligned} \quad (8.4.10)$$

By the adaptedness of  $X_n$ ,  $X_n^M$  and  $\mathbb{E}[\Delta W_{l,n}] = \mathbb{E}[(I_{q,l,t_n,t_{n+1},0} - \frac{\Delta W_{q,n} \Delta W_{l,n}}{2})] = 0$ , we have

$$\mathbb{E}[(X_n^M - X_n)R_i] = 0, \quad i = 2, 3. \quad (8.4.11)$$

Again by the adaptedness of  $X_n$  and  $X_n^M$ , we can have

$$\mathbb{E}[(X_n^M - X_n)R_4] = 0. \quad (8.4.12)$$

In fact, by Lemma 8.2.5, we can represent  $I_{q,l,t_n,t_{n+1},\tau}$  as

$$I_{q,l,t_n,t_{n+1},\tau} = \frac{h}{2} \xi_{q,1}^{(n-m)} \xi_{l,1}^{(n)} + \frac{h}{2\pi} \sum_{p=1}^{\infty} \frac{1}{p} [\xi_{q,2p+1}^{(n)} \xi_{l,2p}^{(n-m)} - \xi_{q,2p}^{(n-m)} \xi_{l,2p+1}^{(n)} - \sqrt{2} \xi_{q,1}^{(n-m)} \xi_{l,2p}^{(n)}]. \quad (8.4.13)$$

Then by the facts  $\mathbb{E}[|(X_n^M - X_n)R_4|] \leq (\mathbb{E}[(X_n^M - X_n)^2])^{1/2} (\mathbb{E}[R_4^2])^{1/2} \leq Ch$  and  $\mathbb{E}[(X_n^M - X_n)\xi_{l,k}^{(n)}] = 0$  for any  $k \geq 1$ , we obtain (8.4.12) from Lebesgue's dominated convergence theorem.

With (8.4.11)-(8.4.12) and Cauchy inequality, from (8.4.5) we have, for  $n \geq m$ ,

$$\begin{aligned} &\mathbb{E}[(X_{n+1}^M - X_{n+1})^2] \\ &\leq \mathbb{E}[(X_n^M - X_n)^2] + 2\mathbb{E}[(X_n^M - X_n)(R_0 + R_1 - \rho_n)] + C \sum_{i=0}^4 \mathbb{E}[R_i^2] + C\mathbb{E}[\rho_n^2] \end{aligned}$$

and further by (8.4.4), (8.4.6)-(8.4.8), and (8.4.9)-(8.4.10), we obtain, for  $n \geq m$ ,

$$\mathbb{E}[X_{n+1}^M - X_{n+1}]^2$$

$$\begin{aligned}
&\leq (1 + Ch)\mathbb{E}[(X_n^M - X_n)^2] + Ch\mathbb{E}[(X_{n-m}^M - X_{n-m})^2] + (C + h^{-1})\mathbb{E}[\rho_n^2] + C \sum_{i=0}^4 \mathbb{E}[R_i^2] \\
&\leq (1 + Ch)\mathbb{E}[(X_n^M - X_n)^2] + Ch\mathbb{E}[(X_{n-m}^M - X_{n-m})^2] + Ch^2,
\end{aligned} \tag{8.4.14}$$

where  $C$  is independent of  $h$ . Similarly, we can also obtain that (8.4.14) holds for  $n = 1, \dots, m-1$ .

Taking the maximum over both sides of (8.4.14) and noting that  $X_i^M - X_i = 0$  for  $-m \leq i \leq 0$ ,

we have

$$\max_{1 \leq i \leq n+1} \mathbb{E}[(X_i^M - X_i)^2] \leq (1 + Ch) \max_{1 \leq i \leq n} \mathbb{E}[(X_i^M - X_i)^2] + Ch^2.$$

Then (8.4.1) follows from discrete Gronwall inequality.  $\square$

**Proof of Lemma 8.2.5.** From (8.2.18), the first formula (8.2.21) can be readily obtained.

Now we consider (8.2.22). For  $l = q$ , it holds that

$$\tilde{I}_{l,l,t_n,t_{n+1},0} = I_{l,l,t_n,t_{n+1},0} = (\Delta W_{l,n})^2/2,$$

if (8.2.5) with piecewise constant basis (8.2.6) or Fourier basis (8.2.8) is used. For any orthogonal expansion (8.2.4), we have  $\mathbb{E}[\int_{t_n}^{t_{n+1}} (\widetilde{W}_q(s) - W_q(s)) dW_l \int_{t_n}^{t_{n+1}} \widetilde{W}_q(s) d(\widetilde{W}_l - W_l)] = 0$  and thus by

$W_q(t_n) = \widetilde{W}_q(t_n)$ , Ito's isometry and integral by parts, we have, when  $l \neq q$ ,

$$\begin{aligned}
&\mathbb{E}[(\tilde{I}_{q,l,t_n,t_{n+1},0} - I_{q,l,t_n,t_{n+1},0})^2] \\
&= \mathbb{E}[(\int_{t_n}^{t_{n+1}} [\widetilde{W}_q(s) - W_q(s)] \circ dW_l + \int_{t_n}^{t_{n+1}} \widetilde{W}_q(s) d[\widetilde{W}_l - W_l])^2] \\
&= \mathbb{E}[(\int_{t_n}^{t_{n+1}} [\widetilde{W}_q(s) - W_q(s)] dW_l)^2] + \mathbb{E}[(\int_{t_n}^{t_{n+1}} \widetilde{W}_q(s) d[\widetilde{W}_l - W_l])^2] \\
&= \int_{t_n}^{t_{n+1}} \mathbb{E}[(\widetilde{W}_q(s) - W_q(s))^2] ds + \mathbb{E}[(\int_{t_n}^{t_{n+1}} [\widetilde{W}_l - W_l] d\widetilde{W}_q(s))^2].
\end{aligned}$$

Then by the mutual independence of all Gaussian random variables  $\xi_{q,i}^{(n)}$ ,  $i = 1, 2, \dots, q = 1, 2, \dots, r$ ,

we obtain  $\mathbb{E}[(\widetilde{W}_q(s) - W_q(s))^2] = \sum_{i=N_h+1}^{\infty} M_i^2(s)$ , where  $M_i(s) = \int_{t_n}^s m_i(\theta) d\theta$  and for  $l \neq q$ ,

$$\begin{aligned} \mathbb{E}\left[\left(\int_{t_n}^{t_{n+1}} [\widetilde{W}_l(s) - W_l(s)] d\widetilde{W}_q\right)^2\right] &= \mathbb{E}\left[\left(\sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \int_{t_n}^{t_{n+1}} M_i(s) m_j(s) ds \xi_{l,i}^{(n)} \xi_{q,j}^{(n)}\right)^2\right] \\ &= \sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \left(\int_{t_n}^{t_{n+1}} M_i(s) m_j(s) ds\right)^2. \end{aligned}$$

Then we have

$$\begin{aligned} &\mathbb{E}[(\widetilde{I}_{q,l,t_n,t_{n+1},0} - I_{q,l,t_n,t_{n+1},0})^2] \\ &= \sum_{i=N_h+1}^{\infty} \int_{t_n}^{t_{n+1}} M_i^2(s) ds + \sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \left(\int_{t_n}^{t_{n+1}} M_i(s) m_j(s) ds\right)^2. \end{aligned} \quad (8.4.15)$$

In (8.4.15), we consider the Fourier basis (8.2.8). Since  $M_i(s)$  ( $i \geq 2$ ) are also sine or cosine functions, we have

$$\sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \left(\int_{t_n}^{t_{n+1}} M_i(s) m_j(s) ds\right)^2 = \left(\int_{t_n}^{t_{n+1}} M_{N_h+1}(s) m_{N_h}(s) ds\right)^2 \quad (8.4.16)$$

when  $N_h$  is even and  $\sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \left(\int_{t_n}^{t_{n+1}} M_i(s) m_j(s) ds\right)^2 = 0$  when  $N_h$  is odd. Moreover, for  $i \geq 2$ , it holds from simple calculations that

$$\int_{t_n}^{t_{n+1}} M_i^2(s) ds = \frac{3\Delta^2}{(2\lfloor i/2 \rfloor \pi)^2}, \quad i \text{ is even and } \frac{\Delta^2}{(2\lfloor i/2 \rfloor \pi)^2} \text{ otherwise.} \quad (8.4.17)$$

Then by (8.4.15), (8.4.16), we have

$$\begin{aligned} &\mathbb{E}[(\widetilde{I}_{q,l,t_n,t_{n+1},0}^F - I_{q,l,t_n,t_{n+1},0})^2] \\ &= \sum_{i=N_h+1}^{\infty} \int_{t_n}^{t_{n+1}} M_i^2(s) ds + \sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \left(\int_{t_n}^{t_{n+1}} M_i(s) m_j(s) ds\right)^2 \\ &= \varsigma(N_h) \frac{\Delta^2}{(N_h \pi)^2} + \sum_{i=N_h+1}^{\infty} \frac{3\varsigma(i) \Delta^2}{(2\lfloor i/2 \rfloor \pi)^2} = \varsigma(N_h) \frac{2\Delta^2}{(N_h \pi)^2} + \sum_{i=M}^{\infty} \frac{\Delta^2}{(i\pi)^2}. \end{aligned}$$

Hence, we arrive at (8.2.22) by the fact  $\sum_{i=M}^{\infty} \frac{1}{i^2} \leq \frac{1}{M}$ . Similarly, we can obtain (8.2.23).  $\square$

**Proof of Theorem 8.2.6.** Subtracting (8.2.17) from (8.2.20) and taking expectation after squaring over both sides, we have

$$\mathbb{E}[(X_{n+1}^M - X_{n+1})^2] = \mathbb{E}[(X_n^M - X_n)^2] + 2 \sum_{i=0}^4 \mathbb{E}[(X_n^M - X_n)R_i] + \sum_{i,j=0}^4 \mathbb{E}[R_i R_j],$$

where we denote  $f_n^M = f(X_n^M, X_{n-m}^M)$  and  $g_{l,n}^M = g_l(X_n^M, X_{n-m}^M)$  and

$$\begin{aligned} R_0 &= h(f_n^M - f_n) + \sum_{l=1}^r (g_{l,n}^M - g_{l,n}) \Delta W_{l,n}, \\ R_1 &= \sum_{l=1}^r \sum_{q=1}^r [\partial_x g_{l,n}^M g_{q,n}^M - \partial_x g_{l,n} g_{q,n}] \tilde{I}_{q,l,t_n,t_{n+1},0}^F, \\ R_2 &= \sum_{l=1}^r \sum_{q=1}^r [\partial_{x_\tau} g_{l,n}^M g_{q,n-m}^M - \partial_{x_\tau} g_{l,n} g_{q,n-m}] \tilde{I}_{q,l,t_n,t_{n+1},\tau}^F, \\ R_3 &= \sum_{l=1}^r \sum_{q=1}^r \partial_x g_{l,n}^M g_{q,n}^M (I_{q,l,t_n,t_{n+1},0} - \tilde{I}_{q,l,t_n,t_{n+1},0}^F), \\ R_4 &= \sum_{l=1}^r \sum_{q=1}^r \partial_{x_\tau} g_{l,n}^M g_{q,n-m}^M (I_{q,l,t_n,t_{n+1},\tau} - \tilde{I}_{q,l,t_n,t_{n+1},\tau}^F). \end{aligned}$$

Similar to the proof of Theorem 8.2.3, we have

$$\begin{aligned} \mathbb{E}[R_0^2] &\leq C(h^2 + h)(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]), \\ \mathbb{E}[R_1^2] &\leq C \max_{1 \leq l, q \leq r} \mathbb{E}[(|X_n^M - X_n|^2 + |X_{n-m}^M - X_{n-m}|^2)] \mathbb{E}[(\tilde{I}_{q,l,t_n,t_{n+1},0}^F)^2], \\ \mathbb{E}[R_2^2] &\leq C \max_{1 \leq l, q \leq r} \mathbb{E}[(|X_n^M - X_n|^2 + |X_{n-m}^M - X_{n-m}|^2)] \mathbb{E}[(\tilde{I}_{q,l,t_n,t_{n+1},\tau}^F)^2], \\ \mathbb{E}[R_3^2] &\leq C \max_{1 \leq l < q \leq r} \mathbb{E}[(1 + |X_n^M|^2 + |X_{n-m}^M|^2)] \mathbb{E}[(I_{q,l,t_n,t_{n+1},0} - \tilde{I}_{q,l,t_n,t_{n+1},0}^F)^2], \\ \mathbb{E}[R_4^2] &\leq C \max_{1 \leq l < q \leq r} \mathbb{E}[(1 + |X_n^M|^2 + |X_{n-m}^M|^2)] \mathbb{E}[(I_{q,l,t_n,t_{n+1},\tau} - \tilde{I}_{q,l,t_n,t_{n+1},\tau}^F)^2]. \end{aligned} \tag{8.4.18}$$

First, we establish the following estimations:

$$\mathbb{E}[R_i^2] \leq Ch^3, \quad i = 3, 4. \quad (8.4.19)$$

The case for  $i = 3$  follows directly from Lemma 8.2.5 and boundedness of moments of  $X_n$  and  $X_n^M$ .

By Lemma 8.2.5 and (8.2.18), we have

$$\begin{aligned} & \mathbb{E}[(I_{q,l,t_n,t_{n+1},\tau} - \tilde{I}_{q,l,t_n,t_{n+1},\tau}^F)^4] \\ = & \mathbb{E}\left[\left(-\frac{\sqrt{2}h}{2\pi}\xi_{q,1}^{(n-m)} \sum_{p=s+1}^{\infty} \frac{1}{p}\xi_{l,2p}^{(n)} + \frac{h}{2\pi} \sum_{p=s_1+1}^{\infty} \frac{1}{p}[\xi_{q,2p+1}^{(n-m)}\xi_{l,2p}^{(n)} - \xi_{q,2p}^{(n-m)}\xi_{l,2p+1}^{(n)}]\right)^4\right] \\ \leq & Ch^4\left[\left(\sum_{p=s+1}^{\infty} \frac{1}{p^2}\right)^2 + \left(\sum_{p=s_1+1}^{\infty} \frac{1}{p^2}\right)^2\right] \leq C\frac{h^4}{N_h^2}, \end{aligned}$$

where  $s = \lfloor \frac{N_h}{2} \rfloor$  and  $s_1 = \lfloor \frac{N_h-1}{2} \rfloor$ . As  $N_h$  is at the order of  $1/h$ , we have

$$\mathbb{E}[(I_{q,l,t_n,t_{n+1},\tau} - \tilde{I}_{q,l,t_n,t_{n+1},\tau}^F)^4] \leq Ch^6. \quad (8.4.20)$$

Then by the fact that  $X_n$  and  $X_n^M$  have bounded fourth-order moment, Cauchy inequality, and (8.4.20), we reach (8.4.19) when  $i = 4$ .

Second, we estimate  $\mathbb{E}[R_i^2]$ ,  $i = 1, 2$ . By (8.2.18), the Lipschitz condition (8.2.2) and  $N_h$  is at the order of  $1/h$ , we have

$$\mathbb{E}[R_1^2] \leq Ch(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]). \quad (8.4.21)$$

Now we require to estimate  $\mathbb{E}[R_2^2]$ . By Lipschitz condition (8.2.2), the adaptedness of  $X_{n-m}^M$  and  $X_{n-m}$  and Cauchy inequality (twice), we have

$$\mathbb{E}[R_2^2] \leq C \max_{1 \leq l, q \leq r} \left\{ \mathbb{E}[|X_n^M - X_n|^2 (\tilde{I}_{q,l,t_n,t_{n+1},\tau}^F)^2] + \mathbb{E}[|X_{n-m}^M - X_{n-m}|^2 (\tilde{I}_{q,l,t_n,t_{n+1},\tau}^F)^2] \right\},$$

$$\begin{aligned} &\leq C \max_{1 \leq l, q \leq r} (\mathbb{E}[|X_n^M - X_n|^4])^{1/4} (\mathbb{E}[(\tilde{I}_{q,l,t_n,t_{n+1},\tau}^F)^8])^{1/4} (\mathbb{E}[|X_n^M - X_n|^2])^{1/2} \\ &\quad + Ch^2 \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]. \end{aligned}$$

It can be readily checked from (8.2.18) that  $\mathbb{E}[(\tilde{I}_{q,l,t_n,t_{n+1},\tau}^F)^8] \leq Ch^8$ . Hence, from boundedness of moments, we have

$$\mathbb{E}[R_2^2] \leq Ch^2 (\mathbb{E}[(X_n^M - X_n)^2])^{1/2} + Ch^2 \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]. \quad (8.4.22)$$

Now estimate  $\mathbb{E}[(X_n^M - X_n)R_i]$ ,  $i = 0, 1, 2, 3, 4$ . By the adaptedness of  $X_n$  and Lipschitz condition of  $f$ , we have

$$\mathbb{E}[(X_n^M - X_n)R_0] \leq Ch \mathbb{E}[ (|X_n^M - X_n|^2 + |X_{n-m}^M - X_{n-m}|^2) ]. \quad (8.4.23)$$

By the adaptedness of  $X_n$  and  $\mathbb{E}[\tilde{I}_{q,l,t_n,t_{n+1},0}] = \delta_{q,l}h/2$  ( $\delta_{q,l}$  is the Kronecker delta) and Lipschitz condition of  $\partial_x g_l g_q$ , we have

$$\mathbb{E}[(X_n^M - X_n)R_1] \leq Ch \mathbb{E}[ (|X_n^M - X_n|^2 + |X_{n-m}^M - X_{n-m}|^2) ]. \quad (8.4.24)$$

By the adaptedness of  $X_n$  and  $\mathbb{E}[\tilde{I}_{q,l,t_n,t_{n+1},0} - I_{q,l,t_n,t_{n+1},0}] = 0$ , we have

$$\mathbb{E}[(X_n^M - X_n)R_3] = 0. \quad (8.4.25)$$

Similar to the proof of (8.4.12), we have

$$\mathbb{E}[(X_n^M - X_n)R_4] = 0. \quad (8.4.26)$$

Then by (8.4.18), (8.4.19)-(8.4.22), (8.4.23)-(8.4.26) and Cauchy inequality, we have

$$\begin{aligned} \mathbb{E}[(X_{n+1}^M - X_{n+1})^2] &\leq (1 + Ch)\mathbb{E}[(X_n^M - X_n)^2] + Ch\mathbb{E}[(X_{n-m}^M - X_{n-m})^2] \\ &\quad + Ch^2(\mathbb{E}[(X_n^M - X_n)^2])^{1/2} + Ch^3, \end{aligned} \tag{8.4.27}$$

where  $n \geq m$ . Similarly, we have that (8.4.27) holds also for  $1 \leq n \leq m - 1$ . From here and by nonlinear Gronwall inequality, we reach the conclusion (8.2.24).  $\square$

# Chapter 9

## Conclusion

### 9.1 Summary

In this work, we developed a recursive multistage Wiener chaos expansion method (WCE) and a recursive multistage stochastic collocation method (SCM) for longer time integration of linear stochastic advection-reaction-diffusion equations with finite dimensional noises. To compute first two moments of the solution with such a recursive multistage procedure, we first compute the covariance matrix of the solution at different physical points at a time step and then recursively compute the covariance matrix of the solution at the next time step using the covariance matrix at the previous time step. We continue this process before we reach the final integration time.

We compared the recursive multistage WCE with methods of characteristics plus a standard Monte Carlo sampling strategy. We showed that WCE is more efficient than standard Monte Carlo methods if high accuracy in first two moments is required.

We also compared WCE and SCM in conjunction with the recursive multistage procedure. Though WCE exhibits higher order convergence than SCM, we show that both methods are comparable in performance, depending on underlying problems. Although the computational cost is proportional to the fourth power of the number of nodes or modes employed in physical space, the

cost can be reduced to the second power if we make full use of the sparsity of the solution.

For SCM, we also investigated a benchmark problem for stochastic solvers—a stochastic piston problem in one-dimensional physical space. We modeled the problem of moving piston with velocity being Brownian motion into a tube with stochastic Euler equations driven by white noise. After splitting the stochastic Euler equations into two parts (by Lie-Trotter splitting), we truncated the Brownian motion with its spectral expansion and applied SCM to obtain variances of the shock locations at different time instants. We showed that SCM is efficient for a short time simulation and quasi-Monte Carlo methods are efficient for a relatively longer time simulation.

We also illustrate the efficiency of SCM with Euler scheme in time through a linear stochastic ordinary differential equations: error estimates show that SCM using sparse grid of Smolyak type is efficient for short time integration and for small magnitudes of noises.

Our conclusion on integration methods in random space is as follows. WCE and SCM are efficient for longer time integration of linear problems using our recursive approach and for a small number of noises within short time simulation. However, if time increases, we have already employed many random variables, either from increments of Brownian motion or from the modes of spectral truncation of Brownian motion. Hence, deterministic integration methods are not efficient any more since their computational cost grows exponentially with the number of random variables. We then have to use randomized sampling strategies, such as Monte Carlo methods or randomized quasi-Monte Carlo methods, possibly together with variance reduction methods to reduce the statistical errors.

For both WCE and SCM, we applied the Wong-Zakai approximation using a spectral approximation of Brownian motion. However, we use different stochastic products for WCE and SCM because of computational efficiency. In practice, WCE is associated with the Ito-Wick product, which yields a weakly coupled system of PDEs for linear equations. SCM is associated with the Stratonovich product, which yields a decoupled system of PDEs. These different formulations lead to different numerical performance but both methods are comparable in performance for linear

problems.

We also used Wong-Zakai approximation to stochastic ordinary differential equations with and without delay. For stochastic differential equations with constant time delay, we derived three schemes from Wong-Zakai approximation using spectral approximation, predictor-corrector scheme, mid-point scheme and Milstein scheme. We found that even when there is one single delay term in the diffusion coefficient, the predictor-corrector and mid-point schemes cannot converge with order one in the mean-square sense. This is quite different from the cases for these two schemes for equations without delay where these two schemes are of strong order one under commutative conditions on the diffusion coefficients. If there is any delay in the diffusion coefficients, we found that *Milstein scheme is substantially slower than the other two schemes unless there is only one single noise. In other words, the most efficient schemes for stochastic delay differential equations are half-order schemes in general.*

For stochastic ordinary differential equations with or without delay, we observed that the convergence order of numerical schemes via the Wong-Zakai approximation is not determined by the Wong-Zakai approximation but relies on further time discretization in time. For example, the Wong-Zakai approximation itself is of order half; however, Milstein scheme based on the Wong-Zakai approximation (called Milstein-like scheme in Chapter 8) is of order one.

Using spectral approximation of Brownian motion, we investigated a semilinear equation with additive spatial white noise. We found that for problems in two or three dimensions in physical space, we cannot expect better convergence from the spectral approximation of Brownian motion than that from piecewise linear approximation. However, we may expect high-order convergence when we have solutions of high regularity. For example, for elliptic equations with additive noise in one-dimensional physical space or even higher-order equations in two- or three-dimensional physical space, we can expect high regularity and benefit from the spectral truncation of Brownian motion.

We also considered stochastic differential equation with non-Lipschitz continuous coefficients both in drift and diffusion. Under a one-sided Lipschitz condition on coefficients, we established a

fundamental limit theorem, i.e., a relationship between the local truncation error and global error in the mean-square sense for numerical schemes for nonlinear stochastic differential equations. We also proposed an explicit balanced scheme so that we can efficiently integrate stochastic differential equation with superlinearly growing coefficients over a finite time interval.

## 9.2 Future work

For long-time integration of nonlinear stochastic differential equations using deterministic integration methods in random space, dimensionality in random space is still the essential difficulty: the number of random variables growth linearly and the number of Wiener chaos modes or stochastic collocation points grows exponentially. For linear equations solved with the recursive multistage WCE or SCM, we will have fast increasing computational cost for some statistics of the solutions other than the first two moments, e.g., the computational cost for third-order moments will be proportional to the sixth power of nodes or modes employed in physical space. For nonlinear equations, the recursive multistage approach fails as nonlinear equations usually have strong dependence on the initial condition and the superposition principle does not work any more.

To lift the curse of dimensionality, we have to suppress the history data and restart from time to time to keep low dimensionality in random space (and thus low computational cost). To suppress the history data, we should employ some reduction methods, such as functional analysis of variance, see e.g. [115, 144, 419], to reduce the number of used random variables before integrating over the next time interval.

For stochastic partial differential equations with space-time noise, deterministic integration methods in random space are too expensive as many random variables will be employed to truncate space-time noise. Monte Carlo methods and associated variance reduction methods including multilevel Monte Carlo method could potentially be applied to resolve this issue.

# Bibliography

- [1] A. ABDULLE, A. BARTH, AND C. SCHWAB, *Multilevel Monte Carlo methods for stochastic elliptic multiscale PDEs*, tech. rep., SAM, ETHZ, 2012.
- [2] P. ACQUISTAPACE AND B. TERRENI, *An approach to Ito linear equations in Hilbert spaces by approximation of white noise with coloured noise*, *Stochastic Anal. Appl.*, 2 (1984), pp. 131–186.
- [3] I. A. ADAMU AND G. J. LORD, *Numerical approximation of multiplicative SPDEs*, *Int. J. Comput. Math.*, 89 (2012), pp. 2603–2621.
- [4] A. ALABERT AND I. GYÖNGY, *On numerical approximation of stochastic Burgers' equation*, in *From stochastic calculus to mathematical finance*, Springer, Berlin, 2006, pp. 1–15.
- [5] E. J. ALLEN, S. J. NOVOSEL, AND Z. ZHANG, *Finite element and difference approximation of some linear stochastic partial differential equations*, *Stochastics Stochastics Rep.*, 64 (1998), pp. 117–142.
- [6] V. V. ANH, W. GRECKSCH, AND A. WADEWITZ, *A splitting method for a stochastic Goursat problem*, *Stochastic Anal. Appl.*, 17 (1999), pp. 315–326.
- [7] L. ARNOLD, *Stochastic differential equations: theory and applications*, Wiley-Interscience, New York, 1974.

- [8] A. ASHYRALYEV AND M. AKAT, *An approximation of stochastic hyperbolic equations: case with Wiener process*, Math. Methods Appl. Sci., 36 (2013), pp. 1095–1106.
- [9] I. BABUSKA, F. NOBILE, AND R. TEMPONE, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034 (electronic).
- [10] I. BABUSKA, R. TEMPONE, AND G. E. ZOURARIS, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825.
- [11] J. BÄCK, F. NOBILE, L. TAMELLINI, AND R. TEMPONE, *Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: A numerical comparison*, in Spectral and High Order Methods for Partial Differential Equations, Springer Berlin Heidelberg, 2011, pp. 43–62.
- [12] C. T. H. BAKER AND E. BUCKWAR, *Numerical analysis of explicit one-step methods for stochastic delay differential equations*, LMS J. Comput. Math., 3 (2000), pp. 315–335.
- [13] V. BALLY, *Approximation for the solutions of stochastic differential equations. I.  $L^p$ -convergence*, Stochastics Stochastics Rep., 28 (1989), pp. 209–246.
- [14] ———, *Approximation for the solutions of stochastic differential equations. II. Strong convergence*, Stochastics Stochastics Rep., 28 (1989), pp. 357–385.
- [15] ———, *Approximation for the solutions of stochastic differential equations. III. Jointly weak convergence*, Stochastics Stochastics Rep., 30 (1990), pp. 171–191.
- [16] V. BALLY, A. MILLET, AND M. SANZ-SOLÉ, *Approximation and support theorem in Hölder norm for parabolic stochastic partial differential equations*, Ann. Probab., 23 (1995), pp. 178–222.

- [17] X. BARDINA, M. JOLIS, AND L. QUER-SARDANYONS, *Weak convergence for the stochastic heat equation driven by Gaussian white noise*, Electron. J. Probab., 15 (2010), pp. no. 39, 1267–1295.
- [18] X. BARDINA, I. NOURDIN, C. ROVIRA, AND S. TINDEL, *Weak approximation of a fractional SDE*, Stochastic Process. Appl., 120 (2010), pp. 39–65.
- [19] A. BARTH AND A. LANG, *Milstein approximation for advection-diffusion equations driven by multiplicative noncontinuous martingale noises*, Appl. Math. Optim., 66 (2012), pp. 387–413.
- [20] ———, *Simulation of stochastic partial differential equations using finite element methods*, Stochastics, 84 (2012), pp. 217–231.
- [21] ———,  *$L^p$  and almost sure convergence of a Milstein scheme for stochastic partial differential equations*, Stochastic Process. Appl., 123 (2013), pp. 1563 – 1587.
- [22] A. BARTH, A. LANG, AND C. SCHWAB, *Multilevel Monte Carlo method for parabolic stochastic partial differential equations*, BIT, 53 (2013), pp. 3–27.
- [23] A. BARTH, C. SCHWAB, AND N. ZOLLINGER, *Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients*, Numer. Math., 119 (2011), pp. 123–161.
- [24] M. BARTON-SMITH, A. DEBUSSCHE, AND L. DI MENZA, *Numerical study of two-dimensional stochastic NLS equations*, Numer. Methods Partial Differential Equations, 21 (2005), pp. 810–842.
- [25] C. BAUZET, *On a time-splitting method for a scalar conservation law with a multiplicative stochastic perturbation and numerical experiments*, J. Evol. Equ., (To appear).
- [26] C. BAYER AND P. K. FRIZ, *Cubature on Wiener space: pathwise convergence*, Appl. Math. Optim., 67 (2013), pp. 261–278.

- [27] S. BECKER, A. JENTZEN, AND P. E. KLOEDEN, *An exponential Wagner-Platen type scheme for SPDEs*, ArXiv e-prints, (2013).
- [28] G. BEN AROUS, M. GRĂDINARU, AND M. LEDOUX, *Hölder norms and the support theorem for diffusions*, Ann. Inst. H. Poincaré Probab. Statist., 30 (1994), pp. 415–436.
- [29] A. BENSOUSSAN, R. GLOWINSKI, AND A. RĂȘCANU, *Approximation of the Zakai equation by the splitting up method*, SIAM J. Control Optim., 28 (1990), pp. 1420–1431.
- [30] ———, *Approximation of the Zakai equation by the splitting up method*, SIAM J. Control Optim., 28 (1990), pp. 1420–1431.
- [31] ———, *Approximation of some stochastic differential equations by the splitting up method*, Appl. Math. Optim., 25 (1992), pp. 81–106.
- [32] F. E. BENTH AND J. GJERDE, *Convergence rates for finite element approximations of stochastic partial differential equations*, Stochastics Stochastics Rep., 63 (1998), pp. 313–326.
- [33] C. BERNARDI AND Y. MADAY, *Spectral methods*, vol. 5 of Handbook of Numerical Analysis, Elsevier, Amsterdam, 1997.
- [34] M. BIERI AND C. SCHWAB, *Sparse high order FEM for elliptic sPDEs*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 1149–1170.
- [35] D. BLÖMKER AND A. JENTZEN, *Galerkin approximations for the stochastic Burgers equation*, SIAM J. Numer. Anal., 51 (2013), pp. 694–715.
- [36] D. BLÖMKER, M. KAMRANI, AND S. M. HOSSEINI, *Full discretization of the stochastic Burgers equation with correlated noise*, IMA J. Numer. Anal., (2013).
- [37] N. BOU-RABEE AND E. VANDEN-EIJNDEN, *A patch that imparts unconditional stability to explicit integrators for Langevin-like equations*, J. Comput. Phys., 231 (2012), pp. 2565–2580.

- [38] H. BRECKNER, *Approximation of the solution of the stochastic Navier-Stokes equation*, Optimization, 49 (2001), pp. 15–38.
- [39] N. BRUTI-LIBERATI AND E. PLATEN, *Strong predictor-corrector Euler methods for stochastic differential equations*, Stoch. Dyn., 8 (2008), pp. 561–581.
- [40] Z. BRZEŹNIAK, M. CAPIŃSKI, AND F. FLANDOLI, *Stochastic partial differential equations and turbulence*, Math. Models Methods Appl. Sci., 1 (1991), pp. 41–59.
- [41] Z. BRZEŹNIAK AND A. CARROLL, *Approximations of the Wong-Zakai type for stochastic differential equations in  $M$ -type 2 Banach spaces with applications to loop spaces*, in Séminaire de Probabilités XXXVII, Springer, Berlin, 2003, pp. 251–289.
- [42] Z. BRZEŹNIAK AND F. FLANDOLI, *Almost sure approximation of Wong-Zakai type for stochastic partial differential equations*, Stochastic Process. Appl., 55 (1995), pp. 329–358.
- [43] Z. BRZEŹNIAK AND A. MILLET, *On the splitting method for some complex-valued quasilinear evolution equations*, in Stochastic Analysis and Related Topics, L. Decreasefond and J. Najim, eds., Springer Berlin Heidelberg, 2012, pp. 57–90.
- [44] Z. BRZENIAK, E. CARELLI, AND A. PROHL, *Finite-element-based discretizations of the incompressible Navier-Stokes equations with multiplicative random forcing*, IMA J. Numer. Anal., (2013).
- [45] R. BUCKDAHN AND É. PARDOUX, *Monotonicity methods for white noise driven quasi-linear SPDEs*, in Diffusion processes and related problems in analysis, Vol. I (Evanston, IL, 1989), Birkhäuser, Boston, 1990, pp. 219–233.
- [46] E. BUCKWAR AND T. SICKENBERGER, *A comparative linear mean-square stability analysis of Maruyama- and Milstein-type methods*, Math. Comput. Simulation, 81 (2011), pp. 1110–1127.

- [47] E. BUCKWAR AND R. WINKLER, *Multistep methods for SDEs and their application to problems with small noise*, SIAM J. Numer. Anal., 44 (2006), pp. 779–803.
- [48] ———, *Multi-step Maruyama methods for stochastic delay differential equations*, Stoch. Anal. Appl., 25 (2007), pp. 933–959.
- [49] A. BUDHIRAJA AND G. KALLIANPUR, *Approximations to the solution of the Zakai equation using multiple Wiener and Stratonovich integral expansions*, Stochastics Stochastics Rep., 56 (1996), pp. 271–315.
- [50] ———, *The Feynman-Stratonovich semigroup and Stratonovich integral expansions in nonlinear filtering*, Appl. Math. Optim., 35 (1997), pp. 91–116.
- [51] ———, *Two results on multiple Stratonovich integrals*, Statist. Sinica, 7 (1997), pp. 907–922.
- [52] R. H. CAMERON AND W. T. MARTIN, *The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals*, Ann. of Math. (2), 48 (1947), pp. 385–392.
- [53] W. CAO AND Z. ZHANG, *Simulations of two-step Maruyama methods for nonlinear stochastic delay differential equations*, Adv. Appl. Math. Mech., 4 (2012), pp. 821–832.
- [54] ———, *On exponential mean-square stability of two-step Maruyama methods for stochastic delay differential equations*, J. Comput. Appl. Math., 245 (2013), pp. 182–193.
- [55] Y. CAO, H. YANG, AND L. YIN, *Finite element methods for semilinear elliptic stochastic partial differential equations*, Numer. Math., 106 (2007), pp. 181–198.
- [56] Y. CAO AND L. YIN, *Spectral Galerkin method for stochastic wave equations driven by space-time white noise*, Commun. Pure Appl. Anal., 6 (2007), pp. 607–617.
- [57] ———, *Spectral method for nonlinear stochastic partial differential equations of elliptic type*, Numer. Math. Theory Methods Appl., 4 (2011), pp. 38–52.

- [58] Y. CAO, R. ZHANG, AND K. ZHANG, *Finite element and discontinuous Galerkin method for stochastic Helmholtz equation in two- and three-dimensions*, J. Comput. Math., 26 (2008), pp. 702–715.
- [59] ———, *Finite element method and discontinuous Galerkin method for stochastic scattering problem of Helmholtz type in  $\mathbb{R}^d$  ( $d = 2, 3$ )*, Potential Anal., 28 (2008), pp. 301–319.
- [60] E. CARELLI AND A. PROHL, *Rates of convergence for discretizations of the stochastic incompressible Navier-Stokes equations*, SIAM J. Numer. Anal., 50 (2012), pp. 2467–2496.
- [61] T. CASS AND C. LITTERER, *On the error estimate for cubature on Wiener space*, ArXiv e-prints, (2011).
- [62] H. D. CENICEROS AND G. O. MOHLER, *A practical splitting method for stiff SDEs with applications to problems with small noise*, Multiscale Model. Simul., 6 (2007), pp. 212–227 (electronic).
- [63] M. CHALEYAT-MAUREL AND D. MICHEL, *A Stroock Varadhan support theorem in nonlinear filtering theory*, Probab. Theory Related Fields, 84 (1990), pp. 119–139.
- [64] J. CHARRIER, *Strong and weak error estimates for elliptic partial differential equations with random coefficients*, SIAM J. Numer. Anal., 50 (2012), pp. 216–246.
- [65] J. CHARRIER AND A. DEBUSSCHE, *Weak truncation error estimates for elliptic PDEs with lognormal coefficients*, Stoch PDE: Anal Comp, 1 (2013), pp. 63–93.
- [66] J. CHARRIER, R. SCHEICHL, AND A. L. TECKENTRUP, *Finite Element Error Analysis of Elliptic PDEs with Random Coefficients and Its Application to Multilevel Monte Carlo Methods*, SIAM J. Numer. Anal., 51 (2013), pp. 322–352.
- [67] G.-Q. CHEN, Q. DING, AND K. H. KARLSEN, *On nonlinear stochastic balance laws*, Arch. Rational Mech. Anal., 204 (2012), pp. 707–743.

- [68] P. L. CHOW, J.-L. JIANG, AND J.-L. MENALDI, *Pathwise convergence of approximate solutions to Zakai's equation in a bounded domain*, in Stochastic partial differential equations and applications (Trento, 1990), Longman Sci. Tech., Harlow, 1992, pp. 111–123.
- [69] I. CHUESHOV AND A. MILLET, *Stochastic two-dimensional hydrodynamical systems: Wong-Zakai approximation and support theorem*, Stoch. Anal. Appl., 29 (2011), pp. 570–611.
- [70] P. G. CIARLET, *The finite element method for elliptic problems*, SIAM, Philadelphia, PA, 2002.
- [71] Z. CIESIELSKI, *Hölder conditions for realizations of Gaussian processes*, Trans. Amer. Math. Soc., 99 (1961), pp. 403–413.
- [72] K. A. CLIFFE, M. B. GILES, R. SCHEICHL, AND A. L. TECKENTRUP, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15.
- [73] F. COMTE, V. GENON-CATALOT, AND Y. ROZENHOLC, *Penalized nonparametric mean square estimation of the coefficients of diffusion processes*, Bernoulli, 13 (2007), pp. 514–543.
- [74] S. COX AND J. VAN NEERVEN, *Convergence rates of the splitting scheme for parabolic linear stochastic Cauchy problems*, SIAM J. Numer. Anal., 48 (2010), pp. 428–451.
- [75] ———, *Pathwise Hölder convergence of the implicit-linear Euler scheme for semi-linear SPDEs with multiplicative noise*, Numer. Math., 125 (2013), pp. 259–345.
- [76] D. CRISAN, *Exact rates of convergence for a branching particle approximation to the solution of the Zakai equation*, Ann. Probab., 31 (2003), pp. 693–718.
- [77] D. CRISAN, *Particle approximations for a class of stochastic partial differential equations*, Appl. Math. Optim., 54 (2006), pp. 293–314.

- [78] D. CRISAN, J. GAINES, AND T. LYONS, *Convergence of a branching particle method to the solution of the Zakai equation*, SIAM J. Appl. Math., 58 (1998), pp. 1568–1590 (electronic).
- [79] D. CRISAN AND T. LYONS, *A particle approximation of the solution of the Kushner-Stratonovitch equation*, Probab. Theory Related Fields, 115 (1999), pp. 549–578.
- [80] D. CRISAN AND J. XIONG, *Numerical solutions for a class of SPDEs over bounded domains*, in Conference Oxford sur les méthodes de Monte Carlo séquentielles, EDP Sci., Les Ulis, 2007, pp. 121–125.
- [81] G. DA PRATO, *Kolmogorov equations for stochastic PDEs*, Birkhäuser Verlag, Basel, 2004.
- [82] G. DA PRATO, A. DEBUSSCHE, AND R. TEMAM, *Stochastic Burgers' equation*, NoDEA Nonlinear Differential Equations Appl., 1 (1994), pp. 389–402.
- [83] G. DA PRATO AND J. ZABCZYK, *Stochastic equations in infinite dimensions*, Cambridge University Press, Cambridge, 1992.
- [84] A. M. DAVIE AND J. G. GAINES, *Convergence of numerical schemes for the solution of parabolic stochastic partial differential equations*, Math. Comp., 70 (2001), pp. 121–134 (electronic).
- [85] P. J. DAVIS AND P. RABINOWITZ, *Methods of numerical integration*, Computer Science and Applied Mathematics, Academic Press Inc., Orlando, FL, second ed., 1984.
- [86] D. A. DAWSON AND E. A. PERKINS, *Measure-valued processes and renormalization of branching particle systems*, in Stochastic partial differential equations: six perspectives, R. A. Carmona and B. Rozovskii, eds., AMS, Providence, RI, 1999, pp. 45–106.
- [87] A. DE BOUARD AND A. DEBUSSCHE, *A stochastic nonlinear Schrödinger equation with multiplicative noise*, Comm. Math. Phys., 205 (1999), pp. 161–181.

- [88] A. DE BOUARD AND A. DEBUSSCHE, *A semi-discrete scheme for the stochastic nonlinear Schrödinger equation*, Numer. Math., 96 (2004), pp. 733–770.
- [89] A. DE BOUARD AND A. DEBUSSCHE, *Weak and strong order of convergence of a semidiscrete scheme for the stochastic nonlinear Schrödinger equation*, Appl. Math. Optim., 54 (2006), pp. 369–399.
- [90] A. DE BOUARD AND A. DEBUSSCHE, *Random modulation of solitons for the stochastic Korteweg-de Vries equation*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 24 (2007), pp. 251–278.
- [91] A. DE BOUARD AND A. DEBUSSCHE, *The nonlinear Schrödinger equation with white noise dispersion*, J. Funct. Anal., 259 (2010), pp. 1300–1321.
- [92] A. DE BOUARD, A. DEBUSSCHE, AND L. DI MENZA, *Theoretical and numerical aspects of stochastic nonlinear Schrödinger equations*, Monte Carlo Methods Appl., 7 (2001), pp. 55–63.
- [93] A. DE BOUARD, A. DEBUSSCHE, AND Y. TSUTSUMI, *White noise driven Korteweg-de Vries equation*, J. Funct. Anal., 169 (1999), pp. 532–558.
- [94] A. DE BOUARD AND A. DEBUSSCHE, *On the stochastic Korteweg-de Vries equation*, J. Funct. Anal., 154 (1998), pp. 215–251.
- [95] A. DEBUSSCHE, *The 2D-Navier-Stokes equations perturbed by a delta correlated noise*, in Probabilistic methods in fluids, World Sci. Publ., River Edge, NJ, 2003, pp. 115–129.
- [96] A. DEBUSSCHE, *Weak approximation of stochastic partial differential equations: the nonlinear case*, Math. Comp., 80 (2011), pp. 89–117.
- [97] A. DEBUSSCHE AND J. PRINTEMPS, *Numerical simulation of the stochastic Korteweg-de Vries equation*, Phys. D, 134 (1999), pp. 200–226.

- [98] ———, *Convergence of a semi-discrete scheme for the stochastic Korteweg-de Vries equation*, Discrete Contin. Dyn. Syst. Ser. B, 6 (2006), pp. 761–781 (electronic).
- [99] ———, *Weak order for the discretization of the stochastic heat equation*, Math. Comp., 78 (2009), pp. 845–863.
- [100] A. DEBUSSCHE AND J. VOVELLE, *Scalar conservation laws with stochastic forcing*, J. Funct. Anal., 259 (2010), pp. 1014–1042.
- [101] A. DEYA, M. JOLIS, AND L. QUER-SARDANYONS, *The Stratonovich heat equation: a continuity result and weak approximations*, Electron. J. Probab., 18 (2013), pp. no. 3, 34.
- [102] J. DICK, F. Y. KUO, AND I. H. SLOAN, *High-dimensional integration: the quasi-Monte Carlo way*, Acta Numer., 22 (2013), pp. 133–288.
- [103] P. DÖRSEK, *Semigroup splitting and cubature approximations for the stochastic Navier-Stokes equations*, SIAM J. Numer. Anal., 50 (2012), pp. 729–746.
- [104] H. DOSS, *Liens entre équations différentielles stochastiques et ordinaires*, C. R. Acad. Sci. Paris Sér. A-B, 283 (1976), pp. Ai, A939–A942.
- [105] Q. DU AND T. ZHANG, *Numerical approximation of some linear stochastic partial differential equations driven by special additive noises*, SIAM J. Numer. Anal., 40 (2002), pp. 1421–1445 (electronic).
- [106] Y. DUAN AND X. YANG, *On the convergence of a full discretization scheme for the stochastic Navier-Stokes equations*, J. Comput. Anal. Appl., 13 (2011), pp. 485–498.
- [107] ———, *The finite element method of a Euler scheme for stochastic Navier-Stokes equations involving the turbulent component*, Int. J. Numer. Anal. Model., 10 (2013), pp. 727–744.

- [108] M. A. EL-TAWIL AND A.-H. A. EL-SHIKHIPY, *Approximations for some statistical moments of the solution process of stochastic Navier-Stokes equation using WHEP technique*, Appl. Math. Inf. Sci., 6 (2012), pp. 1095–1100.
- [109] J. FENG AND D. NUALART, *Stochastic scalar conservation laws*, J. Funct. Anal., 255 (2008), pp. 313 – 373.
- [110] G. FERREYRA, *A Wong-Zakai-type theorem for certain discontinuous semimartingales*, J. Theoret. Probab., 2 (1989), pp. 313–323.
- [111] F. FLANDOLI AND V. M. TORTORELLI, *Time discretization of Ornstein-Uhlenbeck equations and stochastic Navier-Stokes equations with a generalized noise*, Stochastics Stochastics Rep., 55 (1995), pp. 141–165.
- [112] P. FLORCHINGER AND F. LE GLAND, *Time-discretization of the Zakai equation for diffusion processes observed in correlated noise*, in Analysis and optimization of systems (Antibes, 1990), Springer, Berlin, 1990, pp. 228–237.
- [113] ———, *Time-discretization of the Zakai equation for diffusion processes observed in correlated noise*, Stochastics Stochastics Rep., 35 (1991), pp. 233–256.
- [114] ———, *Particle approximation for first order stochastic partial differential equations*, in Applied stochastic analysis (New Brunswick, NJ, 1991), Springer, Berlin, 1992, pp. 121–133.
- [115] J. FOO AND G. E. KARNIADAKIS, *Multi-element probabilistic collocation method in high dimensions*, J. Comput. Phys., 229 (2010), pp. 1536–1557.
- [116] P. FRIZ AND H. OBERHAUSER, *Rough path limits of the Wong-Zakai type with a modified drift term*, J. Funct. Anal., 256 (2009), pp. 3236–3256.
- [117] ———, *On the splitting-up method for rough (partial) differential equations*, J. Differential Equations, 251 (2011), pp. 316–338.

- [118] J. G. GAINES, *Numerical experiments with  $S(P)DE$ 's*, in Stochastic partial differential equations (Edinburgh, 1994), Cambridge Univ. Press, Cambridge, 1995, pp. 55–71.
- [119] J. GALVIS AND M. SARKIS, *Approximating infinity-dimensional stochastic Darcy's equations without uniform ellipticity*, SIAM J. Numer. Anal., 47 (2009), pp. 3624–3651.
- [120] J. GALVIS AND M. SARKIS, *Regularity results for the ordinary product stochastic pressure equation*, SIAM J. Math. Anal., 44 (2012), pp. 2637–2665.
- [121] A. GANGULY, *Wong-Zakai type convergence in infinite dimensions*, Electron. J. Probab., 18 (2013), pp. no. 31, 34.
- [122] J. GARCIA, *Convergence of stochastic integrals and SDE's associated to approximations of the Gaussian white noise*, Adv. Appl. Stat., 10 (2008), pp. 155–177.
- [123] K. GAWĘDZKI AND M. VERGASSOLA, *Phase transition in the passive scalar advection*, Phys. D, 138 (2000), pp. 63–90.
- [124] M. GEISSERT, M. KOVÁCS, AND S. LARSSON, *Rate of weak convergence of the finite element method for the stochastic heat equation with additive noise*, BIT, 49 (2009), pp. 343–356.
- [125] A. GENZ AND B. D. KEISTER, *Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight*, J. Comput. Appl. Math., 71 (1996), pp. 299–309.
- [126] M. GERENCSÉR AND I. GYÖNGY, *Finite difference schemes for stochastic partial differential equations in Sobolev spaces*, ArXiv e-prints, (2013).
- [127] A. GERMANI, L. JETTO, AND M. PICCIONI, *Galerkin approximation for optimal linear filtering of infinite-dimensional linear systems*, SIAM J. Control Optim., 26 (1988), pp. 1287–1305.
- [128] A. GERMANI AND M. PICCIONI, *A Galerkin approximation for the Zakai equation*, in System modelling and optimization (Copenhagen, 1983), Springer, Berlin, 1984, pp. 415–423.

- [129] ———, *Semidiscretization of stochastic partial differential equations on  $\mathbf{R}^d$  by a finite-element technique*, *Stochastics*, 23 (1988), pp. 131–148.
- [130] T. GERSTNER, *Sparse Grid Quadrature Methods for Computational Finance*, PhD thesis, University of Bonn, Habilitation, 2007.
- [131] T. GERSTNER AND M. GRIEBEL, *Numerical integration using sparse grids*, *Numer. Algorithms*, 18 (1998), pp. 209–232.
- [132] R. G. GHANEM AND P. D. SPANOS, *Stochastic finite elements: a spectral approach*, Springer-Verlag, New York, 1991.
- [133] Ī. Ī. GĪHMAN AND A. V. SKOROHOD, *Stochastic differential equations*, Springer-Verlag, New York, 1972.
- [134] M. B. GILES, *Multilevel Monte Carlo path simulation*, *Oper. Res.*, 56 (2008), pp. 607–617.
- [135] ———, *Multilevel Monte Carlo methods*, in *Monte Carlo and quasi-Monte Carlo methods 2012*, Springer, 2013.
- [136] M. B. GILES AND C. REISINGER, *Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance*, *SIAM J. Financial Math.*, 3 (2012), pp. 572–592.
- [137] E. GOBET, G. PAGÈS, H. PHAM, AND J. PRINTEMS, *Discretization and simulation of the Zakai equation*, *SIAM J. Numer. Anal.*, 44 (2006), pp. 2505–2538 (electronic).
- [138] N. Y. GONCHARUK AND P. KOTELENEZ, *Fractional step method for stochastic evolution equations*, *Stochastic Process. Appl.*, 73 (1998), pp. 1–45.
- [139] I. G. GRAHAM, F. Y. KUO, D. NUYENS, R. SCHEICHL, AND I. H. SLOAN, *Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications*, *J. Comput. Phys.*, 230 (2011), pp. 3668–3694.

- [140] ———, *Quasi-Monte carlo finite element methods for elliptic PDEs with log-normal random coefficients*, tech. rep., SAM, ETHZ, 2013.
- [141] W. GRECKSCH AND P. E. KLOEDEN, *Time-discretised Galerkin approximations of parabolic stochastic PDEs*, Bull. Austral. Math. Soc., 54 (1996), pp. 79–85.
- [142] W. GRECKSCH AND H. LISEI, *Approximation of stochastic nonlinear equations of Schrödinger type by the splitting method*, Stoch. Anal. Appl., 31 (2013), pp. 314–335.
- [143] W. GRECKSCH AND B. SCHMALFUSS, *Approximation of the stochastic Navier-Stokes equation*, Mat. Apl. Comput., 15 (1996), pp. 227–239.
- [144] M. GRIEBEL, *Sparse grids and related approximation schemes for higher dimensional problems*, in Foundations of computational mathematics, Santander 2005, Cambridge Univ. Press, Cambridge, 2006, pp. 106–161.
- [145] M. GRIEBEL AND M. HOLTZ, *Dimension-wise integration of high-dimensional functions with applications to finance*, J. Complexity, 26 (2010), pp. 455–489.
- [146] M. GRIGORIU, *Control of time delay linear systems with Gaussian white noise*, PrEM., 12 (1997), pp. 89 – 96.
- [147] C. GUGG, H. KIELHÖFER, AND M. NIGGEMANN, *On the approximation of the stochastic Burgers equation*, Comm. Math. Phys., 230 (2002), pp. 181–199.
- [148] Q. GUO, W. XIE, AND T. MITSUI, *Convergence and stability of the split-step  $\theta$ -Milstein method for stochastic delay Hopfield neural networks*, Abstr. Appl. Anal., (2013), pp. no. 169214, 12.
- [149] S. J. GUO, *On the mollifier approximation for solutions of stochastic differential equations*, J. Math. Kyoto Univ., 22 (1982), pp. 243–254.

- [150] I. GYÖNGY, *On the approximation of stochastic differential equations*, Stochastics, 23 (1988), pp. 331–352.
- [151] —, *On the approximation of stochastic partial differential equations. I*, Stochastics, 25 (1988), pp. 59–85.
- [152] —, *On the approximation of stochastic partial differential equations. II*, Stochastics Stochastics Rep., 26 (1989), pp. 129–164.
- [153] —, *The stability of stochastic partial differential equations and applications. I*, Stochastics Stochastics Rep., 27 (1989), pp. 129–150.
- [154] —, *The stability of stochastic partial differential equations. II*, Stochastics Stochastics Rep., 27 (1989), pp. 189–233.
- [155] —, *The approximation of stochastic partial differential equations and applications in non-linear filtering*, Comput. Math. Appl., 19 (1990), pp. 47–63.
- [156] —, *On the support of the solutions of stochastic differential equations*, Teor. Veroyatnost. i Primenen., 39 (1994), pp. 649–653.
- [157] I. GYÖNGY, *Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise. I*, Potential Anal., 9 (1998), pp. 1–25.
- [158] —, *A note on Euler’s approximations*, Potential Anal., 8 (1998), pp. 205–216.
- [159] —, *Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise. II*, Potential Anal., 11 (1999), pp. 1–37.
- [160] —, *Approximations of stochastic partial differential equations*, in Stochastic partial differential equations and applications (Trento, 2002), Dekker, New York, 2002, pp. 287–307.
- [161] I. GYÖNGY AND N. KRYLOV, *On the rate of convergence of splitting-up approximations for SPDEs*, in Stochastic inequalities and applications, Birkhäuser, Basel, 2003, pp. 301–321.

- [162] ———, *On the splitting-up method and stochastic partial differential equations*, Ann. Probab., 31 (2003), pp. 564–591.
- [163] ———, *An accelerated splitting-up method for parabolic equations*, SIAM J. Math. Anal., 37 (2005), pp. 1070–1097.
- [164] ———, *Accelerated finite difference schemes for linear stochastic partial differential equations in the whole space*, SIAM J. Math. Anal., 42 (2010), pp. 2275–2296.
- [165] I. GYÖNGY AND T. MARTÍNEZ, *On numerical solution of stochastic partial differential equations of elliptic type*, Stochastics, 78 (2006), pp. 213–231.
- [166] I. GYÖNGY AND G. MICHALETZKY, *On Wong-Zakai approximations with  $\delta$ -martingales*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 460 (2004), pp. 309–324.
- [167] I. GYÖNGY AND D. NUALART, *Implicit scheme for quasi-linear parabolic partial differential equations perturbed by space-time white noise*, Stochastic Process. Appl., 58 (1995), pp. 57–72.
- [168] ———, *Implicit scheme for stochastic parabolic partial differential equations driven by space-time white noise*, Potential Anal., 7 (1997), pp. 725–757.
- [169] I. GYÖNGY, D. NUALART, AND M. SANZ-SOLÉ, *Approximation and support theorems in modulus spaces*, Probab. Theory Related Fields, 101 (1995), pp. 495–509.
- [170] I. GYÖNGY AND T. PRÖHLE, *On the approximation of stochastic differential equation and on Stroock-Varadhan's support theorem*, Comput. Math. Appl., 19 (1990), pp. 65–70.
- [171] I. GYÖNGY AND A. SHMATKOV, *Rate of convergence of Wong-Zakai approximations for stochastic partial differential equations*, Appl. Math. Optim., 54 (2006), pp. 315–341.
- [172] I. GYÖNGY AND P. R. STINGA, *Rate of convergence of Wong-Zakai approximations for stochastic partial differential equations*, ArXiv e-prints, (2012).

- [173] L. G. GYURKÓ AND T. J. LYONS, *Efficient and practical implementations of cubature on Wiener space*, in *Stochastic analysis 2010*, Springer, Heidelberg, 2011, pp. 73–111.
- [174] M. HAIRER AND J. MAAS, *A spatial version of the Itô-Stratonovich correction*, *Ann. Probab.*, 40 (2012), pp. 1675–1714.
- [175] M. HAIRER AND J. VOSS, *Approximations to the stochastic Burgers equation*, *J. Nonlinear Sci.*, 21 (2011), pp. 897–920.
- [176] E. J. HALL, *Accelerated spatial approximations for time discretized stochastic partial differential equations*, *SIAM J. Math. Anal.*, 44 (2012), pp. 3162–3185.
- [177] ———, *Higher Order Spatial Approximations for Degenerate Parabolic Stochastic Partial Differential Equations*, *SIAM J. Math. Anal.*, 45 (2013), pp. 2071–2098.
- [178] R. Z. HAS'MINSKIĀ, *Stochastic stability of differential equations*, Sijthoff & Noordhoff, 1980.
- [179] E. HAUSENBLAS, *Numerical analysis of semilinear stochastic evolution equations in Banach spaces*, *J. Comput. Appl. Math.*, 147 (2002), pp. 485–516.
- [180] ———, *Approximation for semilinear stochastic evolution equations*, *Potential Anal.*, 18 (2003), pp. 141–186.
- [181] E. HAUSENBLAS, *Weak approximation for semilinear stochastic evolution equations*, in *Stochastic analysis and related topics VIII*, Birkhäuser, Basel, 2003, pp. 111–128.
- [182] ———, *Wong-Zakai type approximation of SPDEs of Lévy noise*, *Acta Appl. Math.*, 98 (2007), pp. 99–134.
- [183] ———, *Weak approximation of the stochastic wave equation*, *J. Comput. Appl. Math.*, 235 (2010), pp. 33–58.

- [184] J. HE, *Numerical analysis for stochastic age-dependent population equations with diffusion*, in Advances in Electronic Commerce, Web Application and Communication, D. Jin and S. Lin, eds., Springer Berlin Heidelberg, 2012, pp. 37–43.
- [185] R. L. HERMAN AND A. ROSE, *Numerical realizations of solutions of the stochastic KdV equation*, Math. Comput. Simulation, 80 (2009), pp. 164–172.
- [186] D. J. HIGHAM, X. MAO, AND A. M. STUART, *Strong convergence of Euler-type methods for nonlinear stochastic differential equations*, SIAM J. Numer. Anal., 40 (2002), pp. 1041–1063.
- [187] D. J. HIGHAM, X. MAO, AND L. SZPRUCH, *Convergence, non-negativity and stability of a new Milstein scheme with applications to finance*, ArXiv e-prints, (2012).
- [188] D. G. HOBSON AND L. C. G. ROGERS, *Complete models with stochastic volatility*, Math. Finance, 8 (1998), pp. 27–48.
- [189] N. HOFMANN AND T. MÜLLER-GRONBACH, *A modified Milstein scheme for approximation of stochastic delay differential equations with constant time lag*, J. Comput. Appl. Math., 197 (2006), pp. 89–121.
- [190] H. HOLDEN, B. ØKSENDAL, J. UBOE, AND T. ZHANG, *Stochastic partial differential equations*, Birkhäuser Boston Inc., Boston, MA, 1996.
- [191] H. HOLDEN AND N. H. RISEBRO, *Conservation laws with a random source*, Appl. Math. Optim., 36 (1997), pp. 229–241.
- [192] T. Y. HOU, W. LUO, B. ROZOVSKII, AND H.-M. ZHOU, *Wiener chaos expansions and numerical solutions of randomly forced equations of fluid mechanics*, J. Comput. Phys., 216 (2006), pp. 687–706.
- [193] Y. HU, *Semi-implicit Euler-Maruyama scheme for stiff stochastic equations*, in Stochastic analysis and related topics, Birkhäuser Boston, Boston, MA, 1996, pp. 183–202.

- [194] Y. HU, G. KALLIANPUR, AND J. XIONG, *An approximation for the Zakai equation*, Appl. Math. Optim., 45 (2002), pp. 23–44.
- [195] Y. HU, S.-E. A. MOHAMMED, AND F. YAN, *Discrete-time approximations of stochastic delay equations: the Milstein scheme*, Ann. Probab., 32 (2004), pp. 265–314.
- [196] C. HUANG, S. GAN, AND D. WANG, *Delay-dependent stability analysis of numerical methods for stochastic delay differential equations*, J. Comput. Appl. Math., 236 (2012), pp. 3514–3527.
- [197] M. HUTZENTHALER AND A. JENTZEN, *Numerical approximation of stochastic differential equations with non-globally Lipschitz continuous coefficients*, ArXiv e-prints, (2012).
- [198] M. HUTZENTHALER, A. JENTZEN, AND P. E. KLOEDEN, *Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients*, Ann. Appl. Probab., 22 (2012), pp. 1611–1641.
- [199] M. HUTZENTHALER, A. JENTZEN, AND P. E. KLOEDEN, *Divergence of the multilevel Monte Carlo Euler method for nonlinear stochastic differential equations*, Ann. Appl. Probab., 23 (2013), pp. 1913–1966.
- [200] N. IKEDA, S. NAKAO, AND Y. YAMATO, *A class of approximations of Brownian motion*, Publ. Res. Inst. Math. Sci., 13 (1977/78), pp. 285–300.
- [201] N. IKEDA AND S. TANIGUCHI, *The Itô-Nisio theorem, quadratic Wiener functionals, and 1-solitons*, Stochastic Process. Appl., 120 (2010), pp. 605–621.
- [202] N. IKEDA AND S. WATANABE, *Stochastic differential equations and diffusion processes*, North-Holland Publishing Co., Amsterdam, 1981.
- [203] K. ITO, *Approximation of the Zakai equation for nonlinear filtering*, SIAM J. Control Optim., 34 (1996), pp. 620–634.

- [204] K. ITÔ AND M. NISIO, *On the convergence of sums of independent Banach space valued random variables*, Osaka J. Math., 5 (1968), pp. 35–48.
- [205] K. ITO AND B. ROZOVSKII, *Approximation of the Kushner equation for nonlinear filtering*, SIAM J. Control Optim., 38 (2000), pp. 893–915 (electronic).
- [206] A. JENTZEN, *Pathwise numerical approximation of SPDEs with additive noise under non-global Lipschitz coefficients*, Potential Anal., 31 (2009), pp. 375–404.
- [207] ———, *Taylor expansions of solutions of stochastic partial differential equations*, Discrete Contin. Dyn. Syst. Ser. B, 14 (2010), pp. 515–557.
- [208] ———, *Higher order pathwise numerical approximations of SPDEs with additive noise*, SIAM J. Numer. Anal., 49 (2011), pp. 642–667.
- [209] A. JENTZEN AND P. E. KLOEDEN, *The numerical approximation of stochastic partial differential equations*, Milan J. Math., 77 (2009), pp. 205–244.
- [210] A. JENTZEN AND P. E. KLOEDEN, *Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 465 (2009), pp. 649–667.
- [211] A. JENTZEN AND M. RÖCKNER, *A Milstein scheme for SPDEs*, ArXiv e-prints, (2012).
- [212] G.-S. JIANG AND C.-W. SHU, *Efficient implementation of weighted ENO schemes*, J. Comput. Phys., 126 (1996), pp. 202–228.
- [213] E. KALPINELLI, N. FRANGOS, AND A. YANNAKOPOULOS, *Numerical methods for hyperbolic spdes: a wiener chaos approach*, Stoch PDE: Anal Comp., (2013), pp. 1–28.
- [214] I. KARATZAS AND S. E. SHREVE, *Brownian motion and stochastic calculus*, Springer-Verlag, New York, second ed., 1991.

- [215] M. A. KATSOUKAKIS, G. T. KOSSIORIS, AND O. LAKKIS, *Noise regularization and computations for the 1-dimensional stochastic Allen-Cahn problem*, *Interfaces Free Bound.*, 9 (2007), pp. 1–30.
- [216] P. E. KLOEDEN AND A. JENTZEN, *Pathwise convergent higher order numerical schemes for random ordinary differential equations*, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 463 (2007), pp. 2929–2944.
- [217] P. E. KLOEDEN, G. J. LORD, A. NEUENKIRCH, AND T. SHARDLOW, *The exponential integrator scheme for stochastic partial differential equations: pathwise error bounds*, *J. Comput. Appl. Math.*, 235 (2011), pp. 1245–1260.
- [218] P. E. KLOEDEN AND E. PLATEN, *Numerical solution of stochastic differential equations*, Springer-Verlag, Berlin, 1992.
- [219] P. E. KLOEDEN AND T. SHARDLOW, *The Milstein scheme for stochastic delay differential equations without using anticipative calculus*, *Stoch. Anal. Appl.*, 30 (2012), pp. 181–202.
- [220] P. E. KLOEDEN AND S. SHOTT, *Linear-implicit strong schemes for Itô-Galerkin approximations of stochastic PDEs*, *J. Appl. Math. Stochastic Anal.*, 14 (2001), pp. 47–53. Special issue: Advances in applied stochastic.
- [221] O. M. KNIO AND O. P. LE MAÎTRE, *Uncertainty propagation in CFD using polynomial chaos decomposition*, *Fluid Dynam. Res.*, 38 (2006), pp. 616–640.
- [222] L. KOCIS AND W. J. WHITEN, *Computational investigations of low-discrepancy sequences*, *ACM Trans. Math. Softw.*, 23 (1997), pp. 266–294.
- [223] F. KONECNY, *On Wong-Zakai approximation of stochastic differential equations*, *J. Multivariate Anal.*, 13 (1983), pp. 605–611.

- [224] R. KORN, E. KORN, AND G. KROISANDT, *Monte Carlo methods and models in finance and insurance*, CRC Press, Boca Raton, FL, 2010.
- [225] G. T. KOSSIORIS AND G. E. ZOURARIS, *Fully-discrete finite element approximations for a fourth-order linear stochastic parabolic equation with additive space-time white noise*, M2AN Math. Model. Numer. Anal., 44 (2010), pp. 289–322.
- [226] G. T. KOSSIORIS AND G. E. ZOURARIS, *Finite element approximations for a linear Cahn-Hilliard-Cook equation driven by the space derivative of a space-time white noise*, Discrete Contin. Dyn. Syst. Ser. B, 18 (2013), pp. 1845–1872.
- [227] G. T. KOSSIORIS AND G. E. ZOURARIS, *Finite element approximations for a linear fourth-order parabolic SPDE in two and three space dimensions with additive space-time white noise*, Appl. Numer. Math., 67 (2013), pp. 243–261.
- [228] P. KOTELENEZ, *Stochastic ordinary and stochastic partial differential equations*, Springer, New York, 2008.
- [229] M. KOVÁCS, S. LARSSON, AND F. LINDGREN, *Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise*, BIT, 52 (2012), pp. 85–108.
- [230] ———, *Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise ii. fully discrete schemes*, BIT, 53 (2013), pp. 497–525.
- [231] M. KOVÁCS, S. LARSSON, AND A. MESFORUSH, *Finite element approximation of the Cahn-Hilliard-Cook equation*, SIAM J. Numer. Anal., 49 (2011), pp. 2407–2429.
- [232] M. KOVÁCS, S. LARSSON, AND F. SAEDPANAH, *Finite element approximation of the linear stochastic wave equation with additive noise*, SIAM J. Numer. Anal., 48 (2010), pp. 408–427.
- [233] R. H. KRAICHNAN, *Small-scale structure of a scalar field convected by turbulence*, Phys. Fluids, 11 (1968), pp. 945–953.

- [234] I. KRÖKER AND C. ROHDE, *Finite volume schemes for hyperbolic balance laws with multiplicative noise*, Appl. Numer. Math., 62 (2012), pp. 441 – 456.
- [235] R. KRUSE, *Consistency and stability of a Milstein-Galerkin finite element scheme for semi-linear SPDE*, ArXiv e-prints, (2013).
- [236] R. KRUSE, *Optimal error estimates of Galerkin finite element methods for stochastic partial differential equations with multiplicative noise*, IMA J. Numer. Anal., (2013).
- [237] N. V. KRYLOV, *Introduction to the theory of diffusion processes*, AMS, Providence, RI, 1995.
- [238] U. KÜCHLER AND E. PLATEN, *Strong discrete time approximation of stochastic differential equations with time delay*, Math. Comput. Simulation, 54 (2000), pp. 189–205.
- [239] F. Y. KUO, C. SCHWAB, AND I. H. SLOAN, *Multi-level quasi-Monte carlo finite element methods for a class of elliptic partial differential equations with random coefficients*, ArXiv e-prints, (2012).
- [240] F. Y. KUO, C. SCHWAB, AND I. H. SLOAN, *Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients*, SIAM J. Numer. Anal., 50 (2012), pp. 3351–3374.
- [241] T. G. KURTZ AND P. PROTTER, *Weak limit theorems for stochastic integrals and stochastic differential equations*, Ann. Probab., 19 (1991), pp. 1035–1070.
- [242] ———, *Wong-Zakai corrections, random evolutions, and simulation schemes for SDEs*, in Stochastic analysis, Academic Press, Boston, MA, 1991, pp. 331–346.
- [243] T. G. KURTZ AND J. XIONG, *Numerical solutions for a class of SPDEs with application to filtering*, in Stochastics in finite and infinite dimensions, Trends Math., Birkhäuser Boston, Boston, MA, 2001, pp. 233–258.

- [244] H. J. KUSHNER, *On the differential equations satisfied by conditional probability densities of Markov processes, with applications*, J. Soc. Indust. Appl. Math. Ser. A Control, 2 (1964), pp. 106–119.
- [245] A. LANG, *A Lax equivalence theorem for stochastic differential equations*, J. Comput. Appl. Math., 234 (2010), pp. 3387–3396.
- [246] ———, *Almost sure convergence of a Galerkin approximation for SPDEs of Zakai type driven by square integrable martingales*, J. Comput. Appl. Math., 236 (2012), pp. 1724–1732.
- [247] A. LANG, P.-L. CHOW, AND J. POTTHOFF, *Almost sure convergence for a semidiscrete Milstein scheme for SPDEs of Zakai type*, Stochastics, 82 (2010), pp. 315–326.
- [248] S. LARSSON AND A. MESFORUSH, *Finite-element approximation of the linearized Cahn-Hilliard-Cook equation*, IMA J. Numer. Anal., 31 (2011), pp. 1315–1333.
- [249] F. LE GLAND, *Splitting-up approximation for SPDEs and SDEs with application to nonlinear filtering*, in Stochastic partial differential equations and their applications, Springer, Berlin, 1992, pp. 177–187.
- [250] H. W. LIEPMANN AND A. ROSHKO, *Elements of gasdynamics*, John Wiley & Sons Inc., New York, 1957.
- [251] G. LIN, C. H. SU, AND G. E. KARNIADAKIS, *The stochastic piston problem*, Proc. Natl. Acad. Sci. USA, 101 (2004), pp. 15840–15845 (electronic).
- [252] F. LINDNER AND R. SCHILLING, *Weak order for the discretization of the stochastic heat equation driven by impulsive noise*, Potential Anal., 38 (2013), pp. 345–379.
- [253] C. LITTERER AND T. LYONS, *High order recombination and an application to cubature on Wiener space*, Ann. Appl. Probab., 22 (2012), pp. 1301–1327.

- [254] D. LIU, *Convergence of the spectral method for stochastic Ginzburg-Landau equation driven by space-time white noise*, Commun. Math. Sci., 1 (2003), pp. 361–375.
- [255] J. LIU, *A mass-preserving splitting scheme for the stochastic Schrödinger equation with multiplicative noise*, IMA J. Numer. Anal., (2013).
- [256] ———, *Order of convergence of splitting schemes for both deterministic and stochastic nonlinear Schrödinger equations*, SIAM J. Numer. Anal., 51 (2013), pp. 1911–1932.
- [257] M. LIU, W. CAO, AND Z. FAN, *Convergence and stability of the semi-implicit Euler method for a linear stochastic differential delay equation*, J. Comput. Appl. Math., 170 (2004), pp. 255–268.
- [258] J. A. LONDOÑO AND A. M. RAMIREZ, *Numerical performance of some Wong-Zakai type approximations for stochastic differential equations*, tech. rep., Department of Mathematics, National University of Colombia, Bogota, Colombia, 2006.
- [259] G. J. LORD AND J. ROUGEMONT, *A numerical scheme for stochastic PDEs with Gevrey regularity*, IMA J. Numer. Anal., 24 (2004), pp. 587–604.
- [260] G. J. LORD AND T. SHARDLOW, *Postprocessing for stochastic parabolic partial differential equations*, SIAM J. Numer. Anal., 45 (2007), pp. 870–889 (electronic).
- [261] G. J. LORD AND A. TAMBUE, *A modified semi-implicit Euler-Maruyama scheme for finite element discretization of SPDEs*, ArXiv e-prints, (2010).
- [262] G. J. LORD AND A. TAMBUE, *Stochastic exponential integrators for the finite element discretization of SPDEs for multiplicative and additive noise*, IMA J. Numer. Anal., 33 (2013), pp. 515–543.
- [263] G. J. LORD AND V. THÜMMLER, *Computing stochastic traveling waves*, SIAM J. Sci. Comput., 34 (2012), pp. B24–B43.

- [264] S. V. LOTOTSKIĬ AND B. L. ROZOVSKIĬ, *The passive scalar equation in a turbulent incompressible Gaussian velocity field*, Uspekhi Mat. Nauk, 59 (2004), pp. 105–120.
- [265] S. LOTOTSKY, R. MIKULEVICIUS, AND B. L. ROZOVSKII, *Nonlinear filtering revisited: a spectral approach*, SIAM J. Control Optim., 35 (1997), pp. 435–461.
- [266] S. LOTOTSKY AND B. ROZOVSKII, *Stochastic differential equations: a Wiener chaos approach*, in From stochastic calculus to mathematical finance, Springer, Berlin, 2006, pp. 433–506.
- [267] S. V. LOTOTSKY, *Wiener chaos and nonlinear filtering*, Appl. Math. Optim., 54 (2006), pp. 265–291.
- [268] S. V. LOTOTSKY AND B. L. ROZOVSKII, *Wiener chaos solutions of linear stochastic evolution equations*, Ann. Probab., 34 (2006), pp. 638–662.
- [269] T. LYONS AND N. VICTOIR, *Cubature on Wiener space*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 460 (2004), pp. 169–198.
- [270] V. MACKEVIČIUS, *On Ikeda-Nakao-Yamato type approximations*, Litovsk. Mat. Sb., 30 (1990), pp. 752–757.
- [271] ———, *On approximation of stochastic differential equations with coefficients depending on the past*, Liet. Mat. Rink., 32 (1992), pp. 285–298.
- [272] V. MACKYAVICHYUS, *Symmetric stochastic differential equations with nonsmooth coefficients*, Mat. Sb. (N.S.), 116(158) (1981), pp. 585–592, 608.
- [273] P. MALLIAVIN, *Stochastic calculus of variation and hypoelliptic operators*, in Proc. Int. Symp. SDE. Kyoto 1976, K. Ito, ed., Springer, Kinokuniya, 1978, pp. 195–263.
- [274] H. MANOUZI, *A finite element approximation of linear stochastic PDEs driven by multiplicative white noise*, Int. J. Comput. Math., 85 (2008), pp. 527–546.

- [275] H. MANOUZI, *Numerical solutions of SPDEs involving white noise*, Int. J. Tomogr. Stat., 10 (2008), pp. 96–108.
- [276] H. MANOUZI, *SPDEs driven by additive and multiplicative white noises: a numerical study*, Math. Comput. Simulation, 81 (2011), pp. 2234–2243.
- [277] H. MANOUZI AND M. SEAID, *Solving wick-stochastic water waves using a Galerkin finite element method*, Math. Comput. Simulation, 79 (2009), pp. 3523 – 3533.
- [278] H. MANOUZI, M. SEAID, AND M. ZAHRI, *Wick-stochastic finite element solution of reaction-diffusion problems*, J. Comput. Appl. Math., 203 (2007), pp. 516–532.
- [279] H. MANOUZI AND T. G. THETING, *Mixed finite element approximation for the stochastic pressure equation of Wick type*, IMA J. Numer. Anal., 24 (2004), pp. 605–634.
- [280] H. MANOUZI AND T. G. THETING, *Numerical analysis of the stochastic Stokes equations of Wick type*, Numer. Methods Partial Differential Equations, 23 (2007), pp. 73–92.
- [281] X. MAO, *Stochastic differential equations and their applications*, Horwood Publishing Limited, Chichester, 1997.
- [282] X. MAO AND L. SZPRUCH, *Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients*, J. Comput. Appl. Math., 238 (2013), pp. 14–28.
- [283] X. MAO AND L. SZPRUCH, *Strong convergence rates for backward Euler-Maruyama method for non-linear dissipative-type stochastic differential equations with super-linear diffusion coefficients*, Stochastics, 85 (2013), pp. 144–171.
- [284] T. MARTÍNEZ AND M. SANZ-SOLÉ, *A lattice scheme for stochastic partial differential equations of elliptic type in dimension  $d \geq 4$* , Appl. Math. Optim., 54 (2006), pp. 343–368.

- [285] R. MARTY, *On a splitting scheme for the nonlinear Schrödinger equation in a random medium*, Commun. Math. Sci., 4 (2006), pp. 679–705.
- [286] G. MASTROIANNI AND G. MONEGATO, *Error estimates for Gauss-Laguerre and Gauss-Hermite quadrature formulas*, in Approximation and computation, Birkhäuser Boston, 1994, pp. 421–434.
- [287] J. MATOUŠEK, *On the  $L^2$ -discrepancy for anchored boxes*, J. Complex., 14 (1998), pp. 527–556.
- [288] J. C. MATTINGLY, A. M. STUART, AND D. J. HIGHAM, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stochastic Process. Appl., 101 (2002), pp. 185–232.
- [289] E. J. MCSHANE, *Stochastic differential equations and models of random processes*, in Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory, Berkeley, Calif., 1972, Univ. California Press, pp. 263–294.
- [290] R. MIKULEVICIUS AND B. ROZOVSKII, *Separation of observations and parameters in nonlinear filtering*, in Decision and Control, 1993., Proceedings of the 32nd IEEE Conference on, vol. 2, dec 1993, pp. 1564–1569.
- [291] R. MIKULEVICIUS AND B. ROZOVSKII, *Linear parabolic stochastic PDEs and Wiener chaos*, SIAM J. Math. Anal., 29 (1998), pp. 452–480 (electronic).
- [292] R. MIKULEVICIUS AND B. L. ROZOVSKII, *Stochastic Navier-Stokes equations for turbulent flows*, SIAM J. Math. Anal., 35 (2004), pp. 1250–1310.
- [293] A. MILLET AND P.-L. MORIEN, *On implicit and explicit discretization schemes for parabolic SPDEs in any dimension*, Stochastic Process. Appl., 115 (2005), pp. 1073–1106.

- [294] A. MILLET AND M. SANZ-SOLÉ, *A simple proof of the support theorem for diffusion processes*, in Séminaire de Probabilités, XXVIII, vol. 1583 of Lecture Notes in Math., Springer, Berlin, 1994, pp. 36–48.
- [295] ———, *Approximation and support theorem for a wave equation in two space dimensions*, Bernoulli, 6 (2000), pp. 887–915.
- [296] G. N. MIL'SHTEĬN, *A theorem on the order of convergence of mean-square approximations of solutions of systems of stochastic differential equations*, Teor. Veroyatnost. i Primenen., 32 (1987), pp. 809–811.
- [297] G. N. MILSTEIN, *Numerical integration of stochastic differential equations*, Kluwer Academic Publishers Group, Dordrecht, 1995.
- [298] G. N. MILSTEIN, E. PLATEN, AND H. SCHURZ, *Balanced implicit methods for stiff stochastic systems*, SIAM J. Numer. Anal., 35 (1998), pp. 1010–1019.
- [299] G. N. MILSTEIN, Y. M. REPIN, AND M. V. TRETYAKOV, *Numerical methods for stochastic systems preserving symplectic structure*, SIAM J. Numer. Anal., 40 (2002), pp. 1583–1604.
- [300] G. N. MILSTEIN AND M. V. TRETYAKOV, *Numerical methods in the weak sense for stochastic differential equations with small noise*, SIAM J. Numer. Anal., 34 (1997), pp. 2142–2167.
- [301] ———, *Stochastic numerics for mathematical physics*, Springer-Verlag, Berlin, 2004.
- [302] ———, *Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients*, SIAM J. Numer. Anal., 43 (2005), pp. 1139–1154.
- [303] ———, *Computing ergodic limits for Langevin equations*, Phys. D, 229 (2007), pp. 81–95.
- [304] ———, *Solving parabolic stochastic partial differential equations via averaging over characteristics*, Math. Comp., 78 (2009), pp. 2075–2106.

- [305] J. MING AND M. GUNZBURGER, *Efficient numerical methods for stochastic partial differential equations through transformation to equations driven by correlated noise*, Int. J. Uncertain. Quantif., 3 (2013), pp. 321–339.
- [306] S. MISHRA AND C. SCHWAB, *Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data*, Math. Comp., 81 (2012), pp. 1979–2018.
- [307] S. MISHRA, C. SCHWAB, AND J. ŠUKYS, *Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions*, J. Comput. Phys., 231 (2012), pp. 3365–3388.
- [308] ———, *Multilevel Monte Carlo finite volume methods for shallow water equations with uncertain topography in multi-dimensions*, SIAM J. Sci. Comput., 34 (2012), pp. B761–B784.
- [309] Y. S. MISHURA AND G. M. SHEVCHENKO, *Approximation schemes for stochastic differential equations in a Hilbert space*, Teor. Veroyatn. Primen., 51 (2006), pp. 476–495.
- [310] D. S. MITRINOVIĆ, J. E. PEČARIĆ, AND A. M. FINK, *Inequalities involving functions and their integrals and derivatives*, Kluwer Academic Publishers Group, Dordrecht, 1991.
- [311] K. MOHAMED, M. SEAID, AND M. ZAHRI, *A finite volume method for scalar conservation laws with stochastic timespace dependent flux functions*, J. Comput. Appl. Math., 237 (2013), pp. 614 – 632.
- [312] S. E. A. MOHAMMED, *Stochastic functional differential equations*, Pitman Publishing Ltd, Boston, MA, 1984.
- [313] C. M. MORA, *Numerical solution of conservative finite-dimensional stochastic Schrödinger equations*, Ann. Appl. Probab., 15 (2005), pp. 2144–2171.

- [314] M. MOTAMED, F. NOBILE, AND R. TEMPONE, *A stochastic collocation method for the second order wave equation with a discontinuous random speed*, Numer. Math., (2012), pp. 1–44.
- [315] T. MÜLLER-GRONBACH AND K. RITTER, *An implicit Euler scheme with non-uniform time discretization for heat equations with multiplicative noise*, BIT, 47 (2007), pp. 393–418.
- [316] ———, *Lower bounds and nonuniform time discretization for approximation of stochastic heat equations*, Found. Comput. Math., 7 (2007), pp. 135–181.
- [317] T. MULLER-GRONBACH, K. RITTER, AND L. YAROSLAVTSEVA, *Derandomization of the Euler scheme for scalar stochastic differential equations*, J. Complexity, 28 (2012), pp. 139–153.
- [318] H. N. NAJM, *Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics*, in Annu. Rev. Fluid Mech., Vol. 41, Annual Reviews, Palo Alto, CA, 2009, pp. 35–52.
- [319] A. NEUENKIRCH AND L. SZPRUCH, *First order strong approximations of scalar SDEs with values in a domain*, ArXiv e-prints, (2012).
- [320] H. NIEDERREITER, *Random number generation and quasi-Monte Carlo methods*, SIAM, Philadelphia, PA, 1992.
- [321] M. NINOMIYA, *Application of the Kusuoka approximation with a tree-based branching algorithm to the pricing of interest-rate derivatives under the HJM model*, LMS J. Comput. Math., 13 (2010), pp. 208–221.
- [322] F. NOBILE AND R. TEMPONE, *Analysis and implementation issues for the numerical approximation of parabolic equations with random coefficients*, Internat. J. Numer. Methods Engrg., 80 (2009), pp. 979–1006.

- [323] F. NOBILE, R. TEMPONE, AND C. G. WEBSTER, *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2411–2442.
- [324] E. NOVAK AND K. RITTER, *Simple cubature formulas with high polynomial exactness*, Constr. Approx., 15 (1999), pp. 499–522.
- [325] A. NOWAK, *A Wong-Zakai type theorem for stochastic systems of Burgers equations*, Panamer. Math. J., 16 (2006), pp. 1–25.
- [326] A. NOWAK AND K. TWARDOWSKA, *On support and invariance theorems for a stochastic system of Burgers equations*, Demonstratio Math., 39 (2006), pp. 691–710.
- [327] D. NUALART AND M. ZAKAI, *On the relation between the Stratonovich and Ogawa integrals*, Ann. Probab., 17 (1989), pp. 1536–1540.
- [328] S. OGAWA, *Quelques propriétés de l'intégrale stochastique du type noncausal*, Japan J. Appl. Math., 1 (1984), pp. 405–416.
- [329] ———, *On a deterministic approach to the numerical solution of the SDE*, Math. Comput. Simulation, 55 (2001), pp. 209–214.
- [330] B. ØKSENDAL, *Stochastic differential equations*, Universitext, Springer-Verlag, Berlin, sixth ed., 2003. An introduction with applications.
- [331] G. PAGÈS AND H. PHAM, *Optimal quantization methods for nonlinear filtering with discrete-time observations*, Bernoulli, 11 (2005), pp. 893–932.
- [332] G. PAGÈS AND J. PRINTEMS, *Optimal quadratic quantization for numerics: the Gaussian case*, Monte Carlo Methods Appl., 9 (2003), pp. 135–165.
- [333] R. PALEY AND N. WIENER, *Fourier transforms in the complex domain*, AMS, 1934.

- [334] M. D. PAOLA AND A. PIRROTTA, *Time delay induced effects on control of linear systems under random excitation*, PrEM., 16 (2001), pp. 43–51.
- [335] K. PETRAS, *SmolPack: a software for smolyak quadrature with clenshaw-curtis basis-sequence*. [http://people.sc.fsu.edu/~jburkardt/c\\_src/smolpack/smolpack.html](http://people.sc.fsu.edu/~jburkardt/c_src/smolpack/smolpack.html), 2003.
- [336] J. PICARD, *Approximation of nonlinear filtering problems and order of convergence*, in Filtering and control of random processes (Paris, 1983), Springer, Berlin, 1984, pp. 219–236.
- [337] ———, *Approximation of stochastic differential equations and application of the stochastic calculus of variations to the rate of convergence*, in Stochastic analysis and related topics (Silivri, 1986), Springer, Berlin, 1988, pp. 267–287.
- [338] F. PIZZI, *Stochastic collocation and option pricing*, Master’s thesis, Politecnico di Milano, 2012.
- [339] J. PRINTEMS, *On the discretization in time of parabolic stochastic partial differential equations*, M2AN Math. Model. Numer. Anal., 35 (2001), pp. 1055–1078.
- [340] P. PROTTER, *Approximations of solutions of stochastic differential equations driven by semimartingales*, Ann. Probab., 13 (1985), pp. 716–743.
- [341] R. QI AND X. YANG, *Weak convergence of finite element method for stochastic elastic equation driven by additive noise*, J. Sci. Comput., 56 (2013), pp. 450–470.
- [342] L. QUER-SARDANYONS AND M. SANZ-SOLÉ, *Space semi-discretisations for a stochastic wave equation*, Potential Anal., 24 (2006), pp. 303–332.
- [343] C. REISINGER, *Mean-square stability and error analysis of implicit time-stepping schemes for linear parabolic SPDEs with multiplicative Wiener noise in the first derivative*, Int. J. Comput. Math., 89 (2012), pp. 2562–2575.

- [344] A. J. ROBERTS, *A step towards holistic discretisation of stochastic partial differential equations*, ANZIAM J., 45 (2003/04), pp. C1–C15.
- [345] C. ROTH, *A combination of finite difference and Wong-Zakai methods for hyperbolic stochastic partial differential equations*, Stoch. Anal. Appl., 24 (2006), pp. 221–240.
- [346] ———, *Weak approximations of solutions of a first order hyperbolic stochastic partial differential equation*, Monte Carlo Methods Appl., 13 (2007), pp. 117–133.
- [347] B. L. ROZOVSKIĬ, *Stochastic evolution systems*, Kluwer, 1990.
- [348] G. ROZZA, D. B. P. HUYNH, AND A. T. PATERA, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: application to transport and continuum mechanics*, Arch. Comput. Methods Eng., 15 (2008), pp. 229–275.
- [349] M. SANGO, *Splitting-up scheme for nonlinear stochastic hyperbolic equations*, Forum Math., 25 (2013), pp. 931–965.
- [350] B. SAUSSEREAU AND I. L. STOICA, *Scalar conservation laws with fractional stochastic forcing: Existence, uniqueness and invariant measure*, Stochastic Process. Appl., 122 (2012), pp. 1456 – 1486.
- [351] B. SCHMALFUSS, *On approximation of the stochastic Navier-Stokes equations*, Wiss. Z. Tech. Hochsch. Leuna-Merseburg, 27 (1985), pp. 605–612.
- [352] C. SCHWAB AND C. J. GITTELSON, *Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs*, Acta Numer., 20 (2011), pp. 291–467.
- [353] T. SHARDLOW, *Numerical methods for stochastic parabolic PDEs*, Numer. Funct. Anal. Optim., 20 (1999), pp. 121–145.

- [354] ———, *Weak convergence of a numerical method for a stochastic heat equation*, BIT, 43 (2003), pp. 179–193.
- [355] G. SHEVCHENKO, *Rate of convergence of discrete approximations of solutions to stochastic differential equations in a Hilbert space*, Theory Probab. Math. Statist., 69 (2004), pp. 187–199.
- [356] I. H. SLOAN AND S. JOE, *Lattice methods for multiple integration*, Oxford University Press, New York, 1994.
- [357] I. H. SLOAN AND H. WOŹNIAKOWSKI, *When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals?*, J. Complexity, 14 (1998), pp. 1–33.
- [358] S. A. SMOLYAK, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Soviet Math. Dokl., 4 (1963), pp. 240–243.
- [359] A. R. SOHEILI, M. B. NIASAR, AND M. AREZOOMANDAN, *Approximation of stochastic parabolic differential equations with two different finite difference schemes*, Bull. Iranian Math. Soc., 37 (2011), pp. 61–83.
- [360] V. N. STANCIULESCU AND M. V. TRETAKOV, *Numerical solution of the Dirichlet problem for linear parabolic SPDEs based on averaging over characteristics*, in Stochastic analysis 2010, Springer, Heidelberg, 2011, pp. 191–212.
- [361] R. L. STRATONOVİČ, *A new form of representing stochastic integrals and equations*, Vestnik Moskov. Univ. Ser. I Mat. Meh., 1964 (1964), pp. 3–12.
- [362] D. W. STROOCK AND S. R. S. VARADHAN, *On the support of diffusion processes with applications to the strong maximum principle*, in Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. III, Berkeley, CA, 1972, Univ. California Press, pp. 333–359.

- [363] M. SUN AND R. GLOWINSKI, *Pathwise approximation and simulation for the Zakai filtering equation through operator splitting*, *Calcolo*, 30 (1993), pp. 219–239 (1994).
- [364] H. J. SUSSMANN, *On the gap between deterministic and stochastic ordinary differential equations*, *Ann. Probability*, 6 (1978), pp. 19–41.
- [365] ———, *Limits of the Wong-Zakai type with a modified drift term*, in *Stochastic analysis*, Academic Press, Boston, MA, 1991, pp. 475–493.
- [366] D. TALAY, *Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme*, *Markov Process. Related Fields*, 8 (2002), pp. 163–198.
- [367] D. M. TARTAKOVSKY, M. DENTZ, AND P. C. LICHTNER, *Probability density functions for advective-reactive transport with uncertain reaction rates*, *Water Resources Research*, 45 (2009).
- [368] M. TATANG AND G. MCRAE, *Direct treatment of uncertainty in models of reaction and transport*, tech. rep., Department of Chemical Engineering, MIT, 1994.
- [369] A. TECKENTRUP, R. SCHEICHL, M. GILES, AND E. ULLMANN, *Further analysis of multilevel monte carlo methods for elliptic pdes with random coefficients*, *Numer. Math.*, (2013), pp. 1–32.
- [370] A. L. TECKENTRUP, *Multilevel Monte Carlo methods for highly heterogeneous media*, in *Proceedings of the Winter Simulation Conference*, Winter Simulation Conference, 2012, pp. 32:1–32:12.
- [371] G. TESSITORE AND J. ZABCZYK, *Wong-Zakai approximations of stochastic evolution equations*, *J. Evol. Equ.*, 6 (2006), pp. 621–655.

- [372] T. G. THETING, *Solving Wick-stochastic boundary value problems using a finite element method*, Stochastics Stochastics Rep., 70 (2000), pp. 241–270.
- [373] ———, *Solving parabolic Wick-stochastic boundary value problems using a finite element method*, Stoch. Stoch. Rep., 75 (2003), pp. 49–77.
- [374] ———, *Numerical solution of Wick-stochastic partial differential equations*, in Proceedings of the International Conference on Stochastic Analysis and Applications, Dordrecht, 2004, Kluwer Acad. Publ., pp. 303–349.
- [375] V. THOMÉE, *Galerkin finite element methods for parabolic problems*, Springer-Verlag, Berlin, second ed., 2006.
- [376] M. V. TRET'YAKOV AND Z. ZHANG, *A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications*, SIAM J. Numer. Anal., (To appear).
- [377] L. S. TSIMRING AND A. PIKOVSKY, *Noise-induced dynamics in bistable systems with delay*, Phys. Rev. Lett., 87 (2001), p. 250602.
- [378] C. TUDOR, *Wong-Zakai type approximations for stochastic differential equations driven by a fractional Brownian motion*, Z. Anal. Anwend., 28 (2009), pp. 165–182.
- [379] C. TUDOR AND M. TUDOR, *On approximation of solutions for stochastic delay equations*, Stud. Cerc. Mat., 39 (1987), pp. 265–274.
- [380] M. TUDOR, *Approximation schemes for stochastic equations with hereditary argument*, Stud. Cerc. Mat., 44 (1992), pp. 73–85.
- [381] K. TWARDOWSKA, *On the approximation theorem of the Wong-Zakai type for the functional stochastic differential equations*, Probab. Math. Statist., 12 (1991), pp. 319–334 (1992).

- [382] ———, *An extension of the Wong-Zakai theorem for stochastic evolution equations in Hilbert spaces*, Stochastic Anal. Appl., 10 (1992), pp. 471–500.
- [383] ———, *An approximation theorem of Wong-Zakai type for nonlinear stochastic partial differential equations*, Stochastic Anal. Appl., 13 (1995), pp. 601–626.
- [384] ———, *An approximation theorem of Wong-Zakai type for stochastic Navier-Stokes equations*, Rend. Sem. Mat. Univ. Padova, 96 (1996), pp. 15–36.
- [385] ———, *Wong-Zakai approximations for stochastic differential equations*, Acta Appl. Math., 43 (1996), pp. 317–359.
- [386] K. TWARDOWSKA, T. MARNIK, AND M. PASŁAWASKA-POLUNIAK, *Approximation of the Zakai equation in a nonlinear filtering problem with delay*, Int. J. Appl. Math. Comput. Sci., 13 (2003), pp. 151–160.
- [387] K. VASILAKOS AND A. BEUTER, *Effects of noise on a delayed visual feedback system*, J. Theor. Biol., 165 (1993), pp. 389 – 407.
- [388] D. VENTURI, X. WAN, R. MIKULEVICIUS, B. L. ROZOVSKII, AND G. E. KARNIADAKIS, *Wick-Malliavin approximation to nonlinear stochastic partial differential equations: analysis and simulations*, Proc. Roy. Soc. Edinburgh Sect. A., 469 (2013).
- [389] J. B. WALSH, *A stochastic model of neural response*, Adv. in Appl. Probab., 13 (1981), pp. 231–281.
- [390] J. B. WALSH, *Finite element methods for parabolic stochastic PDE's*, Potential Anal., 23 (2005), pp. 1–43.
- [391] J. B. WALSH, *On numerical solutions of the stochastic wave equation*, Illinois J. Math., 50 (2006), pp. 991–1018 (electronic).

- [392] X. WAN AND G. KARNIADAKIS, *Beyond Wiener-Askey expansions: Handling arbitrary pdfs*, J. Sci. Comput., 27 (2006), pp. 455–464.
- [393] X. WAN, B. ROZOVSKII, AND G. E. KARNIADAKIS, *A stochastic modeling methodology based on weighted Wiener chaos and Malliavin calculus*, Proc. Natl. Acad. Sci. USA, 106 (2009), pp. 14189–14194.
- [394] X. WANG AND S. GAN, *B-convergence of split-step one-leg theta methods for stochastic differential equations*, J. Appl. Math. Comput., 38 (2012), pp. 489–503.
- [395] ———, *A Runge-Kutta type scheme for nonlinear stochastic partial differential equations with multiplicative trace class noise*, Numer. Algorithms, 62 (2013), pp. 193–223.
- [396] ———, *Weak convergence analysis of the linear implicit Euler method for semilinear stochastic partial differential equations with additive noise*, J. Math. Anal. Appl., 398 (2013), pp. 151–169.
- [397] X. WANG, S. GAN, AND J. TANG, *Higher order strong approximations of semilinear stochastic wave equation with additive space-time white noise*, ArXiv e-prints, (2013).
- [398] G. W. WASILKOWSKI AND H. WOŹNIAKOWSKI, *Explicit cost bounds of algorithms for multivariate tensor product problems*, J. Complexity, 11 (1995), pp. 1–56.
- [399] E. WONG AND M. ZAKAI, *On the convergence of ordinary integrals to stochastic integrals*, Ann. Math. Statist., 36 (1965), pp. 1560–1564.
- [400] ———, *On the relation between ordinary and stochastic differential equations*, Internat. J. Engrg. Sci., 3 (1965), pp. 213–229.
- [401] F. WU, X. MAO, AND L. SZPRUCH, *Almost sure exponential stability of numerical solutions for stochastic delay differential equations*, Numer. Math., 115 (2010), pp. 681–697.

- [402] D. XIU, *Fast numerical methods for stochastic computations: a review*, Commun. Comput. Phys., 5 (2009), pp. 242–272.
- [403] D. XIU AND J. HESTHAVEN, *High-order collocation methods for differential equations with random inputs*, SIAM J. Sci. Comput., 27 (2005), pp. 1118–1139.
- [404] D. XIU AND G. KARNIADAKIS, *Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 4927–4948.
- [405] D. B. XIU AND G. E. KARNIADAKIS, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644.
- [406] J. XU AND J. LI, *Sparse Wiener chaos approximations of Zakai equation for nonlinear filtering*, in Proceedings of the 21st annual international conference on Chinese Control and Decision Conference, CCDC'09, Piscataway, NJ, USA, 2009, IEEE Press, pp. 910–913.
- [407] Y. YAN, *Semidiscrete Galerkin approximation for a linear stochastic parabolic partial differential equation driven by an additive noise*, BIT, 44 (2004), pp. 829–847.
- [408] ———, *Galerkin finite element methods for stochastic parabolic partial differential equations*, SIAM J. Numer. Anal., 43 (2005), pp. 1363–1384.
- [409] X. YANG, Y. DUAN, AND Y. GUO, *A posteriori error estimates for finite element approximation of unsteady incompressible stochastic Navier-Stokes equations*, SIAM J. Numer. Anal., 48 (2010), pp. 1579–1600.
- [410] X. YANG, W. WANG, AND Y. DUAN, *The approximation of a Crank-Nicolson scheme for the stochastic Navier-Stokes equations*, J. Comput. Appl. Math., 225 (2009), pp. 31–43.

- [411] R.-M. YAO AND L.-J. BO, *Discontinuous Galerkin method for elliptic stochastic partial differential equations on two and three dimensional spaces*, Sci. China Ser. A, 50 (2007), pp. 1661–1672.
- [412] H. YOO, *Semi-discretization of stochastic partial differential equations on  $\mathbf{R}^1$  by a finite-difference method*, Math. Comp., 69 (2000), pp. 653–666.
- [413] N. YOSHIDA, *Stochastic shear thickening fluids: strong convergence of the Galerkin approximation and the energy equality*, Ann. Appl. Probab., 22 (2012), pp. 1215–1242.
- [414] M. ZAHRI, M. SEPID, H. MANOUZI, AND M. EL-AMRANI, *Wiener-Itô chaos expansions and finite-volume solution of stochastic advection equations*, in Finite volumes for complex applications IV, ISTE, London, 2005, pp. 525–538.
- [415] M. ZAKAI, *On the optimal filtering of diffusion processes*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 11 (1969), pp. 230–243.
- [416] G. ZHANG AND M. GUNZBURGER, *Error analysis of a stochastic collocation method for parabolic partial differential equations with random input data*, SIAM J. Numer. Anal., 50 (2012), pp. 1922–1940.
- [417] H. ZHANG, S. GAN, AND L. HU, *The split-step backward Euler method for linear stochastic delay differential equations*, J. Comput. Appl. Math., 225 (2009), pp. 558–568.
- [418] L. ZHANG AND Q. M. ZHANG, *Convergence of numerical solutions for the stochastic Navier-Stokes equation*, Math. Appl. (Wuhan), 21 (2008), pp. 504–509.
- [419] Z. ZHANG, M. CHOI, AND G. E. KARNIADAKIS, *Error estimates for the ANOVA method with polynomial chaos interpolation: tensor product functions*, SIAM J. Sci. Comput., 34 (2012), pp. A1165–A1186.

- [420] Z. ZHANG, B. ROZOVSKII, M. V. TRETYAKOV, AND G. E. KARNIADAKIS, *A multistage Wiener chaos expansion method for stochastic advection-diffusion-reaction equations*, SIAM J. Sci. Comput., 34 (2012), pp. A914–A936.
- [421] Z. ZHANG, M. V. TRETYAKOV, B. ROZOVSKII, AND G. E. KARNIADAKIS, *A recursive stochastic collocation method for stochastic differential equations with white noise*, Submitted, (2013).
- [422] Z. ZHANG, X. YANG, G. LIN, AND G. E. KARNIADAKIS, *Numerical solution of the Stratonovich- and Ito-Euler equations: Application to the stochastic piston problem*, J. Comput. Phys., 236 (2013), pp. 15–27.
- [423] B. ŽIBAITIS AND V. MATSKYAVICHYUS, *Gaussian approximations of Brownian motion in a stochastic integral*, Liet. Mat. Rink., 33 (1993), pp. 508–526.
- [424] B. ŽIBAITIS, *Mollifier approximation of Brownian motion in stochastic integral*, Litovsk. Mat. Sb., 30 (1990), pp. 717–727.