Analysis of Various Estimators for Multi-dimensional Zakai Equations



Zhenru Wang

Somerville College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2019

To my Parents:

Shili Wang & *Xiaohua Huang*

for their unending support, encouragement and love.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Christoph Reisinger, for his superb guidance, patience, and ceaseless encouragement. His deep knowledge in maths, so many years of experience, passion for research, and nice personality help and influence me to move forward and become who I am now.

Moreover, I would like to thank Prof. Mike Giles and Dr. Abdul-Lateef Haji-Ali, for the valuable feedback and insightful comments in my presentations in the MLMC reading group, Transfer, and Confirmation of status.

I also very much want to thank my friends, Wei Fang, Dunhong Jin, Jing Ren, Yixuan Wang, who make these years over DPhil study full of joy. I am so thankful for them sharing my happiness and low spirits, and discussing with me about my research problems. I cannot imagine what my life would be like without them.

I would like thank also to my colleagues Vadim Kaushansky, Andrei Cozma, Matthieu Mariapragassam, Yufei Zhang, for their relaxing coffee chats.

Finally, I am infinitely grateful to my parents and my aunt, who believe in my DPhil study from the very beginning till the end, and support me financially and mentally so much. Though separated between 8-hour time zone, the love and bond from family is only stronger. I know more and more what family means, and I can never thank them repay them enough.

Statement of Originality

I hereby declare that this thesis contains no material which has been accepted or is currently being submitted for any other degree, diploma, certificate or other qualifications at the University of Oxford or elsewhere. To the best of my knowledge, this thesis contains no material previously published and precise reference is made when a previously published result is used or discussed.

This thesis includes two papers published, the first one in the Journal of Computational Mathematics [94] (Chapter 2); the second one accepted in the BIT Numerical Mathematics [93] (Chapter 3), both of which are co-authored with my supervisor Prof. Christoph Reisinger.

The remainder of the thesis contains two additional unpublished papers, also co-authored with my supervisor Prof. Christoph Reisinger.

The ideas, development and writing up of the thesis, and the papers included in it, were the principal responsibility of myself, Zhenru Wang, working under the supervision of Prof. Christoph Reisinger.

Abstract

We first consider a one-dimensional stochastic partial differential equation (SPDE) of Zakai type describing a large credit portfolio. Specifically, we construct estimators of linear functionals of the solution from an implicit Milstein scheme on a space-time mesh. We compare the complexity of a multi-index Monte Carlo (MIMC) approach with the multilevel Monte Carlo (MLMC) method of Giles and Reisinger (2012), and find, by means of Fourier analysis, that the MIMC method has slightly improved complexity $O(\varepsilon^{-2}|\log\varepsilon|)$ for a root mean square error (RMSE) ε if a carefully adapted discretisation is used.

Then, we propose an implicit finite difference scheme for a two-dimensional parabolic SPDE of Zakai type, based on a Milstein approximation to the stochastic integral and an alternating direction implicit (ADI) discretisation of the elliptic term. We prove its mean-square stability and convergence in L_2 of first order in time and second order in space, by Fourier analysis, in the presence of Dirac initial data.

Next, we analyse the accuracy and computational complexity of estimators for linear functionals of the solution to the 2-d SPDE, coupled with the sparse combination technique and MLMC. We find, by detailed Fourier analysis, that for a RMSE ε, MLMC with sparse combination has the optimal complexity $O(\varepsilon^{-2})$, whereas MLMC on regular grids has $O(\varepsilon^{-2}(\log\varepsilon)^2)$, standard MC with sparse combination $O(\varepsilon^{-7/2}(|\log\varepsilon|)^{5/2})$, and MC on regular grids $O(\varepsilon^{-4})$. We give a discussion of the higher-dimensional setting without detailed proofs, which suggests that MLMC with sparse combination always leads to the optimal complexity.

Finally, we consider a particular two-dimensional SPDE with fast mean-reverting volatility on a timescale $O(\epsilon^{-1})$, and study the small ϵ asymptotics. We find an asymptotic expansion of the solution to the SPDE as $\epsilon \to 0$ and conclude from numerical experiments the convergence order $1/2$ of the leading term and order 1 after inclusion of the first correction term.

Contents

1 Introduction **1**

1.1 Zakai equations . . . 1

1.1.1 A motivating example: large portfolio credit model . . . 1

1.1.2 General form of Zakai equations . . . 3

1.2 Literature review on numerical approximations . . . 4

1.3 Relevant background results . . . 7

1.3.1 Multilevel Monte Carlo . . . 7

1.3.2 Sparse combination technique . . . 9

1.3.3 Multi-index Monte Carlo . . . 11

1.3.4 Asymptotics for PDEs . . . 13

1.4 Outline and contribution . . . 14

2 A one-dimensional Zakai SPDE **21**

2.1 Two implicit Milstein finite difference schemes . . . 22

2.2 Multi-index Monte Carlo discretisation on a space-time mesh . . . 27

2.3 Fourier analysis of MIMC estimators . . . 29

2.3.1 Main results . . . 29

2.3.2 Fourier transform of the solution . . . 31

2.3.3 Fourier analysis of MIMC estimators (proof of Theorem 2.3.1) . . . 33

2.3.4 Low wave number region (proof of Lemma 2.3.1) . . . 35

2.3.5 High wave number region (proof of Lemma 2.3.2) . . . 40

2.4 Optimal index set and complexity . . . 41

2.5 Some sub-optimal approximations . . . 44

2.5.1 The alternative discretisation (2.6) . . . 44

2.5.2 An alternative functional approximation . . . 46

2.6 Numerical implementation . . . 48

2.6.1 Mean and variance of $\Delta P_{(l_1,l_2)}$ . . . 49

2.6.2 Complexity . . . 53

2.7 Summary . . . 54

3 A two-dimensional Zakai SPDE **55**
3.1 Approximation and main results . . . 55
3.1.1 Semi-implicit Milstein finite difference scheme . . . 55
3.1.2 Main convergence results . . . 59
3.2 Fourier analysis of mean-square stability . . . 61
3.3 Fourier analysis of L_2-convergence . . . 68
3.3.1 Low wave number region (proof of Lemma 3.3.1) . . . 70
3.3.2 High wave number region (proof of Lemma 3.3.2) . . . 74
3.3.3 Convergence of the ADI scheme (proof of Theorem 3.1.2) . . . 80
3.4 Numerical tests . . . 81
3.5 Scheme and test for a variable coefficient SPDE . . . 85
3.6 Summary . . . 92

4 Sparse combination with multilevel estimators **93**
4.1 Approximation and main results . . . 94
4.1.1 Semi-implicit Milstein finite difference scheme . . . 94
4.1.2 Sparse combination estimators . . . 96
4.1.3 Sparse combination MLMC estimators . . . 98
4.2 Fourier analysis of the sparse combination error expansion . . . 99
4.2.1 Fourier transform of the solution . . . 100
4.2.2 Fourier transform of the sparse combination estimator . . . 101
4.2.3 Proof of Theorem 4.1.2 . . . 103
4.2.4 Proof of Theorem 4.1.3 . . . 111
4.3 Numerical tests . . . 111
4.3.1 Mean and variance of hierarchical increments . . . 111
4.3.2 Mean and variance of sparse grid increments . . . 112
4.4 Generalisation to higher dimensions . . . 116
4.5 Summary . . . 119

5 Asymptotic expansion of a two-dimensional Zakai SPDE **120**
5.1 SPDE asymptotic expansion . . . 121
5.1.1 Zero order term . . . 121
5.1.2 Correction terms . . . 124
5.1.3 Approximation to the marginal density of X . . . 125
5.2 Adaptation of numerical schemes . . . 126
5.3 Numerical results . . . 129
5.3.1 Convergence of $\mathbb{E}[P_T(v) - P_T(v_{0,\epsilon})]$ . . . 130
5.3.2 Convergence of $\mathbb{E}[P_T(v_{1,\epsilon})]$ . . . 131

5.3.3 Convergence of $\mathbb{E}[P_T(v) - P_T(v_{0,\epsilon}) - P_T(v_{1,\epsilon})]$ 135
5.4 Summary 138

6 Conclusions and future work **139**
6.1 Conclusions 139
6.2 Future work 141

A Proof of Proposition 2.1.1 **142**

B Particle system and the limiting SPDE **148**

Bibliography **151**

List of Figures

2.1 Optimal index set. 43
2.2 Optimal index set, with schemes (2.6) and (2.63). 48
2.3 Numerical approximation to the SPDE (2.1) and the pointwise error to the analytic solution (2.3). 48
2.4 θ vs. ρ 49
2.5 Rate verification: contour plots of log values of sample mean (left) and variance (right) of mixed differences used in MIMC. 51
2.6 Log values of mean and variance for MLMC and MIMC with $l_1 = 0$ or $l_2 = 0$. 51
2.7 Log values of mean and variance by the other discretisation scheme, given by (2.6) and (2.11). 52
2.8 Contour plots of profit. 52
2.9 Total cost versus α, and ε^2Cost with different schemes, where "alternative" denotes the scheme (2.6) and (2.11). 53

3.1 Numerical approximation to the SPDE (3.1) with the Milstein-ADI approximation (3.4) for one Brownian path, and the pointwise error to the analytic solution (1.15). 81
3.2 Verifies the L_2-convergence order in h and k from (3.11), with coarsest level $h = 1,\ k = 1/4$ and finest level $h = 2^{-5},\ k = 2^{-10}$, using 100 Monte Carlo samples. 82
3.3 L_2 error in space as function of correlation parameters for fixed $h_x = h_y = 2^{-3}$ and $k = 2^{-9}$, for a fixed path. The dark blue areas correspond to the stability region from Assumption 3.1.1. 83
3.4 Unstable solution for the Milstein-ADI scheme (3.4) for one Brownian path, $k = 2^{-2}$ and $h_x = h_y = 2^{-9}$, $\rho_x = \rho_y = 0.6,\ \rho_{xy} = 0.1$. 84
3.5 Comparing the L_2 error in space with fixed k and letting $h \to 0$ under various senarios: $k = 2^{-2}$; $k = 2^{-4}$; $k = 2^{-2}$ and $\rho_x = \rho_y = 0$; $k = 2^{-2}$ and smooth initial condition (3.32), all with a same Brownian path. 84
3.6 Single path realisation of the density to the SPDE (3.33) and associated numerical cost (CPU time in sec). 88

3.7 Test the convergence order in (h_x, h_y) and k for the schemes (3.34) and (3.36), in L_2 in space and probability, using 100 paths. 90
3.8 Single path realisation of the solution at (x_0, y_0) over time, of SPDE (3.33) with variable coefficients and (3.35) with frozen coefficients. 91
3.9 Numerical error at (x_0, y_0) by taking difference between (h_x, h_y, k) and $(h_x/2, h_y/2, k/4)$ to the schemes (3.34) and (3.36), where $h_x = \frac{5}{32},\ h_y = \frac{1}{160},\ k = \frac{1}{1024}$. 91

4.1 Contour plots of log values of sample mean and variance of $\Delta P^N_{(l_1\, l_2)}$ for $l_1, l_2 > 0$ used in the sparse combination method. 113
4.2 Log values of mean, variance and cost for δP_l. 114
4.3 Comparison of total cost among different schemes. 115

5.1 Convergence of $\Delta \widehat{P}_0$. 131
5.2 $\widehat{P}_1$ defined in (5.34) for $h_x = h_y = 2^{-l},\ k = 0.5 \cdot 4^{-l},\ \epsilon = 0.1 \times 2^{-3}, \cdots, 0.1 \times 2^{-7}$. 132
5.3 Comparing $\widehat{P}_1$ with $\widehat{P}^x_1$ from (5.34). 132
5.4 Comparing $\widehat{P}^x_1$ with regard to ν and ρ 134
5.5 Convergence of $\Delta \widehat{P}_0 - \widehat{P}_1$. 136

A.1 Integral over the high wave region with k fixed. 147

Chapter 1

Introduction

The topic of the thesis is the construction and analysis of finite difference approximations to the solutions of stochastic partial differential equations (SPDE) of Zakai type, and of optimal complexity estimators of associated expected functionals.

In this chapter, we first give an overview of the general form of Zakai equations, and review related research. Then, we introduce the multilevel Monte Carlo (MLMC), sparse combination method, and multi-index Monte Carlo (MIMC), which form the basis of the approach in this thesis. Finally, we give a summary of the whole thesis and highlight the main contributions to the field.

1.1 Zakai equations

1.1.1 A motivating example: large portfolio credit model

The growing of the credit derivatives market and the development of increasingly complex credit instruments require mathematical models for pricing and risk management. One popular class of credit models are structural models, based on the premise that a default is triggered when a company's asset value falls below a certain threshold barrier.

We start by introducing a structural credit model for a large portfolio of credit risky assets [14], which extends the standard models by introducing dynamics and working with

the infinite dimensional limit. In the model, each firm is assigned a distance to default process, combined with a market factor shared by all the firms, and its own idiosyncratic factor. Then, by the law of large numbers, we can approximate the proportion of the firms which have defaulted, and this can be used for credit derivatives pricing.

In [14], it is assumed that the market consists of N different firms whose value at time t satisfies the stopped diffusion process

$$\begin{aligned}
\frac{\mathrm{d}V_t^{i,N}}{V_t^{i,N}} &= r\,\mathrm{d}t + \sigma\sqrt{1-\rho}\,\mathrm{d}W_t^i + \sigma\sqrt{\rho}\,\mathrm{d}W_t, \qquad t < T_0^{i,N} \\
V_t^{i,N} &= B^i\,, \quad t \geqslant T_0^{i,N} \\
V_0^{i,N} &= b^i > B^i,
\end{aligned}$$

for $i = 1, \ldots, N$, where B^i is the default barrier of the i-th firm, $T_0^{i,N} = \inf\{t : V_t^{i,N} = B^i\}$ is the stopping time when the firm defaults, and W^i and W are independent standard Brownian motions.

Then we define the "distance-to-default" processes $X_t^i := (\log V_t^i - \log B^i)/\sigma$, which satisfies the following dynamics

$$\begin{aligned}
\mathrm{d}X_t^{i,N} &= \mu\,\mathrm{d}t + \sqrt{1-\rho}\,\mathrm{d}W_t^i + \sqrt{\rho}\,\mathrm{d}W_t, \qquad t < T_0^{i,N} \\
X_t^{i,N} &= 0, \qquad t \geqslant T_0^{i,N} \\
X_0^{i,N} &= x^{i,N} > 0,
\end{aligned}$$

where $\mu = (r - \frac{1}{2}\sigma^2)/\sigma$.

We study the sequence of equally weighted empirical measures for the entire portfolio:

$$\bar{\nu}_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}\,.$$

We can write this measure as

$$\bar{\nu}_t^N = P_t^N\delta_0 + \nu_t^N\,,$$

where

$$\nu_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}\mathbb{1}_{\{t<T_0^{i,N}\}}\,, \qquad P_t^N = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{t\geq T_0^{i,N}\}}\,.$$

As $N \to \infty$, the limit empirical measure $\bar{\nu}_t = \lim_{N\to\infty} \nu_t^N$ exists and is a probability measure which can be written as $\bar{\nu}_t = P_t\delta_0 + \nu_t$ (see [14]).

The measure ν_t is supported on $(0,\infty)$ with density $v(t,x)$, which is the unique solution of the SPDE (see [14])

$$\begin{aligned} &\mathrm{d}v = -\mu v_x\,\mathrm{d}t + \frac{1}{2}\,v_{xx}\,\mathrm{d}t - \sqrt{\rho}\,v_x\,\mathrm{d}W_t, \\ &v(0,x) = v_0(x), \\ &v(t,0) = 0. \end{aligned}$$

Then the functional of interest is the loss which represents the mass lost at the absorbing boundary,

$$P_t = 1 - \int_0^\infty v(t,x)\,\mathrm{d}x. \tag{1.1}$$

In the credit risk application of [10], P_t describes the loss in a structural credit model, i.e., the fraction of firms whose values have crossed zero and which are considered defaulted. The values of credit products are often functions of the loss P_t at time t.

If we simplify the setting by considering the firm value without the barrier condition, then we arrive at the following 1-d SPDE,

$$\mathrm{d}v = -\mu\frac{\partial v}{\partial x}\,\mathrm{d}t + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}\,\mathrm{d}t - \sqrt{\rho}\frac{\partial v}{\partial x}\,\mathrm{d}W_t, \qquad (t,x)\in(0,T)\times\mathbb{R}, \tag{1.2}$$

where $T>0$, W is a standard Brownian motion, and μ and $0\le\rho<1$ are real-valued parameters.

1.1.2 General form of Zakai equations

The analysis and numerical computation of Zakai equations and other types of SPDEs have been extensively studied in recent years. A general form of Zakai equation (see [2, 45]) is given by

$$\begin{aligned} \mathrm{d}v(t,x) = &\left(\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i\partial x_j}\big[a_{ij}(x)v(t,x)\big] - \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\big[b_i(x)v(t,x)\big]\right)\mathrm{d}t \\ &- \sum_{l=1}^{m}\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\gamma_{i,l}(x)v(t,x))\,\mathrm{d}W_t^k, \end{aligned} \tag{1.3}$$

where $W_t=(W_t^1,\dots,W_t^m)'$ is an m-dimensional standard Brownian motion, a is a $d\times d$ matrix-valued function, b is a $\mathbb{R}^d$-valued function, and $\gamma=(\gamma_{i,l}(x))$ is a $d\times m$ matrix-valued function. This Zakai equation arises from a nonlinear filtering problem: given an

m-dimensional observation process W and a d-dimensional signal process Z, the goal is to estimate the conditional distribution of Z given W. If Z satisfies

$$Z_t = Z_0 + \int_0^t \beta(Z_s)\,\mathrm{d}s + \int_0^t \varsigma(Z_s)\,\mathrm{d}B_s + \int_0^t \gamma(Z_s)\,\mathrm{d}W_s, \tag{1.4}$$

where B is a d-dimensional standard Brownian motion independent of W, ς is a $d \times d$-matrix valued function, and β is a $\mathbb{R}^d$-valued function, then the conditional distribution function of Z given W has a density v in L_2, and it is proved (Theorem 3.1 in [77]) that if the following conditions are satisfied,

- there exists a constant K such that for each $x \in \mathbb{R}^d$,

$$|\varsigma(x)|^2 + |\beta(x)|^2 + |\gamma(x)|^2 \leq K^2;$$

- for each x_1, $x_2 \in \mathbb{R}^d$,

$$|\varsigma(x_1) - \varsigma(x_2)|^2 + |\beta(x_1) - \beta(x_2)|^2 + |\gamma(x_1) - \gamma(x_2)|^2 \leq K^2\big(|x_1 - x_2|^2\big);$$

-

$$\mathbb{E}[Z_0^2] < \infty,$$

then v satisfies (1.3) in a weak sense with

$$a = \varsigma\varsigma^\top + \gamma\gamma^\top, \qquad b = \beta. \tag{1.5}$$

Moreover (see also [77]), the solution v to (1.3) can be interpreted as the density — if it exists — of the limit empirical measure $\nu_t = \lim_{N\to\infty} N^{-1} \sum_{i=1}^N \delta_{Z_t^i}$ for

$$Z_t^i = Z_0 + \int_0^t \beta(Z_s^i)\,\mathrm{d}s + \int_0^t \varsigma(Z_s^i)\,\mathrm{d}B_s^i + \int_0^t \gamma(Z_s^i)\,\mathrm{d}W_s, \tag{1.6}$$

where B^i are independent Brownian motions, independent of W, and the rest as above. In fact, these results are derived in [77] for a more general class of SPDEs, which includes Zakai equations, as well as McKean–Vlasov equations.

1.2 Literature review on numerical approximations

There are several approaches to the numerical approximation of Zakai equations.

Florchinger and Le Gland [30] provided an early discretisation scheme with respect to the space variables, using a particle approximation. The error analysis was an extended result of Raviart [90] on first order deterministic PDE. Hereby, they represented the solution in respect of stochastic characteristics, and approximated the initial condition using a convex linear combination of Dirac masses. Crisan and Lyons [22, 21] constructed a sequence of branching particle systems, the empirical measure of which converges to the solution of the Zakai equation. The major benefit of this Monte Carlo approximation to estimate the conditional distribution is that it is efficient for high dimensional problems. A similar approach and corresponding analysis also appears in [23, 20, 24, 45].

Florchinger and Le Gland [29] provided a time discretisation scheme for diffusion processes as a result of a split-up approximation and Trotter-type product formula. This splitting-up method is first introduced in [6], which decomposes the original SPDE into two parts: a second order deterministic PDE related to a prediction step, and a degenerate second-order stochastic PDE related to a correlation step. Gyöngy and Krylov further estimated the error with the Stratonovich differential [52, 51].

The other main approach is to directly solve the Zakai SPDE by spatial approximation methods and time stepping schemes, coupled again with Monte Carlo sampling, which is the subject of this paper. In terms of numerical analysis to solving SPDEs, there are many studies on the following parabolic SPDE

$$\begin{aligned} \mathrm{d}u(t,x) &= \left[\frac{\partial^2}{\partial x^2}u(t,x) + f(t,x,u(t,x))\right]\mathrm{d}t + \sigma(t,x,u(t,x))\,\mathrm{d}W_t, \\ u(t,0) &= u(t,1) = 0, \qquad t \geq 0, \\ u(0,x) &= u_0(x), \qquad x \in [0,1]. \end{aligned} \tag{1.7}$$

Here, u_0 is a continuous function on $[0,1]$ vanishing at 0 and 1. The coefficients f and σ are locally bounded Borel functions, and W is a standard Brownian motion. When f and σ satisfy Lipschitz and linear growth conditions, there exists a unique strong solution (see [36, 84, 98]).

Gyöngy and Nualart introduced an implicit scheme for the equation (1.7) in [53]. They proved almost sure convergence uniformly in time and space with nonlinear terms being Lipschitz, and convergence in probability provided the nonlinear drift is only measurable. In [49, 50], Gyöngy approximated the quasi-linear SPDE by a finite difference method with several time-stepping schemes, including explicit and implicit Euler schemes. Under smooth initial conditions, the convergence order is $1/2$ in space and $1/4$ in time. Davie and Gaines showed in [27] that this convergence rate is optimal, due to the lack of smoothness of the solution process. Walsh used finite element methods to approximate

(1.7), and the rate of convergence is the same as using finite difference methods (see [99]). Kruse analysed this type of SPDE with multiplicative noise using finite element methods in [73]. He derived strong orders of convergence both under a standard finite element semi-discretisation and a spatio-temporal discretisation.

To achieve a better convergence rate of the approximation to (1.7), it is natural to consider using stochastic Taylor schemes. When solving stochastic ordinary differential equations (SODEs), one could use an iterated application of Itô's formula as in [71], following an idea originally due to Milstein (see [86]). However, it does not apply to SPDEs with infinite-dimensional Brownian motions directly (see [46]). Jentzen and Kloeden overcame this problem in [66, 67] by using the exponential Euler scheme. Consider parabolic SPDEs of the form

$$\mathrm{d}U_t = [AU_t + F(U_t)]\,\mathrm{d}t + B\,\mathrm{d}W_t, \quad U_0 = u_0, \quad t \in [0,T] \tag{1.8}$$

in a separable Hilbert space H, where A is an unbounded linear operator, F is a nonlinear smooth function, B is a bounded linear operator, and W is an infinite-dimensional Brownian motion. Then, (1.8) can be interpreted in the mild sense (see [26]), i.e., as an integral equation of the form

$$U_t = \mathrm{e}^{At}u_0 + \int_0^t \mathrm{e}^{A(t-s)}F(U_s)\,\mathrm{d}s + \int_0^t \mathrm{e}^{A(t-s)}B\,\mathrm{d}W_s.$$

The exponential factor $\mathrm{e}^{A(t-s)}$ in every integral of the mild form acts as a mollifier such that the iterated version of the terms above remain in the Hilbert space H. By using the classical Taylor expansion for the nonlinear term F in the mild integral, one can replace the higher order terms recursively by lower order terms in the Taylor expansion. However, this scheme needs the nonlinear term F to be Fréchet differentiable with bounded derivatives. Later, Jentzen and Röckner applied the Milstein scheme to (1.7), where f and σ are smooth functions satisfying appropriate Lipschitiz assumptions (see details in [69]). The Milstein scheme requires $O(\varepsilon^{-2})$ computational cost for RMSE ε, whereas the linear implicit Euler scheme requires $O(\varepsilon^{-8/3})$.

The Lax equivalence theorem in this case has also been studied in [78]. Millet and Morien [85] derived strong convergence orders of explicit and implicit discretisation schemes of such parabolic SPDEs with space-time white noise. In [69, 3], the Milstein scheme is further applied to the class of SPDEs (1.8). Higher order convergence of schemes for nonlinear parabolic SPDEs with additive noise has been considered in [68].

In terms of Zakai SPDEs, Lang used the finite element method to analyse the almost sure convergence (see [80]). Barth and Lang further analysed L^p convergence with the Milstein finite element scheme [4]. Ito studied several time discretisations of stochastic evolution

equations including Zakai equations [64], and later he analysed, under Euler and Milstein time discretisation, Galerkin approximation in the spatial coordinates [65].

Under suitable regularity and some conditions on the operators A, F, B for equations of type (1.8), mean-square convergence of order $1/2$ is shown for an Euler semi-discretisation in [79] for square-integrable (not necessarily continuous), infinite-dimensional martingale drivers. In contrast, Barth and Lang in [4] allow only for continuous martingales but prove convergence of higher order in space and up to 1 in time, in L^p and almost surely, for a Milstein scheme and spatial Galerkin approximation of sufficiently high order; this is extended to advection-diffusion equations with possibly discontinuous martingales in [3].

As for the Zakai SPDE (1.2), Giles and Reisinger [42] used an explicit Milstein finite difference approximation to the solution of (1.2). By using Fourier analysis, this scheme can be shown to give first order of convergence in the timestep and second order in the spatial mesh size. One constraint in this paper is that the timestep needs to be small enough to ensure stability. Inspired by the numerical analysis of SDEs in [96, 9], Reisinger [92] extended the discretisation to an implicit method on the basis of the σ-θ time-stepping scheme, where the drift and the deterministic part of the double stochastic integral are taken implicit. Fourier analysis shows that the convergence order is the same as in the explicit Milstein scheme, however, this scheme is unconditionally mean-square stable under a constraint on the correlation ρ in (1.2).

1.3 Relevant background results

We briefly review now some methodology and results which this thesis builds on.

1.3.1 Multilevel Monte Carlo

The multilevel Monte Carlo (MLMC) method, established by Giles [38, 39], utilises a sequence of approximations $P_0, P_1, \dots, P_{l^*}$ to a random variable P with increasing accuracy but also higher cost for increasing l. In the simulation of an SDE, l would typically be the refinement level of the time mesh, with 2^l time steps. The MLMC estimator is based on recursive control variates embedded in the identity

$$\mathbb{E}\left[P_{l^*}\right] = \mathbb{E}\big[P_0\big] + \sum_{l=1}^{l^*} \mathbb{E}\big[P_l - P_{l-1}\big],$$

where l^* is a maximum refinement level. The goal is to estimate $\mathbb{E}[P_{l^*}]$ by independent estimation of the summands, in a way that the root mean square error (RMSE) is comparable to the bias, but with a much reduced computational complexity. If fewer samples are needed for higher levels, the total computational cost is much lower than when using the standard Monte Carlo method (see Theorem 3.1 in [38]). In a second step, l^* is adapted for a given RMSE target.

Theorem 1.3.1 (Theorem 3.1 in [38])**.** *Let P denote a functional of the solution of the SDE*

$$\mathrm{d}S_t = a(S_t, t)\,\mathrm{d}t + b(S_t, t)\,\mathrm{d}W_t$$

for a given Brownian path W, and let $\widehat{P}_l$ denote the corresponding approximation using a numerical discretisation with timestep $h_l = M^{-l}T$. If there exist independent estimators $\widehat{Y}_l$ based on N_l Monte Carlo samples, and positive constants $\alpha \geq \frac{1}{2}$, β, c_1, c_2, c_3, such that

(i) $\mathbb{E}[\widehat{P}_l - P] \leq c_1\, h_l^{\alpha}$,

(ii) $$\mathbb{E}[\widehat{Y}_l] = \begin{cases} \mathbb{E}[\widehat{P}_0], & l = 0, \\ \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}], & l > 0, \end{cases}$$

(iii) $\mathbb{V}[\widehat{Y}_l] \leq c_2\, N_l^{-1} h_l^{\beta}$,

(iv) C_l, the computational complexity of $\widehat{Y}_l$, is bounded by $C_l \leq c_3\, N_l h_l^{-1}$,

then there exists a positive constant c_4 such that for any $\varepsilon < \mathrm{e}^{-1}$, there are values l^* and N_l for which the multilevel estimator $\widehat{Y} = \sum_{l=0}^{l^*} \widehat{Y}_l$ has an MSE with bound

$$MSE \coloneqq \mathbb{E}\big[(\widehat{Y} - \mathbb{E}[P])^2\big] < \varepsilon,$$

with a computational complexity C with bound

$$C \leq \begin{cases} c_4 \varepsilon^{-2}, & \beta > 1, \\ c_4 \varepsilon^{-2} (\log \varepsilon)^2, & \beta = 1, \\ c_4 \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

The original MLMC work [38] studied SDEs satisfying the usual conditions (Theorem 10.2.2 in [70]) with Lipschitz payoff under Euler-Maruyama and Milstein schemes. Giles, Debrabant and Rößler [40] studied a digital option, which is a challenge in the sense that the difference between fine path and coarse path can be large. Giles and Szpruch [44] further developed an antithetic treatment which can achieve a small variance even with $O(h^{1/2})$ strong convergence. This approach has been applied to lookback and barrier options [43]. Some other related work includes the sensitivities [13], and the exit time [58]. MLMC also applies to nonlinear SDEs where the drift or the volatility can have super-linear growth. One of the solutions is to use adaptive time-stepping [61, 62, 28]. A comprehensive survey of MLMC research up to 2015 is found in [39].

The MLMC method has been extended to SPDEs, in which case it gives even better savings due to the additional spatial dimensions. MLMC has been applied to elliptic PDEs with random coefficients, see, e.g., [18, 16, 97]. Kuo, Schwab and Sloan applied Quasi-Monte Carlo [75] and Multilevel Quasi Monte Carlo methods [76] to finite element discretizations of elliptic partial differential equations with a random coefficient represented by a countably infinite number of terms. Croci et al. [25] applied MLMC to elliptic SPDEs combining it with finite element methods on non-nested meshes.

Closest to this thesis, Giles and Reisinger [42] applied MLMC to simulate the SPDE (1.2) and the functional of its solution given in (1.1). Both timestep k and mesh size h decrease geometrically on different levels of refinement $l = 0, 1, \ldots, l^*$, with fixed k/h^2. Thus, the variance of the estimators for $\mathbb{E}\big[P_l - P_{l-1}\big]$ is small for large l, compared to the estimator for $\mathbb{E}\big[P_{l^*}\big]$. As a result, for a fixed accuracy ε, the cost can be reduced significantly to $O(\varepsilon^{-2})$ instead of $O(\varepsilon^{-7/2})$ by the standard Monte Carlo estimator.

1.3.2 Sparse combination technique

Sparse grids [102] were first introduced to solve high dimensional PDEs on a tensor product space using the finite element method. The error bounds in [11] and the analysis

in [12] show that they reduce the curse of dimensionality in the numerical implementation of PDEs. In contrast to the finite element method, the combination technique, first introduced in [48], decomposes the solution into contributions from simple tensor product grids with different resolutions in each dimension. The analysis in [89] and [91] shows that the computational cost for the combination technique applied to the Poisson problem with sufficient regularity is independent of the dimension, up to a logarithmic term, for finite elements and finite differences, respectively. Hendricks, Ehrhardt and Günther combine the sparse grid combination technique with the ADI scheme for diffusion equations in [56].

Let P denote a functional of a solution to a PDE on the d-dimensional unit cube. Let $P_{\boldsymbol{l}}$ be an approximation of P, compared with mesh size $h_{\boldsymbol{l}} = (h_{1,l_1}, \ldots, h_{d,l_d}) = (2^{-l_1}, 2^{-l_2}, \ldots, 2^{-l_d})$. We define the first order difference operators $\Delta_i,\ 1 \leq i \leq d$, such that

$$\Delta_i P_{\boldsymbol{l}} = P_{\boldsymbol{l}} - P_{\boldsymbol{l}-\boldsymbol{e}_i} \mathbb{1}_{l_i > 0}, \tag{1.9}$$

where $\boldsymbol{e}_i$ is the unit vector in direction i. Then the d-dimensional mixed first difference is

$$\Delta P_{\boldsymbol{l}} = \left(\prod_{i=1}^{d} \Delta_i \right) P_{\boldsymbol{l}}.$$

It can be shown that

$$P_{\boldsymbol{l}} = \sum_{0 \leq \boldsymbol{i} \leq \boldsymbol{l}} \Delta P_{\boldsymbol{i}}.$$

Assume the pointwise expansion of the error on tensor product grids with mesh sizes $h = (h_1, \ldots, h_d)$ has the form

$$P - P_{\boldsymbol{l}} = \sum_{i=1}^{d} \sum_{\{j_1, \ldots, j_m\} \subset \{1, \ldots, d\}} c_{j_1, \ldots, j_m} h_{j_1}^p \cdot h_{j_2}^p \ldots \cdot h_{j_d}^p, \tag{1.10}$$

where

$$|c_{j_1, \ldots, j_m}| \leq K < \infty, \qquad \forall 1 \leq m \leq d, \quad \forall \{j_1, \ldots j_m\} \subset \{1, \ldots, d\}.$$

The sparse combination method uses the approximation

$$P_L \coloneqq \sum_{|\boldsymbol{l}|_1 \leq L} \Delta P_{\boldsymbol{l}},$$

where $|\boldsymbol{l}|_1 = l_1 + l_2 + \ldots + l_d$.

We state the theorem under the form in the survey paper [41], where the original proof is in [91].

Theorem 1.3.2 (Sparse combination method [41]). *If we approximate P using the sparse combination method with the index set specified by $\{\boldsymbol{l} \leq L\}$, then the computational cost is $O(L^{d-1}2^L)$. Moreover, if the underlying PDE approximation has second order accuracy and the solution u has sufficient mixed regularity* (1.10)*, then $\Delta P_{\boldsymbol{l}}$ has magnitude $O(2^{-2|\boldsymbol{l}|_1})$ so the error in P_L is $O(L^{d-1}2^{-2L})$. Hence the cost to achieve a functional accuracy of ε is $O(\varepsilon^{-1/2}|\log\varepsilon|^{3(d-1)/2})$.*

1.3.3 Multi-index Monte Carlo

Multi-index Monte Carlo (MIMC) can be viewed as a combination of sparse grids [102] and MLMC methods. The level $\boldsymbol{l}$ is now a vector of integer indices, corresponding to the refinement level in the different dimensions. Sparse grids, and hence MIMC, use high order mixed differences instead of only taking first order differences to construct a hierarchical approximation. More specifically, let P denote a functional applied to the solution of an underlying d-dimensional stochastic model. We define the first order difference operators $\Delta_i,\ 1 \leq i \leq d$, the same as (1.9),

$$\Delta_i P_{\boldsymbol{l}} = P_{\boldsymbol{l}} - P_{\boldsymbol{l}-\boldsymbol{e}_i}\mathbb{1}_{l_i>0},$$

where $\boldsymbol{e}_i$ is the unit vector in direction i and $P_{\boldsymbol{l}}$ is the estimator of level $\boldsymbol{l}$, with mesh size $h_{\boldsymbol{l}} = O(2^{-\boldsymbol{l}})$. Then the telescoping sum becomes

$$\mathbb{E}[P] = \sum_{\boldsymbol{l}\in\mathbb{N}^d} \mathbb{E}\left[\left(\prod_{i=1}^{d}\Delta_i\right)P_{\boldsymbol{l}}\right].$$

The strategy is similar to the sparse combination method, which is to neglect terms with large $\boldsymbol{l}$ where the cost is large and the contribution to the estimator is small. Instead of choosing an index l^* as in MLMC, one now has to choose an index set $\mathcal{I} \subset \mathbb{N}^d$ of terms to include.

Here, we include partial cases from Theorem 2.2 in [54], which are applied in this thesis. For the other situations, we refer to the original paper [54].

Theorem 1.3.3 (MIMC work estimate [54]). *We assume the expected value, the variance,*

and the average computational cost of $\Delta P_{\boldsymbol{l}}$ *satisfy*

$$|\mathbb{E}[\Delta P_{\boldsymbol{l}}]| \leq C_1 \prod_{i=1}^{d} \beta_i^{-l_i w_i}, \quad Var[\Delta P_{\boldsymbol{l}}] \leq C_2 \prod_{i=1}^{d} \beta_i^{-l_i s_i}, \quad Cost(\Delta P_{\boldsymbol{l}}) \leq C_3 \prod_{i=1}^{d} \beta_i^{-l_i \gamma_i},$$

for constants C_1, C_2, C_3, *and* $w_i > 0$, $0 < s_i \leq 2w_i$, $\gamma_i > 0$ *for* $i = 1, \ldots, d$.

For a fixed root mean square error (RMSE) ε, *let the approximation index set be* $\mathcal{I}_\delta(L) = \{\boldsymbol{l} \in \mathbb{N}^d : \boldsymbol{\delta} \cdot \boldsymbol{l} \leq L\}$ *for* $\boldsymbol{\delta} \in (0,1]^d$ *given by*

$$\delta_i = \frac{\log(\beta_i)(w_i + \frac{\gamma_i - s_i}{2})}{C_{\boldsymbol{\delta}}}, \qquad i = 1, \ldots, d,$$

where $C_{\boldsymbol{\delta}} = \sum_{j=1}^{d} \log(\beta_j)(w_j + \frac{\gamma_j - s_j}{2})$, *and* L *is a constant such that*

$$Bias(\mathcal{I}_\delta(L)) = \left| \sum_{\boldsymbol{l} \neq \mathcal{I}_\delta(L)} \mathbb{E}[\Delta P_{\boldsymbol{l}}] \right| \leq \sum_{\boldsymbol{l} \neq \mathcal{I}_\delta(L)} \left| \mathbb{E}[\Delta P_{\boldsymbol{l}}] \right| \leq (1 - \theta)\, \varepsilon,$$

where $\theta \in (0,1)$.

The MIMC estimator can be written as

$$\widehat{Y} = \sum_{\boldsymbol{l} \in \mathcal{I}_\delta(L)} \frac{1}{M_{\boldsymbol{l}}} \sum_{m=1}^{M_{\boldsymbol{l}}} \Delta P_{\boldsymbol{l}}^{(m)}.$$

If $\gamma_i < s_i$, $\forall\, i = 1, \ldots, d$, *and the variance of the MIMC estimator* $\widehat{Y}$ *satisfies*

$$Var[\widehat{Y}] = \sum_{\boldsymbol{l} \in \mathcal{I}} \frac{Var[\Delta P_{\boldsymbol{l}}]}{M_{\boldsymbol{l}}} \leq \theta^2 \varepsilon^2,$$

then there exists a constant $C_4 > 0$, *such that*

$$\textit{Total cost of } \widehat{Y} \leq C_4\, \varepsilon^{-2}.$$

The appeal of MIMC methods applied to higher-dimensional SPDEs, similar to sparse grid methods for high-dimensional PDEs, is that they can give a computational complexity with a fixed polynomial order, multiplied by a log term with an exponent which grows with the dimension d. Both the sparse grid combination technique and the MIMC framework are based on certain specific expansions of the discretisation error in the mesh parameters. These have been proven for a small but growing class of PDEs [89, 91, 47]; Haji-Ali, Nobile, and Tempone [54] study PDEs with random coefficients and refers to these PDE results in a path-wise sense, while in [39], MIMC has been applied to nested simulation.

1.3.4 Asymptotics for PDEs

Asymptotic expansions of PDE solutions in the presence of small model parameters have a long tradition in applied mathematics, see, e.g., [60]. The application of such expansions in derivative pricing in the presence of a fast evolving stochastic volatility process is first proposed in Fouque et al. [32]. More specifically, they present a separation of time scales between the stock process and the volatility driving process, such that the rate of mean reversion in the volatility process is proportional to $O(1/\epsilon)$. We will later apply a related approach to an SPDE with fast volatility process.

The stochastic volatility model in [32] is as follows,

$$\begin{aligned} \mathrm{d}X_t &= rX_t\,\mathrm{d}t + f(Y_t)X_t\,\mathrm{d}W_t, \\ \mathrm{d}Y_t &= \Big[\frac{1}{\epsilon}(m - Y_t) - \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\Lambda(Y_t)\Big]\,\mathrm{d}t + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\,\mathrm{d}Z_t, \end{aligned}$$

where W and Z are two standard Brownian motions with correlation ρ, $f(Y_t)$ is the stochastic volatility for a given positive function f, and the function Λ is given by

$$\Lambda(y) = \rho\frac{\mu - r}{f(y)} + \gamma(y)\sqrt{1-\rho^2},$$

assuming that the market price of volatility risk γ is a bounded function of y alone.

Consider a European derivative with payoff $h(X_T)$ at maturity time T for a given function h, then the price at time $t < T$ is a function of the present value of the stock price $X_t = x$ and the volatility $Y_t = y$. By Feynman-Kac formula, the price satisfies the following PDE,

$$\begin{aligned} \frac{\partial P}{\partial t} + \frac{1}{2}f(y)^2x^2\frac{\partial^2 P}{\partial x^2} + \frac{\rho\nu\sqrt{2}}{\sqrt{\epsilon}}xf(y)\frac{\partial^2 P}{\partial x\partial y} + \frac{\nu^2}{\epsilon}\frac{\partial^2 P}{\partial y^2} \\ + r\Big(x\frac{\partial P}{\partial x} - P\Big) + \Big[\frac{1}{\epsilon}(m-y) - \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\Lambda(y)\Big]\frac{\partial P}{\partial y} = 0, \\ P(T,x,y) = h(x). \end{aligned} \tag{1.11}$$

Gathering the terms of order $1/\epsilon$, $1/\sqrt{\epsilon}$, and 1, we can rewrite the PDE (1.11) as

$$\left(\frac{1}{\epsilon}\mathcal{L}_0 + \frac{1}{\sqrt{\epsilon}}\mathcal{L}_1 + \mathcal{L}_2\right)P = 0,$$

with

$$\begin{aligned} \mathcal{L}_0 &= \nu^2\frac{\partial^2}{\partial y^2} + (m-y)\frac{\partial}{\partial y}, \\ \mathcal{L}_1 &= \sqrt{2}\rho\nu x f(y)\frac{\partial^2}{\partial x\partial y} - \sqrt{2}\nu\Lambda(y)\frac{\partial}{\partial y}, \\ \mathcal{L}_2 &= \frac{\partial}{\partial t} + \frac{1}{2}f(y)^2x^2\frac{\partial^2}{\partial x^2} + r\Big(x\frac{\partial}{\partial x} - \cdot\Big). \end{aligned}$$

Then, expanding the option price in powers of $\sqrt{\epsilon}$ yields

$$P = P_0 + \sqrt{\epsilon}P_1 + \epsilon P_2 + \epsilon\sqrt{\epsilon}P_3 + \cdots, \tag{1.12}$$

where P_0, $P_1, \ldots$ are functions of (t, x, y) with $P_0(T, x, y) = h(x)$, and can be determined by substituting (1.12) into (1.11), formally collection terms of the same power of ϵ, and solving the resulting differential equations. Thus, by performing a singular perturbation analysis of the pricing PDE, it follows that the Black-Scholes price with a constant volatility is the zero order term.

Later Fouque et al. also studied the effect of the non-smoothness of the payoff function in option pricing in [34], and further introduce a slowly varying factor in the model for the stochastic volatility to combine regular and singular perturbations in [33]. Some applications of the multiscale stochastic volatility model are given in [35, 31].

1.4 Outline and contribution

We start by studying the one dimensional Zakai SPDE (1.2) in Chapter 2. Giles and Reisinger [42] used an explicit Milstein finite difference approximation to the solution of (1.2), and apply multilevel Monte Carlo method to estimate $\mathbb{E}[P_t]$, the expected value of the loss functional, where P_t is defined in (1.1). Later, Reisinger extended the discretisation to an implicit method in [92].

Instead of using the MLMC method, we apply MIMC to the SPDE (1.2) on a space-time mesh, with an implicit Milstein time-stepping scheme and central spatial differences. We demonstrate that the expectation and variance of the multi-index estimators can be analysed by Fourier analysis, even if the required error expansions do not hold pathwise. We then show that an extra term appears in the error expansion which explodes for small spatial mesh sizes, but by a suitable adaptation of the MIMC estimator, a theoretical complexity of $O(\varepsilon^{-2}|\log\varepsilon|)$ can be obtained, while the practically observed complexity is still $O(\varepsilon^{-2})$ as in the MLMC method [42]. We give a seemingly innocuous variant of this approximation scheme, i.e., with the spatial approximation studied in [42], and show that the "standard" assumptions on the error expansion are only satisfied with lower order, such that the MIMC estimator is less efficient with complexity $O(\varepsilon^{-2}|\log\varepsilon|^3)$. Finally, we give an example of a discontinuous functional and a simple approximation where we show that multiple leading order error terms appear, such that the optimal index set is not triangular, and demonstrate how MIMC can be adapted to this case.

Hence, we show that although MIMC is expected to perform favourably in multi-dimensional situations compared to MLMC, this is not always true. In Chapter 2, we only consider cases where MLMC gives optimal complexity order. Therefore, the results herein can be considered negative in the practical sense that worse performance is obtained for MIMC than MLMC. In addition to these "negative" results being interesting in their own right, we provide a methodology to analyse MIMC estimators for a class of SPDEs. This method carries over to the higher-dimensional setting, where MIMC has complexity advantages over MLMC (as explained further in Section 2.7).

In Chapter 3, we consider the two-dimensional SPDE

$$\begin{aligned} \mathrm{d}v = &-\mu_x \frac{\partial v}{\partial x}\,\mathrm{d}t - \mu_y \frac{\partial v}{\partial y}\,\mathrm{d}t + \frac{1}{2}\left(\frac{\partial^2 v}{\partial x^2} + 2\sqrt{\rho_x \rho_y}\rho_{xy}\frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 v}{\partial y^2}\right)\mathrm{d}t \\ &- \sqrt{\rho_x}\frac{\partial v}{\partial x}\,\mathrm{d}W_t^x - \sqrt{\rho_y}\frac{\partial v}{\partial y}\,\mathrm{d}W_t^y, \end{aligned} \tag{1.13}$$

for $x, y \in \mathbb{R},\ 0 < t \leq T$, where μ_x, μ_y and $0 \leq \rho_x, \rho_y < 1$, $-1 \leq \rho_{xy} \leq 1$ are real-valued parameters, subject to the Dirac initial data

$$v(0, x, y) = \delta(x - x_0) \otimes \delta(y - y_0), \tag{1.14}$$

where x_0 and y_0 are given. It is derived from the special case where the signal processes $Z = (X, Y)'$ satisfies (1.4) with

$$\beta = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad \varsigma = \begin{bmatrix} \sqrt{1-\rho_x} & 0 \\ 0 & \sqrt{1-\rho_y} \end{bmatrix}, \quad \gamma = \begin{bmatrix} \sqrt{\rho_x} & 0 \\ \sqrt{\rho_y}\rho_{xy} & \sqrt{\rho_y(1-\rho_{xy}^2)} \end{bmatrix}.$$

A classical result states that, for a class of SPDEs including (1.13), with initial condition in L_2, there exists a unique solution $v \in L_2(\Omega \times (0,T), \mathcal{F}, L_2(\mathbb{R}^2))$ [74]. This does not include Dirac initial data (1.14), but in fact, the solution to (1.13) and (1.14) is analytically known to be the smooth (in x and y) function

$$v(T, x, y) = \frac{\exp\left(-\frac{\left(x - x_0 - \mu_x T - \sqrt{\rho_x}W_t^x\right)^2}{2(1-\rho_x)T} - \frac{\left(y - y_0 - \mu_y T - \sqrt{\rho_y}W_t^y\right)^2}{2(1-\rho_y)T}\right)}{2\pi\sqrt{(1-\rho_x)(1-\rho_y)}\,T}. \tag{1.15}$$

The availability of a closed-form solution in this case helps us check the validity of our numerical scheme and its convergence rate, although the scheme itself is more widely applicable.

For the SPDE (1.13), we consider both explicit and implicit Milstein schemes. We study the mean-square stability and mean-square convergence. This can give us a bound for the expected error. The advantage over the simpler Euler scheme is that the strong

convergence order is improved from 1/2 to 1. As expected, we find that the explicit scheme is stable in the mean-square sense only under a strong CFL-type condition on the timestep, $k \leq Ch^2$ for timestep k, mesh size h, and a constant C, while the implicit scheme is mean-square stable under the very mild and somewhat unusual CFL condition $k \leq C|\log h|^{-1}$ (provided also some constraints on $\rho_x, \rho_y, \rho_{xy}$).

We therefore focus on the implicit scheme, for which we prove first order convergence in the timestep and second order in the spatial mesh size. The analysis is made more difficult by the Dirac initial datum compared to, say, L_2 initial data. We adapt the approach used in [15] for the heat equation by studying the convergence for different wave number regions in Fourier space and then assemble the contributions to the error by the inverse transform.

Furthermore, we use an Alternating Direction Implicit (ADI) scheme to approximately factorise the discretisation matrix for the implicit elliptic part in (1.13). This concept is well established for PDEs (see, e.g., [88, 19, 63]). It is well known that in the multi-dimensional case standard implicit schemes result in sparse banded linear systems, which cannot be solved by direct elimination in a computational cost which scales linearly with the number of unknowns like in the one-dimensional, tridiagonal case. An alternative to advanced iterative linear solvers such as multigrid methods is to reduce the large sparse linear system approximately to a sequence of tridiagonal linear systems, which are computationally easier to handle, by ADI factorisation. To our knowledge, the present work is the first application of ADI to SPDEs. We show that the ADI approximation is also mean-square stable under the same conditions as the original implicit scheme and has the same convergence order.

We note that published analysis of ADI schemes for parabolic PDEs in the presence of mixed spatial derivative terms is currently restricted to constant coefficients (through the use of von Neumann stability analysis; see e.g. [101]), and we restrict the analysis in the SPDE case to constant coefficients as well. Notwithstanding this, the empirical evidence overwhelmingly suggests that the conclusions drawn there extend to most cases of variable coefficients.

We give the natural extension to the SPDE (1.3) with variable coefficients. In that case, additional iterated stochastic integrals (the *Lévy area*) appear in the Milstein approximation. The efficient, accurate simulation has been studied in the context of SDEs in [72, 37] and invariably leads to relatively complicated schemes. As the computational effort in the context of the SPDE (1.3) is dominated by the matrix computations from the finite difference scheme, we perform a simple approximation of the stochastic integrals

$\int_t^{t+k}(W_s - W_t)\,\mathrm{d}B_s$, for correlated Brownian motions W and B, by simple Euler integration with step k^2. This is sufficiently accurate not to spoil the first order convergence and does not increase the complexity order.

As a specific application, we approximate the equation

$$\begin{aligned}\mathrm{d}u \;=\; &\Big[\kappa_1 u - \Big(r_1 - \frac{1}{2}y - \xi_1\rho_3\rho_{1,1}\rho_{2,1}\Big)\frac{\partial u}{\partial x} - \Big(\kappa_1(\theta_1 - y) - \xi_1^2\Big)\frac{\partial u}{\partial y} + \frac{1}{2}y\frac{\partial^2 u}{\partial x^2} \\ &+\xi_1\rho_3\rho_{1,1}\rho_{2,1}y\frac{\partial^2 u}{\partial x\partial y} + \frac{\xi_1^2}{2}y\frac{\partial^2 u}{\partial y^2}\Big]\,\mathrm{d}t - \rho_{1,1}\sqrt{y}\frac{\partial u}{\partial x}\,\mathrm{d}W_t - \xi_1\rho_{2,1}\frac{\partial}{\partial y}(\sqrt{y}u)\,\mathrm{d}B_t,\end{aligned} \tag{1.16}$$

taken from [55], with the scheme presented in this paper. Although our analysis (based on Fourier transforms) does not directly apply in this case, the scheme preserves first order convergence in time and second order convergence in space in our numerical tests.

The sharp estimates we derive give a precise description of the error given Dirac initial data. This is achieved by a Fourier analysis originating from Carter and Giles in [15], where they estimated the error arising from explicit and implicit approximations of the constant-coefficient 1-d convection-diffusion equation with Dirac initial data.

Moreover, the stability and error analysis in the constant coefficient case provide an accurate description of the local behaviour at $(t, x) = (0, x_0)$ in the variable coefficient case.[1] This connection is established by "freezing" the coefficients at the point $(0, x_0)$ in a more general SPDE of the form (1.13), such that we obtain

$$\begin{aligned}\mathrm{d}v_0(t,x) = &\Big(\frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x_0)\frac{\partial^2}{\partial x_i\partial x_j}v_0(t,x) \\ &- \sum_{i=1}^{d} b_i(x_0)\frac{\partial}{\partial x_i}v_0(t,x)\Big)\,\mathrm{d}t - \sum_{i=1}^{d}\Big(\sum_{l=1}^{m}\gamma_{i,l}(x_0)\Big)\frac{\partial}{\partial x_i}v_0(t,x)\,\mathrm{d}M_t^l.\end{aligned} \tag{1.17}$$

It can be seen by elementary calculus that (1.17) in the case $d = 2$ can be brought in the form (1.13) by a linear coordinate transformation, noting that $a - \gamma\gamma^\top$ is symmetric positive definite by (1.5). This reduction of a broader class of problems to a test equation with a known solution is in the spirit of the well-established paradigm of stability analysis for ODEs described in [59].

We give a numerical justification for this reduction in Section 3.4, by presenting numerical tests for the SPDEs (1.16) and (1.17) to compare the error at (x_0, y_0) between the original SPDE and SPDE with frozen coefficients at (x_0, y_0). We find that the behaviour is extremely close, especially for small time intervals, which confirms the broader usefulness of the method.

[1] Note that for simplicity the coefficients in (1.13) have no t-dependence.

After analysing mean-square convergence of the SPDE (1.13), next we consider in Chapter 4 the expectation of the following functional of the solution,

$$P_t(v) = \int_0^\infty \int_0^\infty v(t,x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{1.18}$$

a straightforward generalisation of the functional (1.1) in the 1-d case. In the context of filtering, P_t is related to the cumulative distribution function of the filter given the signal.

A classical approach to approximating $\mathbb{E}[P_t]$ is the standard Monte Carlo method with an estimator

$$\widehat{Y} = \frac{1}{M}\sum_{m=1}^{M} \widehat{P}^{(m)},$$

where $\widehat{P}$ arises from a single sample of P, using a suitable approximation scheme for (1.13) and sampling of W on a discrete time mesh. For the SPDE (1.13), to achieve a root mean square error ε, this requires an overall computational cost $O(\varepsilon^{-4})$, as we require $O(\varepsilon^{-2})$ samples, $O(\varepsilon^{-1})$ time steps (e.g., for the Euler-Maruyama scheme with weak order 1), and $O(\varepsilon^{-1/2})$ mesh points in each direction (e.g., for central difference schemes with order 2). One way to reduce the cost is the MLMC method (see [38]) by using the SPDE solution on paths with a coarse timestep and spatial mesh as a control variate of solutions on paths with a fine timestep and mesh. Let P_l be an approximation to the functional P using a numerical discretisation with timestep k_l and mesh sizes $h_{x,l}$ and $h_{y,l}$. The MLMC estimator is based on recursive control variates embedded in the identity

$$\mathbb{E}[P_{l^*}] = \mathbb{E}[P_0] + \sum_{l=1}^{l^*} \mathbb{E}[P_l - P_{l-1}], \tag{1.19}$$

where l^* is a maximum refinement level. The variance is then reduced and less computation time is needed for a fixed accuracy. In this case $\mathbb{E}\big[|P_l - P_{l-1}|^2\big] = O\big((h_{x,l}^2 + h_{y,l}^2 + k_l)^2\big)$, which is small for large l, compared to the variance of the estimator for $\mathbb{E}\big[P_{l^*}\big]$. As a result, for a fixed accuracy ε, the cost can be reduced significantly to $O(\varepsilon^{-2}(\log\varepsilon)^2)$ by standard MLMC methodology as in [42]. However, this complexity of the MLMC method is not optimal as in the one-dimensional case in [42]. The reason is that the cost of each sample on higher levels increases with the same order as the variance decays. Moreover, the total cost of MLMC increases exponentially in the dimension.

The approach taken here is to approximate the SPDE (1.13) by the sparse grid combination technique. More specifically, for a d-dimensional problem, we define first order difference operators $\Delta_i,\ 1 \le i \le d$, such that

$$\Delta_i P_{\boldsymbol{l}}^N = P_{\boldsymbol{l}}^N - P_{\boldsymbol{l}-\boldsymbol{e}_i}^N \mathbb{1}_{l_i>0},$$

where $\boldsymbol{e}_i$ is the unit vector in direction i, N is the fixed number of time steps, and $P_{\boldsymbol{l}}^N$ is the estimator of level $\boldsymbol{l}$ (see Section 1.3.2). Then the telescoping sum becomes

$$\sum_{\boldsymbol{l}\in\mathbb{N}^d} \Big(\prod_{i=1}^{d} \Delta_i \Big) P_{\boldsymbol{l}}^N. \tag{1.20}$$

The strategy is to neglect terms with large $\boldsymbol{l}$, where the cost is large and the contribution to the estimator is small, by carefully choosing an index set $\mathcal{I} \subset \mathbb{N}^d$ of terms to include. In this way we get one approximation of $\widehat{P}$ with all $P_{\boldsymbol{l}}^N$ using the same Brownian path. Then, to evaluate $\mathbb{E}[P_t]$, we use M independent samples of $\widehat{P}$ and calculate the average. The benefit here is that with a RMSE ε, the total cost is fixed with $O(\varepsilon^{-7/2})$ up to a logarithmic term, as the cost for one realisation of (1.20) is $O(\varepsilon^{-3/2}|\log\varepsilon|^{5(d-1)/2})$, and the number of samples needed is $M = O(\varepsilon^{-2})$. Hence, this will outperform the MLMC method, whose total cost is $O(\varepsilon^{-1-d/2})$, when the dimension $d > 5$.

To recover the optimal complexity $O(\varepsilon^{-2})$, we further combine the sparse combination technique and MLMC, in a different way from MIMC. Under the MLMC setting, the level difference $P_l - P_{l-1}$ in (1.19) is replaced with the difference of the solution under the sparse combination method, where the timestep and index set vary with levels. In this way, the total cost is optimal, $O(\varepsilon^{-2})$, independent of dimension.

In Chapter 5, we consider the small ϵ asymptotics[2] of the two-dimensional SPDE

$$\begin{aligned} \mathrm{d}v = {} & \frac{1}{\epsilon}\Big(\nu^2(1-\rho_y^2)\partial_{yy}v + \kappa\partial_y(yv)\Big)\,\mathrm{d}t \\ & + \Big(\frac{1}{2}\sigma^2\big(y+U_t\big)\partial_{xx}v - \mu\,\partial_x v\Big)\,\mathrm{d}t - \rho_x\sigma\big(y+U_t\big)\partial_x v\,\mathrm{d}W_t^x, \\ v(0,x,y) = {} & \delta(x-x_0)\otimes\delta(y-y_0), \end{aligned} \tag{1.21}$$

where $\nu > 0,\ \kappa > 0,\ \mu \in \mathbb{R},\ \rho_x \in (-1,1),\ 0 < \epsilon \ll 1$ are fixed constants, $\sigma(\cdot)$ is a real-valued function, and U_t follows the dynamics

$$\mathrm{d}U_t = -\frac{\kappa}{\epsilon}U_t\,\mathrm{d}t + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\rho_y\,\mathrm{d}W_t^y, \qquad U_0 = 0.$$

It is derived from the special case where the signal processes $Z = (X,Y)'$ satisfies (1.4) with

$$\beta = \begin{bmatrix}\mu\\ \kappa y\end{bmatrix}, \quad \varsigma = \begin{bmatrix}\sqrt{1-\rho_x^2}\sigma(y+U_t) & 0\\ 0 & \sqrt{\frac{2(1-\rho_y^2)}{\epsilon}}\nu\end{bmatrix}, \quad \gamma = \begin{bmatrix}\rho_x\sigma(y+U_t) & 0\\ 0 & 0\end{bmatrix}.$$

[2]This ϵ is different from the ε in the previous chapters, which denotes the root mean square error.

Specifically, the underlying particle system is

$$\begin{aligned}
\mathrm{d}X_t^i &= \mu\,\mathrm{d}t + \sigma(Y_t^i + U_t)\Big(\rho_x\,\mathrm{d}W_t^x + \sqrt{1-\rho_x^2}\,\mathrm{d}W_t^{x,i}\Big), \quad && X_0^i = x_0, \quad 1 \le i \le N,\\
\mathrm{d}Y_t^i &= -\frac{\kappa}{\epsilon}Y_t^i\,\mathrm{d}t + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\sqrt{1-\rho_y^2}\,\mathrm{d}W_t^{y,i}, && Y_0^i = y_0, \quad 1 \le i \le N,\\
\mathrm{d}U_t &= -\frac{\kappa}{\epsilon}U_t\,\mathrm{d}t + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\rho_y\,\mathrm{d}W_t^{y}, && U_0 = 0,
\end{aligned}$$

where N represents the number of particles, $W^{x,i}$ and $W^{y,i}$ are standard Brownian motions, independent from each other, and (W^x, W^y) is a bivariate Brownian motion with correlation ρ_{xy}, and $\sigma(\cdot)$ satisfies $\mathbb{E}\big[\int_0^T \sigma^2(Y_s + U_s)\,\mathrm{d}s\big] < \infty$.

This particle system may be used to model a large portfolio of firms, with X^i representing the dynamics of the log-asset price of i-th firm, $\sigma(Y^i + U)$ representing the stochastic volatility driven by a common factor U and an independent component W^y; W^x and W^y denoting the market Brownian drivers, and $W^{x,i}$ and $W^{y,i}$ being idiosyncratic Brownian noise. This is a stochastic volatility extension of the large portfolio model established in [14] (see also the extensions in [82, 1, 10], and especially [55] for a model with stochastic volatility).

Our goal is to find an asymptotic expansion of the solution $v(t,x,y)$ to the SPDE (1.21) as $\epsilon \to 0$, and consider the functional

$$P_t(v) = \int_0^\infty \int_{-\infty}^\infty v(t,x,y)\,\mathrm{d}y\,\mathrm{d}x, \tag{1.22}$$

which represents the proportion of firms which have not defaulted. We can also regard this integral over a step functional as a building block of a more generalised functional class, since some functions can be represented as convergent series of step functions.

We here present a conjecture of the expansion to the SPDE (1.21). We let $\epsilon \to 0$ in the x-terms of the SPDE, but keep ϵ fixed in the y-terms. Thus, we average the coefficients of the x-terms by the invariant distribution of Y and U. We further test numerically the zero term $v_{0,\epsilon}$ and first order term $v_{1,\epsilon}$, using an ADI implicit scheme, as explained in previous chapters. Numerical results show $\mathbb{E}[P_T(v) - v_{0,\epsilon})] = O(\sqrt{\epsilon})$, and $\mathbb{E}[P_T(v - v_{0,\epsilon} - v_{1,\epsilon})] = O(\epsilon)$ empirically.

We also solve a one-dimensional SPDE satisfied by the marginal density of X for the first order expansion term $v_{1,\epsilon}$, and the result coincides with numerical integration of $v_{1,\epsilon}$ over the y-dimension. Hence, if we only care about the marginal density of X, which is the case in the financial model, we just need to compute the 1-d marginal SPDE, which is much cheaper than computing the 2-d SPDE.

Chapter 2

A one-dimensional Zakai SPDE

In this chapter, we propose a space-time Multi-Index Monte Carlo (MIMC) estimator for a one-dimensional parabolic stochastic partial differential equation (SPDE) of Zakai type. We compare the complexity with the Multilevel Monte Carlo (MLMC) method of Giles and Reisinger (2012), and find, by means of Fourier analysis, that the MIMC method: (i) has suboptimal complexity of $O(\varepsilon^{-2}|\log\varepsilon|^3)$ for a root mean square error (RMSE) ε if the same spatial discretisation as in the MLMC method is used; (ii) has a better complexity of $O(\varepsilon^{-2}|\log\varepsilon|)$ if a carefully adapted discretisation is used; (iii) has to be adapted for non-smooth functionals. Numerical tests confirm these findings empirically.

This chapter is structured as follows. We define two implicit Milstein finite difference schemes in Section 2.1, and the MIMC estimators in Section 2.2. The main theoretical result is provided in Section 2.3, where we analyse the convergence order of a MIMC estimator using Fourier analysis. Section 2.4 gives an "optimal index set" for this MIMC estimator and derives the total complexity for fixed accuracy. Section 2.5 derives further results for alternative approximations, while Section 2.6 shows numerical experiments confirming the above findings. Section 2.7 offers summaries of the chapter.

2.1 Two implicit Milstein finite difference schemes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, on which there is given a one-dimensional standard Brownian motion W. We study the parabolic stochastic partial differential equation

$$\mathrm{d}v = -\mu \frac{\partial v}{\partial x}\,\mathrm{d}t + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}\,\mathrm{d}t - \sqrt{\rho}\frac{\partial v}{\partial x}\,\mathrm{d}W_t, \qquad (t,x) \in (0,T)\times\mathbb{R}, \tag{2.1}$$

where $T > 0$, M is a standard Brownian motion, and μ and $0 \le \rho < 1$ are real-valued parameters, subject to the Dirac initial data

$$v(0,x) = \delta(x - x_0), \tag{2.2}$$

where $x_0 \in \mathbb{R}$ is given. We study Dirac initial data here because in the corresponding particle system this models the important situation where all particles start from the same position. Moreover, this is computationally the most challenging scenario. Other initial data with more regularity may be treated by convolution. The problem is posed on an unbounded domain and hence there is no boundary condition imposed (apart from the usual assumption that the solution does not grow exponentially at infinity). This will be convenient for the application of Fourier analysis. For practical computations, we truncate the domain large enough and apply a zero Dirichlet boundary condition. This is valid because of the diffusion effect such that at any fixed time t, the solution has a Gaussian distribution.

A classical result states that, for a class of SPDEs including (2.1), with initial condition in $L_2(\mathbb{R}) = \{f : \int_{\mathbb{R}} f^2\,\mathrm{d}x < \infty\}$, there exists a unique solution $v \in L_2(\Omega\times(0,T), \mathcal{F}, L_2(\mathbb{R})) = \{f(\omega,t,\cdot) :\ f(\omega,t,\cdot) \in L_2(\mathbb{R}),\ f(\omega,t,\cdot) \text{ is } \mathcal{F}_t\text{-measurable},\ \mathbb{E}\int_0^T \|f(\omega,t)\|_{L_2}^2\,\mathrm{d}t < \infty\}$ [74]. This does not include Dirac initial data (2.2), but in fact, the solution to (2.1) and (2.2) at time T is analytically known (see [42]) to be the smooth (in x) function

$$v(T,x) = \frac{1}{\sqrt{2\pi(1-\rho)T}} \exp\left(-\frac{\big(x - x_0 - \mu T - \sqrt{\rho}W_T\big)^2}{2(1-\rho)T}\right). \tag{2.3}$$

Integrating (2.1) over a time interval $[t, t+k]$, we have

$$v(t+k,x) = v(t,x) + \int_t^{t+k} -\mu\frac{\partial v}{\partial x} + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}\,\mathrm{d}s - \int_t^{t+k} \sqrt{\rho}\frac{\partial v}{\partial x}\,\mathrm{d}W_s.$$

We then use a spatial grid with uniform spacing h and a timestep k, and let V_j^n be the approximation to $v\big(nk, jh\big)$, $n = 1,\dots,N$, $j \in \mathbb{Z}$, where $N = T/k$. We approximate $v(0,x)$ by

$$V_j^0 = \begin{cases} h^{-1}, & j = [x_0/h], \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

Here $[x]$ denotes the closest integer to x. To improve the accuracy of the approximation of v in the present case of Dirac initial data, we subsequently choose h such that x_0 is on the grid, then x_0/h is an integer.

By an explicit Milstein scheme [42] with standard central difference approximation to the spatial derivatives, we have

$$V^{n+1} = V^n - \frac{\mu k + \sqrt{\rho k} Z_n}{2h} D_1 V^n + \frac{(1-\rho)k + \rho k Z_n^2}{2h^2} D_2 V^n,$$

where Z_n are independent standard normal random variables, D_1 and D_2 are first and second central difference operators,

$$(D_1 V^n)_j = V^n_{j+1} - V^n_{j-1}, \quad (D_2 V^n)_j = V^n_{j+1} - 2V^n_j + V^n_{j-1}.$$

However, there is a condition for the mean-square stability of this explicit scheme (see [42]), $(1+2\rho^2)\frac{k}{h^2} \leq 1$. This is an obstacle for the use of the MIMC scheme where timestep and mesh size are varied independently. We can avoid the constraint on the timestep by using an implicit scheme instead [92]. Two implicit Milstein finite difference discretisations are conceivable for the SPDE (2.1):

1. Discretise the spatial derivatives first, then apply the Milstein scheme to the resulting system of SDEs, i.e.,

$$V^{n+1} = V^n - \frac{\mu k}{2h} D_1 V^{n+1} + \frac{k}{2h^2} D_2 V^{n+1} - \frac{\sqrt{\rho k} Z_n}{2h} D_1 V^n + \frac{\rho k (Z_n^2 - 1)}{8h^2} D_1^2 V^n, \quad (2.5)$$

   where $(D_1^2 V)_j = V_{j+2} - 2V_j + V_{j-2}$.

2. Apply the Milstein scheme first, then discretise the spatial derivatives. This gives

$$V^{n+1} = V^n - \frac{\mu k}{2h} D_1 V^{n+1} + \frac{k}{2h^2} D_2 V^{n+1} - \frac{\sqrt{\rho k} Z_n}{2h} D_1 V^n + \frac{\rho k (Z_n^2 - 1)}{2h^2} D_2 V^n. \quad (2.6)$$

These two schemes have the same effect on the multilevel algorithm, and (2.6) is used in [42]. However, we will show that (2.6) is less efficient using the multi-index algorithm, as these two schemes result in different orders of variance for the MIMC estimators. Since the scheme (2.5) is more efficient, we will explore it in detail in Section 2.3, and the analysis of (2.6) in Section 2.5.1 is then similar.

Let $\ell_2(\mathbb{Z}) = \{(\ldots, V_{-1}, V_0, V_1, \ldots) : (\sum_{n=-\infty}^{\infty} |V_n|^2)^{\frac{1}{2}} < \infty\}$ and $L_2(\Omega \times (0,T), \mathcal{G}, \ell_2(\mathbb{Z})) = \{V(\omega,t,\cdot) : V(\omega,t,\cdot) \in \ell_2(\mathbb{Z}), V(\omega,t,\cdot) \text{ is } \mathcal{F}_t\text{-measurable}, \mathbb{E}\int_0^T \|V(\omega,t)\|_{\ell_2}^2 \,\mathrm{d}t < \infty\}$.

The implicit schemes are unconditionally stable and converge in mean-square sense (ℓ_2 sense) under a constraint on the correlation [92], as the following theorems describe for (2.6), and an analogous result holds for (2.5).

Theorem 2.1.1 (Theorem 2.1 in [92]). *The implicit Milstein scheme* (2.6) *is unconditionally stable in the mean-square sense, provided* $0 < \rho \leq 1/\sqrt{2}$.

Theorem 2.1.2 (Theorem 2.2 in [92]). *Assume* $0 < \rho \leq 1/\sqrt{2}$, $T > 0$, $k = T/N$ *and* $\frac{k}{h^2} = \lambda > 0$ *is kept fixed. The implicit Milstein scheme* (2.6) *has the error expansion, for Dirac initial data* (2.2),

$$V_j^N - v(T, x_j) = k\, E_1(T, x_j) + h^2\, E_2(T, x_j) + o(k, h^2)\, R(T, x_j),$$

where $x_j = jh$, E_1, E_2 *and* R *are random variables with bounded moments.*

Proposition 2.1.1. *Under the conditions of Theorem 2.1.2 but without fixed* k/h^2, *then for* $h^2 \geq k$, *Theorem 2.1.2 still holds; for* $h^2 < k$, *the implicit Milstein scheme* (2.5) *has the error expansion*

$$V_j^N - v(T, x_j) = k\, E_1(T, x_j) + h^2\, E_2(T, x_j) + \theta^N h^{-1}\, E_3(T, x_j) + o(k, h^2)\, R(T, x_j), \quad (2.7)$$

where $N = T/k$, $0 < \theta < 1$, *and* θ *is independent of* h *and* k.

Proof. See Appendix A. □

Remark 2.1.1. *If for some* $\beta > 0$, *and a constant* $C_0 > 0$ *independent of* h *and* k,

$$\theta^{\frac{T}{k}} \leq 2^{-C_0} \cdot h^{3+\beta}, \quad (2.8)$$

or equivalently,

$$k \leq \frac{T \log_2(\theta^{-1})}{C_0 + (3+\beta)\log_2(h^{-1})}, \quad (2.9)$$

then the implicit Milstein scheme (2.5) *has the error expansion*

$$V_j^N - v(T, x_j) = k\, E_1(T, x_j) + h^2\, E_2(T, x_j) + o(k, h^2)\, R(T, x_j).$$

Remark 2.1.2. *Theorem 2.1.1, Theorem 2.1.2, Proposition 2.1.1 and Remark 2.1.1 are valid for both implicit Milstein schemes* (2.5) *and* (2.6).

We consider a specific linear functional of v for fixed T, the random variable

$$P = \int_{-\infty}^{0} v(T,x)\,\mathrm{d}x = 1 - \int_{0}^{\infty} v(T,x)\,\mathrm{d}x, \tag{2.10}$$

as discussed in the introduction, where v is the solution to (2.1) and (2.2).

By introducing integer multi-indices $\boldsymbol{l} = (l_1, l_2) \in \mathbb{N}^2$ as the index of space and time separately, we denote $P_{(l_1,l_2)}$ as the discrete approximations to P, with mesh size $h = h_0 \cdot 2^{-l_1}$, and timestep $k = k_0 \cdot 2^{-2l_2}$,

$$P_{(l_1,l_2)} = 1 - h\sum_{j=1}^{\infty} V_j^N - \frac{h}{2}V_0^N, \tag{2.11}$$

where the integral is approximated by the trapezoidal rule and V_j^N is determined by (2.5).

Proposition 2.1.2. *Let $P : \Omega \to \mathbb{R}^+$ be the random variable given by* (2.10)*. Let $P_{(l_1,l_2)} : \Omega \to \mathbb{R}^+$ be the approximation to P given by* (2.11)*. Assume $0 < \rho \leq 1/\sqrt{2}$. Then there exists a real number $C > 0$, such that for any $l_1^* \in \mathbb{N}$, any $h_0 > 0$, and any $k_0 > 0$ such that*

$$k_0 \leq \frac{T\log_2(\theta^{-1})}{C_0 + (3+\beta)(l_1^* + \log_2 h_0^{-1})},$$

where $0 < \theta < 1$, $\beta > 0$, $C_0 > 0$ are the same constants as in Remark 2.1.1, the following holds:

For any $0 \leq l_1 \leq l_1^*$, $l_2 \geq 0$,

$$\sqrt{\mathbb{E}\left[|P_{(l_1,l_2)} - P|^2\right]} \leq C\left(h^2 + k\right),$$

where $h = h_0 \cdot 2^{-l_1}$, $k = k_0 \cdot 2^{-2l_2}$.

Proof. Denote the trapezoidal approximation of $1 - \int_0^\infty v(T,x)\,\mathrm{d}x$ with mesh size h by

$$\overline{P} = 1 - h\sum_{j=1}^{\infty} v(T,x_j) - \frac{h}{2}v(T,0),$$

where $x_j = jh$. Since both P and $\overline{P}$ use the theoretical, smooth $v(T,x)$, we have

$$\begin{aligned}\left|\overline{P} - P\right| &= \left|\sum_{j=1}^{\infty}\left[\int_{x_{j-1}}^{x_j} v(T,x)\,\mathrm{d}x - \frac{1}{2}h\left(v(T,x_{j-1}) + v(T,x_j)\right)\right]\right| \\ &\leq \frac{1}{12}h^3\sum_{j=1}^{\infty}\max_{\zeta\in[x_{j-1},x_j]}\left|\frac{\partial^2 v}{\partial x^2}(T,\zeta)\right|.\end{aligned}$$

Given $v(T,x)$ from (2.3),

$$v(T,x) = \frac{1}{\sqrt{2\pi(1-\rho)T}}\exp\left(-\frac{\left(x - x_0 - \mu T - \sqrt{\rho}W_T\right)^2}{2(1-\rho)T}\right),$$

for all h_0, there exists a constant C_0, such that for $h \leq h_0$,

$$\mathbb{E}\left[\left|h\sum_{i=1}^{\infty}\max_{\zeta\in[x_{j-1},x_j]}\left|\frac{\partial^2 v}{\partial x^2}(T,\zeta)\right|\right|^2\right] \leq C_0.$$

Therefore

$$\mathbb{E}\left[\left|\overline{P} - P\right|^2\right] \leq \frac{1}{144}C_0\,h^4.$$

From the proof of Theorem 2.2 in [92], there exist random variables $\widetilde{C}_j$, and a constant $C_1 > 0$, all of which are independent of (l_1, l_2), such that

$$\left|V_j^N - v(T,x_j)\right| \leq \widetilde{C}_j(h^2 + k),$$

where

$$\mathbb{E}\left[\left|h\sum_{j=0}^{\infty}\widetilde{C}_j\right|^2\right] < C_1 < \infty.$$

Combining above equations, we have

$$\begin{aligned}&\mathbb{E}\left[\left|P_{(l_1,l_2)} - P\right|^2\right] = \mathbb{E}\left[\left|P_{(l_1,l_2)} - \overline{P} + \overline{P} - p\right|^2\right] \leq 2\,\mathbb{E}\left[\left|P_{(l_1,l_2)} - \overline{P}\right|^2\right] + 2\,\mathbb{E}\left[\left|\overline{P} - P\right|^2\right] \\ &= 2\mathbb{E}\left[\left|h\sum_{j=1}^{\infty}\left(v(T,x_j) - V_j^N\right) + \frac{h}{2}\left(v(T,0) - V_0^N\right)\right|^2\right] + 2\,\mathbb{E}\left[\left|\overline{P} - P\right|^2\right] \\ &\leq 4\left(h^4 + k^2\right)\mathbb{E}\left[\left|h\sum_{j=0}^{\infty}\widetilde{C}_j\right|^2\right] + \frac{1}{72}C_0\,h^4.\end{aligned}$$

Therefore, by choosing $C = \sqrt{\frac{1}{72}C_0 + 4C_1}$, we have for all $0 \le l_1 \le l_1^*$, $l_2 \ge 0$,

$$\sqrt{\mathbb{E}\left[|P_{(l_1,l_2)} - P|^2\right]} \le C\left(h^2 + k\right).$$

□

2.2 Multi-index Monte Carlo discretisation on a space-time mesh

Following the idea in [54], we introduce Δ_i, the first order difference operator along directions $i = 1, 2$, defined as

$$\Delta_i P_{\boldsymbol{l}} = \begin{cases} P_{\boldsymbol{l}} - P_{\boldsymbol{l}-\boldsymbol{e}_i}, & \text{if } l_i > 0, \\ P_{\boldsymbol{l}} & \text{if } l_i = 0, \end{cases} \tag{2.12}$$

with $\boldsymbol{e}_i$ being the canonical vectors in $\mathbb{R}^2$, i.e., $(\boldsymbol{e}_i)_j = \delta_{ij}$. Then define the first order mixed difference operator $\Delta = \Delta_1 \otimes \Delta_2$. Hence, for $l_1 > 0,\ l_2 > 0$, we have

$$\begin{aligned} \Delta P_{(l_1,l_2)} &= \left(P_{(l_1,l_2)} - P_{(l_1,l_2-1)}\right) - \left(P_{(l_1-1,l_2)} - P_{(l_1-1,l_2-1)}\right) \\ &= P_{(l_1,l_2)} - P_{(l_1,l_2-1)} - P_{(l_1-1,l_2)} + P_{(l_1-1,l_2-1)}. \end{aligned} \tag{2.13}$$

A telescoping sum then gives, for any l_1^*, $l_2^* \ge 0$,

$$P_{(l_1^*,l_2^*)} = \sum_{l_1=0}^{l_1^*}\sum_{l_2=0}^{l_2^*} \Delta P_{(l_1,l_2)}, \qquad \mathbb{E}\big[P_{(l_1^*,l_2^*)}\big] = \sum_{l_1=0}^{l_1^*}\sum_{l_2=0}^{l_2^*} \mathbb{E}\big[\Delta P_{(l_1,l_2)}\big].$$

By setting $\mathcal{I}_0 = \{(l_1, l_2) \in \mathbb{N}^2 : l_1 \le l_1^*,\ l_2 \le l_2^*\}$ and $P_{\mathcal{I}_0} = \sum_{(l_1,l_2)\in\mathcal{I}_0} \Delta P_{(l_1,l_2)}$, with h_{l^*}, k_0 satisfy the condition (2.8). Then the weak error between $P_{(l_1^*,l_2^*)}$ can be formed as

$$\mathbb{E}\big[P_{(l_1^*,l_2^*)} - P\big] = \sum_{l_1=0}^{l_1^*}\sum_{l_2=0}^{l_2^*} \mathbb{E}\big[\Delta P_{(l_1,l_2)}\big] < \varepsilon_0. \tag{2.14}$$

However, this is not necessarily the most efficient way of approximating P, since the approximations $P_{(l_1,l_2)}$ with $(l_1,\ l_2)$ close to $(l_1^*,\ l_2^*)$ are highly accurate but also costly to compute. Instead, we only compute the levels in a minimal index set $\mathcal{I} \subset \mathcal{I}_0$ (adapted to the convergence in h and k) which satisfies

$$\left| \sum_{(l_1,l_2)\in\mathcal{I}_0\setminus\mathcal{I}} \mathbb{E}\big[\Delta P_{(l_1,l_2)}\big] \right| < \varepsilon_1, \tag{2.15}$$

for a given error ε_1. Then it can be easily proved that

$$\left| \sum_{(l_1,l_2)\in\mathcal{I}} \mathbb{E}\big[\Delta P_{(l_1,l_2)}\big] - \mathbb{E}\big[P\big] \right| \leq \varepsilon_0 + \varepsilon_1. \tag{2.16}$$

The MIMC estimator is now defined by (see [54])

$$\widehat{P}_{\mathcal{I}} \coloneqq \sum_{(l_1,l_2)\in\mathcal{I}} \frac{1}{M_{(l_1,l_2)}} \sum_{m=1}^{M_{(l_1,l_2)}} \Delta P^{(m)}_{(l_1,l_2)}, \tag{2.17}$$

where $\mathcal{I} \subset \mathbb{N}^2$ is an index set, and $M_{(l_1,l_2)}$ is the number of samples for each $(l_1, l_2) \in \mathcal{I}$. The key point in the algorithm is that the quantity $\Delta P_{(l_1,l_2)}$ comes from four discrete approximations using the same Brownian path, therefore the variance of $\Delta P_{(l_1,l_2)}$ is small, in a sense made precise in Theorem 2.3.1. Thus the moments of $\Delta P_{(l_1,l_2)}$ are small not only if both k and h are small, but it suffices that either k or h is small. This allows the omission of computationally costly indices. The MIMC algorithm is based on a good choice of $\mathcal{I}$ and, for given $(l_1,\, l_2) \in \mathcal{I}$, $M_{(l_1,l_2)}$ to balance bias and variance for a given accuracy target.

Denote $E_{(l_1,l_2)} \coloneqq \big|\mathbb{E}\left[\Delta P_{(l_1,l_2)}\right]\big|$, $V_{(l_1,l_2)} \coloneqq \mathrm{Var}\big[\Delta P_{(l_1,l_2)}\big]$, and $W_{(l_1,l_2)}$ the work required for a realization of $\Delta P_{(l_1,l_2)}$ for a single path. Then the total work corresponding to the estimator is

$$W = \sum_{(l_1,l_2)\in\mathcal{I}} W_{(l_1,l_2)} M_{(l_1,l_2)}. \tag{2.18}$$

As the paths are chosen independently for different $(l_1,\, l_2)$, the variance of the estimator is

$$\mathrm{Var}\big[\widehat{P}_{\mathcal{I}}\big] = \sum_{(l_1,l_2)\in\mathcal{I}} \frac{V_{(l_1,l_2)}}{M_{(l_1,l_2)}}. \tag{2.19}$$

The mean square error can be expressed as the sum of two contributions, bias and variance,

$$\mathbb{E}\Big[\big(\widehat{P}_{\mathcal{I}} - \mathbb{E}[P]\big)^2\Big] = \mathrm{Var}\big[\widehat{P}_{\mathcal{I}}\big] + \Big(\mathbb{E}\big[\widehat{P}_{\mathcal{I}}\big] - \mathbb{E}[P]\Big)^2. \tag{2.20}$$

To achieve a root mean square error of ε, we split the accuracy as follows,

$$\big|\mathbb{E}\big[\widehat{P}_{\mathcal{I}} - P\big]\big| \leq \alpha\varepsilon, \qquad \mathrm{Var}\big[\widehat{P}_{\mathcal{I}}\big] \leq \big(1-\alpha^2\big)\varepsilon^2,$$

where $0 < \alpha < 1$. By optimising the total work with respect to $M_{\boldsymbol{l}}$ given the variance constraint (2.19) and a fixed $\mathcal{I}$, we can derive (see also [54]) the optimal number of samples for each level $\boldsymbol{l} = (l_1,\, l_2)$ to be

$$M_{\boldsymbol{l}} = (1-\alpha^2)^{-2}\varepsilon^{-2}\Big(\sum_{\boldsymbol{l}\in\mathcal{I}} \sqrt{V_{\boldsymbol{l}} W_{\boldsymbol{l}}}\Big)\sqrt{\frac{V_{\boldsymbol{l}}}{W_{\boldsymbol{l}}}}. \tag{2.21}$$

In the numerical implementation, we take the integer ceiling of $M_{(l_1,l_2)}$ in (2.21), as a result we assume the bound

$$M_{\boldsymbol{l}} \leq 1 + (1-\alpha^2)^{-2}\varepsilon^{-2}\Big(\sum_{\boldsymbol{l}\in\mathcal{I}}\sqrt{V_{\boldsymbol{l}}W_{\boldsymbol{l}}}\Big)\sqrt{\frac{V_{\boldsymbol{l}}}{W_{\boldsymbol{l}}}}, \qquad \forall\, \boldsymbol{l}\in\mathcal{I}.$$

Therefore the total work is bounded by

$$W \leq (1-\alpha^2)^{-2}\varepsilon^{-2}\Big(\sum_{\boldsymbol{l}\in\mathcal{I}}\sqrt{V_{\boldsymbol{l}}W_{\boldsymbol{l}}}\Big)^2 + \sum_{\boldsymbol{l}\in\mathcal{I}} W_{\boldsymbol{l}}, \tag{2.22}$$

where the first term dominates the second in our problem as $\varepsilon \to 0$ (see [54], Lemma 2.3).

In [54], Theorem 2.2 shows the total computational cost using an optimal index set with the MIMC method given bounds on $E_{(l_1,l_2)}$ and $V_{(l_1,l_2)}$. Based on these ideas, we first consider the Fourier analysis of each level $\Delta P_{(l_1,l_2)}$, then find an index set $\mathcal{I}$ to achieve the optimal order of the total work. Though the complexity is the same for both schemes (2.5) and (2.6) when using the multilevel method, we will see that it is different using the multi-index method.

2.3 Fourier analysis of MIMC estimators

In this section, we show that the MIMC increments (2.13) for scheme (2.5) satisfy the moment conditions required for the analysis of [54]. We then derive the optimal index set and the resulting complexity of the MIMC estimator in Section 2.4. In what follows, we will use C to stand for generic positive real constants dependent only on model parameters, and their values may change between occurrences.

2.3.1 Main results

The following theorem is concerned with the first and second moments of $\Delta P_{(l_1,l_2)}$ with schemes (2.5) and (2.11).

Theorem 2.3.1. *Consider $P_{(l_1,l_2)}$ from (2.11) with the implicit Milstein scheme (2.5). Assume $0 < \rho \leq 1/\sqrt{2}$. Then there exists a real number $C > 0$, such that for any $l_1^* \in \mathbb{N}$,*

any $h_0 > 0$, and any $k_0 > 0$ such that

$$k_0 \leq \frac{T \log_2(\theta^{-1})}{C_0 + (3+\beta)(l_1^* + \log_2 h_0^{-1})},$$

where $0 < \theta < 1$, $\beta > 0$, $C_0 > 0$ are the same constants as in Remark 2.1.1, the following holds:

For any $0 \leq l_1 \leq l_1^*$, $l_2 \geq 0$, the first and second moments of $\Delta P_{(l_1,l_2)}$ satisfy

$$\left|\mathbb{E}\left[\Delta P_{(l_1,l_2)}\right]\right| \leq C\, h^2 k, \qquad \mathbb{E}\left[\left|\Delta P_{(l_1,l_2)}\right|^2\right] \leq C\, h^4 k^2, \tag{2.23}$$

where $h = h_0 \cdot 2^{-l_1}$, $k = k_0 \cdot 2^{-2l_2}$.

Proof. See Section 2.3.3. □

Remark 2.3.1. *The condition $0 < \rho \leq 1/\sqrt{2}$ is the same already required for the mean-square stability in Theorem 2.1.1. It will be used for the damping of high wave number components (see Section 2.3.5).*

Remark 2.3.2. *In Theorem 2.3.1, to make sure the condition* (2.8) *holds, k_0 is not fixed. Instead, k_0 depends on l_1^*.*

It follows from Theorem 2.3.1 that the MIMC increments (2.13) (for scheme (2.5) and approximation (2.11) of P) satisfy

$$E_{(l_1,l_2)} = \mathbb{E}[\Delta P_{(l_1,l_2)}] \leq C_1^*\, h_0^2\, k_0 \cdot 2^{-2(l_1+l_2)}, \tag{2.24}$$

$$V_{(l_1,l_2)} = \mathrm{Var}[\Delta P_{(l_1,l_2)}] \leq C_2^*\, h_0^4\, k_0^2 \cdot 2^{-4(l_1+l_2)}, \tag{2.25}$$

$$W_{(l_1,l_2)} \leq C_3^*\, h_0^{-1} k_0^{-1} \cdot 2^{l_1+2l_2}, \tag{2.26}$$

where C_1^*, C_2^*, C_3^* are positive constants, and the constants will be used in the numerical implementation.

These are the Assumptions 1, 2, 3 in [54]. We will follow their approach in Section 2.4 to construct the index set $\mathcal{I}$, choose the number $M_{(l_1,l_2)}$ of samples and derive the complexity.

Similar to the proof of Theorem 2.3.1, we have the following.

Remark 2.3.3. *Assume $0 < \rho \leq 1/\sqrt{2}$. Then there exists a real number $C > 0$, such that for any $l_1^* \in \mathbb{N}$, any $h_0 > 0$, and any $k_0 > 0$ such that*

$$k_0 \leq \frac{T \log_2(\theta^{-1})}{C_0 + (3+\beta)(l_1^* + \log_2 h_0^{-1})},$$

where $0 < \theta < 1$, $\beta > 0$, $C_0 > 0$ are the same constants as in Remark 2.1.1, the following holds:

For any $0 \leq l_1 \leq l_1^*$, $l_2 \geq 0$ with $h = h_0 \cdot 2^{-l_1}$, $k = k_0 \cdot 2^{-2l_2}$, the first order differences of $P_{(l_1,l_2)}$ derived from (2.5) and (2.11) have the first and second moments

$$\left|\mathbb{E}\left[\Delta_1 P_{(l_1,l_2)}\right]\right| \leq C\,h^2, \qquad \mathbb{E}\left[\left|\Delta_1 P_{(l_1,l_2)}\right|^2\right] \leq C\,h^4,$$
$$\left|\mathbb{E}\left[\Delta_2 P_{(l_1,l_2)}\right]\right| \leq C\,k, \qquad \mathbb{E}\left[\left|\Delta_2 P_{(l_1,l_2)}\right|^2\right] \leq C\,k^2,$$

using the notation from (2.12).

In the remainder of this section, we give a proof of Theorem 2.3.1. The idea of the proof follows [15]. By applying the Fourier transform to the original SPDE (2.1) and the discretisation scheme (2.5), and splitting the Fourier domain into a low wave region and a high wave region, we calculate the error contribution from each wave number range. We then prove that the leading order error comes from the low wave region, and the error in the high wave region is negligible.

2.3.2 Fourier transform of the solution

Define the Fourier transform pair

$$\widetilde{v}(t,\gamma) = \int_{-\infty}^{\infty} v(t,x)\mathrm{e}^{-\mathrm{i}\gamma x}\,\mathrm{d}x,$$
$$v(t,x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{v}(t,\gamma)\mathrm{e}^{\mathrm{i}\gamma x}\,\mathrm{d}\gamma.$$

The Fourier transform of (2.1) yields

$$\mathrm{d}\widetilde{v} = -\widetilde{v}\Big(\mathrm{i}\mu\gamma\,\mathrm{d}t + \frac{1}{2}\gamma^2\,\mathrm{d}t + \mathrm{i}\sqrt{\rho}\gamma\,\mathrm{d}W_T\Big), \tag{2.27}$$

subject to the initial data $\widetilde{v}(0,\gamma) = \mathrm{e}^{-\mathrm{i}\gamma x_0}$ from (2.2). We assume that $\mu = 0$ in the following for simplicity, since the results will not change when a drift term appears. (See Remark 2.3 in [92].)

The solution to (2.27) is then

$$\widetilde{v}(t) = X(t)\mathrm{e}^{-\mathrm{i}\gamma x_0},$$

where

$$X(t) = \exp\Big(-\frac{1}{2}(1-\rho)\gamma^2 t - \mathrm{i}\gamma\sqrt{\rho}W_T\Big). \tag{2.28}$$

For the discretised equation, we can use a discrete-continuous Fourier decomposition

$$V_j^0 = \frac{1}{2\pi h}\int_{-\pi}^{\pi}\widetilde{V}^0(\gamma)\mathrm{e}^{\mathrm{i}(j-j_0)\gamma}\,\mathrm{d}\gamma,$$

for j_0 defined above (2.4), where

$$\widetilde{V}^0(\gamma) = h\sum_{j=-\infty}^{\infty}V_j^0\mathrm{e}^{-\mathrm{i}(j-j_0)\gamma}.$$

Since we approximate the Dirac initial data by (2.4), it follows that $\widetilde{V}^0(\gamma) = 1$. Then by linearity of the equation, we have

$$V_j^n = \frac{1}{2\pi h}\int_{-\pi}^{\pi}\widetilde{V}^n(\gamma)\mathrm{e}^{\mathrm{i}(j-j_0)\gamma}\,\mathrm{d}\gamma = \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h}\widetilde{V}^n(\gamma)\mathrm{e}^{\mathrm{i}(j-j_0)\gamma h}\,\mathrm{d}\gamma, \tag{2.29}$$

where we make the ansatz (i.e., we assume the solution has the following form, which in turn does satisfy the discretisation scheme)

$$\widetilde{V}^n(\gamma) = X_n(\gamma)\widetilde{V}^0(\gamma) = X_n(\gamma). \tag{2.30}$$

We can write (2.5) as

$$\Big(I + \frac{\mu k}{2h}D_1 - \frac{k}{2h^2}D_2\Big)V^{n+1} = \Big(I - \frac{\sqrt{\rho k}Z_n}{2h}D_1 + \frac{\rho k(Z_n^2-1)}{8h^2}D_1^2\Big)V^n. \tag{2.31}$$

Inserting in (2.31), a simple calculation gives

$$X_{n+1}(\gamma) = \frac{1 - \mathrm{i}c\sqrt{\rho k}Z_n + a\rho k(Z_n^2-1)}{1-\hat{a}k}X_n(\gamma), \tag{2.32}$$

where

$$a = -\frac{\sin^2\gamma h}{2h^2}, \qquad \hat{a} = -\frac{2\sin^2(\gamma h/2)}{h^2}, \qquad c = \frac{\sin\gamma h}{h}. \tag{2.33}$$

Following [57, 92], we say that the implicit Milstein scheme is mean-square stable, provided

$$\lim_{n\to\infty}\mathbb{E}\left[|X_n|^2\right] = 0 \quad \text{for all } \gamma.$$

Therefore, by stationarity we need

$$\mathbb{E}\big[|X_{n+1}|^2\big] < \mathbb{E}\big[|X_n|^2\big],$$

i.e.,

$$\mathbb{E}\left[\left|\frac{1 - \mathrm{i}c\sqrt{\rho k}Z_n + a\rho k(Z_n^2 - 1)}{1 - \hat{a}k}\right|^2\right] < 1.$$

This gives a condition on the correlation for mean-square stability (see [92]), stated previously as Theorem 2.1.1.

Now that the stability is ensured, we can approximate the functional of the solutions.

2.3.3 Fourier analysis of MIMC estimators (proof of Theorem 2.3.1)

From (2.11), (2.29) and (2.30), we can give a discrete approximation of the functional $P(v(T,\cdot))$ of the form

$$P_{(l_1,l_2)} = 1 - h\sum_{j=1}^{\infty} V_j^N - \frac{h}{2}V_0^N = 1 - \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} hX_N(\gamma)\left(\sum_{j=-j_0+1}^{\infty} \mathrm{e}^{\mathrm{i}jh\gamma} + \frac{1}{2}\mathrm{e}^{-\mathrm{i}j_0h\gamma}\right)\mathrm{d}\gamma.$$

To simplify the notation, we define

$$\chi(h,\gamma) = \sum_{j=-j_0+1}^{\infty} \mathrm{e}^{\mathrm{i}jh\gamma} + \frac{1}{2}\mathrm{e}^{-\mathrm{i}j_0h\gamma}, \tag{2.34}$$

where $j_0 = x_0/h$ as before. Note that χ is defined in a distributional sense, and it only appears multiplied by the smooth, fast decaying function X_N and in integral form, hence this is well-defined. Then $P_{(l_1,l_2)}$ has the form

$$P_{(l_1,l_2)} = 1 - \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} hX_N(\gamma)\chi(h,\gamma)\,\mathrm{d}\gamma. \tag{2.35}$$

Now we derive the leading order term of $\Delta P_{(l_1,l_2)}$, assuming that $\rho \le 1/\sqrt{2}$ holds.

Let $N' = N/4$ be the final timestep for level $(l_1,\, l_2 - 1)$. We denote $X_N^{l_1,l_2}$ as $X_N(\gamma)$ with $h = h_0 \cdot 2^{-l_1},\ k = k_0 \cdot 2^{-2l_2}$. From (2.35) we have

$$\begin{aligned}
\Delta P_{(l_1,l_2)} &= P_{(l_1,l_2)} - P_{(l_1,l_2-1)} - P_{(l_1-1,l_2)} + P_{(l_1-1,l_2-1)} \\
&= \frac{1}{2\pi}\int_{|\gamma|<\frac{\pi}{2h}} -h\big(X_N^{l_1,l_2} - X_{N'}^{l_1,l_2-1}\big)\chi(h,\gamma) + 2h\big(X_N^{l_1-1,l_2} - X_{N'}^{l_1-1,l_2-1}\big)\chi\big(2h,\gamma\big)\,\mathrm{d}\gamma \\
&\quad - \frac{1}{2\pi}\int_{\frac{\pi}{2h}<|\gamma|<\frac{\pi}{h}} h\big(X_N^{l_1,l_2} - X_{N'}^{l_1,l_2-1}\big)\chi(h,\gamma)\,\mathrm{d}\gamma.
\end{aligned} \tag{2.36}$$

We introduce $\Delta\widehat{P}_{(l_1,l_2)}(\gamma)$ as the integrand in (2.36),

$$\Delta\widehat{P}_{(l_1,l_2)}(\gamma)=\begin{cases}-h\big(X_N^{l_1,l_2}-X_{N'}^{l_1,l_2-1}\big)\chi(h,\gamma)+2h\big(X_N^{l_1-1,l_2}-X_{N'}^{l_1-1,l_2-1}\big)\chi(2h,\gamma), & \text{for } |\gamma|<\frac{\pi}{2h}\\ -h\big(X_N^{l_1,l_2}-X_{N'}^{l_1,l_2-1}\big)\chi(h,\gamma), & \text{for } \frac{\pi}{2h}<|\gamma|<\frac{\pi}{h}.\end{cases} \tag{2.37}$$

Hence we can express $\Delta P_{(l_1,l_2)}$ as

$$\Delta P_{(l_1,l_2)}=\frac{1}{2\pi}\int_{|\gamma|<\frac{\pi}{h}}\Delta\widehat{P}_{(l_1,l_2)}(\gamma)\,\mathrm{d}\gamma. \tag{2.38}$$

Moreover, after some computation we have for $|\gamma|<\frac{\pi}{2h}$,

$$\begin{aligned}\Delta\widehat{P}_{(l_1,l_2)}(\gamma)=&-2h\cdot\chi(2h,\gamma)\cdot\left(X_N^{l_1,l_2}-X_{N'}^{l_1,l_2-1}-X_N^{l_1-1,l_2}+X_{N'}^{l_1-1,l_2-1}\right)\\&-\big(h\cdot\chi(h,\gamma)-2h\cdot\chi(2h,\gamma)\big)\cdot\left(X_N^{l_1,l_2}-X_{N'}^{l_1,l_2-1}\right).\end{aligned} \tag{2.39}$$

Following the analysis in [15], we compare the numerical solution to the analytical solution by splitting the domain into two wave number regions. Assume p is a constant satisfying $0<p<\frac{1}{4}$. Then we define the low wave number region by

$$\Omega_{\text{low}}=\big\{\gamma:|\gamma|\le\min\{h^{-2p},\,k^{-p}\}\big\}, \tag{2.40}$$

and the high wave number region by

$$\Omega_{\text{high}}=\big\{\gamma:|\gamma|>\min\{h^{-2p},\,k^{-p}\}\big\}\cap\left[-\frac{\pi}{h},\frac{\pi}{h}\right]. \tag{2.41}$$

Then we have the following lemmas.

Lemma 2.3.1 (Low wave region)**.** *For $\Delta\widehat{P}_{(l_1,l_2)}$ introduced in* (2.37)*, there exists a constant $C>0$, such that for any $l_1,l_2\in\mathbb{N}$,*

$$\int_{\Omega_{low}}\Delta\widehat{P}_{(l_1,l_2)}(\gamma)\,\mathrm{d}\gamma=h^2k\cdot R(T),$$

where $R(T)$ is a random variable with bounded first and second moments satisfying

$$\Big|\mathbb{E}\big[R(T)\big]\Big|\le C,\quad \mathbb{E}\Big[\big|R(T)\big|^2\Big]\le C.$$

Proof. See Section 2.3.4. □

Lemma 2.3.2 (High wave region)**.** *There exists a real number $C > 0$, such that for any $l_1^* \in \mathbb{N}$, any $h_0 > 0$, and any $k_0 > 0$ such that*

$$k_0 \leq \frac{T \log_2(\theta^{-1})}{C_0 + (3+\beta)(l_1^* + \log_2 h_0^{-1})},$$

where $0 < \theta < 1$, $\beta > 0$, $C_0 > 0$ are the same constants as in Remark 2.1.1, the following holds for $h_0\, 2^{-l_1^*} \leq h \leq h_0$:

$$\int_{\Omega_{\text{high}}} \Delta\widehat{P}_{(l_1,l_2)}(\gamma)\, \mathrm{d}\gamma = h^{2+\frac{1}{2}\beta}\, k^r \cdot \widetilde{R}(T), \quad \forall\ r > 0,$$

where $\widetilde{R}(T)$ is a random variable with bounded first and second moments satisfying

$$\left|\mathbb{E}\big[\widetilde{R}(T)\big]\right| \leq C, \quad \mathbb{E}\Big[\big|\widetilde{R}(T)\big|^2\Big] \leq C.$$

Proof. See Section 2.3.5. □

Combining Lemma 2.3.1 and Lemma 2.3.2, we get

$$\Delta P_{(l_1,l_2)} = \frac{1}{2\pi} \int_{|\gamma| < \frac{\pi}{h}} \Delta\widehat{P}_{(l_1,l_2)}(\gamma)\, \mathrm{d}\gamma = h^2 k \cdot R(T) + h^{2+\frac{1}{2}\beta}\, k^r \cdot \widetilde{R}(T).$$

Hence there exists a constant $C > 0$ independent of (l_1, l_2), such that the first and second moments of $\Delta P_{(l_1,l_2)}$ satisfy

$$\left|\mathbb{E}\left[\Delta P_{(l_1,l_2)}\right]\right| \leq C\, h^2 k, \qquad \mathbb{E}\left[\left|\Delta P_{(l_1,l_2)}\right|^2\right] \leq C\, h^4 k^2.$$

This concludes the proof of Theorem 2.3.1. We will prove Lemma 2.3.1 and Lemma 2.3.2 in the remainder of this section.

2.3.4 Low wave number region (proof of Lemma 2.3.1)

First consider the case where γ is small, such that $\gamma \in \Omega_{\text{low}}$. To analyse the mean and the variance of $\Delta\widehat{P}_{(l_1,l_2)}(\gamma)$ in (2.39), it suffices to analyse the mean and the variance of

$$X_N^{l_1,l_2} - X_{N'}^{l_1,l_2-1} - X_N^{l_1-1,l_2} + X_{N'}^{l_1-1,l_2-1} \qquad \text{and} \qquad X_N^{l_1,l_2} - X_{N'}^{l_1,l_2-1}.$$

From (2.28) it follows that the exact solution of $X(t_{n+1})$ given $X(t_n)$ is

$$X(t_{n+1}) = X(t_n)\exp\Big(-\frac{1}{2}(1-\rho)\gamma^2 k - \mathrm{i}\gamma\sqrt{\rho}W_n\Big), \tag{2.42}$$

where $M_{t_{n+1}} - M_{t_n} \coloneqq W_n$ is the Brownian increment.

Now we consider the numerical approximation of (2.42), written as

$$X_{n+1}^{l_1,l_2} = C_n^{l_1,l_2} X_n^{l_1,l_2},$$

where

$$C_n^{l_1,l_2} \coloneqq \exp\Big(-\frac{1}{2}(1-\rho)\gamma^2 k - \mathrm{i}\gamma\sqrt{\rho}W_n + e_n^{l_1,l_2}\Big), \tag{2.43}$$

and $e_n^{l_1,l_2}$ is the logarithmic error between the numerical solution and the exact solution introduced during $[nk,(n+1)k]$. Aggregating over N timesteps, at $t_N = kN = T$,

$$X_N^{l_1,l_2} = X(T)\exp\Big(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\Big),$$

where $X(T) = \exp(-\frac{1}{2}(1-\rho)\gamma^2 T - \mathrm{i}\gamma\sqrt{\rho}\,W_T)$ is the exact solution at time T. Therefore

$$\begin{aligned}
&X_N^{l_1,l_2} - X_{N'}^{l_1,l_2-1}\\
&= X(T)\Big[\exp\Big(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\Big) - \exp\Big(\sum_{n=0}^{N'-1} e_n^{l_1,l_2-1}\Big)\Big],\\
&X_N^{l_1,l_2} - X_{N'}^{l_1,l_2-1} - X_N^{l_1-1,l_2} + X_{N'}^{l_1-1,l_2-1}\\
&= X(T)\Big[\exp\Big(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\Big) - \exp\Big(\sum_{n=0}^{N'-1} e_n^{l_1,l_2-1}\Big) - \exp\Big(\sum_{n=0}^{N-1} e_n^{l_1-1,l_2}\Big) + \exp\Big(\sum_{n=0}^{N'-1} e_n^{l_1-1,l_2-1}\Big)\Big].
\end{aligned} \tag{2.44}$$

From (2.43), we have

$$\sum_{n=0}^{N-1} e_n^{l_1,l_2} = \sum_{n=0}^{N-1} \log C_n^{l_1,l_2} + \frac{1}{2}(1-\rho)\gamma^2 T + \mathrm{i}\gamma\sqrt{\rho}\,W_T, \tag{2.45}$$

in which $C_n^{l_1,l_2}$ has the same form as in (2.32),

$$C_n^{l_1,l_2} = \frac{1 - \mathrm{i}c\sqrt{\rho k}Z_n + a\rho k(Z_n^2 - 1)}{1 - \hat{a}k},$$

where a, $\hat{a}$ and c are as in (2.33).

In the following, for a small parameter k, $\eta_1(k) = o(k)$, $\eta_2(k) = O(k)$, $\eta_3(k) = o(1)$, $\eta_4(k) = O(1)$ denote, respectively,

$$\begin{aligned}
\limsup_{k\to 0}\mathbb{E}\bigg[\Big(\frac{\eta_1(k)}{k}\Big)^2\bigg] = 0, \qquad \limsup_{k\to 0}\mathbb{E}\bigg[\Big(\frac{\eta_2(k)}{k}\Big)^2\bigg] < \infty,\\
\limsup_{k\to 0}\mathbb{E}\Big[\big(\eta_3(k)\big)^2\Big] = 0, \qquad \limsup_{k\to 0}\mathbb{E}\Big[\big(\eta_4(k)\big)^2\Big] < \infty.
\end{aligned}$$

Then one can derive from Taylor expansion of $\log C_n^{l_1,l_2}$,

$$\begin{aligned}\sum_{n=0}^{N-1} e_n^{l_1,l_2} &= \mathrm{i}(\gamma-c)\sqrt{\rho k}\sum_{n=0}^{N-1} Z_n + \big(a+\frac{1}{2}c^2\big)\rho k\sum_{n=0}^{N-1} Z_n^2 + \big(a+\frac{1}{2}\gamma^2\big)(1-\rho)T + (\hat{a}-a)T \\ &\quad + \mathrm{i}\big(ac+\frac{1}{3}c^3\big)\rho k\sqrt{\rho k}\sum_{n=0}^{N-1} Z_n^3 - \mathrm{i}ac\rho k\sqrt{\rho k}\sum_{n=0}^{N-1} Z_n - \big(\frac{1}{2}a^2+ac^2+\frac{1}{4}c^4\big)\rho^2k^2\sum_{n=0}^{N-1} Z_n^4 \\ &\quad + (a^2+ac^2)\rho^2k^2\sum_{n=0}^{N-1} Z_n^2 + \frac{1}{2}(\hat{a}^2-a^2\rho^2)kT + \gamma^4 o(k),\end{aligned}$$

where we have

$$\mathbb{E}\left[\left(\sum_{n=0}^{N-1} Z_n^{2m}\right)^2\right] \le \frac{(4m-1)!!\,T^2}{k^2}, \quad \mathbb{E}\left[\left(\sum_{n=0}^{N-1} Z_n^{2m-1}\right)^2\right] = \frac{(4m-3)!!T}{k}.$$

Moreover, this discretisation scheme satisfies

$$a+\frac{1}{2}c^2 = 0.$$

This is important since the term $(a+\frac{1}{2}c^2)\rho k\sum Z_n^2$ vanishes in this case, and it follows that

$$\begin{aligned}\exp\left(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\right) &= \exp\left(\Big(\hat{a}-a+(a+\frac{1}{2}\gamma^2)(1-\rho)\Big)T\right)\cdot\Bigg[1+\mathrm{i}(\gamma-c)\sqrt{k}\sum_{n=0}^{N-1} Z_n + \frac{1}{2}(\hat{a}^2 \\ &\quad - a^2\rho^2)kT - \frac{1}{2}(\gamma-c)^2\rho k\Big(\sum_{n=0}^{N-1} Z_n\Big)^2 + \mathrm{i}\gamma^3 k\sqrt{\rho k}\sum_{n=0}^{N-1}\phi_1(Z_n) + \gamma^4k^2\sum_{n=0}^{N-1}\phi_2(Z_n) + \gamma^4 k\cdot r_0(k)\Bigg],\end{aligned}$$

where $\phi_1(\cdot)$ is an odd degree polynomial function, $\phi_2(\cdot)$ is an even degree polynomial function, and $r_0(k)$ is a random variable with bounded second moments such that

$$\lim_{l_2\to\infty}\mathbb{E}\big[|r_0(k)|^2\big] \to 0.$$

Note that the level $(l_1,\, l_2-1)$ has mesh size h and timestep $4k$, such that it has $N'=N/4$ time steps. To reduce the variance, we use the same Wiener process in these two levels, hence the Brownian increment $\widetilde{Z}_n$ on level $(l_1,\, l_2-1)$ satisfies

$$\widetilde{Z}_n = \frac{1}{2}(Z_{4n-3}+Z_{4n-2}+Z_{4n-1}+Z_{4n}), \quad \text{thus} \quad \sum_{n=0}^{N-1} Z_n = 2\sum_{n=0}^{N'-1}\widetilde{Z}_n.$$

As a result, we have

$$\begin{aligned}
&\exp\Big(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\Big) - \exp\Big(\sum_{n=0}^{N'-1} e_n^{l_1,l_2-1}\Big) \\
&= \exp\bigg(\Big(\hat{a} - a + (a+\frac{1}{2}\gamma^2)(1-\rho)\Big)T\bigg)\cdot\bigg[-\frac{3}{2}(\hat{a}^2 - a^2\rho^2)kT + \mathrm{i}\gamma^3 k\sqrt{k}\Big(\sum_{n=0}^{N-1}\phi_1(Z_n) \\
&\quad - \sum_{n=0}^{N'-1}\phi_1(\widetilde{Z}_n)\Big) + \gamma^4 k^2\Big(\sum_{n=0}^{N-1}\phi_2(Z_n) - \sum_{n=0}^{N'-1}\phi_2(\widetilde{Z}_n)\Big) + \gamma^4 k\cdot r_1(k)\bigg] \\
&:= k\cdot E_1(T,\gamma),
\end{aligned} \tag{2.46}$$

where

$$\lim_{l_2\to\infty}\mathbb{E}\big[|r_1(k)|^2\big] \to 0,$$

$$\begin{aligned}
E_1(T,\gamma) = \exp\bigg(\Big(\hat{a} - a + (a+\frac{1}{2}\gamma^2)(1-\rho)\Big)T\bigg)\cdot\bigg[-\frac{3}{2}(\hat{a}^2 - a^2\rho^2)T + \mathrm{i}\gamma^3\sqrt{k}\Big(\sum_{n=0}^{N-1}\phi_1(Z_n) \\
- \sum_{n=0}^{N'-1}\phi_1(\widetilde{Z}_n)\Big) + \gamma^4 k\Big(\sum_{n=0}^{N-1}\phi_2(Z_n) - \sum_{n=0}^{N'-1}\phi_2(\widetilde{Z}_n)\Big) + \gamma^4\, r_1(k)\bigg].
\end{aligned}$$

Since

$$\begin{aligned}
a &= -\frac{\gamma^2}{2}\cdot\frac{\sin^2 h\gamma}{h^2\gamma^2} = -\frac{\gamma^2}{2} + \frac{\gamma^4}{6}h^2 + \gamma^6 O(h^4), \\
\hat{a} &= -\frac{\gamma^2}{2}\cdot\frac{\sin^2(h\gamma/2)}{(h\gamma/2)^2} = -\frac{\gamma^2}{2} + \frac{\gamma^4}{24}h^2 + \gamma^6 O(h^4), \\
c &= \gamma\cdot\frac{\sin h\gamma}{h\gamma} = \gamma - \frac{\gamma^3}{6}h^2 + \gamma^5 O(h^4),
\end{aligned}$$

it follows that

$$\exp\bigg(\Big(\hat{a} - a + (a+\frac{1}{2}\gamma^2)(1-\rho)\Big)T\bigg) = 1 + \frac{1-4\rho}{24}T\cdot\gamma^4 h^2 + \gamma^6 O(h^4).$$

Therefore we derive

$$\begin{aligned}
&\exp\Big(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\Big) - \exp\Big(\sum_{n=0}^{N'-1} e_n^{l_1,l_2-1}\Big) - \exp\Big(\sum_{n=0}^{N-1} e_n^{l_1-1,l_2}\Big) + \exp\Big(\sum_{n=0}^{N'-1} e_n^{l_1-1,l_2-1}\Big) \\
&= \mathrm{i}\gamma^7 k\sqrt{k}\Big(\sum_{n=0}^{N-1}\phi_1(Z_n) - \sum_{n=0}^{N'-1}\phi_1(\widetilde{Z}_n)\Big)\cdot O(h^2) + \gamma^8 k^2\Big(\sum_{n=0}^{N-1}\phi_2(Z_n) \\
&\quad - \sum_{n=0}^{N'-1}\phi_2(\widetilde{Z}_n)\Big)\cdot O(h^2) + \gamma^8 h^2 k\frac{(1-\rho^2)(4\rho-1)T^2}{64} + \gamma^8 h^2 k\cdot r_2(k) \\
&:= h^2 k\cdot E_2(T,\gamma),
\end{aligned} \tag{2.47}$$

where

$$\lim_{l_2\to\infty}\mathbb{E}\big[|r_2(k)|^2\big]\to 0,$$

$$E_2(T,\gamma)=\mathrm{i}\gamma^7\sqrt{k}\Big(\sum_{n=0}^{N-1}\phi_1(Z_n)-\sum_{n=0}^{N'-1}\phi_1(\widetilde{Z}_n)\Big)\cdot O(1)+\gamma^8 k\Big(\sum_{n=0}^{N-1}\phi_2(Z_n)-\sum_{n=0}^{N'-1}\phi_2(\widetilde{Z}_n)\Big)\cdot O(1)+\gamma^8\frac{(1-\rho^2)(4\rho-1)T^2}{64}+\gamma^8\cdot r_2(k).$$

So we have

$$X_N^{l_1,l_2}-X_{N'}^{l_1,l_2-1}=k\cdot X(T)\cdot E_1(T,\gamma),$$
$$X_N^{l_1,l_2}-X_{N'}^{l_1,l_2-1}-X_N^{l_1-1,l_2}+X_{N'}^{l_1-1,l_2-1}=h^2k\cdot X(T)\cdot E_2(T,\gamma).$$

Here, for $i=1,2$,

$$\int_{\Omega_{\mathrm{low}}}h\,\chi(h,\gamma)\cdot X(T)\,E_i(T,\gamma)\,\mathrm{d}\gamma$$

is the numerical approximation to

$$\int_0^\infty\int_{\Omega_{\mathrm{low}}}\mathrm{e}^{\mathrm{i}\gamma(x-x_0)}\cdot X(T)\,E_i(T,\gamma)\,\mathrm{d}\gamma\,\mathrm{d}x.$$

Because of the exponential decay of $X(T)$ in γ, there exists C_0 independent of (l_1,l_2), such that

$$\mathbb{E}\left|\int_{\Omega_{\mathrm{low}}}h\cdot\chi(h,\gamma)\Big(X(T)\cdot E_i(T,\gamma)\Big)\,\mathrm{d}\gamma-\int_0^\infty\int_{\Omega_{\mathrm{low}}}\mathrm{e}^{\mathrm{i}\gamma(x-x_0)}X(T)\cdot E_i(T,\gamma)\,\mathrm{d}\gamma\,\mathrm{d}x\right|\le C_0\,h^2.$$

Then there exists a constant $C_1>0$ independent of l_1 and l_2, such that

$$\begin{aligned}&\int_{\Omega_{\mathrm{low}}}\big(h\cdot\chi(h,\gamma)-2h\cdot\chi(2h,\gamma)\big)\cdot\big(X_N^{l_1,l_2}-X_{N'}^{l_1,l_2-1}\big)\,\mathrm{d}\gamma\\&=h^2k\int_{\Omega_{\mathrm{low}}}\big(h\cdot\chi(h,\gamma)-2h\cdot\chi(2h,\gamma)\big)\cdot X(T)\,E_1(T,\gamma)\,\mathrm{d}\gamma\coloneqq h^2k\cdot R_1(T),\\&\int_{\Omega_{\mathrm{low}}}2h\,\chi(2h,\gamma)\cdot\big(X_N^{l_1,l_2}-X_{N'}^{l_1,l_2-1}-X_N^{l_1-1,l_2}+X_{N'}^{l_1-1,l_2-1}\big)\,\mathrm{d}\gamma\\&=h^2k\int_{\Omega_{\mathrm{low}}}2h\,\chi(2h,\gamma)\cdot X(T)\,E_2(T,\gamma)\,\mathrm{d}\gamma\coloneqq h^2k\cdot R_2(T),\end{aligned}$$

where $R_1(T)$, $R_2(T)$ are random variables with bounded moments satisfying

$$\Big|\mathbb{E}\big[R_i(T)\big]\Big|\le C_1,\quad\mathbb{E}\Big[\big|R_i(T)\big|^2\Big]\le C_1,\qquad i=1,2.$$

Then Lemma 2.3.1 follows by letting

$$R(T)=-\big(R_1(T)+R_2(T)\big),\quad C=4C_1.$$

2.3.5 High wave number region (proof of Lemma 2.3.2)

Now we consider the case where γ is large, such that $\gamma \in \Omega_{\text{high}}$. By (2.32), we have

$$\begin{aligned} X_N &= X_0 \prod_{n=0}^{N-1} \frac{1 - \mathrm{i}\sqrt{k}cZ_n + k\rho a(Z_n^2 - 1)}{1 - k\hat{a}} \\ &= X_0 \prod_{n=0}^{N-1} \left(1 - \mathrm{i}\gamma\sqrt{\rho k}Z_n \frac{\sin h\gamma}{h\gamma} - \frac{\rho}{2}\gamma^2 k(Z_n^2 - 1)\frac{\sin^2 h\gamma}{(h\gamma)^2}\right)\left(1 + \frac{1}{2}\gamma^2 k \frac{\sin^2(h\gamma/2)}{(h\gamma/2)^2}\right)^{-1} \\ &= X_0 \prod_{n=0}^{N-1} \left(1 - \mathrm{i}\gamma\sqrt{\rho k}Z_n \frac{\sin h\gamma}{h\gamma} - \frac{\rho}{2}\gamma^2 ku(Z_n^2 - 1)\cos^2\frac{h\gamma}{2}\right)\left(1 + \frac{1}{2}\gamma^2 ku\right)^{-1}, \end{aligned}$$

denoting

$$u = \frac{\sin^2 \frac{h\gamma}{2}}{(\frac{h\gamma}{2})^2} \Longrightarrow \frac{\sin^2 h\gamma}{(h\gamma)^2} = \cos^2\frac{h\gamma}{2} \cdot u.$$

Note that $u \in [4/\pi^2, 1]$ when $|h\gamma| < \pi$. We have

$$\lim_{N\to\infty} \mathbb{E}\big[X_N\big] = \lim_{N\to\infty} X_0\left(1 + \frac{1}{2}\gamma^2 ku\right)^{-N} = X_0 \cdot \exp\left(-\frac{1}{2}\gamma^2 uT\right), \tag{2.48}$$

and

$$\begin{aligned} \lim_{N\to\infty} \mathbb{E}\big[|X_N|^2\big] &= \lim_{N\to\infty} X_0^2\left(\frac{1 + \rho\gamma^2 ku\cos^2\frac{h\gamma}{2} + \frac{\rho^2}{2}\gamma^4 k^2u^2\cos^4\frac{h\gamma}{2}}{1 + \gamma^2 ku + \frac{1}{4}\gamma^4 k^2u^2}\right)^N, \\ &= \lim_{N\to\infty} X_0^2\left(1 - \gamma^2 ku\frac{1 - \rho\cos^2\frac{h\gamma}{2} + \frac{1}{4}(1 - 2\rho^2\cos^4\frac{h\gamma}{2})\gamma^2 ku}{1 + \gamma^2 ku + \frac{1}{4}\gamma^4 k^2u^2}\right)^N \\ &= X_0^2 \cdot \exp\left(-\gamma^2 uT\frac{1 - \rho\cos^2\frac{h\gamma}{2} + \frac{1}{4}(1 - 2\rho^2\cos^4\frac{h\gamma}{2})\gamma^2 ku}{1 + \gamma^2 ku + \frac{1}{4}\gamma^4 k^2u^2}\right). \end{aligned}$$

Since we have assumed that $2\rho^2 \le 1$, the numerator $1 - \rho\cos^2\frac{h\gamma}{2} + \frac{1}{4}(1 - 2\rho^2\cos^4\frac{h\gamma}{2})\gamma^2 ku$ is positive.

When $h^2 \ge k$, the high wave region is

$$h^{-2p} < |\gamma| < \pi h^{-1}.$$

In this case,

$$\mathbb{E}\left[\left|\int_{\Omega_{\text{high}}} \Delta\widehat{P}_{(l_1,l_2)}(\gamma)\,\mathrm{d}\gamma\right|^2\right] = o(h^r), \qquad \forall r > 0.$$

When $h^2 < k$, following the proof of Proposition 2.1.1 in Appendix A, we have

$$\int_{\Omega_{\text{high}}} \mathbb{E}\,|X_N(\gamma)|^2\,\mathrm{d}\gamma \le \widetilde{C}\,\theta^N h^{-1}, \tag{2.49}$$

where $\widetilde{C} > 0$, $0 < \theta < 1$ are constants independent of h and k. As k_0 satisfies the condition

$$k_0 \leq \frac{T \log_2(\theta^{-1})}{C_0 + (3+\beta)(l_1^* + \log_2 h_0^{-1})},$$

or, equivalently,

$$\theta^{\frac{T}{k_0}} \leq 2^{-C_0} \cdot h^{3+\beta}, \qquad \text{for all } h \geq h_0\, 2^{-l_1^*},$$

by using (2.39) and a similar argument as in Appendix A, it follows that there exists $C > 0$ independent of h and k, such that for all $r > 0$,

$$\left|\mathbb{E}\left[\int_{\Omega_{\text{high}}} \Delta\widehat{P}_{(l_1,l_2)}(\gamma)\,\mathrm{d}\gamma\right]\right| \leq C \cdot h^{2+\frac{1}{2}\beta} k^r, \quad \mathbb{E}\left[\left|\int_{\Omega_{\text{high}}} \Delta\widehat{P}_{(l_1,l_2)}(\gamma)\,\mathrm{d}\gamma\right|^2\right] \leq C \cdot h^{2+\frac{1}{2}\beta} k^r,$$

since in (2.49),

$$\theta^N h^{-1} = \theta^{N_0} h^{-1} \theta^{\frac{k_0}{k}} \leq C \cdot h^{2+\frac{1}{2}\beta} \theta^{\frac{k_0}{k}}.$$

This concludes the proof of Lemma 2.3.2.

2.4 Optimal index set and complexity

So far, we have derived the first and second moments of $\Delta P_{(l_1,l_2)}$. Based on these, we now need to find an appropriate index set $\mathcal{I}$ for the MIMC estimator (2.17). We first do this for scheme (2.5) and approximation (2.11).

Finding the optimal (with respect to the computational cost required for a given prescribed accuracy $\alpha\varepsilon$) index set $\mathcal{I}$ is equivalent to solving the optimisation problem

$$\min_{\mathcal{I} \subset \mathbb{N}^2} W(\mathcal{I}) \qquad \text{such that} \quad \left|\mathbb{E}\big[\widehat{P}_{\mathcal{I}} - P\big]\right| < \alpha\varepsilon, \tag{2.50}$$

where W is the total work as in (2.18) and $\widehat{P}_{\mathcal{I}}$ the MIMC estimator for P as in (2.17), $\alpha \in (0,1)$. From (2.16), we have

$$\left|\sum_{(l_1,l_2)\in\mathcal{I}} \mathbb{E}\big[\Delta P_{(l_1,l_2)}\big] - \mathbb{E}\big[P\big]\right| \leq \varepsilon_0 + \varepsilon_1,$$

where ε_0 defined in (2.14) is the error between $P_{(l^*,l^*)}$ and P, and ε_1 defined in (2.15) is the error between two index sets. We let $\varepsilon_0 = (\alpha\varepsilon)^{1+r}$ with $r > 0$, $0 < \alpha < 1$, and $\varepsilon_1 = \alpha\varepsilon$. Then the weak error between $P_{(l_1^*,l_2^*)}$ and P has a higher order, and the dominant weak error of the MIMC estimator would be $\varepsilon_1 = \alpha\varepsilon$.

Firstly we find the l_1^*, l_2^* and k_0. From Proposition 2.1.2, we have

$$\sqrt{\mathbb{E}\big[|P_{(l_1,l_2)} - P|^2\big]} \leq C(h^2 + k).$$

For

$$h_{l_1} = h_0 \cdot 2^{-l_1}, \quad k_{l_2} = k_0 \cdot 2^{-2l_2},$$

it is sufficient to have

$$\begin{aligned} C\, h_{l_1^*}^2 = C h_0^2 \cdot 2^{-2l_1^*} &\leq (\alpha\varepsilon)^{1+r}, \\ C\, k_{l_2^*} = C\, k_0 \cdot 2^{-2l_2^*} &\leq (\alpha\varepsilon)^{1+r} \\ \theta^{T/k_0} &= \varepsilon^{2(1+r)}. \end{aligned} \tag{2.51}$$

Hence we have

$$\begin{aligned} l_1^* &\geq \frac{1+r}{2}\log_2(\varepsilon^{-1}) + \frac{1+r}{2}\log_2(\alpha^{-1}) + \frac{1}{2}\log_2(C\, h_0^2) = O(\log_2(\varepsilon^{-1})), \\ l_2^* &\geq \frac{1+r}{2}\log_2(\varepsilon^{-1}) + \frac{1+r}{2}\log_2(\alpha^{-1}) + \frac{1}{2}\log_2(C\, k_0) = O(\log_2(\varepsilon^{-1})), \\ k_0 &= \frac{T\log_2(\theta^{-1})}{2(1+r)\log_2(\varepsilon^{-1})} = O\bigg(\frac{1}{\log(\varepsilon^{-1})}\bigg). \end{aligned} \tag{2.52}$$

We follow the idea in [54], and introduce the term "profit" as

$$P_{\boldsymbol{l}} = \frac{E_{\boldsymbol{l}}}{\sqrt{V_{\boldsymbol{l}} W_{\boldsymbol{l}}}}. \tag{2.53}$$

Here $E_{\boldsymbol{l}}$, defined as the expectation of $\Delta P_{\boldsymbol{l}}$ as before, can be regarded as the contribution to the solution (if included) or error (if not included) originating from level (l_1, l_2). The denominator $\sqrt{V_{\boldsymbol{l}} W_{\boldsymbol{l}}}$ is proportional to the total work on level (l_1, l_2), which is given by

$$M_{\boldsymbol{l}} W_{\boldsymbol{l}} = (1-\alpha^2)^{-2}\varepsilon^{-2}\bigg(\sum_{\boldsymbol{l}\in\mathcal{I}} \sqrt{V_{\boldsymbol{l}} W_{\boldsymbol{l}}}\bigg)\sqrt{V_{\boldsymbol{l}} W_{\boldsymbol{l}}}$$

from (2.21). This can be regarded as a knapsack problem, where we try to include those levels with small work and large bias. This gives rise to the following heuristic optimisation.

Following Lemma 2.1 in [54], we define a candidate index set as

$$\{\boldsymbol{l} \in \mathbb{N}^2 : P_{\boldsymbol{l}} > \nu\} \qquad \text{for some } \nu.$$

Note that this class may not contain the optimal index set, but we will see that it is sufficiently rich to obtain optimal complexity order. From (2.24) to (2.26), we deduce

$$P_{\boldsymbol{l}} = \frac{E_{\boldsymbol{l}}}{\sqrt{V_{\boldsymbol{l}} W_{\boldsymbol{l}}}} \approx C_P \cdot 2^{-(w_1 l_1 + w_2 l_2)}, \tag{2.54}$$

for some constant $C_P > 0$, where

$$w_1 = \frac{1}{2}, \qquad w_2 = 1.$$

We introduce strictly positive normalized weights as

$$\delta_1 = \frac{w_1}{w_1 + w_2}, \qquad \delta_2 = \frac{w_2}{w_1 + w_2}.$$

Then $\delta_1 + \delta_2 = 1$.

$$\delta_1 = \frac{1}{3}, \qquad \delta_2 = \frac{2}{3}.$$

Now the index set is denoted as

$$\mathcal{I}(l^*) = \left\{(l_1,\, l_2) \in \mathbb{N}^2 : \delta_1 l_1 + \delta_2 l_2 \leq l^*\right\} \cap \mathcal{I}_0, \tag{2.55}$$

where l^* can be derived from the bias constraint

$$\left| \sum_{(l_1,\, l_2) \in \mathcal{I}_0 \setminus \mathcal{I}} E_{(l_1,\, l_2)} \right| \leq \alpha\varepsilon. \tag{2.56}$$

We try to find the minimal l^* such that this bias constraint holds. See Figure 2.1.

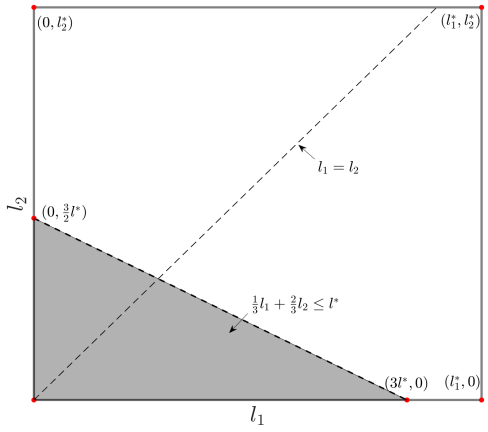


Figure 2.1: Optimal index set.

To this end, we approximate $E_{(l_1,\, l_2)}$ by $C_1^*\, h_0^2\, k_0 \cdot 2^{-2l_1 - 2l_2}$ from (2.24). Inserting this in (2.56), we choose l^* such that

$$2^{-3l^*} = \frac{3\,\alpha}{8C_1^*\, h_0^2\, k_0}\varepsilon, \quad \Rightarrow l^* = \frac{1}{3}\log_2\left(\frac{8C_1^*\, h_0^2\, k_0}{3\,\alpha}\varepsilon^{-1}\right). \tag{2.57}$$

Now we compute the computational cost under this index set. Recall from (2.22) that

$$W \leq (1 - \alpha^2)^{-2}\varepsilon^{-2}\left(\sum_{\boldsymbol{l} \in \mathcal{I}} \sqrt{V_{\boldsymbol{l}} W_{\boldsymbol{l}}}\right)^2 + \sum_{\boldsymbol{l} \in \mathcal{I}} W_{\boldsymbol{l}}.$$

Since the first term dominates the second one, we only need to compute

$$\Big(\sum_{\boldsymbol{\iota}\in\mathcal{I}}\sqrt{V_{\boldsymbol{\iota}}W_{\boldsymbol{\iota}}}\Big)^2 \sim k_0^{-1}\Big(\sum_{l_1=0}^{3l^*}\sum_{l_2=0}^{(-l_1+3l^*)/2} 2^{-\frac{3}{2}l_1-l_2}\Big)^2 = \; C\cdot k_0^{-1} = O\big(\big|\log\varepsilon\big|\big),$$

where C is a positive constant. Therefore, using discretisation (2.5) and $P_{(l_1,l_2)}$ (2.11), the order of the total computational cost W is

$$W = O\left(\varepsilon^{-2}\big|\log\varepsilon\big|\right). \tag{2.58}$$

Although theoretically there is a log term in the complexity, it usually does not affect the numerical tests, as θ is small, and θ^N dominates h^{-1}.

2.5 Some sub-optimal approximations

2.5.1 The alternative discretisation (2.6)

Recall the alternative discretisation scheme given in (2.6), in which we apply the Milstein scheme first, then apply a second finite difference with step size h. This makes no difference in the convergence order of the multilevel scheme, of which the second moment of $\Delta P_l \coloneqq P_{(l,l)} - P_{(l-1,l-1)}$ satisfies

$$\mathbb{E}\big[\,|\Delta P_l|^2\,\big] = O(h^4) + O(k^2) + O\left(h^4 k\right), \tag{2.59}$$

where $h = h_0\cdot 2^{-l}$, $k = k_0\cdot 4^{-l}$. As $O(h^2) = O(k)$ in order to balance the bias, $O(h^4k)$ is a higher order term in (2.59). However, in the multi-index scheme, $O(h^4k)$ is no longer a higher order term due to the independence of h and k. This leads to the difference in complexity between the MLMC and MIMC methods.

Proposition 2.5.1. *Consider the approximation* (2.6) *and* (2.11). *Assume* $\rho \le 1/\sqrt{2}$. *Then there exists a real number* $C > 0$, *such that for any* $l_1^* \in \mathbb{N}$, *any* $h_0 > 0$, *and any* $k_0 > 0$ *such that*

$$k_0 \le \frac{T\log_2(\theta^{-1})}{C_0 + (3+\beta)(l_1^* + \log_2 h_0^{-1})},$$

where $0 < \theta < 1$, $\beta > 0$, $C_0 > 0$ *are the same constants as in Remark 2.1.1, the following holds:*

For any $0 \le l_1 \le l_1^*$, $l_2 \ge 0$ *with* $h = h_0 \cdot 2^{-l_1}$, $k = k_0 \cdot 2^{-2l_2}$, *the first and second moments of* $\Delta P_{(l_1,l_2)}$ *satisfy*

$$\mathbb{E}\left[\Delta P_{(l_1,l_2)}\right] \le C\,h^2 k, \quad \mathbb{E}\big[\,\big|\Delta P_{(l_1,l_2)}\big|^2\,\big] \le C\,h^4 k. \tag{2.60}$$

Proof. In this discretisation scheme,

$$\left(I + \frac{\mu k}{2h} D_1 - \frac{k}{2h^2} D_2\right) V^{n+1} = \left(I - \frac{\sqrt{\rho k} Z_n}{2h} D_1 + \frac{\rho k (Z_n^2 - 1)}{2h^2} D_2\right) V^n.$$

Hence we have

$$X_{n+1}(\gamma) = \frac{1 - \mathrm{i}c\sqrt{\rho k} Z_n + \hat{a}\rho k (Z_n^2 - 1)}{1 - \hat{a}k} X_n(\gamma), \tag{2.61}$$

where

$$\hat{a} = -\frac{2\sin^2(\gamma h/2)}{h^2}, \qquad c = \frac{\sin \gamma h}{h}.$$

Following the previous analysis, in the low wave region (see Section 2.3.4) we can derive that

$$\begin{aligned}
\sum_{n=0}^{N-1} e_n^{l_1,l_2} &= \sum_{n=0}^{N-1} \log C_n^{l_1,l_2} + \frac{1}{2}(1-\rho)\gamma^2 T + \mathrm{i}\sqrt{\rho}\gamma W_T \\
&= \mathrm{i}(\gamma - c)\sqrt{\rho k} \sum_{n=0}^{N-1} Z_n + \Big(\hat{a} + \frac{1}{2}c^2\Big)\rho k \sum_{n=0}^{N-1} Z_n^2 + \Big(\hat{a} + \frac{1}{2}\gamma^2\Big)(1-\rho)T \\
&\quad + \mathrm{i}\Big(\hat{a}c + \frac{1}{3}c^3\Big)\rho k\sqrt{\rho k} \sum_{n=0}^{N-1} Z_n^3 - \mathrm{i}\hat{a}c\rho k\sqrt{\rho k} \sum_{n=0}^{N-1} Z_n - \Big(\frac{1}{2}\hat{a}^2 + \hat{a}c^2 + \frac{1}{4}c^4\Big)\rho^2 k^2 \sum_{n=0}^{N-1} Z_n^4 \\
&\quad + \big(\hat{a}^2 + \hat{a}c^2\big)\rho^2 k^2 \sum_{n=0}^{N-1} Z_n^2 + \gamma^4 o(k).
\end{aligned}$$

In this case, $\hat{a} + \frac{1}{2}c^2$ is no longer zero, so that

$$\exp\left(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\right) = \mathrm{e}^{(\hat{a} + \frac{1}{2}\gamma^2)(1-\rho)T} \cdot \left(\mathrm{i}(\gamma - c)\sqrt{\rho k} \sum_{n=0}^{N-1} Z_n + \Big(\hat{a} + \frac{1}{2}c^2\Big)\rho k \sum_{n=0}^{N-1} Z_n^2 + r(k)\right),$$

where $r(k)$ satisfies

$$\mathbb{E}[r(k)] = O(k), \qquad \mathbb{E}\big[\big|r(k)\big|^2\big] = O(k^2).$$

Now we have

$$\exp\left(\sum_{n=0}^{N-1} e_n^{l_1,\,l_2}\right) - \exp\left(\sum_{n=0}^{N'-1} e_n^{l_1,\,l_2-1}\right)$$
$$= \mathrm{e}^{(\hat{a}+\frac{1}{2}\gamma^2)(1-\rho)T}\cdot\left(\left(\hat{a}+\frac{1}{2}c^2\right)\rho k\left(\sum_{n=0}^{N-1} Z_n^2 - 4\sum_{n=0}^{N'-1}\widetilde{Z}_n^2\right) + r(k)\right).$$

However,

$$\mathbb{E}\left(\sum_{n=0}^{N-1} Z_n^2 - 4\sum_{n=0}^{N'-1}\widetilde{Z}_n^2\right)^2 = C_0\cdot k^{-1},$$

for some constant C_0. Hence similar to the proof of Theorem 2.3.1, there exists a constant C, such that

$$\left|\mathbb{E}\left[\int_{\Omega_{\text{low}}} \Delta\widehat{P}_{(l_1,\,l_2)}(\gamma)\,\mathrm{d}\gamma\right]\right| \le C\,h^2 k, \qquad \mathbb{E}\left[\left|\int_{\Omega_{\text{low}}} \Delta\widehat{P}_{(l_1,\,l_2)}(\gamma)\,\mathrm{d}\gamma\right|^2\right] \le C\,h^4 k.$$

For γ in the high wave region, the proof is the same as in Theorem 2.3.1. □

By a similar computation to Section 2.4, the index set is now

$$\mathcal{I}(l^*) = \left\{(l_1,\,l_2)\in\mathbb{N}^2 : \frac{1}{5}l_1 + \frac{4}{5}l_2 \le l^*\right\}\cap\mathcal{I}_0,$$

and the total computational cost W satisfies

$$W = O\left(\varepsilon^{-2}\left|\log\varepsilon\right|^3\right). \tag{2.62}$$

The corresponding numerical results are demonstrated at the end of Section 2.6.1.

2.5.2 An alternative functional approximation

Instead of approximating the loss by (2.11), consider a simpler approximation,

$$P_{(l_1,\,l_2)} = 1 - h\sum_{j=0}^{\infty} V_j^N = 1 - \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} hX_N(\gamma)\left(\sum_{j=-j_0}^{\infty}\mathrm{e}^{\mathrm{i}jh\gamma}\right)\mathrm{d}\gamma, \tag{2.63}$$

where $h = h_0\cdot 2^{-l_1}, k = k_0\cdot 2^{-2l_2}$. Using the discretisation from the previous subsection, we have the following results.

Proposition 2.5.2. *Consider the approximation* (2.6) *and* (2.63). *Assume* $\rho \leq 1/\sqrt{2}$. *Then there exists a real number* $C > 0$, *such that for any* $l_1^* \in \mathbb{N}$, *any* $h_0 > 0$, *and any* $k_0 > 0$ *such that*

$$k_0 \leq \frac{T \log_2(\theta^{-1})}{C_0 + (3+\beta)(l_1^* + \log_2 h_0^{-1})},$$

where $0 < \theta < 1$, $\beta > 0$, $C_0 > 0$ *are the same constants as in Remark 2.1.1, the following holds:*

For any $0 \leq l_1 \leq l_1^*$, $l_2 \geq 0$ *with* $h = h_0 \cdot 2^{-l_1}$, $k = k_0 \cdot 2^{-2l_2}$, *the first and second moments of* $\Delta P_{(l_1,l_2)}$ *satisfy*

$$\mathbb{E}\left[\Delta P_{(l_1,l_2)}\right] \leq C\, hk, \quad \mathbb{E}\left[\left|\Delta P_{(l_1,l_2)}\right|^2\right] \leq C\left(h^4 k + h^2 k^2\right). \tag{2.64}$$

Proof. Similar to the proof of Theorem 2.3.1. □

By the choice of h and k above, we get

$$O(h^4 k) = 2^{-2(l_1+l_2)-2l_1}, \quad O(h^2 k^2) = 2^{-2(l_1+l_2)-2l_2},$$

which gives

$$O\left(h^4 k\right) + O\left(h^2 k^2\right) = 2^{-2(l_1+l_2)-2\min\{l_1,l_2\}}.$$

As a consequence, the convergence rate varies between -2 and -3, depending on which refinement path is chosen.

The peculiarity of this scheme is that it has two leading terms of different orders in the variance, such that the assumptions of the framework in [54] (in this case, Assumption 2) are not satisfied. However, the fundamental principles of the MIMC concept still apply and we can split the domain into two parts: $\{(l_1,l_2) : l_1 \leq l_2\} \cap \mathcal{I}_0$ and $\{(l_1,l_2) : l_1 > l_2\} \cap \mathcal{I}_0$, and the optimal index set is as shown in Figure 2.2, where the analysis from Section 2.4 can be applied separately in the two parts. The total computational cost W satisfies

$$W = O\left(\varepsilon^{-2}\left|\log \varepsilon\right|^3\right). \tag{2.65}$$

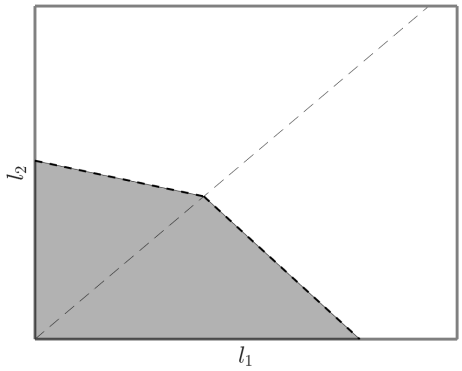
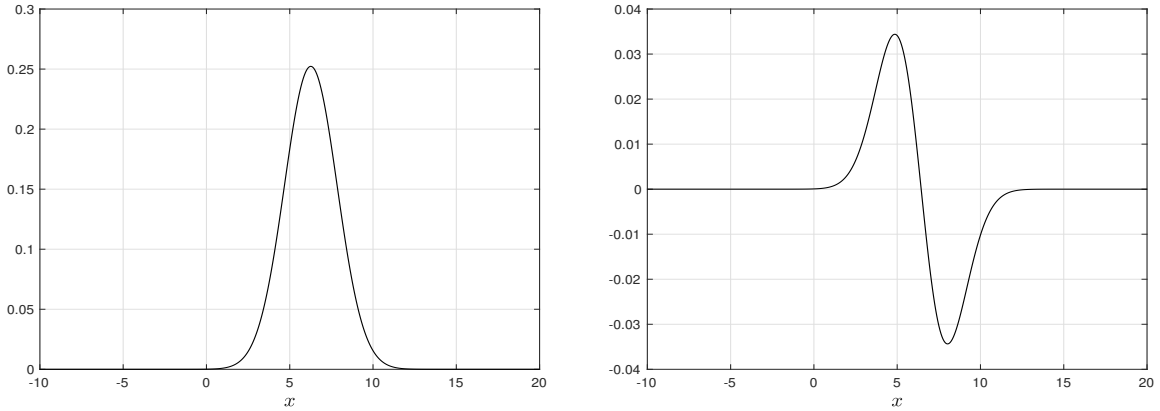


Figure 2.2: Optimal index set, with schemes (2.6) and (2.63).

2.6 Numerical implementation

Here, we present numerical tests for the schemes above to illustrate the theoretical convergence results and compare the complexity among MLMC and two MIMC schemes. Moreover we solve the SPDE on an interval $(x_{\min}, x_{\max})$ with zero boundary conditions. We choose the parameters $\mu = 0.081,\ \rho = 0.2,\ x_{\min} = -10,\ x_{\max} = 20$, and the initial density is $v(0, x) = \delta(x - x_0)$, where $x_0 = 5$.

First of all, we show in Figure 2.3 the numerical solution to (2.1), for a single Brownian path, with $h = 2^{-4},\ k = 4^{-5}$.



(a) Solution at $T = 5$.

(b) Pointwise error: $V_j^n - v(t_n, x_j)$

Figure 2.3: Numerical approximation to the SPDE (2.1) and the pointwise error to the analytic solution (2.3).

Figure 2.4 shows the values of θ from (2.7) corresponding to different correlation ρ, where θ is derived from the following equation

$$h \int_{\Omega_{\text{high}}} \left(\frac{1 + \rho\gamma^2 ku \cos^2 \frac{h\gamma}{2} + \frac{\rho^2}{2}\gamma^4 k^2 u^2 \cos^4 \frac{h\gamma}{2}}{1 + \gamma^2 ku + \frac{1}{4}\gamma^4 k^2 u^2} \right)^N \mathrm{d}\gamma = \theta^N. \tag{2.66}$$

The proof of (2.66) is given in Appendix A.

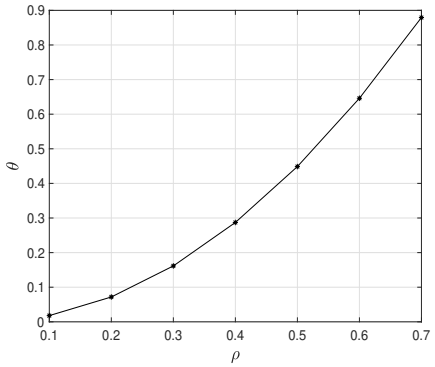


Figure 2.4: θ vs. ρ

For $\rho = 0.2$ and $T = 5$, $\theta = 0.0678$. So if $h_0 = 1$, we can choose $k_0 = 1/2$ for $l_1^* \leq 12$, and $k_0 = 1/4$ for $l_1^* \leq 25$, as then $\theta^{N_0} \leq 2^{-C_0}\, h^4$, where $h = h_0 \cdot 2^{-l_1^*}$ and $N_0 = T/k_0$. Therefore k_0 does not matter much in this problem.

2.6.1 Mean and variance of $\Delta P_{(l_1, l_2)}$

In this section, we conduct numerical tests for $\mathbb{E}\Big[\Delta P_{(l_1,l_2)}\Big]$ and $\mathrm{Var}\Big[\Delta P_{(l_1,l_2)}\Big]$. The analysis shows that for (2.5) and (2.11),

$$\mathbb{E}\Big[\Delta P_{(l_1\, l_2)}\Big] = O(h^2 k), \qquad \mathrm{Var}\Big[\Delta P_{(l_1,l_2)}\Big] = O(h^4 k^2),$$

where

$$h = h_0 \cdot 2^{-l_1}, \quad k = k_0 \cdot 2^{-2l_2},$$

such that

$$O\big(h^2 k\big) = 2^{-2(l_1+l_2)}, \quad O\big(h^4 k^2\big) = 2^{-4(l_1+l_2)}.$$

Table 2.1 shows $\log_2 \left|\mathbb{E}\Delta P_{(l_1,l_2)}\right|$ with different levels of timestep and mesh size, using 10^4 Monte Carlo samples.

The data in Table 2.1 suggest that for $l_1^* = 25$ $k_0 = 1/4$, from given $l \coloneqq l_1 + l_2$ to $l+1$ (i.e., from one diagonal to the next), the values decrease by around 2, in line with (2.24). Figure 2.5(a) is the contour plot of the values of $\log_2 \left|\mathbb{E}\Delta P_{(l_1,\, l_2)}\right|$.

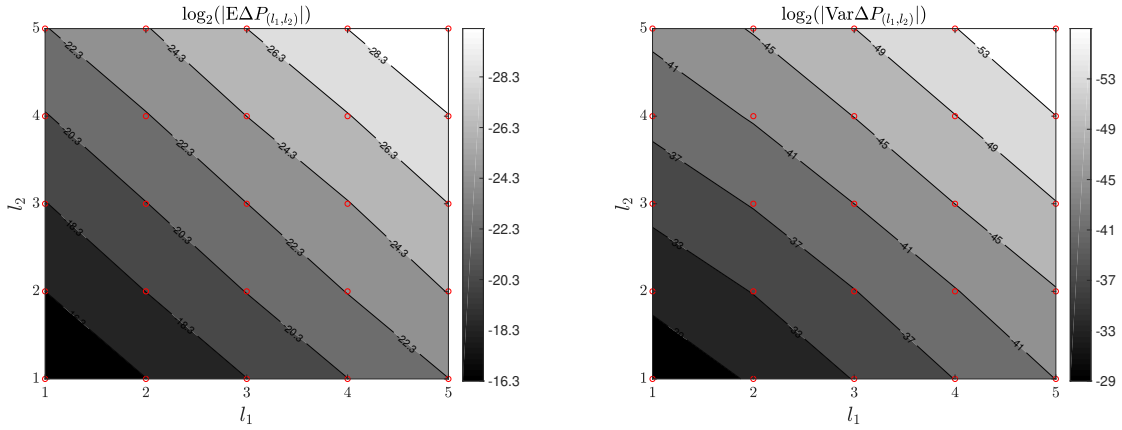
	$l_2 = 1$	$l_2 = 2$	$l_2 = 3$	$l_2 = 4$	$l_2 = 5$
$l_1 = 1$	-14.3384	-16.2991	-18.2382	-20.1987	-22.2193
$l_1 = 2$	-16.2725	-18.3163	-20.2538	-22.1928	-24.2061
$l_1 = 3$	-18.2749	-20.3508	-22.2862	-24.2709	-26.2991
$l_1 = 4$	-20.2770	-22.3616	-24.2059	-26.2203	-28.3113
$l_1 = 5$	-22.2776	-24.3644	-26.2984	-28.2543	-30.2064

Table 2.1: $\log_2 \left(\left|\mathbb{E}\Delta\left[P_{(l_1,\, l_2)}\right]\right|\right)$ with $h = h_0 \cdot 2^{-l_1}$ and $k = k_0 \cdot 2^{-2l_2}$, for approximation (2.5) and (2.11).

Table 2.2 shows the corresponding $\log_2 \left(\mathrm{Var}\left[\Delta P_{(l_1\, l_2)}\right]\right)$. We can see from the table that from given $l \coloneqq l_1 + l_2$ to $l+1$, the values decrease by approximately 4, in line with (2.25). Figure 2.5(b) is the contour plot of the values of $\log_2 \mathrm{Var}[\Delta P_{(l_1\, l_2)}]$.

	$l_2 = 1$	$l_2 = 2$	$l_2 = 3$	$l_2 = 4$	$l_2 = 5$
$l_1 = 1$	-26.2614	-30.0205	-34.0963	-38.1821	-42.0136
$l_1 = 2$	-29.3889	-33.1355	-37.2427	-41.3587	-45.2778
$l_1 = 3$	-33.1243	-36.8837	-41.0030	-44.9869	-48.9696
$l_1 = 4$	-37.0547	-40.8191	-44.9404	-48.9230	-52.9080
$l_1 = 5$	-41.0371	-44.8028	-48.8718	-52.9715	-56.9993

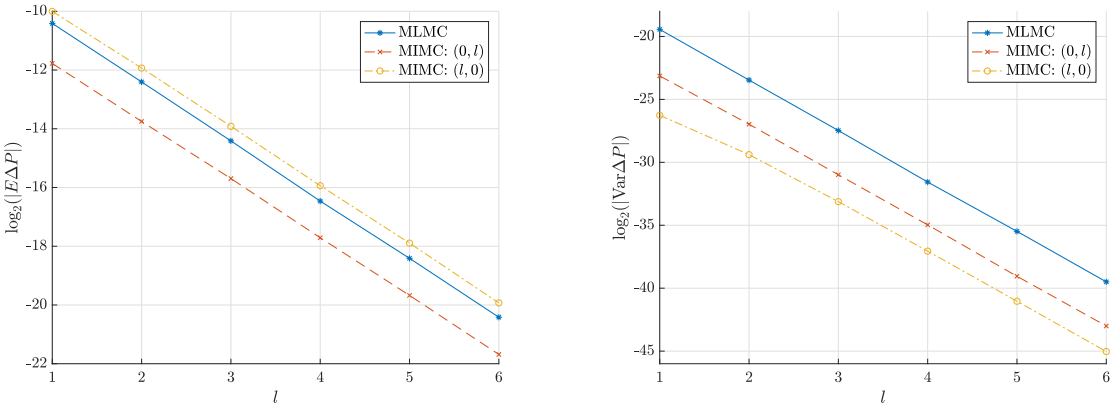
Table 2.2: $\log_2 \left(\mathrm{Var}[\Delta P_{(l_1\, l_2)}]\right)$ with $h = h_0 \cdot 2^{-l_1}$ and $k = k_0 \cdot 2^{-2l_2}$, for approximation (2.5) and (2.11).



(a) $\log_2 \left|\mathbb{E}\Delta P_{(l_1\, l_2)}\right|$ (b) $\log_2 \mathrm{Var}(\Delta P_{(l_1,\, l_2)})$

Figure 2.5: Rate verification: contour plots of log values of sample mean (left) and variance (right) of mixed differences used in MIMC.

Figure 2.6 verifies the convergence rate of MLMC and MIMC at boundary levels, i.e., $l_1 = 0$ or $l_2 = 0$. We can see from the plots that they have the same rate of convergence, which is consistent with Remark 2.3.3.



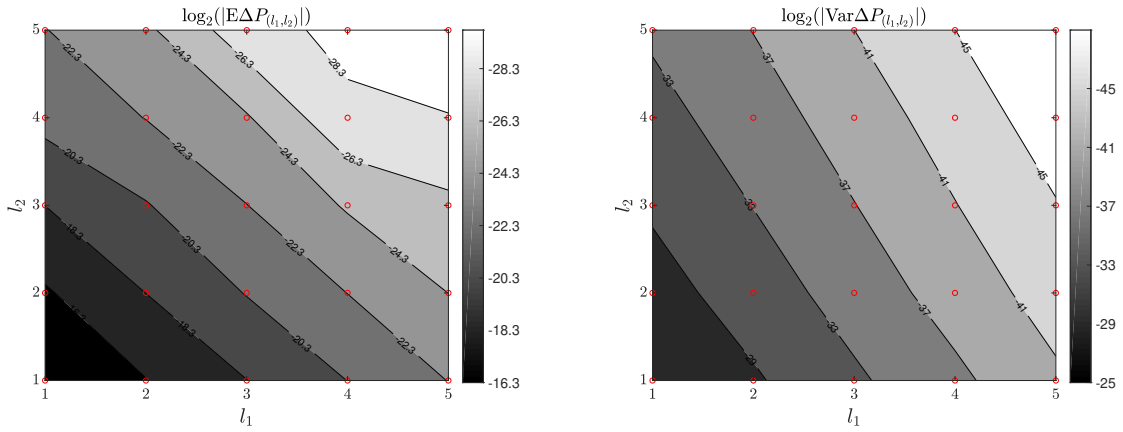
(a) $\log_2 \left|\mathbb{E}\Delta P_{(l_1\, l_2)}\right|$ (b) $\log_2 \mathrm{Var}(\Delta P_{(l_1,\, l_2)})$

Figure 2.6: Log values of mean and variance for MLMC and MIMC with $l_1 = 0$ or $l_2 = 0$.

Now we compare the results with the alternative discretisation scheme (2.6) but the same $P_{(l_1,\, l_2)}$ from (2.11). The analysis in Section 2.5.1 shows that under that scheme,

$$\mathbb{E}\left[\Delta P_{(l_1\, l_2)}\right] = O(h^2 k), \qquad \mathrm{Var}\left[\Delta P_{(l_1,\, l_2)}\right] = O(h^4 k).$$

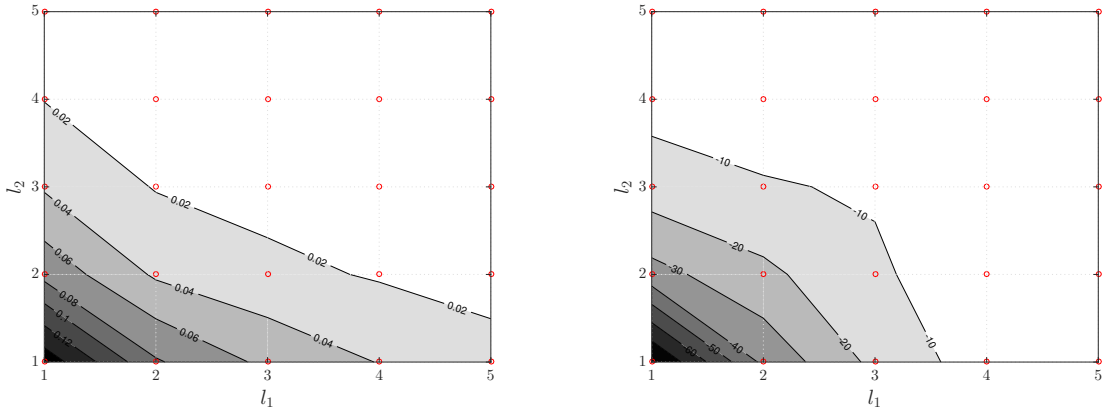
Figure 2.7 shows the contour plots of $\log_2 \left|\mathbb{E}\Delta P_{(l_1,l_2)}\right|$ and $\log_2 \left(\mathrm{Var}[\Delta P_{(l_1\, l_2)}]\right)$ using the alternative discretisation scheme (2.6) and (2.11), with $h = h_0 \cdot 2^{-l_1}$ and $k = k_0 \cdot 2^{-2l_2}$.



(a) $\log_2 \left|\mathbb{E}\Delta P_{(l_1\, l_2)}\right|$ (b) $\log_2 \mathrm{Var}\left(\Delta P_{(l_1,\, l_2)}\right)$

Figure 2.7: Log values of mean and variance by the other discretisation scheme, given by (2.6) and (2.11).

We compute the optimal index set based on the profit (2.53). Figure 2.8(a) shows the profit using scheme (2.5) and (2.11). This provides the basis for the optimal index set (2.55). For scheme (2.6) and (2.63), there are two leading orders of variance, hence the index set is not triangular, as one can deduce from Figure 2.8(b).



(a) Profit of scheme (2.5) and (2.11). (b) Profit of scheme (2.6) and (2.63).

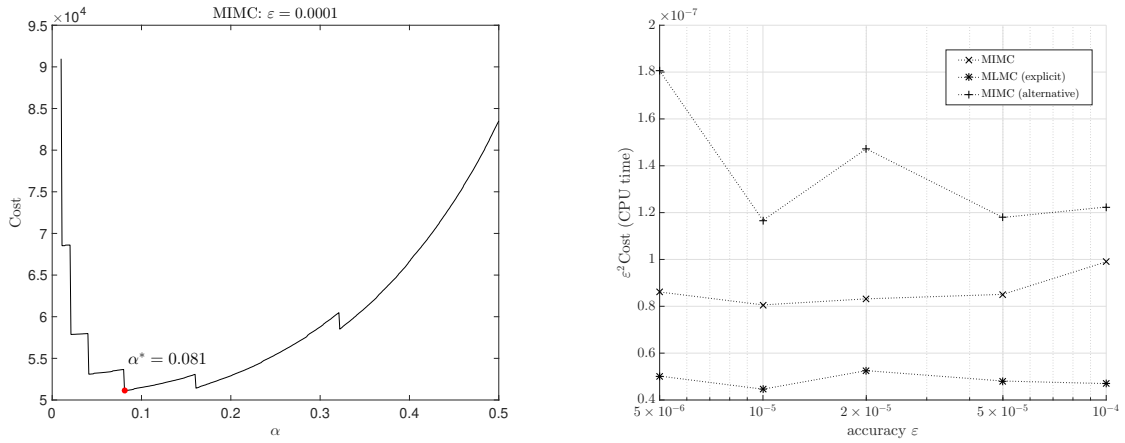
Figure 2.8: Contour plots of profit.

2.6.2 Complexity

The analysis shows that the total computational cost of the scheme (2.5) and (2.11) for a RMSE ε is $W = O\left(\varepsilon^{-2}\left|\log\varepsilon\right|\right)$. As k_0 does not matter much in this problem, we can choose $k_0 = 1/4$ fixed for $l_1^* \leq 25$, and the computational cost is expected to have the order $O\left(\varepsilon^{-2}\right)$. We test the total cost given different accuracy ε, and compare the results with the multilevel algorithm. The mean square error for the estimator can be expressed as the sum of the variance of the estimator and the square of the weak error, as in (2.20),

$$\left|\mathbb{E}\left[\widehat{P}_{\mathcal{I}} - P\right]\right| \leq \alpha\varepsilon, \qquad \mathrm{Var}\left[\widehat{P}_{\mathcal{I}}\right] \leq \left(1-\alpha^2\right)\varepsilon^2.$$

Since $V_{(l_1,l_2)}W_{(l_1,l_2)} \sim 2^{-3l_1-2l_2}$, the variance decays more rapidly with the level than the cost increases, so that the dominant cost is on level 0. For each fixed accuracy, we find the global minimum of the total cost with respect to α and choose that optimal α^*, thus reducing the total cost. Figure 2.9(a) shows how the total cost varies with α when $\varepsilon = 10^{-4}$. Figure 2.9(b) plots the CPU time of $\varepsilon^2 W$ of two multi-index and the multilevel algorithms. As k_0 does not affect much in this problem, the total computational costs of the MLMC algorithm and MIMC with schemes (2.5) and (2.11) are approximately proportional to ε^{-2}, hence $\varepsilon^2 W$ should not vary much with different accuracy ε. However, the cost of the alternative discretisation scheme (2.6) with MIMC has the order $\varepsilon^{-2}(\log\varepsilon)^2$. We can see from the figure that $\varepsilon^2 W$ is no longer a constant but is slightly increasing as $\varepsilon \to 0$.



(a) Cost versus α with scheme (2.5) and (2.11).

(b) ε^2Cost of different schemes.

Figure 2.9: Total cost versus α, and ε^2Cost with different schemes, where "alternative" denotes the scheme (2.6) and (2.11).

The costs of the MLMC and MIMC schemes are very similar across a wide range of parameters (not reported here), which is to be expected as the dominant cost comes from the coarsest level. The MLMC scheme is slightly more efficient here, because it allows explicit time-stepping with a slightly lower computational cost. When an implicit scheme (2.5) is used for MLMC as well, giving better stability properties, the MIMC scheme outperforms. This is useful for locally refined meshes.

2.7 Summary

We have analysed the accuracy and complexity of a MIMC estimator for a one-dimensional parabolic SPDE using Fourier analysis. Specifically, we analysed a functional $P_T(v)$ of the solution. We showed that, by using the implicit Milstein finite difference discretisation (2.5), the order of the first and second moments of $\Delta P_{(l_1,l_2)}$ are $O(h^2k)$ and $O(h^4k^2)$, with $h = h_0 \cdot 2^{-l_1}$, $k = k_0 \cdot 2^{-2l_2}$. For a fixed RMSE ε, we achieved the complexity $O(\varepsilon^{-2}|\log\varepsilon|)$ in (2.58). Moreover, under a different discretisation (2.6), the first and second moments of $\Delta P_{(l_1,l_2)}$ are $O(h^2k)$ and $O(h^4k)$, and the total complexity of this scheme is $O(\varepsilon^{-2}|\log\varepsilon|^3)$ from (2.62). We further used a simpler approximation of P_T as (2.63) with discretisation (2.6). This gave two leading terms of different orders in the variance, which violated the simplest form of assumptions of the framework in [54]. After some adaptation, we achieved a complexity of $O\big(\varepsilon^{-2}\,|\log\varepsilon|^3\big)$ as in (2.65).

Although MLMC achieved an optimal complexity $O(\varepsilon^{-2})$ in this case, one shortcoming of MLMC is that the efficiency still depends on the dimensionality. For high-dimensional problems, when the level increases, the decay of the variance will be slower than the increase of the cost. In that case, the total computational cost is no longer $O(\varepsilon^{-2})$. For example, consider a two-dimensional version of the SPDE,

$$\begin{aligned} \mathrm{d}v = &-\mu_x\frac{\partial v}{\partial x}\,\mathrm{d}t - \mu_y\frac{\partial v}{\partial y}\,\mathrm{d}t + \frac{1}{2}\left(\frac{\partial^2 v}{\partial x^2} + 2\sqrt{\rho_x\rho_y}\rho_{xy}\frac{\partial^2 v}{\partial x\partial y} + \frac{\partial^2 v}{\partial y^2}\right)\mathrm{d}t \\ &- \sqrt{\rho_x}\frac{\partial v}{\partial x}\,\mathrm{d}W_t^x - \sqrt{\rho_y}\frac{\partial v}{\partial y}\,\mathrm{d}W_t^y, \end{aligned}$$

with $\rho_x, \rho_y \in [0,1)$, $\rho_{xy} \in [-1,1]$, μ_x, μ_y real parameters. Assuming second order convergence in both spatial directions and first order in time, the total MLMC cost given a fixed accuracy ε is expected to be $O\left(\varepsilon^{-2}(\log\varepsilon)^2\right)$. Although this is much lower than the $O(\varepsilon^{-4})$ expected for the standard Monte Carlo method, the order is not optimal as in the one-dimensional case. Next we will analyse high-dimensional SPDEs.

Chapter 3

A two-dimensional Zakai SPDE

In this chapter, we propose an implicit finite difference scheme for a two-dimensional parabolic stochastic partial differential equation (SPDE) of Zakai type. The scheme is based on a Milstein approximation to the stochastic integral and an alternating direction implicit (ADI) discretisation of the elliptic term. We prove its mean-square stability and convergence in L_2 of first order in time and second order in space, by Fourier analysis, in the presence of Dirac initial data. Numerical tests confirm these findings empirically.

The rest of this chapter is structured as follows. We define the approximation schemes in Section 3.1. Then we analyse the mean-square stability and L_2-convergence in Sections 3.2 and 3.3 in the constant coefficient case of (1.13). Section 3.4 shows numerical experiments confirming the above findings. Section 3.4 extends the scheme to variable coefficients as in (1.3) and presents tests for the example (1.16). Section 3.6 offers summaries and directions for further research.

3.1 Approximation and main results

3.1.1 Semi-implicit Milstein finite difference scheme

First, we introduce the numerical scheme to the SPDE (1.13), repeated here for convenience,

$$\begin{aligned} \mathrm{d}v = &-\mu_x \frac{\partial v}{\partial x}\,\mathrm{d}t - \mu_y \frac{\partial v}{\partial y}\,\mathrm{d}t + \frac{1}{2}\left(\frac{\partial^2 v}{\partial x^2} + 2\sqrt{\rho_x\rho_y}\rho_{xy}\frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 v}{\partial y^2}\right)\mathrm{d}t \\ &- \sqrt{\rho_x}\frac{\partial v}{\partial x}\,\mathrm{d}W_t^x - \sqrt{\rho_y}\frac{\partial v}{\partial y}\,\mathrm{d}W_t^y, \end{aligned} \tag{3.1}$$

with Dirac initial $v(0,x,y) = \delta(x-x_0) \otimes \delta(y-y_0)$. We use a spatial grid with uniform spacing h_x, $h_y > 0$, and, for $T > 0$ fixed, N time steps of size $k = T/N$. Let $V^n_{i,j}$ be the approximation to $v(nk, ih_x, jh_y)$, $n = 1, \dots, N$, $i, j \in \mathbb{Z}$, where $i_0 \coloneqq [x_0/h_x]$, $j_0 \coloneqq [y_0/h_y]$, the closest integers to x_0/h_x and y_0/h_y. We approximate $v(0,x,y)$ by

$$V^0_{i,j} = h_x^{-1} h_y^{-1} \delta_{(i_0, j_0)} = \begin{cases} h_x^{-1} h_y^{-1}, & i = i_0,\ j = j_0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

To improve the accuracy of the approximation of v in the present case of Dirac initial data, we subsequently choose h_x and h_y such that x_0/h_x and y_0/h_y are integers and therefore x_0 and y_0 are on the grid.

Extending the implicit Euler scheme in [92] for the 1-d case, it is natural to take the drift term

$$-\mu_x \frac{\partial v}{\partial x} - \mu_y \frac{\partial v}{\partial y} + \frac{1}{2}\left(\frac{\partial^2 v}{\partial x^2} + 2\sqrt{\rho_x \rho_y}\rho_{xy} \frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 v}{\partial y^2} \right)$$

implicit, and the terms driven by W^x and W^y explicit. We will prove later that in this way we obtain better stability (compare Proposition 3.2.2 to Theorem 3.2.1). For computational simplicity, in the following, we take the mixed derivative term therein explicit. This is also in preparation for the ADI splitting schemes we will study later.

Using such a semi-implicit Euler scheme, the SPDE (3.1) can be approximated by

$$\begin{aligned} V^{n+1} = V^n &- \frac{\mu_x k}{2h_x} D_x V^{n+1} - \frac{\mu_y k}{2h_y} D_y V^{n+1} + \frac{k}{2h_x^2} D_{xx} V^{n+1} + \frac{k}{2h_y^2} D_{yy} V^{n+1} \\ &+ \sqrt{\rho_x \rho_y}\rho_{xy} \frac{k}{4h_x h_y} D_{xy} V^n - \frac{\sqrt{\rho_x k} Z_{n,x}}{2h_x} D_x V^n - \frac{\sqrt{\rho_y k}\widetilde{Z}_{n,y}}{2h_y} D_y V^n, \end{aligned}$$

where

$$\begin{aligned} (D_x V)_{i,j} &= V_{i+1,j} - V_{i-1,j}, & (D_y V)_{i,j} &= V_{i,j+1} - V_{i,j-1}, \\ (D_{xx} V)_{i,j} &= V_{i+1,j} - 2V_{i,j} + V_{i-1,j}, & (D_{yy} V)_{i,j} &= V_{i,j+1} - 2V_{i,j} + V_{i,j-1}, \\ (D_{xy} V)_{i,j} &= V_{i+1,j+1} - V_{i-1,j+1} - V_{i+1,j-1} + V_{i-1,j-1}, \end{aligned}$$

and $\widetilde{Z}^y_n = \rho_{xy} Z^x_n + \sqrt{1-\rho^2_{xy}} Z^y_n$, with $Z^x_n, Z^y_n \sim N(0,1)$ being independent normal random variables. To achieve a higher order of convergence, we introduce the Milstein scheme. Integrating (3.1) over the time interval $[nk, (n+1)k]$,

$$\begin{aligned} &v(nk+k, x, y) \\ &= v(nk, x, y) + \int_{nk}^{nk+k} \left(-\mu_x \frac{\partial v}{\partial x} - \mu_y \frac{\partial v}{\partial y} + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} + \sqrt{\rho_x \rho_y}\rho_{xy} \frac{\partial^2 v}{\partial x \partial y} + \frac{1}{2}\frac{\partial^2 v}{\partial y^2} \right) \mathrm{d}s \\ &\quad - \int_{nk}^{nk+k} \sqrt{\rho_x} \frac{\partial v}{\partial x}\, \mathrm{d}W^x_s - \int_{nk}^{nk+k} \sqrt{\rho_y} \frac{\partial v}{\partial y}\, \mathrm{d}W^y_s. \end{aligned}$$

In the Euler scheme, we approximate all integrands by their value at time nk or $(n+1)k$, which is a zero-order expansion in time. By contrast, in the Milstein scheme, we use a first-order expansion for the stochastic integrals, such that we make the approximation $v(s,x,y) \approx v(nk,x,y)$ for $nk < s < (n+1)k$ in the first integral and

$$v(s,x,y) \approx v(nk,x,y) - \sqrt{\rho_x}\frac{\partial v}{\partial x}(nk,x,y)(W_s^x - W_{nk}^x) - \sqrt{\rho_y}\frac{\partial v}{\partial y}(nk,x,y)(W_s^y - W_{nk}^y)$$

in the second and third. We denote nk as t, and it follows

$$\begin{aligned}
&- \int_t^{t+k} \sqrt{\rho_x}\frac{\partial v}{\partial x}(s,x,y)\,\mathrm{d}W_s^x - \int_t^{t+k} \sqrt{\rho_y}\frac{\partial v}{\partial y}(s,x,y)\,\mathrm{d}W_s^y \\
\approx\ & -\sqrt{\rho_x}\frac{\partial v}{\partial x}(t,x,y)\Delta W_n^x - \sqrt{\rho_y}\frac{\partial v}{\partial y}(t,x,y)\Delta W_n^y \\
&+\ \rho_x\frac{\partial^2 v}{\partial x^2}(t,x,y)\int_t^{t+k}(W_s^x - W_t^x)\,\mathrm{d}W_s^x\ + \rho_y\frac{\partial^2 v}{\partial y^2}(t,x,y)\int_t^{t+k}(W_s^y - W_t^y)\,\mathrm{d}W_s^y \\
&+\ \sqrt{\rho_x\rho_y}\frac{\partial^2 v}{\partial x \partial y}(t,x,y)\bigg(\int_t^{t+k}(W_s^x - W_t^x)\,\mathrm{d}W_s^y + \int_t^{t+k}(W_s^y - W_t^y)\,\mathrm{d}W_s^x\bigg),
\end{aligned}$$

where

$$\Delta W_n^x = W_{t+k}^x - W_t^x = \sqrt{k}Z_n^x, \quad \Delta W_n^y = W_{t+k}^y - W_t^y = \sqrt{k}\widetilde{Z}_n^y.$$

From standard Itô calculus, we have

$$\int_t^{t+k}(W_s^x - W_t^x)\,\mathrm{d}W_s^x = \frac{1}{2}\Big((\Delta W_n^x)^2 - k\Big), \qquad \int_t^{t+k}(W_s^y - W_t^y)\,\mathrm{d}W_s^y = \frac{1}{2}\Big((\Delta W_n^y)^2 - k\Big),$$
$$\int_t^{t+k}(W_s^x - W_t^x)\,\mathrm{d}W_s^y + \int_t^{t+k}(W_s^y - W_t^y)\,\mathrm{d}W_s^x = \Delta W_n^x \Delta W_n^y - \rho_{xy}k.$$

We see that the mixed-derivative terms cancel, and we derive the implicit Milstein scheme as follows,

$$\begin{aligned}
&\bigg(I + \frac{\mu_x k}{2h_x}D_x + \frac{\mu_y k}{2h_y}D_y - \frac{k}{2h_x^2}D_{xx} - \frac{k}{2h_y^2}D_{yy}\bigg)V^{n+1} \\
&\quad = \bigg(I - \frac{\sqrt{\rho_x k}Z_{n,x}}{2h_x}D_x - \frac{\sqrt{\rho_y k}\widetilde{Z}_{n,y}}{2h_y}D_y + \frac{\rho_x k(Z_{n,x}^2 - 1)}{8h_x^2}D_x^2 \\
&\qquad + \frac{\rho_y k(\widetilde{Z}_{n,y}^2 - 1)}{8h_y^2}D_y^2 + \frac{\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}}{4h_x h_y}D_{xy}\bigg)V^n.
\end{aligned} \tag{3.3}$$

To facilitate its implementation, we combine the scheme with an Alternating Direction Implicit (ADI) factorisation, which has been introduced in [88] for parabolic PDEs to approximately factorise the system matrix by matrices which correspond to derivatives in individual directions and which can thus more easily be inverted, while the consistency

order is maintained. Applying this principle to the implicit terms on the left-hand side of (3.3),

$$\begin{aligned}
&\left(I + \frac{\mu_x k}{2h_x} D_x + \frac{\mu_y k}{2h_y} D_y - \frac{k}{2h_x^2} D_{xx} - \frac{k}{2h_y^2} D_{yy}\right) V \\
&\quad = \left(I + \frac{\mu_x k}{2h_x} D_x - \frac{k}{2h_x^2} D_{xx}\right)\left(I + \frac{\mu_y k}{2h_y} D_y - \frac{k}{2h_y^2} D_{yy}\right) V \\
&\qquad - k^2 \left(\frac{\mu_x}{2h_x} D_x - \frac{1}{2h_x^2} D_{xx}\right)\left(\frac{\mu_y}{2h_y} D_y - \frac{1}{2h_y^2} D_{yy}\right) V.
\end{aligned}$$

Since the second term on right-hand side proportional to k^2 is of higher order than the overall scheme (which is of first order), it is justifiable in terms of the consistency order to neglect it. This yields the ADI scheme as follows,

$$\begin{aligned}
&\left(I + \frac{\mu_x k}{2h_x} D_x - \frac{k}{2h_x^2} D_{xx}\right)\left(I + \frac{\mu_y k}{2h_y} D_y - \frac{k}{2h_y^2} D_{yy}\right) V^{n+1} \\
&\quad = \left(I - \frac{\sqrt{\rho_x k} Z_{n,x}}{2h_x} D_x - \frac{\sqrt{\rho_y k}\widetilde{Z}_{n,y}}{2h_y} D_y + \frac{\rho_x k (Z_{n,x}^2 - 1)}{8h_x^2} D_x^2 \right. \\
&\qquad \left. + \frac{\rho_y k (\widetilde{Z}_{n,y}^2 - 1)}{8h_y^2} D_y^2 + \frac{\sqrt{\rho_x \rho_y} k Z_{n,x} \widetilde{Z}_{n,y}}{4h_x h_y} D_{xy}\right) V^n.
\end{aligned} \tag{3.4}$$

Note that there is no substantial benefit in second order accurate splitting schemes (such as Craig–Sneyd [19] or Hundsdorfer–Verwer [63]) as the overall order is limited to 1 by the Milstein approximation to the stochastic integral. As an alternative to ADI scheme, which effectively approximates the solution to certain sparse non-tridiagonal linear systems resulting from multidimensional PDEs by the solution to a sequence of tridiagonal linear systems, other methods which solve the original systems of equations directly are widely applied to PDEs and might also work efficiently for this problem. For example, the algebraic multigrid technique, introduced in [7, 95], is an efficient tool for the solution of symmetric, positive definite linear algebraic systems, with $O(N)$ complexity.

We approximate the second derivative on the right-hand side with D_x^2 and D_y^2, but the results for D_{xx} and D_{yy} would be similar.

We can also use the explicit Milstein finite difference scheme to approximate the SPDE (3.1)

$$\begin{aligned}
V^{n+1} = &\left(I - \frac{\mu_x k + \sqrt{\rho_x k} Z_{n,x}}{2h_x} D_x - \frac{\mu_y k + \sqrt{\rho_y k}\widetilde{Z}_{n,y}}{2h_y} D_y + \frac{k}{2h_x^2} D_{xx} + \frac{k}{2h_y^2} D_{yy} \right. \\
&\left. + \frac{\rho_x k (Z_{n,x}^2 - 1)}{8h_x^2} D_x^2 + \frac{\rho_y k (\widetilde{Z}_{n,y}^2 - 1)}{8h_y^2} D_y^2 + \frac{\sqrt{\rho_x \rho_y} k Z_{n,x} \widetilde{Z}_{n,y}}{4h_x h_y} D_{xy}\right) V^n,
\end{aligned} \tag{3.5}$$

but, as we will see, this scheme is stable only under a restrictive condition on the timestep.

3.1.2 Main convergence results

The following theorems describe the mean-square stability and convergence of the implicit finite difference scheme (3.3) and the ADI scheme (3.4). We make the following assumption:

Assumption 3.1.1. *Let* $0 \leq \rho_x, \rho_y < 1$, $-1 \leq \rho_{xy} \leq 1$ *such that*

$$2\rho_x^2(1 + 2|\rho_{xy}|) < 1, \tag{3.6a}$$

$$2\rho_y^2(1 + 2|\rho_{xy}|) < 1, \tag{3.6b}$$

$$2\rho_x\rho_y(3\rho_{xy}^2 + 2|\rho_{xy}| + 1) < 1. \tag{3.6c}$$

In Section 3.2 we show that Assumption 3.1.1 is a sufficient condition for stability of the schemes (3.3) and (3.4).[1] If $\rho_{xy} = 0$, these conditions reduce to $2\rho_x^2 \leq 1$ and $2\rho_y^2 \leq 1$, which is analogous to the condition for mean-square stability in the 1-dimensional case in [92]. In the worst case, $|\rho_{xy}| = 1$, sufficient conditions are $\rho_x, \rho_y \leq 1/\sqrt{6}$, and $\rho_x\rho_y \leq 1/12$.

First, we recall the setting of our numerical schemes. For $T > 0$ fixed, we discretise $[0, T]$ by N steps, and let $k = T/N$ be the timestep. We discretise space $\mathbb{R}^2$ with mesh sizes h_x and h_y.

The following theorem shows the convergence of the implicit Milstein difference scheme (3.3). The constant $\theta \in (0, 1)$ therein is determined by the parameters ρ_x, ρ_y and ρ_{xy}. The proof of Lemma 3.3.5 gives an explicit value; though not sharp, this is sufficient to highlight the divergence for $h_x, h_y \to 0$ when k is fixed.

Theorem 3.1.1. *Let* $T > 0$, $k = T/N$, $h_x > 0$ *and* $h_y > 0$ *be mesh sizes. Then, under Assumption 3.1.1, there exists* $\theta \in (0, 1)$, *independent of* h_x, h_y *and* k, *such that the implicit Milstein finite difference scheme* (3.3) *has the error expansion*

$$\begin{aligned} V_{i,j}^N - v(T, x_i, y_j) = k\, E_1(T, x_i, y_j) + h_x^2\, E_2(T, x_i, y_j) + h_y^2\, E_3(T, x_i, y_j) \\ + \theta^N h_x^{-2}\, E_4(T, x_i, y_j) + \theta^N h_y^{-2}\, E_5(T, x_i, y_j) \\ + o(k, h_x^2, h_y^2, \theta^N h_x^{-2}, \theta^N h_y^{-2})\, R(T, x_i, y_j), \end{aligned} \tag{3.7}$$

where $x_i = ih_x$, $y_j = jh_y$, $E_1, \ldots, E_5$, *and* R *are random variables with bounded first and second moments, all independent of* h_x, h_y *and* k.

[1] We do not believe it to be necessary, due to estimates made in the derivation and as evidenced by numerical tests.

Proof. See Section 3.3. □

Remark 3.1.1. *In the setting of Theorem 3.1.1 with* $\widehat{h} = \min\{h_x, h_y\}$*, if*

$$\theta^{\frac{T}{k}} \leq 2^{-C_0} \cdot \widehat{h}^{4+\beta},$$

for some $\beta, C_0 > 0$ *independent of* h_x*,* h_y *and* k*, or, equivalently,*

$$k \leq \frac{T \log_2(\theta^{-1})}{C_0 + (4+\beta)\log_2(\widehat{h}^{-1})}, \tag{3.8}$$

then the implicit Milstein scheme (3.3) *has the error expansion*

$$V_{i,j}^N - v(T, x_i, y_j) = k\, E_1(T, x_i, y_j) + h_x^2\, E_2(T, x_i, y_j) + h_y^2\, E_3(T, x_i, y_j) + o(k, h_x^2, h_y^2)\, R(T, x_i, y_j).$$

Corollary 3.1.1. *Under the conditions of Theorem 3.1.1 and Remark 3.1.1, the error of the implicit Milstein scheme* (3.3) *at time* T *satisfies, for all* $i, j \in \mathbb{Z}$*,*

$$\sqrt{\mathbb{E}\big[|V_{i,j}^N - v(T, ih_x, jh_y)|^2\big]} = O(h_x^2) + O(h_y^2) + O(k). \tag{3.9}$$

For the ADI discretisation scheme (3.4), a similar convergence result holds.

Theorem 3.1.2. *Under the conditions of Remark 3.1.1, the error of the ADI scheme* (3.4) *has the same order as for the implicit Milstein scheme,*

$$V_{i,j}^N - v(T, x_i, y_j) = k\, E_1(T, x_i, y_j) + h_x^2\, E_2(T, x_i, y_j) + h_y^2\, E_3(T, x_i, y_j) + o(k, h_x^2, h_y^2)\, R(T, x_i, y_j),$$

where E_1*,* E_2*,* E_3 *and* R *are random variables with bounded first and second moments.*

Proof. See Section 3.3. □

Theorem 3.1.1 and Theorem 3.1.2 state the convergence pointwise in space and L_2 in probability. The discrete L_2 error in space and (continuous) in probability is defined as

$$\sqrt{\mathbb{E}\left[\, \left| \sqrt{\sum_{i,j} \left[\left| V_{i,j}^N - v(T, ih_x, jh_y) \right|^2 \right] h_x h_y} \, \right|^2 \,\right]} = \sqrt{\sum_{i,j} \mathbb{E}\left[\left| V_{i,j}^N - v(T, ih_x, jh_y) \right|^2 \right] h_x h_y}, \tag{3.10}$$

where the term within the outer $|\cdot|$ is the discrete spatial L_2 norm. Applying Parseval's theorem to (3.10), we get:

Corollary 3.1.2. *Under the conditions of Theorem 3.1.1, the discrete L_2 error in space and L_2 in probability of the implicit Milstein scheme* (3.3) *at time T, defined in* (3.10)*, satisfies,*

$$\sqrt{\sum_{i,j} \mathbb{E}\Big[\big|V_{i,j}^N - v(T, ih_x, jh_y)\big|^2\Big] h_x h_y} = O(h_x^2) + O(h_y^2) + O(k) + O(\theta^N h_x^{-1/2} h_y^{-1/2}). \quad (3.11)$$

If the initial condition lies in L_2, then

$$\sqrt{\sum_{i,j} \mathbb{E}\Big[\big|V_{i,j}^N - v(T, ih_x, jh_y)\big|^2\Big] h_x h_y} = O(h_x^2) + O(h_y^2) + O(k).$$

Proof. See Section 3.3. □

3.2 Fourier analysis of mean-square stability

Recall the SPDE (3.1),

$$\begin{aligned} \mathrm{d}v = &-\mu_x \frac{\partial v}{\partial x}\,\mathrm{d}t - \mu_y \frac{\partial v}{\partial y}\,\mathrm{d}t + \frac{1}{2}\left(\frac{\partial^2 v}{\partial x^2} + 2\sqrt{\rho_x \rho_y}\rho_{xy}\frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 v}{\partial y^2}\right)\mathrm{d}t \\ &- \sqrt{\rho_x}\frac{\partial v}{\partial x}\,\mathrm{d}W_t^x - \sqrt{\rho_y}\frac{\partial v}{\partial y}\,\mathrm{d}W_t^y, \end{aligned} \quad . \quad (3.12)$$

Define the Fourier transform pair

$$\begin{aligned} \widetilde{v}(t, \xi, \eta) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v(t, x, y)\mathrm{e}^{-\mathrm{i}\xi x - \mathrm{i}\eta y}\,\mathrm{d}x\,\mathrm{d}y, \\ v(t, x, y) &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{v}(t, \xi, \eta)\mathrm{e}^{\mathrm{i}\xi x + \mathrm{i}\eta y}\,\mathrm{d}\xi\,\mathrm{d}\eta. \end{aligned}$$

The Fourier transform of (3.12) yields

$$\mathrm{d}\widetilde{v} = -\left(\left(\mathrm{i}\mu_x\xi + \mathrm{i}\mu_y\eta + \frac{1}{2}\xi^2 + \sqrt{\rho_x\rho_y}\rho_{xy}\xi\eta + \frac{1}{2}\eta^2\right)\mathrm{d}t + \mathrm{i}\sqrt{\rho_x}\xi\,\mathrm{d}W_t^x + \mathrm{i}\sqrt{\rho_y}\eta\,\mathrm{d}W_t^y\right)\widetilde{v}, \quad (3.13)$$

subject to the initial data $\widetilde{v}(0) = \mathrm{e}^{-\mathrm{i}\xi x_0 - \mathrm{i}\eta y_0}$. For the remainder of the analysis, we take $\mu_x = \mu_y = 0$. This does not alter the results (see Remark 2.3 in [92] for the 1-d case).

The solution to (3.13) is

$$\widetilde{v}(t) = X(t)\mathrm{e}^{-\mathrm{i}\xi x_0 - \mathrm{i}\eta y_0}, \quad (3.14)$$

where

$$X(t) = \exp\left(-\frac{1}{2}(1-\rho_x)\xi^2 t - \frac{1}{2}(1-\rho_y)\eta^2 t - \mathrm{i}\xi\sqrt{\rho_x}W_t^x - \mathrm{i}\eta\sqrt{\rho_y}W_t^y\right). \quad (3.15)$$

From this we see that $\mathbb{E}\big[|X(t)|^2\big]$ goes to zero exponentially as $t \to \infty$, as well as $\mathbb{E}\big[\|\tilde{v}(t,\cdot,\cdot)\|_{L_2}^2\big]$ and, by isometry, $\mathbb{E}\big[\|v(t,\cdot,\cdot)\|_{L_2}^2\big]$. The latter holds not only for Dirac initial data, but also for initial data in L_2. We therefore say that $v = 0$ is a mean-square stable equilibrium solution.

For the numerical solution, we can use a discrete-continuous Fourier decomposition (note that we approximate $v(nk, ih_x, jh_y)$ by $V_{i,j}^n$)

$$V_{i,j}^0 = \frac{1}{4\pi^2 h_x h_y} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \widetilde{V}^0(u,v) \mathrm{e}^{\mathrm{i}\left((i-i_0)u+(j-j_0)v\right)} \,\mathrm{d}u \,\mathrm{d}v,$$

where $i_0 = x_0/h_x$, $j_0 = y_0/h_y$, and

$$\widetilde{V}^0(u,v) = h_x h_y \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} V_{i,j}^0 \mathrm{e}^{\mathrm{i}\left(-(i-i_0)u-(j-j_0)v\right)}.$$

From (3.2), $V_{i,j}^0 = h_x^{-1} h_y^{-1} \delta_{(i_0,j_0)}$, we have $\widetilde{V}^0(u,v) = 1$ for all $(u,v) \in \mathbb{R}^2$. Similarly for n-th time-step,

$$\begin{aligned} V_{i,j}^n &= \frac{1}{4\pi^2 h_x h_y} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \widetilde{V}^n(u,v) \mathrm{e}^{\mathrm{i}\left((i-i_0)u+(j-j_0)v\right)} \,\mathrm{d}u \,\mathrm{d}v \\ &= \frac{1}{4\pi^2} \int_{-\frac{\pi}{h_y}}^{\frac{\pi}{h_y}} \int_{-\frac{\pi}{h_x}}^{\frac{\pi}{h_x}} \widetilde{V}^n(\xi,\eta) \mathrm{e}^{\mathrm{i}\left((i-i_0)\xi h_x+(j-j_0)\eta h_y\right)} \,\mathrm{d}\xi \,\mathrm{d}\eta. \end{aligned} \tag{3.16}$$

In the last step, we integrate by substitution, $\xi = u/h_x,\ \eta = v/h_y$.

By analogy with the theoretical solution $\widetilde{v}(t) = X(t)\widetilde{v}(0)$, we make the ansatz

$$\widetilde{V}^n(\xi,\eta) = X_n(\xi,\eta)\widetilde{V}^0(\xi,\eta), \tag{3.17}$$

but as $\widetilde{V}^0(\xi,\eta) = 1$ we simply have $\widetilde{V}^n(\xi,\eta) = X_n(\xi,\eta)$. We can regard $X_n(\xi,\eta)$ as the numerical approximation to $X(nk)$ in (3.15).

We say that the equilibrium solution zero of the scheme is asymptotically mean-square stable, provided for any $(\xi,\eta) \in [-\pi/h_x, \pi/h_x] \times [-\pi/h_y, \pi/h_y]$,

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n(\xi,\eta)|^2\right] = 0. \tag{3.18}$$

This concept has been defined in the context of systems of SDEs in Definition 2.2, 2.3, in [8], and we apply it here to a fixed wave number in the Fourier domain. A generalisation to SPDEs is analysed in [81] (see Definition 2.1 therein). We will show convergence in L_2 (for fixed T) directly under the same conditions; see [92] for mean-square stability and convergence of a 1-d parabolic SPDE. If (3.18) holds without any restriction between h_x, h_y and k, we call it unconditionally stable. This leads to three conditions summarised in Assumption 3.1.1, as shown by the following.

Theorem 3.2.1. *The implicit Milstein finite difference scheme* (3.3) *is unconditionally stable in the mean-square sense of (3.18) provided Assumption 3.1.1 holds.*

Proof. By inserting (3.16) and (3.17) in (3.3), we have

$$\begin{aligned} X_{n+1}(\xi,\eta) = \frac{1}{1-(a_x+a_y)k}\Big(1 - \mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} \\ + b_x\rho_x k(Z_{n,x}^2-1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\Big)X_n(\xi,\eta), \end{aligned} \tag{3.19}$$

where

$$a_x = -\frac{2\sin^2\frac{\xi h_x}{2}}{h_x^2}, \quad b_x = -\frac{\sin^2\xi h_x}{2h_x^2}, \quad c_x = \frac{\sin\xi h_x}{h_x}, \quad d = -\frac{\sin\xi h_x\sin\eta h_y}{h_x h_y}, \tag{3.20a}$$

$$a_y = -\frac{2\sin^2\frac{\eta h_y}{2}}{h_y^2}, \quad b_y = -\frac{\sin^2\eta h_y}{2h_y^2}, \quad c_y = \frac{\sin\eta h_y}{h_y}. \tag{3.20b}$$

Given the time-homogeneity of (3.19) and noting that the bracketed term in (3.19) and X_n are independent, $\mathbb{E}|X_{n+1}|^2/\mathbb{E}|X_n|^2$ is independent of n. Hence, to ensure mean-square stability, it is necessary and sufficient that for any (ξ,η) we have

$$\mathbb{E}\left|\frac{1-\mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x}-\mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y}+b_x\rho_x k(Z_{n,x}^2-1)+b_y\rho_y k(\widetilde{Z}_{n,y}^2-1)+d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}}{1-k(a_x+a_y)}\right|^2 < 1.$$

This is equivalent to

$$\begin{aligned} \mathbb{E}\Big[\big(1 + b_x\rho_x k(Z_{n,x}^2-1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\big)^2 \\ + \big(c_x\sqrt{\rho_x k}Z_{n,x} + c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y}\big)^2\Big] < \big(1-k(a_x+a_y)\big)^2. \end{aligned} \tag{3.21}$$

Note that $\widetilde{Z}_n^y = \rho_{xy}Z_n^x + \sqrt{1-\rho_{xy}^2}Z_n^y$, with $Z_n^x, Z_n^y \sim N(0,1)$ being independent normal random variables, hence

$$\mathbb{E}\big[Z_{n,x}^2\widetilde{Z}_{n,y}^2\big] = 1+2\rho_{xy}^2, \quad \mathbb{E}\big[Z_{n,x}\widetilde{Z}_{n,y}\big] = \rho_{xy}, \quad \mathbb{E}\big[Z_{n,x}^3\widetilde{Z}_{n,y}\big] = \mathbb{E}\big[Z_{n,x}\widetilde{Z}_{n,y}^3\big] = 3\rho_{xy}\ ,$$

and

$$\begin{aligned}
&\mathbb{E}\left[\left(1+b_x\rho_x k(Z_{n,x}^2-1)+b_y\rho_y k(\widetilde{Z}_{n,y}^2-1)+d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\right)^2\right]\\
=&1+b_x^2\rho_x^2k^2\,\mathbb{E}\big[(Z_{n,x}^2-1)^2\big]+b_y^2\rho_y^2k^2\,\mathbb{E}\big[(\widetilde{Z}_{n,y}^2-1)^2\big]+d^2\rho_x\rho_yk^2\,\mathbb{E}\big[Z_{n,x}^2\widetilde{Z}_{n,y}^2\big]\\
&+2b_x\rho_xk\,\mathbb{E}\big[Z_{n,x}^2-1\big]+2b_y\rho_yk\,\mathbb{E}\big[\widetilde{Z}_{n,y}^2-1\big]+2d\sqrt{\rho_x\rho_y}k\,\mathbb{E}\big[Z_{n,x}\widetilde{Z}_{n,y}\big]\\
&+2b_xb_y\rho_x\rho_yk^2\,\mathbb{E}\big[(Z_{n,x}^2-1)(\widetilde{Z}_{n,y}^2-1)\big]+2b_xd\rho_x\sqrt{\rho_x\rho_y}k^2\,\mathbb{E}\big[(Z_{n,x}^2-1)Z_{n,x}\widetilde{Z}_{n,y}\big]\\
&+2b_yd\rho_y\sqrt{\rho_x\rho_y}k^2\,\mathbb{E}\big[(\widetilde{Z}_{n,y}^2-1)Z_{n,x}\widetilde{Z}_{n,y}\big]\\
=&1+2b_x^2\rho_x^2k^2+2b_y^2\rho_y^2k^2+d^2\rho_x\rho_y(1+2\rho_{xy}^2)k^2+2d\sqrt{\rho_x\rho_y}\rho_{xy}k+4b_xb_y\rho_x\rho_y\rho_{xy}^2k^2\\
&+4b_xd\rho_x\sqrt{\rho_x\rho_y}\rho_{xy}k^2+4b_yd\rho_y\sqrt{\rho_x\rho_y}\rho_{xy}k^2,
\end{aligned}$$

and

$$\mathbb{E}\left[\left(c_x\sqrt{\rho_xk}Z_{n,x}+c_y\sqrt{\rho_yk}\widetilde{Z}_{n,y}\right)^2\right]=c_x^2\rho_xk+c_y^2\rho_yk+2c_xc_y\sqrt{\rho_x\rho_y}\rho_{xy}k.$$

Note that $c_xc_y+d=0,\ 4b_xb_y=d^2$, and

$$\begin{aligned}
&b_x=\cos^2\frac{\xi h_x}{2}a_x,\quad b_y=\cos^2\frac{\eta h_y}{2}a_y,\quad c_x^2=-2\cos^2\frac{\xi h_x}{2}a_x,\\
&c_y^2=-2\cos^2\frac{\eta h_y}{2}a_y,\quad d^2=4\cos^2\frac{\xi h_x}{2}\cos^2\frac{\eta h_y}{2}a_xa_y.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{E}\Big[\Big(1+b_x\rho_x k(Z_{n,x}^2-1)+b_y\rho_y k(\widetilde{Z}_{n,y}^2-1)+d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\Big)^2\\
&\qquad+\Big(c_x\sqrt{\rho_x k}Z_{n,x}+c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y}\Big)^2\Big]\\
=&1+2b_x^2\rho_x^2k^2+2b_y^2\rho_y^2k^2+d^2\rho_x\rho_y(1+2\rho_{xy}^2)k^2+4b_xb_y\rho_x\rho_y\rho_{xy}^2k^2\\
&+4b_xd\rho_x\sqrt{\rho_x\rho_y}\rho_{xy}k^2+4b_yd\rho_y\sqrt{\rho_x\rho_y}\rho_{xy}k^2+c_x^2\rho_xk+c_y^2\rho_yk\\
=&1+2\cos^4\frac{\xi h_x}{2}a_x^2\rho_x^2k^2+2\cos^4\frac{\eta h_y}{2}a_y^2\rho_y^2k^2+4\cos^2\frac{\xi h_x}{2}\cos^2\frac{\eta h_y}{2}\rho_x\rho_y(1+3\rho_{xy}^2)a_xa_yk^2\\
&+4d\sqrt{\rho_x\rho_y}\rho_{xy}k^2\big(\cos^2\frac{\xi h_x}{2}a_x\rho_x+\cos^2\frac{\eta h_y}{2}a_y\rho_y\big)-2\cos^2\frac{\xi h_x}{2}a_x\rho_xk-2\cos^2\frac{\eta h_y}{2}a_y\rho_yk\\
\le&1+2\cos^4\frac{\xi h_x}{2}a_x^2\rho_x^2k^2+2\cos^4\frac{\eta h_y}{2}a_y^2\rho_y^2k^2+4\cos^2\frac{\xi h_x}{2}\cos^2\frac{\eta h_y}{2}\rho_x\rho_y(1+3\rho_{xy}^2)a_xa_yk^2\\
&+4|\rho_{xy}|k^2\big(\cos^2\frac{\xi h_x}{2}a_x\rho_x+\cos^2\frac{\eta h_y}{2}a_y\rho_y\big)^2-2\cos^2\frac{\xi h_x}{2}a_x\rho_xk-2\cos^2\frac{\eta h_y}{2}a_y\rho_yk\\
=&1+2\cos^4\frac{\xi h_x}{2}\rho_x^2(1+2|\rho_{xy}|)a_x^2k^2+2\cos^4\frac{\eta h_y}{2}\rho_y^2(1+2|\rho_{xy}|)a_y^2k^2\\
&+4\cos^2\frac{\xi h_x}{2}\cos^2\frac{\eta h_y}{2}\rho_x\rho_y(3\rho_{xy}^2+2|\rho_{xy}|+1)a_xa_yk^2-2\cos^2\frac{\xi h_x}{2}a_x\rho_xk-2\cos^2\frac{\eta h_y}{2}a_y\rho_yk\\
\le&1+2\rho_x^2(1+2|\rho_{xy}|)a_x^2k^2+2\rho_y^2(1+2|\rho_{xy}|)a_y^2k^2+4\rho_x\rho_y(3\rho_{xy}^2+2|\rho_{xy}|+1)a_xa_yk^2\\
&-2\cos^2\frac{\xi h_x}{2}a_x\rho_xk-2\cos^2\frac{\eta h_y}{2}a_y\rho_yk,
\end{aligned}$$

and $(1-k(a_x+a_y))^2=1+k^2a_x^2-2ka_x+k^2a_y^2-2ka_y+2k^2a_xa_y$.

One sufficient condition for (3.21) to hold is

$$\begin{aligned}
2\rho_x^2(1+2|\rho_{xy}|)a_x^2k^2-2\cos^2\frac{\xi h_x}{2}a_x\rho_xk<k^2a_x^2-2ka_x,\\
2\rho_y^2(1+2|\rho_{xy}|)a_y^2k^2-2\cos^2\frac{\eta h_y}{2}a_y\rho_yk<k^2a_y^2-2ka_y,\\
4\rho_x\rho_y(3\rho_{xy}^2+2|\rho_{xy}|+1)a_xa_yk^2\le 2k^2a_xa_y.
\end{aligned}$$

Replacing a_x, a_y with their expressions in (3.20), the above is equivalent to

$$\begin{aligned}
\frac{k}{h_x^2}\Big(2\rho_x^2(1+2|\rho_{xy}|)-1\Big)\sin^2\frac{\xi h_x}{2}+\rho_x\cos^2\frac{\xi h_x}{2}<1,\\
\frac{k}{h_y^2}\Big(2\rho_y^2(1+2|\rho_{xy}|)-1\Big)\sin^2\frac{\eta h_y}{2}+\rho_y\cos^2\frac{\eta h_y}{2}<1,\\
2\rho_x\rho_y(3\rho_{xy}^2+2|\rho_{xy}|+1)\le 1.
\end{aligned}$$

A sufficient condition for this is that $\rho_x, \rho_y, \rho_{xy}$ satisfy (3.6).

□

Now we prove the stability of the ADI scheme.

Proposition 3.2.1. *Under Assumption 3.1.1, mean-square stability (3.18) also holds for the ADI scheme* (3.4).

Proof. By insertion in (3.4), we have

$$\begin{aligned} X_{n+1}(\xi,\eta) = \frac{1}{(1-a_x k)(1-a_y k)}\Big(1 - \mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} \\ + b_x\rho_x k(Z_{n,x}^2-1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\Big)X_n(\xi,\eta). \end{aligned} \tag{3.22}$$

As in the proof of Theorem 3.2.1, we need $\mathbb{E}|X_{n+1}|^2/\mathbb{E}|X_n|^2 < 1$ to achieve stability. Since $|1-(a_x+a_y)k| \le |(1-a_x k)(1-a_y k)|$ for $a_x, a_y \le 0$, the stability also holds for the ADI scheme. □

Proposition 3.2.2. *The explicit Milstein (finite difference) scheme* (3.5) *is stable in the mean-square sense provided*

$$\frac{k}{h_x^2} \le \big(2+2\rho_x^2+2\rho_x\rho_y+\big(3\rho_x+\rho_y+4\rho_x^2+4\rho_x\rho_y\big)|\rho_{xy}|+6\rho_x\rho_y\rho_{xy}^2\big)^{-1}, \tag{3.23a}$$

$$\frac{k}{h_y^2} \le \big(2+2\rho_y^2+2\rho_x\rho_y+\big(\rho_x+3\rho_y+4\rho_y^2+4\rho_x\rho_y\big)|\rho_{xy}|+6\rho_x\rho_y\rho_{xy}^2\big)^{-1}. \tag{3.23b}$$

Proof. By insertion in (3.5), we have

$$\begin{aligned} X_{n+1}(\xi,\eta) = \Big(1 + (a_x+a_y)k - \mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + b_x\rho_x k(Z_{n,x}^2-1) \\ + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\Big)X_n(\xi,\eta). \end{aligned} \tag{3.24}$$

To ensure mean-square stability in this case, similar to before, we need

$$\begin{aligned} \mathbb{E}\Big[\Big|1 + (a_x+a_y)k - \mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} \\ + b_x\rho_x k(Z_{n,x}^2-1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\Big|^2\Big] < 1. \end{aligned}$$

For simplicity, we denote $u = |\sin\frac{\xi h_x}{2}|$, $v = |\sin\frac{\eta h_y}{2}|$, then we have

$$
\begin{aligned}
&\mathbb{E}\Big[\Big|1+(a_x+a_y)k - \mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + b_x\rho_x k(Z_{n,x}^2-1)\\
&\quad + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\Big|^2\Big]\\
&= \mathbb{E}\Big[\Big(1+(a_x+a_y)k + b_x\rho_x k(Z_{n,x}^2-1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\Big)^2\\
&\quad + \Big(c_x\sqrt{\rho_x k}Z_{n,x} + c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y}\Big)^2\Big]\\
&= 1 + \big(a_x^2k^2 + 2a_xk + 2b_x^2\rho_x^2k^2 + c_x^2\rho_xk\big) + \big(a_y^2k^2 + 2a_yk + 2b_y^2\rho_y^2k^2 + c_y^2\rho_yk\big)\\
&\quad + \big(2a_xa_yk^2 + \rho_x\rho_y(1+3\rho_{xy}^2)d^2k^2\big) + 2d(a_x + 2b_x\rho_x + a_y + 2b_y\rho_y)\sqrt{\rho_x\rho_y}\rho_{xy}k^2\\
&= 1 - 4\frac{k}{h_x^2}u^2\bigg(1-\rho_x(1-u^2) - \frac{k}{h_x^2}u^2(1+2\rho_x^2(1-u^2)^2)\bigg)\\
&\quad - 4\frac{k}{h_y^2}v^2\bigg(1-\rho_y(1-v^2) - \frac{k}{h_y^2}v^2(1+2\rho_y^2(1-v^2)^2)\bigg)\\
&\quad + 8\frac{k^2}{h_x^2h_y^2}u^2v^2\bigg(1+2(1-u^2)(1-v^2)\rho_x\rho_y(1+3\rho_{xy}^2)\bigg)\\
&\quad - 4d\sqrt{\rho_x\rho_y}\rho_{xy}k\bigg(\frac{k}{h_x^2}u^2(1+2(1-u^2)\rho_x) + \frac{k}{h_y^2}v^2(1+2(1-v^2)\rho_y)\bigg)\\
&\le 1 - 4\frac{k}{h_x^2}u^2\bigg(1-\rho_x(1-u^2) - \frac{k}{h_x^2}u^2(1+2\rho_x^2)\bigg)\\
&\quad - 4\frac{k}{h_y^2}v^2\bigg(1-\rho_y(1-v^2) - \frac{k}{h_y^2}v^2(1+2\rho_y^2)\bigg)\\
&\quad + 8\frac{k^2}{h_x^2h_y^2}u^2v^2\bigg(1+2\rho_x\rho_y(1+3\rho_{xy}^2)\bigg)\\
&\quad + 8|\rho_{xy}|\bigg(\frac{k}{h_x^2}u^2\rho_x + \frac{k}{h_y^2}v^2\rho_y\bigg)\bigg(\frac{k}{h_x^2}u^2(1+2\rho_x) + \frac{k}{h_y^2}v^2(1+2\rho_y)\bigg)\\
&= 1 - 4\frac{k}{h_x^2}u^2\bigg[1-\rho_x(1-u^2) - \frac{k}{h_x^2}u^2\Big(1+2\rho_x^2+2|\rho_{xy}|\big(\rho_x+2\rho_x^2\big)\Big)\bigg]\\
&\quad - 4\frac{k}{h_y^2}v^2\bigg[1-\rho_y(1-v^2) - \frac{k}{h_y^2}v^2\Big(1+2\rho_y^2+2|\rho_{xy}|(\rho_y+2\rho_y^2)\Big)\bigg]\\
&\quad + 8\frac{k^2}{h_x^2h_y^2}u^2v^2\Big[1+2\rho_x\rho_y(1+3\rho_{xy}^2) + |\rho_{xy}|\big(\rho_x+\rho_y+4\rho_x\rho_y\big)\Big]\\
&\le 1 - 4\frac{k}{h_x^2}u^2\bigg[1-\rho_x(1-u^2) - \frac{k}{h_x^2}u^2\Big(1+2\rho_x^2+2\rho_{xy}\big(\rho_x+\rho_x^2\big)\Big)\bigg]\\
&\quad - 4\frac{k}{h_y^2}v^2\bigg[1-\rho_y(1-v^2) - \frac{k}{h_y^2}v^2\Big(1+2\rho_y^2+2\rho_{xy}(\rho_y+\rho_y^2)\Big)\bigg]\\
&\quad + 4\bigg(\frac{k^2}{h_x^4}u^4 + \frac{k^2}{h_y^4}v^4\bigg)\Big[1+2\rho_x\rho_y(1+3\rho_{xy}^2) + |\rho_{xy}|\big(\rho_x+\rho_y+2\rho_x\rho_y\big)\Big]
\end{aligned}
$$

$$\leq 1 - 4\frac{k}{h_x^2}u^2\Bigg[1 - \rho_x(1-u^2) - \frac{k}{h_x^2}u^2\Bigg(2 + 2\rho_x^2 + 2\rho_x\rho_y + \big(3\rho_x + \rho_y + 4\rho_x^2 + 4\rho_x\rho_y\big)|\rho_{xy}| + 6\rho_x\rho_y\rho_{xy}^2\Bigg)\Bigg]$$
$$- 4\frac{k}{h_y^2}v^2\Bigg[1 - \rho_y(1-v^2) - \frac{k}{h_y^2}v^2\Bigg(2 + 2\rho_y^2 + 2\rho_x\rho_y + \big(\rho_x + 3\rho_y + 4\rho_y^2 + 4\rho_x\rho_y\big)|\rho_{xy}| + 6\rho_x\rho_y\rho_{xy}^2\Bigg)\Bigg]$$
$$< 1, \qquad \text{for all } 0 \leq u, v < 1.$$

This leads to the two sufficient conditions in (3.23). □

It follows that if $\rho_{xy} = 0$, the stability conditions are

$$\frac{k}{h_x^2} \leq (2 + 2\rho_x^2 + 2\rho_x\rho_y)^{-1}, \qquad \frac{k}{h_y^2} \leq (2 + 2\rho_y^2 + 2\rho_x\rho_y)^{-1}.$$

So it is sufficient that $k/h_x^2 \leq 1/6$, and $k/h_y^2 \leq 1/6$. If $|\rho_{xy}| = 1$, the worst case in (3.23), the stability conditions are

$$\frac{k}{h_x^2} \leq (2 + 3\rho_x + \rho_y + 6\rho_x^2 + 12\rho_x\rho_y)^{-1}, \quad \frac{k}{h_y^2} \leq (2 + \rho_x + 3\rho_y + 6\rho_y^2 + 12\rho_x\rho_y)^{-1}.$$

So it is sufficient to ensure $k/h_x^2 \leq 1/24$, and $k/h_y^2 \leq 1/24$.

3.3 Fourier analysis of L_2-convergence

Extending the analysis in [15] for the standard 1-d (deterministic) heat equation to our 2-d SPDE, we compare the numerical solution to the exact solution in Fourier space first by splitting the Fourier domain into two wave number regions. Assume p is a constant satisfying $0 < p < \frac{1}{4}$. Then we define the low wave number region by

$$\Omega_{\text{low}} = \big\{(\xi, \eta) : |\xi| \leq \min\{h_x^{-2p},\, k^{-p}\} \text{ and } |\eta| \leq \min\{h_y^{-2p},\, k^{-p}\}\big\},$$

and the high wave number region by

$$\Omega_{\text{high}} = \big\{(\xi, \eta) : |\xi| > \min\{h_x^{-2p},\, k^{-p}\} \text{ or } |\eta| > \min\{h_y^{-2p},\, k^{-p}\}\big\} \cap [-\pi h_x^{-1}, \pi h_x^{-1}] \times [-\pi h_y^{-1}, \pi h_y^{-1}].$$

Note that both X_n and $X(nk)$ are functions of ξ and η. The idea of the convergence proof is that X_n is a good approximation to $X(nk)$ in the low wave region, and they both damp exponentially in the high wave region.

Lemma 3.3.1. *For $(\xi, \eta) \in \Omega_{low}$, we have*

$$X_N - X(T) = X(T) \cdot \Big(h_x^2 f_1(\xi) + h_y^2 f_2(\eta) + k\, f_3(\xi, \eta) + o(k, h_x^2, h_y^2) \cdot \varphi(T, h_x\xi, h_y\eta) \Big),$$

where $f_1(\xi), f_2(\eta), f_3(\xi, \eta), \varphi(T, h_x\xi, h_y\eta)$ are random variables such that after multiplication by $X(T)$, the integral over Ω_{low} has bounded first and second moments independent of N.

Proof. See Section 3.3.1. □

Lemma 3.3.2. *Under Assumption 3.1.1, there exists $C > 0$ independent of h_x, h_y, and k, such that*

$$\sqrt{\mathbb{E}\Big[\Big| \iint_{\Omega_{high}} X(T, \xi, \eta) - X_N(\xi, \eta)\, \mathrm{d}\xi \mathrm{d}\eta \Big|^2 \Big]} \leq C h_x^{-2} \theta^N + C h_y^{-2} \theta^N,$$

where $0 < \theta < 1$ is independent of h_x, h_y, and k.

Proof. See Section 3.3.2. □

The following proof of Theorem 3.1.1 now deduces mean square convergence of the implicit finite difference scheme (3.3).

Proof of Theorem 3.1.1. By Lemma 3.3.1 and Lemma 3.3.2, the inverse Fourier transform gives

$$\begin{aligned}
& V_{i,j}^N - v(T, ih_x, jh_y) \\
&= \frac{1}{4\pi^2} \iint_{\Omega_{\mathrm{low}} \cup \Omega_{\mathrm{high}}} \big(X_N - X(T) \big) \mathrm{e}^{\mathrm{i}\left((i - i_0)\xi h_x + (j - j_0)\eta h_y \right)}\, \mathrm{d}\xi\, \mathrm{d}\eta \\
&= k\, E_1(T, x_i, y_j) + h_x^2\, E_2(T, x_i, y_j) + h_y^2\, E_3(T, x_i, y_j) + \theta^N h_x^{-2}\, E_4(T, x_i, y_j) \\
&\quad + \theta^N h_y^{-2}\, E_5(T, x_i, y_j) + o(k, h_x^2, h_y^2, \theta^N h_x^{-2}, \theta^N h_y^{-2})\, R(T, x_i, y_j),
\end{aligned}$$

where $x_i = ih_x$, $y_j = jh_y$, $E_1, \ldots, E_5$, and R are random variables with bounded first and second moments, $N = T/k$, $0 < \theta < 1$, and θ is independent of h_x, h_y and k. □

Next we give a proof of Corollary 3.1.2, the L_2 convergence in space and probability of the implicit finite difference scheme (3.3).

Proof of Corollary 3.1.2. We apply Parseval's theorem to $V_{i,j}^N - v(T, ih_x, jh_y)$ and its Fourier transform. It follows

$$\sum_{i,j} \left| V_{i,j}^N - v(T, ih_x, jh_y) \right|^2 h_x h_y = \iint |\widetilde{v}(0,\xi,\eta)|^2 |X(T,\xi,\eta) - X_N(\xi,\eta)|^2 \,\mathrm{d}\xi \,\mathrm{d}\eta.$$

In Lemma 3.3.1, we have proved $|X(T,\xi,\eta) - X_N(\xi,\eta)| = X(T)\big(1 + O(h_x^2) + O(h_y^2) + O(k)\big)$ for (ξ,η) in the low wave region. In Lemma 3.3.2, we have proved $|X(T,\xi,\eta) - X_N(\xi,\eta)|^2 \leq C\theta^N$, for some $C > 0$ and $\theta \in (0,1)$, (ξ,η) in the high wave region.

As for Dirac initial datum, $|\widetilde{v}(0,\xi,\eta)| = 1$, we have

$$\begin{aligned}
&\iint |X(T,\xi,\eta) - X_N(\xi,\eta)|^2 \,\mathrm{d}\xi \,\mathrm{d}\eta \\
&= \iint_{\Omega_{\text{low}}} |X(T,\xi,\eta) - X_N(\xi,\eta)|^2 \,\mathrm{d}\xi \,\mathrm{d}\eta + \iint_{\Omega_{\text{high}}} |X(T,\xi,\eta) - X_N(\xi,\eta)|^2 \,\mathrm{d}\xi \,\mathrm{d}\eta \\
&= O(h_x^2) + O(h_y^2) + O(k) + O(h_x^{-1} h_y^{-1} \theta^N).
\end{aligned}$$

For initial data in L_2, $\iint |\widetilde{v}(0,\xi,\eta)|^2 \,\mathrm{d}\xi \,\mathrm{d}\eta < \infty$, and therefore

$$\iint_{\Omega_{\text{high}}} |\widetilde{v}(0,\xi,\eta)|^2 |X(T,\xi,\eta) - X_N(\xi,\eta)|^2 \,\mathrm{d}\xi \,\mathrm{d}\eta = o(k^r) \quad \text{for any } r > 0,$$

and consequently

$$\iint |X(T,\xi,\eta) - X_N(\xi,\eta)|^2 \,\mathrm{d}\xi \,\mathrm{d}\eta = O(h_x^2) + O(h_y^2) + O(k).$$

□

3.3.1 Low wave number region (proof of Lemma 3.3.1)

For the low wave region, we consider the case where both ξ, η are small. It follows from (3.15) that the exact solution of $X(t_{n+1})$ given $X(t_n)$ is

$$X(t_{n+1}) = X(t_n) \exp\left(-\frac{1}{2}(1-\rho_x)\xi^2 k - \frac{1}{2}(1-\rho_y)\eta^2 k - \mathrm{i}\xi\sqrt{\rho_x k} Z_{n,x} - \mathrm{i}\eta\sqrt{\rho_y k}\widetilde{Z}_{n,y} \right),$$

where $W_{t_{n+1}}^x - W_{t_n}^x \equiv \sqrt{k} Z_{n,x}$, $W_{t_{n+1}}^y - W_{t_n}^y \equiv \sqrt{k}\widetilde{Z}_{n,y}$ are the Brownian increments.

Now we consider X_n, the numerical approximation of $X(nk)$. Let

$$X_{n+1} = C_n \, X_n,$$

where

$$C_n = \exp\bigg(-\frac{1}{2}(1-\rho_x)\xi^2 k - \frac{1}{2}(1-\rho_y)\eta^2 k - \mathrm{i}\xi\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}\eta\sqrt{\rho_y k}\widetilde{Z}_{n,y} + e_n\bigg), \quad (3.25)$$

and e_n is the logarithmic error between the numerical solution and the exact solution introduced during $[nk,(n+1)k]$. Aggregating over N time steps, at $t_N = kN = T$,

$$X_N = X(T)\exp\bigg(\sum_{n=0}^{N-1} e_n\bigg), \quad (3.26)$$

where

$$X(T) = \exp\bigg(-\frac{1}{2}(1-\rho_x)\xi^2 T - \frac{1}{2}(1-\rho_y)\eta^2 T - \mathrm{i}\xi\sqrt{\rho_x k}\sum_{n=0}^{N-1} Z_{n,x} - \mathrm{i}\eta\sqrt{\rho_y k}\sum_{n=0}^{N-1}\widetilde{Z}_{n,y}\bigg)$$

is the exact solution at time T.

From (3.25), we have

$$e_n = \log C_n + \frac{1}{2}(1-\rho_x)\xi^2 k + \frac{1}{2}(1-\rho_y)\eta^2 k + \mathrm{i}\xi\sqrt{\rho_x k}Z_{n,x} + \mathrm{i}\eta\sqrt{\rho_y k}\widetilde{Z}_{n,y},$$

hence

$$\sum_{n=0}^{N-1} e_n = \sum_{n=0}^{N-1}\log C_n + \frac{1}{2}(1-\rho_x)\xi^2 T + \frac{1}{2}(1-\rho_y)\eta^2 T + \mathrm{i}\xi\sqrt{\rho_x k}\sum_{n=0}^{N-1} Z_{n,x} + \mathrm{i}\eta\sqrt{\rho_y k}\sum_{n=0}^{N-1}\widetilde{Z}_{n,y}.$$

From (3.19), C_n has the form

$$C_n = \frac{1 - \mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + b_x\rho_x k(Z_{n,x}^2-1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}}{1-(a_x+a_y)k},$$

where

$$\begin{aligned}
a_x &= -\frac{\xi^2}{2}\cdot\frac{\sin^2\frac{\xi h_x}{2}}{(\frac{\xi h_x}{2})^2} = -\frac{\xi^2}{2} + \frac{\xi^4}{24}h_x^2 + O(h_x^4\xi^6),\\
a_y &= -\frac{\eta^2}{2}\cdot\frac{\sin^2\frac{\eta h_y}{2}}{(\frac{\eta h_y}{2})^2} = -\frac{\eta^2}{2} + \frac{\eta^4}{24}h_y^2 + O(h_y^4\eta^6),\\
b_x &= -\frac{\xi^2}{2}\cdot\frac{\sin^2\xi h_x}{\xi^2 h_x^2} = -\frac{\xi^2}{2} + \frac{\xi^4}{6}h_x^2 + O(h_x^4\xi^6),\\
b_y &= -\frac{\eta^2}{2}\cdot\frac{\sin^2\eta h_y}{\eta^2 h_y^2} = -\frac{\eta^2}{2} + \frac{\eta^4}{6}h_y^2 + O(h_y^4\eta^6),\\
c_x &= \xi\cdot\frac{\sin\xi h_x}{\xi h_x} = \xi - \frac{\xi^3}{6}h_x^2 + O(h_x^4\xi^5),\\
c_y &= \eta\cdot\frac{\sin\eta h_y}{\eta h_y} = \eta - \frac{\eta^3}{6}h_y^2 + O(h_y^4\eta^5),\\
d &= -\xi\eta\cdot\frac{\sin\xi h_x\sin\eta h_y}{\xi h_x\eta h_y}.
\end{aligned}$$

Note that $c_xc_y + d = 0$, $b_x + \frac{1}{2}c_x^2 = 0$, $b_y + \frac{1}{2}c_y^2 = 0$, then one can derive by Taylor expansion (by lengthy, but elementary calculations),

$$\begin{aligned}
\log C_n &= -\mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + (a_x + a_y - b_x\rho_x - b_y\rho_y)k + (b_x + \frac{1}{2}c_x^2)\rho_x k Z_{n,x}^2 \\
&+ (b_y + \frac{1}{2}c_y^2)\rho_y k\widetilde{Z}_{n,y}^2 + (c_x c_y + d)\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y} + O\big((|\xi|+|\eta|)^3 k\sqrt{k}\big)\cdot \mathrm{i}\phi_1(Z_{n,x},\widetilde{Z}_{n,y}) \\
&+ O\big((|\xi|+|\eta|)^4k^2\big)\cdot\phi_2(Z_{n,x},\widetilde{Z}_{n,y}) + o\big((|\xi|+|\eta|)^4k^2\big) \\
&= -\mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + (a_x + a_y - b_x\rho_x - b_y\rho_y)k \\
&+ O\big((|\xi|+|\eta|)^3 k\sqrt{k}\big)\cdot\mathrm{i}\phi_1(Z_{n,x},\widetilde{Z}_{n,y}) + O\big((|\xi|+|\eta|)^4k^2\big)\cdot\phi_2(Z_{n,x},\widetilde{Z}_{n,y}) + o\big((|\xi|+|\eta|)^4k^2\big),
\end{aligned}$$

where $\phi_1(\cdot,\cdot)$ is an odd and $\phi_2(\cdot,\cdot)$ an even degree polynomial. Therefore

$$\begin{aligned}
\sum_{n=0}^{N-1} e_n &= \sum_{n=0}^{N-1}\log C_n + \frac{1}{2}(1-\rho_x)\xi^2 T + \frac{1}{2}(1-\rho_y)\eta^2 T + \mathrm{i}\xi\sqrt{\rho_x k}\sum_{n=0}^{N-1}Z_{n,x} + \mathrm{i}\eta\sqrt{\rho_y k}\sum_{n=0}^{N-1}\widetilde{Z}_{n,y} \\
&= \mathrm{i}(\xi - c_x)\sqrt{\rho_x k}\sum_{n=0}^{N-1}Z_{n,x} + \mathrm{i}(\eta - c_y)\sqrt{\rho_y k}\sum_{n=0}^{N-1}\widetilde{Z}_{n,y} \\
&\quad + \Big(a_x + a_y - b_x\rho_x - b_y\rho_y + \frac{1-\rho_x}{2}\xi^2 + \frac{1-\rho_y}{2}\eta^2\Big)T \\
&\quad + O\big((|\xi|+|\eta|)^3 k\sqrt{k}\big)\cdot\mathrm{i}\sum_{n=0}^{N-1}\phi_1(Z_{n,x},\widetilde{Z}_{n,y}) \\
&\quad + O\big((|\xi|+|\eta|)^4k^2\big)\cdot\sum_{n=0}^{N-1}\phi_2(Z_{n,x},\widetilde{Z}_{n,y}) + o\big((|\xi|+|\eta|)^4k\big),
\end{aligned}$$

so we have

$$\begin{aligned}
\exp\left(\sum_{n=0}^{N-1}e_n\right) &= \exp\bigg(\Big(a_x + a_y - b_x\rho_x - b_y\rho_y + \frac{1-\rho_x}{2}\xi^2 + \frac{1-\rho_y}{2}\eta^2\Big)T\bigg) \\
&\quad\cdot\exp\bigg(\mathrm{i}(\xi - c_x)\sqrt{\rho_x k}\sum_{n=0}^{N-1}Z_{n,x} + \mathrm{i}(\eta - c_y)\sqrt{\rho_y k}\sum_{n=0}^{N-1}\widetilde{Z}_{n,y} \\
&\qquad + O\big((|\xi|+|\eta|)^3 k\sqrt{k}\big)\cdot\mathrm{i}\sum_{n=0}^{N-1}\phi_1(Z_{n,x},\widetilde{Z}_{n,y}) \\
&\qquad + O\big((|\xi|+|\eta|)^4k^2\big)\cdot\sum_{n=0}^{N-1}\phi_2(Z_{n,x},\widetilde{Z}_{n,y}) + o\big((|\xi|+|\eta|)^4k\big)\bigg).
\end{aligned}$$

Here

$$\begin{aligned}
&\exp\left(\Big(a_x + a_y - b_x\rho_x - b_y\rho_y + \frac{1-\rho_x}{2}\xi^2 + \frac{1-\rho_y}{2}\eta^2\Big)T\right) \\
&= 1 + \frac{\xi^4}{24}h_x^2(1-4\rho_x)T + \frac{\eta^4}{24}h_y^2(1-4\rho_y)T + O(\xi^6 h_x^4) + O(\eta^6 h_y^4),
\end{aligned}$$

and

$$\begin{aligned}
&\exp\bigg(\mathrm{i}(\xi-c_x)\sqrt{\rho_x k}\sum_{n=0}^{N-1}Z_{n,x}+\mathrm{i}(\eta-c_y)\sqrt{\rho_y k}\sum_{n=0}^{N-1}\widetilde{Z}_{n,y}\\
&\qquad+O\big((|\xi|+|\eta|)^3k\sqrt{k}\big)\cdot\mathrm{i}\sum_{n=0}^{N-1}\phi_1(Z_{n,x},\widetilde{Z}_{n,y})+O\big((|\xi|+|\eta|)^4k^2\big)\cdot\sum_{n=0}^{N-1}\phi_2(Z_{n,x},\widetilde{Z}_{n,y})\bigg)\\
&=1+\mathrm{i}(\xi-c_x)\sqrt{\rho_x k}\sum_{n=0}^{N-1}Z_{n,x}+\mathrm{i}(\eta-c_y)\sqrt{\rho_y k}\sum_{n=0}^{N-1}\widetilde{Z}_{n,y}-\frac{1}{2}(\xi-c_x)^2\rho_x k\bigg(\sum_{n=0}^{N-1}Z_{n,x}\bigg)^2\\
&\quad-\frac{1}{2}(\eta-c_y)^2\rho_y k\bigg(\sum_{n=0}^{N-1}\widetilde{Z}_{n,y}\bigg)^2+O\big((|\xi|+|\eta|)^3k\sqrt{k}\big)\cdot\mathrm{i}\sum_{n=0}^{N-1}\widehat{\phi}_1(Z_{n,x},\widetilde{Z}_{n,y})\\
&\quad+O\big((|\xi|+|\eta|)^4k^2\big)\cdot\sum_{n=0}^{N-1}\widehat{\phi}_2(Z_{n,x},\widetilde{Z}_{n,y})+o\big((|\xi|+|\eta|)^4k\big),
\end{aligned}$$

where $\widehat{\phi}_1(\cdot,\cdot)$ is a polynomial function with odd degree, and $\widehat{\phi}_2(\cdot,\cdot)$ are with even degree, and

$$\begin{aligned}
&\mathbb{E}\bigg[X(T)\sum_n\widehat{\phi}_1(Z_{n,x},\widetilde{Z}_{n,y})\bigg]=O(k^{-\frac{1}{2}})\exp\bigg(-\frac{1}{2}\big(\xi^2+\eta^2+2\xi\eta\sqrt{\rho_x\rho_y}\rho_{xy}\big)T\bigg),\\
&\mathbb{E}\bigg[X(T)\sum_n\widehat{\phi}_2(Z_{n,x},\widetilde{Z}_{n,y})\bigg]=O(k^{-1})\exp\bigg(-\frac{1}{2}\big(\xi^2+\eta^2+2\xi\eta\sqrt{\rho_x\rho_y}\rho_{xy}\big)T\bigg),\\
&\mathbb{E}\bigg|X(T)\sum_n\widehat{\phi}_1(Z_{n,x},\widetilde{Z}_{n,y})\bigg|^2=O(k^{-1})\exp\bigg(-\big(\xi^2+\eta^2+2\xi\eta\sqrt{\rho_x\rho_y}\rho_{xy}\big)T\bigg),\\
&\mathbb{E}\bigg|X(T)\sum_n\widehat{\phi}_2(Z_{n,x},\widetilde{Z}_{n,y})\bigg|^2=O(k^{-2})\exp\bigg(-\big(\xi^2+\eta^2+2\xi\eta\sqrt{\rho_x\rho_y}\rho_{xy}\big)T\bigg).
\end{aligned}$$

Hence we have in the low wave number region,

$$\begin{aligned}
X_N-X(T)&=X(T)\cdot\bigg(\exp\Big(\sum_{n=0}^{N-1}e_n\Big)-1\bigg)=X(T)\cdot\bigg\{\frac{\mathrm{i}}{6}\sqrt{\rho_x}\xi^3h_x^2W_T^x+\frac{1}{24}(1-4\rho_x)\xi^4h_x^2T\\
&+\frac{\mathrm{i}}{6}\sqrt{\rho_y}\eta^3h_y^2\widetilde{W}_T^y+\frac{1}{24}(1-4\rho_y)\eta^4h_y^2T+O\big((|\xi|+|\eta|)^3k\sqrt{k}\big)\cdot\mathrm{i}\sum_{n=0}^{N-1}\widehat{\phi}_1(Z_{n,x},\widetilde{Z}_{n,y})\\
&+O\big((|\xi|+|\eta|)^4k^2\big)\cdot\sum_{n=0}^{N-1}\widehat{\phi}_2(Z_{n,x},\widetilde{Z}_{n,y})+o(k,h_x^2,h_y^2)\cdot\varphi(T,h_x\xi,h_y\eta)\bigg\},
\end{aligned}$$

where $\varphi(T,h_x\xi,h_y\eta)$ is a random variable with bounded moments.

Remark 3.3.1. *We can derive the exact leading order term by taking the inverse Fourier transform. For instance, the leading order error in h_x is*

$$\bigg(-\frac{1}{6}\sqrt{\rho_x}W_T^x\frac{\partial^3}{\partial x^3}v(T,x,y)+\frac{1}{24}(1-4\rho_x)T\frac{\partial^4}{\partial x^4}v(T,x,y)\bigg)\cdot h_x^2,$$

and similar for h_y (replacing 'x' by 'y'); the leading order error in k can be found by the same technique but is significantly lengthier and hence omitted.

3.3.2 High wave number region (proof of Lemma 3.3.2)

Now we consider the case when either ξ or η is large.

First we calculate the upper bound of $\mathbb{E}\big[\big|X_N(\xi,\eta)\big|^2\big]$. To simplify the proof, we take $h_x = h_y = h$, and the case where $h_x \neq h_y$ is similar. Write $\lambda = \frac{k}{h^2}$.

Lemma 3.3.3. *For $(\xi,\eta) \notin \Omega_{low}$,*

$$\mathbb{E}\big[\big|X_N(\xi,\eta)\big|^2\big] \leq |X_0|^2\bigg(1 - 4\beta\frac{\lambda\big(\sin^2\frac{\xi h}{2} + \sin^2\frac{\eta h}{2}\big) + \lambda^2\big(\sin^2\frac{\xi h}{2} + \sin^2\frac{\eta h}{2}\big)^2}{\Big(1 + 2\lambda\Big(\sin^2\frac{\xi h}{2} + \sin^2\frac{\eta h}{2}\Big)\Big)^2}\bigg)^N,$$

where

$$\beta = \min\big\{1-\rho_x, 1-\rho_y, 1-2\rho_x^2(1+2|\rho_{xy}|), 1-2\rho_y^2(1+2|\rho_{xy}|), 1-2\rho_x\rho_y(1+2|\rho_{xy}|+3\rho_{xy}^2)\big\} \in (0,1).$$

Proof. By (3.19), we have $X_N = X_0\prod_{n=0}^{N-1} C_n$, where

$$C_n = \frac{1-\mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + b_x\rho_x k(Z_{n,x}^2-1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}}{1-(a_x+a_y)k},$$

$$\begin{aligned}
a_x &= -\frac{2\sin^2\frac{\xi h_x}{2}}{h_x^2}, \qquad b_x = -\frac{\sin^2\xi h_x}{2h_x^2}, \qquad c_x = \frac{\sin\xi h_x}{h_x}, \qquad d = -\frac{\sin\xi h_x\sin\eta h_y}{h_x h_y},\\
a_y &= -\frac{2\sin^2\frac{\eta h_y}{2}}{h_y^2}, \qquad b_y = -\frac{\sin^2\eta h_y}{2h_y^2}, \qquad c_y = \frac{\sin\eta h_y}{h_y}.
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E}\big[\big|C_n\big|^2\big] \leq 1 &- \frac{4\frac{k}{h^2}\sin^2\frac{\xi h}{2}\big(1-\rho_x\cos^2\frac{\xi h}{2} + \frac{k}{h^2}\sin^2\frac{\xi h}{2}\,(1-2\rho_x^2(1+2|\rho_{xy}|))\big)}{\Big[1+2\frac{k}{h^2}\Big(\sin^2\frac{\xi h}{2}+\sin^2\frac{\eta h}{2}\Big)\Big]^2}\\
&- \frac{4\frac{k}{h^2}\sin^2\frac{\eta h}{2}\big(1-\rho_y\cos^2\frac{\eta h}{2} + \frac{k}{h^2}\sin^2\frac{\eta h}{2}\,(1-2\rho_y^2(1+2|\rho_{xy}|))\big)}{\Big[1+2\frac{k}{h^2}\Big(\sin^2\frac{\eta h}{2}+\sin^2\frac{\eta h}{2}\Big)\Big]^2}\\
&- \frac{8\frac{k^2}{h^4}\sin^2\frac{\xi h}{2}\sin^2\frac{\eta h}{2}\,\big(1-2\rho_x\rho_y(1+2|\rho_{xy}|+3\rho_{xy}^2)\big)}{\Big[1+2\frac{k}{h^2}\Big(\sin^2\frac{\eta h}{2}+\sin^2\frac{\eta h}{2}\Big)\Big]^2}.
\end{aligned}$$

Denote $\lambda = \frac{k}{h^2},\ a = \sin^2\frac{\xi h}{2},\ b = \sin^2\frac{\eta h}{2}$. It follows that

$$\begin{aligned}\mathbb{E}\big[\big|C_n\big|^2\big] \le 1 &- \frac{4\lambda a\Big(1-\rho_x\cos^2\frac{\xi h}{2}+\lambda a\big(1-2\rho_x^2(1+2|\rho_{xy}|)\big)\Big)}{\big(1+2\lambda(a+b)\big)^2}\\ &- \frac{4\lambda b\Big(1-\rho_y\cos^2\frac{\eta h}{2}+\lambda b\big(1-2\rho_y^2(1+2|\rho_{xy}|)\big)\Big)}{\big(1+2\lambda(a+b)\big)^2}\\ &- \frac{8\lambda^2 ab\Big(1-2\rho_x\rho_y(1+2|\rho_{xy}|+3\rho_{xy}^2)\Big)}{\big(1+2\lambda(a+b)\big)^2}.\end{aligned}$$

By Assumption 3.1.1,

$$\begin{aligned}&0\le\rho_x<1,\quad 0\le\rho_y<1,\quad 0<1-2\rho_x^2(1+2|\rho_{xy}|)\le 1,\\ &0<1-2\rho_y^2(1+2|\rho_{xy}|)\le 1,\quad 0<1-2\rho_x\rho_y(1+2|\rho_{xy}|+3\rho_{xy}^2)\le 1,\end{aligned}$$

we write

$$\begin{aligned}\beta = \min\big\{&1-\rho_x, 1-\rho_y, 1-2\rho_x^2(1+2|\rho_{xy}|), 1-2\rho_y^2(1+2|\rho_{xy}|),\\ &1-2\rho_x\rho_y(1+2|\rho_{xy}|+3\rho_{xy}^2)\big\}\in(0,1),\\ d = a+b &= \sin^2\frac{\xi h}{2}+\sin^2\frac{\eta h}{2}.\end{aligned}$$

Consequently,

$$\mathbb{E}\big[\big|C_n\big|^2\big] \le 1-\frac{4\lambda a\big(\beta+\lambda a\beta\big)+4\lambda b\big(\beta+\lambda b\beta\big)+8\lambda^2 ab\beta}{\big(1+2\lambda(a+b)\big)^2} = 1-\frac{4\beta\big(\lambda d+\lambda^2 d^2\big)}{\big(1+2\lambda d\big)^2}.$$

Therefore we have

$$\begin{aligned}\mathbb{E}\big[\big|X_N\big|^2\big] &= \big|X_0\big|^2\prod_{n=0}^{N-1}\mathbb{E}\big[\big|C_n\big|^2\big]\\ &\le |X_0|^2\left(1-4\beta\frac{\lambda\big(\sin^2\frac{\xi h}{2}+\sin^2\frac{\eta h}{2}\big)+\lambda^2\big(\sin^2\frac{\xi h}{2}+\sin^2\frac{\eta h}{2}\big)^2}{\Big(1+2\lambda\Big(\sin^2\frac{\xi h}{2}+\sin^2\frac{\eta h}{2}\Big)\Big)^2}\right)^N.\end{aligned}\tag{3.27}$$

□

Then we consider two scenarios: $0<\lambda<1$ and $\lambda\ge 1$. For $0<\lambda<1$,

$$\Omega_{\text{high}} = \big\{(\xi,\eta):|\xi|>h^{-2p},\ \text{or}\ \ |\eta|>h^{-2p}\big\}\cap[-\pi h^{-1},\pi h^{-1}]\times[-\pi h^{-1},\pi h^{-1}].$$

Lemma 3.3.4. *For $0 < \lambda < 1$ (i.e., $k < h^2$),*

$$\mathbb{E}\bigg[\Big|\iint_{\Omega_{high}} X(T,\xi,\eta) - X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\bigg] = o(h^r), \quad \forall r > 0.$$

Proof. Note that

$$X(T) = \exp\Big(-\frac{1}{2}(1-\rho_x)\xi^2 T - \frac{1}{2}(1-\rho_y)\eta^2 T - \mathrm{i}\xi\sqrt{\rho_x}W_t^x - \mathrm{i}\eta\sqrt{\rho_y}W_t^y\Big).$$

Then

$$\begin{aligned}
&\mathbb{E}\bigg[\Big|\iint_{\Omega_{\mathrm{high}}} X(T,\xi,\eta) - X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\bigg] \\
&< 4\pi^2 h^{-2}\mathbb{E}\bigg[\iint_{\Omega_{\mathrm{high}}}\Big|X(T,\xi,\eta) - X_N(\xi,\eta)\Big|^2\,\mathrm{d}\xi\mathrm{d}\eta\bigg] \\
&\le 8\pi^2 h^{-2}\iint_{\Omega_{\mathrm{high}}}\mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2 + \big|X(T,\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta \\
&= 8\pi^2 h^{-2}\iint_{\Omega_{\mathrm{high}}}\mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta + f_0(k),
\end{aligned}$$

where

$$\begin{aligned}
f_0(k) &= 8\pi^2 h^{-2}\iint_{\Omega_{\mathrm{high}}} \mathrm{e}^{-(1-\rho_x)\xi^2 T-(1-\rho_y)\eta^2 T}\,\mathrm{d}\xi\mathrm{d}\eta \\
&= 8\pi^2 h^{-2}\int_0^{\pi/h}\int_{h^{-2p}}^{\pi/h} \mathrm{e}^{-(1-\rho_x)\xi^2 T-(1-\rho_y)\eta^2 T}\,\mathrm{d}\xi\mathrm{d}\eta \\
&\quad + 8\pi^2 h^{-2}\int_{h^{-2p}}^{\pi/h}\int_0^{\pi/h} \mathrm{e}^{-(1-\rho_x)\xi^2 T-(1-\rho_y)\eta^2 T}\,\mathrm{d}\xi\mathrm{d}\eta \\
&\quad - 8\pi^2 h^{-2}\int_{h^{-2p}}^{\pi/h}\int_{h^{-2p}}^{\pi/h} \mathrm{e}^{-(1-\rho_x)\xi^2 T-(1-\rho_y)\eta^2 T}\,\mathrm{d}\xi\mathrm{d}\eta \\
&\le C\cdot h^{-2+2p}\big(\mathrm{e}^{-(1-\rho_x)Th^{-2p}} + \mathrm{e}^{-(1-\rho_y)Th^{-2p}}\big) = o(h^r), \quad \forall r > 0.
\end{aligned}$$

Denote $d = \sin^2\frac{\xi h}{2} + \sin^2\frac{\eta h}{2}$. From (3.27) and $\lambda = k/h^2$,

$$\mathbb{E}\big[\big|X_N\big|^2\big] \le |X_0|^2\bigg(1 - \frac{4\beta\big(d+\lambda d^2\big)}{\big(1+2\lambda d\big)^2}\cdot\frac{T}{Nh^2}\bigg)^N < |X_0|^2\exp\Big(-\frac{4\beta\big(d+\lambda d^2\big)T}{\big(1+2\lambda d\big)^2}\cdot h^{-2}\Big).$$

In this case, as at least one of ξ and η belongs to $(h^{-2p}, \pi/h)$, we have

$$d = a + b = \sin^2\frac{\xi h}{2} + \sin^2\frac{\eta h}{2} \ge \sin^2\frac{h^{1-2p}}{2} = \frac{h^{2-4p}}{4} - \frac{h^{4-8p}}{48} + O(h^{5-10p}).$$

Therefore,

$$\mathbb{E}\big[\big|X_N\big|^2\big] < |X_0|^2 \exp\Big(- \frac{4\beta\big(d+\lambda d^2\big)T}{\big(1+2\lambda d\big)^2} \cdot h^{-2}\Big)$$
$$< |X_0|^2 \exp\big(- 4\beta d T h^{-2}\big) < |X_0|^2 \exp\big(- \beta T h^{-4p}\big) = o(h^r).$$

As a result,

$$8\pi^2 h^{-2} \iint_{\Omega_{\text{high}}} \mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta < 16\pi^4 |X_0|^2 h^{-4} \exp\big(- \beta T h^{-4p}\big) = o(h^r), \quad \forall r > 0.$$

Therefore, for all $r > 0$,

$$\mathbb{E}\Bigg[\,\Big|\iint_{\Omega_{\text{high}}} X(T,\xi,\eta) - X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\,\Bigg] < 8\pi^2 h^{-2} \iint_{\Omega_{\text{high}}} \mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta + f_0(k) = o(h^r).$$

□

For $\lambda \geq 1$, we further separate the domain Ω_{high} into a middle wave region

$$\Omega^1_{\text{high}} = \big\{(|\xi|,|\eta|) \in [k^{-p}, k^{-\frac{1}{2}}] \times [0, k^{-\frac{1}{2}}] \cup [0, k^{-\frac{1}{2}}] \times [k^{-p}, k^{-\frac{1}{2}}]\big\}, \tag{3.28}$$

and a high wave region

$$\Omega^2_{\text{high}} = \big\{(|\xi|,|\eta|) \in [k^{-\frac{1}{2}}, \pi/h] \times [0, \pi/h] \cup [0, \pi/h] \times [k^{-\frac{1}{2}}, \pi/h]\big\}, \tag{3.29}$$

such that $\Omega_{\text{high}} = \Omega^1_{\text{high}} \cup \Omega^2_{\text{high}}$.

Lemma 3.3.5. *For $\lambda \geq 1$ (i.e., $k \geq h^2$), there exists $\theta \in (0,1)$ independent of h and k such that*

$$\mathbb{E}\Bigg[\,\Big|\iint_{\Omega_{high}} X(T,\xi,\eta) - X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\,\Bigg] \leq C h^{-4} \theta^N.$$

Proof.

$$
\begin{aligned}
&\mathbb{E}\bigg[\Big|\iint_{\Omega_{\text{high}}} X(T,\xi,\eta)-X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\bigg]\\
&\le 2\,\mathbb{E}\bigg[\Big|\iint_{\Omega^1_{\text{high}}} X(T,\xi,\eta)-X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\bigg]+2\,\mathbb{E}\bigg[\Big|\iint_{\Omega^2_{\text{high}}} X(T,\xi,\eta)-X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\bigg]\\
&< 8k^{-1}\mathbb{E}\bigg[\iint_{\Omega^1_{\text{high}}}\Big|X(T,\xi,\eta)-X_N(\xi,\eta)\Big|^2\,\mathrm{d}\xi\mathrm{d}\eta\bigg]\\
&\quad+8\pi^2h^{-2}\mathbb{E}\bigg[\iint_{\Omega^2_{\text{high}}}\Big|X(T,\xi,\eta)-X_N(\xi,\eta)\Big|^2\,\mathrm{d}\xi\mathrm{d}\eta\bigg]\\
&\le 16k^{-1}\iint_{\Omega^1_{\text{high}}}\mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2+\big|X(T,\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta\\
&\quad+16\pi^2h^{-2}\iint_{\Omega^2_{\text{high}}}\mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2+\big|X(T,\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta\\
&= 16k^{-1}\iint_{\Omega^1_{\text{high}}}\mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta+16\pi^2h^{-2}\iint_{\Omega^2_{\text{high}}}\mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta+f_1(k),
\end{aligned}
$$

where

$$
\begin{aligned}
f_1(k)&=16k^{-1}\iint_{\Omega^1_{\text{high}}}\mathrm{e}^{-(1-\rho_x)\xi^2T-(1-\rho_y)\eta^2T}\,\mathrm{d}\xi\mathrm{d}\eta+16\pi^2h^{-2}\iint_{\Omega^2_{\text{high}}}\mathrm{e}^{-(1-\rho_x)\xi^2T-(1-\rho_y)\eta^2T}\,\mathrm{d}\xi\mathrm{d}\eta\\
&\le C\cdot k^{p-1}\exp\big(-\beta Tk^{-p}\big)+C\cdot kh^{-2}\exp\big(-2\beta Tk^{-1}\big).
\end{aligned}
$$

Denote $d=\sin^2\frac{\xi h}{2}+\sin^2\frac{\eta h}{2}$, from (3.27),

$$
\mathbb{E}\big[\big|X_N\big|^2\big]\le|X_0|^2\bigg(1-\frac{4\beta\big(\lambda d+\lambda^2d^2\big)}{\big(1+2\lambda d\big)^2}\bigg)^N.
$$

For $\lambda\ge1$, and $(\xi,\eta)\in\Omega^1_{\text{high}}$,

$$
\mathbb{E}\big[\big|X_N\big|^2\big]\le|X_0|^2\bigg(1-\frac{4\beta\big(d+\lambda d^2\big)}{\big(1+2\lambda d\big)^2}\cdot\frac{T}{Nh^2}\bigg)^N<|X_0|^2\exp\Big(-\frac{4\beta\big(d+\lambda d^2\big)T}{\big(1+2\lambda d\big)^2}\cdot h^{-2}\Big).
$$

In this case, as at least one of ξ and η belongs to $(k^{-p},k^{-1/2})$, we have

$$
d=a+b=\sin^2\frac{\xi h}{2}+\sin^2\frac{\eta h}{2}\ge\sin^2\frac{k^{-p}h}{2}=\frac{k^{-2p}h^2}{4}-\frac{k^{-4p}h^4}{48}+O(k^{-5p}h^5).
$$

$$
\begin{aligned}
\mathbb{E}\big[\big|X_N\big|^2\big]&<|X_0|^2\exp\Big(-\frac{4\beta\big(d+\lambda d^2\big)T}{\big(1+2\lambda d\big)^2}\cdot h^{-2}\Big)\\
&<|X_0|^2\exp\big(-4\beta dTh^{-2}\big)\\
&<|X_0|^2\exp\big(-\beta Tk^{-2p}\big).
\end{aligned}
$$

Since $|\Omega^1_{\text{high}}| < 4k^{-1}$,

$$16k^{-1}\iint_{\Omega^1_{\text{high}}} \mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta < 64|X_0|^2k^{-2}\exp\big(-\beta Tk^{-2p}\big).$$

For $\lambda \geq 1$, and $(\xi,\eta) \in \Omega^2_{\text{high}}$, $d = \sin^2\frac{\xi h}{2} + \sin^2\frac{\eta h}{2} \in [\sin^2\frac{1}{2\sqrt{\lambda}}, 2]$,

$$\max_d\left(1-\frac{4\beta\big(\lambda d+\lambda^2d^2\big)}{\big(1+2\lambda d\big)^2}\right) = 1-\beta\min_d\left(1-\frac{1}{\big(1+2\lambda d\big)^2}\right) = 1-\beta\left(1-\max_d\frac{1}{\big(1+2\lambda d\big)^2}\right).$$

As

$$\max_d\frac{1}{\big(1+2\lambda d\big)^2} = \frac{1}{\big(1+2\lambda d\big)^2}\bigg|_{d=\sin^2\frac{1}{2\sqrt{\lambda}}} = \frac{1}{\big(1+2\lambda\sin^2\frac{1}{2\sqrt{\lambda}}\big)^2} \leq \frac{1}{\big(1+2\sin^2\frac{1}{2}\big)^2} < 0.5,$$

we have

$$\begin{aligned}1-\frac{4\beta\big(\lambda d+\lambda^2d^2\big)}{\big(1+2\lambda d\big)^2} &\leq 1-\beta\Big(1-\frac{1}{\big(1+2\lambda\sin^2\frac{1}{2\sqrt{\lambda}}\big)^2}\Big)\\ &= 1-\beta+\frac{\beta}{\big(1+2\lambda\sin^2\frac{1}{2\sqrt{\lambda}}\big)^2} < 1-\frac{1}{2}\beta.\end{aligned}$$

Denote $\theta_0 \coloneqq 1-\frac{1}{2}\beta \in (0,1)$, then

$$\mathbb{E}\big[\big|X_N\big|^2\big] \leq |X_0|^2\left(1-\frac{4\beta\big(\lambda d+\lambda^2d^2\big)}{\big(1+2\lambda d\big)^2}\right)^N \leq |X_0|^2\cdot\theta_0^N.$$

So

$$16\pi^2h^{-2}\iint_{\Omega^2_{\text{high}}} \mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta < 64\pi^4|X_0|^2h^{-4}\theta_0^N.$$

Hence

$$\begin{aligned}&\mathbb{E}\left[\left|\iint_{\Omega_{\text{high}}} X(T,\xi,\eta)-X_N(\xi,\eta)\mathrm{d}\xi\mathrm{d}\eta\right|^2\right]\\ &\leq Ck^{p-1}\exp\big(-\beta Tk^{-p}\big)+C\,kh^{-2}\exp\big(-2\beta Tk^{-1}\big)\\ &\quad+64|X_0|^2k^{-2}\exp\big(-\beta Tk^{-2p}\big)+64\pi^4|X_0|^2h^{-4}\theta_0^N.\end{aligned}\tag{3.30}$$

As the first three terms in (3.30) are terms of higher order than $h^4\theta^N$, we have, for $\lambda \geq 1$,

$$\mathbb{E}\left[\left|\iint_{\Omega_{\text{high}}} X(T,\xi,\eta)-X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\right|^2\right] \leq Ch^{-4}\theta_0^N.$$

Letting $\theta = \sqrt{\theta_0}$ the result follows. □

3.3.3 Convergence of the ADI scheme (proof of Theorem 3.1.2)

Theorem 3.1.2 states that the error of the ADI method (3.4) has the same order as the implicit Milstein scheme (3.3). We now give a proof as follows.

Proof of Theorem 3.1.2. Let

$$X_{n+1} = C_n\, X_n,$$

where

$$C_n \equiv \exp\Big(-\frac{1}{2}(1-\rho_x)\xi^2 k - \frac{1}{2}(1-\rho_y)\eta^2 k - \mathrm{i}\xi\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}\eta\sqrt{\rho_y k}\widetilde{Z}_{n,y} + e_n\Big),$$

and e_n is the logarithmic error between the numerical solution and the exact solution introduced during $[nk,(n+1)k]$. From (3.22), C_n has the form

$$C_n = \frac{1-\mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + b_x\rho_x k(Z_{n,x}^2-1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}}{(1-a_x k)(1-a_y k)}.$$

In the low wave region, the numerical solutions are close to the exact solutions. We get from Taylor expansion that

$$\begin{aligned} X_N - X(T) = X(T)\cdot\Big\{ &\frac{\mathrm{i}}{6}\sqrt{\rho_x}\xi^3 h_x^2 W_T^x + \frac{1}{24}(1-4\rho_x)\xi^4 h_x^2 T + \frac{\mathrm{i}}{6}\sqrt{\rho_y}\eta^3 h_y^2 \widetilde{W}_T^y \\ &+ \frac{1}{24}(1-4\rho_y)\eta^4 h_y^2 T + \mathrm{i}k\sqrt{k}\sum_{n=0}^{N-1}\widehat{\phi}_1(Z_{n,x},\widetilde{Z}_{n,y}) \\ &+ k^2\sum_{n=0}^{N-1}\widehat{\phi}_2(Z_{n,x},\widetilde{Z}_{n,y}) + o(k,h_x^2,h_y^2)\Big\}. \end{aligned}$$

In the high wave region, we have

$$X_N = X_0\prod_{n=0}^{N-1}\frac{1-\mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + b_x\rho_x k(Z_{n,x}^2-1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}}{(1-a_x k)(1-a_y k)}.$$

Then

$$\begin{aligned} &\lim_{N\to\infty}\mathbb{E}[X_N] = X_0\exp\Big(-\frac{1}{2}\big(\xi^2 u + \eta^2 v + \frac{1}{2}\xi^2\eta^2 uvk + 2\xi\eta\frac{\sin\xi h_x\sin\eta h_y}{\xi h_x\eta h_y}\sqrt{\rho_x\rho_y}\rho_{xy}\big)\,T\Big), \\ &\lim_{N\to\infty}\mathbb{E}[|X_N|^2] \le |X_0|^2\exp\Big(-\frac{1}{2}\xi^2 uT\big(1-\rho_x+\frac{1}{4}\xi^2 uk(1-2\rho_x(1+\rho_{xy}))\big) \\ &\qquad -\frac{1}{2}\eta^2 vT\big(1-\rho_y+\frac{1}{4}\eta^2 vk(1-2\rho_y(1+\rho_{xy}))\big) \\ &\qquad -\frac{1}{4}\xi^2\eta^2 uvkT\big(1-\rho_{xy}(1+\rho_{xy}+3\rho_{xy}^2)\big)\Big), \end{aligned}$$

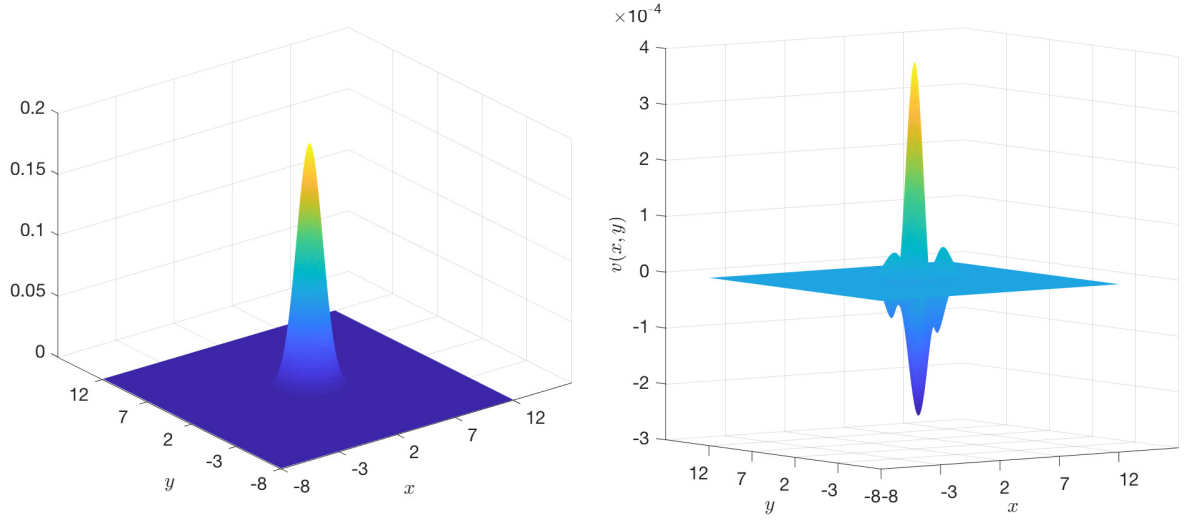
where $u = \sin^2 \frac{h_x \xi}{2} / (\frac{h_x \xi}{2})^2$, $v = \sin^2 \frac{h_y \eta}{2} / (\frac{h_y \eta}{2})^2$. By the same reasoning as for the implicit scheme (3.3), the integration over the high wave region is of higher order than h_x^2 and h_y^2 given condition (3.8). Then the inverse Fourier transform gives the result. □

3.4 Numerical tests

In this section, we illustrate the stability and convergence results from the previous section by way of empirical tests.

Unless stated otherwise, we choose parameters $T = 1$, $x_0 = y_0 = 2$, $\mu_x = \mu_y = 0.0809$, $\rho_x = \rho_y = 0.2$, $\rho_{xy} = 0.45$. For the computations, we truncate the domain to $[-8, 12] \times [-8, 12]$, chosen large enough such that the effect of zero Dirichlet boundary conditions on the solution is negligible.

Figure 3.1(a) shows the numerical solution for one Brownian path, with $h_x = h_y = 2^{-4}$, $k = 2^{-10}$. Figure 3.1(b) plots the pointwise error between the Milstein-ADI approximation (3.4) and the analytic solution (1.15). We also compared the time cost between the implicit scheme (3.3) and ADI scheme (3.4) under the same mesh sizes and parameters, and find that the ADI scheme is about 12 times faster than the pure implicit scheme.



(a) Solution at $t = 1$

(b) Error of the approximation: $V_{i,j}^n - v(t_n, x_i, y_j)$.

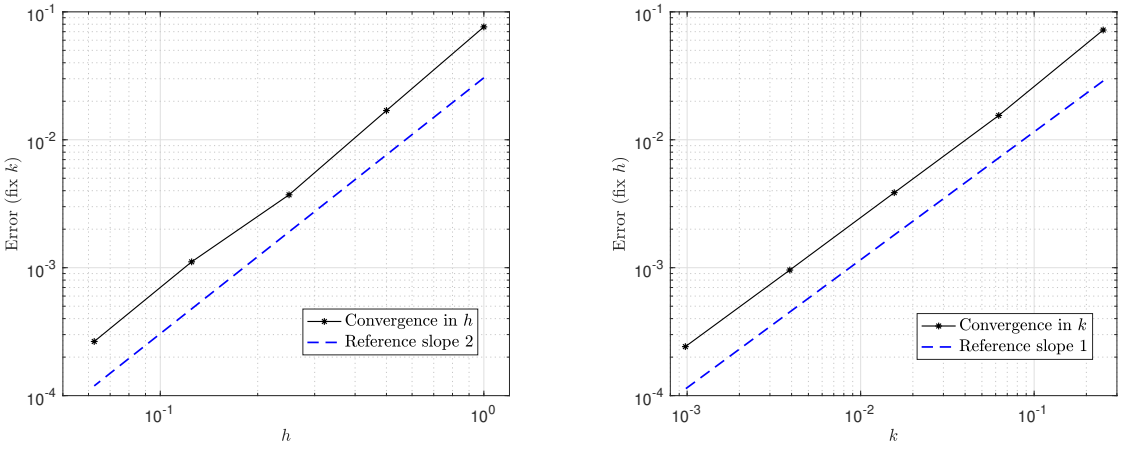
Figure 3.1: Numerical approximation to the SPDE (3.1) with the Milstein-ADI approximation (3.4) for one Brownian path, and the pointwise error to the analytic solution (1.15).

Figure 3.2 verifies the L_2-convergence order in h and k from (3.11). We approximate the error by

$$\begin{aligned}&\Big(\sum_{i,j} h_x h_y E_L\big[|V_{i,j}^N - v(T,x_i,y_j)|^2\big]\Big)^{1/2}\\ &\quad= \Big(\sum_{i,j}\frac{h_x h_y}{L}\sum_{l=1}^{L}|V_{i,j}^N(W^{(l)}) - v(T,x_i,y_j;W^{(l)})|^2\Big)^{1/2},\end{aligned} \tag{3.31}$$

where $W^{(l)}$ are independent Brownian motions and E_L is the empirical mean with $L = 100$ samples.

Here, Figure 3.2(a) shows the convergence in $h = h_x = h_y$ with $k = 2^{-12}$ small enough, which demonstrates second order convergence in h. Figure 3.2(b) shows the convergence in k with $h = h_x = h_y = 2^{-6}$ small enough to ensure sufficient accuracy of the spatial approximation. One can clearly observe first order convergence in k.



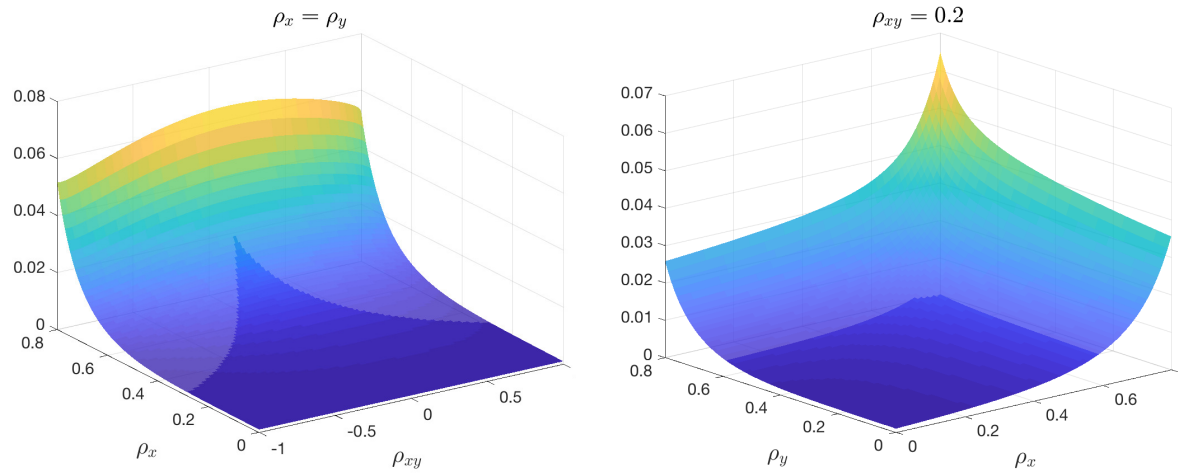
(a) Convergence in h with fixed $k = 2^{-12}$.

(b) Convergence in k with fixed $h = 2^{-6}$.

Figure 3.2: Verifies the L_2-convergence order in h and k from (3.11), with coarsest level $h = 1,\ k = 1/4$ and finest level $h = 2^{-5},\ k = 2^{-10}$, using 100 Monte Carlo samples.

In Figure 3.3, we illustrate the dependence of the approximation error in the L_2-norm on the correlation parameters. The error increases as a function of ρ_x and ρ_y. The error for $\rho_x = \rho_y \leq 0.3$ (see Figure 3.3(a)) varies between roughly 10^{-3} and $3\cdot 10^{-3}$, the error being smallest for $\rho_x = \rho_y = 0$ (the PDE case), and largest for large $\rho_x = \rho_y$ and ρ_{xy} between 0.1 and 0.4. For larger ρ_x and ρ_y, the error increases sharply. The stability region from Assumption 3.1.1 is marked in dark blue, which shows that stable results are obtained even outside the region where mean-square stability is proven. We found problems only for $\rho_x = \rho_y \geq 0.8$. This discrepancy is partly due to the fact that Assumption 3.1.1 is

sufficient, but not necessary, as some of the estimates are not sharp. Figure 3.3(b) shows a similar behaviour when varying ρ_x and ρ_y independently for fixed ρ_{xy}.



(a) Error as function of $\rho_x = \rho_y$, and ρ_{xy}.

(b) Error as function of ρ_x and ρ_y for fixed ρ_{xy}.

Figure 3.3: L_2 error in space as function of correlation parameters for fixed $h_x = h_y = 2^{-3}$ and $k = 2^{-9}$, for a fixed path. The dark blue areas correspond to the stability region from Assumption 3.1.1.

Figure 3.4 shows the singular behaviour of the solution for large k and small h, as predicted by Theorem 3.1.1. Figure 3.5 investigates the behaviour of the error in this regime further, with $h_x = h_y = h$ in Figure 3.5(a), and $h_y = 2^{-1}$ fixed, $h_x = h$ in Figure 3.5(b), letting h go to zero. We calculate the L_2 error in space, and compare different scenarios under the same Brownian path. The top black line shows the error with $k = 2^{-2}$ fixed, and $\rho_x = \rho_y = 0.6$, $\rho_{xy} = 0.1$. We can see that as h goes to zero, the error diverges with rate $h^{-1/2}$ (choosing $h_y = 2^{-1}$ fixed enables more refinements in h_x to show the asymptotic behaviour better). The blue line second from top plots the error with $k = 2^{-4}$ fixed instead. Note that in this case the error will eventually diverge for h going to zero, but this is not visible yet for this level of h. The next red line plots the error for $\rho_x = \rho_y = 0$, with $k = 2^{-2}$ fixed. Then the SPDE (3.1) becomes a PDE and divergence does not appear. Finally, the bottom blue dotted line plots the error for the SPDE (3.1) with initial condition

$$v(0,x,y) = \frac{1}{2\pi\sqrt{(1-\rho_x)(1-\rho_y)}} \exp\Big(-\frac{(x-x_0-\mu_x)^2}{2(1-\rho_x)} - \frac{(y-y_0-\mu_y)^2}{2(1-\rho_y)}\Big). \tag{3.32}$$

For this smooth initial condition, the solution does not diverge for large k and small h. Hence, this verifies that the divergence is a result of the interplay of singular data and stochastic terms only, as shown in Corollary 3.1.2. We emphasise that the instability is

so mild that it is only visible in artificial numerical tests, while for reasonably small k, in particular for $k \sim h^2$ as would be chosen in practice, no instabilities occur.

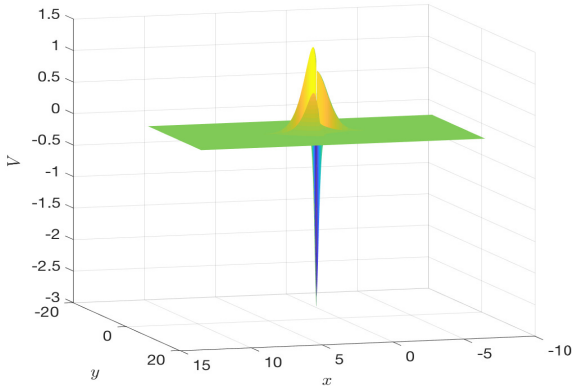
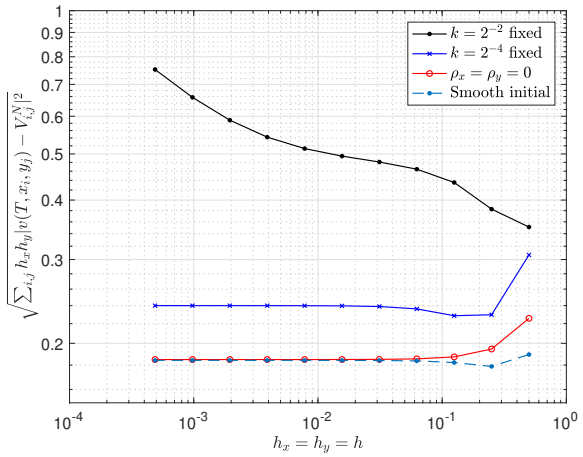
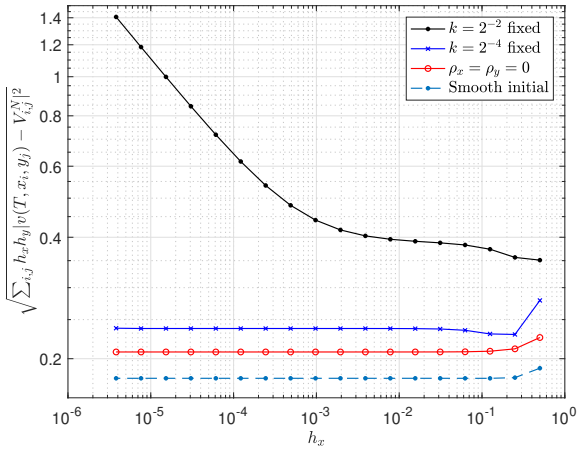


Figure 3.4: Unstable solution for the Milstein-ADI scheme (3.4) for one Brownian path, $k = 2^{-2}$ and $h_x = h_y = 2^{-9}$, $\rho_x = \rho_y = 0.6,\ \rho_{xy} = 0.1$.



(a) L_2 error in space with $h_x = h_y = h \to 0$.



(b) L_2 error in space with $h_y = 2^{-1}$ fixed, and $h_x \to 0$.

Figure 3.5: Comparing the L_2 error in space with fixed k and letting $h \to 0$ under various senarios: $k = 2^{-2}$; $k = 2^{-4}$; $k = 2^{-2}$ and $\rho_x = \rho_y = 0$; $k = 2^{-2}$ and smooth initial condition (3.32), all with a same Brownian path.

3.5 Scheme and test for a variable coefficient SPDE

To approximate the general Zakai SPDE (1.3),

$$\begin{aligned}\mathrm{d}v(t,x) = \left(\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}\big[a_{ij}(x)v(t,x)\big] - \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\big[b_i(x)v(t,x)\big]\right)\mathrm{d}t \\ - \sum_{l=1}^{m}\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\gamma_{i,l}(x)v(t,x))\,\mathrm{d}W_t^k,\end{aligned}$$

by the Milstein scheme, we approximate in the last term

$$v(s) \approx v(t) - \sum_{l=1}^{m}\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\Big(\gamma_{i,l}(x)v(t,x)\Big)\left(W_s^l - W_t^l\right),$$

then

$$\begin{aligned}&-\sum_{l=1}^{m}\sum_{i=1}^{d}\int_t^{t+k}\frac{\partial}{\partial x_i}\Big(\gamma_{i,l}(x)v(s,x)\Big)\,\mathrm{d}W_s^l \\ &= -\sum_{l=1}^{m}\sum_{i=1}^{d}\int_t^{t+k}\frac{\partial}{\partial x_i}\Big(\gamma_{i,l}(x)\Big[v(t) - \sum_{p=1}^{m}\sum_{j=1}^{d}\frac{\partial}{\partial x_j}\Big(\gamma_{j,p}(x)v(t,x)\Big)\left(W_s^p - W_t^p\right)\Big]\Big)\,\mathrm{d}W_s^l \\ &= -\sum_{l=1}^{m}\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\Big(\gamma_{i,l}(x)v(t,x)\Big)\,\Delta W_n^l \\ &\quad + \sum_{l=1}^{m}\sum_{p=1}^{m}\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\Big(\gamma_{il}(x)\sum_{j=1}^{d}\frac{\partial}{\partial x_j}\big(\gamma_{jp}(x)v(t,x)\big)\Big)\int_t^{t+k} W_s^p - W_t^p\,\mathrm{d}W_s^l.\end{aligned}$$

The corresponding ADI implicit Milstein scheme is

$$\begin{aligned}&\prod_{i=1}^{d}\left(I + \frac{k}{2h_i}D_i b_i(X) - \frac{1}{2}\frac{k}{h_i^2}D_{ii}a_{ii}(X)\right)V^{n+1} \\ &= \Bigg\{I + \frac{1}{2}\sum_{i\neq j}\frac{k}{4h_ih_j}D_iD_ja_{ij}(X) - \sum_{l=1}^{m}\Delta W_n^l\sum_{i=1}^{d}\frac{1}{2h_i}D_i\gamma_{i,l}(X) \\ &\quad + \sum_{l=1}^{m}\sum_{p=1}^{m}\Big(\int_{nk}^{(n+1)k}\left(W_s^p - W_t^p\right)\mathrm{d}W_s^l\Big)\sum_{i=1}^{d}\Big(\frac{1}{2h_i}D_i\gamma_{il}(X)\sum_{j=1}^{d}\frac{1}{2h_j}D_j\gamma_{jp}(X)\Big)\Bigg\}V^n.\end{aligned}$$

Here, $\{D_i\}_{1\le i\le d}$ are first order difference operators, and $\{D_{ij}\}_{1\le i,j\le d}$ are second order difference operators, X is the vector of mesh points ordered the same way as V, and, by slight abuse of notation, we denote by $a(X), b(X), \gamma(X)$ the diagonal matrices such that each element of the diagonal corresponds to the function evaluated at the corresponding mesh point.

Notice the presence of an iterated Itô integral $\int_t^{t+k}(W_s^p - W_t^s)\,\mathrm{d}W_s^l$, called Lévy area, is common in multi-dimensional Milstein schemes. It has been proved in [17, 87] that there is no way to achieve a better order of strong convergence than for the Euler scheme by using solely the discrete increments of the driving Brownian motions. Giles and Szpruch [44] proposed an antithetic multilevel Monte Carlo method for high dimensional SDEs avoiding Lévy area simulation. An efficient algorithm for the approximate simulation of the Lévy area has been proposed in [100], building on earlier work in [72, 37] and based on an approximation of the distribution of the tail-sum in a truncated infinite series representation derived from the characteristic functions of these integrals, achieving the complexity $\varepsilon^{-3/2}$ of sampling a single path to obtain a strong error ε. Malham and Wiese proposed another algorithm for the simulation of the Lévy area conditioned on the endpoints, in [83], representing the Lévy area by an infinite weighted sum of independent logistic random variables, further reducing the complexity to $\varepsilon^{-1}|\log(\varepsilon^{-1})|$.

However, the algorithms mentioned above are fairly complex. What we do instead is to further divide the interval $[t, t+k]$ into $O(k^{-1})$ steps and perform a simple Euler approximation to the Lévy area. Note that we always use the timestep k for the SPDE computation. Overall, this still leads to first order convergence in time, and second order in space. To balance the leading order error, the optimal choice is $O(k) = O(h^2) = O(\varepsilon)$. Therefore, the estimate of the Lévy area in each time-step increases the computation time by $O(k^{-1}) = O(\varepsilon^{-1})$ for one step, whereas the matrix calculation for each time step is also $O(\varepsilon^{-1})$, and hence the order of the total complexity does not change.

Moreover, the path simulation including the Lévy areas can be performed separately beforehand using vectorisation, leading to further speed-up.

Now we apply this method to an SPDE from [55],

$$\begin{aligned}\mathrm{d}u = \Big[&\kappa_1 u - \Big(r_1 - \frac{1}{2}y - \xi_1\rho_3\rho_{1,1}\rho_{2,1}\Big)\frac{\partial u}{\partial x} - \Big(\kappa_1(\theta_1 - y) - \xi_1^2\Big)\frac{\partial u}{\partial y} + \frac{1}{2}y\frac{\partial^2 u}{\partial x^2}\\ &+ \xi_1\rho_3\rho_{1,1}\rho_{2,1}y\frac{\partial^2 u}{\partial x\partial y} + \frac{\xi_1^2}{2}y\frac{\partial^2 u}{\partial y^2}\Big]\,\mathrm{d}t - \rho_{1,1}\sqrt{y}\frac{\partial u}{\partial x}\,\mathrm{d}W_t - \xi_1\rho_{2,1}\frac{\partial}{\partial y}(\sqrt{y}u)\,\mathrm{d}B_t,\end{aligned} \tag{3.33}$$

with Dirac initial $u(0, x, y) = \delta(x - x_0) \otimes \delta(y - y_0)$. This SPDE models the limit empirical measure of a large portfolio of defaultable assets in which the asset value processes are modelled by Heston-type stochastic volatility models with common and idiosyncratic factors in both the asset values and the variances, and default is triggered by hitting a lower boundary.

Similar to before, we implement the SPDE (3.33) with a Milstein ADI scheme as follows:

$$\begin{aligned} A_x\,A_y\,U^{n+1} \;=\; & \Big((1+\kappa_1\,k)I - \frac{k}{4h_x}\xi_1\rho_{1,1}\rho_{2,1}\rho_3 D_x - \frac{\sqrt{k}Z_{n,x}}{2h_x}\rho_{1,1}\sqrt{Y}D_x \\ & + \frac{k(Z_{n,x}^2-1)}{8h_x^2}\rho_{1,1}^2 Y D_x^2 + \frac{k}{4h_x}\xi_1\rho_{1,1}\rho_{2,1}Z_{n,x}\widetilde{Z}_{n,y}\Big(D_x + \frac{1}{h_y}YD_{xy}\Big) \\ & + \frac{1}{4h_x}\xi_1\rho_{2,1}\rho_{1,1}\Big(\int_t^{t+k}(W_s - W_t)\,\mathrm{d}B_s\Big)D_x\Big)U^n \\ & - \frac{\sqrt{k}\widetilde{Z}_{n,y}}{2h_y}\xi_1\rho_{2,1}D_y(\sqrt{Y}U^n) + \frac{k(\widetilde{Z}_{n,y}^2-1)}{8h_y^2}\xi_1^2\rho_{2,1}^2 D_y\Big(\sqrt{Y}\big(D_y(\sqrt{Y}U^n)\big)\Big), \end{aligned} \tag{3.34}$$

where the notation for Y follows the same principle as above for X, $\widetilde{Z}_n^y = \rho_{xy}Z_n^x + \sqrt{1-\rho_{xy}^2}Z_n^y$, with $Z_n^x, Z_n^y \sim N(0,1)$ being independent normal random variables, and

$$\begin{aligned} A_x &= I + \frac{k}{2h_x}\Big(r_1 - \xi_1\rho_3\rho_{1,1}\rho_{2,1}\Big)D_x - \frac{k}{4h_x}YD_x - \frac{k}{2h_x^2}YD_{xx}, \\ A_y &= I + \frac{k}{2h_y}(\kappa_1\theta_1 - \xi_1^2)D_y - \frac{k}{2h_y}\kappa_1 YD_y - \frac{k}{2h_y^2}\xi_1^2 YD_{yy}. \end{aligned}$$

Furthermore, we also implement the SPDE (3.33) with coefficients frozen at (x_0, y_0), i.e.,

$$\begin{aligned} \mathrm{d}u = & \Big(\kappa_1 u - \Big(r_1 - \frac{1}{2}y_0 - \xi_1\rho_3\rho_{1,1}\rho_{2,1}\Big)\frac{\partial u}{\partial x} - \Big(\kappa_1(\theta_1 - y_0) - \xi_1^2\Big)\frac{\partial u}{\partial y} + \frac{1}{2}y_0\frac{\partial^2 u}{\partial x^2} \\ & + \xi_1\rho_3\rho_{1,1}\rho_{2,1}y_0\frac{\partial^2 u}{\partial x\partial y} + \frac{\xi_1^2}{2}y_0\frac{\partial^2 u}{\partial y^2}\Big)\,\mathrm{d}t \\ & - \rho_{1,1}\sqrt{y_0}\frac{\partial u}{\partial x}\,\mathrm{d}W_t - \xi_1\rho_{2,1}\frac{\partial}{\partial y}(\sqrt{y_0}u)\,\mathrm{d}B_t, \end{aligned} \tag{3.35}$$

with Dirac initial datum $u(0,x,y) = \delta(x-x_0)\otimes\delta(y-y_0)$. We can apply a Milstein ADI scheme similar to (3.3),

$$\begin{aligned} & \Big(I + \frac{k}{2h_x}\Big(r_1 - \xi_1\rho_3\rho_{1,1}\rho_{2,1} - \frac{1}{2}y_0\Big)D_x - \frac{k}{2h_x^2}y_0 D_{xx}\Big) \\ & \cdot\Big(I + \frac{k}{2h_y}(\kappa_1(\theta_1 - y_0) - \xi_1^2)D_y - \frac{k}{2h_y^2}\xi_1^2 y_0 D_{yy}\Big)U^{n+1} \\ = & \Big((1+\kappa_1\,k)I - \frac{\sqrt{k}Z_{n,x}}{2h_x}\rho_{1,1}\sqrt{y_0}D_x + \frac{k(Z_{n,x}^2-1)}{8h_x^2}\rho_{1,1}^2 y_0 D_x^2 \\ & - \frac{\sqrt{k}\widetilde{Z}_{n,y}}{2h_y}\xi_1\rho_{2,1}\sqrt{y_0}D_y + \frac{k(\widetilde{Z}_{n,y}^2-1)}{8h_y^2}\xi_1^2\rho_{2,1}^2 y_0 D_y^2 \\ & + \frac{k\,Z_{n,x}\widetilde{Z}_{n,y}}{4h_x h_y}\xi_1\rho_{1,1}\rho_{2,1}D_{xy}\Big)U^n. \end{aligned} \tag{3.36}$$

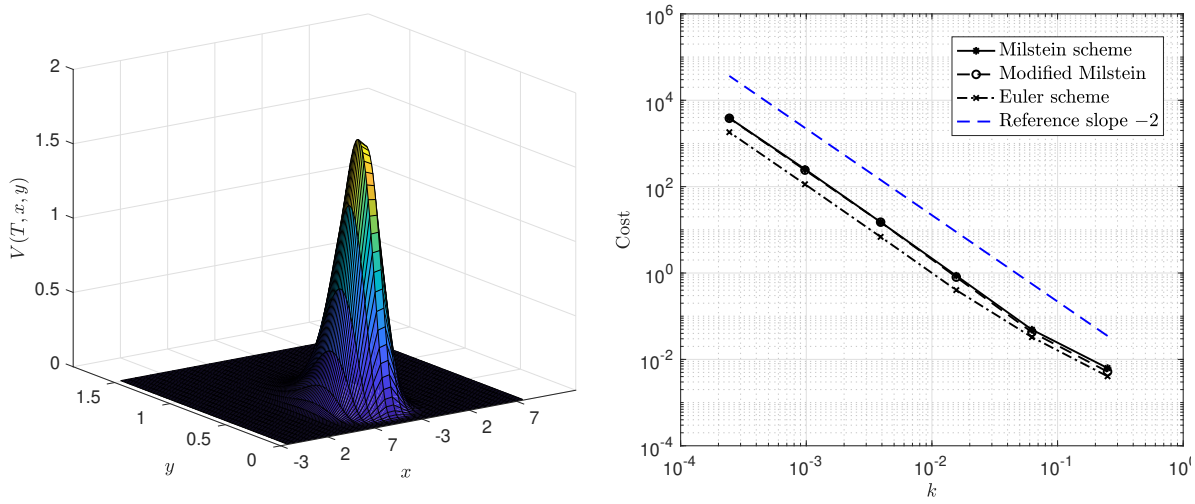
We will test the behaviour of (3.35) at (x_0, y_0) and compare with the SPDE (3.33).

We choose parameters $T = 1, x_0 = 2, y_0 = 1.4, r = 0.05, \xi = 0.5, \theta = 0.4, \kappa = 2, \rho_{1,1} = 0.3$, $\rho_{2,1} = 0.2$, $\rho_3 = 0.5$. We truncate the domain to $[-3, 7] \times [0, 1.5]$ for the scheme (3.34), and $[-3, 7] \times [-1, 1.5]$ for the scheme (3.36), sufficiently large in this setting.

Figure 3.6(a) shows the density for a single Brownian path, with $k = 2^{-8}$, $h_x = 5/16$, and $h_y = 1/80$.

Figure 3.6(b) compares the computational cost under different time-stepping schemes: the Milstein scheme, a "modified" Milstein scheme, and the Euler scheme. Here, for the Milstein scheme we approximate the Lévy area by sub-timestepping as explained above, while in the "modified" Milstein scheme we drop $\int_t^{t+k}(W_s - W_t)\,\mathrm{d}B_s$ but keep the one-dimensional iterated integrals as they are known analytically. We expect that the latter will lead to a worse convergence in time (for non-zero $\xi_1, \rho_{2,1}, \rho_{1,1}$), which is verified in Figure 3.7(b).

In Figure 3.6(b), from a coarsest mesh with $h_x = 0.625$, $h_y = 0.025$, and $k = 0.25$, we keep decreasing the time-step k by a factor of 4, and the spatial mesh width by a factor of 2. This shows that the cost, measured by time elapsed in simulating one path, increases by a factor of 16, demonstrating that these three schemes result in the same order of complexity.



(a) Sample density, $k = 2^{-8}$, $h_x = 5/16$, $h_y = 1/80$. (b) Cost in k with $h_x = \frac{5}{4}k^{1/2}$ and $h_y = \frac{1}{20}k^{1/2}$.

Figure 3.6: Single path realisation of the density to the SPDE (3.33) and associated numerical cost (CPU time in sec).

Figure 3.7 verifies the L_2 convergence order in h and k. In absence of an exact solution,

we compute a proxy to the error in $h_x,\ h_y$ by

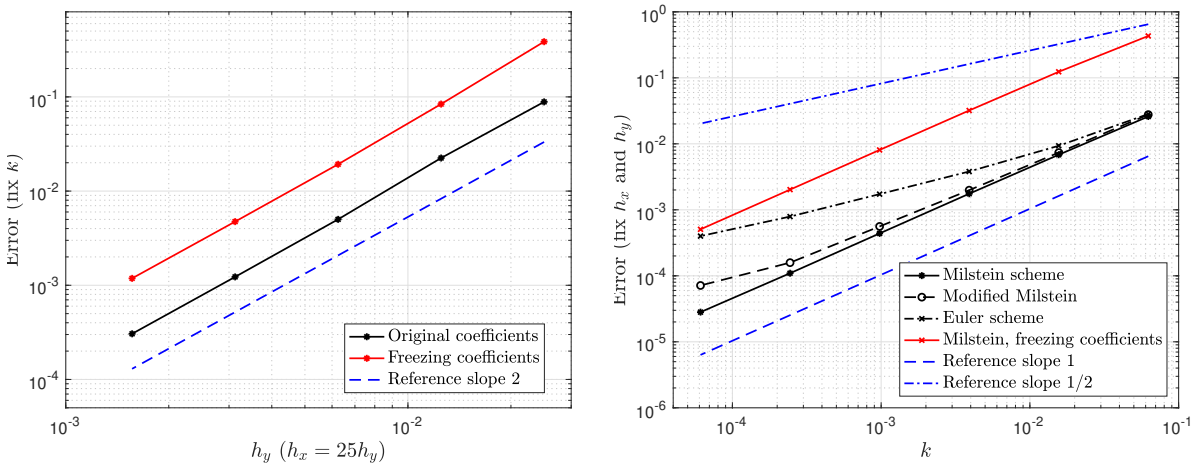
$$\Big(\sum_{i,j} h_x h_y E_L\Big[\Big|V^N_{2i,2j}\Big(k,\frac{h_x}{2},\frac{h_y}{2}\Big) - V^N_{i,j}(k,h_x,h_y)\Big|^2\Big]\Big)^{1/2}$$
$$= \Big(\sum_{i,j}\frac{h_x h_y}{L}\sum_{l=1}^{L}\Big|V^N_{2i,2j}\Big(k,\frac{h_x}{2},\frac{h_y}{2};W^{(l)},B^{(l)}\Big) - V^N_{i,j}(k,h_x,h_y;W^{(l)},B^{(l)})\Big|^2\Big)^{1/2},$$

where $V^N_{i,j}(k,h_x,h_y)$ is the numerical approximation to $v(T,ih_x,jh_y)$ with mesh widths h_x and h_y, $V^N_{2i,2j}(k,h_x/2,h_y/2)$ uses a fine mesh with $h_x/2$ and $h_y/2$, and E_L is the empirical mean with L samples of the Brownian path W. Both share the same Brownian path and same time step k, thus the univariate error in k cancels and we should see the correct convergence order in h_x and h_y. We could alternatively solve on a very fine mesh and treat the result as the exact solution to test the convergence as in (5.1). It is clear from the construction that the observed convergence order will be the same. However, since the computation of a highly accurate solution is extremely expensive, and such a numerical reference solution will introduce a bias, we do not perform this test here. Here, Figure 3.7(a) shows the convergence in $h = h_x = h_y$ with $k = 2^{-4}$ fixed and $L = 100$, which demonstrates second order convergence in h.

Similarly, we study the error in k by using the difference between two solutions with same mesh size, same Brownian path, but different time-steps,

$$\Big(\sum_{i,j} h_x h_y E_L\big[|V^N_{i,j}(k,h_x,h_y) - V^{2N}_{i,j}(k/2,h_x,h_y)|^2\big]\Big)^{1/2}$$
$$= \Big(\sum_{i,j}\frac{h_x h_y}{L}\sum_{l=1}^{L}\Big|V^N_{i,j}\Big(k,h_x,h_y;W^{(l)},B^{(l)}\Big) - V^{2N}_{i,j}(k/2,h_x,h_y;W^{(l)},B^{(l)})\Big|^2\Big)^{1/2}.$$

Figure 3.7(b) shows the convergence in k with fixed $h_x = 5/8,\ h_y = 1/40$, under the Milstein scheme, "modified" Milstein scheme, and Euler scheme. The timestep k decreases by a factor of 4 from one level to the next. We also plot two blue dashed lines with slope $1/2$ and 1 as reference. One can clearly observe first order convergence in k for the Milstein scheme, and half order convergence in k for the Euler scheme. As for the "modified" Milstein scheme, although it appears to converge with first order on coarse levels (due to dominance of the terms which converge with first order for this level of accuracy), the asymptotic order is seen to be lower. However, the benefit of the "modified" Milstein scheme is that it does not require the Lévy area simulation, and at the same time has a smaller error than the Euler scheme for the same mesh sizes.



(a) Convergence in h with fixed $k = 2^{-4}$.

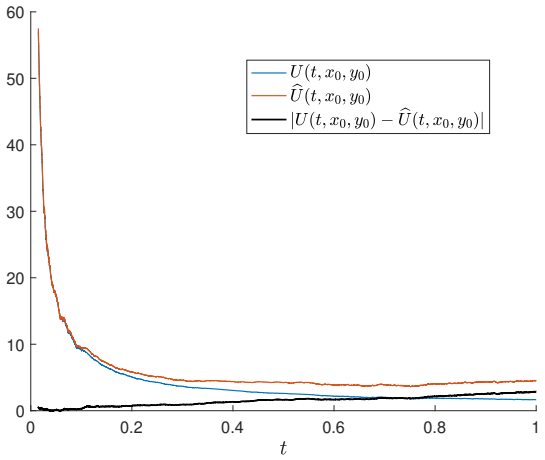
(b) Convergence in k with fixed $h_x = \frac{5}{8}$, $h_y = \frac{1}{40}$.

Figure 3.7: Test the convergence order in (h_x, h_y) and k for the schemes (3.34) and (3.36), in L_2 in space and probability, using 100 paths.

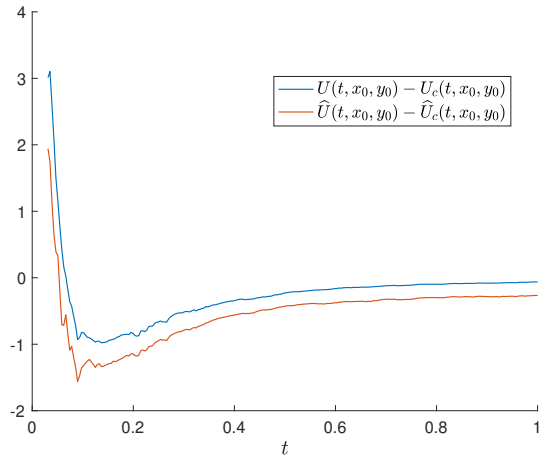
In Figure 3.7, we also compare the convergence of the scheme (3.36) for the frozen coefficient SPDE (3.35) with the scheme (3.34) for the SPDE (3.33). We find in the following that the behaviour of two solutions at (x_0, y_0) is extremely close, especially for small time intervals. Hence, even though the analysis does not apply to the SPDE (3.33), the numerical scheme applies, and we can also study the local behaviour at (x_0, y_0) via (3.35), for which we have an exact error analysis. We can see that the scheme (3.36) has second order convergence in h, and first order convergence in k, as expected.

We now illustrate more carefully the behaviour at (x_0, y_0) over time. Figure 3.8(a) shows one path realisation of the solution $U(t, x_0, y_0)$ with the scheme (3.34) as a function of t, and $\widehat{U}(t, x_0, y_0)$ with the scheme (3.36). The mesh sizes and time step are $h_x = 5 \times 2^{-8}$, $h_y = 0.2 \times 2^{-8}$, $k = 4^{-8}$. Note that the plots below start from $t = 0.015$ in Figure 3.8(a) and $t = 0.03$ in Figure 3.8(b) rather than zero, because of the Dirac initial datum, however, even under this singular initial condition, $U(t, x_0, y_0)$ and $\widehat{U}(t, x_0, y_0)$ are very close over an initial time period. Figure 3.8(b) shows the error of the schemes (3.34) and (3.36) for one realisation of the path, where we use $h_x = 5/16$, $h_y = 1/80$, $k = 1/256$ for "coarse" solutions U_c and $\widehat{U}_c$, and we treat the fine approximations U and $\widehat{U}$ as the exact solution. The errors for the SPDEs (3.33) and (3.35) track each other closely.

In Figure 3.9, we use another approach to show the error by taking the difference between the solutions with mesh sizes and time steps (h_x, h_y, k) and $(h_x/2, h_y/2, k/4)$. We show the error both for one path, as well as L_2 in probability $(E_L[|U_f(t, x_0, y_0) - U_c(t, x_0, y_0)|^2])^{1/2}$, where E_L is the empirical mean with $L = 100$ Monte Carlo samples.

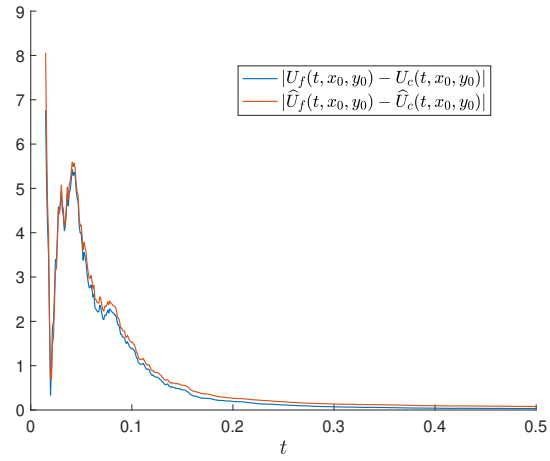


(a) Solutions U from (3.33) and $\widehat{U}$ from (3.35) at (x_0, y_0), with $h_x = 5\times2^{-8}$, $h_y = 0.2\times2^{-8}$, $k = 4^{-8}$, and t from 0.015.

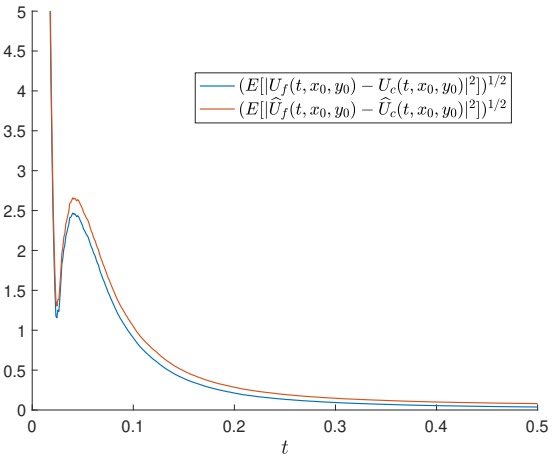


(b) Numerical error for U_c from (3.33) and $\widehat{U}_c$ from (3.35) at (x_0, y_0), with $h_x = \frac{5}{16}$, $h_y = \frac{1}{80}$, $k = \frac{1}{256}$, and t from 0.03.

Figure 3.8: Single path realisation of the solution at (x_0, y_0) over time, of SPDE (3.33) with variable coefficients and (3.35) with frozen coefficients.



(a) Numerical error at (x_0, y_0) between (3.33) and (3.35), using a single path, and t from 0.015.



(b) Numerical error at (x_0, y_0) between (3.33) and (3.35), taking the L_2 average, and t from 0.03.

Figure 3.9: Numerical error at (x_0, y_0) by taking difference between (h_x, h_y, k) and $(h_x/2, h_y/2, k/4)$ to the schemes (3.34) and (3.36), where $h_x = \frac{5}{32}$, $h_y = \frac{1}{160}$, $k = \frac{1}{1024}$.

3.6 Summary

We studied a two-dimensional parabolic SPDE arising from a filtering problem. We proved mean-square stability and convergence for a semi-implicit Milstein discretisation scheme in terms of L_2 in probability, and pointwise as well as L_2 in space (as defined previously). To reduce the complexity, we also implemented an ADI version of the scheme, and provided corresponding convergence results.

Chapter 4

Sparse combination with multilevel estimators

In this chapter, we analyse the accuracy and computational complexity of estimators for linear functionals of the solution to multidimensional parabolic stochastic partial differential equations (SPDE) of Zakai-type. Here, we use the Milstein scheme for time integration and an alternating direction implicit (ADI) splitting of the spatial finite difference discretisation, coupled with the sparse combination technique and multilevel Monte Carlo sampling (MLMC). In the two-dimensional case, we find by detailed Fourier analysis that for a root-mean-square error (RMSE) ε, MLMC on sparse combination has the optimal complexity $O(\varepsilon^{-2})$, whereas MLMC on regular grids has $O(\varepsilon^{-2}(\log \varepsilon)^2)$, standard MC on sparse combination $O(\varepsilon^{-7/2}(|\log \varepsilon|)^{5/2})$, and MC on regular grids $O(\varepsilon^{-4})$. Numerical tests confirm these findings empirically. We give a discussion of the higher-dimensional setting without detailed proofs, which suggests that MLMC with sparse combination always leads to the optimal complexity, standard MC with sparse combination has a fixed complexity order independent of the dimension (up to a logarithmic term), whereas the cost of MLMC and MC on regular grids increases exponentially with the dimension.

The rest of this chapter is structured as follows. We define the approximation schemes in Section 4.1. Section 4.2 shows the Fourier analysis of sparse combination estimators. Section 4.4 generalises the problem to higher dimensions, and Section 4.3 shows numerical experiments confirming the above findings. Section 4.5 offers summaries and directions for further research.

4.1 Approximation and main results

For simplicity of presentation, we initially restrict ourselves with the description of the schemes and their analysis to the two-dimensional case. The extension to higher dimensions is discussed in Section 4.4.

Moreover, we focus on the case of constant coefficients as in (1.13) and the functional (1.18). While the numerical method itself is directly applicable to the variable coefficient case and more general functionals, the Fourier analysis we perform is tailored to the present setting.

We repeat the numerical scheme to the SPDE (1.13), repeated here for convenience,

$$\begin{aligned} \mathrm{d}v = &-\mu_x \frac{\partial v}{\partial x}\,\mathrm{d}t - \mu_y \frac{\partial v}{\partial y}\,\mathrm{d}t + \frac{1}{2}\left(\frac{\partial^2 v}{\partial x^2} + 2\sqrt{\rho_x\rho_y}\rho_{xy}\frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 v}{\partial y^2}\right)\mathrm{d}t \\ &- \sqrt{\rho_x}\frac{\partial v}{\partial x}\,\mathrm{d}W_t^x - \sqrt{\rho_y}\frac{\partial v}{\partial y}\,\mathrm{d}W_t^y, \end{aligned} \tag{4.1}$$

for $x, y \in \mathbb{R},\ 0 < t \leq T$, subject to the Dirac initial datum

$$v(0, x, y) = \delta(x - x_0) \otimes \delta(y - y_0), \tag{4.2}$$

for given x_0 and y_0.

4.1.1 Semi-implicit Milstein finite difference scheme

We use a spatial grid with uniform spacing $h_x,\ h_y > 0$, and let $V_{i,j}^n$ be the approximation to $v(nk, ih_x, jh_y)$, $n = 1, \ldots, N,\ i, j \in \mathbb{Z}$. We assume for simplicity that $i_0 := x_0/h_x$ and $j_0 := y_0/h_y$ are integers. Then we approximate $v(0, x, y) = \delta(x - x_0) \otimes \delta(y - y_0)$ by

$$V_{i,j}^0 = h_x^{-1} h_y^{-1} \delta_{(i_0, j_0)} = \begin{cases} h_x^{-1} h_y^{-1}, & i = i_0,\ j = j_0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

We use the semi-implicit Milstein scheme to approximate (4.1), proposed in Chapter 3:

$$\begin{aligned} &\left(I + \frac{\mu_x k}{2h_x}D_x + \frac{\mu_y k}{2h_y}D_y - \frac{k}{2h_x^2}D_{xx} - \frac{k}{2h_y^2}D_{yy}\right)V^{n+1} \\ &\qquad = \left(I - \frac{\sqrt{\rho_x k}Z_{n,x}}{2h_x}D_x - \frac{\sqrt{\rho_y k}\widetilde{Z}_{n,y}}{2h_y}D_y + \frac{\rho_x k(Z_{n,x}^2 - 1)}{8h_x^2}D_x^2 \right. \\ &\qquad\qquad \left. + \frac{\rho_y k(\widetilde{Z}_{n,y}^2 - 1)}{8h_y^2}D_y^2 + \frac{\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}}{4h_x h_y}D_{xy}\right)V^n \\ &\qquad := M_n V^n, \end{aligned} \tag{4.4}$$

where M is a random operator and

$$
\begin{aligned}
&(D_x V)_{i,j} = V_{i+1,j} - V_{i-1,j}, && (D_y V)_{i,j} = V_{i,j+1} - V_{i,j-1},\\
&(D_{xx} V)_{i,j} = V_{i+1,j} - 2V_{i,j} + V_{i-1,j}, && (D_{yy} V)_{i,j} = V_{i,j+1} - 2V_{i,j} + V_{i,j-1},\\
&(D_{xy} V)_{i,j} = V_{i+1,j+1} - V_{i-1,j+1} - V_{i+1,j-1} + V_{i-1,j-1},
\end{aligned}
$$

and $\widetilde{Z}_n^y = \rho_{xy} Z_n^x + \sqrt{1-\rho_{xy}^2} Z_n^y$ with $Z_n^x, Z_n^y \sim N(0,1)$ independent normals. Briefly, the terms on the left-hand side of (4.4) correspond to the implicit approximation of the drift term in (4.1); the second and third terms on the right-hand side are the Euler-Maruyama approximation of the stochastic integral; and the last three terms the Milstein correction for strong first order 1. Note that the cross-derivative term on the left-hand side cancels out with a Milstein term, as detailed in Chapter 3.

To save computational cost, we combine the scheme with an alternating direction implicit (ADI) factorisation [88],

$$
\left(I + \frac{\mu_x k}{2h_x} D_x - \frac{k}{2h_x^2} D_{xx}\right)\left(I + \frac{\mu_y k}{2h_y} D_y - \frac{k}{2h_y^2} D_{yy}\right) V^{n+1} = M_n V^n. \tag{4.5}
$$

A detailed analysis of the L_2 stability and convergence of these schemes can be found in Chapter 3. Here, we state the main result which is relevant here. We make the following technical assumptions throughout[1]:

$$
2\rho_x^2(1+2|\rho_{xy}|) \le 1, \quad 2\rho_y^2(1+2|\rho_{xy}|) \le 1, \quad 2\rho_x\rho_y(3\rho_{xy}^2 + 2|\rho_{xy}| + 1) \le 1. \tag{4.6}
$$

First, we repeat Corollary 3.1.2 as a theorem needed in this chapter, for the special case of k satisfying a CFL condition.

Theorem 4.1.1. *Let $\lambda > 0$ be fixed and (4.6) be satisfied. Then there exists $C > 0$ such that for any k, h_x, h_y with*

$$
k/\min\{h_x^2, h_y^2\} \le \lambda, \tag{4.7}
$$

for the solution to the implicit Milstein scheme (4.4),

$$
\sqrt{\mathbb{E}\left[|V_{i,j}^N - v(T, ih_x, jh_y)|^2\right]} \le C(h_x^2 + h_y^2).
$$

This convergence result also holds for the ADI scheme (4.5).

[1] They are used in the proof of Lemma 4.2.3.

4.1.2 Sparse combination estimators

Now we consider the specific functional

$$P_T(v) = \int_0^\infty \int_0^\infty v(T, x, y)\,\mathrm{d}x\,\mathrm{d}y, \tag{4.8}$$

as discussed in the introduction, where v is the solution to (4.1) and (4.2).

By introducing integer multi-indices $\boldsymbol{l} = (l_1, l_2) \in \mathbb{N}^2$ as the refinement levels of the spatial mesh in dimensions x and y, respectively, we denote by $P^N_{(l_1,l_2)}$ the discrete approximations to P with mesh sizes $h_x = h_0 \cdot 2^{-l_1}$, $h_y = h_0 \cdot 2^{-l_2}$, and fixed timestep k with $k \leq \lambda \min\{h_x^2, h_y^2\}$. Let $N = T/k$. We then use the trapezoidal approximation

$$P^N_{(l_1,l_2)} = h_x h_y \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} V^N_{i,j} + \frac{h_x}{2}\sum_{j=1}^{\infty} V^N_{0,j} + \frac{h_y}{2}\sum_{i=1}^{\infty} V^N_{i,0} + \frac{1}{4}h_x h_y V^N_{0,0}, \tag{4.9}$$

where $V^N_{i,j}$ is the solution to (4.4).

Proposition 4.1.1. *Assume* (4.6) *holds. Let* P_T *be given by* (4.8) *and* $P^N_{(l_1,l_2)}$ *by* (4.9). *Then for fixed* $\lambda > 0$ *there exists* $C > 0$ *such that for any* $(l_1, l_2) \in \mathbb{N}_0^2$, $k \leq \lambda \min\{h_x^2, h_y^2\}$,

$$\sqrt{\mathbb{E}\left[|P^N_{(l_1,l_2)} - P_T|^2\right]} \leq C(h_x^2 + h_y^2), \tag{4.10}$$

where $h_x = h_0 \cdot 2^{-l_1}$, $h_y = h_0 \cdot 2^{-l_2}$, $N = T/k$.

Proof. Similar to Proposition 2.1.2. □

Let Δ_i be the first-order difference operator along directions $i = 1, 2$, defined as

$$\Delta_i P^N_{\boldsymbol{l}} = \begin{cases} P^N_{\boldsymbol{l}} - P^N_{\boldsymbol{l}-\boldsymbol{e}_i}, & \text{if } l_i > 0, \\ P^N_{\boldsymbol{l}} & \text{if } l_i = 0, \end{cases} \tag{4.11}$$

with $\boldsymbol{e}_i$ the canonical unit vectors in $\mathbb{R}^2$, i.e., $(\boldsymbol{e}_i)_j = \delta_{ij}$, and $\boldsymbol{l} = (l_1, l_2)$. We also define the first-order mixed difference operator $\Delta = \Delta_1 \otimes \Delta_2$. Hence, for $l_1 > 0, l_2 > 0$,

$$\Delta P^N_{(l_1,l_2)} = P^N_{(l_1,l_2)} - P^N_{(l_1,l_2-1)} - P^N_{(l_1-1,l_2)} + P^N_{(l_1-1,l_2-1)}. \tag{4.12}$$

We will prove in Section 4.2.3 the following theorem.

Theorem 4.1.2. *Assume* (4.6) *holds. Let* $h_x = h_0 \cdot 2^{-l_1}$, $h_y = h_0 \cdot 2^{-l_2}$, *and for any fixed* $\lambda > 0$, $k/\min\{h_x^2, h_y^2\} \leq \lambda$, $N = T/k$. *Then for* $\Delta P^N_{(l_1,l_2)}$ *from* (4.12),

$$\left|\mathbb{E}\left[\Delta P^N_{(l_1,l_2)}\right]\right| = O(h_x^2 h_y^2), \qquad \mathbb{E}\left[\left|\Delta P^N_{(l_1,l_2)}\right|^2\right] = O(h_x^4 h_y^4). \tag{4.13}$$

Proof. See Section 4.2.3. □

Remark 4.1.1. *We emphasise that* (4.13) *does not follow from* (4.10), *but ascertains essentially that the difference operator cancels out any terms in the error expansion which depend on* h_x *or* h_y *alone. Establishing this property, which depends on the regularity of the problem, is the crucial step for the application of the sparse combination technique in the deterministic case of PDEs (see [48, 12, 91]). It is precisely the condition required for the Multi-Index Monte Carlo method of [54]. Indeed, the approach here is interpretable as a specific multi-index decomposition applied to the spatial variables, treating time separately.*

Corollary 4.1.1. *With the same setting as in Theorem 4.1.2, the first order differences of* $P^N_{(l_1,l_2)}$ *derived from* (4.9) *have the first and second moments*

$$\left|\mathbb{E}\left[\Delta_1 P^N_{(l_1,l_2)}\right]\right| = O(h_x^2), \qquad \mathbb{E}\left[\left|\Delta_1 P^N_{(l_1,l_2)}\right|^2\right] = O(h_x^4),$$
$$\left|\mathbb{E}\left[\Delta_2 P^N_{(l_1,l_2)}\right]\right| = O(h_y^2), \qquad \mathbb{E}\left[\left|\Delta_2 P^N_{(l_1,l_2)}\right|^2\right] = O(h_y^4).$$

Following the ideas in [12, 54], given a sequence of index sets $\mathcal{I}_l = \{(l_1, l_2) \in \mathbb{N}_0^2 : l_1 + l_2 \leq l + 1\}$, the approximation on level l is defined as

$$P_l^N = \sum_{(l_1,l_2)\in\mathcal{I}_l} \Delta P^N_{(l_1,l_2)}. \tag{4.14}$$

Note that we use the same k for all $\Delta P^N_{(l_1,l_2)}$, $(l_1, l_2) \in \mathcal{I}_l$. We have the following.

Theorem 4.1.3. *Assume* (4.6) *holds. Let* $k \leq \lambda\, h_0^2\, 2^{-2l}$ *for a fixed* $\lambda > 0$, $N = T/k$, P *given by* (4.8) *and* P_l^N *given by* (4.14). *Then*

$$\sqrt{\mathbb{E}\left[|P_l^N - P_T|^2\right]} = O(l\, 2^{-2l}). \tag{4.15}$$

Proof. See Section 4.2.4. □

The standard Monte Carlo estimator for $\mathbb{E}\big[P_l^N\big]$ using M samples is defined as

$$\widehat{Y} = \frac{1}{M}\sum_{m=1}^{M}\sum_{(l_1,l_2)\in\mathcal{I}_l}\Big(\Delta\widehat{P}^N_{(l_1,l_2)}\Big)^{(m)}, \tag{4.16}$$

where $\Big(\Delta\widehat{P}^N_{(l_1,l_2)}\Big)^{(m)}$ is the m-th sample for the difference on spatial grid level $(l_1,\, l_2)$ of the SPDE approximation using N time steps.

To reduce the bias below ε, we can choose $l = \big[\frac{1}{2}(-\log_2\varepsilon + \log_2|\log\varepsilon|)\big]$, where $[\,\cdot\,]$ is the closest integer. Since $k \le \lambda h_0^2 2^{-2l} = O(2^{-2l})$, the computational cost for one sample of $P^N_{(l_1,l_2)}$ is

$$W = \sum_{l_1+l_2\le l+1} 2^{l_1+l_2}\cdot O(k^{-1}) = O(l\,2^{3l}) = O\big(\varepsilon^{-\frac{3}{2}}|\log\varepsilon|^{\frac{5}{2}}\big).$$

Using standard Monte Carlo sampling, we need $M = O(\varepsilon^{-2})$ samples to reduce the variance below ε^2, hence the total computational cost to achieve a RMSE ε is

$$W = O\big(\varepsilon^{-\frac{7}{2}}|\log\varepsilon|^{\frac{5}{2}}\big).$$

4.1.3 Sparse combination MLMC estimators

Next, we combine the sparse grid combination technique with the MLMC method. We will show that this combination leads to the optimal complexity $O(\varepsilon^{-2})$.

Let $P_l \coloneqq P_l^{N_l}$ be an approximation to P as in (4.14) using a numerical discretisation with timestep k_l and index set $\mathcal{I}_l$, where

$$\mathcal{I}_l = \{(l_1,\, l_2)\in\mathbb{N}_0^2 :\, l_1+l_2\le l+1\}, \qquad k_l = k_0\cdot 2^{-2l}, \qquad N_l = \frac{T}{k_l}.$$

For $(l_1,\, l_2)\in\mathcal{I}_l$, the mesh sizes are $h_x^{(l_1)} = h_0\cdot 2^{-l_1}$, $h_y^{(l_2)} = h_0\cdot 2^{-l_2}$. By using different levels of refinement, we have the following identity:

$$\mathbb{E}[P_T] = \mathbb{E}[P_0] + \sum_{l=1}^{\infty}\mathbb{E}[P_l - P_{l-1}].$$

Then we define the difference operators (acting on the level)

$$\delta P_l = P_l - P_{l-1} = \sum_{(l_1,l_2)\in\mathcal{I}_l}\Delta P^{N_l}_{(l_1,l_2)} - \sum_{(l_1,l_2)\in\mathcal{I}_{l-1}}\Delta P^{N_{l-1}}_{(l_1,l_2)}, \quad l\ge 0, \tag{4.17}$$

where we denote $P_{-1} \coloneqq 0$. Thus, the approximation $\mathbb{E}[P_{l^*}]$ to $\mathbb{E}[P_T]$ at level l^* is

$$\sum_{l=0}^{l^*} \mathbb{E}[\delta P_l] = \mathbb{E}\Big[\sum_{(l_1,l_2)\in\mathcal{I}_0} \Delta P^{N_0}_{(l_1,l_2)}\Big] + \sum_{l=1}^{l^*} \mathbb{E}\Big[\sum_{(l_1,l_2)\in\mathcal{I}_l} \Delta P^{N_l}_{(l_1,l_2)} - \sum_{(l_1,l_2)\in\mathcal{I}_{l-1}} \Delta P^{N_{l-1}}_{(l_1,l_2)}\Big].$$

Therefore, instead of directly simulating P_{l^*}, we simulate δP_l, $l = 0, 1, \ldots, l^*$ separately. The key point is that we use the same Brownian path for P_l and P_{l-1} to calculate $\delta P_l = P_l - P_{l-1}$ such that the variance is considerably reduced. We have:

$$\mathbb{E}[\delta P_l] \leq C_1 \cdot l\, 2^{-2l}, \quad \mathrm{Var}[\delta P_l] \leq C_2 \cdot l^2\, 2^{-4l}, \quad \mathrm{Cost}[\delta P_l] \leq C_3 \cdot l\, 2^{3l}, \tag{4.18}$$

where the first two inequalities follow from Theorem 4.1.3 and the third is immediate.

Let $\widehat{Y}_l$ be an estimator for $\mathbb{E}\big[\delta P_l\big]$ using M_l independent samples $\delta \widehat{P}_l^{(m)}$ of δP_l,

$$\widehat{Y}_l = \frac{1}{M_l} \sum_{m=1}^{M_l} \delta \widehat{P}_l^{(m)}, \quad l = 1, \ldots, l^*.$$

The MLMC estimator is then defined as $\widehat{P}_{l^*} \coloneqq \sum_{l=0}^{l^*} \widehat{Y}_l$ or

$$\begin{aligned} \widehat{P}_{l^*} \;=\; & \frac{1}{M_0} \sum_{m=1}^{M_0} \sum_{(l_1,l_2)\in\mathcal{I}_0} \Big(\Delta \widehat{P}^{N_0}_{(l_1,l_2)}\Big)^{(m)} + \\ & \sum_{l=1}^{l^*} \frac{1}{M_l} \sum_{m=1}^{M_l} \Big[\sum_{(l_1,l_2)\in\mathcal{I}_l} \Big(\Delta \widehat{P}^{N_l}_{(l_1,l_2)}\Big)^{(m)} - \sum_{(l_1,l_2)\in\mathcal{I}_{l-1}} \Big(\Delta \widehat{P}^{N_{l-1}}_{(l_1,l_2)}\Big)^{(m)}\Big], \end{aligned}$$

where $\Big(\Delta \widehat{P}^{N_l}_{(l_1,l_2)}\Big)^{(m)}$ is the m-th sample for the difference operator on spatial grid level (l_1, l_2) using N_l time steps. Following [38, Theorem 3.1], choosing M_l to minimise the computational cost for a fixed variance, using (4.18) and noticing that the variance decreases strictly faster than the cost increases (the polynomial terms in l being negligible), we achieve the optimal complexity $O(\varepsilon^{-2})$.

4.2 Fourier analysis of the sparse combination error expansion

We will prove Theorem 4.1.2 and Theorem 4.1.3 in this section. We employ a Fourier transform and then analyse the different wave number regions separately, a technique that has been successfully used to derive error expansions for numerical approximations to PDEs (see [15]).

4.2.1 Fourier transform of the solution

Define the Fourier transform pair

$$\widetilde{v}(t,\xi,\eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v(t,x,y)\mathrm{e}^{-\mathrm{i}\xi x-\mathrm{i}\eta y}\,\mathrm{d}x\,\mathrm{d}y,$$
$$v(t,x,y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{v}(t,\xi,\eta)\mathrm{e}^{\mathrm{i}\xi x+\mathrm{i}\eta y}\,\mathrm{d}\xi\,\mathrm{d}\eta.$$

The Fourier transform of (4.1) yields

$$\mathrm{d}\widetilde{v} = -\bigg(\big(\mathrm{i}\mu_x\xi + \mathrm{i}\mu_y\eta + \frac{\xi^2}{2} + \sqrt{\rho_x\rho_y}\rho_{xy}\xi\eta + \frac{\eta^2}{2}\big)\,\mathrm{d}t + \mathrm{i}\sqrt{\rho_x}\xi\,\mathrm{d}W_t^x + \mathrm{i}\sqrt{\rho_y}\eta\,\mathrm{d}W_t^y\bigg)\widetilde{v}, \quad (4.19)$$

subject to the initial data $\widetilde{v}(0) = \mathrm{e}^{-\mathrm{i}\xi x_0-\mathrm{i}\eta y_0}$. For the remainder of the analysis, we take $\mu_x = \mu_y = 0$. This does not alter the results (see Remark 2.3 in [92] for 1-d case).

The solution to (4.19) is

$$\widetilde{v}(t) = X(t)\mathrm{e}^{-\mathrm{i}\xi x_0-\mathrm{i}\eta y_0},$$

where

$$X(t) = \exp\bigg(-\frac{1}{2}(1-\rho_x)\xi^2 t - \frac{1}{2}(1-\rho_y)\eta^2 t - \mathrm{i}\xi\sqrt{\rho_x}W_t^x - \mathrm{i}\eta\sqrt{\rho_y}W_t^y\bigg). \quad (4.20)$$

For the numerical solution, we can use a discrete-continuous Fourier pair

$$\begin{aligned} V_{i,j}^0 &= \frac{1}{4\pi^2 h_x h_y}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\widetilde{V}^0(\xi,\eta)\mathrm{e}^{\mathrm{i}\left((i-i_0)\xi+(j-j_0)\eta\right)}\,\mathrm{d}\xi\,\mathrm{d}\eta,\\ \widetilde{V}^0(\xi,\eta) &= h_x h_y\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} V_{i,j}^0\mathrm{e}^{\mathrm{i}\left(-(i-i_0)\xi-(j-j_0)\eta\right)}, \end{aligned}$$

where $i_0 = x_0/h_x$, $j_0 = y_0/h_y$. It follows from (4.3) that $\widetilde{V}^0(\xi,\eta) = 1$. Then we have

$$V_{i,j}^n = \frac{1}{4\pi^2}\int_{-\frac{\pi}{h_y}}^{\frac{\pi}{h_y}}\int_{-\frac{\pi}{h_x}}^{\frac{\pi}{h_x}}\widetilde{V}^n(\xi,\eta)\mathrm{e}^{\mathrm{i}\left((i-i_0)\xi h_x+(j-j_0)\eta h_y\right)}\,\mathrm{d}\xi\,\mathrm{d}\eta, \quad (4.21)$$

where we make the ansatz

$$\widetilde{V}^n(\xi,\eta) = X_n(\xi,\eta)\widetilde{V}^0(\xi,\eta) = X_n(\xi,\eta). \quad (4.22)$$

It follows from (4.20) that the exact solution of $X(t_{n+1})$ given $X(t_n)$ is

$$X(t_{n+1}) = X(t_n)\exp\bigg(-\frac{\xi^2}{2}(1-\rho_x)k - \frac{\eta^2}{2}(1-\rho_y)k - \mathrm{i}\xi\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}\eta\sqrt{\rho_y k}\widetilde{Z}_{n,y}\bigg),$$

where $W_{t_{n+1}}^x - W_{t_n}^x =: \sqrt{k}Z_{n,x}$, $W_{t_{n+1}}^y - W_{t_n}^y =: \sqrt{k}\widetilde{Z}_{n,y}$ are the Brownian increments.

Now we consider the numerical approximation of X_n. Let

$$\begin{aligned}
&X_{n+1} = C_n\, X_n,\\
&C_n := \exp\left(-\frac{\xi^2}{2}(1-\rho_x)k - \frac{\eta^2}{2}(1-\rho_y)k - \mathrm{i}\xi\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}\eta\sqrt{\rho_y k}\widetilde{Z}_{n,y} + e_n\right),
\end{aligned}$$

and e_n is the logarithmic error between the numerical solution and the exact solution introduced during $[nk, (n+1)k]$. Aggregating over N time steps, at $t_N = kN = T$,

$$X_N = \prod_{n=0}^{N-1} C_n = X(T)\exp\left(\sum_{n=0}^{N-1} e_n\right), \tag{4.23}$$

where

$$X(T) = \exp\left(-\frac{\xi^2}{2}(1-\rho_x)T - \frac{\eta^2}{2}(1-\rho_y)T - \mathrm{i}\xi\sqrt{\rho_x k}\sum_{n=0}^{N-1} Z_{n,x} - \mathrm{i}\eta\sqrt{\rho_y k}\sum_{n=0}^{N-1}\widetilde{Z}_{n,y}\right)$$

is the exact solution at time T. Moreover, inserting $V_{i,j}^n$ from (4.4) to (4.21), we have

$$\begin{aligned}
X_{n+1}(\xi,\eta) = \frac{1}{1-(a_x+a_y)k}\Big(&1 - \mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y}\\
&+ b_x\rho_x k(Z_{n,x}^2 - 1) + b_y\rho_y k(\widetilde{Z}_{n,y}^2 - 1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\Big)X_n(\xi,\eta),
\end{aligned} \tag{4.24}$$

where

$$\begin{aligned}
&a_x = -\frac{2\sin^2\frac{\xi h_x}{2}}{h_x^2}, \quad b_x = -\frac{\sin^2\xi h_x}{2h_x^2}, \quad c_x = \frac{\sin\xi h_x}{h_x}, \quad d = -\frac{\sin\xi h_x\sin\eta h_y}{h_x h_y},\\
&a_y = -\frac{2\sin^2\frac{\eta h_y}{2}}{h_y^2}, \quad b_y = -\frac{\sin^2\eta h_y}{2h_y^2}, \quad c_y = \frac{\sin\eta h_y}{h_y}.
\end{aligned}$$

Hence, C_n can also be expressed as

$$\frac{1-\mathrm{i}\sqrt{k}\big(c_x\sqrt{\rho_x}Z_{n,x}+c_y\sqrt{\rho_y}\widetilde{Z}_{n,y}\big)+k\big(b_x\rho_x(Z_{n,x}^2-1)+b_y\rho_y(\widetilde{Z}_{n,y}^2-1)+d\sqrt{\rho_x\rho_y}Z_{n,x}\widetilde{Z}_{n,y}\big)}{1-(a_x+a_y)k}.$$

4.2.2 Fourier transform of the sparse combination estimator

With the notation from Section 4.1.2, omitting N in $P_{(l_1,l_2)}$ to simplify the notation,

$$\begin{aligned}
P_{(l_1,l_2)} &= h_x h_y\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} V_{i,j}^N + \frac{h_x}{2}\sum_{j=1}^{\infty} V_{0,j}^N + \frac{h_y}{2}\sum_{i=1}^{\infty} V_{i,0}^N + \frac{1}{4}h_x h_y V_{0,0}^N\\
&= \frac{1}{4\pi^2}\int_{-\frac{\pi}{h_y}}^{\frac{\pi}{h_y}}\int_{-\frac{\pi}{h_x}}^{\frac{\pi}{h_x}} h_x h_y X_N^{(l_1,l_2)}(\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta,
\end{aligned} \tag{4.25}$$

where $\chi(h_x,\xi),\ \chi(h_y,\eta)$ is defined in a distributional sense as

$$\chi(h_x,\xi)=\sum_{j=-i_0+1}^{\infty}\mathrm{e}^{\mathrm{i}jh_x\xi}+\frac{1}{2}\mathrm{e}^{-\mathrm{i}i_0h_x\xi},\qquad \chi(h_y,\eta)=\sum_{j=-j_0+1}^{\infty}\mathrm{e}^{\mathrm{i}jh_y\eta}+\frac{1}{2}\mathrm{e}^{-\mathrm{i}j_0h_y\eta}. \tag{4.26}$$

Note that χ only appears multiplied by the smooth, fast decaying function X_N and in integral form, such that this is well-defined.

We recall from (4.12) the sparse combination estimator

$$\Delta P_{(l_1,l_2)}=P_{(l_1,l_2)}-P_{(l_1,l_2-1)}-P_{(l_1-1,l_2)}+P_{(l_1-1,l_2-1)}. \tag{4.27}$$

We assume $h_x,\ h_y<1$. Even though $P_{(l_1,l_2)},\ P_{(l_1,l_2-1)},\ P_{(l_1-1,l_2)},\ P_{(l_1-1,l_2-1)}$ have different Fourier domains, we define Ω_{low} shared by all of them, for $0<p<1/2$,

$$\Omega_{\text{low}}=\big\{(\xi,\eta):|\xi|\le h_x^{-p}\text{ and }|\eta|\le h_y^{-p}\big\}.$$

Then we define $I_{(l_1,l_2)}$ as

$$I_{(l_1,l_2)}=\frac{1}{4\pi^2}\iint_{\Omega_{(l_1,l_2)}\setminus\Omega_{\text{low}}}h_xh_yX_N^{(l_1,l_2)}(\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta,$$

where $\Omega_{(l_1,l_2)}=[-\pi/h_x,\pi/h_x]\times[-\pi/h_y,\pi/h_y]$. Then

$$P_{(l_1,l_2)}=\frac{1}{4\pi^2}\iint_{\Omega_{\text{low}}}h_xh_yX_N^{(l_1,l_2)}(\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta+I_{(l_1,l_2)},$$

and

$$\begin{aligned}
&\Delta P_{(l_1,l_2)}=P_{(l_1,l_2)}-P_{(l_1,l_2-1)}-P_{(l_1-1,l_2)}+P_{(l_1-1,l_2-1)}\\
&=\frac{1}{4\pi^2}\iint_{\Omega_{\text{low}}}\Big(h_xh_yX_N^{(l_1,l_2)}\chi(h_x,\xi)\chi(h_y,\eta)-2h_xh_yX_N^{(l_1-1,l_2)}\chi(2h_x,\xi)\chi(h_y,\eta)\\
&-2h_xh_yX_N^{(l_1,l_2-1)}\chi(h_x,\xi)\chi(2h_y,\eta)+4h_xh_yX_N^{(l_1-1,l_2-1)}\chi(2h_x,\xi)\chi(2h_y,\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta\\
&+I_{(l_1,l_2)}-I_{(l_1-1,l_2)}-I_{(l_1,l_2-1)}+I_{(l_1-1,l_2-1)}.
\end{aligned}$$

Denote

$$\begin{aligned}
G_{(l_1,l_2)}(\xi,\eta)\coloneqq\ &h_xh_yX_N^{(l_1,l_2)}\chi(h_x,\xi)\chi(h_y,\eta)-2h_xh_yX_N^{(l_1-1,l_2)}\chi(2h_x,\xi)\chi(h_y,\eta)\\
&-2h_xh_yX_N^{(l_1,l_2-1)}\chi(h_x,\xi)\chi(2h_y,\eta)+4h_xh_yX_N^{(l_1-1,l_2-1)}\chi(2h_x,\xi)\chi(2h_y,\eta).
\end{aligned} \tag{4.28}$$

Then we have

$$\begin{aligned}
G_{(l_1,l_2)}(\xi,\eta)=\ &4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\big(X_N^{l_1,l_2}-X_N^{l_1,l_2-1}-X_N^{l_1-1,l_2}+X_N^{l_1-1,l_2-1}\big)\\
&+\big(2h_xh_y\chi(2h_x,\xi)\chi(h_y,\eta)-4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\big)\big(X_N^{l_1,l_2}-X_N^{l_1-1,l_2}\big)\\
&+\big(2h_xh_y\chi(h_x,\xi)\chi(2h_y,\eta)-4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\big)\big(X_N^{l_1,l_2}-X_N^{l_1,l_2-1}\big)\\
&+\big(h_xh_y\chi(h_x,\xi)\chi(h_y,\eta)-2h_xh_y\chi(h_x,\xi)\chi(2h_y,\eta)-2h_xh_y\chi(2h_x,\xi)\chi(h_y,\eta)\\
&+4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\big)X_N^{l_1,l_2},
\end{aligned} \tag{4.29}$$

and

$$\Delta P_{(l_1,l_2)} = \frac{1}{4\pi^2}\iint_{\Omega_{\text{low}}} G_{(l_1,l_2)}(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta + I_{(l_1,l_2)} - I_{(l_1-1,l_2)} - I_{(l_1,l_2-1)} + I_{(l_1-1,l_2-1)}.$$

4.2.3 Proof of Theorem 4.1.2

In this section, we give a proof of Theorem 4.1.2. The splitting into different wave number regions is motivated by the analysis of the heat equation in [15]. We further separate $\Omega_{(l_1,l_2)} \setminus \Omega_{\text{low}} = \Omega_{\text{mid}} \cup \Omega_{\text{high}}$. Hence $\Omega_{(l_1,l_2)}$ is divided into three regions,

$$\begin{aligned}
\Omega_{\text{low}} &= \big\{(\xi,\eta) : |\xi| \le h_x^{-p} \text{ and } |\eta| \le h_y^{-p}\big\},\\
\Omega_{\text{mid}} &= \big\{(\xi,\eta) : |\xi| \le h_x^{-p} \text{ and } h_y^{-p} < |\eta| \le \pi/h_y\big\} \cup\\
&\qquad \big\{(\xi,\eta) : h_x^{-p} < |\xi| \le \pi/h_x \text{ and } |\eta| \le h_y^{-p}\big\},\\
\Omega_{\text{high}} &= \big\{(\xi,\eta) : h_x^{-p} < |\xi| \le \pi/h_x \text{ and } h_y^{-p} < |\eta| \le \pi/h_y\big\}.
\end{aligned}$$

We state three lemmas without proof, before giving the proof of Theorem 4.1.2.

Lemma 4.2.1 (Low wave region)**.** *For $G_{(l_1,l_2)}(\xi,\eta)$ introduced in (4.28), there exists a constant $C > 0$, such that for any $l_1, l_2 \in \mathbb{N}$,*

$$\iint_{\Omega_{low}} G_{(l_1,l_2)}(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta = h_x^2 h_y^2 \cdot R(T),$$

where $R(T)$ is a random variable satisfying $|\mathbb{E}[R(T)]| \le C$, $\mathbb{E}[|R(T)|^2] \le C$.

Proof. For $(\xi,\eta) \in \Omega_{\text{low}}$, we have $G_{(l_1,l_2)}(\xi,\eta)$ as in (4.29). From (4.23), we have

$$\begin{aligned}
X_N^{l_1,l_2} - X_N^{l_1,l_2-1} &= X(T)\Big(\exp\big(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\big) - \exp\big(\sum_{n=0}^{N-1} e_n^{l_1,l_2-1}\big)\Big),\\
X_N^{l_1,l_2} - X_N^{l_1-1,l_2} &= X(T)\Big(\exp\big(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\big) - \exp\big(\sum_{n=0}^{N-1} e_n^{l_1-1,l_2}\big)\Big),\\
X_N^{l_1,l_2} - X_N^{l_1,l_2-1} - X_N^{l_1-1,l_2} + X_N^{l_1-1,l_2-1} &= X(T)\Big(\exp\big(\sum_{n=0}^{N-1} e_n^{l_1,l_2}\big) - \exp\big(\sum_{n=0}^{N-1} e_n^{l_1,l_2-1}\big)\\
&\qquad - \exp\big(\sum_{n=0}^{N-1} e_n^{l_1-1,l_2}\big) + \exp\big(\sum_{n=0}^{N-1} e_n^{l_1-1,l_2-1}\big)\Big),
\end{aligned} \tag{4.30}$$

where

$$X(T)=\exp\bigg(-\frac{1}{2}(1-\rho_x)\xi^2T-\frac{1}{2}(1-\rho_y)\eta^2T-\mathrm{i}\xi\sqrt{\rho_xk}\sum_{n=0}^{N-1}Z_{n,x}-\mathrm{i}\eta\sqrt{\rho_yk}\sum_{n=0}^{N-1}\widetilde{Z}_{n,y}\bigg).$$

By a similar proof to Lemma 3.3.1, we have

$$\begin{aligned}
X_N^{l_1,l_2}&=X(T)\cdot R_0(T,h_x\xi,h_y\eta),\\
X_N^{l_1,l_2}-X_N^{l_1,l_2-1}&=O(h_y^2)\cdot X(T)\cdot R_1(T,h_x\xi,h_y\eta),\\
X_N^{l_1,l_2}-X_N^{l_1-1,l_2-1}&=O(h_x^2)\cdot X(T)\cdot R_2(T,h_x\xi,h_y\eta),\\
X_N^{l_1,l_2}-X_N^{l_1,l_2-1}-X_N^{l_1-1,l_2}+X_N^{l_1-1,l_2-1}&=O(h_x^2h_y^2)\cdot X(T)\cdot R_3(T,h_x\xi,h_y\eta).
\end{aligned}$$

Here, $R_0,\ R_1,\ R_2,\ R_3$ are random variables satisfying

$$\mathbb{E}\Big[\big|X(T)\cdot R_i(T,h_x\xi,h_y\eta)\big|^n\Big]=f_i(\xi,\eta)\mathrm{e}^{-\frac{n}{2}(\xi^2+\eta^2+2\xi\eta\sqrt{\rho_x\rho_y}\rho_{xy})T},\qquad i=0,1,2,3,$$

where $f_i(\cdot)$ are polynomial functions. So in the low wave region, we get

$$\begin{aligned}
&\mathbb{E}\bigg[\iint_{\Omega_{\text{low}}}G_{(l_1,l_2)}(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\bigg]\\
&=O(h_x^2h_y^2)\,\mathbb{E}\bigg[\iint_{\Omega_{\text{low}}}4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\Big(X(T)\,R_3(T,h_x\xi,h_y\eta)\Big)\mathrm{d}\xi\,\mathrm{d}\eta\bigg]\\
&\quad+O(h_x^2)\,\mathbb{E}\bigg[\iint_{\Omega_{\text{low}}}\Big(2h_xh_y\chi(2h_x,\xi)\chi(h_y,\eta)-4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\Big)\\
&\qquad\qquad\cdot\Big(X(T)\,R_2(T,h_x\xi,h_y\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg]\\
&\quad+O(h_y^2)\,\mathbb{E}\bigg[\iint_{\Omega_{\text{low}}}\Big(2h_xh_y\chi(h_x,\xi)\chi(2h_y,\eta)-4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\Big)\\
&\qquad\qquad\cdot\Big(X(T)\,R_1(T,h_x\xi,h_y\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg]\\
&\quad+\mathbb{E}\bigg[\iint_{\Omega_{\text{low}}}\Big(h_xh_y\chi(h_x,\xi)\chi(h_y,\eta)-2h_xh_y\chi(h_x,\xi)\chi(2h_y,\eta)-2h_xh_y\chi(2h_x,\xi)\chi(h_y,\eta)\\
&\qquad\qquad+4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\Big)\Big(X(T)\,R_0(T,h_x\xi,h_y\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg].
\end{aligned}$$

Here,

$$\iint_{\Omega_{\text{low}}} 4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\Big(X(T)\cdot R_3(T,h_x\xi,h_y\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta$$

is the numerical approximation to

$$\int_0^\infty\int_0^\infty\bigg(\iint_{\Omega_{\text{low}}}\mathrm{e}^{\mathrm{i}\xi(x-x_0)+\mathrm{i}\eta(y-y_0)}\Big(X(T)\cdot R_3(T,h_x\xi,h_y\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg)\,\mathrm{d}x\,\mathrm{d}y$$

by the trapezoidal rule with error $O(h_x^2)+O(h_y^2)$. Therefore,

$$\mathbb{E}\bigg[\iint_{\Omega_{\text{low}}} 4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\Big(X(T)\cdot R_3(T,h_x\xi,h_y\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg]=O(1).$$

By the same reason,

$$\begin{aligned}
&\mathbb{E}\bigg[\iint_{\Omega_{\text{low}}}\Big(2h_xh_y\chi(2h_x,\xi)\chi(h_y,\eta)-4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\Big)\\
&\qquad\cdot\Big(X(T)\cdot R_2(T,h_x\xi,h_y\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg]=O(h_y^2),\\
&\mathbb{E}\bigg[\iint_{\Omega_{\text{low}}}\Big(2h_xh_y\chi(h_x,\xi)\chi(2h_y,\eta)-4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\Big)\\
&\qquad\cdot\Big(X(T)\cdot R_1(T,h_x\xi,h_y\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg]=O(h_x^2),\\
&\mathbb{E}\bigg[\iint_{\Omega_{\text{low}}}\Big(h_xh_y\chi(h_x,\xi)\chi(h_y,\eta)-2h_xh_y\chi(h_x,\xi)\chi(2h_y,\eta)-2h_xh_y\chi(2h_x,\xi)\chi(h_y,\eta)\\
&\qquad+4h_xh_y\chi(2h_x,\xi)\chi(2h_y,\eta)\Big)\Big(X(T)\cdot R_0(T,h_x\xi,h_y\eta)\Big)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg]=O(h_x^2h_y^2).
\end{aligned}$$

Therefore, we get

$$\mathbb{E}\left[\iint_{\Omega_{\text{low}}}G_{(l_1,l_2)}(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\right]=O\big(h_x^2h_y^2\big).$$

Similarly,

$$\mathbb{E}\left[\left|\iint_{\Omega_{\text{low}}}G_{(l_1,l_2)}(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\right|^2\right]=O\big(h_x^4h_y^4\big).$$

□

Lemma 4.2.2 (Middle wave region)**.** *For the middle wave region, we have,*

$$\mathbb{E}\bigg[\bigg|\iint_{\Omega_{mid}}h_xh_yX_N(\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg|^2\bigg]=o(h_x^r)+o(h_y^r),\qquad\forall r>0.$$

Proof. As

$$\Omega_{\text{mid}} = \big\{(\xi,\eta) : |\xi| \le h_x^{-p} \text{ and } h_y^{-p} < |\eta| \le \pi/h_y\big\} \cup \big\{(\xi,\eta) : h_x^{-p} < |\xi| \le \pi/h_x \text{ and } |\eta| \le h_y^{-p}\big\},$$

it is enough to show for $\Omega^1_{\text{mid}} = \{0 < |\xi| < h_x^{-p},\ h_y^{-p} < |\eta| < \pi/h_y\}$,

$$\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\text{mid}}} h_x h_y X_N(\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\Big|^2\,\bigg] = o(h_y^r),$$

We have

$$\begin{aligned}
&\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\text{mid}}} h_x h_y X_N(\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\Big|^2\,\bigg] \\
&\quad\le 2\,\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\text{mid}}} h_x h_y X(T,\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\Big|^2\,\bigg] \\
&\qquad + 2\,\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\text{mid}}} h_x h_y \Big(X(T,\xi,\eta) - X_N(\xi,\eta)\Big)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\Big|^2\,\bigg],
\end{aligned}$$

where $X(T)$ is introduced in (4.20), and X_N is given by (4.24). Specifically,

$$\begin{aligned}
X(T) &= \prod_{n=0}^{N-1} \exp\Big(-\frac{1}{2}(1-\rho_x)\xi^2 k - \frac{1}{2}(1-\rho_y)\eta^2 k - \mathrm{i}\xi\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}\eta\sqrt{\rho_y k}\widetilde{Z}_{n,y}\Big), \\
X_N &= \prod_{n=0}^{N-1} \frac{1}{1-(a_x+a_y)k}\Big(1 - \mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} - \mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + b_x\rho_x k(Z_{n,x}^2 - 1) \\
&\qquad\qquad + b_y\rho_y k(\widetilde{Z}_{n,y}^2 - 1) + d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}\Big).
\end{aligned}$$

As $X(T)$ is given in closed form, a direct calculation gives

$$\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\text{mid}}} h_x h_y X(T,\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\Big|^2\,\bigg] = o(h_y^r), \quad \forall r > 0.$$

Hence, in the following we focus on

$$\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\text{mid}}} h_x h_y \Big(X(T,\xi,\eta) - X_N(\xi,\eta)\Big)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\Big|^2\,\bigg].$$

Since $h_x h_y \chi(h_x,\xi)\chi(h_y,\eta)$ has finite 2-norm, it is justifiable to consider

$$\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\text{mid}}} X(T,\xi,\eta) - X_N(\xi,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\Big|^2\,\bigg].$$

First we denote

$$X_1(T,\xi) \coloneqq \exp\bigg(-\frac{1}{2}(1-\rho_x)\xi^2 T - \mathrm{i}\xi\sqrt{\rho_x k}\sum_{n=0}^{N-1} Z_{n,x}\bigg),$$
$$X_2(T,\eta) \coloneqq \exp\bigg(-\frac{1}{2}(1-\rho_y)\eta^2 T - \mathrm{i}\eta\sqrt{\rho_y k}\sum_{n=0}^{N-1} \widetilde{Z}_{n,y}\bigg),$$

so that

$$X(T) = X_1(T,\xi)X_2(T,\eta).$$

Then we introduce $\widetilde{X}_n$ such that

$$\widetilde{X}_N = \prod_{n=0}^{N-1}\widetilde{C}_n^x \prod_{n=0}^{N-1}\widetilde{C}_n^y,$$

where

$$\widetilde{C}_n^x = \frac{1-\mathrm{i}c_x\sqrt{\rho_x k}Z_{n,x} + b_x\rho_x k(Z_{n,x}^2-1)}{1-a_x k}, \quad \widetilde{C}_n^y = \frac{1-\mathrm{i}c_y\sqrt{\rho_y k}\widetilde{Z}_{n,y} + b_y\rho_y k(\widetilde{Z}_{n,y}^2-1)}{1-a_y k}.$$

It has been proved in [94] that

$$X_1(T,\xi) - \prod_{n=0}^{N-1}\widetilde{C}_n^x = X_1(T,\xi)\cdot O(h_x^2),$$
$$\mathbb{E}\bigg[\int_{h_y^{-p}}^{\pi/h_y}\bigg|\prod_{n=0}^{N-1}\widetilde{C}_n^y\bigg|^2\mathrm{d}\eta\bigg] = o(h_y^r), \quad \mathbb{E}\bigg[\int_{h_y^{-p}}^{\pi/h_y}\Big|X_2(T,\eta)\Big|^2\mathrm{d}\eta\bigg] = o(h_y^r), \qquad \forall r>0.$$

Thus there exists a constant C such that

$$\int_{-h_x^{-p}}^{h_x^{-p}} \widetilde{X}_N - X(T,\xi,\eta)\,\mathrm{d}\xi = C\cdot\big(1+O(h_x^2)\big)\prod_{n=0}^{N-1}\widetilde{C}_n^y - C\cdot X_2(T,\eta),$$

and as a result,

$$\begin{aligned}
&\mathbb{E}\bigg[\bigg|\iint_{\Omega_{\mathrm{mid}}^1} X(T,\xi,\eta) - \widetilde{X}_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\bigg|^2\bigg]\\
&\le \frac{\pi}{h_y}\int_{h_y^{-p}}^{\pi/h_y}\mathbb{E}\bigg[\bigg|\int_{-h_x^{-p}}^{h_x^{-p}} X(T,\xi,\eta) - \widetilde{X}_N(\xi,\eta)\,\mathrm{d}\xi\bigg|^2\bigg]\mathrm{d}\eta\\
&= o(h_y^r).
\end{aligned}$$

Since

$$\mathbb{E}\Big[\Big|X_n-\widetilde{X}_n\Big|^2\Big]=O(h_x)\mathbb{E}\big[\big|\widetilde{X}_n\big|^2\big],$$

it follows that

$$\begin{aligned}\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\mathrm{mid}}}\widetilde{X}_N(\xi,\eta)-X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\,\bigg]&\le\pi h_x^{-p}h_y^{-1}\iint_{\Omega^1_{\mathrm{mid}}}\mathbb{E}\Big[\Big|\widetilde{X}_N(\xi,\eta)-X_N(\xi,\eta)\Big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta\\&=\pi h_x^{-p}h_y^{-1}\iint_{\Omega^1_{\mathrm{mid}}}O(h_x)\mathbb{E}\Big[\big|\widetilde{X}_n\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta=o(h_y^r).\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\mathrm{mid}}}X(T,\xi,\eta)-X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\,\bigg]\\&\le 2\,\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\mathrm{mid}}}X(T,\xi,\eta)-\widetilde{X}_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\,\bigg]+2\,\mathbb{E}\bigg[\,\Big|\iint_{\Omega^1_{\mathrm{mid}}}\widetilde{X}_N(T,\xi,\eta)-X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\,\bigg]\\&=o(h_y^r).\end{aligned}$$

□

Lemma 4.2.3 (High wave region)**.** *For the high wave region, we have*

$$\mathbb{E}\bigg[\,\Big|\iint_{\Omega_{high}}h_xh_yX_N(\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\Big|^2\,\bigg]=o(h_x^rh_y^r),\qquad\forall r>0.$$

Proof. For the same reason as in the proof of Lemma 4.2.2, it is sufficient to prove

$$\mathbb{E}\bigg[\,\Big|\iint_{\Omega_{\mathrm{high}}}X(T,\xi,\eta)-X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\Big|^2\,\bigg]=o(h_x^rh_y^r).$$

Note that

$$\Omega_{\mathrm{high}}=\big\{(\xi,\eta):h_x^{-p}<|\xi|\le\frac{\pi}{2h_x}\text{ and }h_y^{-p}<|\eta|\le\frac{\pi}{2h_y}\big\}.$$

By (4.24), we have

$$X_N=X_0\prod_{n=0}^{N-1}\frac{1-\mathrm{i}c_x\sqrt{\rho_xk}Z_{n,x}-\mathrm{i}c_y\sqrt{\rho_yk}\widetilde{Z}_{n,y}+b_x\rho_xk(Z_{n,x}^2-1)+b_y\rho_yk(\widetilde{Z}_{n,y}^2-1)+d\sqrt{\rho_x\rho_y}kZ_{n,x}\widetilde{Z}_{n,y}}{1-(a_x+a_y)k}.$$

We denote

$$u = \frac{\sin^2 \frac{h_x \xi}{2}}{(\frac{h_x \xi}{2})^2} = O(1), \qquad v = \frac{\sin^2 \frac{h_y \eta}{2}}{(\frac{h_y \eta}{2})^2} = O(1).$$

In this case,

$$\begin{aligned}
\mathbb{E}\big[X_N\big] &= X_0 \prod_{n=0}^{N-1} \mathbb{E}\big[C_n\big] = X_0 \bigg(\frac{1 + d\sqrt{\rho_x \rho_y}\rho_{xy} k}{1 + \frac{1}{2}(\xi^2 u + \eta^2 v)k} \bigg)^N \\
&< X_0 \exp\bigg(-\frac{1}{2}\big(\xi^2 u + \eta^2 v + 2\xi\eta \frac{\sin \xi h_x \sin \eta h_y}{\xi h_x \eta h_y} \sqrt{\rho_x \rho_y}\rho_{xy}\big)\, T \bigg), \\
\mathbb{E}\big[\big|X_N\big|^2\big] &= \big|X_0\big|^2 \prod_{n=0}^{N-1} \mathbb{E}\big[\big|C_n\big|^2\big] \\
&\le \big|X_0\big|^2 \bigg(1 - \frac{k\xi^2 u\big(1 - \rho_x \cos^2 \frac{\xi h_x}{2} + \frac{1}{4}\xi^2 u k(1 - 2\rho_x^2(1 + 2|\rho_{xy}|))\big)}{\big(1 - (a_x + a_y)k\big)^2} \\
&\qquad - \frac{k\eta^2 v\big(1 - \rho_y \cos^2 \frac{\eta h_y}{2} + \frac{1}{4}\eta^2 v k(1 - 2\rho_y^2(1 + 2|\rho_{xy}|))\big)}{\big(1 - (a_x + a_y)k\big)^2} \\
&\qquad - \frac{k^2\xi^2\eta^2 uv\big(1 - 2\rho_x\rho_y(1 + |\rho_{xy}| + 3\rho_{xy}^2)\big)}{2\big(1 - (a_x + a_y)k\big)^2} \bigg)^N \\
&\le \big|X_0\big|^2 \bigg(1 - \frac{k\xi^2 u\big(\beta + \frac{1}{4}\xi^2 uk\beta\big)}{\big(1 - (a_x + a_y)k\big)^2} - \frac{k\eta^2 v\big(\beta + \frac{1}{4}\eta^2 vk\beta\big)}{\big(1 - (a_x + a_y)k\big)^2} - \frac{k^2\xi^2\eta^2 uv\beta}{2\big(1 - (a_x + a_y)k\big)^2} \bigg)^N \\
&< \big|X_0\big|^2 \exp\bigg(-\xi^2 u\beta T \frac{1 + \frac{1}{4}\beta\xi^2 uk}{\big(1 + \frac{1}{2}(\xi^2 u + \eta^2 v)k\big)^2} - \eta^2 v\beta T \frac{1 + \frac{1}{4}\beta\eta^2 vk}{\big(1 + \frac{1}{2}(\xi^2 u + \eta^2 v)k\big)^2} \bigg),
\end{aligned}$$

where

$$\beta = \min\big\{1-\rho_x, 1-\rho_y, 1-2\rho_x^2(1+2|\rho_{xy}|), 1-2\rho_y^2(1+2|\rho_{xy}|), 1-2\rho_x\rho_y(1+2|\rho_{xy}|+3\rho_{xy}^2)\big\} \in (0,1).$$

As we have $k < \lambda \min\{h_x^2, h_y^2\}$,

$$0 < \xi^2 k < \lambda\pi^2/4, \quad 0 < \eta^2 k < \lambda\pi^2/4.$$

Also,

$$\min\bigg\{ \frac{u(1 + \frac{1}{4}\beta\xi^2 uk)}{\big(1 + \frac{1}{2}(\xi^2 u + \eta^2 v)k\big)^2}, \frac{v(1 + \frac{1}{4}\beta\eta^2 vk)}{\big(1 + \frac{1}{2}(\xi^2 u + \eta^2 v)k\big)^2} \bigg\} > \frac{8}{\pi^2\big(1 + \frac{1}{4}\lambda\pi^2\big)^2}.$$

We define a constant

$$\kappa \coloneqq \frac{8\beta T}{\pi^2\big(1+\frac{1}{4}\lambda\pi^2\big)^2},$$

then

$$\mathbb{E}\big[\big|X_N\big|^2\big] < |X_0|^2 \exp\big(-\kappa\xi^2-\kappa\eta^2\big).$$

Therefore we have

$$\begin{aligned}
\mathbb{E}\bigg[\,\bigg|\iint_{\Omega_{\text{high}}} X_N(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta\bigg|^2\,\bigg] &< \frac{\pi^2}{h_x h_y}\iint_{\Omega_{\text{high}}} \mathbb{E}\Big[\big|X_N(\xi,\eta)\big|^2\Big]\,\mathrm{d}\xi\mathrm{d}\eta \\
&< \frac{\pi^2|X_0|^2}{h_x h_y}\iint_{\Omega_{\text{high}}} \mathrm{e}^{-\kappa\xi^2-\kappa\eta^2}\,\mathrm{d}\xi\mathrm{d}\eta \\
&= \frac{\pi^2|X_0|^2}{h_x h_y}\int_{h_y^{-p}}^{\pi/h_y}\int_{h_x^{-p}}^{\pi/h_x} \mathrm{e}^{-\kappa\xi^2-\kappa\eta^2}\,\mathrm{d}\xi\mathrm{d}\eta \\
&\le C\cdot(h_x h_y)^{-1+p}\exp\big(-\kappa h_x^{-p}-\kappa h_y^{-p}\big) \\
&= o(h_x^r h_y^r), \quad \forall r>0.
\end{aligned}$$

□

Theorem 4.1.2. We have derived in Section 4.2.2 that

$$\Delta P_{(l_1,l_2)} = \frac{1}{4\pi^2}\iint_{\Omega_{\text{low}}} G_{(l_1,l_2)}(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta + I_{(l_1,l_2)} - I_{(l_1-1,l_2)} - I_{(l_1,l_2-1)} + I_{(l_1-1,l_2-1)}.$$

Lemma 4.2.2 and Lemma 4.2.3 give that

$$\begin{aligned}
\mathbb{E}\Big[\big|I_{(l_1,l_2)}\big|^2\Big] \le{}& 2\,\mathbb{E}\bigg[\,\bigg|\iint_{\Omega_{\text{mid}}} h_x h_y X_N(\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg|^2\,\bigg] \\
&+ 2\,\mathbb{E}\bigg[\,\bigg|\iint_{\Omega_{\text{high}}} h_x h_y X_N(\xi,\eta)\chi(h_x,\xi)\chi(h_y,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\bigg|^2\,\bigg] = o(h_x^r)+o(h_y^r)
\end{aligned}$$

for all $r>0$. Similar estimates hold for $I_{(l_1-1,l_2)}$, $I_{(l_1,l_2-1)}$, and $I_{(l_1-1,l_2-1)}$. Combining this with Lemma 4.2.1, the two inequalities in (4.13) follow. □

4.2.4 Proof of Theorem 4.1.3

Theorem 4.1.3. Since from Proposition 4.1.1,

$$\lim_{l_1,l_2\to\infty} \sqrt{\mathbb{E}\left[|P^N_{(l_1,l_2)} - P_T|^2\right]} = O(k) = O(2^{-2l}),$$

we have

$$\begin{aligned}
\sqrt{\mathbb{E}\left[|P^N_l - P_T|^2\right]} &= \sqrt{\mathbb{E}\Big[\Big|\sum_{l_1+l_2>l+1} \Delta P_{(l_1,l_2)}\Big|^2\Big]} + O(2^{-2l}) \\
&\le \sum_{l_1+l_2>l+1} \sqrt{\mathbb{E}\Big[\big|\Delta P_{(l_1,l_2)}\big|^2\Big]} + O(2^{-2l}) \\
&\le C\cdot\left[\sum_{l_1=0}^{l}\sum_{l_2=l+1-l_1}^{\infty} 2^{-2l_1-2l_2} + \sum_{l_1=l+1}^{\infty}\sum_{l_2=0}^{\infty} 2^{-2l_1-2l_2}\right] + O(2^{-2l}) \\
&= C\Big(\frac{l}{3}2^{-2l} + \frac{7}{9}2^{-2l}\Big) + O(2^{-2l}) = O(l\,2^{-2l}).
\end{aligned}$$

□

4.3 Numerical tests

In this section, we test the theoretical convergence results empirically.

4.3.1 Mean and variance of hierarchical increments

First, we verify numerically for $l_1,\, l_2 > 0$, $h_x = h_0 \cdot 2^{-l_1}$, $h_y = h_0 \cdot 2^{-l_2}$, the result from Theorem 4.1.2 that

$$\Big|\mathbb{E}\left[\Delta P^N_{(l_1,l_2)}\right]\Big| = O(h_x^2 h_y^2) \sim 2^{-2(l_1+l_2)}, \qquad \mathbb{E}\Big[\,\big|\Delta P^N_{(l_1,l_2)}\big|^2\,\Big] = O(h_x^4 h_y^4) \sim 2^{-4(l_1+l_2)}.$$

We choose parameters $h_0 = 1,\ T = 1,\ x_0 = y_0 = 2,\ \mu_x = \mu_y = 0.0809,\ \rho_x = \rho_y = 0.2,\ \rho_{xy} = 0.45$, and truncate the domain to $(x, y) \in [-8, 12] \times [-8, 12]$.

Table 4.1 and Table 4.2 show $\log_2 \left|\mathbb{E}[\Delta P^N_{(l_1,l_2)}]\right|$ and $\log_2 \mathrm{Var}[\Delta P^N_{(l_1\,l_2)}]$ with fixed timestep $k = 4^{-3}$, and different levels of mesh refinement. We can see from the table that the mean

decreases by around 2 going from level l to $l+1$, and the variance decreases by approximately 4, consistent with the theoretical prediction. Figure 4.1 depicts the corresponding contour plots.

	$l_2=0$	$l_2=1$	$l_2=2$	$l_2=3$	$l_2=4$
$l_1=0$	-0.0819	-7.28	-9.04	-11.00	-12.98
$l_1=1$	-7.30	-13.73	-15.49	-17.44	-19.43
$l_1=2$	-9.05	-15.48	-17.24	-19.19	-21.17
$l_1=3$	-10.99	-17.43	-19.19	-21.13	-23.12
$l_1=4$	-12.98	-19.42	-21.17	-23.13	-25.11

Table 4.1: $\log_2 \left|\mathbb{E}[\Delta P^N_{(l_1,l_2)}]\right|$ with $h_x = 2^{-l_1}$, $h_y = 2^{-l_2}$ and $k = 4^{-3}$.

	$l_2=0$	$l_2=1$	$l_2=2$	$l_2=3$	$l_2=4$
$l_1=0$	-8.34	-13.90	-17.36	-21.20	-25.15
$l_1=1$	-13.91	-25.29	-29.05	-33.03	-36.99
$l_1=2$	-17.35	-28.34	-32.26	-36.19	-40.24
$l_1=3$	-21.19	-32.30	-36.02	-39.95	-43.98
$l_1=4$	-25.14	-36.16	-39.97	-43.96	-47.85

Table 4.2: $\log_2 \mathrm{Var}[\Delta P^N_{(l_1\,l_2)}]$ with $h_x = 2^{-l_1}$, $h_y = 2^{-l_2}$ and $k = 4^{-3}$.

4.3.2 Mean and variance of sparse grid increments

Next, we estimate $\mathbb{E}\big[\delta P_l\big]$ and $\mathrm{Var}\big[\delta P_l\big]$, where δP_l is given by (4.17). From Table 4.1 and Table 4.2 we deduce that for the index set $I_l = \{(l_1,l_2) : l_1 + l_2 \leq l+1\}$, the terms on the 'boundary' (i.e., $l_1 = 0$ or $l_2 = 0$) will dominate. Although this does not affect the total order of complexity, we can further optimise the cost by modifying the index set such that the contribution from the boundary and 'interior' are similar.

Define therefore $\widehat{I}_l$ as the indices for interior meshes, and $\widetilde{I}_l$ for boundary meshes,

$$\widehat{I}_l = \{l_1 > 0, l_2 > 0 : l_1 + l_2 \leq l\}, \quad \widetilde{I}_l = \{(l_1,0) : 0 \leq l_1 \leq l\} \cup \{(0,l_2) : 0 \leq l_2 \leq l\}.$$

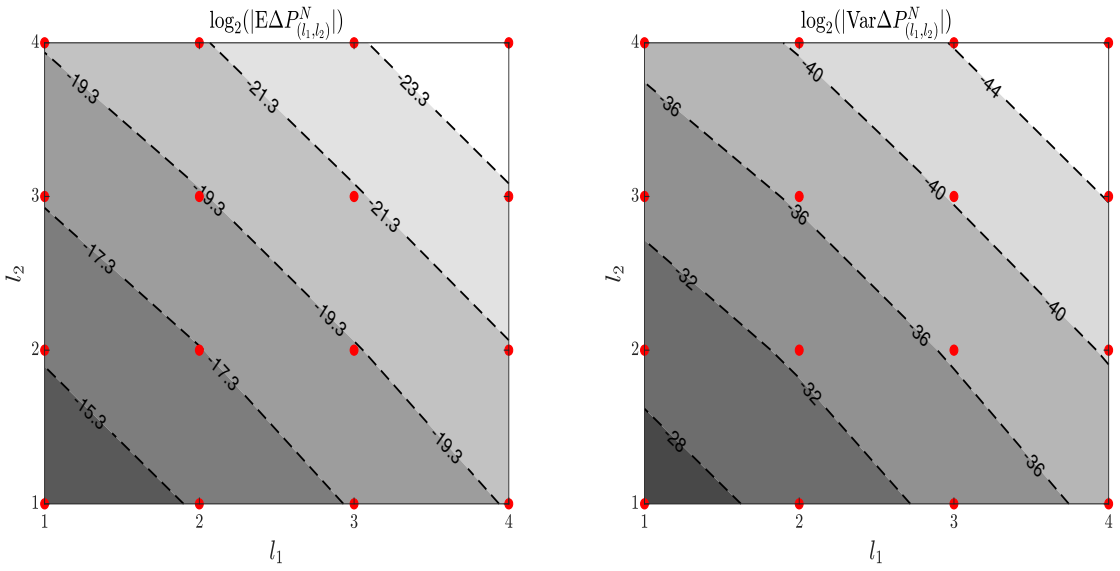


Figure 4.1: Contour plots of log values of sample mean and variance of $\Delta P^N_{(l_1\, l_2)}$ for $l_1, l_2 > 0$ used in the sparse combination method.

We balance the contributions from these two sets by finding, for some fixed N,

$$l^* = \max\left\{ \operatorname*{argmin}_{l_x \geq 0} \left| \log_2 \left|\mathbb{E}\Delta P^N_{(1,1)}\right| - \log_2 \left|\mathbb{E}\Delta P^N_{(2+l_x,\,0)}\right| \right|, \right.$$
$$\left. \operatorname*{argmin}_{l_y \geq 0} \left| \log_2 \left|\mathbb{E}\Delta P^N_{(1,1)}\right| - \log_2 \left|\mathbb{E}\Delta P^N_{(0,2+l_y)}\right| \right| \right\}.$$

and using the index set

$$\mathcal{I}_l = \begin{cases} \widetilde{I}_l & \text{if } l < 2 + l^*, \\ \widetilde{I}_l \cup \widehat{I}_{l-l^*} & \text{if } l \geq 2 + l^*. \end{cases}$$

For example, from Table 4.1, $l^* = 2$. So we have

$$\mathcal{I}_0 = \{(0,0)\}, \qquad \mathcal{I}_l = \mathcal{I}_{l-1} \cup \{(l,0),(0,l)\}, \quad l \in \{1,2,3\},$$
$$\mathcal{I}_4 = \mathcal{I}_3 \cup \{(4,0),(0,4),(1,1)\}, \qquad \mathcal{I}_5 = \mathcal{I}_4 \cup \{(5,0),(0,5),(1,2),(2,1)\}, \cdots$$

We use $k_l = k_0 \cdot 2^{-2l}$ and $\mathcal{I}_l$ for the construction of δP_l. Table 4.3 verifies

$$\mathbb{E}[\delta P_l] = O(l\, 2^{-2l}), \quad \mathrm{Var}[\delta P_l] = O(l^2\, 2^{-4l}), \quad \mathrm{Cost}[\delta P_l] = O(l\, 2^{3l}).$$

	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
$\log_2\left\|\mathbb{E}\left[\delta P_l\right]\right\|$	-0.0647	-7.62	-9.44	-11.42	-13.30	-15.22
$\log_2\left(\mathrm{Var}\left[\delta P_l\right]\right)$	-9.15	-16.93	-20.71	-24.66	-28.20	-31.83
$\log_2\left(\mathrm{Cost}[\delta P_l]\right)$	13.64	18.50	22.02	25.22	28.44	31.65

Table 4.3: Log values of mean, variance, cost for δL_l, $h_0 = 1/2,\ k_0 = 1/8$.

Figure 4.2 are the corresponding plots. The fitted slopes in Figure 4.2 are -1.91, -3.73, and 3.27, respectively, compared to the theoretical asymptotic values of -2, -4, and 3 (neglecting logarithmic terms, which will play a role for low levels).

The cost is counted here as the total number of operations. Specifically, under the ADI scheme (4.5), the cost of one numerical realisation of (4.1) with mesh size h_x, h_y and timestep k is

$$\text{cost/path} \ \sim \ \frac{x_{\max} - x_{\min}}{h_x} \cdot \frac{y_{\max} - y_{\min}}{h_y} \cdot \frac{T}{k}.$$

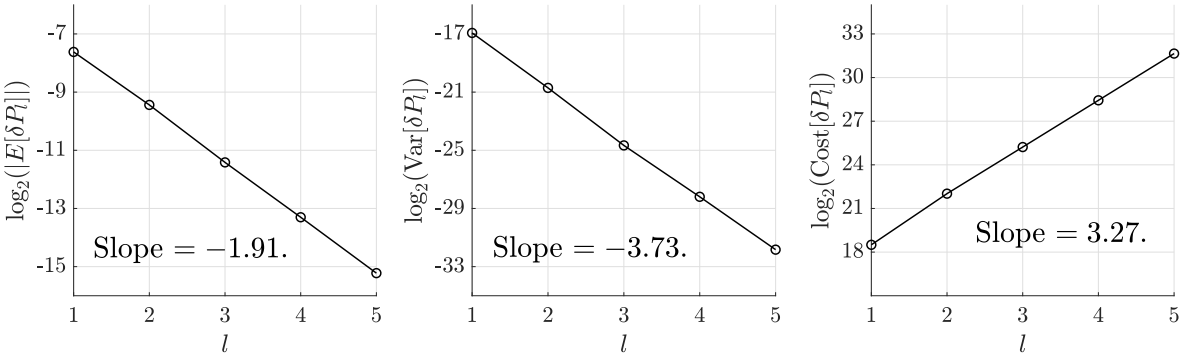
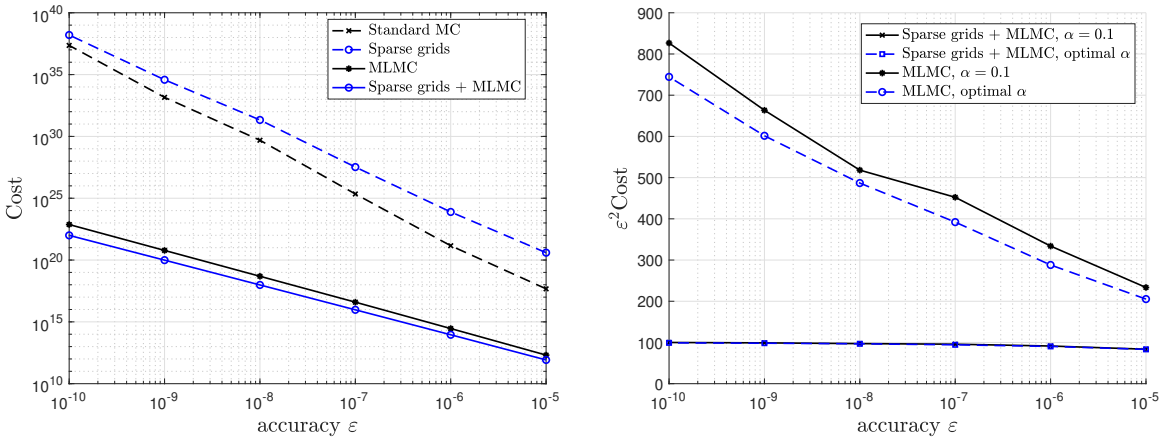


Figure 4.2: Log values of mean, variance and cost for δP_l.

Since the mean square error for the estimator can be expressed as sum of the variance and the square of the weak error, we split the accuracy "budget" as

$$\left|\mathbb{E}\left[P_l^N - P_T\right]\right| \le \alpha\varepsilon, \qquad \mathrm{Var}\left[P_l^N\right] \le (1-\alpha^2)\varepsilon^2, \tag{4.31}$$

and optimize over α. Since $V_l C_l \sim l^3\, 2^{-l}$, the variance decays with levels more rapidly than the cost increases with levels, and thus the dominant cost is on level 0. Therefore, we choose α relatively small to reduce the cost on level 0, and hence we reduce the total cost.



(a) MC, Sparse MC, MLMC, Sparse MLMC.

(b) Sparse and full grid MLMC.

Figure 4.3: Comparison of total cost among different schemes.

To find the optimal α, one approach is to approximate the total cost given different α, and choose the one which minimise the complexity.

Figure 4.3(a) is the loglog plot of total cost among all the methods mentioned before: standard Monte Carlo, sparse combination, multilevel Monte Carlo, sparse combination with MLMC. All the schemes use the optimal α for each accuracy. We can see from the graph that standard MC gives the cost $O(\varepsilon^{-4})$, and sparse combination gives $O(\varepsilon^{-7/2})$, as expected. As for MLMC and sparse combination with MLMC methods, both yield approximately $O(\varepsilon^{-2})$, which verifies our proof as the log term in MLMC cost is negligible in this plot.

Figure 4.3(b) compares $\varepsilon^2 W$ between sparse combination with MLMC and MLMC alone (without sparse combination). The total computational cost of sparse combination with MLMC is approximately proportional to ε^{-2}, hence $\varepsilon^2 W$ does not vary significantly for different accuracy ε. However, as the total cost of the multilevel scheme is proportional to $\varepsilon^{-2}(\log \varepsilon)^2$, and we can see from the figure that $\varepsilon^2 W$ increases as ε goes to zero. For the black line, we use $\alpha = 0.1$. For the blue dotted line, each scheme uses the optimal α for different accuracies. As a result, sparse combination with MLMC achieves the optimal order of complexity.

4.4 Generalisation to higher dimensions

We can generalise the SPDE to d dimensions, where $d \geq 3$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where there is given a d-dimensional standard Brownian motion W with correlation matrix $\Sigma = (\rho_{k,m})$. The natural extension of the SPDE (4.1) is

$$\mathrm{d}v = -\sum_{k=1}^{d} \mu_k \frac{\partial v}{\partial x_k}\,\mathrm{d}t + \frac{1}{2}\sum_{k,m=1}^{d} \sqrt{\rho_k \rho_m}\,\rho_{k,m} \frac{\partial^2 v}{\partial x_k \partial x_m}\,\mathrm{d}t - \sum_{k=1}^{d} \sqrt{\rho_k}\frac{\partial v}{\partial x_k}\,\mathrm{d}W_t^k, \tag{4.32}$$

for $x \in \mathbb{R}^d,\ 0 < t \leq T$, where $\mu \in \mathbb{R}^d$ and $\rho \in (0,1)^d$ are parameters, subject to the Dirac initial datum $v(0,x) = \delta(x_1 - x_{1,0}) \otimes \cdots \otimes \delta(x_d - x_{d,0})$, where $x_0 \in \mathbb{R}^d_+$ is given.

We use a spatial grid with uniform spacing $h_{x_i} > 0$, $i = 1, \ldots, d$, and let $V^m_{i_1,\ldots,i_d}$ be the approximation to $v(mk, i_1 h_{x_1}, \ldots, i_d h_{x_d})$, $m = 1, \ldots, N$, $i_j \in \mathbb{Z}$, where $i_{0,j} \coloneqq [x_{0,j}/h_{x_j}]$, the closest integers to $x_{0,j}/h_{x_j}$. We approximate $v(0, x_1, \ldots, x_d)$ by

$$V^0_{i_1,\ldots,i_d} = \prod_{j=1}^{d} h_{x_j}^{-1} \delta_{(i_{0,1},\ldots,i_{0,d})} = \begin{cases} \prod_{j=1}^{d} h_{x_j}^{-1}, & i_j = i_{0,j}, \quad \forall j \in \{1,\ldots,d\}, \\ 0, & \text{otherwise.} \end{cases}$$

In analogy with (4.8), we define a linear functional by

$$P_T(v) = \int_{\mathbb{R}^d_+} v(T,x)\,\mathrm{d}x. \tag{4.33}$$

Similar as before, we use an implicit finite difference scheme to approximate (4.32), and we use the trapezoidal rule for P with a truncation of the domain. We have the following conjectures and results.

Conjecture 4.4.1. *Assume the implicit finite difference scheme is stable* [2]*, and the timestep k and mesh size h_{x_i} satisfy*

$$k \leq \lambda \min\{h_{x_1}^2, h_{x_2}^2, \cdots, h_{x_d}^2\}, \tag{4.34}$$

for arbitrary fixed $\lambda > 0$. Then we have the error expansion

$$\sqrt{\mathbb{E}\big[|V^N_{i_1,\ldots,i_d} - v(T, x_1, \ldots, x_d)|^2\big]} = O(h_{x_1}^2) + O(h_{x_2}^2) + \ldots + O(h_{x_d}^2),$$

where $N = T/k$ is the number of time steps.

[2] We expect this to be true for 'small enough' correlations as in the two-dimensional case for (4.6), but it is not obvious what the conditions will be for specific $d > 2$ without performing the analysis.

Next we apply the sparse combination method to (4.33) with fixed k satisfying (4.34). Similar to Section 4.1.2, let Δ_i be the first-order difference operator along directions $i = 1, \ldots, d$, defined as in (4.11), and $\Delta = \Delta_1 \otimes \cdots \otimes \Delta_d$.

Conjecture 4.4.2. *Assume the implicit finite difference scheme is stable. Let* $h_{x_i} = h_0 \cdot 2^{-l_i}$, $i = 1, \cdots, d$, *and* k *be the timestep such that for an arbitrary fixed* $\lambda > 0$, $k \leq \lambda \min\{h_{x_1}^2, h_{x_2}^2, \cdots, h_{x_d}^2\}$, *where* $N = T/k$ *the number of time steps. Then the first and second moments of* $\Delta P_{\boldsymbol{l}}^N$ *satisfy*

$$\left| \mathbb{E}\left[\Delta P_{\boldsymbol{l}}^N\right] \right| = O(h_{x_1}^2 \cdots h_{x_d}^2), \qquad \mathbb{E}\Big[\left|\Delta P_{\boldsymbol{l}}^N\right|^2 \Big] = O(h_{x_1}^4 \cdots h_{x_d}^4). \tag{4.35}$$

Given a sequence of index sets

$$\mathcal{I}_l = \{\boldsymbol{l} \in \mathbb{N}_0^d : \, l_1 + \cdots + l_d \leq l + 1\},$$

the approximation on level l is defined (similar to (4.14)) as

$$P_l^N = \sum_{\boldsymbol{l} \in \mathcal{I}_l} \Delta P_{\boldsymbol{l}}^N. \tag{4.36}$$

Note that we use the same k for all $\Delta P_{\boldsymbol{l}}^N$, $(l_1, \cdots, l_d) \in \mathcal{I}_l$. Then we have:

Proposition 4.4.1. *Assume Conjectures 4.4.1 and 4.4.2 to be true. Then, for* P *given by* (4.33) *and* P_l^N *by* (4.36), *we have*

$$\sqrt{\mathbb{E}\left[|P_l^N - P_T|^2\right]} = O(l^{d-1}\, 2^{-2l}).$$

Moreover, by choosing $l = \left[\frac{1}{2}(-\log_2 \varepsilon + (d-1)\log_2 |\log \varepsilon|)\right]$, *the computational cost* W *to achieve a RMSE* ε *using standard Monte Carlo estimation is*

$$W = O\big(\varepsilon^{-\frac{7}{2}} |\log \varepsilon|^{\frac{5}{2}(d-1)}\big). \tag{4.37}$$

Next, we combine MLMC with the sparse combination scheme. Similar to Section 4.1.3, let $P_l \coloneqq P_l^{N_l}$ be an approximation to P as in (4.14) using a discretisation with timestep k_l and index set $\mathcal{I}_l$ defined above. Then we define

$$\delta P_l = P_l - P_{l-1}, \quad l \geq 0, \tag{4.38}$$

where we denote $P_{-1} \coloneqq 0$. Thus the approximation to P at level l^* has the form

$$\mathbb{E}[P_{l^*}] = \sum_{l=0}^{l^*} \mathbb{E}[\delta P_l] = \sum_{l=0}^{l^*} \mathbb{E}\bigg[\sum_{\boldsymbol{l} \in \mathcal{I}_l} \Delta P_{\boldsymbol{l}}^{N_l} - \sum_{\boldsymbol{l} \in \mathcal{I}_{l-1}} \Delta P_{\boldsymbol{l}}^{N_{l-1}} \bigg].$$

In this way, we simulate δP_l, $l = 0, 1, \ldots, l^*$ instead of directly simulating P_{l^*}, such that the variance of $\delta P_l = P_l - P_{l-1}$ is considerably reduced by using the same Brownian path for P_l and P_{l-1}.

Let $\widehat{Y}_l$ be an estimator for $\mathbb{E}\big[\delta P_l\big]$ using M_l samples. Each estimator is an average of M_l independent samples, where $\delta \widehat{P}_l^{(m)}$ is the m-th sample arising from a single SPDE approximation,

$$\widehat{Y}_l = \frac{1}{M_l} \sum_{m=1}^{M_l} \delta \widehat{P}_l^{(m)}, \quad l = 0, \ldots, l^*.$$

The MLMC estimator is defined as

$$\widehat{P}_{l^*} \coloneqq \sum_{l=0}^{l^*} \widehat{Y}_l = \frac{1}{M_l} \sum_{m=1}^{M_l} \bigg[\sum_{\boldsymbol{l} \in \mathcal{I}_l} \Big(\Delta \widehat{P}_{\boldsymbol{l}}^{N_l} \Big)^{(m)} - \sum_{\boldsymbol{l} \in \mathcal{I}_{l-1}} \Big(\Delta \widehat{P}_{\boldsymbol{l}}^{N_{l-1}} \Big)^{(m)} \bigg], \tag{4.39}$$

where $\Big(\Delta \widehat{P}_{\boldsymbol{l}}^{N_l} \Big)^{(m)}$ is the m-th sample for the difference on spatial grid level $\boldsymbol{l} = (l_1, \ldots, l_d)$ of the SPDE approximation using N_l time steps. Following [38], through optimising M_l to minimise the computational cost for a fixed variance, we can achieve the optimal complexity $O(\varepsilon^{-2})$ in this case.

Proposition 4.4.2. *Assume Conjectures 4.4.1 and 4.4.2 to be true. Given δP_l from* (4.38)*, there exist C_1, C_2, $C_3 > 0$, such that*

$$\mathbb{E}[\delta P_l] \leq C_1 \cdot l^{d-1}\, 2^{-2l}, \qquad \mathrm{Var}[\delta P_l] \leq C_2 \cdot l^{2d-2}\, 2^{-4l}, \qquad \mathrm{Cost}[\delta P_l] \leq C_3 \cdot l^{d-1}\, 2^{3l}.$$

Then, given a RMSE ε, the MLMC estimator (4.39) *leads to a total complexity $O(\varepsilon^{-2})$.*

Proof. The first inequalities follow directly from (4.35). The complexity then is a small modification of [38, Theorem 3.1]. □

If we use MLMC without sparse combination to estimate $\mathbb{E}[P_T]$, by letting $h_{x_j} = h_0 \cdot 2^{-l}$, $k = k_0 \cdot 2^{-2l}$, then for constants $C_1, C_2, C_3 > 0$, independent of h and k,

$$E_l \leq C_1 \cdot 2^{-2l}, \quad V_l \leq C_2 \cdot 2^{-4l}, \quad W_l \leq C_3 \cdot 2^{(d+2)l}.$$

Similarly, we have the following result.

Proposition 4.4.3. *Given a RMSE ε, the total computational cost W for P_T in* (4.33) *using MLMC satisfies*

$$W = O\big(\varepsilon^{-1-\frac{d}{2}}\big), \quad d \geq 3. \tag{4.40}$$

Comparing (4.40) with (4.37), we can see that when $d > 5$, the sparse combination scheme with standard MC performs better than MLMC on regular grids.

4.5 Summary

We considered a two-dimensional parabolic SPDE and a functional of the solution. We analysed the accuracy and complexity of a sparse combination multilevel Monte Carlo estimator, and showed that, by using a semi-implicit Milstein finite difference discretisation (4.4), we achieved the order of complexity $O(\varepsilon^{-2})$ for a RMSE ε, whereas the cost using standard Monte Carlo is $O(\varepsilon^{-4})$, and that using MLMC is $O(\varepsilon^{-2}(\log \varepsilon)^2)$. When generalising to higher-dimensional problems, sparse combination with MLMC maintains the optimal complexity, whereas MLMC has an increasing total cost as the dimension increases.

Chapter 5

Asymptotic expansion of a two-dimensional Zakai SPDE

In this chapter, we will study the two dimensional Zakai SPDE

$$\begin{aligned} \mathrm{d}v &= \frac{1}{\epsilon}\Big(\nu^2(1-\rho_y^2)\partial_{yy}v + \kappa\partial_y(yv)\Big)\,\mathrm{d}t \\ &\quad + \Big(\frac{1}{2}\sigma^2\big(y+U_t\big)\partial_{xx}v - \mu\,\partial_x v\Big)\,\mathrm{d}t - \rho_x\sigma\big(y+U_t\big)\partial_x v\,\mathrm{d}W_t^x, \\ v(0,x,y) &= \delta(x-x_0)\otimes\delta(y-y_0), \end{aligned} \tag{5.1}$$

where $\nu > 0$, $\kappa > 0$, $\mu \in \mathbb{R}$, $\rho_x \in (-1,1)$, $0 < \epsilon \ll 1$ are fixed constants, $\sigma(\cdot)$ is a real-valued function (we will specify the condition needed later), and U_t follows the dynamics

$$\mathrm{d}U_t = -\frac{\kappa}{\epsilon}U_t\,\mathrm{d}t + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\rho_y\,\mathrm{d}W_t^y, \qquad U_0 = 0. \tag{5.2}$$

In fact, this SPDE (5.1) is derived from a large portfolio system with fast mean reverting stochastic volatility process in [55],

$$\begin{aligned} \mathrm{d}X_t^i &= \mu\,\mathrm{d}t + \sigma(Y_t^i+U_t)\Big(\rho_x\,\mathrm{d}W_t^x + \sqrt{1-\rho_x^2}\,\mathrm{d}W_t^{x,i}\Big), && X_0^i = x_0, \quad 1\le i\le N, \\ \mathrm{d}Y_t^i &= -\frac{\kappa}{\epsilon}Y_t^i\,\mathrm{d}t + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\sqrt{1-\rho_y^2}\,\mathrm{d}W_t^{y,i}, && Y_0^i = y_0, \quad 1\le i\le N, \\ \mathrm{d}U_t &= -\frac{\kappa}{\epsilon}U_t\,\mathrm{d}t + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\rho_y\,\mathrm{d}W_t^y, && U_0 = 0, \end{aligned} \tag{5.3}$$

where N represents the number of firms, $W^{x,i}$ and $W^{y,i}$ are standard Brownian motions, independent from each other, and (W^x, W^y) is a bivariate Brownian motion with correlation ρ_{xy}, independent of the other Brownian motions, and $\sigma(\cdot)$ satisfies $\mathbb{E}\big[\int_0^T \sigma^2(Y_s + U_s)\,\mathrm{d}s\big] < \infty$. We can also generalise this model as a particle system, where each X^i represents a particle, with volatility process $(Y^i + U^i)$. The derivation of the SPDE from the particle system can be found in the Appendix B. Note that this ϵ here measures how fast the process reverts to its mean, which is different from the ε in the previous chapters (note the different symbol "\varepsilon"), which denotes the root mean square error.

We aim to find an asymptotic expansion of the solution $v(T, x, y)$ to the SPDE (5.1) as $\epsilon \to 0$, and consider the functional

$$P_T(v) = \int_0^\infty \int_{-\infty}^\infty v(T, x, y)\,\mathrm{d}y\,\mathrm{d}x, \qquad T < \infty. \tag{5.4}$$

5.1 SPDE asymptotic expansion

In this section, we heuristically derive the leading order terms and verify the convergence numerically in the following sections. We rewrite the SPDE (5.1) as

$$\begin{aligned}\mathrm{d}v &= \frac{1}{\epsilon}\Big(\nu^2(1-\rho_y^2)\partial_{yy}v + \kappa\partial_y(yv)\Big)\,\mathrm{d}t \\ &\quad + \Big(\frac{1}{2}\sigma^2\big(y+U_t\big)\partial_{xx}v - \mu\,\partial_x v\Big)\,\mathrm{d}t - \rho_x\sigma\big(y+U_t\big)\partial_x v\,\mathrm{d}W_t^x \\ &= \Big(\frac{1}{\epsilon}\mathcal{L}_0 + \mathcal{L}_2\Big)v\,\mathrm{d}t - \rho_x\sigma(y+U_t)\partial_x v\,\mathrm{d}W_t^x,\end{aligned} \tag{5.5}$$

where

$$\mathcal{L}_0(v) \coloneqq \nu^2(1-\rho_y^2)\partial_{yy}v + \kappa\partial_y(yv), \qquad \mathcal{L}_2(v) \coloneqq \frac{1}{2}\sigma^2\big(y+U_t\big)\partial_{xx}v - \mu\,\partial_x v.$$

5.1.1 Zero order term

We first ignore the transition from the initial condition to the stationary distribution in a short period after $t = 0$. We will later include the initial transient of the process Y_t^i and U_t.

We expect that as $\epsilon \to 0$, Y_t^i and U_t in the system (5.3) converge to the invariant distribution for fixed $t > 0$,

$$\begin{aligned} Y_t^i &\sim N\bigg(0, \frac{(1-\rho_y^2)\nu^2}{\kappa}\bigg), \\ U_t &\sim N\bigg(0, \frac{\rho_y^2\nu^2}{\kappa}\bigg), \end{aligned} \tag{5.6}$$

where Y_t^i are i.i.d. across i and independent of U_t, and both Y_t^i and U_t are independent across t.

If we first focus on the marginal density in x and consider the ergodic average of the coefficients in (5.5) in the y-component, we arrive at

$$\mathrm{d}v_0^x = \Big(\frac{1}{2}\big\langle\sigma^2(\cdot + U_t)\big\rangle_Y \partial_{xx} v_0^x - \mu\,\partial_x v_0^x\Big)\mathrm{d}t - \rho_x\big\langle\sigma(\cdot + U_t)\big\rangle_Y \partial_x v_0^x\,\mathrm{d}W_t^x,$$

where

$$\langle f\rangle_Y = \int_{-\infty}^{\infty} f(y)\Phi_\infty^Y(y)\,\mathrm{d}y \tag{5.7}$$

is the average over the stationary distribution of Y^i, with density

$$\Phi_\infty^Y(y) = \frac{\sqrt{\kappa}}{\nu\sqrt{2(1-\rho_y^2)\pi}}\exp\left(-\frac{\kappa y^2}{2(1-\rho_y^2)\nu^2}\right). \tag{5.8}$$

Now we also include the ergodic average over the process U_t. Given the invariant distribution of Y and U in (5.6),

$$\begin{aligned} \big\langle\langle f(\cdot + U)\rangle_Y\big\rangle_U &= \iint_{\mathbb{R}^2} f(y+u)\Phi_\infty^Y(y)\Phi_\infty^U(u)\,\mathrm{d}y\,\mathrm{d}u \\ &= \int_{-\infty}^{\infty} f(x)\int_{-\infty}^{\infty}\Phi_\infty^Y(y)\Phi_\infty^U(x-y)\,\mathrm{d}y\,\mathrm{d}x. \end{aligned}$$

Since the convolution of two Gaussian functions is also a Gaussian, with variance being the sum of the original variances, as a result,

$$\Phi(x) = \int_{-\infty}^{\infty}\Phi_\infty^Y(y)\Phi_\infty^U(x-y)\,\mathrm{d}y = \frac{\kappa}{\nu\sqrt{2\pi}}\exp\Big(-\frac{\kappa x^2}{2\nu^2}\Big).$$

The average over both stationary distributions has the form as follows,

$$\overline{f} \coloneqq \big\langle\langle f(\cdot + U)\rangle_Y\big\rangle_Z = \int_{-\infty}^{\infty} f(x)\Phi(x)\,\mathrm{d}x, \tag{5.9}$$

where

$$\Phi(x) = \frac{\kappa}{\nu\sqrt{2\pi}}\exp\Big(-\frac{\kappa x^2}{2\nu^2}\Big).$$

Averaging over both stationary distributions, we have the following conjectured limit SPDE

$$\begin{aligned} \mathrm{d}v_0^x &= \left(\frac{1}{2}\overline{\sigma^2}\partial_{xx}v_0^x - \mu\,\partial_x v_0^x\right)\mathrm{d}t - \rho_x\overline{\sigma}\partial_x v_0^x\,\mathrm{d}W_t^x \\ &\coloneqq \overline{\mathcal{L}}_2 v_0^x\,\mathrm{d}t - \rho_x\overline{\sigma}\partial_x v_0^x\,\mathrm{d}W_t^X, \\ v_0^x(0,x) &= \delta(x-x_0), \end{aligned} \tag{5.10}$$

where $\overline{\sigma}$ and $\overline{\sigma^2}$ are defined according to (5.9), and the operator $\overline{\mathcal{L}}_2$ is defined as

$$\overline{\mathcal{L}}_2(v) \coloneqq \frac{1}{2}\overline{\sigma^2}\partial_{xx}v - \mu\,\partial_x v. \tag{5.11}$$

Moreover, the SPDE (5.10) has the analytic solution $v_0^x(t,x) = \Psi(t,x)$, where

$$\Psi(t,x) = \frac{1}{\sqrt{2\pi\big(\overline{\sigma^2} - \rho_x^2\overline{\sigma}^2\big)t}}\exp\left(-\frac{(x - x_0 - \mu t - \rho_x\overline{\sigma}W_t^x)^2}{2\big(\overline{\sigma^2} - \rho_x^2\overline{\sigma}^2\big)t}\right). \tag{5.12}$$

Now we include the initial transient of the process Y_t^i to its stationary distribution. The marginal distribution of the OU process Y^i at time t is known to be

$$Y_t^i \sim N(\mu_y(t), \sigma_y^2(t)),$$

where $\mu_y(t)$ and $\sigma_y^2(t)$ have the form

$$\mu_y(t) = y_0\mathrm{e}^{-\kappa t/\epsilon}, \quad \sigma_y^2(t) = \frac{(1-\rho_y^2)\nu^2}{\kappa}\left(1 - \mathrm{e}^{-2\kappa t/\epsilon}\right). \tag{5.13}$$

Writing down the density Φ^Y explicitly,

$$\Phi^Y(t,y) = \frac{1}{\sqrt{2\pi\sigma_y^2(t)}}\exp\left(-\frac{\big(y - \mu_y(t)\big)^2}{2\sigma_y^2(t)}\right), \tag{5.14}$$

and it satisfies the PDE

$$\begin{aligned} \partial_t\Phi^Y &= \frac{1}{\epsilon}\mathcal{L}_0\Phi^Y \\ &= \frac{1}{\epsilon}\left(\nu^2(1-\rho_y^2)\partial_{yy}\Phi^Y + \kappa\partial_y\left(y\Phi^Y\right)\right), \\ \Phi^Y(0,y) &= \delta(y - y_0). \end{aligned}$$

Next, we define an approximation $v_{0,\epsilon}$, where we let $\epsilon \to 0$ in the x-terms of the SPDE, but keep ϵ fixed in the y-terms. In other words, we keep the marginal density of Y exact,

but approximate the density of X given Y, and hence the joint density. Therefore, we consider the equation

$$\begin{aligned} \mathrm{d}v_{0,\epsilon} &= \frac{1}{\epsilon}\left(\nu^2(1-\rho_y^2)\partial_{yy}v_{0,\epsilon} + \kappa\partial_y\left(yv_{0,\epsilon}\right)\right)\mathrm{d}t \\ &\quad + \left(\frac{1}{2}\overline{\sigma^2}\partial_{xx}v_{0,\epsilon} - \partial_x(\mu v_{0,\epsilon})\right)\mathrm{d}t - \rho_x\overline{\sigma}\partial_x v_{0,\epsilon}\,\mathrm{d}W_t^X \\ &:= \left(\frac{1}{\epsilon}\mathcal{L}_0 + \overline{\mathcal{L}}_2\right)v_{0,\epsilon}\,\mathrm{d}t - \rho_x\overline{\sigma}\partial_x v_{0,\epsilon}\,\mathrm{d}W_t^X, \end{aligned} \tag{5.15}$$

where $\overline{\sigma}$ and $\overline{\sigma^2}$ are defined according to (5.9), and $\overline{\mathcal{L}}_2$ is defined in (5.11).

Moreover, equation (5.15) has analytic solution

$$v_{0,\epsilon} = \Psi(t,x)\Phi^Y(t,y), \tag{5.16}$$

where $\Psi(t,x)$ is defined in (5.12), and $\Phi^Y(t,y)$ is defined in (5.14).

We will later show numerically that $\mathbb{E}\big[P_T(v_{0,\epsilon}) - P_T(v)\big] \to 0$ as $\epsilon \to 0$, where $P_T(\cdot)$ is the functional (5.4).

5.1.2 Correction terms

We now derive an inhomogeneous SPDE for $v - v_{0,\epsilon}$. Taking the difference between (5.5) and (5.15), we obtain an SPDE for $v - v_{0,\epsilon}$ as

$$\begin{aligned} \mathrm{d}(v - v_{0,\epsilon}) &= \Big(\frac{1}{\epsilon}\mathcal{L}_0 + \mathcal{L}_2\Big)(v - v_{0,\epsilon})\,\mathrm{d}t - \rho_x\sigma(y+U_t)\partial_x(v - v_{0,\epsilon})\,\mathrm{d}W_t^x \\ &\quad - \frac{1}{2}\Big(\overline{\sigma^2} - \sigma^2(y+U_t)\Big)\partial_{xx}v_{0,\epsilon}\,\mathrm{d}t + \rho_x\big(\overline{\sigma} - \sigma(y+U_t)\big)\partial_x v_{0,\epsilon}\,\mathrm{d}W_t^x, \\ (v - v_{0,\epsilon})(0,x,y) &= 0. \end{aligned} \tag{5.17}$$

Replacing $\mathcal{L}_2$ by $\overline{\mathcal{L}}_2$ and $\sigma(y+U_t)$ by $\overline{\sigma}$ in the first line of (5.17) as previously,

$$\begin{aligned} \mathrm{d}v_{1,\epsilon} &= \Big(\frac{1}{\epsilon}\mathcal{L}_0 + \overline{\mathcal{L}}_2\Big)v_{1,\epsilon}\,\mathrm{d}t - \rho_x\overline{\sigma}\partial_x v_{1,\epsilon}\,\mathrm{d}W_t^x \\ &\quad - \frac{1}{2}\Big(\overline{\sigma^2} - \sigma^2(y+U_t)\Big)\partial_{xx}v_{0,\epsilon}\,\mathrm{d}t + \rho_x\big(\overline{\sigma} - \sigma(y+U_t)\big)\partial_x v_{0,\epsilon}\,\mathrm{d}W_t^x, \\ v_{1,\epsilon}(0,x,y) &= 0. \end{aligned} \tag{5.18}$$

In this definition, coming from (5.17) which describes the exact error, we have approximated $\mathcal{L}$ by $\overline{\mathcal{L}}_2$ and σ by $\overline{\sigma}$.

Given $v_{0,\epsilon}$ in (5.15) and $v_{1,\epsilon}$ in (5.18), we can recursively approximate the higher order corrections as follows. The density $v_{n+1,\epsilon}(t,x,y)$ is defined as the leading order approximation to $\big(v - \sum_{i=0}^{n} v_{i,\epsilon}\big)$,

$$\begin{aligned} \mathrm{d}v_{n+1,\epsilon} &= \Big(\frac{1}{\epsilon}\mathcal{L}_0 + \overline{\mathcal{L}}_2\Big)v_{n+1,\epsilon}\,\mathrm{d}t - \rho_x\overline{\sigma}\partial_x v_{n+1,\epsilon}\,\mathrm{d}W_t^x \\ &\quad - \frac{1}{2}\Big(\overline{\sigma^2} - \sigma^2(y+U_t)\Big)\partial_{xx}v_{n,\epsilon}\,\mathrm{d}t + \rho_x\big(\overline{\sigma} - \sigma(y+U_t)\big)\partial_x v_{n,\epsilon}\,\mathrm{d}W_t^x, \\ v_{n+1,\epsilon}(0,x,y) &= 0, \end{aligned} \tag{5.19}$$

5.1.3 Approximation to the marginal density of X

One would hope that the expansion allows the computation of an approximate solution cheaper than solving the original two-dimensional SPDE.

This is not clear when considering (5.18) directly, as the solution still depends on x and y. Even if dropping $\mathcal{L}_0$ in (5.18), the SPDE is still parametrised by y, so essentially two-dimensional if the solution for all x and y is needed.

However, if one is only interested in the marginal density of X,

$$v^x(t,x) = \int_{-\infty}^{\infty} v(t,x,y)\,\mathrm{d}y,$$

integrating (5.15) over y gives the zero order approximation

$$v_0^x = \int_{-\infty}^{\infty} v_{0,\epsilon}(t,x,y)\,\mathrm{d}y = \Psi(t,x),$$

where $\Psi(t,x)$ is given by (5.12).

As for the first order term $v_{1,\epsilon}$, we similarly define v_1^x as

$$v_1^x = \int_{-\infty}^{\infty} v_{1,\epsilon}(t,x,y)\,\mathrm{d}y.$$

Assuming $\lim_{y\to\pm\infty} v_{1,\epsilon}(t,x,y) = 0$, $\lim_{y\to\pm\infty}\partial_y v_{1,\epsilon}(t,x,y) = 0$, it follows that

$$\int_{-\infty}^{\infty} \mathcal{L}_0 v_{1,\epsilon}\,\mathrm{d}y = 0.$$

Moreover, from (5.16), (5.14), and (5.8),

$$\begin{aligned} \int_{-\infty}^{\infty} \sigma^2(y+U_t)\partial_{xx}v_{0,\epsilon}(t,x,y)\,\mathrm{d}y &= \int_{-\infty}^{\infty} \sigma^2(y+U_t)\partial_{xx}\Psi(t,x)\Phi^Y(t,y)\,\mathrm{d}y \\ &= \partial_{xx}\Psi(t,x)\int_{-\infty}^{\infty}\sigma^2(y+U_t)\Phi_\infty^Y(y)\,\mathrm{d}y + o(\epsilon^r) \\ &= \langle\sigma^2(\cdot+U_t)\rangle_Y\partial_{xx}\Psi(t,x) + o(\epsilon^r), \qquad \forall r>0, \\ \int_{-\infty}^{\infty} \sigma(y+U_t)\partial_x v_{0,\epsilon}(t,x,y)\,\mathrm{d}y &= \int_{-\infty}^{\infty} \sigma(y+U_t)\partial_x\Psi(t,x)\Phi_\infty^Y(y)\,\mathrm{d}y + o(\epsilon^r) \\ &= \langle\sigma(\cdot+U_t)\rangle_Y\partial_x\Psi(t,x) + o(\epsilon^r), \qquad \forall r>0. \end{aligned} \tag{5.20}$$

Integrating (5.18) over y yields an SPDE for a first order approximation to $v^x(t,x)$,

$$\begin{aligned} \mathrm{d}v_1^x &= \overline{\mathcal{L}}_2 v_1^x \,\mathrm{d}t - \rho_x \overline{\sigma} \partial_x v_1^x \,\mathrm{d}W_t^x \\ &\quad - \frac{1}{2}\left(\overline{\sigma^2} - \langle \sigma^2(\cdot + U_t)\rangle_Y\right) \partial_{xx}\Psi(t,x)\,\mathrm{d}t \\ &\quad + \rho_x \left(\overline{\sigma} - \langle \sigma(\cdot + U_t)\rangle_Y\right) \partial_x \Psi(t,x)\,\mathrm{d}W_t^x, \\ v_1^x(0,x) &= 0. \end{aligned} \tag{5.21}$$

Corresponding to the functional $P_T(v)$ in (5.4), we consider the following functional for v_1^x,

$$P_T^x(v_1^x) = \int_0^\infty v_1^x(T,x)\,\mathrm{d}x, \qquad T < \infty. \tag{5.22}$$

5.2 Adaptation of numerical schemes

In this section, we present numerical schemes for solving the SPDEs from the previous section. We apply a Milstein ADI scheme similar to that in Section 3.1.1. First, we introduce the numerical scheme for the SPDE (5.5), repeated here for convenience,

$$\begin{aligned} \mathrm{d}v &= \frac{1}{\epsilon}\Big(\nu^2(1-\rho_y^2)\partial_{yy}v + \kappa\partial_y(yv)\Big)\,\mathrm{d}t \\ &\quad + \Big(\frac{1}{2}\sigma^2\big(y+U_t\big)\partial_{xx}v - \mu\,\partial_x v\Big)\,\mathrm{d}t - \rho_x\sigma\big(y+U_t\big)\partial_x v\,\mathrm{d}W_t^x. \end{aligned}$$

We use a spatial grid with uniform spacing h_x, $h_y > 0$, and, for $T > 0$ fixed, N time steps of size $k = T/N$. Let $V_{i,j}^n$ be the approximation to $v(nk, ih_x, jh_y)$, $n = 1, \ldots, N$, $i, j \in \mathbb{Z}$.

We apply a Milstein ADI scheme to the SPDE (5.5), and an Euler scheme to the process U_t (5.2).

We simulate the Ornstein–Uhlenbeck process (5.2) with timestep $k\epsilon$ such that the simulated process has the strong error $O(k)$ for a small ϵ.

First, we simulate the bivariate standard normal random variables $(Z_{m,x}, Z_{m,y})$ with correlation ρ_{xy}, where $m = 1, 2, \ldots, NN_\epsilon$, and $N_\epsilon = 1/\epsilon$, where we assume for simplicity $1/\epsilon \in \mathbb{Z}$. The discrete-time approximation of U_t is thus

$$\begin{aligned} \widehat{U}_m &= \widehat{U}_{m-1} - \frac{\kappa}{\epsilon}\,k\,\widehat{U}_{m-1} + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\rho_y\Big(W_m^y - W_{m-1}^y\Big) \\ &= \widehat{U}_{m-1} - \frac{\kappa}{\epsilon}\,k\,\widehat{U}_{m-1} + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\rho_y\sqrt{k}\Big(Z_{m,y} - Z_{m-1,y}\Big) \\ \widehat{U}_0 &= 0, \end{aligned} \tag{5.23}$$

where $m = 1, 2, \ldots, NN_\epsilon$, $t_m - t_{m-1} = k\epsilon$. Then we take

$$U^n = \widehat{U}_{nN_\epsilon}, \qquad n = 1, 2, \cdots, N, \tag{5.24}$$

as the approximation to the process U_t at time $t = nk$, i.e., at the n-th timestep of the coarser time mesh with width k, on which we approximate the SPDE.

The Brownian increment of W^x over a "large" timestep k is thus, in law,

$$W^x_{nk} - W^x_{(n-1)k} = \sqrt{k}\widetilde{Z}_{n,x} \coloneqq \sum_{i=1}^{N_\epsilon} \sqrt{k\epsilon} Z_{N_\epsilon(n-1)+i,x}.$$

Hence, we get

$$\widetilde{Z}_{n,x} = \sqrt{\epsilon} \sum_{i=1}^{N_\epsilon} Z_{N_\epsilon(n-1)+i,x}.$$

To simplify the notation, we write $Z_{n,x}$ instead of $\widetilde{Z}_{n,x}$ in the following. We implement the SPDE (5.5) with a Milstein ADI scheme as below:

$$\begin{aligned} &\left(I - \frac{\nu^2(1-\rho_y^2)}{\epsilon}\frac{k}{h_y^2}D_{yy} - \frac{\kappa}{\epsilon}\frac{k}{2h_y}D_y Y\right)\left(I - \frac{\sigma^2(Y+U^{n+1})}{2}\frac{k}{h_x^2}D_{xx} + \mu\frac{k}{2h_x}D_x\right)V^{n+1} \\ &= \left(I - \rho_x\sigma(Y+U^n)\frac{\sqrt{k}Z_{n,x}}{2h_x}D_x + \rho_x^2\sigma^2(Y+U^n)\frac{k(Z_{n,x}^2-1)}{2h_x^2}D_{xx}\right)V^n, \end{aligned} \tag{5.25}$$

with V^0 being the approximation of the initial condition to the SPDE (5.5), and the process U_t simulated as U^n, by (5.23), (5.24), and

$$\begin{aligned} (D_x V)_{i,j} &= V_{i+1,j} - V_{i-1,j}, & (D_y V)_{i,j} &= V_{i,j+1} - V_{i,j-1}, \\ (D_{xx} V)_{i,j} &= V_{i+1,j} - 2V_{i,j} + V_{i-1,j}, & (D_{yy} V)_{i,j} &= V_{i,j+1} - 2V_{i,j} + V_{i,j-1}, \\ (D_{xy} V)_{i,j} &= V_{i+1,j+1} - V_{i-1,j+1} - V_{i+1,j-1} + V_{i-1,j-1}, \end{aligned}$$

and where Y is the diagonal matrix such that each element of the diagonal corresponds to the y-coordinate of the mesh point ordered the same way as V.

For the zero order term $v_{0,\epsilon}$, we denote the corresponding numerical solution as V_0, with the scheme

$$\begin{aligned} &\left(I - \frac{\nu^2(1-\rho_y^2)}{\epsilon}\frac{k}{h_y^2}D_{yy} - \frac{\kappa}{\epsilon}\frac{k}{2h_y}D_y Y\right)\left(I - \frac{\overline{\sigma^2}}{2}\frac{k}{h_x^2}D_{xx} + \mu\frac{k}{2h_x}D_x\right)V_0^{n+1} \\ &= \left(I - \rho_x\overline{\sigma}\frac{\sqrt{k}Z_{n,x}}{2h_x}D_x + \rho_x^2\overline{\sigma}^2\frac{k(Z_{n,x}^2-1)}{2h_x^2}D_{xx}\right)V_0^n, \end{aligned} \tag{5.26}$$

where the initial condition V_0^0 is an approximation of the initial condition to the SPDE (5.5). Note that the closed form solution to V_0 is (5.16), and $\overline{\sigma}$, $\overline{\sigma^2}$ will be computed analytically for specific choices of $\sigma(\cdot)$.

Alternatively, we can directly solve the SPDE for $v - v_{0,\epsilon}$ from (5.17). Denoting the solution as $\widetilde{V}_0$, we have the scheme

$$\begin{aligned}
&\Big(I-\frac{\nu^2(1-\rho_y^2)}{\epsilon}\frac{k}{h_y^2}D_{yy}-\frac{\kappa}{\epsilon}\frac{k}{2h_y}D_yY\Big)\Big(I-\frac{\sigma^2(Y+U^{n+1})}{2}\frac{k}{h_x^2}D_{xx}+\mu\frac{k}{2h_x}D_x\Big)\widetilde{V}_0^{n+1}\\
&=\Big(I-\rho_x\sigma(Y+U^n)\frac{\sqrt{k}Z_{n,x}}{2h_x}D_x+\rho_x^2\sigma^2(Y+U^n)\frac{k(Z_{n,x}^2-1)}{2h_x^2}D_{xx}\Big)\widetilde{V}_0^{n+1}\\
&\quad-\frac{1}{2}\Big(\overline{\sigma^2}-\sigma^2(Y+U_t)\Big)k\big(\partial_{xx}V_0^n\big)+\rho_x\Big(\overline{\sigma}-\sigma(Y+U_t)\Big)\big(\partial_xV_0^n\big)\sqrt{k}Z_{n,x}\\
&\quad+\frac{1}{2}\rho_x^2\Big(\overline{\sigma}-\sigma(Y+U_t)\Big)^2\big(\partial_{xx}V_0^n\big)k(Z_{n,x}^2-1),
\end{aligned}\tag{5.27}$$

where we use the analytic solution for $\partial_x V_0$ and $\partial_{xx}V_0$ from (5.16) in the scheme for $\widetilde{V}_0$, with initial condition $\widetilde{V}_0^0 = 0$.

Similarly, the scheme for the approximation of the first order term (5.18), denoted by V_1, is

$$\begin{aligned}
&\Big(I-\frac{\nu^2(1-\rho_y^2)}{\epsilon}\frac{k}{h_y^2}D_{yy}-\frac{\kappa}{\epsilon}\frac{k}{2h_y}D_yY\Big)\Big(I-\frac{\overline{\sigma^2}}{2}\frac{k}{h_x^2}D_{xx}+\mu\frac{k}{2h_x}D_x\Big)V_1^{n+1}\\
&=\Big(I-\rho_x\overline{\sigma}\frac{\sqrt{k}Z_{n,x}}{2h_x}D_x+\rho_x^2\overline{\sigma}^2\frac{k(Z_{n,x}^2-1)}{2h_x^2}D_{xx}\Big)V_1^n\\
&\quad-\frac{1}{2}\Big(\overline{\sigma^2}-\sigma^2(Y+U_t)\Big)k\big(\partial_{xx}V_0^n\big)+\rho_x\Big(\overline{\sigma}-\sigma(Y+U_t)\Big)\big(\partial_xV_0^n\big)\sqrt{k}Z_{n,x}\\
&\quad+\frac{1}{2}\rho_x^2\Big(\overline{\sigma}-\sigma(Y+U_t)\Big)^2\big(\partial_{xx}V_0^n\big)k(Z_{n,x}^2-1),
\end{aligned}\tag{5.28}$$

with zero initial condition, $V_1^0 = 0$. We also use the analytic solution for $\partial_x V_0$ and $\partial_{xx}V_0$ in the scheme (5.28).

As for the marginal SPDE for $v_1^x(t,x)$ from (5.21), the scheme we use is

$$\begin{aligned}
&\Big(I-\frac{\overline{\sigma^2}}{2}\frac{k}{h_x^2}D_{xx}+\mu\frac{k}{2h_x}D_x\Big)V_{1,x}^{n+1}\\
&=\Big(I-\rho_x\overline{\sigma}\frac{\sqrt{k}Z_{n,x}}{2h_x}D_x+\rho_x^2\overline{\sigma}^2\frac{k(Z_{n,x}^2-1)}{2h_x^2}D_{xx}\Big)V_{1,x}^n\\
&\quad-\frac{1}{2}\Big(\overline{\sigma^2}-\langle\sigma^2(\cdot+U_t)\rangle_Y\Big)k\Psi_{xx}(nk,X)+\rho_x\Big(\overline{\sigma}-\langle\sigma(\cdot+U_t)\rangle_Y\Big)\Psi_x(nk,X)\sqrt{k}Z_{n,x}\\
&\quad+\frac{1}{2}\rho_x^2\Big(\overline{\sigma}-\langle\sigma(\cdot+U_t)\rangle_Y\Big)^2\Psi_{xx}(nk,X)k(Z_{n,x}^2-1),
\end{aligned}\tag{5.29}$$

where $\Psi_x(nk,X)$ and $\Psi_{xx}(nk,X)$ are the value of the functions applying on the grid X.

In the computations below, we make some further specifications.

Here we choose $\sigma(x) = \exp(\alpha x)$, where $\alpha > 0$, such that the particle system follows an

exponential OU process. Then from (5.9), we have

$$\begin{aligned}\overline{\sigma} &= \int_{-\infty}^{\infty} \frac{\kappa}{\nu\sqrt{2\pi}} \exp\Big(\alpha x - \frac{\kappa x^2}{2\nu^2}\Big)\,\mathrm{d}x \;\; = \exp\Big(\frac{\alpha^2\nu^2}{2\kappa}\Big),\\ \overline{\sigma^2} &= \int_{-\infty}^{\infty} \frac{\kappa}{\nu\sqrt{2\pi}} \exp\Big(2\alpha x - \frac{\kappa x^2}{2\nu^2}\Big)\,\mathrm{d}x = \exp\Big(\frac{4\alpha^2\nu^2}{2\kappa}\Big).\end{aligned} \tag{5.30}$$

For $\sigma(x) = \exp(\alpha x)$, from (5.7) we have

$$\begin{aligned}\langle\sigma(\cdot + U_t)\rangle_Y &= \int_{-\infty}^{\infty} \frac{\kappa}{\nu\sqrt{2\pi}} \exp\Big(\alpha(x+U_t) - \frac{\kappa x^2}{2\nu^2(1-\rho_y^2)}\Big)\,\mathrm{d}x \;\; = \exp\left(\frac{\alpha^2\nu^2(1-\rho_y^2)}{2\kappa} + \alpha U_t\right),\\ \langle\sigma^2(\cdot + U_t)\rangle_Y &= \int_{-\infty}^{\infty} \frac{\kappa}{\nu\sqrt{2\pi}} \exp\Big(2\alpha(x+U_t) - \frac{\kappa x^2}{2\nu^2(1-\rho_y^2)}\Big)\,\mathrm{d}x = \exp\left(\frac{4\alpha^2\nu^2(1-\rho_y^2)}{2\kappa} + 2\alpha U_t\right).\end{aligned}$$

To approximate $P_t(v)$ in (5.4), we use the trapezoidal rule for the numerical integration.

5.3 Numerical results

In this section, we illustrate the convergence of the expansion for the model (5.3) by way of empirical tests. Instead of using the Dirac initial condition, we use the following Gaussian distribution as the initial condition

$$v(0,x,y) = \Psi\big(t_0, x\,; W_{t_0}^x = 0\big)\Phi^Y(t_0, y), \tag{5.31}$$

where $t_0 > 0$ is a fixed constant, and Ψ, Φ are defined in (5.12) and (5.14). We choose this smooth initial condition so that when we numerically approximate $v_{1,\epsilon}$ in (5.18), the approximation is stable when taking the partial derivatives of $v_{0,\epsilon}$. Moreover, we can still obtain the analytic solution to $v_{0,\epsilon}$ in (5.15),

$$v_{0,\epsilon}(T,x,y) = \Psi\big(t_0+T, x\,; W_{t_0}^x = 0\big)\Phi^Y(t_0+T, y). \tag{5.32}$$

As a base case, we choose parameters $T = 1,\ x_0 = y_0 = 2,\ \mu = 0.05,\ \rho_x = 0.3,\ \rho_y = 0.2,$ $\rho_{xy} = 0.5,\ \kappa = 0.2,\ \nu = 0.5,\ \alpha = 0.1$, and t_0 in the initial condition (5.31) is chosen as 0.2. We then vary ϵ and estimate the contribution of $v_{0,\epsilon}$ and $v_{1,\epsilon}$ to expected functionals of the solution. Later on, we will also test the effect of different parameters, in particular negative ρ_{xy}, and different ratios of κ and ν.

For the computations, we truncate the domain to $[-10,10]\times[-10,10]$, chosen large enough that the effect of zero Dirichlet boundary conditions on the solution is negligible for the tested parameter values.

5.3.1 Convergence of $\mathbb{E}[P_T(v) - P_T(v_{0,\epsilon})]$

We first analyse numerical results for the convergence of $\mathbb{E}[P_T(v) - P_T(v_{0,\epsilon})]$, where P_T is the functional from (5.4). Since P_T is linear, we have $P_T(v) - P_T(v_{0,\epsilon}) = P_T(v - v_{0,\epsilon})$. We use the scheme (5.27) to directly approximate $v - v_{0,\epsilon}$, and use a standard Monte Carlo method to estimate $\mathbb{E}[P_T(v-v_{0,\epsilon})]$ as detailed below. While simulating $(v-v_{0,\epsilon})$ from (5.17) and simulating v in (5.5) and $v_{0,\epsilon}$ in (5.15) separately lead to the same $\mathbb{E}[P_T(v - v_{0,\epsilon})]$ for $h_x, h_y, k \to 0$, the former choice has computational savings and is faster by a constant factor.

Given M Brownian paths, the Monte Carlo estimator for $\mathbb{E}[P_T(v - v_{0,\epsilon})]$ is

$$\Delta\widehat{P}_0 = \frac{1}{M}\sum_{i=1}^{M} P_T(\widetilde{V}_0^{(i)}),$$

where $\widetilde{V}_0^{(i)}$ is the numerical solution to $v - v_{0,\epsilon}$, with the scheme (5.27) for the i-th path.

For each ϵ, the numerical error between $\Delta\widehat{P}_0$ and $\mathbb{E}[P_T(v - v_{0,\epsilon})]$ consists of discretisation error in h and k, and Monte Carlo error. If $h,\ k \to 0$, and $M \to \infty$, $\Delta\widehat{P}_0$ is thus an accurate approximation to $\mathbb{E}[P_T(v - v_{0,\epsilon})]$, and we treat $\Delta\widehat{P}_0$ as the weak error in ϵ for small $h,\ k$ and large M.

Figure 5.1 shows the convergence to zero of $\Delta\widehat{P}_0$, with $\epsilon = 0.1\times 2^{-3},\ 0.1\times 2^{-4}, \ldots, 0.1\times 2^{-7}$, and the error bar with 3 standard deviations, where the exact numbers are shown in Table 5.3 in Section 5.3.3. For each ϵ, we let $h_x = h_y = 2^{-l}$, and $k = 0.5 \cdot 4^{-l}$, where $l = 1,\ 2,\ 3,\ 4$. We use $M = 10^6$ for $l = 1,\ 2$, and $M = 10^5$ for $l = 3,\ 4$, as the computational cost for finer meshes is large. For $l = 4$ with $M = 10^5$, the run time of the Matlab code is up to around 72 hours with 36 cores in parallel (speed 2300 RPMs, RAM 768GB, Linux system). We take the results from $l = 4$ as numerical approximation to $\mathbb{E}[P_T(v - v_{0,\epsilon})]$, shown as the black solid line in the loglog plot in Figure 5.1, comparing it to a dashed line with slope 1/2. We can see in the figure that the slope is slightly less than 1/2, and a linear regression shows that the slope is 0.4237. The reason why it is less than 0.5 is that $l = 4$ is not large enough, hence the error in h and k is not small enough to be neglected. We deduce empirically that the weak error $\mathbb{E}[P_T(v - v_{0,\epsilon})]$ is $O(\sqrt{\epsilon})$, which is stated in the Conjecture 5.3.1 below.

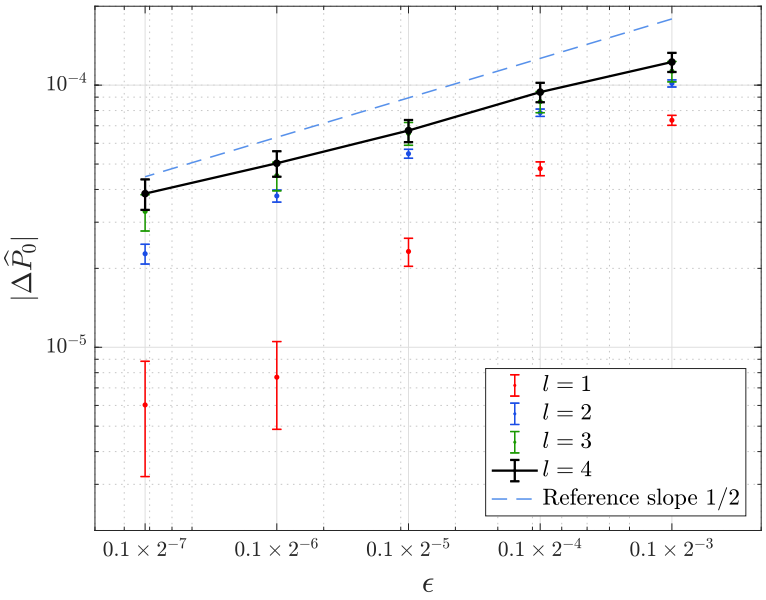


Figure 5.1: Convergence of $\Delta \widehat{P}_0$.

Conjecture 5.3.1. *The zero order approximation* $v_{0,\epsilon}(T, x, y)$ *in* (5.15) *converges to* $v(T, x, y)$ *in* (5.5) *of order* $1/2$ *in* ϵ *in the weak sense, such that*

$$\mathbb{E}\Big[P_T(v) - P_T(v_{0,\epsilon})\Big] = \mathbb{E}\Big[\int_0^\infty \int_{-\infty}^\infty v(T, x, y) - v_{0,\epsilon}(T, x, y)\,\mathrm{d}y\,\mathrm{d}x\Big] = O(\sqrt{\epsilon}\,). \tag{5.33}$$

5.3.2 Convergence of $\mathbb{E}[P_T(v_{1,\epsilon})]$

Next, we analyse the convergence of $\mathbb{E}[P_T(v_{1,\epsilon})]$ to zero, where $v_{1,\epsilon}$ satisfies the SPDE (5.18). We also compare it with $P_T^x(v_1^x)$, where v_1^x is the solution to the marginal SPDE (5.21), $P_T^x(\cdot)$ is defined in (5.22), and we expect these two values should be approximately the same.

From the derivation of the SPDE for v_1^x in Section 5.1.3, when we integrate over the y-dimension in (5.20), we replace $\Phi^Y(T, y)$ in the analytic solution (5.16) by the invariant distribution $\Phi_\infty^Y(y)$, which yields a more concise analytical form. This does not change the convergence order, as for $\epsilon \to 0$, $\Phi^Y(T, y)$ converges to $\Phi_\infty^Y(y)$ exponentially fast. Hence, to make sure $P_T(v_{1,\epsilon})$ gives the same results as $P_T(v_1^x)$, we also use $v_0(T, x, y) = \Psi(T, x)\Phi_\infty^Y(y)$ as the analytic solution to the zero order term $v_{0,\epsilon}$ in the schemes (5.27) and (5.28).

Given M Brownian paths, the Monte Carlo estimators for $\mathbb{E}[P_T(v_{1,\epsilon})]$ and $\mathbb{E}[P_T(v_1^x)]$ are

$$\widehat{P}_1 = \frac{1}{M}\sum_{i=1}^{M} P_T(V_1^{(i)}), \qquad \widehat{P}_1^x = \frac{1}{M}\sum_{i=1}^{M} P_T\big((V_1^x)^{(i)}\big), \tag{5.34}$$

where V_1 follows the scheme (5.28), and V_1^x follows the scheme (5.29).

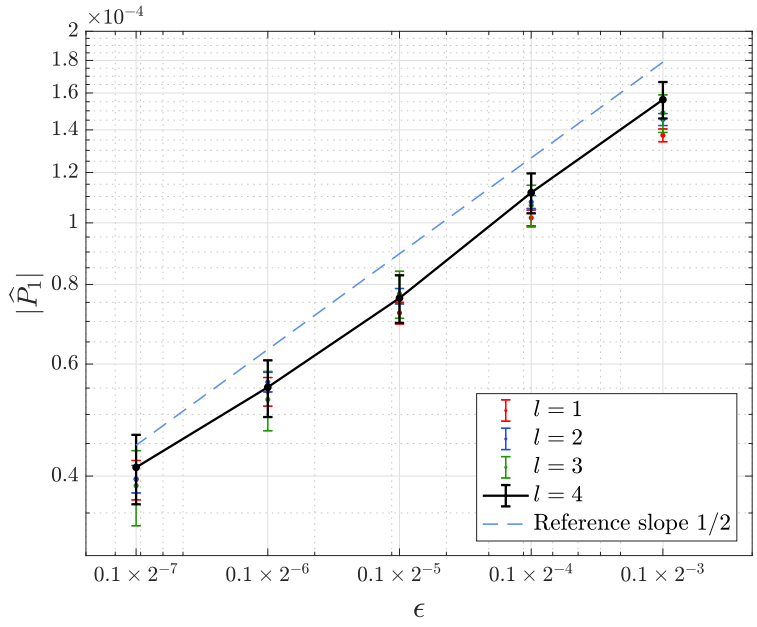


Figure 5.2: $\widehat{P}_1$ defined in (5.34) for $h_x = h_y = 2^{-l}$, $k = 0.5{\cdot}4^{-l}$, $\epsilon = 0.1{\times}2^{-3}, \cdots, 0.1{\times}2^{-7}$.

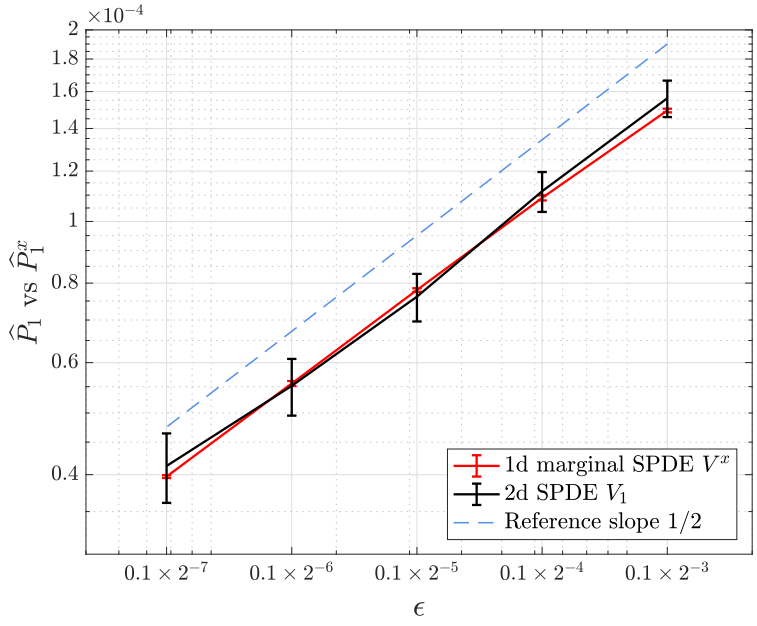


Figure 5.3: Comparing $\widehat{P}_1$ with $\widehat{P}_1^x$ from (5.34).

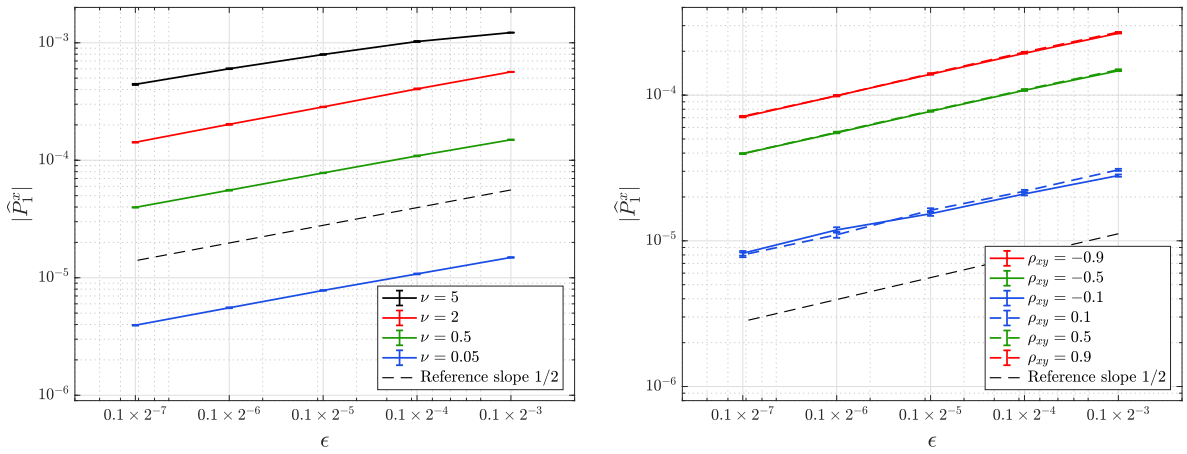
Figure 5.2 shows $\widehat{P}_1$, with $\epsilon = 0.1 \times 2^{-3},\ 0.1 \times 2^{-4}, \ldots, 0.1 \times 2^{-7}$, and the error bars with 3 standard deviations. Similarly as in Figure 5.1, for each ϵ, we let $h_x = h_y = 2^{-l}$, and $k = 0.5 \cdot 4^{-l}$, where $l = 1,\ 2,\ 3,\ 4$. We use $M = 10^6$ for $l = 1,\ 2$, and $M = 10^5$ for $l = 3,\ 4$. We take the results from $l = 4$ as numerical approximation to $\mathbb{E}[P_T(v_{1,\epsilon})]$, shown as the black solid line in the loglog plot in Figure 5.2, comparing it to a dashed line with slope 1/2. The exact numbers for Figure 5.2 are in Table 5.3 in Section 5.3.3.

We further include the approximation $\widehat{P}_1^x$ from the marginal SPDE (5.21), and compare $\widehat{P}_1^x$ with $\widehat{P}_1$ in Figure 5.3. As for $\widehat{P}_1^x$, we use a multilevel Monte Carlo method, with prescribed 1% relative error (the ratio between root mean-square error and true value). The details of the MLMC estimator are similar to previous chapters, so we do not repeat them here. However, unlike approximating $\mathbb{E}[P_T(v)]$ by a standard Monte Carlo method, and discretising with mesh size $h_l = h_0 \times 2^{-l}$ and timestep $k_l = k_0 \times 4^{-l}$, approximating by MLMC requires a good coupling between fine path and coarse path. Therefore, when we apply MLMC to estimate $\mathbb{E}[P_T(v_1^x)]$, the timestep is set as $k_l = k_0 \times \epsilon \times 4^{-l}$, proportional to ϵ.

Linear regression yields that the fitted slope for $\widehat{P}_1$ is 0.4855, and for $\widehat{P}_1^x$ it is 0.4792. We have made the conjecture that $P_T(v_{1,\epsilon})$ is the leading order approximation to $P_T(v) - P_T(v_{0,\epsilon})$, hence a necessary condition is $\mathbb{E}[P_T(v_{1,\epsilon})] = O(\sqrt{\epsilon}\,)$, which is satisfied from the numerical results. Moreover, since v_1^x is the solution to a one-dimensional SPDE, the computational cost is much lower for $\mathbb{E}[P_T(v_1^x)]$ with the same accuracy than for $\mathbb{E}[P_T(v_{1,\epsilon})]$, which shows the benefit of this asymptotic expansion.

We also test the effect of the ratio between ν and κ, and of the correlation ρ_{xy}. Figure 5.4(a) and Table 5.1 show the effect of different ν/κ, where we keep the relative error less than 1%. We test by changing ν among 0.05, 0.5, 2, 5, while keeping other parameters fixed. We can see from Figure 5.4(a) that $|\widehat{P}_1^x|$ increases as ν/κ increases, and from Table 5.1, we can even deduce that for a fixed ϵ, $|\widehat{P}_1^x(\nu/\kappa)| = O(\nu/\kappa)$, which is consistent with the previous tests where ϵ varies: $O((\nu/\sqrt{\epsilon})/(\kappa/\epsilon)) = O(\sqrt{\epsilon}\,\nu/\kappa)$. However, one point is that the values for $\nu = 5$ are negative, whereas the other cases are positive.

Figure 5.4(b) and Table 5.2 show the effect of changing the parameter ρ_{xy}, where we keep the relative error less than 5%. We can see from Figure 5.4(b) that $|\widehat{P}_1^x|$ increases as $|\rho_{xy}|$ increases, and from Table 5.2 that $\widehat{P}_1^x$ is positive when ρ_{xy} is positive, and vice versa. This effect is similar to the asymptotic expansion of the backward PDE for the stochastic volatility model [32], where the leading order correction term is proportional to the correlation between the two Brownian motions involved.



(a) $\widehat{P}_1^x$ with different ν.

(b) $\widehat{P}_1^x$ with different ρ_{xy}.

Figure 5.4: Comparing $\widehat{P}_1^x$ with regard to ν and ρ .

ϵ	$\nu = 0.05$	$\nu = 0.5$	$\nu = 2$	$\nu = 5$
0.1×2^{-3}	1.49×10^{-5} $\pm\ 1 \times 10^{-7}$	1.50×10^{-4} $\pm\ 1 \times 10^{-6}$	5.65×10^{-4} $\pm\ 2 \times 10^{-6}$	-1.22×10^{-3} $\pm\ 2 \times 10^{-6}$
0.1×2^{-4}	1.08×10^{-5} $\pm\ 1 \times 10^{-7}$	1.08×10^{-4} $\pm\ 1 \times 10^{-6}$	4.05×10^{-4} $\pm\ 4 \times 10^{-6}$	-1.03×10^{-3} $\pm\ 1 \times 10^{-5}$
0.1×2^{-5}	7.80×10^{-6} $\pm\ 5 \times 10^{-8}$	7.80×10^{-5} $\pm\ 5 \times 10^{-7}$	2.85×10^{-4} $\pm\ 2 \times 10^{-6}$	-7.95×10^{-4} $\pm\ 5 \times 10^{-6}$
0.1×2^{-6}	5.55×10^{-6} $\pm\ 5 \times 10^{-8}$	5.56×10^{-5} $\pm\ 5 \times 10^{-7}$	2.02×10^{-4} $\pm\ 2 \times 10^{-6}$	-6.01×10^{-4} $\pm\ 5 \times 10^{-6}$
0.1×2^{-7}	3.95×10^{-6} $\pm\ 3 \times 10^{-8}$	3.97×10^{-5} $\pm\ 2 \times 10^{-7}$	1.42×10^{-4} $\pm\ 1 \times 10^{-6}$	-4.43×10^{-4} $\pm\ 5 \times 10^{-6}$

Table 5.1: Test $\widehat{P}_1$ with confidence intervals (three standard deviations) for various ν, with other parameters kept fixed, 1% relative error.

ϵ	$\rho_{xy} = -0.9$	$\rho_{xy} = -0.5$	$\rho_{xy} = -0.1$	$\rho_{xy} = 0.1$	$\rho_{xy} = 0.5$	$\rho_{xy} = 0.9$
0.1×2^{-3}	-2.66×10^{-4} $\pm 1 \times 10^{-6}$	-1.47×10^{-4} $\pm 1 \times 10^{-6}$	-2.81×10^{-5} $\pm 5 \times 10^{-7}$	3.07×10^{-5} $\pm 5 \times 10^{-7}$	1.49×10^{-4} $\pm 1 \times 10^{-6}$	2.70×10^{-4} $\pm 1 \times 10^{-6}$
0.1×2^{-4}	-1.94×10^{-4} $\pm 1 \times 10^{-6}$	-1.08×10^{-4} $\pm 1 \times 10^{-6}$	-2.10×10^{-5} $\pm 5 \times 10^{-7}$	2.19×10^{-5} $\pm 5 \times 10^{-7}$	1.09×10^{-4} $\pm 1 \times 10^{-6}$	1.97×10^{-4} $\pm 1 \times 10^{-6}$
0.1×2^{-5}	-1.39×10^{-4} $\pm 1 \times 10^{-6}$	-7.74×10^{-5} $\pm 4 \times 10^{-7}$	-1.54×10^{-5} $\pm 5 \times 10^{-7}$	1.62×10^{-5} $\pm 5 \times 10^{-7}$	7.80×10^{-5} $\pm 5 \times 10^{-7}$	1.40×10^{-4} $\pm 1 \times 10^{-6}$
0.1×2^{-6}	-9.90×10^{-5} $\pm 9 \times 10^{-7}$	-5.52×10^{-5} $\pm 5 \times 10^{-7}$	-1.19×10^{-5} $\pm 5 \times 10^{-7}$	1.10×10^{-5} $\pm 5 \times 10^{-7}$	5.56×10^{-5} $\pm 5 \times 10^{-7}$	9.90×10^{-5} $\pm 9 \times 10^{-7}$
0.1×2^{-7}	-7.08×10^{-5} $\pm 5 \times 10^{-7}$	-3.96×10^{-5} $\pm 2 \times 10^{-7}$	-8.22×10^{-6} $\pm 3 \times 10^{-7}$	8.01×10^{-6} $\pm 3 \times 10^{-7}$	3.97×10^{-5} $\pm 2 \times 10^{-7}$	7.16×10^{-5} $\pm 5 \times 10^{-7}$

Table 5.2: Test $\widehat{P}_1$ with confidence intervals (three standard deviations) for various ρ_{xy}, with other parameters kept fixed, 5% relative error.

5.3.3 Convergence of $\mathbb{E}[P_T(v) - P_T(v_{0,\epsilon}) - P_T(v_{1,\epsilon})]$

To verify that $P_T(v_{1,\epsilon})$ is indeed the leading order approximation to $P_T(v) - P_T(v_{0,\epsilon})$, we further show the result of $\Delta\widehat{P}_0 - \widehat{P}_1$, which is the Monte Carlo estimator for $P_T(v - v_{0,\epsilon} - v_{1,\epsilon})$. Note that we have used the same Brownian paths for $\Delta\widehat{P}_0$ and $\widehat{P}_1$.

Figure 5.5 shows the loglog plot of the convergence of $\mathbb{E}[\Delta\widehat{P}_0 - \widehat{P}_1]$ to zero, with respect to ϵ, with the error bars being three standard deviations away. The exact numbers are given in Table 5.3. Similar as previous tests, here we stop at $l = 5$ with 1000 Monte Carlo samples, and linear regression yields the fitted slope as 1.0092, which empirically indicates first order convergence. Furthermore, in Table 5.3, we compare the convergence of $\Delta\widehat{P}_0$, $\widehat{P}_1$, and $\Delta\widehat{P}_0 - \widehat{P}_1$. It shows that under the same Brownian paths, the standard deviation is much smaller for $\Delta\widehat{P}_0 - \widehat{P}_1$ compared to $\Delta\widehat{P}_0$ and $\widehat{P}_1$.

We can deduce empirically that $P_T(v_{1,\epsilon})$ is the leading order approximation to $P_T(v - v_{0,\epsilon})$, with $P_T(v - v_{0,\epsilon} - v_{1,\epsilon}) = O(\epsilon)$. This is stated as the Conjecture 5.3.2 below.

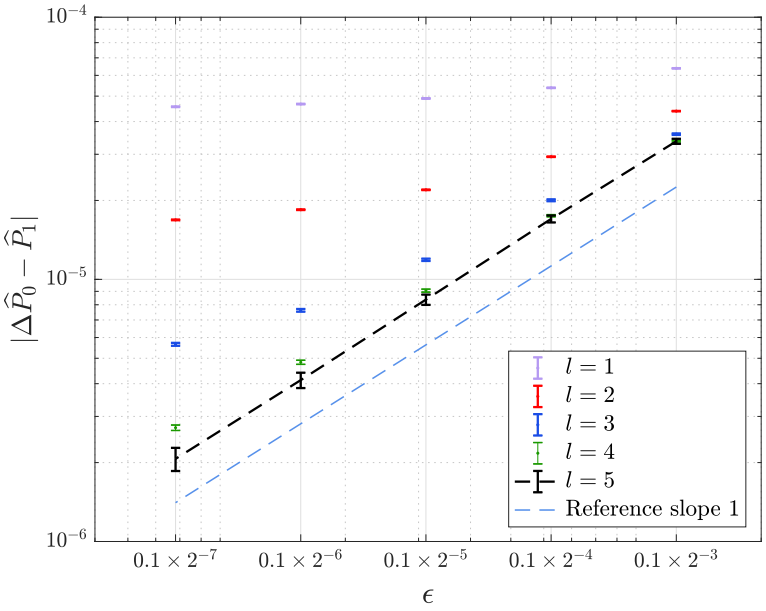


Figure 5.5: Convergence of $\Delta\widehat{P}_0 - \widehat{P}_1$.

ϵ	l	$\Delta\widehat{P}_0$		$\widehat{P}_1$		$\Delta\widehat{P}_0 - \widehat{P}_1$	
		mean	sd	mean	sd	mean	sd
0.1×2^{-3}	1	7.35×10^{-5}	1.07×10^{-6}	1.37×10^{-4}	1.06×10^{-6}	-6.38×10^{-5}	3.65×10^{-8}
	2	1.02×10^{-4}	1.05×10^{-6}	1.45×10^{-4}	1.05×10^{-6}	-4.38×10^{-5}	2.69×10^{-8}
	3	1.13×10^{-4}	3.36×10^{-6}	1.49×10^{-4}	3.37×10^{-6}	-3.58×10^{-5}	7.79×10^{-8}
	4	1.23×10^{-4}	3.42×10^{-6}	1.56×10^{-4}	3.43×10^{-6}	-3.37×10^{-5}	7.85×10^{-8}
	5					-3.36×10^{-5}	2.48×10^{-7}
0.1×2^{-4}	1	4.80×10^{-5}	9.71×10^{-7}	1.02×10^{-4}	9.64×10^{-7}	-5.38×10^{-5}	3.78×10^{-8}
	2	7.85×10^{-5}	8.43×10^{-7}	1.08×10^{-4}	8.40×10^{-7}	-2.94×10^{-5}	2.16×10^{-8}
	3	8.65×10^{-5}	2.67×10^{-6}	1.07×10^{-4}	2.67×10^{-6}	-2.00×10^{-5}	5.56×10^{-8}
	4	9.41×10^{-5}	2.67×10^{-6}	1.12×10^{-4}	2.68×10^{-6}	-1.75×10^{-5}	5.53×10^{-8}
	5					-1.70×10^{-5}	1.73×10^{-7}
0.1×2^{-5}	1	2.32×10^{-5}	9.47×10^{-7}	7.22×10^{-5}	9.41×10^{-7}	-4.90×10^{-5}	3.93×10^{-8}
	2	5.48×10^{-5}	7.16×10^{-7}	7.67×10^{-5}	7.13×10^{-7}	-2.20×10^{-5}	1.89×10^{-8}
	3	6.55×10^{-5}	2.18×10^{-6}	7.74×10^{-5}	2.19×10^{-6}	-1.19×10^{-5}	4.03×10^{-8}
	4	6.71×10^{-5}	2.17×10^{-6}	7.62×10^{-5}	2.18×10^{-6}	-9.06×10^{-6}	3.96×10^{-8}
	5					-8.37×10^{-6}	1.26×10^{-7}
0.1×2^{-6}	1	7.69×10^{-6}	9.42×10^{-7}	5.43×10^{-5}	9.38×10^{-7}	-4.66×10^{-5}	4.04×10^{-8}
	2	3.78×10^{-5}	6.67×10^{-7}	5.62×10^{-5}	6.64×10^{-7}	-1.84×10^{-5}	1.78×10^{-8}
	3	4.52×10^{-5}	1.88×10^{-6}	5.28×10^{-5}	1.88×10^{-6}	-7.62×10^{-6}	3.01×10^{-8}
	4	5.03×10^{-5}	1.87×10^{-6}	5.52×10^{-5}	1.88×10^{-6}	-4.83×10^{-6}	2.90×10^{-8}
	5					-4.13×10^{-6}	9.27×10^{-8}
0.1×2^{-7}	1	-6.03×10^{-6}	9.38×10^{-7}	3.95×10^{-5}	9.36×10^{-7}	-4.55×10^{-5}	4.11×10^{-8}
	2	2.27×10^{-5}	6.57×10^{-7}	3.96×10^{-5}	6.55×10^{-7}	-1.69×10^{-5}	1.74×10^{-8}
	3	3.30×10^{-5}	1.74×10^{-6}	3.86×10^{-5}	1.74×10^{-6}	-5.65×10^{-6}	2.42×10^{-8}
	4	3.86×10^{-5}	1.71×10^{-6}	4.13×10^{-5}	1.72×10^{-6}	-2.72×10^{-6}	2.16×10^{-8}
	5					-2.07×10^{-6}	6.98×10^{-8}

Table 5.3: Convergence of $\Delta\widehat{P}_0$, $\widehat{P}_1$, and $\Delta\widehat{P}_0 - \widehat{P}_1$.

Conjecture 5.3.2. *The first order term* $v_{1,\epsilon}(T, x, y)$ *in* (5.18) *converges to* $(v - v_{0,\epsilon})$ *in* (5.17) *with order* 1 *in* ϵ *in the weak sense, such that*

$$\begin{aligned}\mathbb{E}\Big[P_T(v) - P_T(v_{0,\epsilon}) - P_T(v_{1,\epsilon})\Big] &= \mathbb{E}\Big[\int_0^\infty\int_{-\infty}^\infty v(T,x,y) - v_{0,\epsilon}(T,x,y) - v_{1,\epsilon}(T,x,y)\,\mathrm{d}y\,\mathrm{d}x\Big] \\ &= O(\epsilon).\end{aligned} \tag{5.35}$$

We further conjecture that $v - \sum_{i=0}^{n+1} v_{i,\epsilon}$ converges to zero in the weak sense as follows.

Conjecture 5.3.3. *The $(n+1)$-th order term $v_{n+1,\epsilon}(T,x,y)$ from (5.19) converges to $v - \sum_{i=0}^{n} v_{i,\epsilon}$ of order $(n+1)/2$ in ϵ in the weak sense, such that*

$$\mathbb{E}\Big[P_T(v) - \sum_{i=0}^{n} P_T(v_{i,\epsilon})\Big] = \mathbb{E}\Big[\int_0^{\infty}\int_{-\infty}^{\infty}\Big(v(T,x,y) - \sum_{i=0}^{n} v_{i,\epsilon}(T,x,y)\Big)\mathrm{d}y\,\mathrm{d}x\Big] = O\big(\epsilon^{\frac{n+1}{2}}\big). \quad (5.36)$$

5.4 Summary

We studied the SPDE (5.1) derived from a fast mean-reverting particle system (5.3). We presented an asymptotic expansion to the SPDE, by averaging over the stationary distributions of the volatility process. We conjectured that the expansion converges to the exact solution in a weak sense. We further tested the convergence numerically, and the rates observed are as expected. Finally, we hypothesized the convergence order from the numerical results.

Chapter 6

Conclusions and future work

6.1 Conclusions

In this thesis, we focus on the numerical computation of Zakai type SPDEs, and use Fourier analysis to find a precise description of the error given Dirac initial data. We also apply MLMC, MIMC, and sparse grids to estimate an expected functional of the solution, and achieve the optimal complexity independent of the dimension in this setting. Though the analysis is restricted to the constant coefficient case, the scheme has a wider application. In the end, we also study the asymptotic expansion of a 2-d Zakai SPDE, and provided some conjectures from numerical experiments.

We started from the one dimensional Zakai SPDE in Chapter 2,

$$\mathrm{d}v = -\mu\frac{\partial v}{\partial x}\,\mathrm{d}t + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}\,\mathrm{d}t - \sqrt{\rho}\frac{\partial v}{\partial x}\,\mathrm{d}W_t, \qquad (t,x) \in (0,T)\times\mathbb{R}.$$

Specifically, we analysed a functional $P_T(v)$ of the solution. We proposed a space-time MIMC estimator and compared the complexity with the MLMC method. We showed that, by using the implicit Milstein finite difference discretisation (2.6), the first and second moments of $\Delta P_{(l_1,l_2)}$ are $O(h^2k)$ and $O(h^4k)$, and the total complexity of this scheme is $O\big(\varepsilon^{-2}\left|\log\varepsilon\right|^3\big)$. However, we achieved a better complexity $O\big(\varepsilon^{-2}\left|\log\varepsilon\right|\big)$ if we use a carefully adapted discretisation scheme (2.5). Although MLMC achieved an optimal complexity $O(\varepsilon^{-2})$ in this case, the optimal complexity is generally lost for a

higher dimensional problem, as the decay of the variance can be slower than the increase of the cost. Thus we further study the 2-d SPDE,

$$\begin{aligned} \mathrm{d}v = &-\mu_x \frac{\partial v}{\partial x}\,\mathrm{d}t - \mu_y \frac{\partial v}{\partial y}\,\mathrm{d}t + \frac{1}{2}\left(\frac{\partial^2 v}{\partial x^2} + 2\sqrt{\rho_x\rho_y}\rho_{xy}\frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 v}{\partial y^2}\right)\mathrm{d}t \\ &- \sqrt{\rho_x}\frac{\partial v}{\partial x}\,\mathrm{d}W_t^x - \sqrt{\rho_y}\frac{\partial v}{\partial y}\,\mathrm{d}W_t^y. \end{aligned} \tag{6.1}$$

We gave a rigorous stability and error analysis in L_2 for a Milstein finite difference scheme for the SPDE (6.1), deriving sharp leading order error terms in Chapter 3. We derived pointwise errors for Dirac initial data, which reveal a mild instability for large implicit timesteps and small spatial mesh sizes in this case, not seen in previous studies for L_2 data. Then, we extended the analysis to an alternating direction implicit (ADI) factorisation, which, to our knowledge, is the first application of an ADI scheme to stochastic PDEs. Finally, we proposed a modification for the more general equation (1.3) through sub-simulation of the Lévy area, which is empirically shown to be of first order.

In Chapter 4, we applied the sparse combination method and MLMC method to the SPDE (6.1) on a space-time mesh with an implicit ADI Milstein time-stepping scheme, and used Fourier analysis to derive the first and second moment of the level estimator. We showed that the optimal complexity $O(\varepsilon^{-2})$ can be achieved for a fixed RMSE ε, whereas MLMC on regular grids has $O(\varepsilon^{-2}(\log\varepsilon)^2)$, and standard MC gives $O(\varepsilon^{-4})$. Finally, we gave a discussion in the higher-dimensional setting that sparse combination with MLMC always leads to the optimal complexity, standard MC with sparse combination has a fixed complexity order independent of the dimension (up to a logarithmic term), whereas the cost of MLMC and MC on regular grids increases exponentially with the dimension.

In Chapter 5, we considered the small ϵ asymptotic of the two-dimensional SPDE

$$\begin{aligned} \mathrm{d}v = &\frac{1}{\epsilon}\Big(\nu^2(1-\rho_y^2)\partial_{yy}v + \kappa\partial_y(yv)\Big)\,\mathrm{d}t \\ &+ \Big(\frac{1}{2}\sigma^2\big(y+U_t\big)\partial_{xx}v - \mu\,\partial_x v\Big)\,\mathrm{d}t - \rho_x\sigma\big(y+U_t\big)\partial_x v\,\mathrm{d}W_t^x, \\ v(0,x,y) = &\,\delta(x-x_0)\otimes\delta(y-y_0). \end{aligned}$$

We conjectured an asymptotic expansion of the solution $v(t,x,y)$ to the SPDE (5.1) as $\epsilon \to 0$, and, from numerical experiments, we concluded the convergence of the zero order term $\mathbb{E}[P_T(v) - P_T(v_{0,\epsilon})]$ and of higher order terms.

6.2 Future work

A further research question is to analyse almost sure convergence of equation (6.1), which is of interest for filtering applications and does not follow directly from our analysis. Analysing a Zakai type SPDE with non-constant coefficients is also of interest.

Another open question is a complete analysis of the numerical approximation of initial-boundary value problems (as opposed to problems posed on $\mathbb{R}^d$) for the considered SPDE, when the regularity at the boundary is lost. For example, for the 1-d SPDE with constant coefficients on the half-line,

$$\begin{aligned} \mathrm{d}v &= -\mu\frac{\partial v}{\partial x}\,\mathrm{d}t + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}\,\mathrm{d}t - \sqrt{\rho}\frac{\partial v}{\partial x}\,\mathrm{d}W_t, \qquad (t,x) \in (0,T)\times\mathbb{R}^+, \\ v(t,0) &= 0, \end{aligned}$$

with initial condition $v(0,\cdot) \in H^1$, the second derivative can be unbounded, i.e., $v(t,\cdot) \notin H^2$. This and more general forms have been studied in [74, 14]. In such cases, the assumptions on Galerkin approximations in papers previously mentioned such as in [3, 5] are not established in the literature, hence a new approach for the numerical analysis is to be developed.

One more future project is to mathematically prove the conjectures in Chapter 5. A possible starting point could be to analyse $(v - v_{0,\epsilon})$, which satisfies the inhomogeneous SPDE (5.17). We conjecture that the solution $(v - v_{0,\epsilon})$ has the form

$$\begin{aligned} (v - v_{0,\epsilon})(t,x,y) = &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^t \left(\overline{\sigma^2} - \sigma^2(\eta + U_\tau)\right)\partial_{xx}v_{0,\epsilon}(\xi,\eta,\tau)v(x-\xi,y,t-\tau;\eta)\,\mathrm{d}\tau\,\mathrm{d}\xi\,\mathrm{d}\eta \\ &+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^t \left(\overline{\sigma} - \sigma(\eta + U_\tau)\right)\partial_{xx}v_{0,\epsilon}(\xi,\eta,\tau)v(x-\xi,y,t-\tau;\eta)\,\mathrm{d}W_\tau^X\,\mathrm{d}\xi\,\mathrm{d}\eta. \end{aligned}$$

It may be possible to derive the convergence order by carefully estimating $\mathbb{E}[P_T(v - v_{0,\epsilon})]$ from this exact form.

Appendix A

Proof of Proposition 2.1.1

As for the mean-square convergence, Proposition 2.1.1, we compare the numerical solution to the analytical solution by splitting the domain into two wave number regions. Assume p is a constant satisfying $0 < p < \frac{1}{4}$. Then, we define the low wave number region as in (2.40),

$$\Omega_{\text{low}} = \left\{\gamma : |\gamma| \leq \min\{h^{-2p},\, k^{-p}\}\right\},$$

and the high wave number region as in (2.41),

$$\Omega_{\text{high}} = \left\{\gamma : |\gamma| > \min\{h^{-2p},\, k^{-p}\}\right\} \cap \left[-\frac{\pi}{h}, \frac{\pi}{h}\right].$$

For γ in the low wave region, the analysis is the same as [92], Theorem 2.1, that the Fourier transform of the numerical solution is very close to the theoretical solution, with error $O(h^2) + O(k)$. However, when $h^2 < k$, there is an extra term $O(\theta^N h^{-1})$ in the high wave region. So here we only analyse the contribution from the high wave region.

Proposition 2.1. For the case $h^2 < k$, we divide the high wave region into two parts, $\{k^{-p} < |\gamma| < k^{-1/2}\}$ and $\{k^{-1/2} < |\gamma| < \pi/h\}$, and we integrate separately. Note that

$X(T,\gamma) = \exp(-\frac{1}{2}(1-\rho)\gamma^2 T - \mathrm{i}\gamma\sqrt{\rho}\, W_T)$.

$$
\begin{aligned}
&\mathbb{E}\left[\,\Big|\int_{\Omega_{\text{high}}} X(T,\gamma) - X_N(\gamma)\,\mathrm{d}\gamma\Big|^2\,\right] \\
&\le 4k^{-1/2}\mathbb{E}\left[\int_{k^{-p}}^{k^{-1/2}} \Big|X(T,\gamma) - X_N(\gamma)\Big|^2\,\mathrm{d}\gamma\right] + 4\pi h^{-1}\mathbb{E}\left[\int_{k^{-1/2}}^{\pi/h} \Big|X(T,\gamma) - X_N(\gamma)\Big|^2\,\mathrm{d}\gamma\right] \\
&\le 8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} \mathbb{E}\left[|X_N(\gamma)|^2 + |X(T,\gamma)|^2\right]\,\mathrm{d}\gamma + 8\pi h^{-1}\int_{k^{-1/2}}^{\pi/h} \mathbb{E}\left[|X_N(\gamma)|^2 + |X(T,\gamma)|^2\right]\,\mathrm{d}\gamma \\
&= 8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} \left(\frac{1+\rho\gamma^2 ku\cos^2\frac{h\gamma}{2} + \frac{\rho^2}{2}\gamma^4k^2u^2\cos^4\frac{h\gamma}{2}}{1+\gamma^2ku+\frac{1}{4}\gamma^4k^2u^2}\right)^N \mathrm{d}\gamma \\
&\quad + 8\pi h^{-1}\int_{k^{-1/2}}^{\pi/h} \left(\frac{1+\rho\gamma^2 ku\cos^2\frac{h\gamma}{2} + \frac{\rho^2}{2}\gamma^4k^2u^2\cos^4\frac{h\gamma}{2}}{1+\gamma^2ku+\frac{1}{4}\gamma^4k^2u^2}\right)^N \mathrm{d}\gamma + f_0(k),
\end{aligned}
$$

where $u = \frac{\sin^2\frac{h\gamma}{2}}{(\frac{h\gamma}{2})^2}$, and

$$
\begin{aligned}
f_0(k) &\coloneqq 8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} \exp\big(-(1-\rho)\gamma^2 T\big)\,\mathrm{d}\gamma + 8\pi h^{-1}\int_{k^{-1/2}}^{\pi/h} \exp\big(-(1-\rho)\gamma^2 T\big)\,\mathrm{d}\gamma \\
&\le \frac{4k^{p-1/2}}{(1-\rho)T}\exp\big(-(1-\rho)Tk^{-2p}\big) + \frac{4\pi k^{1/2}}{(1-\rho)T\,h}\exp\big(-(1-\rho)Tk^{-1}\big).
\end{aligned}
\tag{A.1}
$$

Denote $\lambda \coloneqq k/h^2 \in [1,\infty)$, and

$$
f(\gamma) \coloneqq \frac{1+\rho\gamma^2 ku\cos^2\frac{h\gamma}{2} + \frac{\rho^2}{2}\gamma^4k^2u^2\cos^4\frac{h\gamma}{2}}{1+\gamma^2ku+\frac{1}{4}\gamma^4k^2u^2} = \frac{1+\rho\cdot\lambda\sin^2(h\gamma) + \frac{1}{2}\rho^2\lambda^2\sin^4(h\gamma)}{1+4\lambda\sin^2\frac{h\gamma}{2}+4\lambda^2\sin^4\frac{h\gamma}{2}}.
$$

we have

$$
f(\gamma) \le \frac{1+\rho\gamma^2 ku + \frac{\rho^2}{2}\gamma^4k^2u^2}{1+\gamma^2ku+\frac{1}{4}\gamma^4k^2u^2} = \frac{1+\rho\cdot 4\lambda\sin^2\frac{h\gamma}{2} + 8\rho^2\lambda^2\sin^4\frac{h\gamma}{2}}{1+4\lambda\sin^2\frac{h\gamma}{2}+4\lambda^2\sin^4\frac{h\gamma}{2}} \coloneqq g(\gamma).
$$

So

$$
\begin{aligned}
g(\gamma) &= \frac{1+\rho\cdot 4\lambda\sin^2\frac{h\gamma}{2} + 8\rho^2\lambda^2\sin^4\frac{h\gamma}{2}}{1+4\lambda\sin^2\frac{h\gamma}{2}+4\lambda^2\sin^4\frac{h\gamma}{2}} = \frac{1+\rho\cdot 4\lambda d + 8\rho^2\lambda^2 d^2}{1+4\lambda d+4\lambda^2 d^2} \\
g'(\gamma) &= -\frac{4\rho(1-2\rho)\lambda^2 d^2 + 2(1-2\rho^2)\lambda d + (1-\rho)}{(1+4\lambda d+4\lambda^2 d^2)^2}\cdot 2\lambda h\sin(\gamma h),
\end{aligned}
$$

where $d = \sin^2\frac{h\gamma}{2}$. Therefore we have

$$
g'(\gamma^*) = 0 \quad\Longleftrightarrow\quad 4\rho(1-2\rho)\lambda^2 d^2 + 2(1-2\rho^2)\lambda d + (1-\rho) = 0.
$$

The equation has solutions γ_1^*, γ_2^* as follows:

$$\sin^2 \frac{h\gamma_1^*}{2} = -\frac{1}{2\lambda} < 0 \quad \text{(not possible)}, \tag{A.2}$$

$$\sin^2 \frac{h\gamma_2^*}{2} = \frac{1-\rho}{2\rho(2\rho-1)\lambda} \quad \text{(need } \rho > 1/2 \text{ and } \gamma^* < \tfrac{\pi}{h}\text{)}, \tag{A.3}$$

$$\sin^2 \frac{h\gamma^*}{2} = -\frac{1-\rho}{2(1-2\rho^2)\lambda} < 0 \quad \text{(for } \rho = 1/2 \text{ and this is not possible.)} \tag{A.4}$$

So for $\gamma \in \Omega_{\text{high}}$ and $0 < \rho \le 1/2$, $g(\gamma)$ decreases over the high wave region:

$$g'(\gamma) < 0.$$

1. For $k^{-p} < \gamma < k^{-1/2}$,

$$\begin{aligned} g(\gamma) < g(k^{-p}) &= \frac{1 + \rho \cdot 4\lambda \sin^2 \frac{h}{2k^p} + 8\rho^2\lambda^2 \sin^4 \frac{h}{2k^p}}{1 + 4\lambda \sin^2 \frac{h}{2k^p} + 4\lambda^2 \sin^4 \frac{h}{2k^p}} \\ &= \frac{1 + \rho \cdot 4\lambda \frac{h^2}{4k^{2p}} + (8\rho^2\lambda^2 - \frac{4}{3}\rho\lambda)\frac{h^4}{16k^{4p}} + O(\frac{h}{k^{2p}})^6}{1 + 4\lambda \frac{h^2}{4k^{2p}} + (4\lambda^2 - \frac{4}{3}\lambda)\frac{h^4}{16k^{4p}} + O(\frac{h}{k^{2p}})^6} \\ &= 1 - \frac{(1-\rho)k^{1-2p} + \frac{1}{4}(1-2\rho^2)k^{2-4p} + \frac{1}{12}(1-\rho)hk^{1-4p}}{1 + k^{1-2p} + \frac{1}{4}k^{2-4p} - \frac{1}{12}\rho hk^{1-4p}} + o(k^{3(1-4p)}), \end{aligned}$$

$$\begin{aligned} g^N(\gamma) < g^N(k^{-p}) &= \left(1 - \frac{(1-\rho)k^{1-2p} + \frac{1}{4}(1-2\rho^2)k^{2-4p} + \frac{1}{12}(1-\rho)hk^{1-4p}}{1 + k^{1-2p} + \frac{1}{4}k^{2-4p} - \frac{1}{12}\rho hk^{1-4p}} + o(k^{3(1-4p)})\right)^N \\ &< \exp\left(-(1-\rho)T\Big(k^{-2p} + \frac{1}{4}(1-2\rho^2)k^{-4p} + \frac{1}{12}(1-\rho)hk^{-4p}\Big) + o(k^{-4p})\right). \end{aligned}$$

Therefore,

$$\begin{aligned} &8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} g^N(\gamma)\,\mathrm{d}\gamma \\ &< 8k^{-1}\exp\left(-(1-\rho)T\Big(k^{-2p} + \frac{1-2\rho^2}{4}k^{-4p} + \frac{1-\rho}{12}hk^{-4p}\Big) + o(k^{-4p})\right). \end{aligned}$$

This is $o(k^r)$ for any $r > 0$ as $k \to 0$, i.e.,

$$\lim_{k\to 0} \frac{1}{k^r} \cdot 8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} g^N(\gamma)\,\mathrm{d}\gamma = 0, \qquad \text{for all } r > 0.$$

2. For $k^{-1/2} < \gamma < \pi/h$,

$$g(\gamma) < g(k^{-1/2}) = \frac{1+\rho\tilde{u}+\frac{1}{2}\rho^2\tilde{u}^2}{1+\tilde{u}+\frac{1}{4}\tilde{u}^2} \coloneqq q(\rho,\tilde{u}), \qquad \text{where } \tilde{u} = \left(\sin\left(\tfrac{h}{2\sqrt{k}}\right)\Big/\tfrac{h}{2\sqrt{k}}\right)^2 \in [4/\pi^2, 1].$$

The function $q(\rho,\tilde{u})$ is strictly decreasing in $\tilde{u}$ and increasing in ρ for $4/\pi^2 \le \tilde{u} \le 1,\ 0 < \rho \le 1/2$. So we have,

$$g(k^{-1/2}) < q\Big(\frac{1}{2}, \frac{4}{\pi^2}\Big) = \frac{1+\frac{2}{\pi^2}+\frac{2}{\pi^4}}{1+\frac{4}{\pi^2}+\frac{4}{\pi^4}} \coloneqq \theta_0 < 0.8457.$$

Therefore,

$$8\pi h^{-1}\int_{k^{-1/2}}^{\pi/h} g^N(\gamma)\,\mathrm{d}\gamma < 8\pi^2 h^{-2}\cdot\theta_0^N.$$

Consequently, for $0 < \rho \le 1/2$,

$$\begin{aligned}
&\mathbb{E}\left[\,\Big|\int_{\Omega_{\text{high}}} X(T,\gamma) - X_N(\gamma)\,\mathrm{d}\gamma\Big|^2\,\right] \\
&\le 8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} f^N(\gamma)\,\mathrm{d}\gamma + 8\pi h^{-1}\int_{k^{-1/2}}^{\pi/h} f^N(\gamma)\,\mathrm{d}\gamma + f_0(k) \\
&\le 8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} g^N(\gamma)\,\mathrm{d}\gamma + 8\pi h^{-1}\int_{k^{-1/2}}^{\pi/h} g^N(\gamma)\,\mathrm{d}\gamma + f_0(k) \\
&\le C\cdot\theta_0^N\cdot h^{-2},
\end{aligned}$$

where C and θ_0 are constants independent of h and k.

For $\gamma \in \Omega_{\text{high}}$ and $1/2 < \rho < 1/\sqrt{2}$, $g(\gamma)$ is decreasing before γ^*, and increasing afterwards, where γ^* solves the equation (A.3) such that

$$\gamma^{*2}k = \frac{2(1-\rho)}{\rho(2\rho-1)}\cdot\left(\frac{\sin\frac{h\gamma^*}{2}}{\frac{h\gamma^*}{2}}\right)^{-2} \ge \frac{2(1-\rho)}{\rho(2\rho-1)}.$$

So

$$\gamma^* \ge C_0\cdot k^{-1/2}, \qquad \text{where } C_0 = \tfrac{2(1-\rho)}{\rho(2\rho-1)} \ge 2.$$

We analyse separately the two parts, $\{k^{-p} < |\gamma| < k^{-1/2}\}$ and $\{k^{-1/2} < |\gamma| < \pi/h\}$.

1. When $k^{-p} < |\gamma| < k^{-1/2}$, $g(\gamma)$ is strictly decreasing. Similar to the previous analysis, we have

$$8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} g^N(\gamma)\,\mathrm{d}\gamma < 8k^{-1}\exp\left(-(1-\rho)T\Big(k^{-2p}+\frac{1}{4}(1-2\rho^2)k^{-4p}+\frac{1}{12}(1-\rho)hk^{-4p}\Big)\right).$$

   This is $o(k^r)$ for any $r > 0$ as $k \to 0$.

2. For $k^{-1/2} < |\gamma| < \pi/h$, $g(\gamma)$ decreases first, then increases if $\gamma^* < \pi/h$. Therefore,

$$g(\gamma) < \max\{\, g(k^{-1/2}),\, g(\pi/h)\}.$$

   We will show both terms are strictly less than 1.

   Similar to the previous analysis,

$$g(k^{-1/2}) = \frac{1+\rho\tilde{u}+\frac{1}{2}\rho^2\tilde{u}^2}{1+\tilde{u}+\frac{1}{4}\tilde{u}^2} = q(\rho,\tilde{u}) < q\left(\frac{1}{\sqrt{2}},\frac{4}{\pi^2}\right) \coloneqq \theta_1 < 0.918.$$

   On the other hand,

$$g(\pi/h) = \frac{1+4\rho\lambda+8\rho^2\lambda^2}{1+4\lambda+4\lambda^2},$$

   and this is strictly decreasing in λ and increasing in ρ, so

$$g(\pi/h) \le \frac{1+4\rho\lambda+8\rho^2\lambda^2}{1+4\lambda+4\lambda^2}\bigg|_{\lambda=1,\,\rho=1/\sqrt{2}} < 0.8699 \coloneqq \theta_2.$$

   Therefore, if we let $\theta_0 = \max\{\theta_1,\,\theta_2\} = \theta_1$, we have

$$8\pi h^{-1}\int_{k^{-1/2}}^{\pi/h} g^N(\gamma)\,\mathrm{d}\gamma < 8\pi^2 h^{-2}\cdot\theta_0^N.$$

Consequently, for $0 < \rho \leq 1/2$,

$$\begin{aligned}
&\mathbb{E}\left[\left|\int_{\Omega_{\text{high}}} X(T,\gamma) - X_N(\gamma)\,\mathrm{d}\gamma\right|^2\right] \\
&\leq 8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} f^N(\gamma)\,\mathrm{d}\gamma + 8\pi h^{-1}\int_{k^{-1/2}}^{\pi/h} f^N(\gamma)\,\mathrm{d}\gamma + f_0(k) \\
&\leq 8k^{-1/2}\int_{k^{-p}}^{k^{-1/2}} g^N(\gamma)\,\mathrm{d}\gamma + 8\pi h^{-1}\int_{k^{-1/2}}^{\pi/h} g^N(\gamma)\,\mathrm{d}\gamma + f_0(k) \\
&\leq C\cdot\theta_0^N\cdot h^{-2},
\end{aligned}$$

where C and θ_0 are constants independent of h and k.

Numerical results confirm this as well. Figure A.1 shows the integral over the high wave region with $T = 1$, $k = 1/4$ fixed, and $h = 2^{-l}$, $l = 7, 8, \ldots, 20$.

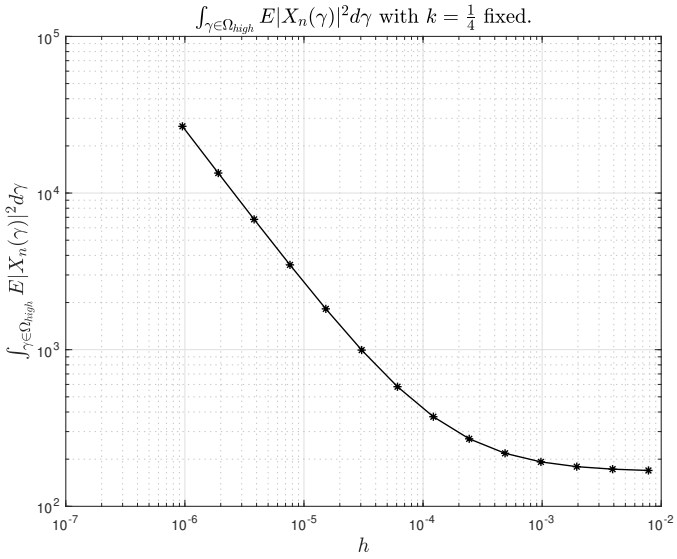


Figure A.1: Integral over the high wave region with k fixed.

□

Appendix B

Particle system and the limiting SPDE

We start from the particle system

$$\begin{aligned}
\mathrm{d}X_t^i &= \mu\,\mathrm{d}t + \sigma(Y_t^i + U_t)\Big(\rho_x\,\mathrm{d}W_t^x + \sqrt{1-\rho_x^2}\,\mathrm{d}W_t^{x,i}\Big), \quad & X_0^i &= x_0, \quad 1 \le i \le N, \\
\mathrm{d}Y_t^i &= -\frac{\kappa}{\epsilon}Y_t^i\,\mathrm{d}t + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\sqrt{1-\rho_y^2}\,\mathrm{d}W_t^{y,i}, & Y_0^i &= y_0, \quad 1 \le i \le N, \\
\mathrm{d}U_t &= -\frac{\kappa}{\epsilon}U_t\,\mathrm{d}t + \frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\rho_y\,\mathrm{d}W_t^y, & U_0 &= 0,
\end{aligned} \tag{B.1}$$

where N represents the number of particles, $W^{x,i}$ and $W^{y,i}$ are standard Brownian motion, independent from each other, and W^x, W^y have correlation ρ_{xy}.

We consider the two dimensional empirical measure for (X_t^i, Y_t^i),

$$\nu_{N,t} = \frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i, Y_t^i}.$$

It has been proved in [77, 55] for a similar system that there exists a limit empirical measure

$$\nu_t = \lim_{N\to\infty}\nu_{N,t},$$

and there exists density v for the limit empirical measure, where

$$\nu_t(\mathrm{d}x, \mathrm{d}y) = v(t,x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Moreover, for any Borel set A,

$$\nu_t(A) = \mathbb{P}\big((X_t^1, Y_t^1) \in A \mid \mathcal{F}_t^{x,y}\big),$$

where $\mathcal{F}^{x,y}$ is the filtration generated by W^x and W^y.

Here we adapt the proof stated in [77, 55], assume there exists a unique limit empirical measure, and present a heuristic derivation to the density of the limit empirical measure.

Let C_{test} be the class of infinitely differentiable functions with compact support on $\mathbb{R}^2$. For any $f \in C_{\text{test}}$, we write

$$\langle f, \nu_t \rangle = \iint_{\mathbb{R}^2} f(x,y) v(t,x,y)\,\mathrm{d}x\,\mathrm{d}y = \mathbb{E}\big[f(X_t^1, Y_t^1) \mid \mathcal{F}_t^{x,y}\big]. \tag{B.2}$$

We apply Itô's formula, and get

$$\begin{aligned}
\mathrm{d}f(X_t^1, Y_t^1) = \Big(&\mu\,\partial_x f(X_t^1, Y_t^1) + \frac{1}{2}\sigma^2\big(Y_t^1 + U_t\big)\partial_{xx} f(X_t^1, Y_t^1) \\
&-\frac{\kappa}{\epsilon} Y_t^1 \partial_y f(X_t^1, Y_t^1) + \frac{\nu^2}{\epsilon}\big(1-\rho_y^2\big)\partial_{yy} f(X_t^1, Y_t^1)\Big)\,\mathrm{d}t \\
&+\rho_x \sigma\big(Y_t^1 + U_t\big)\partial_x f(X_t^1, Y_t^1)\,\mathrm{d}W_t^x + \sqrt{1-\rho_x^2}\,\sigma\big(Y_t^1 + U_t\big)\partial_x f(X_t^1, Y_t^1)\,\mathrm{d}W_t^{x,1} \\
&+\frac{\nu\sqrt{2}}{\sqrt{\epsilon}}\sqrt{1-\rho_y^2}\,\partial_y f(X_t^1, Y_t^1)\,\mathrm{d}W_t^{y,1}.
\end{aligned} \tag{B.3}$$

From Lemma 5.2 in [1], for any real-valued adapted process H with $\mathbb{E}\Big[\int_0^T H_s^2\,\mathrm{d}s\Big] < \infty$, then the following equations hold with probability 1,

$$\begin{aligned}
\mathbb{E}\bigg[\int_0^t H_s\,\mathrm{d}W_s^x \;\bigg|\; \mathcal{F}_t^{x,y}\bigg] &= \int_0^t \mathbb{E}\big[H_s \mid \mathcal{F}_s^{x,y}\big]\,\mathrm{d}W_s^x, \\
\mathbb{E}\bigg[\int_0^t H_s\,\mathrm{d}W_s^{x,1} \;\bigg|\; \mathcal{F}_t^{x,y}\bigg] &= \mathbb{E}\bigg[\int_0^t H_s\,\mathrm{d}W_s^{y,1} \;\bigg|\; \mathcal{F}_t^{x,y}\bigg] = 0.
\end{aligned}$$

Under an L^2-integrability condition on σ, which holds in the case of Ornstein-Uhlenbeck (OU) or an exponential OU stochastic volatility model, taking the conditional expectation of (B.3) with respect to $\mathcal{F}^{x,y}$ gives

$$\begin{aligned}
\langle f, \nu_t \rangle = \langle f, \nu_0 \rangle &+ \int_0^t \langle \mu\,\partial_x f, \nu_s \rangle\,\mathrm{d}s + \frac{1}{2}\int_0^t \langle \sigma^2 \partial_{xx} f, \nu_s \rangle\,\mathrm{d}s - \frac{\kappa}{\epsilon}\int_0^t \langle y \partial_y f, \nu_s \rangle\,\mathrm{d}s \\
&+ \frac{\nu^2}{\epsilon}(1-\rho_y^2)\int_0^t \langle \partial_{yy} f, \nu_s \rangle\,\mathrm{d}s + \rho_x \int_0^t \langle \sigma \partial_x f, \nu_s \rangle\,\mathrm{d}W_s^x
\end{aligned}$$

From (B.2), we can rewrite this and apply integration by parts,

$$
\begin{aligned}
&\iint_{\mathbb{R}^2} f(x,y)v(t,x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \iint_{\mathbb{R}^2} f(x,y)v(0,x,y)\,\mathrm{d}x\,\mathrm{d}y + \int_0^t \iint_{\mathbb{R}^2} \mu\,\partial_x f(x,y)v(s,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \\
&\quad + \frac{1}{2}\int_0^t \iint_{\mathbb{R}^2} \sigma^2(y+U_s)\partial_{xx} f(x,y)v(s,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \\
&\quad - \frac{\kappa}{\epsilon}\int_0^t \iint_{\mathbb{R}^2} y\partial_y f(x,y)v(s,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \\
&\quad + \frac{\nu^2}{\epsilon}(1-\rho_y^2)\int_0^t \iint_{\mathbb{R}^2} \partial_{yy} f(x,y)v(s,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \\
&\quad + \rho_x \int_0^t \iint_{\mathbb{R}^2} \sigma(y+U_s)\partial_x f(x,y)v(s,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}W_s^x
\end{aligned}
$$

$$
\begin{aligned}
&= \iint_{\mathbb{R}^2} f(x,y)v(0,x,y)\,\mathrm{d}x\,\mathrm{d}y - \int_0^t \iint_{\mathbb{R}^2} f(x,y)\mu\,\partial_x v(s,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \\
&\quad + \frac{1}{2}\int_0^t \iint_{\mathbb{R}^2} f(x,y)\sigma^2(y+U_s)\partial_{xx} v(s,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \\
&\quad + \frac{\kappa}{\epsilon}\int_0^t \iint_{\mathbb{R}^2} f(x,y)\partial_y\big(yv(s,x,y)\big)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \\
&\quad + \frac{\nu^2}{\epsilon}(1-\rho_y^2)\int_0^t \iint_{\mathbb{R}^2} f(x,y)\partial_{yy} v(s,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \\
&\quad - \rho_x \int_0^t \iint_{\mathbb{R}^2} f(x,y)\sigma(y+U_s)\partial_x v(s,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}W_s^x
\end{aligned}
$$

Since this satisfies for all test function $f \in C_{\text{test}}$, it yields to an SPDE for the density

$$
\begin{aligned}
\mathrm{d}v = \frac{1}{\epsilon}\Big(\nu^2(1-\rho_y^2)\partial_{yy} v + \kappa\partial_y(yv)\Big)\,\mathrm{d}t \\
+ \Big(\frac{1}{2}\sigma^2\big(y+U_t\big)\partial_{xx} v - \mu\,\partial_x v\Big)\,\mathrm{d}t - \rho_x\sigma\big(y+U_t\big)\partial_x v\,\mathrm{d}W_t^x.
\end{aligned}
$$

Bibliography

[1] Ahmad, F., Hambly, B. M., and Ledger, S. A stochastic partial differential equation model for the pricing of mortgage-backed securities. *Stochastic Processes and their Applications 128*, 11 (2018), 3778–3806.

[2] Bain, A., and Crisan, D. *Fundamentals of Stochastic Filtering*, vol. 60 of *Stochastic Modelling and Applied Probability*. Springer, New York, NY, 2009.

[3] Barth, A., and Lang, A. Milstein approximation for advection-diffusion equations driven by multiplicative noncontinuous martingale noises. *Applied Mathematics & Optimization 66*, 3 (2012), 387–413.

[4] Barth, A., and Lang, A. L^p and almost sure convergence of a Milstein scheme for stochastic partial differential equations. *Stochastic Processes and their Applications 123*, 5 (2013), 1563–1587.

[5] Barth, A., Lang, A., and Schwab, C. Multilevel Monte Carlo method for parabolic stochastic partial differential equations. *BIT Numerical Mathematics 53*, 1 (2013), 3–27.

[6] Bensoussan, A., and Glowinski, R. Approximation of Zakai equation by the splitting up method. In *Stochastic Systems and Optimization*. Springer Berlin Heidelberg, 1989, pp. 255–265.

[7] Brandt, A., McCoruick, S., and Huge, J. Algebraic multigrid (amg) for sparse matrix equati0ns. *Sparsity and its Applications* (1984), 257–284.

[8] Buckwar, E., and Kelly, C. Towards a systematic linear stability analysis of numerical methods for systems of stochastic differential equations. *SIAM Journal on Numerical Analysis 48*, 1 (2010), 298–321.

[9] Buckwar, E., and Sickenberger, T. A comparative linear mean-square stability analysis of Maruyama- and Milstein-type methods. *Mathematics and Computers in Simulation 81*, 6 (2011), 1110–1127.

[10] Bujok, K., and Reisinger, C. Numerical valuation of basket credit derivatives in structural jump-diffusion models. *The Journal of Computational Finance 15*, 4 (2012), 115–158.

[11] Bungartz, H.-J. *Finite elements of higher order on sparse grids*. PhD thesis, Technische Universität München, 1998.

[12] Bungartz, H.-J., and Griebel, M. Sparse grids. *Acta Numerica 13* (2004), 147–269.

[13] Burgos, S., and Giles, M. B. Computing Greeks using multilevel path simulation. In *Monte Carlo and Quasi-Monte Carlo Methods 2010* (2012), Springer Berlin Heidelberg, pp. 281–296.

[14] Bush, N., Hambly, B. M., Haworth, H., Jin, L., and Reisinger, C. Stochastic evolution equations in portfolio credit modelling. *SIAM Journal on Financial Mathematics 2*, 1 (2011), 627–664.

[15] Carter, R., and Giles, M. B. Sharp error estimates for discretizations of the 1D convection–diffusion equation with Dirac initial data. *IMA Journal of Numerical Analysis 27*, 2 (2007), 406–425.

[16] Charrier, J., Scheichl, R., and Teckentrup, A. L. Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods. *SIAM Journal on Numerical Analysis 51*, 1 (2013), 322–352.

[17] Clark, J. M. C., and Cameron, R. J. The maximum rate of convergence of discrete approximations for stochastic differential equations. In *Stochastic Differential Systems Filtering and Control* (1980), B. Grigelionis, Ed., Springer Berlin Heidelberg, pp. 162–171.

[18] Cliffe, K. A., Giles, M. B., Scheichl, R., and Teckentrup, A. L. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science 14*, 1 (2011), 3–15.

[19] Craig, I. J. D., and Sneyd, A. D. An alternating-direction implicit scheme for parabolic equations with mixed derivatives. *Computers & Mathematics with Applications 16*, 4 (1988), 341–350.

[20] Crisan, D. Exact rates of convergeance for a branching particle approximation to the solution of the Zakai equation. *The Annals of Probability 31*, 2 (2003), 693–718.

[21] Crisan, D., Gaines, J., and Lyons, T. Convergence of a branching particle method to the solution of the Zakai equation. *SIAM Journal on Applied Mathematics 58*, 5 (1998), 1568–1590.

[22] Crisan, D., and Lyons, T. Nonlinear filtering and measure-valued processes. *Probability Theory and Related Fields 109*, 2 (1997), 217–244.

[23] Crisan, D., and Lyons, T. A particle approximation of the solution of the Kushner–Stratonovitch equation. *Probability Theory and Related Fields 115*, 4 (1999), 549–578.

[24] Crisan, D., and Xiong, J. Approximate McKean–Vlasov representations for a class of SPDEs. *Stochastics: An International Journal of Probability and Stochastics Processes 82*, 1 (2010), 53–68.

[25] Croci, M., Giles, M. B., Rognes, M. E., and Farrell, P. E. Efficient white noise sampling and coupling for multilevel Monte Carlo with nonnested meshes. *SIAM/ASA Journal on Uncertainty Quantification 6*, 4 (2018), 1630–1655.

[26] Da Prato, G., and Zabczyk, J. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, 2014.

[27] Davie, A. M., and Gaines, J. G. Convergence of numerical schemes for the solution of parabolic stochastic partial differential equations. *Mathematics of Computation 70*, 233 (2001), 121–134.

[28] Fang, W., and Giles, M. B. Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift: Part I, finite time interval. *arXiv preprint arXiv:1609.08101* (2016).

[29] Florchinger, P., and Le Gland, F. Time-discretization of the Zakai equation for diffusion processes observed in correlated noise. *Stochastics and Stochastic Reports 35*, 4 (1991), 233–256.

[30] Florchinger, P., and Le Gland, F. Particle approximation for first order stochastic partial differential equations. In *Applied Stochastic Analysis* (1992), Springer Berlin Heidelberg, pp. 121–133.

[31] Forde, M., Jacquier, A., and Lee, R. The small-time smile and term structure of implied volatility under the Heston model. *SIAM Journal on Financial Mathematics 3*, 1 (2012), 690–708.

[32] Fouque, J.-P., Papanicolaou, G., and Sircar, K. R. *Derivatives in Financial Markets With Stochastic Volatility*. Cambridge University Press, 2000.

[33] Fouque, J.-P., Papanicolaou, G., Sircar, R., and Solna, K. Multiscale stochastic volatility asymptotics. *Multiscale Modeling & Simulation 2*, 1 (2003), 22–42.

[34] Fouque, J.-P., Papanicolaou, G., Sircar, R., and Solna, K. Singular perturbations in option pricing. *SIAM Journal on Applied Mathematics 63*, 5 (2003), 1648–1665.

[35] Fouque, J.-P., Papanicolaou, G., Sircar, R., and Solna, K. *Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivatives*. Cambridge University Press, 2011.

[36] Funaki, T. Random motion of strings and related stochastic evolution equations. *Nagoya Mathematical Journal 89* (1983), 129–193.

[37] Gaines, J. G., and Lyons, T. J. Random generation of stochastic area integrals. *SIAM Journal on Applied Mathematics 54*, 4 (1994), 1132–1146.

[38] Giles, M. B. Multilevel Monte Carlo path simulation. *Operations Research 56*, 3 (2008), 607–617.

[39] Giles, M. B. Multilevel Monte Carlo methods. *Acta Numerica 24* (2015), 259–328.

[40] Giles, M. B., Debrabant, K., and Rössler, A. Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation. *Discrete and Continuous Dynamical Systems - Series B* (2019).

[41] Giles, M. B., Kuo, F. Y., and Sloan, I. H. Combining sparse grids, multilevel MC and QMC for elliptic PDEs with random coefficients. In *International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (2018), Springer International Publishing, pp. 265–281.

[42] Giles, M. B., and Reisinger, C. Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance. *SIAM Journal on Financial Mathematics 3*, 1 (2012), 572–592.

[43] Giles, M. B., and Szpruch, L. Antithetic multilevel Monte Carlo estimation for multidimensional SDEs. In *Monte Carlo and Quasi-Monte Carlo Methods 2012* (2013), Springer Berlin Heidelberg, pp. 367–384.

[44] Giles, M. B., and Szpruch, L. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *The Annals of Applied Probability 24*, 4 (2014), 1585–1620.

[45] Gobet, E., Pages, G., Pham, H., and Printems, J. Discretization and simulation of the Zakai equation. *SIAM Journal on Numerical Analysis 44*, 6 (2006), 2505–2538.

[46] Gradinaru, M., Nourdin, I., and Tindel, S. Itô's- and Tanaka's-type formulae for the stochastic heat equation: The linear case. *Journal of Functional Analysis 228*, 1 (2005), 114–143.

[47] Griebel, M., and Harbrecht, H. On the convergence of the combination technique. In *Sparse Grids and Applications – Munich 2012*. Springer International Publishing, 2014, pp. 55–74.

[48] Griebel, M., Schneider, M., and Zenger, C. A combination technique for the solution of sparse grid problems. In *Iterative Methods in Linear Algebra: Proceedings of the IMACS International Symposium on Iterative Methods in Linear Algebra*. North-Holland, 1992, pp. 263–281.

[49] Gyöngy, I. Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise I. *Potential Analysis 9*, 1 (1998), 1–25.

[50] Gyöngy, I. Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise II. *Potential Analysis 11*, 1 (1999), 1–37.

[51] Gyöngy, I., and Krylov, N. On the rate of convergence of splitting-up approximations for SPDEs. In *Stochastic Inequalities and Applications* (2003), Birkhäuser Basel, pp. 301–321.

[52] Gyöngy, I., and Krylov, N. On the splitting-up method and stochastic partial differential equations. *The Annals of Probability 31*, 2 (2003), 564–591.

[53] Gyöngy, I., and Nualart, D. Implicit scheme for stochastic parabolic partial differential equations driven by space-time white noise. *Potential Analysis 7*, 4 (1997), 725–757.

[54] Haji-Ali, A.-L., Nobile, F., and Tempone, R. Multi-index Monte Carlo: when sparsity meets sampling. *Numerische Mathematik 132*, 4 (2016), 767–806.

[55] Hambly, B., and Kolliopoulos, N. Stochastic evolution equations for large portfolios of stochastic volatility models. *SIAM Journal on Financial Mathematics 8*, 1 (2017), 962–1014.

[56] Hendricks, C., Ehrhardt, M., and Günther, M. High-order ADI schemes for diffusion equations with mixed derivatives in the combination technique. *Applied Numerical Mathematics 101* (2016), 36–52.

[57] Higham, D. J. Mean-square and asymptotic stability of the stochastic theta method. *SIAM Journal on Numerical Analysis 38*, 3 (2000), 753–769.

[58] Higham, D. J., Mao, X., Roj, M., Song, Q., and Yin, G. Mean exit times and the multilevel Monte Carlo method. *SIAM/ASA Journal on Uncertainty Quantification 1*, 1 (2013), 2–18.

[59] Higham, D. J., and Trefethen, L. N. Stiffness of ODEs. *BIT Numerical Mathematics 33*, 2 (1993), 285–303.

[60] Hinch, E. J. *Perturbation Methods*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 1991.

[61] Hoel, H., Von Schwerin, E., Szepessy, A., and Tempone, R. Adaptive multilevel Monte Carlo simulation. In *Numerical Analysis of Multiscale Computations* (2012), Springer Berlin Heidelberg, pp. 217–234.

[62] Hoel, H., Von Schwerin, E., Szepessy, A., and Tempone, R. Implementation and analysis of an adaptive multilevel Monte Carlo algorithm. *Monte Carlo Methods and Applications 20*, 1 (2013), 1–41.

[63] Hundsdorfer, W., and Verwer, J. G. *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, vol. 33 of *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 2003.

[64] Ito, K. Approximation of stochastic evolution equations. In *Proceedings of 1995 34th IEEE Conference on Decision and Control* (1995), vol. 4, IEEE, pp. 4035–4040.

[65] Ito, K. Approximation of the Zakai equation for nonlinear filtering. *SIAM Journal on Control and Optimization 34*, 2 (1996), 620–634.

[66] Jentzen, A., and Kloeden, P. E. The numerical approximation of stochastic partial differential equations. *Milan Journal of Mathematics 77*, 1 (2009), 205–244.

[67] Jentzen, A., and Kloeden, P. E. Taylor expansions of solutions of stochastic partial differential equations with additive noise. *The Annals of Probability 38*, 2 (2010), 532–569.

[68] Jentzen, A., Kloeden, P. E., and Winkel, G. Efficient simulation of nonlinear parabolic SPDEs with additive noise. *The Annals of Applied Probability 21*, 3 (2011), 908–950.

[69] Jentzen, A., and Röckner, M. A Milstein scheme for SPDEs. *Foundations of Computational Mathematics 15*, 2 (2015), 313–362.

[70] Kloeden, P., and Platen, E. *Numerical Solution of Stochastic Differential Equations*, vol. 23 of *Applications of Mathematic*. Springer, Berlin, Heidelberg, 1992.

[71] Kloeden, P. E., and Platen, E. Higher-order implicit strong numerical schemes for stochastic differential equations. *Journal of Statistical Physics 66*, 1 (1992), 283–314.

[72] Kloeden, P. E., Platen, E., and Wright, I. W. The approximation of multiple stochastic integrals. *Stochastic Analysis and Applications 10*, 4 (1992), 431–441.

[73] Kruse, R. Optimal error estimates of Galerkin finite element methods for stochastic partial differential equations with multiplicative noise. *IMA Journal of Numerical Analysis 34*, 1 (2014), 217–251.

[74] Krylov, N. V., and Rozovskii, B. L. Stochastic evolution equations. *Journal of Soviet Mathematics 16*, 4 (1981), 1233–1277.

[75] Kuo, F. Y., Schwab, C., and Sloan, I. H. Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients. *SIAM Journal on Numerical Analysis 50*, 6 (2012), 3351–3374.

[76] Kuo, F. Y., Schwab, C., and Sloan, I. H. Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients. *Foundations of Computational Mathematics 15*, 2 (2015), 411–449.

[77] Kurtz, T. G., and Xiong, J. Particle representations for a class of nonlinear SPDEs. *Stochastic Processes and Their Applications 83*, 1 (1999), 103–126.

[78] Lang, A. A Lax equivalence theorem for stochastic differential equations. *Journal of Computational and Applied Mathematics 234*, 12 (2010), 3387–3396.

[79] Lang, A. Mean square convergence of a semidiscrete scheme for SPDEs of Zakai type driven by square integrable martingales. vol. 1, Elsevier, pp. 1615–1623.

[80] Lang, A. Almost sure convergence of a Galerkin approximation for SPDEs of Zakai type driven by square integrable martingales. *Journal of Computational and Applied Mathematics 236*, 7 (2012), 1724–1732.

[81] Lang, A., Petersson, A., and Thalhammer, A. Mean-square stability analysis of approximations of stochastic differential equations in infinite dimensions. *BIT Numerical Mathematics 57*, 4 (2017), 963–990.

[82] Ledger, S. Sharp regularity near an absorbing boundary for solutions to second order SPDEs in a half-line with constant coefficients. *Stochastic Partial Differential Equations: Analysis and Computations 2*, 1 (2014), 1–26.

[83] Malham, S. J. A., and Wiese, A. Efficient almost-exact Lévy area sampling. *Statistics & Probability Letters 88* (2014), 50–55.

[84] Manthey, R. On the Cauchy problem for reaction-diffusion equations with white noise. *Mathematische Nachrichten 136*, 1 (1988), 209–228.

[85] Millet, A., and Morien, P.-L. On implicit and explicit discretization schemes for parabolic SPDEs in any dimension. *Stochastic Processes and their Applications 115*, 7 (2005), 1073–1106.

[86] Milshtein, G. N. Approximate integration of stochastic differential equations. *Theory of Probability and its Applications 19*, 3 (1975), 557–562.

[87] Müller-Gronbach, T. *Strong approximation of systems of stochastic differential equations*. PhD thesis, Darmstadt University of Technology, 2002.

[88] Peaceman, D. W., and Rachford, Jr, H. H. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics 3*, 1 (1955), 28–41.

[89] Pflaum, C., and Zhou, A. Error analysis of the combination technique. *Numerische Mathematik 84*, 2 (1999), 327–350.

[90] Raviart, P.-A. An analysis of particle methods. In *Numerical Methods in Fluid Dynamics*, F. Brezzi, Ed. Springer Berlin Heidelberg, Berlin, Heidelberg, 1985, pp. 243–324.

[91] Reisinger, C. Analysis of linear difference schemes in the sparse grid combination technique. *IMA Journal of Numerical Analysis 33*, 2 (2012), 544–581.

[92] Reisinger, C. Mean-square stability and error analysis of implicit time-stepping schemes for linear parabolic SPDEs with multiplicative Wiener noise in the first derivative. *International Journal of Computer Mathematics 89*, 18 (2012), 2562–2575.

[93] Reisinger, C., and Wang, Z. Stability and error analysis of an implicit Milstein finite difference scheme for a two-dimensional Zakai SPDE. *To appear in BIT Numerical Mathematics*.

[94] Reisinger, C., and Wang, Z. Analysis of multi-index Monte Carlo estimators for a Zakai SPDE. *Journal of Computational Mathematics 36*, 2 (2018), 202–236.

[95] Ruge, J. W., and Stüben, K. Algebraic multigrid. In *Multigrid Methods*. SIAM, 1987, pp. 73–130.

[96] Szpruch, L. *Numerical approximations of nonlinear stochastic systems*. PhD thesis, University of Strathclyde, 2010.

[97] Teckentrup, A. L., Scheichl, R., Giles, M. B., and Ullmann, E. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numerische Mathematik 125*, 3 (2013), 569–600.

[98] Walsh, J. B. An introduction to stochastic partial differential equations. In *École d'Été de Probabilités de Saint Flour XIV-1984* (1986), Springer Berlin Heidelberg, pp. 265–439.

[99] Walsh, J. B. Finite element methods for parabolic stochastic PDE's. *Potential Analysis 23*, 1 (2005), 1–43.

[100] Wiktorsson, M. Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *The Annals of Applied Probability 11*, 2 (2001), 470–487.

[101] Wyns, M., and in 't Hout, K. J. Convergence of the Modified Craig–Sneyd scheme for two-dimensional convection–diffusion equations with mixed derivative term. *Journal of Computational and Applied Mathematics 296* (2016), 170–180.

[102] Zenger, C. Sparse grids. In *Parallel Algorithms for Partial Differential Equations, Proceedings of the Sixth GAMM-Seminar (W. Hackbusch ed.). Notes on Numerical Fluid Dynamics 31. Braunschweig: Vieweg.* (1991).