Adaptive Milstein Methods for Stochastic Differential Equations

Fandi Sun

Submitted for the degree of
Doctor of Philosophy

Department of Mathematics
School of Mathematical and Computer Sciences
Heriot-Watt University

February 2023

Abstract

It was shown in [27] that the Euler-Maruyama (EM) method fails to converge with equidistant timesteps in the strong sense to the solutions of stochastic differential equations (SDEs) when either of the drift or diffusion coefficients is not globally Lipschitz continuous. Higher-order methods or schemes that are developed based on EM, e.g. Milstein method or EM with jumps, inherit the problem.

We introduce an explicit adaptive Milstein method for SDEs with no commutativity condition. The drift and diffusion are separately locally Lipschitz and together satisfy a monotone condition. This method relies on a class of path-bounded timestepping strategies which work by reducing the stepsize as solutions approach the boundary of a sphere, invoking a backstop method in the event that the timestep becomes too small. We prove that such schemes are strongly L_2 convergent of order one. This order is inherited by an explicit adaptive EM scheme in the additive noise case. Moreover, we show that the probability of using the backstop method at any step can be made arbitrarily small. We compare our method to other fixed-step Milstein variants on a range of test problems.

Secondly, we introduce a jump-adapted adaptive Milstein (JAAM) method for SDEs driven by Poisson random measure. With the conditions of drift and diffusion coefficients remaining the same as for the adaptive Milstein method, and the jump coefficient is globally Lipschitz continuous. The corresponding time-stepping strategies that we propose are hence path-bounded and also jump-adapted. We prove the L_2 strong convergence of order one for JAAM and compare its computational efficiency with jump-adapted and fixed-step methods on test models.

In memory of my beloved mother

戴 燕 冰

DAI YAN BING

For her sacrifice, protection, optimism, enthusiasm, joy, resilience, wisdom, and forever love...

Acknowledgements

I would like to express my deepest gratitude to my two excellent supervisors: Professor Gabriel J. Lord and Dr Cónall Kelly, for their knowledgeable, inspirational, encouraging, supportive, and patient guidance. They satiated my aspiration of witnessing the process of mathematics acting as the absolute bedrock on which we could construct a substantial and rigorous structure for exploring nature. Not only did they open the gate of academia and accompany me throughout this journey full of adventures, but they also broadened my perspective on how individual research contributions collaboratively push the edge of human knowledge and jointly enrich the language of the universe.

Further, I am enormously indebted to my dear father Qi Sun. We encountered traumatic life events in 2019, and the sorrow he has borne is beyond my imagination. But with his determination to seek happiness, I found the exit to this cave of doom and gloom. His love and faith in me have been the cornerstone of my strength during my entire PhD time.

Lastly, I would like to thank my partner Zhuowen Zhou who has been my best friend, soulmate, and source of endless joy. Our philosophical discussions sparked many of my eureka moments. Special thanks to Jiaqi and Chunxiao for their generous help in the research and life. Appreciation also goes to my dear friends Hanbing, Yuchen, Emine, and Stuart, for creating precious memories together at Heriot-Watt and in the city of Edinburgh.

HERIOT WATT UNIVERSITY

Research Thesis Submission

Please note this form should be bound into the submitted thesis.

Name:	**Fandi Sun**		
School:	**MACS**		
Version: *(i.e. First, Resubmission, Final)*	**Final**	Degree Sought:	**PHD**

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1. The thesis embodies the results of my own work and has been composed by myself
2. Where appropriate, I have made acknowledgement of the work of others
3. The thesis is the correct version for submission and is the same version as any electronic versions submitted*.
4. My thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5. I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6. I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

ONLY for submissions including published works

Please note you are only required to complete the Inclusion of Published Works Form (page 2) if your thesis contains published works)

7. Where the thesis contains published outputs under Regulation 6 (9.1.2) or Regulation 43 (9) these are accompanied by a critical review which accurately describes my contribution to the research and, for multi-author outputs, a signed declaration indicating the contribution of each author (complete)
8. Inclusion of published outputs under Regulation 6 (9.1.2) or Regulation 43 (9) shall not constitute plagiarism.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:	Fandi Sun	Date:	**28/02/2023**

Submission

Submitted By *(name in capitals)*:	FANDI SUN
Signature of Individual Submitting:	Fandi Sun
Date Submitted:	**28/02/2023**

For Completion in the Student Service Centre (SSC)

Limited Access	Requested	Yes		No		Approved	Yes		No	
E-thesis Submitted (***mandatory for final theses)***										
Received in the SSC by *(name in capitals)*:						Date:				

Inclusion of Published Works

Please note you are only required to complete the Inclusion of Published Works Form if your thesis contains published works under Regulation 6 (9.1.2)

Declaration

This thesis contains one or more multi-author published works. In accordance with Regulation 6 (9.1.2) I hereby declare that the contributions of each author to these publications is as follows:

Citation details	e. g. Author 1 and Author 2, Title of paper, Title of Journal, X, XX-XX (20XX)
Author 1	Contribution….
Author 2	Contribution….
Signature:	
Date:	

Citation details	e. g. Author 1 and Author 2, Title of paper, Title of Journal, X, XX-XX (20XX)
Author 1	Contribution….
Author 2	Contribution….
Signature:	
Date:	

Citation details	e. g. Author 1 and Author 2, Title of paper, Title of Journal, X, XX-XX (20XX)
Author 1	Contribution….
Author 2	Contribution….
Signature:	
Date:	

Please included additional citations as required.

General Notation

a.s. :	almost surely, or with probability 1.
$A := B$:	A is defined by B or A is denoted by B.
$\emptyset$:	the empty set.
$\mathbf{1}_A$:	the indicator function of a set A, i.e. $\mathbf{1}_A(x) = 1$ if $x \in A$ or otherwise 0.
$\mathbb{I}_d$:	the identity matrix with rank d.
A^T :	the transpose of a vector or matrix A.
$\sigma(\mathcal{C})$:	the σ-algebra generated by $\mathcal{C}$.
$a \vee b$:	the maximum of a and b.
$a \wedge b$:	the minimum of a and b.
$f : A \to B$:	the mapping f from A to B.
$\mathbb{R}^+$:	the set of all nonnegative real numbers, i.e. $\mathbb{R}^+ = [0, \infty)$.
$\mathbb{R}^d$:	the d-dimensional Euclidean space.
$\mathbb{R}^{d\times m}$:	the space of real $d \times m$-matrices.
$\|x\|$:	the Euclidean norm of vector x
$\|x\|_{\mathbf{F}}$:	the Frobenius norm of matrix x.
$\|x\|_{\mathbf{T}_3}$:	the operator norm of 3-tensor x.
$\|x\|_{L_2}$:	$= \left(\mathbb{E}[\|x\|^2]\right)^{1/2}$ for $x \in \mathbb{R}^d$.
$C^m(D, \mathbb{R}^d)$:	the family of continuously m-times differentiable $\mathbb{R}^d$-valued functions defined on D.
$\mathbf{D}V(x)$:	$= \left(\frac{\partial V}{\partial x_1}, \cdots, \frac{\partial V}{\partial x_d}\right)$ for $V \in C^1(\mathbb{R}, \mathbb{R}^d)$. It is the Jacobian matrix if $V \in \mathbb{R}^d$.
$\mathbf{D}^2V(x)$:	$= \left(\frac{\partial^2 V}{\partial x_i \partial x_i}\right)_{d\times d}$ for $V \in C^2(\mathbb{R}, \mathbb{R}^d)$. It is a 3-tensor if $V \in \mathbb{R}^d$.
$\mathcal{L}(X, Y)$:	the set of linear operators $L : X \to Y$ for vector spaces X and Y.
$\mathcal{H}_2^T(\mathbb{R}^d)$:	the set of $\mathbb{R}^d$-valued processes $\{X(t) : t \in [0, T]\}$ where each component belongs to the Banach space such that $\mathbb{E}\left[\int_0^T \|X(s)\|^2 ds\right]^{1/2} < \infty$.

Other notations will be explained where they first appear.

Contents

List of Figures

Chapter 1

Introduction

Stochastic differential equations (SDEs) are practical in modelling real-world problems that involve random behaviours in e.g. derivative pricing [51], stochastic volatility [54, 17], telomere shortening in molecular biology [18] and neural networks [58, 4] etc. In simulating event-driven phenomenons, SDEs with jumps are beneficial in e.g. default modelling for credit risk [60], the spread of cancer from one organ to another in cancer metastasis [52], the natural disaster that spikes claims of property insurance in the fields of catastrophe option pricing [16] and insurance aggregate loss modelling [29] etc.

Numerical methods for approximating these models are essential, whereas some standard methods only converge in restricted conditions. It was shown in [27] that the Euler-Maruyama (EM) method fails to converge with equidistant timesteps in the strong sense to the solutions of SDEs when either of the drift or diffusion coefficients is not globally Lipschitz continuous. Higher-order methods or schemes that are developed based on EM, e.g. Milstein method or EM with jumps, inherit the problem. Milstein method or EM with jumps. For instance, by [57] that the Euler-Maruyama and Euler-Milstein methods coincide in the additive noise case, hence the explicit Milstein scheme over a uniform mesh cannot converge in the mean-square sense to solutions of SDEs with superlinearly growing coefficients, i.e. the mean-square-error between the approximations and the true solutions does not tend to 0 when the step number tends to infinity.

Studies of the variants of these methods that can converge with more relaxed conditions are hence popular. Some focus on the method itself, e.g. by taming [57] or truncating [41] the superlinear terms to reach convergence. Alternatively, restructuring the timestep mesh without modifying the method is also a possibility.

A review of methods that adapt the timestep in order to control local error may be found in the introduction to [30]; we cite here [6, 35, 28, 49, 14, 43] and remark that our purpose is instead to handle the nonlinear response of the discrete system see also [11, 12] and discussion in [30, 31]. A common feature of the adaptivity is the use of both a minimum and maximum time step where the magnitude of the minimum step is controlled by a free parameter which requires some a-priori knowledge on the part of the user. The approach of [11, 12], whose adaptive steps were designed to satisfy specific conditions without the need for a backstop method, was recently extended to McKean-Vlasov equations in [46] and include a Milstein approximation. In addition, we note the fully adaptive Milstein method proposed in [26] for a scalar SDE with light constraints on the coefficients. There the authors stated that such a method was easy to implement but hard to analyse and as a result considered a different, but related method.

The models we approximate in this thesis are the d-dimensional SDE of Itô type

$$X(t) = X(0) + \int_0^t f(X(r))dr + \sum_{i=1}^{m} \int_0^t g_i(X(r))dW_i(r), \tag{1.1}$$

and the d-dimensional SDE driven by Poisson random measure as

$$\begin{aligned} X^{\mathtt{J}}(t) = X^{\mathtt{J}}(0) + \int_0^t f\big(X^{\mathtt{J}}(r)\big)dr + \sum_{i=1}^{m} \int_0^t g_i\big(X^{\mathtt{J}}(r)\big)dW_i(r) \\ + \int_0^t \int_Z \gamma\big(z, X^{\mathtt{J}}(r^-)\big) J_\nu(dz \times dr). \end{aligned} \tag{1.2}$$

Both models are for $t \in [0, T]$, $T \geq 0$ and $i = 1, \dots, m \in \mathbb{N}$, where $W = [W_1, \cdots, W_m]^T$ is an m-dimensional Wiener process, the drift coefficient $f : \mathbb{R}^d \to \mathbb{R}^d$ and the diffusion coefficient $g : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ each satisfy a local Lipschitz condition along with a polynomial growth condition and, together, a monotone condition. Both are twice continuously differentiable; see Assumption 3.1.1 and Assumption 3.1.2. For the jump term in (1.2), J_ν is a Poisson random measure with finite intensity measure ν, and the amplitude coeffient $\gamma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ satisfies a globally Lipschitz condition; see Assumption 4.1.1. Throughout, we take the initial vector $X(0) = X^{\mathtt{J}}(0) = X_0 = X_0^{\mathtt{J}} \in \mathbb{R}^d$ to be deterministic. Notice that when the jump intensity in (1.2) is 0, the whole process is a d-dimensional SDE of the form (1.1).

In the construction of a strongly convergent explicit Milstein-type numerical scheme, we investigate the use of adaptive time-stepping strategies for SDEs (1.1) and jump-adapted adaptive time-stepping strategies for SDEs driven by Poisson random measure (1.2), so that both can achieve strong root-mean-square convergence of order one. The strong convergence rate is generally the lower bound of the weak convergence rate, i.e. strong convergence implies weak convergence but not with the optimal rate [32]. This thesis only considers mean-square strong convergence. As an immediate consequence of this, in the case of additive noise an adaptive EM method also has root-mean-square convergence of order one.

To prove our convergence result it is essential to introduce a new variant of the admissible class of time-stepping strategies introduced in [31, 30], which we call path-bounded strategies. To cope with the jumps in (1.2), motivated by [5] we merge all the jump times to the adaptive mesh grid. Each adaptive step is set to be constrained by their next jump time. We call such time-stepping the jump-adapted path-bounded strategies.

Our framework for adaptivity was introduced in [30] for an explicit EM method and has since been extended to SDE systems with monotone coefficients in [31] and to SPDE methods in [7]. With an upper and a lower constraint to the adaptive steps, these schemes all use a backstop method when the chosen strategy attempts to select a stepsize below the lower bound. However, we demonstrate here that with an appropriately small upper bound and an appropriately large ratio between the upper and the lower bound, the probability of using the backstop can be made arbitrarily small for a path-bounded strategy. This is consistent with the observations that we have in Chapter 5, and with the intuitive notion that the use of the backstop should be rare in practice (see (e) and (f) of Figure 5.1).

Several variants on the fixed-step Milstein method for approximating SDE (1.1) have been proposed, see for example the tamed Milstein [57, 34], projected and split-step backward Milstein [2], truncated Milstein [19], implicit Milstein methods [25, 59] and a recent Tamed Stochastic Runge-Kutta (of order one) method of [15], all designed to converge strongly to solutions of SDEs with more general drift and diffusions, such as in (1.1). However, with few exceptions (see [34, 2]) explicit methods of this kind have only examined the case where the diffusion coefficient g_i satisfies a commutativity condition. We do not impose a commutativity restriction and hence must consider the associated Lévy areas (see Lemma 2.3.1).

Empirical fixed-step numerical methods for simulating (1.2) are reviewed. Tamed EM for SDE driven by Lévy noise with monotone f, g_i and γ in [8]. Trucated EM with super-linearly growing coefficients in [9]. Implicit EM with globally Lipschitz f, g_i and γ in [24]. Tamed Milstein with monotone f, g_i and γ in [33]. Compensated projected EM with monotone f, g_i and γ in [37]. Two-step Mistein with globally Lipschitz f, g_i and γ in [47].

The structure of the thesis is as follows. In Chapter 2, we present the mathematical preliminaries that we need for the study, with one new result of the moment bound of Lévy area in Lemma 2.3.1. In Chapter 3, we introduce the adaptive Milstein method for approximating (1.1), with its corresponding path-bounded time-stepping strategies introduced in Section 3.2, and with mean-square strong convergence proof and the probability of using backstop proof in Section 3.3. In Chapter 4, we introduce the jump-adapted adaptive Milstein method for approximating (1.2), with its corresponding jump-adapted path-bounded time-stepping strategies introduced in Section 4.2, and its L_2 strong convergence proof in Section 4.3. We then present in Chapter 5 the numerical results for both methods, with one-dimensional and two-dimensional test systems compared to existing fixed-step Milstein methods. Finally, we summarise and discuss the future work in Chapter 6.

The adaptive Milstein method in Chapter 3 with its numerical experiments in Section 5.1 have been submitted for peer-reviewed publication. The jump-adapted adaptive Milstein method in Chapter 4 and its numerical comparisons in Section 5.2 are being prepared for submission to journal.

Chapter 2

Mathematical preliminaries

In this chapter, we define the notations and theorems that we use in this thesis. First of all, we denote the l^2 Euclidean norm as

$$\|v\| := \left(|v_1|^2 + \cdots + |v_d|^2\right)^{1/2}$$

for a column vector $v = [v_1, \ldots, v_d]^T \in \mathbb{R}^d$, and $|\cdot|$ for the absolute value. Next, we denote the Frobenius norm as

$$\|A\|_{\mathbf{F}(d_1 \times d_2)} := \left(\sum_{i=1}^{d_1} \sum_{j=1}^{d_2} |A_{ij}|^2\right)^{1/2}$$

given a matrix $A \in \mathbb{R}^{d_1 \times d_2}$, where A_{ij} denotes the element in i^{th} row and j^{th} column of matrix A. Throughout this thesis, we write $\|A\|_{\mathbf{F}(d\times d)}$ as $\|A\|_{\mathbf{F}}$ for simplicity.

For a function $u : \mathbb{R}^d \to \mathbb{R}^m$ and we use the notation $\mathbf{D}^n u(x)$ to denote the n^{th} derivative of u with respect to $x \in \mathbb{R}^d$. With $\mathcal{L}(X, Y)$ being the set of bounded linear operators $L : X \to Y$ for vector spaces X and Y, $\mathbf{D}^n u(x)$ is a linear operator in $\mathcal{L}(\mathbb{R}^d \times \cdots \times \mathbb{R}^d, \mathbb{R}^m)$, and the notation $\mathbf{D}^n u(x)[h_1, \ldots, h_n]$ is used to denote the action of the linear operator on $[h_1, \ldots, h_n] \in \mathbb{R}^d \times \cdots \times \mathbb{R}^d$. The abbreviation $[h]^n$ is used for $[h, \ldots, h]$ and stands for the outer product of h and itself for n times.

For example, for all $x \in \mathbb{R}^d$ and for all $\phi(x) \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$, $\mathbf{D}\phi(x) := \mathbf{D}^1\phi(x) \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)$ is the Jacobian matrix of $\phi(x)$; $\mathbf{D}^2\phi(x) \in \mathcal{L}(\mathbb{R}^{d\times d}, \mathbb{R}^d)$ is the second derivative of $\phi(x)$ with respect to a vector x, which forms a 3-tensor in $\mathbb{R}^{d\times d\times d}$. We write the operator norm of $\mathbf{D}^2\phi(x)$ as

$$\|\mathbf{D}^2\phi(x)\|_{\mathbf{T}_3} := \sup_{h_1,h_2\in\mathbb{R}^d;\|h_1\|,\|h_2\|\le 1} \left\|\mathbf{D}^2\phi(x)\,(h_1\otimes h_2)\right\|,$$

from which we have $\|\mathbf{D}^2\phi(x)[h]^2\| \le \|\mathbf{D}^2\phi(x)\|_{\mathbf{T}_3}\|\,\|h\|^2$ (noting that $\|[\cdot]^2\|_{\mathbf{F}} = \|\cdot\|^2$).

Further, for $a, b \in \mathbb{R}$, $a \vee b$ denotes $\max\{a, b\}$ and $a \wedge b$ denotes $\min\{a, b\}$. Next, we state Taylor's theorem with the remainder term in integrated form (see for example [38, Thm. A.1]).

Theorem 2.0.1 (Taylor)**.** *If $u \in C^{n+1}(\mathbb{R}^d, \mathbb{R}^m)$, then we have*

$$u(x+h) = u(x) + \mathbf{D}u(x)[h] + \cdots + \frac{1}{n!}\mathbf{D}^n u(x)[h]^n + R_n,$$

where $x, h \in \mathbb{R}^d$ and the remainder is

$$R_n = \frac{1}{n!}\int_0^1 (1-\epsilon)^n \mathbf{D}^{n+1}u(x+\epsilon h)[h]^{n+1} d\epsilon.$$

2.1 Stochastic calculus

In this section, we define the terms that we use from probability theory and stochastic calculus for Wiener process. We refer to [40, 13, 48, 38] for the formal definitions of this section.

For a set Y, we use $\mathcal{B}(Y)$ to denote the **Borel σ-algebra** that is the smallest σ-algebra containing all open subsets of Y. A σ-algebra $\mathcal{G}$ is a **sub σ-algebra** of $\mathcal{F}$ if $F \in \mathcal{F}$ for every $F \in \mathcal{G}$. We say that a function $u : X \to \mathbb{R}^d$ is **$\mathcal{F}$-measurable** if $\{x \in X : u(x) \le a\} \in \mathcal{F}$ for every $a \in \mathbb{R}^d$.

We use the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ to denote a **probability space**, where the σ-algebra $\mathcal{F}$ is a collection of subsets of the sample space Ω which contains all possible outcomes. Probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a set function that assigns a probability to each event in $\mathcal{F}$, with $\mathbb{P}(\Omega) = 1$.

Given a set $\mathcal{T} \in \mathbb{R}$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an H-valued **stochastic process** is a set of H-valued random variables $\{X(t, \omega) : t \in \mathcal{T}, \omega \in \Omega\}$. We write $X(t)$ to denote the process for simplicity.

Additionally, a **filtration** $\{\mathcal{F}_t\}_{t\ge 0}$ is a family of sub σ-algebra of $\mathcal{F}$ that are increasing along with the increased time; that is $\mathcal{F}_s$ is a sub σ-algebra of $\mathcal{F}_t$ for

$s \leq t$. We call the quadruple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ a **filtered probability space**. A stochastic process $\{X(t) : t \in [0, T]\}$ is $\{\mathcal{F}_t\}$-**adapted** if the random variable $X(t)$ is $\mathcal{F}_t$-measurable for all $t \in [0, T]$.

Next, we give the definition of Wiener process, or known as standard Brownian motion. We say $\{W(t) : t \in \mathbb{R}^+\}$ is an $\{\mathcal{F}_t\}$-Wiener process on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ if

(i) $W(0) = 0$ a.s.,

(ii) $W(t)$ is $\{\mathcal{F}_t\}$-adapted and $W(t) - W(s)$ is independent of $\mathcal{F}_s$, $s < t$,

(iii) $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ for $0 \leq s \leq t$, i.e.

$$\mathbb{P}[W(t) - W(s) \leq x] = \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{x} e^{-x^2/\left(2(t-s)\right)} dx,$$

(iv) $W(t, \omega)$ is a continuous as a function of t, for almost all $\omega \in \Omega$.

Notice that property (iv) is a consequence of (ii) and (iii). With a square-integrable process $X(s) \in \mathbb{R}^d$ adapted to the filtration generated by a Wiener process $W(s)$, we call $\int_0^t X(s)dW(s)$ the **Itô integral**. There are two properties of Itô integral that are frequently used in this thesis, first is martingale property, for $0 \leq r \leq t \leq T$,

$$\mathbb{E}\left[\int_0^t X(s)dW(s)\middle|\mathcal{F}_r\right] = \int_0^r X(s)dW(s), \quad a.s., \tag{2.1}$$

and in particular the integral has mean zero. Moreover, we have

$$\mathbb{E}\left[\int_{t_1}^{t_2} X(s)dW(s)\middle|\mathcal{F}_{t_1}\right] = 0.$$

Second property is the conditional **Itô's isometry** (see [40, Thm. 5.9]). Conditioning on $\mathcal{F}_{t_1}$ we have for $0 \leq t_1 \leq t_2 \leq T$,

$$\mathbb{E}\left[\left\|\int_{t_1}^{t_2} X(s)dW(s)\right\|^2\middle|\mathcal{F}_{t_1}\right] = \int_{t_1}^{t_2} \mathbb{E}\Big[\|X(s)\|^2 \Big|\mathcal{F}_{t_1}\Big]ds. \tag{2.2}$$

We now consider d-dimensional SDEs, which are $\mathbb{R}^d$-valued continuous adapated process $u(t) = [u_1(t), \ldots, u_d(t)]^T$ on $t \geq 0$ of the form

$$u(t) = u(0) + \int_0^t f(u(s))ds + \sum_{i=1}^{m} \int_0^t g_i(u(s))dW_i(s), \tag{2.3}$$

where $f(u) = [f_1(u), \dots, f_d(u)]^T \in C^2(\mathbb{R}^d, \mathbb{R}^d)$ and $g \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^{d\times m})$ with $g_i(u) = [g_{1,i}(u), \dots, g_{d,i}(u)]^T \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$. Written in explicit form we have

$$\begin{pmatrix} u_1(t) \\ \vdots \\ u_d(t) \end{pmatrix} = \begin{pmatrix} u_1(0) \\ \vdots \\ u_d(0) \end{pmatrix} + \int_0^t \begin{pmatrix} f_1(u(s)) \\ \vdots \\ f_d(u(s)) \end{pmatrix} ds + \int_0^t \begin{pmatrix} g_{1,1}(u(s)) & \cdots & g_{1,m}(u(s)) \\ \vdots & \ddots & \vdots \\ g_{d,1}(u(s)) & \cdots & g_{d,m}(u(s) \end{pmatrix} \begin{pmatrix} dW_1(s) \\ \vdots \\ dW_m(s) \end{pmatrix}. \tag{2.4}$$

Further, (2.3) can also be written in the differential form as

$$du = f(u)dt + g(u)dW(t). \tag{2.5}$$

Notice that the above three forms (2.3), (2.4) and (2.5) are equivalent.

Furthermore, the multidimensional Itô formula is frequently used in this thesis, it is applied on the multidimensional SDEs that have a unique solution on all intervals $[0, T]$ for $T < \infty$. To achieve this, the drift and diffusion of (1.1) need to satisfy standard conditions, e.g. globally Lipschitz and linear growth condition. Superlinear coefficients, which is studied in this thesis, is presented in Assumption 3.1.1 in Chapter 3. The Itô formula is as follows (see [13, Thm. 5.3]).

Theorem 2.1.1 (multidimensional Itô formula)**.** *Let $u(t)$ be a d-dimensional SDEs on $t \geq 0$ in the form of* (2.3)*. Let $V \in C^{1,2}(\mathbb{R}^+ \times \mathbb{R}^d)$. Then*

$$V(t, u(t)) = V(0, u(0)) + \int_0^t \left(\frac{\partial}{\partial t} + L\right) V(s, u(s))ds + \sum_{i=1}^m \int_0^t L^i V(s, u(s))dW_i(s) \tag{2.6}$$

where for $x \in \mathbb{R}^d$ and $t > 0$,

$$LV(t, x) := f(x)^T \mathbf{D}V(t, x) + \frac{1}{2}\sum_{i=1}^m g_i(x)^T \mathbf{D}^2 V(t, x) g_i(x), \tag{2.7}$$

$$L^i V(t, x) := \mathbf{D}V(t, x)^T g_i(x). \tag{2.8}$$

Here g_i denotes the i^{th} column of the diffusion matrix g. $\mathbf{D}V$ is the Jacobian matrix *and* $\mathbf{D}^2 V$ *the 3-tensor of* $V(t,x)$ *with respect to* x.

The following example is used in our proof of Theorem 3.3.1 on the strong convergence of our numerical scheme.

Example 2.1.1. *Let* $u(t)$ *be a d-dimensional SDE as in* (2.3)*, and* $V(t,u(t)) = \|u(t)\|^2$. *Then, by the multidimensional Itô formula* (2.6)*, we have*

$$\begin{aligned}\left\|u(t)\right\|^2 =& \left\|u(0)\right\|^2 + 2\int_0^t \Big\langle f(u(s)), u(s)\Big\rangle ds + \sum_{i=1}^m \int_0^t \left\|g_i(u(s))\right\|^2 ds \\ &+ 2\sum_{i=1}^m \int_0^t \Big\langle u(s), g_i(u(s))\Big\rangle dW_i(s).\end{aligned}$$

Proof. We first get the essential elements in (2.7) and (2.8). With $V(s,x(s)) = \|u(s)\|^2$, we have $\frac{\partial V(s,x(s))}{\partial t} = 0$ and

$$\mathbf{D}V(s,x(s)) = \frac{\partial\left(\left\|u(s)\right\|^2\right)}{\partial u(s)} = \left[\frac{\partial\left(\sum_{i=1}^d (u_i)^2\right)}{\partial u_1(s)}, \ldots, \frac{\partial\left(\sum_{i=1}^d (u_i)^2\right)}{\partial u_d(s)}\right] = 2u(s), \tag{2.9}$$

$$\mathbf{D}^2 V(s,x(s)) = \frac{\partial^2\left(\left\|u(s)\right\|^2\right)}{\partial u(s)^2} = 2\frac{\partial u(s)}{\partial u(s)} = 2\,\mathbb{I}_d,$$

where $\mathbb{I}_d$ denotes identity matrix with rank d. Therefore, for (2.7),

$$f(u(s))^T \mathbf{D}V(s,x(s)) = 2\Big\langle f(u(s)), u(s)\Big\rangle, \tag{2.10}$$

$$\frac{1}{2}\sum_{i=1}^m g_i(u(s))^T \mathbf{D}^2 V(s,u(s)) g_i(u(s)) = \frac{2}{2}\sum_{i=1}^m [g_{1,i}, \ldots, g_{d,i}]\,\mathbb{I}_d \begin{bmatrix} g_{1,i} \\ \vdots \\ g_{d,i}\end{bmatrix} = \sum_{i=1}^m \|g_i\|^2,$$

where $g_{j,i}$ stands for the element at i^{th} column and j^{th} row of matrix g. Following (2.9) and (2.10), (2.8) can be derived accordingly. Substituting the corresponding elements in (2.7) and (2.8) back to (2.6), we have the desired result. □

To analyse adaptive time-stepping which are random variables, we further define stochastic calculus and the corresponding proprieties with stopping times below. We refer to [40, Chap.5], [42, Chap.1] and [13, Thm.4.3] for further details.

Definition 2.1.1 (stopping time)**.** A random variable $\mu : \Omega \to [0, \infty]$ is called an $\{\mathcal{F}_t\}$-stopping time (or simply, stopping time) if $\{\omega : \mu(\omega) \le t\} \in \mathcal{F}_t$ for any $t \ge 0$. Let τ_1 and τ_2 be two stopping times with $\tau_1 \le \tau_2$. We define

$$[[\mu_1, \mu_2]] = \big\{(t, \omega) \in \mathbb{R}^+ \times \Omega : \mu_1(\omega) \le t \le \mu_2(\omega)\big\},$$

and call it a stochastic interval. If μ is a stopping time, define

$$\mathcal{F}_\mu := \{A \in \mathcal{F} \,:\, A \cap \{\omega : \mu(\omega) \le t\} \in \mathcal{F}_t, \text{ for all } t \ge 0\}, \tag{2.11}$$

which is a sub-σ-algebra of $\mathcal{F}$.

To define the stochastic integrals with stopping time, we need to have the indicator process $\{\mathbf{1}_{[[0,\mu]]}(t)\}_{t\ge 0}$. If μ is an $\{\mathcal{F}_t\}$-stopping time, then it is a bounded right continuous $\{\mathcal{F}_t\}$-adapted process. Moreover, $\mathbf{1}_{[[0,\mu]]}(t)$ is predictable and $\mathcal{F}_t$-measurable (see [40, Sec. 1.3]). Next, we let $\mathcal{H}_2^T(\mathbb{R})$ be the set of $\mathbb{R}$-valued processes $\{X(t) : t \in [0, T]\}$ such that $\mathbb{E}\left[\int_0^T |X(s)|^2 ds\right]^{1/2} < \infty$, and we define the stochastic Itô integral with stopping time as follows.

Definition 2.1.2 (stochastic integral with stopping time)**.** Let $f \in \mathcal{H}_2^T(\mathbb{R})$, and let μ be an $\mathcal{F}_t$-stopping time such that $0 \le \mu \le T$. Then, $\{\mathbf{1}_{[[0,\mu]]}(t)f(t)\}_{0\le t\le T} \in \mathcal{H}_2^T(\mathbb{R})$, and we define

$$\int_0^\mu f(s)dW(s) = \int_0^T \mathbf{1}_{[[0,\mu]]}(s)f(s)dW(s).$$

Consequently, we have the following properties from (2.1) and (2.2) hold for $X \in \mathcal{H}_2^T(\mathbb{R}^d)$ and stopping times $0 \le \tau_1 \le \tau_2 \le T$ with respect to $\mathcal{F}_t$ that

$$\mathbb{E}\left[\int_{\tau_1}^{\tau_2} X(t)dW(t)\middle|\mathcal{F}_{\tau_1}\right] = 0, \tag{2.12}$$

$$\mathbb{E}\left[\left\|\int_{\tau_1}^{\tau_2} X(t)dW(t)\right\|^2\middle|\mathcal{F}_{\tau_1}\right] = \int_{\tau_1}^{\tau_2} \mathbb{E}\left[\|X(t)\|_{\mathbf{F}}^2\middle|\mathcal{F}_{\tau_1}\right]dt. \tag{2.13}$$

Remark 1 (**Itô formula with stopping time**)**.** Notice that (2.6) shows that its LHS and RHS are stochastically equivalent. Since they are continuous, their sample paths coincide. Therefore, for any $\mathcal{F}_t$-measurable stopping time $\mu \in [0, T]$, the Itô formula (2.6) applies as

$$V(\mu,u(\mu)) = V(0,u(0)) + \int_0^\mu \left(\frac{\partial}{\partial t} + L\right) V(s,u(s))ds + \sum_{i=1}^m \int_0^\mu L^i V(s,u(s))dW_i(s).$$

2.2 SDEs driven by Poisson random measure

In Chapter 4, we introduce the jump-adapted adaptive Milstein method for SDEs driven by Poisson random measure, so in the section we define the settings for jump processes, see for example [53, 45, 48].

First of all, the simplest jump process is the homogeneous **Poisson point process**, or the counting process. Let $(\pi_i)_{i\geq 1}$ be a sequence of independent exponential random variables with parameter $\lambda \in \mathbb{R}^+$ and $\tau_n = \sum_{i=1}^n \pi_i$. The process $(N_t)_{t\geq 0}$ counts the number of jumps between 0 and t as defined by

$$N(t) := \#\big\{i \geq 1,\ \tau_i \in [0,t]\big\}.$$

With an additional finite set $Z := \mathbb{R}^d \backslash \{0\}$ for jump sizes, we define the **Poisson random measure** that we use in this thesis as $J_\nu : Z \times [0,T] \to \mathbb{N}$ with finite intensity ν. Such a measure counts the number of jumps between 0 and T whose jump size $(\zeta_n)_{n\geq 1} \in Z$:

$$J_\nu(Z \times T) := \#\Big\{i \geq 1,\ (\zeta_i, \tau_i) \in Z \times [0,T]\Big\}. \tag{2.14}$$

By integrating an amplitude function $\gamma : Z \times [0,T] \to \mathbb{R}^d$ with respect to the Poisson random measure J_ν we have a **jump process** $X(t)$ as

$$X(t) = \int_0^T \int_Z \gamma(z,t) J_\nu(dz \times dt) = \sum_{\{n,\, \tau_n \in [0,T]\}} \gamma(\zeta_n, \tau_n). \tag{2.15}$$

Whenever there is a jump time $t = \tau \in [0,T]$ and a jump size $z = \zeta \in Z$, the measure $J_\nu(dz \times dt) = 1$ so we have the jump amplitude $\gamma(\zeta,\tau)$ added to the process. We assume that the intensity measure $\nu(dz)dt = \mathbb{E}[J_\nu(dz \times dt)]$ has finite total intensity $\lambda = \nu(Z) < \infty$. Since we will be working on a jump-adapted mesh which requires analysis on the particular jump times, the following notation is useful when the jump process is at a point t:

$$\int_Z \gamma(z,t^-)J_\nu(dz\times\{t\}) := \int_0^t\int_Z \gamma(z,s)J_\nu(dz\times ds) - \int_0^{t^-}\int_Z \gamma(z,s)J_\nu(dz\times ds). \quad (2.16)$$

Remark 2 (**Notation** $J(t)$)**.** By (2.14) the number of jumps between $[0,t]$, whose sizes are in the set Z, is denoted as $J_\nu(Z\times t)$. In this thesis, we omit the Z and ν, and write $J(t)$ for simplicity.

The Poisson random measure $J_\nu(Z\times T)$ in (2.14) generates a sequence of pairs $\{(\zeta_i,\tau_i), i\in\{1,2,\ldots,J(T)\}\}$. In more detail, $(\zeta_i)_{i\geq 1}$, which can be observed at their corresponding jump times, is a sequence of i.i.d. random variables representing the **jump sizes**. $(\tau_i)_{i\geq1}$ is a sequence of increasing non-negative random variables representing the **jump times** of a Poisson process with intensity λ, with the **waiting time** between every two jumps denoted as $\pi_i\sim\exp(1/\lambda)$ for $i=1,\ldots,J(T)$ so that $\tau_i=\sum_{j=1}^i \pi_i$.

Adding the jump process (2.15) to an SDE (2.3) we have the d-dimensional **SDE driven by Poisson random measure** as

$$u(t) = u(0) + \int_0^t f(u(r))dr + \sum_{i=1}^m \int_0^t g_i(u(r))dW_i(r) + \int_0^t\int_Z \gamma\big(z,u(r^-)\big)J_\nu(dz\times dr), \quad (2.17)$$

where drift $f(u)=[f_1(u),\ldots,f_d(u)]^T\in C^2(\mathbb{R}^d,\mathbb{R}^d)$ and diffusion $g\in \mathrm{C}^2(\mathbb{R}^d,\mathbb{R}^{d\times m})$ with $g_i(u) = [g_{1,i}(u),\ldots,g_{d,i}(u)]^T\in\mathrm{C}^2(\mathbb{R}^d,\mathbb{R}^d)$, and jump coefficient $\gamma(z,u) = [\gamma_1(z,u),\ldots,\gamma_d(z,u)]^T\in\mathrm{C}^2\big((\mathbb{R}^d\backslash\{0\})\times\mathbb{R}^d,\mathbb{R}^d\big)$.

One example comparison of the trajectories of Poisson counting process, compound Poisson process and compound Poisson process with Wiener noises is shown in Figure 2.1. All three processes are modeled based on the same jump times, with the counting process (a) adding value 1 on a jump time, and the other two (b)-(c) adding a random value drew from a probability distribution e.g. $\mathcal{N}(0,1)$ here. By including Wiener noises during the waiting time of the jumps to the compound Poisson process, we have the plot (c) that is the SDE driven by Poisson random measure (2.17) when drift $f=0$, diffusion $g_i=1$ and jump amplitude function equals the noises written as $\gamma = z$.

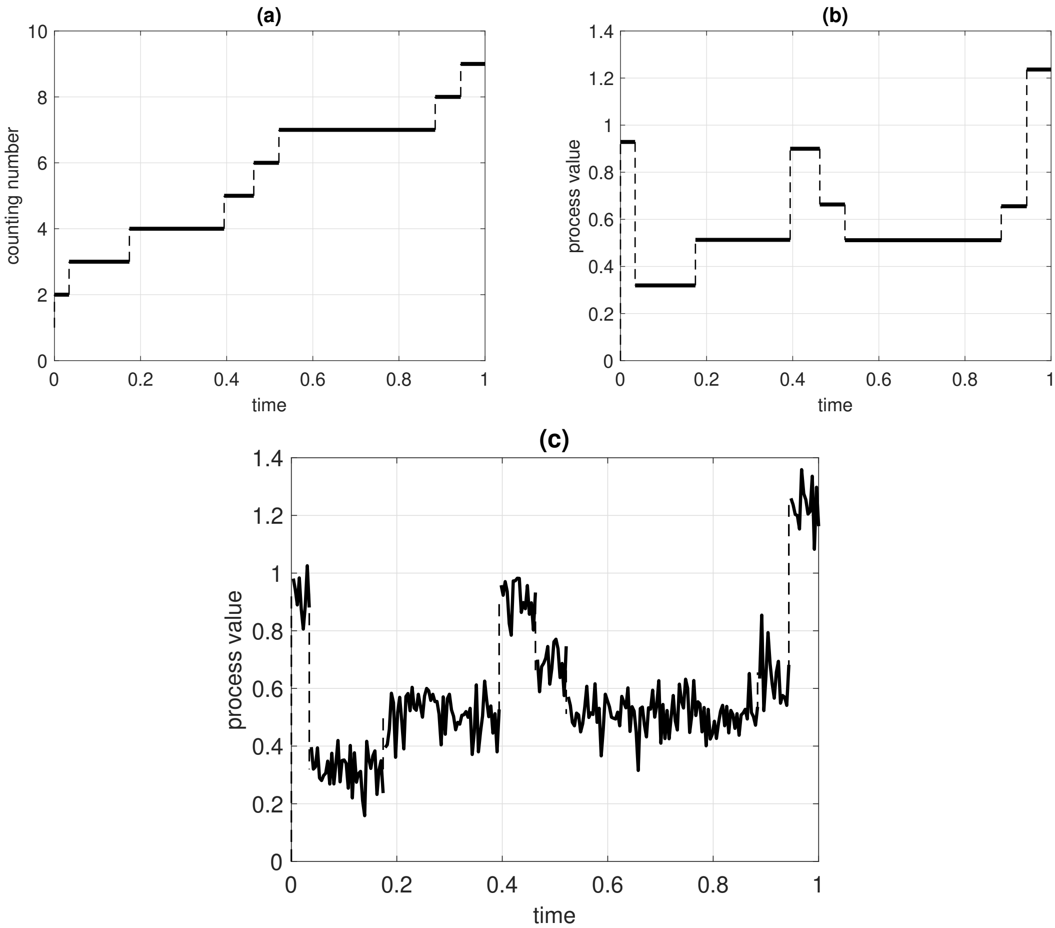

Figure 2.1: Trajectories of counting process (a), compound Poisson process (b), and compound Poisson with Wiener noise (c).

Finally, we state the Itô formula for SDEs with jumps (1.2), (see [53, Chap.8]). The coefficients should be assumed to guarantee that (1.2) has a unique solution on all intervals $[0, T]$ for $T < \infty$. The conditions that we assume in this thesis are in Assumption 3.1.1 and 4.1.1.

Theorem 2.2.1 (Itô formula for SDE driven by Poisson random measure)**.** *Let $u(t)$ be a d-dimensional SDE on $t \geq 0$ in the form of* (2.17) *. Let $V \in C^2(\mathbb{R}^d, \mathbb{R}^+ \times \mathbb{R}^d)$, we have*

$$V(t, u(t)) = V(0, u(0)) + \int_0^t \left(\frac{\partial}{\partial t} + L\right) V(s, u(s))ds + \sum_{i=1}^{m} \int_0^t L^i V(s, u(s))dW_i(s) + \int_0^t \int_Z \Big\{V\Big(s,\ u(s^-) + \Delta u(s)\Big) - V\Big(s, u(s^-)\Big)\Big\} J(dz \times ds) \quad (2.18)$$

where $LV(t, x)$ and $L^iV(t, x)$ are in (2.7) *and* (2.8)*, respectively, from multidimensional Itô's formula for SDE in Theorem* (2.1.1)*. $\Delta u(s) := u(s) - u(s^-)$. The jump* *term in* (2.18) *can be written in the compound Poisson form as*

$$\sum_{i=1}^{J(t)} \left\{ V\Big(\tau_i,\ u(\tau_i^-) + \Delta u(\tau_i)\Big) - V\big(\tau_i, u(\tau_i^-)\big) \right\}.$$

Notice that due to Remark 1, the Itô formula for SDEs with jumps (2.18) can also be applied with stopping times.

2.3 Numerical methods

In this section, we define some existing numerical methods that we consider in this thesis. First of all, for $n \in \mathbb{N}$ we define $\Delta t_n = |t_{n+1} - t_n|$, and following [57, 2], the stochastic integral and the iterated stochastic integral are defined as

$$I_i^{t_n,t_{n+1}} := \int_{t_n}^{t_{n+1}} dW_i(s), \qquad I_{j,i}^{t_n,t_{n+1}} := \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW_j(p) dW_i(s). \tag{2.19}$$

Given Y_n and for $n \in \mathbb{N}$, the discrete-time Euler-Maruyama (EM) method for approximating the SDE (1.1) is given as

$$Y_{n+1} = Y_n + f(Y_n)\Delta t_n + \sum_{i=1}^{m} g_i(Y_n) I_i^{t_n,t_{n+1}}. \tag{2.20}$$

At the n^{th} step, substituting the given current value Y_n into coefficient functions, together with Wiener increment approximated by $\mathcal{N}(0,1)$ random variable scaled by $\sqrt{\Delta t}$, we have the approximated value Y_{n+1}. The EM method has strong convergence of order 1/2. With one extra term, we have the Milstein method with strong convergence of order 1. We define the continuous-time interpolation over one step $[t_n, t_{n+1}]$ of the discrete-time Milstein approximation as (see [57])

Definition 2.3.1 (Explicit Milstein method)**.** For $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$, the fixed-step Milstein scheme for the SDE in (1.1) is

$$\begin{aligned} Y(s) = Y(t_n) + f\big(Y(t_n)\big)|s - t_n| + \sum_{i=1}^{m} g_i\big(Y(t_n)\big) I_i^{t_n,s} \\ + \sum_{i,j=1}^{m} \mathbf{D} g_i\big(Y(t_n)\big) g_j\big(Y(t_n)\big) I_{j,i}^{t_n,s}. \end{aligned} \tag{2.21}$$

Notice that $Y(t_n) = Y_n$ for $n \in \mathbb{N}$, i.e. on grid points. Expanding the last term in (2.21), with $I_{j,i}^{t_n,s}$ defined in (2.19) we have

$$\begin{aligned}
&\sum_{i,j=1}^{m} \mathbf{D}g_i\big(Y(t_n)\big)g_j\big(Y(t_n)\big)I_{j,i}^{t_n,s} \\
=&\frac{1}{2}\sum_{i=1}^{m} \mathbf{D}g_i\big(Y(t_n)\big)g_i\big(Y(t_n)\big)\left(\left(I_i^{t_n,s}\right)^2 - |s-t_n|\right) \\
&+\frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^{m} \Big(\mathbf{D}g_i\big(Y(t_n)\big)g_j\big(Y(t_n)\big) + \mathbf{D}g_j\big(Y(t_n)\big)g_i\big(Y(t_n)\big)\Big) I_i^{t_n,s}I_j^{t_n,s} \\
&+\sum_{\substack{i,j=1\\ i<j}}^{m} \Big(\mathbf{D}g_i\big(Y(t_n)\big)g_j\big(Y(t_n)\big) - \mathbf{D}g_j\big(Y(t_n)\big)g_i\big(Y(t_n)\big)\Big) A_{ij}^{t_n,s},
\end{aligned} \tag{2.22}$$

where the term $A_{ij}^{t_n,s}$ is the Lévy area (see for example [36, Eq. (1.2.2)]) defined by

$$A_{ij}^{t_n,s} := \frac{1}{2}\left(I_{i,j}^{t_n,s} - I_{j,i}^{t_n,s}\right), \tag{2.23}$$

and we have used the relations $I_{i,i}^{t_n,s} = \frac{1}{2}((I_i^{t_n,s})^2 - |t-s|)$ and $I_{i,j}^{t_n,s} + I_{j,i}^{t_n,s} = I_i^{t_n,s}I_j^{t_n,s}$. When only the first term on the RHS of (2.22) exists we say it is an SDE with diagonal noise. If first two terms exist, it is an SDE with commutative noise. If all three terms exist, it is the non-commutative noise. Notice that when the noise is additive, i.e. $\mathbf{D}g_i\big(Y(t_n)\big) = 0$ for $i = 1, ..., m$, the EM method and Milstein method coincide with convergence order 1.

Many authors assume the commutativity condition that for all $i, j = 1, \dots, m$ and $y \in \mathbb{R}^d$, $\mathbf{D}g_i(y)g_j(y) = \mathbf{D}g_j(y)g_i(y)$. When this holds, the last term in (2.22) vanishes, avoiding the need for any analysis of $A_{ij}^{t_n,s}$ defined in (2.23). We do not impose such a condition in this thesis, and therefore make use of the following conditional moment bounds on the Lévy areas.

Lemma 2.3.1 (`Lévy Area`). *For all $i, j = 1, \dots, m$, $0 \le t_n \le s < T$ and for a pair of Wiener process $(W_i(r), W_j(r))^T$ where $r \in [t_n, s]$ and the Lévy area $A_{ij}^{t_n,s}$ defined in (2.23), there exists a finite constant $C_{\mathtt{LA}}$ whose explicit form is in (2.25) such that for $k \ge 1$*

$$\mathbb{E}\left[\left|A_{ij}^{t_n,s}\right|^k \middle| \mathcal{F}_{t_n}\right] \le C_{\mathtt{LA}}(k)\ |s-t_n|^k \quad a.s. \tag{2.24}$$

Proof. Set $\mathtt{i}^2 = -1$. Since the pair of Wiener processes $(W_i(r), W_j(r))^T$, $r \in [t_n, s]$, are mutually independent, by [36, Eq. (1.3.5)] the characteristic function of the

Lévy area (2.23) is given by $\phi(\lambda) = \left(\cosh\left(\frac{1}{2}|s-t_n|\lambda\right)\right)^{-1}$. This was applied in the context of numerical methods for SDEs in [39]. The Taylor expansion of the function $\cosh\left(\frac{1}{2}|s-t_n|\lambda\right)$ around 0 gives

$$\phi(\lambda) = \sum_{N=0}^{\infty} \frac{\mathbf{E}_{2N}}{(2N)!}\left(\frac{1}{2}|s-t_n|\right)^{2N} \lambda^{2N}, \quad \left|\frac{1}{2}|s-t_n|\lambda\right| < \frac{\pi}{2},$$

where $\mathbf{E}_{2N}$ stands for the $2N^{\text{th}}$ Euler number, which may be expressed as

$$\mathbf{E}_{2N} = \mathtt{i}\sum_{b=1}^{2N+1}\sum_{j=0}^{b}\binom{j}{b}\frac{(-1)^j(b-2j)^{2N+1}}{2^b\,\mathtt{i}^b\,b}, \quad N = 0,1,2,3,\ldots.$$

All odd Euler numbers are zero. The k^{th} derivative of the characteristic function with respect to λ is

$$\phi(\lambda)^{(k)}_{\lambda} = \sum_{N=\lceil\frac{k}{2}\rceil}^{\infty}\left(\prod_{B=0}^{k-1}(2N-B)\right)\frac{\mathbf{E}_{2N}}{(2N)!}\left(\frac{1}{2}|s-t_n|\right)^{2N}\lambda^{2N-k}.$$

As $\lambda \to 0$, since all terms vanish unless $k = 2N$, we have

$$\lim_{\lambda\to 0}\phi(\lambda)^{(k)}_{\lambda} = \begin{cases}\left(\prod_{B=0}^{k-1}(k-B)\right)\frac{\mathbf{E}_k}{(k)!}\left(\frac{1}{2}|s-t_n|\right)^{k}, & k \text{ even;}\\ 0, & k \text{ odd.}\end{cases}$$

In the calculation of expectations, we make use of the mutual independence, conditional upon $\mathcal{F}_{t_n}$, of the pair of Wiener increments $(W_i(t), W_j(t))^T$. Therefore, the k^{th} conditional moment of $A_{ij}^{t_n,s}$ is

$$\mathbb{E}\left[\left(A_{ij}^{t_n,s}\right)^k\middle|\mathcal{F}_{t_n}\right] = L_k\,|s-t_n|^k,$$

where for all $a = 1,2,3,\ldots$

$$\begin{aligned} L_k &= \left(\prod_{B=0}^{k-1}(k-B)\right)\frac{\mathbf{E}_k}{(k)!}\left(-\frac{1}{2}\,\mathtt{i}\right)^k \\ &:= \begin{cases}\left(\prod_{B=0}^{k-1}(k-B)\right)\frac{\mathbf{E}_k}{(k)!}\left(\frac{1}{2}\right)^k, & k = 4a = 4,8,12,\ldots\\ -\left(\prod_{B=0}^{k-1}(k-B)\right)\frac{\mathbf{E}_k}{(k)!}\left(\frac{1}{2}\right)^k, & k = 4a-2 = 2,6,10\ldots\\ 0, & k = 2a-1 = 1,3,5,\ldots\end{cases}\end{aligned}$$

which is finite, as a finite product of finite factors. When k is even, we have

$$\mathbb{E}\Big[\big|A_{ij}^{t_n,s}\big|^k\Big|\mathcal{F}_{t_n}\Big] = \mathbb{E}\Big[\big(A_{ij}^{t_n,s}\big)^k\Big|\mathcal{F}_{t_n}\Big] = L_k\,|s-t_n|^k, \quad a.s.$$

When k is odd, i.e. $k = 2c+1$ for all $c = 0,1,2,\dots$, we have a.s.

$$\begin{aligned}
\mathbb{E}\Big[\big|A_{ij}^{t_n,s}\big|^k\Big|\mathcal{F}_{t_n}\Big] &= \mathbb{E}\Big[\big|A_{ij}^{t_n,s}\big|^{2c+1}\Big|\mathcal{F}_{t_n}\Big] \\
&\leq \sqrt{\mathbb{E}\Big[\big(A_{ij}^{t_n,s}\big)^{4c}\Big|\mathcal{F}_{t_n}\Big]\mathbb{E}\Big[\big(A_{ij}^{t_n,s}\big)^{2}\Big|\mathcal{F}_{t_n}\Big]} \\
&= \begin{cases} \sqrt{L_2}\,|s-t_n|, & c=0; \\ \sqrt{L_{4c}\cdot L_2}|s-t_n|^{2c+1}, & c=1,2,3,\dots \end{cases} \\
&= \begin{cases} \sqrt{L_2}\,|s-t_n|, & k=1; \\ \sqrt{L_{2k-2}\cdot L_2}|s-t_n|^{k}, & k=3,5,7,\dots. \end{cases}
\end{aligned}$$

Therefore, in conclusion we have

$$\mathbb{E}\Big[\big|A_{ij}^{t_n,s}\big|^k\Big|\mathcal{F}_{t_n}\Big] \;\leq\; C_{\mathtt{LA}}\,(k)\,|s-t_n|^k, \quad a.s.$$

where

$$C_{\mathtt{LA}}\,(k) = \begin{cases} \sqrt{L_2}, & k=1; \\ \sqrt{L_{2k-2}\cdot L_2}, & k=3,5,7,\dots; \\ L_k, & k=2,4,6,\dots \end{cases} \tag{2.25}$$

□

The following discrete-time method reaches l_2 strong convergence of order 1 for approximating the SDE (1.1), with the drift coefficient is one-sided Lipschitz continuous and the diffusion coefficient is globally Lipschitz continuous (see [57]).

Definition 2.3.2 (Tamed Milstein method)**.** Given Y_n and $n \in \mathbb{N}$, the fixed-step tamed Milstein scheme for the SDE (1.1) is given as

$$Y_{n+1} = Y_n + \frac{f(Y_n)}{1 + \Delta t_n \big\| f(Y_n) \big\|} \Delta t_n + \sum_{i=1}^{m} g_i(Y_n) I_i^{t_n,t_{n+1}} + \sum_{i,j=1}^{m} \mathbf{D} g_i(Y_n) g_j(Y_n) I_{j,i}^{t_n,t_{n+1}}. \quad (2.26)$$

When dealing with the SDE with jumps as in (2.17) with globally Lipschitz coefficients, the following Milstein method can reach strong convergence with order 1, (see [5]). Given Y_n and $n \in \mathbb{N}$, the discrete-time Milstein scheme for the SDE with jumps is given in one-step as

$$\begin{aligned} Y_{n+1} =& Y_n + f(Y_n)\Delta t + \sum_{i=1}^{m} g_i(Y_n) I_i^{t_n,t_{n+1}} + \int_{t_n}^{s} \int_Z \gamma\big(z, Y_n\big) J(dz \times ds) \\ &+ \sum_{i,j=1}^{m} g_j(Y_n))\mathbf{D} g_i(Y_n) I_{j,i}^{t_n,t_{n+1}} \\ &+ \sum_{i=1}^{m} \int_{t_n}^{s} \int_{t_n}^{r} \int_Z \Delta g_i\big(z, Y_n\big) J(dz \times dr) dW_i(s) \\ &+ \sum_{i=1}^{m} \int_{t_n}^{s} \int_Z \int_{t_n}^{r} g_i(Y_n) \mathbf{D}\gamma\big(z, Y_n\big) dW_i(r) J(dz \times ds) \\ &+ \int_{t_n}^{t_{n+1}} \int_Z \int_{t_n}^{r} \int_Z \Delta\gamma\big(z, z_1, Y_n\big) J(dz_1 \times dr) J(dz \times ds), \end{aligned} \quad (2.27)$$

where

$$\begin{aligned} \Delta g_i\big(z, Y_n\big) =& g_i\Big(Y_n^- + \gamma\big(z, Y_n\big)\Big) - g_i\big(Y_n^-\big) \\ \Delta\gamma\big(z, z_1, Y(t_n)\big) =& \gamma\Big(z, Y_n^- + \gamma\big(z_1, Y_n\big)\Big) - \gamma\big(z, Y_n^-\big). \end{aligned}$$

By including all the jumps in the mesh points, the following continuous-time interpolation of the discrete-time jump-adapted Milstein method can also reach strong convergence with order 1 with globally Lipschitz coefficients, (see [5]).

Definition 2.3.3 (Jump-adapted Milstein method for SDE driven by Poisson random measure)**.** Given $Y(t_n)$, for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$, the continuous form of jump-adapted Milstein scheme for approximating (2.17) is given by $Y(s)$ as

$$\widehat{\Upsilon}(s) = Y(t_n) + f\big(Y(t_n)\big)|s - t_n| + \sum_{i=1}^{m} g_i\big(Y(t_n)\big) I_i^{t_n,s} + \sum_{i,j=1}^{m} \mathbf{D} g_i\big(Y(t_n)\big) g_j\big(Y(t_n)\big) I_{j,i}^{t_n,s} \tag{2.28}$$

$$Y(s) = \widehat{\Upsilon}(s) + \int_Z \gamma\big(z, \widehat{\Upsilon}(s)\big) J_\nu(dz \times \{s\}), \tag{2.29}$$

where

$$\int_Z \gamma\big(z, \widehat{\Upsilon}(s)\big) J_\nu(dz \times \{s\}) = \begin{cases} \gamma\big(\zeta_{J(s)}, \widehat{\Upsilon}(s)\big), & s = \tau_{J(s)}, \\ 0, & s \neq \tau_{J(s)}. \end{cases} \tag{2.30}$$

The notation $\widehat{\Upsilon}(s)$ in (2.28) stands for the process that approximates up to the time s without the jump if there is one at s. Further, the jump-adapted method in (2.29) is based on a deterministic mesh that is a superposition of the jump times $\{\tau_i\}_{i\in[1,\ldots,J(s)]}$ and equidistant time steps. By Remark 2, $\tau_{J(s)}$ in (2.30) shows the last jump time in $[0, s]$, so that (2.30) can catch every jump when s is the jump time, with the jump size $\zeta_{J(s)}$. The jump term can be sampled exactly.

Comparing (2.29) with the non-jump-adapted Milstein method (2.27), we can see that by tracking each jump on the mesh, the jump-adapted Milstein method (2.29) becomes much simpler. The last three double integrals coupled with Wiener process and/or Poisson random measure in (2.27) do not present in the the jump-adapted methods. This simplifies the analysis and numerical implementation but can be computational expensive when jump intensity is high.

Finally, the definition of L_2 (root-mean-square) strong convergence is as follows.

Definition 2.3.4 (strong convergence)**.** Let $(X(t))_{t\in[0,T]}$ be the solution of a stochastic differential equation and let $(Y(t))_{t\in[0,T]}$ be an approximation method of the stochastic differential equation. We say $(Y(t))_{t\in[0,T]}$ converges L_2-strongly with order α if the following holds

$$\sup_{0\le t\le T} \Big(\mathbb{E}\Big[\big\|X(t) - Y(t)\big\|^2\Big]\Big)^{1/2} \le \mathcal{O}(\Delta t^\alpha).$$

In contrast, we say $(Y(t))_{t\in[0,T]}$ convergent weakly with order β if the following bound holds

$$\Big\|\mathbb{E}\big[X(t)\big]-\mathbb{E}\big[Y(t)\big]\Big\|\leq\mathcal{O}(\Delta t^{\beta}).$$

We see that strong convergence focuses on the mean of the error, whereas weak convergence is on error of the mean. Strong convergence implies weak convergence with a rate that is not optimal, we in this thesis investigate the strong convergence of the proposed schemes.

2.4 Some useful results

In this section, we list some useful inequality-related results that we frequently use in this thesis. For general references, see [40, 38].

For $a, b, \epsilon > 0$, we have the **Young's inequality** as

$$2ab\leq\epsilon a^2+\frac{1}{\epsilon}b^2. \tag{2.31}$$

Next, if $\phi:\mathbb{R}\to\mathbb{R}$ is a convex function and X is a real-valued random variable with $\mathbb{E}[X]<\infty$, then we have **Jensen's inequality** as

$$\phi\big(\mathbb{E}[X]\big)\leq\mathbb{E}\big[\phi(X)\big]. \tag{2.32}$$

In particular, for $p\geq 1$, $\big(\mathbb{E}[|X|]\big)^p\leq\mathbb{E}\big[|X|^p\big]$. Furthermore, the following two consequences are also commonly used. For $f\in L^1$ and $p\geq 1$,

$$\left|\int_0^t f(s)ds\right|^p\leq t^{p-1}\int_0^t|f(s)|^p ds,\quad t\geq 0. \tag{2.33}$$

For $a_i\in\mathbb{R}$ and $p\geq 1$,

$$\left|\sum_{i=1}^n a_i\right|^p\leq n^{p-1}\sum_{i=1}^n|a_i|^p,\quad n\in\mathbb{N}\backslash\{0\}. \tag{2.34}$$

Moreover, let H be a Hilbert space. Then the **Cauchy-Schwarz inequality** is

$$\|\langle u,v\rangle\|\leq\|u\|\,\|u\|,\quad\forall u,v\in H. \tag{2.35}$$

Let X and Y be random variables, we have

$$\big|\mathbb{E}[XY]\big| \leq \Big(\mathbb{E}\big[X^2\big]\Big)^{1/2}\Big(\mathbb{E}\big[Y^2\big]\Big)^{1/2}. \tag{2.36}$$

Suppose that $(X_k, \mathcal{F}_k, \mu_k)$ for $k = 1, 2$ are σ-finite measure spaces and consider a measurable function $u : X_1 \times X_2 \to Y$. If

$$\int_{X_2}\left(\int_{X_1}\big\|u(x_1, x_2)\big\|_Y d\mu_1(x_1)\right) d\mu_2(x_2) < \infty, \tag{2.37}$$

then by **Fubini's theorem** u is integrable with respect to the product measure $\mu_1 \times \mu_2$ and

$$\begin{aligned}\int_{X_1\times X_2} u(x_1, x_2)d(\mu_1 \times \mu_2)(x_1, x_2) &= \int_{X_2}\left(\int_{X_1} u(x_1, x_2)d\mu_1(x_1)\right) d\mu_2(x_2)\\ &= \int_{X_1}\left(\int_{X_2} u(x_1, x_2)d\mu_2(x_2)\right) d\mu_1(x_1).\end{aligned} \tag{2.38}$$

Finally, let $T > 0$ and $c \geq 0$, $u(\cdot)$ be a Borel measurable bounded nonnegative function on $[0, T]$, and $v(\cdot)$ be a nonnegative integrable function on $[0, T]$. If

$$u(t) \leq c + \int_0^t v(s)u(s)ds, \quad \text{for all } 0 \leq t \leq T,$$

then by **Gronwall's inequality**

$$u(t) \leq c\exp\left(\int_0^t v(s)ds\right), \quad \text{for all } 0 \leq t \leq T. \tag{2.39}$$

Chapter 3

Adaptive Milstein method

In this chapter, we introduce the adaptive Mistein method for approximating (1.1) with L_2-strong convergence of order one. The assumptions of the model are stated in Section 3.1, and the adaptive time-stepping strategies in Section 3.2. The theorems of L_2-strong convergence and the probability of using backstop method are in Section 3.3 with their proofs in Sections 3.3.3 and 3.3.4, respectively.

3.1 Assumptions

We now present our assumptions on f and g_i in (1.1).

Assumption 3.1.1. *Let $f \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ and $g \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^{d\times m})$ with $g_i(x) = [g_{1,i}(x), \ldots, g_{d,i}(x)]^T \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$. For each $\varkappa \geq 1$ there exist $L_\varkappa > 0$ such that*

$$\left\|f(x) - f(y)\right\|^2 + \left\|g(x) - g(y)\right\|^2_{\mathbf{F}(d\times m)} \leq L_\varkappa \left\|x - y\right\|^2, \tag{3.1}$$

for $x, y \in \mathbb{R}^d$ with $\|x\| \vee \|y\| \leq \varkappa$, and there exists $L \geq 0$ such that for some $\eta \geq 2$

$$\left\langle x - y, f(x) - f(y)\right\rangle + \frac{\eta - 1}{2}\left\|g(x) - g(y)\right\|^2_{\mathbf{F}(d\times m)} \leq L\left\|x - y\right\|^2. \tag{3.2}$$

In addition, for some constants $c_{3,4,5,6}$, q_1, $q_2 \geq 0$; $i = 1, \ldots, m$, we have

$$\left\|\mathbf{D}f(x)\right\|_{\mathbf{F}} \leq c_3(1 + \|x\|^{q_1+1}), \qquad \left\|\mathbf{D}g_i(x)\right\|_{\mathbf{F}} \leq c_4(1 + \|x\|^{q_2+1}), \tag{3.3}$$

$$\left\|f(x)\right\| \leq c_5(1 + \|x\|^{q_1+2}), \qquad \left\|g(x)\right\|_{\mathbf{F}(d\times m)} \leq c_6(1 + \|x\|^{q_2+2}). \tag{3.4}$$

Furthermore, for some $c_{1,2} \geq 0$; $i = 1, \ldots, m$, *we have*

$$\left\|\mathbf{D}^2 f(x)\right\|_{\mathbf{T}_3} \leq c_1(1 + \|x\|^{q_1}), \quad \left\|\mathbf{D}^2 g_i(x)\right\|_{\mathbf{T}_3} \leq c_2(1 + \|x\|^{q_2}). \tag{3.5}$$

Under (3.1) and (3.2), the SDE (1.1) has a unique strong solution on any interval $[0, T]$, where $T < \infty$ on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, see [21], [40] and [55].

Assumption 3.1.2. *Suppose that* (3.2) *in Assumption 3.1.1 holds with*

$$\eta \geq 4q + 2q_2 + 10,$$

where $q := q_1 \vee q_2$, q_1 *and* q_2 *are from* (3.4) *in Assumption 3.1.1.*

We now give the following Lemma on moments of the solution.

Lemma 3.1.1. *[41, Lem. 4.2] Let* f *and* g *satisfy* (3.1) *and* (3.2), *and suppose that Assumption 3.1.2 holds. If* g *further satisfies* (3.4), *then there is a constant* $C_{\mathcal{X}} > 0$ *such that the solution of* (1.1) *satisfies*

$$\mathbb{E}\left[\sup_{s\in[0,T]} \|X(s)\|^{\eta-2q_2-2}\right] \leq C_{\mathcal{X}}. \tag{3.6}$$

3.2 Adaptive time-stepping

To deal with the extra terms that arise from Milstein as in Definition 2.3.1 over Euler-Maruyama type discretisations as in Definition 2.20, we introduce a new class of time-stepping strategies in Definition 3.2.4.

Let $\{h_{n+1}\}_{n\in\mathbb{N}}$ be a sequence of strictly positive random timesteps with corresponding random times $\{t_n := \sum_{i=1}^n h_i\}_{n\in\mathbb{N}\setminus\{0\}}$, where $t_0 = 0$.

Assumption 3.2.1. *For the sequence of random timesteps* $\{h_{n+1}\}_{n\in\mathbb{N}}$, *there are constant values* $h_{\max} > h_{\min} > 0$, $\rho > 1$ *such that* $h_{\max} = \rho h_{\min}$, *and*

$$0 < h_{\min} \leq h_{n+1} \leq h_{\max} \leq 1. \tag{3.7}$$

In addition, we assume each h_{n+1} *is* $\mathcal{F}_{t_n}$*-measurable.*

Let $\{\mathcal{F}_t\}$ stand for the natural filtration of W, and if μ is any $\{\mathcal{F}_t\}$-stopping time, then $\mathcal{F}_\mu$ is defined in (2.11) from Definition 2.1.1 (see [42, p. 14]). The following lemma allows us to condition on $\mathcal{F}_{t_n}$ at any point on the random time-set $\{t_n\}_{n\in\mathbb{N}}$.

Lemma 3.2.1. *Let $\{h_{n+1}\}_{n\in\mathbb{N}}$ satisfy Assumption 3.2.1. Each member of $\{t_n = \sum_{i=1}^n h_i\}_{n\in\mathbb{N}\backslash\{0\}}$ is an $\{\mathcal{F}_t\}$-stopping time: i.e. $\{t_n \le t\} \in \mathcal{F}_t$ for all $t \in [0,T]$.*

Proof. We show this by induction for $n \in \mathbb{N}$.

$n = 1$: By Assumption 3.2.1, h_1 is an $\mathcal{F}_{t_0}$-measurable random variable with $t_0 = 0$, that is $\{h_1 \le t\} \in \mathcal{F}_0$ for all $t \in [0,T]$. Since $\mathcal{F}_0$ is a sub-σ algebra of $\mathcal{F}_t$, we have that $\{h_1 \le t\} \in \mathcal{F}_t$ for all $t \in [0,T]$. So $\{t_1 \le t\} = \{h_1 \le t\} \in \mathcal{F}_t$, for all $t \in [0,T]$, and we conclude that t_1 is an $\{\mathcal{F}_t\}$-stopping time.

$n = k$: With the induction hypothesis being that t_k is an $\{\mathcal{F}_t\}$-stopping time, that is $\{t_k \le t\} \in \mathcal{F}_t$ for all $t \in [0,T]$, by (2.11) in Definition 2.1.1 we define the σ-algebra at the stopping time t_k as $\mathcal{F}_{t_k} := \{A \in \mathcal{F} : A \cap \{t_k \le t\} \in \mathcal{F}_t$, for all $t \in [0,T]\}$. Since $t_n = \sum_{i=1}^n h_i$ for $n \in \mathbb{N}\backslash\{0\}$, we have

$$\{t_{k+1} \le t\} = \{h_{k+1} + t_k \le t\} = \bigcup_{r\in[0,t]} \Big(\{h_{k+1} \le t - r\} \cap \{t_k \le r\}\Big).$$

Since h_{k+1} is an $\mathcal{F}_{t_k}$-measurable random variable by Assumption 3.2.1, and by setting $A = \{h_{k+1} \le t - r\}$ and by the definition of $\mathcal{F}_{t_n}$ we have $A \cap \{t_k \le r\} \in \mathcal{F}_r \subseteq \mathcal{F}_t$ (sub-σ algebra) for all $r \in [0,T]$. Therefore, we have $\{t_{k+1} \le t\} \in \mathcal{F}_t$ for all $t \in [0,T]$ and conclude that t_{k+1} is an $\{\mathcal{F}_t\}$-stopping time. The proof is complete.

□

Definition 3.2.1. Let $N^{(t)}$ be a random integer such that

$$N^{(t)} := \max\{n \in \mathbb{N}\backslash\{0\} : t_{n-1} < t\}, \tag{3.8}$$

and let $N = N^{(T)}$ and $t_N = T$, so that T is always the last point on the mesh. Note that $N^{(t)}$ indicates the step number such that $t \in \big[t_{N^{(t)}-1}, t_{N^{(t)}}\big]$. Furthermore, by Assumption 3.2.1, $N^{(t)}$ only takes values in the finite set $\{N^{(t)}_{\min}, \ldots, N^{(t)}_{\max}\}$, where $N^{(t)}_{\min} := \lfloor t/h_{\max} \rfloor$ and $N^{(t)}_{\max} := \lceil t/h_{\min} \rceil$.

In Assumption 3.2.1, the lower bound $h_{\min}$ given by (3.7) ensures that a simulation over the interval $[0,T]$ can be completed in a finite number of time steps. In the event that at time t_n our strategy attempts to select a stepsize $h_{n+1} \le h_{\min}$, we instead apply a single step of a backstop method (φ in Definition 3.2.2 below), a known method that satisfies a mean-square consistency requirement as in (3.11) with deterministic step $h_{n+1} = h_{\min}$ (see also discussion in Remarks 3).

First we recall the Milstein method expressed as a map. Over each step $[t_n, t_{n+1}]$ the Milstein map $\theta : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^d$ is defined as for $s \in [t_n, t_{n+1}]$

$$\theta\big(x, t_n, s - t_n\big) := x + (s - t_n) f(x) + \sum_{i=1}^{m} g_i(x) I_i^{t_n,s} + \sum_{i,j=1}^{m} \mathbf{D} g_i(x) g_j(x) I_{j,i}^{t_n,s}. \quad (3.9)$$

Following [31, Def. 9], we define an adaptive Milstein scheme combining the Milstein method and a backstop method as follows.

Definition 3.2.2 (Adaptive Milstein Scheme)**.** Let $\{h_{n+1}\}_{n\in\mathbb{N}}$ satisfy Assumption 3.2.1. Using indicator functions to distinguish the backstop case when $h_{n+1} = h_{\min}$ (and allowing for the possibility that the final step taken to terminal time T is smaller than $h_{\min}$, in which case the backstop method is also used), we define the continuous form of an *adaptive Milstein scheme* associated with a particular time-stepping strategy $\{h_{n+1}\}_{n\in\mathbb{N}}$

$$\begin{aligned}\widetilde{Y}(s) := \theta\left(\widetilde{Y}(t_n),\, t_n,\, s - t_n\right) \cdot \mathbf{1}_{\{h_{\min} < h_{n+1} \le h_{\max}\}} \\ + \varphi\left(\widetilde{Y}(t_n),\, t_n,\, s - t_n\right) \cdot \mathbf{1}_{\{h_{n+1} \le h_{\min}\}}, \quad (3.10)\end{aligned}$$

for $s \in [t_n, t_{n+1}]$, $n \in \mathbb{N}$, and where $\widetilde{Y}(0) = X(0)$. Thus the scheme is characterised by the sequence of tuples, $\big\{\big(\widetilde{Y}(s)\big)_{s\in[t_n,t_{n+1}]}, h_{n+1}\big\}_{n\in\mathbb{N}}$. The backstop map $\varphi : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times \Omega \to \mathbb{R}^d$ in (3.10) satisfies for $n \in \mathbb{N}$

$$\begin{aligned}\mathbb{E}\left[\left\|X(s) - \varphi\left(\widetilde{Y}(t_n),\, t_n,\, s - t_n\right)\right\|^2 \middle| \mathcal{F}_{t_n}\right] \le \left\|X(t_n) - \widetilde{Y}(t_n)\right\|^2 \\ + C_{B_1} \int_{t_n}^{s} \mathbb{E}\left[\left\|X(r) - \varphi\left(\widetilde{Y}(t_n),\, t_n,\, r - t_n\right)\right\|^2 \middle| \mathcal{F}_{t_n}\right] dr + C_{B_2} h_{\min}^3, \quad (3.11)\end{aligned}$$

a.s, for positive constants C_{B_1} and C_{B_2}.

Throughout the thesis it is notationally convenient to make the following definition based on (3.10).

Definition 3.2.3. Let $\widetilde{Y}$ be as given in Definition 3.2.2 and define for each $n \in \mathbb{N}$

$$Y_\theta(s) := \theta\Big(\widetilde{Y}(t_n), t_n, s - t_n\Big), \quad s \in [t_n, t_{n+1}]. \quad (3.12)$$

Remark 3 (**$h_{\max}$ and $h_{\min}$**)**.** The upper bound $h_{\max}$ prevents stepsizes from becoming too large and allows us to examine strong convergence of the adaptive Milstein method (3.10) to solutions of (1.1) as $h_{\max} \to 0$ (and hence as $h_{\min} \to 0$). Note that φ satisfies (3.11) if the backstop method satisfies a mean-square consistency requirement. In practice, instead of testing (3.11), we choose a backstop method that is strongly convergent with rate 1.

Remark 4 (**Wiener increments**)**.** For all $i, j = 1, 2, \ldots, m$, $I_i^{t_n,t_{n+1}}$ in (2.19) is a Wiener increment taken over a random step of length h_{n+1} , which itself may depend on $\widetilde{Y}(t_n)$ and therefore is not necessarily independent and normally distributed. However, since h_{n+1} is $\mathcal{F}_{t_n}$-measurable, then $I_i^{t_n,t_{n+1}}$ is $\mathcal{F}_{t_n}$-conditionally normally distributed and by the expected stochastic integral with stopping time in (2.12) and (2.13), together with the Optional Sampling Theorem (see for example [50]), for all $p = 0, 1, 2, \ldots$

$$\mathbb{E}\left[I_i^{t_n,t_{n+1}}\Big|\mathcal{F}_{t_n}\right] = 0, \quad a.s.; \tag{3.13}$$

$$\mathbb{E}\left[\left|I_i^{t_n,t_{n+1}}\right|^2\Big|\mathcal{F}_{t_n}\right] = h_{n+1}, \quad a.s.; \tag{3.14}$$

$$\mathbb{E}\left[\left|I_i^{t_n,s}\right|^p\Big|\mathcal{F}_{t_n}\right] = \Upsilon_p|s - t_n|^{\frac{p}{2}}, \quad a.s.; \tag{3.15}$$

where $\Upsilon_p := 2^{p/2}\Gamma\left((p+1)/2\right)\pi^{-1/2}$, and Γ is the Gamma function (see for example [44, p.148]). In implementation, it is sufficient to replace the sequence of Wiener increments with i.i.d. $\mathcal{N}(0,1)$ random variables scaled at each step by the $\mathcal{F}_{t_n}$-measurable random variable $\sqrt{h_{n+1}}$.

We now provide a specific example of a time-stepping strategy that we use in Section 5.1 and that satisfies the assumptions for our convergence proof in Theorem 3.3.1. Suppose that for each $n = 0, \ldots, N-1$ and some fixed constant $\kappa > 0$, we choose constant values $h_{\max} > h_{\min} > 0$, $\rho > 1$ such that $h_{\max} = \rho h_{\min}$ and

$$h_{n+1} = h_{\min} \vee \left(\frac{h_{\max}}{\left\|\widetilde{Y}(t_n)\right\|^{1/\kappa}} \wedge h_{\max}\right). \tag{3.16}$$

Then (3.7) in Assumption 3.2.1 holds for (3.16). Given (3.16), when on the event $\{h_{\min} < h_{n+1} \leq h_{\max}\}$, two situations exist: **Case 1**. $\left(h_{n+1} = h_{\max}/\left\|\widetilde{Y}(t_n)\right\|^{1/\kappa}\right)$, that is

$$h_{\min} < \frac{h_{\max}}{\left\|\widetilde{Y}(t_n)\right\|^{1/\kappa}} \leq h_{\max}.$$

Since $\rho h_{\min} = h_{\max}$ we have

$$1 < \left(\frac{h_{\max}}{h_{\max}}\right)^{\kappa} \leq \left\|\widetilde{Y}(t_n)\right\| < \left(\frac{h_{\max}}{h_{\min}}\right)^{\kappa} = \rho^{\kappa}.$$

Case 2. $\Big(h_{n+1} = h_{\max}\Big)$, we have

$$\frac{h_{\max}}{\left\|\widetilde{Y}(t_n)\right\|^{1/\kappa}} > h_{\max}, \qquad \Longrightarrow \qquad \left\|\widetilde{Y}(t_n)\right\| < \left(\frac{h_{\max}}{h_{\max}}\right)^{\kappa} = 1.$$

The strategy given by (3.16) is admissible in the sense given in [30, 31]. However, it also motivates the following class of time-stepping strategies to which our convergence analysis applies.

Definition 3.2.4 (Path-bounded time-stepping strategies)**.** Let $\left\{\widetilde{Y}(t_n), h_{n+1}\right\}_{n\in\mathbb{N}}$ be a numerical approximation for (1.1) given by (3.10), associated with a timestep sequence $\{h_{n+1}\}_{n\in\mathbb{N}}$ satisfying Assumption 3.2.1. We say that $\{h_{n+1}\}_{n\in\mathbb{N}}$ is a path-bounded time-stepping strategy for (3.10) if there exist real non-negative constants $0 \leq Q < R$ (where R is independent of N and may be infinite if $Q \neq 0$) such that whenever $h_{\min} < h_{n+1} \leq h_{\max}$,

$$Q \leq \left\|\widetilde{Y}(t_n)\right\| < R, \quad n = 0, \ldots, N-1. \tag{3.17}$$

Note that throughout this paper we use a strategy where $Q = 0$ and $R < \infty$. As we will see in Section 5.1.2, a careful choice of the parameter κ can be used to minimise invocations of the backstop method when ρ is fixed.

3.3 Results: Theorem 3.3.1 and Theorem 3.3.2

Our first main result shows strong convergence with order 1 of solutions of (3.10) to solutions of (1.1) when $\{h_{n+1}\}_{n\in\mathbb{N}}$ is a path-bounded time-stepping strategy ensuring that (3.17) holds.

Theorem 3.3.1 (Strong Convergence). *Let $(X(t))_{t\in[0,T]}$ be a solution of* (1.1) *with initial value $X(0) = X_0 \in \mathbb{R}^d$. Suppose that the conditions of Assumptions 3.1.1 and 3.1.2 hold. Let $\left\{\left(\widetilde{Y}(s)\right)_{s\in[t_n,t_{n+1}]}, h_{n+1}\right\}_{n\in\mathbb{N}}$ be the adaptive Milstein scheme given in Definition 3.2.2 with initial value for the first component $\widetilde{Y}_0 = X_0$, and let the path-bounded time-stepping strategy $\{h_{n+1}\}_{n\in\mathbb{N}}$ satisfy the conditions of Definition 3.2.4 for some $R < \infty$. Then there exists a constant $C(R, \rho, T) > 0$ independent of $h_{\max}$ and with the explicit form in* (3.89) *such that*

$$\max_{t\in[0,T]}\left(\mathbb{E}\left[\|X(t)-\widetilde{Y}(t)\|^2\right]\right)^{1/2} \leq C(R,\rho,T)\, h_{\max}. \tag{3.18}$$

Furthermore,

$$\lim_{\rho\to\infty} C(R,\rho,T) = \infty. \tag{3.19}$$

Notice that (3.19) shows the necessity of the lower bound $h_{\min}$ bounded away from 0. The proof of Theorem 3.3.1, which is given in Section 3.3.3, accounts for the properties of the random sequences $\{t_n\}_{n\in\mathbb{N}}$ and $\{h_{n+1}\}_{n\in\mathbb{N}}$ and uses (3.17) to compensate for the non-Lipschitz drift and diffusion.

Our second main result shows that for the specific strategy given by (3.16) the probability of needing a backstop method can be made arbitrarily small by taking ρ sufficiently large for fixed κ, where ρ represents the distance between $h_{\max}$ and $h_{\min}$ by Assumption 3.2.1, and κ prevents the strategy from always taking $h_{\min}$ when initial value is large see Section 5.1.2.

Theorem 3.3.2 (Probability of Backstop). *Let all the conditions of Theorem 3.3.1 hold, and suppose that the path-bounded time-stepping strategy $\{h_{n+1}\}_{n\in\mathbb{N}}$ also satisfies* (3.16) *with $\delta = h_{\max}$.*

Let $C(R, \rho, T)$ be the error constant in estimate (3.18) *from the statement of Theorem 3.3.1.*

For any fixed $\kappa \geq 1$ there exists a constant $C_{prob} = C_{prob}(T, R, h_{\max})$ with its explicit form in (3.94) *such that, for $h_{\max} \leq 1/C(R, \rho, T)$,*

$$\mathbb{P}\left[h_{n+1} = h_{\min}\right] \leq C_{prob}\; \rho^{1-2\kappa}. \tag{3.20}$$

Further for arbitrarily small tolerance $\varepsilon \in (0, 1)$, there exists $\rho > 0$ such that

$$\mathbb{P}\left[h_{n+1} = h_{\min}\right] < \varepsilon, \quad n \in \mathbb{N}.$$

The upper bound of $h_{\max}$ is a technical constraint required for the proof, see Section 3.3.4.

3.3.1 Preliminary Lemmas

We present five lemmas necessary for the proof of Theorem 3.3.1 and Theorem 3.3.2. Throughout this section we assume that coefficients f and g satisfy Assumptions 3.1.1 and that we are on the event $\{h_{\min} < h_{n+1} \leq h_{\max}\}$ (except for Lemma 3.3.6) so that (3.17) holds of Definition 3.2.4. We use (3.3), (3.5) and (3.4) to define some bounded constant coefficients depending on $R < \infty$. The constant bounds in (3.21) are then used in the development of the one-step error bound for the adaptive part of the scheme.

$$\begin{aligned}
\left\|f\big(\widetilde{Y}(t_n)\big)\right\| &\leq c_5(1+R^{q_1+2}) =: C_f;\\
\left\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\right\|_{\mathbf{F}} &\leq c_3(1+R^{q_1+1}) =: C_{Df};\\
\left\|g_i\big(\widetilde{Y}(t_n)\big)\right\| \leq \left\|g\big(\widetilde{Y}(t_n)\big)\right\|_{\mathbf{F}(d\times m)} &\leq c_6(1+R^{q_2+2}) =: C_{g_i};\\
\left\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\right\|_{\mathbf{F}} &\leq c_4(1+R^{q_2+1}) =: C_{Dg_i}.
\end{aligned} \tag{3.21}$$

The following lemma provides a bound for the even conditional moments of the iterated stochastic integral in (2.19).

Lemma 3.3.3 (`Iterated Stochastic Integral`). *Let $\left\{\big(\widetilde{Y}(s)\big)_{s\in[t_n,t_{n+1}]}, h_{n+1}\right\}_{n\in\mathbb{N}}$ be the adaptive Milstein scheme given in Definitions 3.2.2 and 3.2.4. Then there exists a constant C_{ISI} such that for $k \geq 1$, $n \in \mathbb{N}$ and $s \in [t_n, t_{n+1}]$, on the event $\{h_{\min} < h_{n+1} \leq h_{\max}\}$,*

$$\mathbb{E}\left[\left\|\sum_{i,j=1}^{m} \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{j,i}^{t_n,s}\right\|^{2k}\middle|\mathcal{F}_{t_n}\right] \leq C_{\mathit{ISI}}(k,R)\,|s-t_n|^{2k}, \tag{3.22}$$

where

$$C_{\mathit{ISI}}(k,R) := 3^{2k}m^{4k}C_{Dg_i}^{2k}C_{g_i}^{2k}\Big(\Upsilon_{4k} + 1 + \Upsilon_{2k}^2 + C_{\mathit{LA}}(2k)\Big). \tag{3.23}$$

Here, Υ_p is from (3.15), $C_{\mathit{LA}}(2k)$ is from Lemma 2.3.1 with explicit form given in (2.25), and the R dependence in $C_{\mathit{ISI}}(k,R)$ arises from (3.21).

Proof. First of all, for convenience we set

$$G_{\mathtt{ISI}}(s) := \left\| \sum_{i,j=1}^{m} \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{i,j}^{t_n,s} \right\|^{2k}.$$

By (2.22) and (2.34), we have, for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$,

$$\begin{aligned} G_{\mathtt{ISI}}(s) \le 3^{2k-1} \Bigg(&\left\| \frac{1}{2}\sum_{i=1}^{m} \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big(\big(I_i^{t_n,s}\big)^2 - |s-t_n|\Big) \right\|^{2k} \\ &+ \left\| \frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^{m} \Big(\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big) + \mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big) I_i^{t_n,s}I_j^{t_n,s} \right\|^{2k} \\ &+ \left\| \sum_{\substack{i,j=1\\ i<j}}^{m} \Big(\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big) - \mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big) A_{ij}^{t_n,s} \right\|^{2k} \Bigg). \end{aligned}$$

Applying (2.34) again and by submultiplicativity of the Euclidean norm and the fact that the induced matrix 2-norm is bounded above by the Frobenius norm, for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$, we get

$$\begin{aligned} &G_{\mathtt{ISI}}(s) \\ &\le 3^{2k-1} \Bigg(\frac{m^{2k-1}}{2^{2k}} \sum_{i=1}^{m} \big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}^{2k} \big\|g_i\big(\widetilde{Y}(t_n)\big)\big\|^{2k} \Big(\big(I_i^{t_n,s}\big)^2 + |s-t_n|\Big)^{2k} \\ &+ \left(\frac{m(m-1)}{2}\right)^{2k-1} \sum_{\substack{i,j=1\\ i<j}}^{m} \Big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big) + \mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i(Y(t_n)\Big\|^{2k} \\ &\qquad \times \left(\frac{1}{2^{2k}} \Big| I_i^{t_n,s} I_j^{t_n,s} \Big|^{2k} + \big|A_{ij}^{t_n,s}\big|^{2k} \right) \Bigg). \end{aligned}$$

Applying conditional expectations on both sides, together with the pairwise conditional independence of $I_i^{t_n,s}$ and $I_j^{t_n,s}$ for $i \neq j$, (3.3) and (3.21), we have for

$s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$

$$\mathbb{E}\Big[G_{\mathtt{ISI}}(s)\Big|\mathcal{F}_{t_n}\Big] \leq 3^{2k}\Bigg(m^{2k-1}C_{Dg_i}^{2k}C_{g_i}^{2k}\sum_{i=1}^{m}\left(\mathbb{E}\left[\left|I_i^{t_n,s}\right|^{4k}\Big|\mathcal{F}_{t_n}\right] + |s-t_n|^{2k}\right)$$
$$+\left(\frac{m(m-1)}{2}\right)^{2k-1}C_{Dg_i}^{2k}C_{g_i}^{2k}\sum_{\substack{i,j=1\\i<j}}^{m}\left(\mathbb{E}\left[\left|I_i^{t_n,s}\right|^{2k}\Big|\mathcal{F}_{t_n}\right]\mathbb{E}\left[\left|I_j^{t_n,s}\right|^{2k}\Big|\mathcal{F}_{t_n}\right]\right.$$
$$\left.+\,\mathbb{E}\left[\left|A_{ij}^{t_n,s}\right|^{2k}\Big|\mathcal{F}_{t_n}\right]\right)\Bigg).$$

Using (3.14), (3.15) and (2.24) we have

$$\mathbb{E}\Big[G_{\mathtt{ISI}}(s)\Big|\mathcal{F}_{t_n}\Big] \leq C_{\mathtt{ISI}}\left(k,R\right)|s-t_n|^{2k}, \quad a.s.$$

where $C_{\mathtt{ISI}}\left(k,R\right)$ is in (3.23). □

The following lemma provides a bound on the conditional moments of the adaptive Milstein scheme in (3.10) over one step, in the case where the method applies the map θ.

Lemma 3.3.4. *Consider* $\left\{\left(\widetilde{Y}(s)\right)_{s\in[t_n,t_{n+1}]}, h_{n+1}\right\}_{n\in\mathbb{N}}$ *from Definitions 3.2.2 and 3.2.4, and let* $(Y_\theta(s))_{s\in(t_n,t_{n+1}]}$ *be as defined in Definition* (3.2.3)*. Then there exists a constant* $C_{Y_\theta} > 0$ *such that for* $k \geq 1$*,* $n \in \mathbb{N}$ *and* $s \in (t_n, t_{n+1}]$*, on the event* $\{h_{\min} < h_{n+1} \leq h_{\max}\}$*,*

$$\mathbb{E}\Big[\big\|Y_\theta(s)\big\|^k\Big|\mathcal{F}_{t_n}\Big] \leq C_{Y_\theta}\big(k,R\big), \tag{3.24}$$

where

$$C_{Y_\theta}\big(k,R\big) := \; 4^{k-1}\Big(R^k + C_f^k + m^k C_{g_i}^k\;\Upsilon_k + C_{ISI}\left(2k\right)^{1/2}\Big), \tag{3.25}$$

with the constant C_{ISI} *from Lemma 3.3.3.*

Proof. By (3.12), (3.9) and (2.34), we have, for $s \in (t_n, t_{n+1}]$ and $n \in \mathbb{N}$,

$$\begin{aligned}\left\|Y_\theta(s)\right\|^k &= \left\|\theta\left(\widetilde{Y}(t_n),\, t_n,\, s-t_n,\, \omega\right)\right\|^k \\ &\le 4^{k-1}\Bigg(\left\|\widetilde{Y}(t_n)\right\|^k + \left\|f\big(\widetilde{Y}(t_n)\big)\right\|^k |s-t_n|^k + \left\|\sum_{i=1}^m g_i\big(\widetilde{Y}(t_n)\big) I_i^{t_n,s}\right\|^k \\ &\quad + \left(\left\|\sum_{i,j=1}^m \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big) g_j\big(\widetilde{Y}(t_n)\big) I_{j,i}^{t_n,s}\right\|^{2k}\right)^{1/2}\Bigg).\end{aligned}$$

Applying (2.34), (3.17) and (3.21) for $s \in (t_n, t_{n+1}]$ and $n \in \mathbb{N}$, it yields

$$\begin{aligned}\left\|Y_\theta(s)\right\|^k \le 4^{k-1}\Bigg(& R^k + C_f^k |s-t_n|^k + m^{k-1} C_{g_i}^k \sum_{i=1}^m \left|I_i^{t_n,s}\right|^k \\ & + \left(\left\|\sum_{i,j=1}^m \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big) g_j\big(\widetilde{Y}(t_n)\big) I_{j,i}^{t_n,s}\right\|^{2k}\right)^{1/2}\Bigg).\end{aligned}$$

Taking conditional expectation on both sides, with Jensen's inequality on the last term we have for $s \in (t_n, t_{n+1}]$ and $n \in \mathbb{N}$

$$\begin{aligned}\mathbb{E}\Big[\|Y_\theta(s)\|^k \Big| \mathcal{F}_{t_n}\Big] \le\; 4^{k-1}\Bigg(& R^k + C_f^k |s-t_n|^k + m^{k-1} C_{g_i}^k \sum_{i=1}^m \mathbb{E}\Big[\left|I_i^{t_n,s}\right|^k \Big| \mathcal{F}_{t_n}\Big] \\ & + \left(\mathbb{E}\left[\left\|\sum_{i,j=1}^m \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big) g_j\big(\widetilde{Y}(t_n)\big) I_{j,i}^{t_n,s}\right\|^{2k} \middle| \mathcal{F}_{t_n}\right]\right)^{1/2}\Bigg).\end{aligned}$$

Using (3.14), (3.22) from Lemma 3.3.3 and since $|s-t_n| \le h_{\max} \le 1$ (3.7) we have

$$\mathbb{E}\Big[\|Y_\theta(s)\|^k \Big| \mathcal{F}_{t_n}\Big] \le C_{Y_\theta}(k, R), \quad a.s.$$

where $C_{Y_\theta}(k, R)$ is in (3.25). □

The following lemma proves regularity in time of the adaptive Milstein scheme in (3.10) when applying the map θ.

Lemma 3.3.5 (`Scheme Regularity`). *Consider* $\left\{\big(\widetilde{Y}(s)\big)_{s\in[t_n,t_{n+1}]}, h_{n+1}\right\}_{n\in\mathbb{N}}$ *in Definitions 3.2.2 and 3.2.4, and let* $(Y_\theta(s))_{s\in(t_n,t_{n+1}]}$ *be as defined in Definition* (3.2.3). *Then there exists a constant* C_{SR} *such that for* $k \ge 1$, $n \in \mathbb{N}$ *and* $s \in (t_n, t_{n+1}]$, *on the event* $\{h_{\min} < h_{n+1} \le h_{\max}\}$,

$$\mathbb{E}\Big[\left\|Y_\theta(s) - \widetilde{Y}(t_n)\right\|^{2k} \Big| \mathcal{F}_{t_n}\Big] \le C_{\mathit{SR}}(k, R)\, |s-t_n|^k, \tag{3.26}$$

where

$$C_{SR}(k,R) := \; 3^{2k-1}\Big(C_f^{2k} + m^{2k}C_{g_i}^{2k}\;\Upsilon_{2k} + C_{ISI}(2k)\Big), \tag{3.27}$$

with the constant C_{ISI} from Lemma 3.3.3.

The method of proof is similar to the proof of Lemma 3.3.4.

Proof. By (3.12), (3.9) and (2.34), we have, for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$,

$$\begin{aligned}\big\|Y_\theta(s) - \widetilde{Y}(t_n)\big\|^{2k} &= \Big\|\theta\left(\widetilde{Y}(t_n),\, t_n,\, s-t_n,\, \omega\right) - \widetilde{Y}(t_n)\Big\|^{2k}\\ &\le \; 3^{2k-1}\Bigg(\,\Big\|f\big(\widetilde{Y}(t_n)\big)\Big\|^{2k}|s-t_n|^{2k} + \Bigg\|\sum_{i=1}^{m} g_i\big(\widetilde{Y}(t_n)\big)I_i^{t_n,s}\Bigg\|^{2k}\\ &\qquad + \Bigg\|\sum_{i,j=1}^{m}\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{j,i}^{t_n,s}\Bigg\|^{2k}\Bigg).\end{aligned}$$

Applying (2.34), (3.17) and (3.21) for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$, it yields

$$\begin{aligned}\big\|Y_\theta(s) - \widetilde{Y}(t_n)\big\|^{2k} &\le \; 3^{2k-1}\Bigg(C_f^{2k}|s-t_n|^{2k} + m^{2k-1}C_{g_i}^{2k}\sum_{i=1}^{m}\big|I_i^{t_n,s}\big|^{2k}\\ &\qquad + \Bigg\|\sum_{i,j=1}^{m}\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{j,i}^{t_n,s}\Bigg\|^{2k}\Bigg).\end{aligned}$$

Taking conditional expectation on both sides we have

$$\begin{aligned}&\mathbb{E}\Big[\big\|Y_\theta(s) - \widetilde{Y}(t_n)\big\|^{2k}\Big|\mathcal{F}_{t_n}\Big]\\ &\qquad \le \; 3^{2k-1}\Bigg(C_f^{2k}|s-t_n|^{2k} + m^{2k-1}C_{g_i}^{2k}\sum_{i=1}^{m}\mathbb{E}\left[\big|I_i^{t_n,s}\big|^{2k}\Big|\mathcal{F}_{t_n}\right]\\ &\qquad\qquad + \mathbb{E}\Bigg[\Bigg\|\sum_{i,j=1}^{m}\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{j,i}^{t_n,s}\Bigg\|^{2k}\Bigg|\mathcal{F}_{t_n}\Bigg]\Bigg),\end{aligned}$$

for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$. Using (3.15) and (3.22) from Lemma 3.3.3 we have

$$\begin{aligned}\mathbb{E}\Big[\big\|Y_\theta(s) - \widetilde{Y}(t_n)\big\|^{2k}\Big|\mathcal{F}_{t_n}\Big] &\le 3^{2k-1}\Big(C_f^{2k}|s-t_n|^{2k} + m^{2k}C_{g_i}^{2k}\;\Upsilon_{2k}|s-t_n|^{k}\\ &\qquad + C_{\mathtt{ISI}}(2k)\,|s-t_n|^{2k}\Big), \quad a.s.\end{aligned}$$

Since $|s - t_n| \leq h_{\max} \leq 1$ (3.7), we have

$$\mathbb{E}\left[\left\|Y_\theta(s) - \widetilde{Y}(t_n)\right\|^{2k}\middle|\mathcal{F}_{t_n}\right] \leq C_{\mathtt{SR}}(k,R)\ |s-t_n|^k,$$

where the explicit form of $C_{\mathtt{SR}}(k,R)$ is in (3.27). □

Remark 5 (**Superscript**). Our analysis requires a certain number of finite moments for the SDE (1.1), and it is necessary to track exactly what those are in order to see that the conditions of Assumption 3.1.2 are not violated. To this end, we introduce a superscript notation for random variables appearing as conditional expectations at this point. The notation should be interpreted according to the following example: in (3.29) below the random variable $C_{\mathtt{PR}}^{\{2k(q+2)\}}$ requires $2k(q+2)$ finite moments of the SDE (1.1) to have finite expectation.

The following lemma examines the regularity of solutions of the SDE (1.1).

Lemma 3.3.6 (`Path Regularity`). *Let f, g also satisfy Assumption 3.1.2, and let $(X(s))_{s\in[t_n,t_{n+1}]}$ be a solution of (1.1). Then there exists an $\mathcal{F}_{t_n}$-measurable random variable $\overline{C}_{\mathtt{PR}}^{\{2k(q+2)\}}$ such that for $k \geq 1$, $n \in \mathbb{N}$ and $s \in [t_n, t_{n+1}]$ a.s.*

$$\mathbb{E}\left[\|X(s) - X(t_n)\|^{2k}\middle|\mathcal{F}_{t_n}\right] \leq \overline{C}_{\mathtt{PR}}^{\{2k(q+2)\}}\ |s-t_n|^k, \tag{3.28}$$

where $q = q_1 \vee q_2$ is as defined in Assumption 3.1.2. Where a.s.

$$\begin{aligned}\overline{C}_{\mathtt{PR}}^{\{2k(q+2)\}} &= 2^{4k-2}c_5^{2k}\left(1 + \mathbb{E}\left[\sup_{p\in[t_n,t_{n+1}]}\|X(p)\|^{2k(q_1+2)}\middle|\mathcal{F}_{t_n}\right]\right)\\ &\quad + 2^{4k-2}(k(2k-1))^k c_6^{2k}\left(1 + \mathbb{E}\left[\sup_{p\in[t_n,t_{n+1}]}\|X(p)\|^{2k(q_2+2)}\middle|\mathcal{F}_{t_n}\right]\right).\end{aligned} \tag{3.29}$$

where the expectation of $\overline{C}_{\mathtt{PR}}^{\{2k(q+2)\}}$ is denoted $C_{\mathtt{PR}}(k)$, given by

$$C_{\mathtt{PR}}(k) := \mathbb{E}\left[\overline{C}_{\mathtt{PR}}^{\{2k(q+2)\}}\right] \leq 2^{4k-2}\left(1 + C_{\mathtt{X}}\right)\left(c_5^{2k} + (k(2k-1))^k c_6^{2k}\right). \tag{3.30}$$

Proof. The method of proof follows that of [40, Thm. 7.1]. The bound (3.30) follows from (3.6) and Assumption 3.1.2. □

The following lemma provides a bound on the even conditional moments of the remainder term from a Taylor's expansion (see Definition 2.0.1) of either the drift f or diffusion g, around $\widetilde{Y}(t_n)$.

Lemma 3.3.7 (`Taylor Error`). *Consider* $\left\{\left(\widetilde{Y}(s)\right)_{s\in[t_n,t_{n+1}]}, h_{n+1}\right\}_{n\in\mathbb{N}}$ *from Definitions 3.2.2 and 3.2.4, and let* $(Y_\theta(s))_{s\in(t_n,t_{n+1}]}$ *be as defined in Definition (3.2.3). Let* $u\in\{f,g\}$ *and set* $c_{\mathbf{D2}} := c_1 \vee c_2$*. Then there exists a constant* C_{TE} *such that for* $k\geq 1$, $n\in\mathbb{N}$ *and* $s\in(t_n,t_{n+1}]$, *on the event* $\{h_{\min} < h_{n+1} \leq h_{\max}\}$,

$$\mathbb{E}\left[\left\|\int_0^1 (1-\epsilon)\mathbf{D}^2 u\Big(\widetilde{Y}(t_n) - \epsilon\big(Y_\theta(s) - \widetilde{Y}(t_n)\big)\Big)d\epsilon\right\|_{\mathbf{T_3}}^{2k}\middle|\mathcal{F}_{t_n}\right] \leq C_{TE}\big(k,R\big), \qquad (3.31)$$

where $C_{TE}(k,R) := c_{\mathbf{D2}}^{2k}\left(1 + 3^{2kq+1}\left(R^{2kq} + C_{Y_\theta}(k,R)\right)\right)$, *where* $C_{Y_\theta}\big(k,R\big)$ *is from Lemma 3.3.4.*

Proof. By using (2.33), (2.34), (3.5), Lemma 3.3.4, (3.17) and since $c_{\mathbf{D2}} = c_1 \vee c_2$, $q = q_1 \vee q_2$ we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|\int_0^1 (1-\epsilon)\mathbf{D}^2 u\Big(Y(t_n) - \epsilon\big(Y_\theta(s) - \widetilde{Y}(t_n)\big)\Big)d\epsilon\right\|_{\mathbf{T_3}}^{2k}\middle|\mathcal{F}_{t_n}\right]\\
\leq&\mathbb{E}\left[\int_0^1 (1-\epsilon)^{2k}\left\|\mathbf{D}^2 u\Big(\widetilde{Y}(t_n) - \epsilon\big(Y_\theta(s) - \widetilde{Y}(t_n)\big)\Big)\right\|_{\mathbf{T_3}}^{2k}d\epsilon\middle|\mathcal{F}_{t_n}\right]\\
\leq&c_{\mathbf{D2}}^{2k}\mathbb{E}\left[\int_0^1 (1-\epsilon)^{2k}\Big(1 + \left\|\widetilde{Y}(t_n) - \epsilon\cdot\big(Y_\theta(s) - \widetilde{Y}(t_n)\big)\right\|^{2kq}\Big)d\epsilon\middle|\mathcal{F}_{t_n}\right]\\
\leq&c_{\mathbf{D2}}^{2k}\mathbb{E}\Big[1 + 3^{2kq-1}\left\|\widetilde{Y}(t_n)\right\|^{2kq}\\
&\qquad+\int_0^1 (1-\epsilon)^{2k}\epsilon^{2kq}3^{2kq}\left(\|Y_\theta(s)\|^{2kq} + \left\|\widetilde{Y}(t_n)\right\|^{2kq}\right)d\epsilon\Big|\mathcal{F}_{t_n}\Big]\\
\leq&c_{\mathbf{D2}}^{2k}\left(1 + 3^{2kq+1}\left\|\widetilde{Y}(t_n)\right\|^{2kq} + 3^{2kq}\mathbb{E}\Big[\|Y_\theta(s)\|^{2kq}\Big|\mathcal{F}_{t_n}\Big]\right)\\
=&C_{\mathbf{TE}}(k,R),
\end{aligned}$$

where $(1-\epsilon)^{2k}\epsilon^{2kq} \leq 1$ for $k,q\geq 1$ and $\epsilon\in[0,1]$. □

To prove the strong convergence result of Theorem 3.3.1 and Theorem 3.3.2 on the probability of using the backstop and the role of ρ from Assumption 3.2.1, we first set up the error funcion.

3.3.2 Setting up the error function

Notice that $\widetilde{Y}(s)$, from the explicit adaptive Milstein scheme (3.10), takes either the Milstein map θ in (3.9) or the backstop map φ in (3.11) depending on the value of

h_{n+1}. Thus, we define the error by

$$\widetilde{E}(s) := X(s) - \widetilde{Y}(s) = E_\theta(s) + E_\varphi(s), \tag{3.32}$$

for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$. Here

$$E_\varphi(s) := \Big(X(s) - \varphi\left(\widetilde{Y}(t_n), t_n, s - t_n\right)\Big) \mathbf{1}_{\{h_{n+1} \le h_{\min}\}}, \tag{3.33}$$

and $Y_\theta(s)$ is as defined in Definition 3.2.3 and

$$\begin{aligned} E_\theta(s) &:= \big(X(s) - Y_\theta(s)\big) \mathbf{1}_{\{h_{\min} < h_{n+1} \le h_{\max}\}} \\ &= \Bigg(\widetilde{E}(t_n) + \int_{t_n}^{s} \Delta f\big(X(r), \widetilde{Y}(t_n)\big) dr \\ &\quad + \sum_{i=1}^{m} \int_{t_n}^{s} \Delta g_i\big(r, X(r), \widetilde{Y}(t_n)\big) dW_i(r)\Bigg) \mathbf{1}_{\{h_{\min} < h_{n+1} \le h_{\max}\}}, \end{aligned} \tag{3.34}$$

with

$$\Delta f\big(X(r), \widetilde{Y}(t_n)\big) := f(X(r)) - f\big(\widetilde{Y}(t_n)\big); \tag{3.35}$$

$$\Delta g_i\big(r, X(r), \widetilde{Y}(t_n)\big) := g_i(X(r)) - g_i\big(\widetilde{Y}(t_n)\big) - \sum_{j=1}^{m} \mathbf{D} g_i\big(\widetilde{Y}(t_n)\big) g_j\big(\widetilde{Y}(t_n)\big) I_j^{t_n, r}. \tag{3.36}$$

To simplify the proof of Theorem 3.3.1 and Theorem 3.3.2, we require two lemmas. First, we find the second-moment bound of Δg_i in (3.36) on the event $\{h_{\min} < h_{n+1} \le h_{\max}\}$ (so that (3.17) holds).

Lemma 3.3.8. *Let g satisfy Assumption 3.1.1 and Δg_i be as in* (3.36)*. Take $s \in (t_n, t_{n+1}]$, let $X(s)$ be a solution of* (1.1)*, consider $\big(\widetilde{Y}(s), h_{n+1}\big)$ from Definitions 3.2.2 and 3.2.4, and let $Y_\theta(s)$ be as defined in Definition* (3.2.3)*. In this case there exists a constant C_G such that, on the event $\{h_{\min} < h_{n+1} \le h_{\max}\}$,*

$$\begin{aligned} &\mathbb{E}\left[\left\|\Delta g_i\big(s, X(s), \widetilde{Y}(t_n)\big)\right\|^2 \middle| \mathcal{F}_{t_n}\right] \\ &\qquad \le 2\mathbb{E}\Big[\big\|g(X(s)) - g\big(Y_\theta(s)\big)\big\|^2_{\mathbf{F}(d \times m)} \Big| \mathcal{F}_{t_n}\Big] + C_G(R)|s - t_n|^2, \end{aligned} \tag{3.37}$$

where

$$C_G(R) := \; 8C_{Dg_i}^2\big(C_f^2 + C_{\mathit{ISI}}(1,R)\big) + 4C_{\mathit{TE}}(2,R)^{1/2}\, C_{\mathit{SR}}(4,R)^{1/2}\,, \tag{3.38}$$

and C_{ISI}, C_{TE} *and* C_{SR} *are from Lemma 3.3.3, 3.3.7 and 3.3.5, respectively.*

Proof. Substitute Δg_i by (3.36) in the LHS of (3.37), add and subtract $g_i\big(Y_\theta(s)\big)$, and use (2.34) to get

$$\begin{aligned}\mathbb{E}\left[\left\|\Delta g_i\big(s, X(s), \widetilde{Y}(t_n)\big)\right\|^2 \middle| \mathcal{F}_{t_n}\right] &\leq 2\mathbb{E}\left[\left\|g_i(X(s)) - g_i\big(Y_\theta(s)\big)\right\|^2 \middle| \mathcal{F}_{t_n}\right] \\ &+ \underbrace{2\mathbb{E}\left[\left\|g_i\big(Y_\theta(s)\big) - g_i\big(\widetilde{Y}(t_n)\big) - \sum_{j=1}^m \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_j^{t_n,s}\right\|^2 \middle| \mathcal{F}_{t_n}\right]}_{=:G_1}. \end{aligned} \tag{3.39}$$

To analyse G_1, we expand $g_i(Y_\theta(s))$ using Taylor's theorem (see for example [38, A.1]) around $g_i\big(\widetilde{Y}(t_n)\big)$ to get

$$\begin{aligned} g_i\big(Y_\theta(s)\big) - g_i\big(\widetilde{Y}(t_n)\big) =\; & \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big(Y_\theta(s) - \widetilde{Y}(t_n)\big) \\ & + \int_0^1 (1-\epsilon)\mathbf{D}^2 g_i\Big(\widetilde{Y}(t_n) - \epsilon\big(Y_\theta(s) - \widetilde{Y}(t_n)\big)\Big)\Big[Y_\theta(s) - \widetilde{Y}(t_n)\Big]^2 d\epsilon, \end{aligned} \tag{3.40}$$

where we recall from Definition 2.0.1 that $[\cdot]^2$ represents the outer product of a vector with itself. Substituting (3.40) into G_1 in (3.39), taking out $\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)$ as a common factor, and applying (2.34) gives

$$\begin{aligned} G_1 \;\leq\; & 4\mathbb{E}\left[\left\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\Big(Y_\theta(s) - \widetilde{Y}(t_n) - \sum_{j=1}^m g_j\big(\widetilde{Y}(t_n)\big)I_j^{t_n,s}\Big)\right\|^2 \middle| \mathcal{F}_{t_n}\right] \\ & + 4\mathbb{E}\Bigg[\Bigg\| \int_0^1 (1-\epsilon)\mathbf{D}^2 g_i\Big(\widetilde{Y}(t_n) - \epsilon\big(Y_\theta(s) - \widetilde{Y}(t_n)\big)\Big) \\ & \qquad\qquad \times \Big[Y_\theta(s) - \widetilde{Y}(t_n)\Big]^2 d\epsilon\Bigg\|^2 \Bigg| \mathcal{F}_{t_n}\Bigg] \\ =: \; & G_{1.1} + G_{1.2}. \end{aligned} \tag{3.41}$$

For $G_{1.1}$ in (3.41), by submultiplicativity of the Euclidean norm and the fact that the induced matrix 2-norm is bounded above by the Frobenius norm; by (3.12), (3.21) and (3.22) in the statement of Lemma 3.3.3 with $k = 1$, we have

$$\begin{aligned} G_{1.1} &\leq 8\Big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\Big\|_{\mathbf{F}}^2\Bigg(\Big\|f\big(\widetilde{Y}(t_n)\big)\Big\|^2|s-t_n|^2 \\ &\qquad + \mathbb{E}\Bigg[\Bigg\|\sum_{i,j=1}^{m}\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{j,i}^{t_n,s}\Bigg\|^2\Bigg|\mathcal{F}_{t_n}\Bigg]\Bigg) \\ &\leq 8C_{Dg_i}^2\big(C_f^2 + C_{\mathtt{ISI}}\,(1,R)\,\big)|s-t_n|^2. \end{aligned} \tag{3.42}$$

For $G_{1.2}$ in (3.41), we apply (2.33), the Cauchy-Schwarz inequality, then using (3.31) in Lemma 3.3.7 with $k = 2$ and (3.26) in Lemma 3.3.5 with $k = 4$ we get

$$\begin{aligned} G_{1.2} &\leq 4\int_0^1\left(\mathbb{E}\left[\Big\|(1-\epsilon)\mathbf{D}^2g_i\Big(\widetilde{Y}(t_n)-\epsilon\big(Y_\theta(s)-\widetilde{Y}(t_n)\big)\Big)\Big\|_{\mathbf{T_3}}^4\Big|\mathcal{F}_{t_n}\right]\right)^{1/2}d\epsilon \\ &\qquad\qquad \times\left(\mathbb{E}\left[\Big\|Y_\theta(s)-\widetilde{Y}(t_n)\Big\|^8\Big|\mathcal{F}_{t_n}\right]\right)^{1/2} \\ &\leq 4C_{\mathtt{TE}}\,(2,R)^{1/2}\,C_{\mathtt{SR}}\,(4,R)^{1/2}\ |s-t_n|^2. \end{aligned} \tag{3.43}$$

Substituting the bounds (3.42) and (3.43) back to (3.41) before bringing together the terms in (3.39), we have

$$\begin{aligned} \mathbb{E}\left[\Big\|\Delta g_i\big(s,X(s),\widetilde{Y}(t_n)\big)\Big\|^2\Big|\mathcal{F}_{t_n}\right] &\leq 2\mathbb{E}\left[\Big\|g_i(X(s))-g_i\big(Y_\theta(s)\big)\Big\|^2\Big|\mathcal{F}_{t_n}\right] \\ &\qquad + C_G(R)|s-t_n|^2, \end{aligned}$$

where $C_G(R)$ is given in (3.38). By bounding $\|g_i\|^2$ with $\|g\|^2_{\mathbf{F}(d\times m)}$, the statement of Lemma 3.3.8 follows. □

The second lemma in the following gives the conditional second-moment bound of $E_\theta(s)$ as in (3.34), which is the first part of the one-step error in (3.32).

Lemma 3.3.9. *Let f, g satisfy Assumption 3.1.1 and 3.1.2. Let $X(s)$ be a solution of* (1.1) *and $\widetilde{E}(s)$ a solution of* (3.32) *with $E_\theta(s)$ defined in* (3.34)*, with $s \in [t_n, t_{n+1}]$, $n \in \mathbb{N}$. In this case there exists a constant C_E and an $\mathcal{F}_{t_n}$-measurable random variable $\overline{C}_M^{\{4(q+2)\}}$ such that*

$$\begin{aligned} \mathbb{E}\Big[\big\|E_\theta(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \leq\ \big\|\widetilde{E}(t_n)\big\|^2 + C_E(R)\int_{t_n}^{t_{n+1}}\mathbb{E}\Big[\big\|E_\theta(r)\big\|^2\Big|\mathcal{F}_{t_n}\Big]\,dr \\ + \overline{C}_M^{\{4(q+2)\}}(R)\,h_{n+1}^3, \quad a.s. \end{aligned} \tag{3.44}$$

where

$$C_E(R) := \; 2K_1(R) + 2L, \tag{3.45}$$

with constant K_1 *as defined in* (3.72). *The* $\mathcal{F}_{t_n}$*-measurable random variable* $\overline{C}_M^{\{4(q+2)\}}$ *is given by*

$$\overline{C}_M^{\{4(q+2)\}}(R) \; := \; m^4 C_{Df}^2 C_{g_i}^2 + 2\overline{K}_2^{\{4(q+2)\}} + mC_G(R), \tag{3.46}$$

with the $\mathcal{F}_{t_n}$*-measurable random variable* $\overline{K}_2^{\{4(q+2)\}}$ *defined in* (3.73), *constant* C_G *defined in Lemma 3.3.8. We denote* $\mathbb{E}\left[\overline{C}_M^{\{4(q+2)\}}(R)\right] =: C_M(R)$, *the finiteness of which is ensured in* (3.76).

We recall that the superscript notation in (3.46) follows the convention introduced in the statement of Lemma 3.3.6 and indicates the number of finite moments required of the SDE solution.

Proof. Throughout the proof, we restrict attention to trajectories on the event $\{h_{\min} < h_{n+1} \leq h_{\max}\}$, since by (3.34), $E_\theta(s)$ is only nonzero on this event, otherwise (3.44) holds trivially. Applying the stopping time variant of Itô formula (see Mao & Yuan [42, (1.45)] , Remark 1 and Example 2.1.1) to (3.34), we have

$$\begin{aligned} \left\|E_\theta(t_{n+1})\right\|^2 &= \left\|\widetilde{E}(t_n)\right\|^2 + 2\int_{t_n}^{t_{n+1}} \underbrace{\Big\langle E_\theta(r), \Delta f\big(X(r), \widetilde{Y}(t_n)\big)\Big\rangle}_{=:J_f} dr \\ &+ \sum_{i=1}^{m}\int_{t_n}^{t_{n+1}} \Big\| \underbrace{\Delta g_i\big(r, X(r), \widetilde{Y}(t_n)\big)}_{=:J_{g_i}} \Big\|^2 dr + 2\sum_{i=1}^{m}\int_{t_n}^{t_{n+1}} \big\langle E_\theta(r), J_{g_i}\big\rangle dW_i(r). \end{aligned} \tag{3.47}$$

Take expectations on both sides conditional upon $\mathcal{F}_{t_n}$, and since $\int_{t_n}^{t_{n+1}} \big|J_f\big| dr$ has finite expectation (by the boundedness of $\widetilde{Y}(t_n)$ in (3.17) and the finiteness of absolute moments of $X(r)$ see (3.6)), using Fubini's Theorem (see for example [10, Proposition 12.10]) and (3.13) we have,

$$\begin{aligned} \mathbb{E}\Big[\left\|E_\theta(t_{n+1})\right\|^2 \Big| \mathcal{F}_{t_n}\Big] &= \left\|\widetilde{E}(t_n)\right\|^2 + 2\int_{t_n}^{t_{n+1}} \mathbb{E}\big[J_f \big| \mathcal{F}_{t_n}\big] dr \\ &+ \sum_{i=1}^{m}\int_{t_n}^{t_{n+1}} \mathbb{E}\big[\|J_{g_i}\|^2 \big| \mathcal{F}_{t_n}\big] dr, \end{aligned} \tag{3.48}$$

By Lemma 3.3.8, we have the bound of $\|J_{g_i}\|^2$ in (3.48) as

$$\mathbb{E}\big[\|J_{g_i}\|^2\big|\mathcal{F}_{t_n}\big] \le 2\mathbb{E}\Big[\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big] + C_G(R)|r-t_n|^2, \quad a.s. \quad (3.49)$$

For J_f, by substituting Δf with (3.35) with adding in and subtracting out $f(Y_\theta(r))$, we have

$$J_f = \Big\langle E_\theta(r), f(X(r)) - f(Y_\theta(r))\Big\rangle + \underbrace{\Big\langle E_\theta(r), f(Y_\theta(r)) - f\big(\widetilde{Y}(t_n)\big)\Big\rangle}_{=:H}. \quad (3.50)$$

Substituting (3.50) and (3.49) back into (3.48), we have

$$\mathbb{E}\Big[\big\|E_\theta(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \le \big\|\widetilde{E}(t_n)\big\|^2 + mC_G(R)h^3_{n+1} + 2\int_{t_n}^{t_{n+1}}\mathbb{E}\big[J_{f,g}\big|\mathcal{F}_{t_n}\big]dr + 2\int_{t_n}^{t_{n+1}}\mathbb{E}\big[H\big|\mathcal{F}_{t_n}\big]dr, \quad (3.51)$$

where

$$J_{f,g} := \Big\langle E_\theta(r), f(X(r)) - f(Y_\theta(r))\Big\rangle + \big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}. \quad (3.52)$$

For H in (3.50), and in a similar way to (3.40), we expand $f(Y_\theta(r))$ using Taylor's theorem around $\widetilde{Y}(t_n)$ to have

$$f(Y_\theta(r)) - f\big(\widetilde{Y}(t_n)\big) = \mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big(Y_\theta(r) - \widetilde{Y}(t_n)\big) + \int_0^1 (1-\epsilon)\mathbf{D}^2 f\Big(\widetilde{Y}(t_n) - \epsilon\cdot\big(Y_\theta(r) - \widetilde{Y}(t_n)\big)\Big)\Big[Y_\theta(r) - \widetilde{Y}(t_n)\Big]^2 d\epsilon. \quad (3.53)$$

Then, by (3.12) we substitute $Y_\theta(r)$ in the first term on the RHS of (3.53) with (3.9) where we use its expanded form as in (2.22) for $s = r$. Therefore, for the last term on the RHS of (3.51), we have

$$\mathbb{E}\big[H\big|\mathcal{F}_{t_n}\big] \le H_1 + H_2 + H_3 + H_4 + H_5 + H_6, \quad (3.54)$$

where

$$H_1 := \mathbb{E}\Big[\Big\langle E_\theta(r), \mathbf{D}f\big(\widetilde{Y}(t_n)\big)|r-t_n|f\big(\widetilde{Y}(t_n)\big)\Big\rangle\Big|\mathcal{F}_{t_n}\Big];$$

$$H_2 := \mathbb{E}\Big[\Big\langle E_\theta(r), \underbrace{\sum_{i=1}^{m}\mathbf{D}f\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)I_i^{t_n,r}}_{=:H_{2R}}\Big\rangle\Big|\mathcal{F}_{t_n}\Big];$$

$$H_3 := \mathbb{E}\Bigg[\Bigg\langle E_\theta(r), \ \frac{1}{2}\sum_{i=1}^{m}\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big) \times\Big(\big(I_i^{t_n,r}\big)^2-|r-t_n|\Big)\Bigg\rangle\Bigg|\mathcal{F}_{t_n}\Bigg]$$

$$H_4 := \mathbb{E}\Bigg[\Bigg\langle E_\theta(r), \ \frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^{m}\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\Big(\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big) + \mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big)I_i^{t_n,r}I_j^{t_n,r}\Bigg\rangle\Bigg|\mathcal{F}_{t_n}\Bigg];$$

$$H_5 := \mathbb{E}\Bigg[\Bigg\langle E_\theta(r), \ \sum_{\substack{i,j=1\\ i<j}}^{m}\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\Big(\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big) - \mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big)A_{ij}(t_n,r)\Bigg\rangle\Bigg|\mathcal{F}_{t_n}\Bigg];$$

$$H_6 := \mathbb{E}\Bigg[\Bigg\langle E_\theta(r), \ \int_0^1(1-\epsilon)\mathbf{D}^2f\Big(\widetilde{Y}(t_n)-\epsilon\cdot\big(Y_\theta(r)-\widetilde{Y}(t_n)\big)\Big) \times\Big[Y_\theta(r)-\widetilde{Y}(t_n)\Big]^2d\epsilon\Bigg\rangle\Bigg|\mathcal{F}_{t_n}\Bigg].$$

We will now determine suitable upper bounds for each of H_1, H_2, H_3, H_4, H_5, and H_6 in turn. For H_1 in (3.54), by the Cauchy-Schwarz inequality, (2.31), and (3.21), we have

$$\begin{aligned} H_1 &\le \mathbb{E}\Big[\|E_\theta(r)\|\,\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}\,\big\|f\big(\widetilde{Y}(t_n)\big)\big\|\,|r-t_n|\,\Big|\mathcal{F}_{t_n}\Big] \\ &\le \mathbb{E}\left[\frac{1}{2}\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}^2\,\big\|f\big(\widetilde{Y}(t_n)\big)\big\|^2\|E_\theta(r)\|^2+\frac{1}{2}|r-t_n|^2\,\middle|\mathcal{F}_{t_n}\right] \\ &\le \frac{1}{2}C_{Df}^2C_f^2\,\mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big]+\frac{1}{2}|r-t_n|^2. \end{aligned} \tag{3.55}$$

Next, for the analysis of H_2 in (3.54), by (3.13), we firstly have

$$\mathbb{E}\big[H_{2R}\big|\mathcal{F}_{t_n}\big] = \sum_{i=1}^{m}\mathbf{D}f\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\mathbb{E}\Big[I_i^{t_n,r}\Big|\mathcal{F}_{t_n}\Big]=0. \tag{3.56}$$

By (2.34), the Cauchy-Schwarz inequality, (3.21) and (3.14) we also have

$$\mathbb{E}\Big[\big\|H_{2R}\big\|^2\Big|\mathcal{F}_{t_n}\Big] \leq m\sum_{i=1}^{m}\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}^2\big\|g_i\big(\widetilde{Y}(t_n)\big)\big\|^2\mathbb{E}\left[\big|I_i^{t_n,r}\big|^2\Big|\mathcal{F}_{t_n}\right]$$
$$\leq m^2C_{Df}^2C_{g_i}^2|r-t_n|. \tag{3.57}$$

Then, for H_2 in (3.54) we firstly expand $E_\theta(r)$ using (3.34) to have

$$H_2 = \mathbb{E}\left[\Big\langle\widetilde{E}(t_n),\, H_{2R}\Big\rangle\Big|\mathcal{F}_{t_n}\right] + \mathbb{E}\left[\left\langle\int_{t_n}^{r}\Delta f(X(p),Y(t_n))dp,\, H_{2R}\right\rangle\Big|\mathcal{F}_{t_n}\right]$$
$$+\,\mathbb{E}\left[\left\langle\sum_{i=1}^{m}\int_{t_n}^{r}\Delta g_i\big(p,X(p),\widetilde{Y}(t_n)\big)dW_i(p),\, H_{2R}\right\rangle\Big|\mathcal{F}_{t_n}\right]$$
$$=: H_{2.1}+H_{2.2}+H_{2.3}. \tag{3.58}$$

For $H_{2.1}$ in (3.58), by (3.56) we have

$$H_{2.1} = \Big\langle\widetilde{E}(t_n),\, \mathbb{E}\big[H_{2R}\big|\mathcal{F}_{t_n}\big]\Big\rangle = 0. \tag{3.59}$$

For $H_{2.2}$ in (3.58), by adding in and subtracting out $f(X(t_n))$ in Δf in (3.35):

$$H_{2.2} = \mathbb{E}\left[\Big\langle\int_{t_n}^{r} f(X(r))-f(X(t_n))dp,\, H_{2R}\Big\rangle\Big|\mathcal{F}_{t_n}\right]$$
$$+\,\mathbb{E}\left[\Big\langle\int_{t_n}^{r} f(X(t_n))-f\big(\widetilde{Y}(t_n)\big)dp,\, H_{2R}\Big\rangle\Big|\mathcal{F}_{t_n}\right]$$
$$=: H_{2.21}+H_{2.22}. \tag{3.60}$$

Similar to $H_{2.1}$ in (3.59), we have $H_{2.22}=0$. For $H_{2.21}$ in (3.60), using the Cauchy-Schwarz inequality and (3.57) we have

$$H_{2.21} \leq \mathbb{E}\left[\left\|\int_{t_n}^{r} f(X(p))-f(X(t_n))dp\right\|\,\big\|H_{2R}\big\|\Big|\mathcal{F}_{t_n}\right]$$
$$\leq\left(|r-t_n|\int_{t_n}^{r}\mathbb{E}\left[\|f(X(p))-f(X(t_n))\|^2\big|\mathcal{F}_{t_n}\right]dp\;\mathbb{E}\Big[\big\|H_{2R}\big\|^2\Big|\mathcal{F}_{t_n}\Big]\right)^{1/2}$$
$$\leq mC_{Df}C_{g_i}|r-t_n|\left(\int_{t_n}^{r}\mathbb{E}\left[\|f(X(p))-f(X(t_n))\|^2\big|\mathcal{F}_{t_n}\right]dp\right)^{1/2}. \tag{3.61}$$

By Taylor's theorem, we expand $f(X(p))$ around $f(X(t_n))$ to the first order, and using (3.3), the Cauchy-Schwarz inequality, Lemma 3.3.6 with $k = 2$ and (2.34):

$$\begin{aligned}
&\mathbb{E}\left[\left\|f(X(p)) - f(X(t_n))\right\|^2 \Big| \mathcal{F}_{t_n}\right] \\
&= \mathbb{E}\left[\left\|\int_0^1 \mathbf{D}f(X(t_n) - \epsilon \cdot (X(p) - X(t_n))(X(p) - X(t_n))d\epsilon\right\|^2 \Bigg| \mathcal{F}_{t_n}\right] \\
&\leq \Big(\mathbb{E}\left[\|X(p) - X(t_n)\|^4 \big| \mathcal{F}_{t_n}\right]\Big)^{1/2} \\
&\qquad \times \left(\mathbb{E}\left[\left\|\int_0^1 \mathbf{D}f\big(X(t_n) - \epsilon \cdot (X(p) - X(t_n)\big)d\epsilon\right\|_{\mathbf{F}}^4 \Bigg| \mathcal{F}_{t_n}\right]\right)^{1/2} \\
&\leq \overline{C}_{H2.21}^{\{4(q+2)\}} |p - t_n|, \qquad (3.62)
\end{aligned}$$

where

$$\begin{aligned}
\overline{C}_{H2.21}^{\{4(q+2)\}} := &\left(\overline{C}_{\mathtt{PR}}^{\{4(q+2)\}}\right)^{1/2} \\
&\times c_3^2 \Big(1 + 3^{4q_1+4}\mathbb{E}\Big[\sup_{p \in [t_n, t_{n+1}]} \|X(p)\|^{4q_1+4} \Big| \mathcal{F}_{t_n}\Big]\Big)^{1/2}. \qquad (3.63)
\end{aligned}$$

Substituting (3.62) back to (3.61) and using that $H_{2.22} = 0$, we have

$$H_{2.2} \leq mC_{Df}C_{g_i} \left(\overline{C}_{H2.21}^{\{4(q+2)\}}\right)^{1/2} |r - t_n|^2. \qquad (3.64)$$

For $H_{2.3}$ as in (3.58), using the Cauchy-Schwarz inequality, (2.34), (3.21), (3.14) and Itô's isometry in (2.13) we have

$$\begin{aligned}
H_{2.3} \leq &\left(\mathbb{E}\left[\left\|\sum_{i=1}^{m} \int_{t_n}^{r} \Delta g_i\big(p, X(p), \widetilde{Y}(t_n)\big)dW_i(p)\right\|^2 \Bigg| \mathcal{F}_{t_n}\right]\right)^{1/2} \\
&\qquad \times \left(\mathbb{E}\left[\left\|\sum_{i=1}^{m} \mathbf{D}f\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)I_i^{t_n,r}\right\|^2 \Bigg| \mathcal{F}_{t_n}\right]\right)^{1/2} \\
\leq &\left(m\sum_{i=1}^{m} \int_{t_n}^{r} \mathbb{E}\left[\left\|\Delta g_i\big(p, X(p), \widetilde{Y}(t_n)\big)\right\|^2 \Big| \mathcal{F}_{t_n}\right] dp\right)^{1/2} \\
&\qquad \times mC_{Df}C_{g_i}|r - t_n|^{1/2}.
\end{aligned}$$

Then, by Lemma 3.3.8 we have

$$H_{2.3} \leq \left(2m^2 \int_{t_n}^{r} \mathbb{E}\Big[\big\|g(X(p)) - g(Y_\theta(p))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dp + C_G(R)|r-t_n|^3\right)^{1/2} mC_{Df}C_{g_i}|r-t_n|^{1/2}.$$

Since the integrand $\mathbb{E}\Big[\big\|g(X(p)) - g(Y_\theta(p))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]$ is non-negative for all $p \in [t_n, t_{n+1}]$, we can replace the upper limit of integration with t_{n+1}. With $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, we have

$$\begin{aligned} H_{2.3} \leq{}& \sqrt{2}m^2 C_{Df}C_{g_i}|r-t_n|^{1/2} \\ &\times \left(\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dr\right)^{1/2} \\ &+ mC_{Df}C_{g_i}C_G(R)^{1/2}|r-t_n|^2. \end{aligned} \tag{3.65}$$

Notice that we changed the variable of integration from p back to r for consistency. Substituting (3.59), (3.64) and (3.65) back into (3.58), we have

$$\begin{aligned} H_2 \leq{}& mC_{Df}C_{g_i}\left(\left(\overline{C}^{\{4(q+2)\}}_{H2.21}\right)^{1/2} + C_G(R)^{1/2}\right)|r-t_n|^2 \\ &+ \sqrt{2}m^2 C_{Df}C_{g_i}|r-t_n|^{1/2} \\ &\times \left(\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dr\right)^{1/2}. \end{aligned} \tag{3.66}$$

For H_3 in (3.54), by the Cauchy-Schwarz inequality, triangle inequality, (2.34), (2.31), (3.15), (3.3) and (3.21) we have

$$\begin{aligned} H_3 \leq{}& \mathbb{E}\left[\frac{1}{4}\sum_{i=1}^{m}\left(\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|^2_{\mathbf{F}}\big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|^2_{\mathbf{F}}\big\|g_i\big(\widetilde{Y}(t_n)\big)\big\|^2\|E_\theta(r)\|^2 + 2\left|I_i^{t_n,r}\right|^4 + 2|r-t_n|^2\right)\middle|\mathcal{F}_{t_n}\right] \\ \leq{}& \frac{m}{4}C^2_{Df}C^2_{Dg_i}C^2_{g_i}\,\mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big] + \frac{(\Upsilon_4+1)m}{2}|r-t_n|^2. \end{aligned} \tag{3.67}$$

For H_4 in (3.54), by the Cauchy-Schwarz inequality, conditional independence of the Itô integrals, (3.14), triangle inequality, (2.31), Itô's isometry in (2.13), (3.3), and

(3.21), we have

$$
\begin{aligned}
H_4 \leq \mathbb{E}\Bigg[\frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^{m} \|E_\theta(r)\| \big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}} \Big(\big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}} \big\|g_j\big(\widetilde{Y}(t_n)\big)\big\| \\
+ \big\|\mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}} \big\|g_i\big(\widetilde{Y}(t_n)\big)\big\|\Big) \, |I_i^{t_n,r}| \, |I_j^{t_n,r}| \Bigg| \mathcal{F}_{t_n}\Bigg] \\
\leq \frac{1}{4} m(m-1) C_{Df}^2 C_{Dg_i}^2 C_{g_i}^2 \mathbb{E}\Big[\|E_\theta(r)\|^2 \Big| \mathcal{F}_{t_n}\Big] + \frac{1}{8} m(m-1)|r-t_n|^2. \qquad (3.68)
\end{aligned}
$$

For H_5 in (3.54), by the Cauchy-Schwarz inequality, triangle inequality, (2.31), (3.21), (3.3), and Lemma 2.3.1 with $b = 2$, we have

$$
\begin{aligned}
H_5 \leq \mathbb{E}\Bigg[\frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^{m} \Bigg(\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\|_{\mathbf{F}}^2 \|E_\theta(r)\|^2 \Big(\big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}^2 \big\|g_j\big(\widetilde{Y}(t_n)\big)\big\|^2 \\
+ \big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}^2 \big\|g_j\big(\widetilde{Y}(t_n)\big)\big\|^2\Big) + \big|A_{ij}(t_n,r)\big|^2\Bigg) \Bigg| \mathcal{F}_{t_n}\Bigg] \\
\leq \frac{1}{2} m(m-1) C_{Df}^2 C_{Dg_i}^2 C_{g_i}^2 \mathbb{E}\Big[\|E_\theta(r)\|^2 \Big| \mathcal{F}_{t_n}\Big] \\
+ \frac{1}{4} m(m-1) (C_{\mathtt{LA}}(2))^2 |r-t_n|^2. \qquad (3.69)
\end{aligned}
$$

For H_6 in (3.54), by the Cauchy-Schwarz inequality, triangle inequality, and (2.31) we have (noting that $\|[\cdot]^2\|_{\mathbf{F}} = \|\cdot\|^2$)

$$
\begin{aligned}
H_6 \leq & \, \mathbb{E}\Bigg[\|E_\theta(r)\| \big\|Y_\theta(r) - \widetilde{Y}(t_n)\big\|^2 \\
& \times \Big\|\int_0^1 (1-\epsilon)\mathbf{D}^2 f\Big(\widetilde{Y}(t_n) - \epsilon\cdot\big(Y_\theta(r) - \widetilde{Y}(t_n)\big)\Big) d\epsilon\Big\|_{\mathbf{T}_3} \Bigg| \mathcal{F}_{t_n}\Bigg] \\
\leq & \, \frac{1}{2}\mathbb{E}\Big[\|E_\theta(r)\|^2 \Big| \mathcal{F}_{t_n}\Big] + \frac{1}{2} \underbrace{\sqrt{\mathbb{E}\left[\big\|Y_\theta(r) - \widetilde{Y}(t_n)\big\|^8 \Big| \mathcal{F}_{t_n}\right]}}_{H_{6.1}} \\
& \times \underbrace{\sqrt{\mathbb{E}\left[\Big\|\int_0^1 (1-\epsilon)\mathbf{D}^2 f\Big(\widetilde{Y}(t_n) - \epsilon\cdot\big(Y_\theta(r) - \widetilde{Y}(t_n)\big)\Big) d\epsilon\Big\|_{\mathbf{T}_3}^4 \Bigg| \mathcal{F}_{t_n}\right]}}_{H_{6.2}}.
\end{aligned}
$$

From (3.26) in Lemma 3.3.5 with $k = 4$, we have $H_{6.1} \leq C_{\mathtt{SR}}(4,R)^{1/2}|r-t_n|^2$. From (3.31) in Lemma 3.3.7 with $k = 2$, we have $H_{6.2} \leq C_{\mathtt{TE}}(2,R)^{1/2}$. Therefore, H_6 in

(3.54) becomes

$$H_6 \leq \frac{1}{2}\mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big] + \frac{1}{2}C_{\texttt{SR}}\,(4,R)^{1/2}\;C_{\texttt{TE}}\,(2,R)^{1/2}\;|r-t_n|^2, \tag{3.70}$$

Substituting (3.55), (3.66), (3.67), (3.68), (3.69) and (3.70) back into (3.54) for H, we have

$$\begin{aligned}\mathbb{E}[H|\mathcal{F}_{t_n}] \leq\; & K_1(R)\mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big] + \overline{K}_2^{\{4(q+2)\}}(R)|r-t_n|^2 \\ & + \sqrt{2}m^2C_{Df}C_{g_i}|r-t_n|^{1/2} \\ & \times\left(\int_{t_n}^{t_{n+1}}\mathbb{E}\Big[\big\|g(X(r))-g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dr\right)^{1/2}, \end{aligned}\tag{3.71}$$

where

$$K_1(R) := \frac{1}{2} + \frac{1}{2}C_{Df}^2C_f^2 + m(m-1)C_{Df}^2C_{Dg_i}^2C_{g_i}^2, \tag{3.72}$$

and with $\overline{C}_{H2.21}^{\{4(q+2)\}}$ from (3.63)

$$\begin{aligned}\overline{K}_2^{\{4(q+2)\}}(R) :=\; & \frac{1}{2} + mC_{Df}C_{g_i}\Bigg(\left(\overline{C}_{H2.21}^{\{4(q+2)\}}\right)^{1/2} + \frac{1}{2}(\Upsilon_4+1)m \\ & + \frac{1}{4}m(m-1)\big(1+(C_{\texttt{LA}}\,(2))^2\,\big) + \frac{1}{2}C_{\texttt{SR}}\,(4,R)^{1/2}\,C_{\texttt{TE}}\,(2,R)^{1/2}. \end{aligned}\tag{3.73}$$

Substituting $\mathbb{E}[H|\mathcal{F}_{t_n}]$ from (3.71) back into (3.51), we have

$$\begin{aligned}\mathbb{E}\Big[\big\|E_\theta(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \leq\; & \big\|\widetilde{E}(t_n)\big\|^2 + 2K_1(R)\int_{t_n}^{t_{n+1}}\mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big]dr \\ & + mC_G(R)h_{n+1}^3 + \overline{K}_2^{\{4(q+2)\}}(R)h_{n+1}^3 \\ & + 2\int_{t_n}^{t_{n+1}}\mathbb{E}\big[J_{f,g}\big|\mathcal{F}_{t_n}\big]dr + \sqrt{2}m^2C_{Df}C_{g_i}h_{n+1}^{3/2} \\ & \times\left(\int_{t_n}^{t_{n+1}}\mathbb{E}\Big[\big\|g(X(r))-g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dr\right)^{1/2}. \end{aligned}\tag{3.74}$$

Using (2.31) on the last term on the RHS of (3.74), we have

$$\mathbb{E}\Big[\big\|E_\theta(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \leq \ \big\|\widetilde{E}(t_n)\big\|^2 + 2K_1(R)\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big]dr \\ + \overline{C}_M^{\{4(q+2)\}}(R)\, h_{n+1}^3 \\ + 2\int_{t_n}^{t_{n+1}} \mathbb{E}\left[J_{f,g} + \frac{1}{2}\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}\middle|\mathcal{F}_{t_n}\right]dr, \quad (3.75)$$

where $\overline{C}_M^{\{4(q+2)\}}$ is as defined in (3.46). Recall $J_{f,g}$ is given in (3.52) so that

$$J_{f,g} + \frac{1}{2}\Big\|g(X(r)) - g(Y_\theta(r))\Big\|^2_{\mathbf{F}(d\times m)} \\ = \Big\langle E_\theta(r), f(X(r)) - f(Y_\theta(r))\Big\rangle + \frac{3}{2}\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}.$$

By Assumption 3.1.2 we can apply the monotone condition (3.2):

$$\mathbb{E}\Big[\big\|E_\theta(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \leq \ \big\|\widetilde{E}(t_n)\big\|^2 + C_E(R)\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big]dr \\ + \overline{C}_M^{\{4(q+2)\}}(R)\, h_{n+1}^3,$$

where $C_E(R)$ is in (3.45).

To obtain the the final estimate on $C_M(R)$ in the Lemma we use the explicit form of $\overline{C}_M^{\{4(q+2)\}}$, $\overline{K}_2^{\{4(q+2)\}}$, $\overline{C}_{H2.21}^{\{4(q+2)\}}$, given by (3.46), (3.73), and (3.63) respectively, (3.30) in the statement of Lemma 3.3.7, (3.6), and Assumption 3.1.2 we bound the expectation of $\overline{C}_M^{\{4(q+2)\}}$ in (3.46) as follows,

$$C_M(R) := \mathbb{E}\left[\overline{C}_M^{\{4(q+2)\}}(R)\right] \\ \leq m^4 C_{Df}^2 C_{g_i}^2 + 2mC_{Df}C_{g_i}\Big(c_3 C_{\mathtt{PR}}(2)^{1/4}\left(1 + 3^{q_1+1}C_{\mathtt{X}}\right) + C_G(R)^{1/2}\Big) \\ + \frac{1}{2}(\Upsilon_4 + 1)m + \frac{1}{2}m(m-1)\big(1 + (C_{\mathtt{LA}}(2))^2\big) \\ + C_{\mathtt{SR}}(4,R)^{1/2}\, C_{\mathtt{TE}}(2,R)^{1/2} + mC_G(R) + 1. \quad (3.76)$$

□

3.3.3 Proof of Theorem 3.3.1 on strong convergence.

Proof. Firstly, by (3.32) we have the conditional second-moment bound of the one-step error as

$$\mathbb{E}\Big[\big\|\widetilde{E}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] = \mathbb{E}\Big[\big\|E_\theta(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] + \mathbb{E}\Big[\big\|E_\varphi(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big], \tag{3.77}$$

where by (3.11) and (3.33), the one-step error bound of the backstop map yields

$$\mathbb{E}\Big[\big\|E_\varphi(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \le \big\|\widetilde{E}(t_n)\big\|^2 + C_{B_1}\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\|E_\varphi(r)\|^2\Big|\mathcal{F}_{t_n}\Big]dr + C_{B_2}\,h_{n+1}^3, \quad a.s. \tag{3.78}$$

Therefore, by substituting (3.44) and (3.78) into (3.77), and recalling (3.32) we have for any h_{n+1} that satisfies Assumption 3.2.1,

$$\mathbb{E}\Big[\big\|\widetilde{E}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \le \big\|\widetilde{E}(t_n)\big\|^2 + \Gamma_1(R)\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_n}\Big]dr + \overline{\Gamma}_2^{\{4(q+2)\}}\big(R\big)h_{n+1}^3, \quad a.s. \tag{3.79}$$

where with $C_E(R)$ in (3.45) and $C_M(R)$ in (3.76) we define Γ_1, $\overline{\Gamma}_2$ as

$$\Gamma_1(R) := C_E(R) + C_{B_1}; \tag{3.80}$$

$$\overline{\Gamma}_2^{\{4(q+2)\}}\big(R\big) := \overline{C}_M^{\{4(q+2)\}}(R) + C_{B_2};$$

$$\Gamma_2(R) := \mathbb{E}\left[\overline{\Gamma}_2^{\{4(q+2)\}}\big(R\big)\right] \le C_M(R) + C_{B_2}. \tag{3.81}$$

For a fixed $t > 0$, let $N^{(t)}$ be as in Definition 3.2.1, we multiply both sides of (3.79) with the indicator function $\mathbf{1}_{\{N^{(t)}>n+1\}}$ and sum up the steps excluding the last step $N^{(t)}$ to have

$$\sum_{n=0}^{N^{(t)}-2}\mathbb{E}\Big[\big\|\widetilde{E}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big]\mathbf{1}_{\{N^{(t)}>n+1\}} \le \sum_{n=0}^{N^{(t)}-2}\big\|\widetilde{E}(t_n)\big\|^2\mathbf{1}_{\{N^{(t)}>n+1\}} + \Gamma_1(R)\sum_{n=0}^{N^{(t)}-2}\int_{t_n}^{t_{n+1}}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_n}\Big]\mathbf{1}_{\{N^{(t)}>n+1\}}dr + \overline{\Gamma}_2^{\{4(q+2)\}}(R)\sum_{n=0}^{N^{(t)}-2}h_{n+1}^3\mathbf{1}_{\{N^{(t)}>n+1\}}. \tag{3.82}$$

Since $t \in \left[t_{N^{(t)}-1}, t_{N^{(t)}}\right]$, we use (3.79) to express the last step, noting that it holds when t_n, t_{n+1} are replaced by $t_{N^{(t)}-1}$ and t respectively:

$$\mathbb{E}\Big[\big\|\widetilde{E}(t)\big\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big] \le \big\|\widetilde{E}(t_{N^{(t)}-1})\big\|^2 + \Gamma_1(R)\int_{t_{N^{(t)}-1}}^{t} \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big]dr \\ + \overline{\Gamma}_2^{\{4(q+2)\}}(R)\big|t - t_{N^{(t)}-1}\big|^3. \tag{3.83}$$

By adding the both sides of (3.82) and (3.83), and taking an expectation:

$$\left.\begin{aligned}&\mathbb{E}\Bigg[\sum_{n=0}^{N^{(t)}-2}\Big(\mathbb{E}\Big[\big\|\widetilde{E}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] - \big\|\widetilde{E}(t_n)\big\|^2\Big)\mathbf{1}_{\{N^{(t)}>n+1\}} \\ &\qquad + \mathbb{E}\Big[\big\|\widetilde{E}(t)\big\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big] - \big\|\widetilde{E}(t_{N^{(t)}-1})\big\|^2\Bigg]\end{aligned}\right\} =: \mathrm{LHS}$$

$$\left.\begin{aligned}\le\ &\Gamma_1(R)\mathbb{E}\Bigg[\sum_{n=0}^{N^{(t)}-2}\int_{t_n}^{t_{n+1}}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_n}\Big]\mathbf{1}_{\{N^{(t)}>n+1\}}dr \\ &\qquad + \int_{t_{N^{(t)}-1}}^{t}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big]dr\Bigg]\end{aligned}\right\} =: \mathrm{R}_1$$

$$+\mathbb{E}\Bigg[\overline{\Gamma}_2^{\{4(q+2)\}}(R)\Bigg(\sum_{n=0}^{N^{(t)}-2} h_{n+1}^3\mathbf{1}_{\{N^{(t)}>n+1\}} + \big|t - t_{N^{(t)}-1}\big|^3\Bigg)\Bigg]\ \Big\} =: \mathrm{R}_2 \tag{3.84}$$

where we analyse (3.84) (LHS $\le$ R$_1$+R$_2$) below. For the LHS in (7.51), $N^{(t)}$ is a random number taking value from $N_{\min}^{(t)}$ to $N_{\max}^{(t)}$, and $\mathbf{1}_{\{N^{(t)}>n+1\}}$ is an $\mathcal{F}_{t_n}$-measurable random variable. Therefore it is useful decompose the range of n into three parts on each trajectory. First, when $n < N^{(t)} - 1$, then $1_{\{N^{(t)}>n+1\}} = 1_{\{N^{(t)}>n\}} = 1$. Second, when $n = N^{(t)} - 1$, then $1_{\{N^{(t)}>n+1\}} = 0$ and $1_{\{N^{(t)}>n\}} = 1$. Finally, when $n > N^{(t)} - 1$, then $1_{\{N^{(t)}>n+1\}} = 1_{\{N^{(t)}>n\}} = 0$. Hence we obtain a telescoping sum with the appropriate cancellation that terminates at $\mathbb{E}\big[\|\widetilde{E}(t_{N^{(t)}-1})\|^2\, 1_{\{N^{(t)}>N^{(t)}-1\}}\big] = \mathbb{E}\big[\|\widetilde{E}(t_{N^{(t)}-1})\|^2\big]$. Applying this with the tower property for conditional expectations, and using the fact that $\|\widetilde{E}(t_0)\|^2 = 0$, we have

$$\begin{aligned}\mathrm{LHS} &= \sum_{n=0}^{N_{\max}^{(t)}-2}\mathbb{E}\Big[\big\|\widetilde{E}(t_{n+1})\big\|^2\mathbf{1}_{\{N^{(t)}>n+1\}} - \big\|\widetilde{E}(t_n)\big\|^2\mathbf{1}_{\{N^{(t)}>n+1\}}\Big] \\ &\qquad + \mathbb{E}\Big[\mathbb{E}\Big[\big\|\widetilde{E}(t)\big\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big] - \big\|\widetilde{E}(t_{N^{(t)}-1})\big\|^2\Big] \\ &= \mathbb{E}\Big[\big\|\widetilde{E}(t_{N^{(t)}-1})\big\|^2\Big] - \mathbb{E}\Big[\big\|\widetilde{E}(t_0)\big\|^2\Big] + \mathbb{E}[\|\widetilde{E}(t)\|^2] - \mathbb{E}\Big[\big\|\widetilde{E}(t_{N^{(t)}-1})\big\|^2\Big] \\ &= \mathbb{E}\Big[\big\|\widetilde{E}(t)\big\|^2\Big].\end{aligned} \tag{3.85}$$

For R_1 in (3.84), since by Definition 3.2.1 $n = N^{(r)} - 1$, we first write the condition $\mathcal{F}_{t_n}$ as $\mathcal{F}_{t_{N^{(r)}-1}}$, then the indicator function as $\mathbf{1}_{\{N^{(t)}>N^{(r)}\}}$. By summing up all the steps, we have an integration from 0 to $t_{N^{(t)}-1}$ that

$$\begin{aligned}\mathrm{R}_1 = \Gamma_1(R)\mathbb{E}\Bigg[\int_0^{t_{N^{(t)}-1}} \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2 \mathbf{1}_{\{N^{(t)}>N^{(r)}\}}\Big|\mathcal{F}_{t_{N^{(r)}-1}}\Big]dr \\ + \int_{t_{N^{(t)}-1}}^{t} \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big]dr\Bigg] \\ \leq \Gamma_1(R)\int_0^t \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big]dr. \end{aligned} \tag{3.86}$$

For R_2 in (3.84), by (3.81), Definition 3.2.1 and $\rho h_{\min} = h_{\max}$ we have

$$\mathrm{R}_2 \leq \Gamma_2(R) N_{\max}^{(t)} h_{\max}^3 \leq \Gamma_2(R)\,(\rho t + 1)\, h_{\max}^2. \tag{3.87}$$

We see that $4(q+2)$ is the minimum number of finite SDE moments required for a finite R_2, and this is guaranteed by Assumption 3.1.2. Combining (3.85), (3.86) and (3.87) back into (3.84), for all $t \in [0, T]$ we have

$$\mathbb{E}\Big[\big\|\widetilde{E}(t)\big\|^2\Big] \;\leq\; \Gamma_1(R)\int_0^t \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big]dr + \Gamma_2(R)\,(\rho t + 1)\, h_{\max}^2.$$

By Gronwall's inequality (see [40, Thm. 8.1]), we have for all $t \in [0, T]$

$$\left(\mathbb{E}\Big[\big\|\widetilde{E}(t)\big\|^2\Big]\right)^{\frac{1}{2}} \leq C(R, \rho, t)\, h_{\max}. \tag{3.88}$$

Taking the maximum over t on the both sides, with $\Gamma_1(R)$ and $\Gamma_2(R)$ as defined in (3.80) and (3.81) respectively, the proof follows with

$$C(R, \rho, T) := \sqrt{\Gamma_2(R)\,(\rho T + 1)\exp\big(T\Gamma_1(R)\big)}. \tag{3.89}$$

□

3.3.4 Proof of Theorem 3.3.2 on the probability of using the backstop.

Proof. By (3.16) and by the Markov inequality we have

$$\mathbb{P}\left[h_{n+1}=h_{\min}\right]=\mathbb{P}\left[\frac{h_{\max}}{\left\|\widetilde{Y}(t_n)\right\|^{1/\kappa}}\le h_{\min}\right]\le\frac{\mathbb{E}\left[\left\|\widetilde{Y}(t_n)\right\|^2\right]}{\rho^{2\kappa}}. \tag{3.90}$$

By adding in and subtracting out $X(t_n)$ together with the tower property of conditional expectation, (2.34), (3.32) and (3.6) we have

$$\begin{aligned}\mathbb{E}\left[\left\|\widetilde{Y}(t_n)\right\|^2\right]&\le 2\mathbb{E}\left[\left\|X(t_n)-\widetilde{Y}(t_n)\right\|^2\right]+2\mathbb{E}\left[\|X(t_n)\|^2\right]\\&\le 2\mathbb{E}\left[\mathbb{E}\left[\left\|X(t_n)-\widetilde{Y}(t_n)\right\|^2\middle|\mathcal{F}_{t_{n-1}}\right]\right]+2\mathbb{E}\left[\sup_{t_n\in[0,T]}\|X(t_n)\|^2\right]\\&\le 2\mathbb{E}\left[\mathbb{E}\left[\left\|\widetilde{E}(t_n)\right\|^2\middle|\mathcal{F}_{t_{n-1}}\right]\right]+2C_{\mathtt{X}}.\end{aligned} \tag{3.91}$$

Next, we repeatedly substitute (3.79) into the RHS of (3.91) for decreasing values of n until $n=0$. Then by Definition 3.2.1, (3.16) and (3.81) we have

$$\begin{aligned}\mathbb{E}\left[\left\|\widetilde{Y}(t_n)\right\|^2\right]&\le 2\mathbb{E}\left[\left\|\widetilde{E}(t_{n-1})\right\|^2\right]+2\Gamma_1(R)\mathbb{E}\left[\int_{t_{n-1}}^{t_n}\mathbb{E}\left[\left\|\widetilde{E}(r)\right\|^2\middle|\mathcal{F}_{t_{n-1}}\right]dr\right]\\&\quad+2\mathbb{E}\left[\overline{\Gamma}_2^{\{4(q+2)\}}(R)h_n^3\right]+2C_{\mathtt{X}}\\&\le 2\mathbb{E}\left[\left\|\widetilde{E}(t_0)\right\|^2\right]+2\Gamma_1(R)\mathbb{E}\left[\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\mathbb{E}\left[\left\|\widetilde{E}(r)\right\|^2\middle|\mathcal{F}_{t_{j-1}}\right]dr\right]\\&\quad+2N_{\max}^{(T)}\Gamma_2(R)h_{\max}^3+2C_{\mathtt{X}}\\&\le 2\Gamma_1(R)\mathbb{E}\left[\int_0^{t_n}\mathbb{E}\left[\left\|\widetilde{E}(r)\right\|^2\middle|\mathcal{F}_{t_{N(r)-1}}\right]dr\right]\\&\quad+2\left(\rho T+1\right)\Gamma_2(R)h_{\max}^3+2C_{\mathtt{X}}.\end{aligned} \tag{3.92}$$

Since the integrand $\mathbb{E}\left[\left\|\widetilde{E}(r)\right\|^2\middle|\mathcal{F}_{t_{N(r)-1}}\right]$ in the second term on the RHS of (3.92) is almost surely non-negative for all $r\in[0,T]$, we can replace the upper limit of integration with T. Using $\widetilde{E}(t_0)=0$, (3.16), the tower property of conditional expectation, and (3.88) from Theorem 3.3.1, we have

$$\mathbb{E}\left[\int_0^{t_n} \mathbb{E}\left[\left\|\widetilde{E}(r)\right\|^2 \Big| \mathcal{F}_{t_{N(r)-1}}\right] dr\right]$$
$$\leq \int_0^T \mathbb{E}\left[\left\|\widetilde{E}(r)\right\|^2\right] dr \leq T \max_{r\in[0,T]} \mathbb{E}\left[\left\|\widetilde{E}(r)\right\|^2\right] \leq T\, C^2(R,\rho,T) h_{\max}^2. \tag{3.93}$$

By choosing $h_{\max} \leq 1/C(R,\rho,T)$, we substitute (3.93) into (3.92) and then (3.90) to get the desired result in (3.20) with

$$C_{\text{prob}} := 2\Big(\Gamma_1(R)T + (T+1)\,\Gamma_2(R)h_{\max}^3 + C_{\mathtt{X}}\Big), \tag{3.94}$$

where $\Gamma_1(R)$, $\Gamma_2(R)$ and $C_{\mathtt{X}}$ as defined in (3.80), (3.81) and (3.6) respectively. □

Chapter 4

Jump-adapted adaptive Milstein method (JAAM)

In this chapter, we introduce JAAM for approximating SDEs driven by Poisson random measure (1.2) and prove its L_2 strong convergence of order one. We first state the assumptions in Section 3.1, then the jump-adapted adaptive time-stepping strategies in Section 4.2. Finally, the L_2 strong convergence and its proof are in Section 4.3.

4.1 Assumptions

We now present our assumptions on f, g_i and γ. Since the jump-adapted adaptive Milstein method is based on the adaptive Milstein method in Chapter 3, so Assumptions 3.1.1 and 3.1.2 hold through out this chapter. Detailed discussion on Assumption 3.1.2 can be found in Remark 6. In addition, we assume for jump coefficient γ that

Assumption 4.1.1. *Let $f \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ and $g \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^{d\times m})$ with each $g_i(x) = [g_{1,i}(x), \ldots, g_{d,i}(x)]^T \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ for $i = 1, \ldots, d$. Let $\gamma \in \mathrm{C}^2\big((\mathbb{R}^d \backslash \{0\}) \times \mathbb{R}^d, \mathbb{R}^d\big)$. Let $x, y \in \mathbb{R}^d$ with $\|x\| \vee \|y\| \leq \varkappa$ and for $\varkappa > 1$, there exists a constant $L(\varkappa) < \infty$ such that*

$$2\big\langle x - y, f(x) - f(y)\big\rangle + \big\|g(x) - g(y)\big\|^2_{\mathbf{F}(d\times m)} + \int_Z \big\|\gamma(z, x) - \gamma(z, y)\big\|^2 \nu(dz) \leq L(\varkappa)\big\|x - y\big\|^2. \quad (4.1)$$

Furthermore, there exist a constant $L(1)$ such that

$$\int_Z \big\|\gamma(z,x) - \gamma(z,y)\big\|^2 \nu(dz) \leq L(1)\big\|x - y\big\|^2. \tag{4.2}$$

Under Assumption 4.1.1 and (3.2) from Assumption 3.1.1, the SDE driven by Poisson random measure (1.2) has a unique strong solution on any interval $[0,T]$, where $T < \infty$, see [20, Thm. 2].

Furthermore, by (3.2) and (4.2) we have the following corollary, which is useful for the lemma later.

Corollary 4.1.0.1. *For $x \in \mathbb{R}^d$*

$$\big\langle x, f(x)\big\rangle + \frac{\eta - 3}{2}\big\|g(x)\big\|^2_{\mathbf{F}(d\times m)} \leq L_0 + L_1\big\|x\big\|^2 \tag{4.3}$$

$$\int_Z \big\|\gamma(z,x)\big\|^2 \nu(dz) \leq L_2 + 2L(1)\big\|x\big\|^2, \tag{4.4}$$

where $L_0 := \|f(0)\|^2/2 + (\eta-1)(\eta-5)\sum_{i=1}^m \|g_i(0)\|^2_{\mathbf{F}(d\times m)}/8$, $L_1 := L + 1/2$ and $L_2 := \int_Z \|\gamma(z,0)\|^2\nu(dz)$, with η and L from (3.2)*.*

Proof. For (4.3), by (3.2) with $y = 0$ we have

$$\begin{aligned}\big\langle x, f(x) - f(0)\big\rangle + \frac{\eta-1}{2}\big\|g(x) - g(0)\big\|^2_{\mathbf{F}(d\times m)} &\leq L\big\|x\big\|^2 \\ \big\langle x, f(x)\big\rangle + \frac{\eta-1}{2}\big\|g(x) - g(0)\big\|^2_{\mathbf{F}(d\times m)} &\leq L\big\|x\big\|^2 + \big\langle x, f(0)\big\rangle.\end{aligned}$$

Expanding the g part on the LHS, we have

$$\begin{aligned}\big\langle x, f(x)\big\rangle + \frac{\eta-1}{2}\Big(\big\|g(x)\big\|^2_{\mathbf{F}(d\times m)} - 2\big\langle g(x), g(0)\big\rangle + \big\|g(0)\big\|^2_{\mathbf{F}(d\times m)}\Big) \\ \leq\ L\big\|x\big\|^2 + \big\langle x, f(0)\big\rangle.\end{aligned}$$

By rearranging both sides we have

$$\begin{aligned}\big\langle x, f(x)\big\rangle + \frac{\eta-1}{2}\big\|g(x)\big\|^2_{\mathbf{F}(d\times m)} \leq L\big\|x\big\|^2 + \big\langle x, f(0)\big\rangle + (\eta-1)\big\langle g(x), g(0)\big\rangle \\ - \frac{\eta-1}{2}\big\|g(0)\big\|^2_{\mathbf{F}(d\times m)}.\end{aligned}$$

For the 3rd term on the RHS, by Cauchy–Schwarz inequality (2.31) we have

$$
\begin{aligned}
(\eta-1)\big\langle g(x), g(0)\big\rangle &\leq 2\|g(x)\|_{\mathbf{F}(d\times m)}\frac{\eta-1}{2}\|g(0)\|_{\mathbf{F}(d\times m)} \\
&\leq \|g(x)\|^2_{\mathbf{F}(d\times m)} + \frac{(\eta-1)^2}{8}\|g(0)\|^2_{\mathbf{F}(d\times m)}.
\end{aligned}
$$

Substituting back we have

$$
\begin{aligned}
\big\langle x, f(x)\big\rangle + \frac{\eta-3}{2}\|g(x)\|^2_{\mathbf{F}(d\times m)} \leq \left(L+\frac{1}{2}\right)\|x\|^2 + \frac{1}{2}\|f(0)\|^2 \\
+ \frac{(\eta-1)(\eta-5)}{8}\|g(0)\|^2_{\mathbf{F}(d\times m)},
\end{aligned}
$$

so that the form follows in (4.3). For (4.4), by (4.2) with $y=0$ we have

$$
\int_Z \big\|\gamma(z,x)-\gamma(z,0)\big\|^2\nu(dz) \leq L(1)\|x\|^2.
$$

Expanding the LHS, we have

$$
\int_Z \big\|\gamma(z,x)\big\|^2\nu(dz) - 2\int_Z \big\langle\gamma(z,x),\gamma(z,0)\big\rangle\nu(dz) + \int_Z \big\|\gamma(z,0)\big\|^2\nu(dz) \leq L(1)\|x\|^2.
$$

By rearranging both sides with (2.31) we have

$$
\begin{aligned}
\int_Z \big\|\gamma(z,x)\big\|^2\nu(dz) &\leq L(1)\|x\|^2 + 2\int_Z \big\langle\gamma(z,x),\gamma(z,0)\big\rangle\nu(dz) + \int_Z \big\|\gamma(z,0)\big\|^2\nu(dz) \\
&\leq L(1)\|x\|^2 + \frac{1}{2}\int_Z \big\|\gamma(z,x)\big\|^2\nu(dz) \\
&\quad + \frac{3}{2}\int_Z \big\|\gamma(z,0)\big\|^2\nu(dz) - \int_Z \big\|\gamma(z,0)\big\|^2\nu(dz).
\end{aligned}
$$

Rearranging both sides we have

$$
\int_Z \big\|\gamma(z,x)\big\|^2\nu(dz) \leq 2L(1)\|x\|^2 + \int_Z \big\|\gamma(z,0)\big\|^2\nu(dz),
$$

so the form follows in (4.4). □

Moreover the following moment bounds apply over any finite interval $[0,T]$:

Lemma 4.1.1. *[9, Lem. 3.5] Let* (3.1),(3.2), (4.1) *and* (4.2) *hold. Then the SDE driven by Poisson random measure* (1.2) *has a unique global solution* $X^J(t)$ *such* *that for any* $p \in (2, \eta)$,

$$\sup_{0 \le t \le T} \mathbb{E}\big[\|X^J(t)\|^p\big] < \infty, \quad \forall T < \infty.$$

Similar to Lemma 3.1.1, we require a stronger bound on the moments with the extra condition (4.2) for our convergence proof. The number of moments we get are restricted by the set of parameters, so we keep track of this through the constant C_X^J. It is inspired by [41, Lem. 4.2] and [8, Lem. 2.2].

Lemma 4.1.2. *Let* f, g *and* γ *satisfy* (3.1),(3.2), (4.1), (4.2) *and* (3.4). *Then the SDE driven by Poisson random measure* (1.2) *has a unique global solution such that there exists a constant*

$$\mathbb{E}\Bigg[\sup_{s \in [0,T]} \|X^J(s)\|^{\eta - 2q_2 - 2}\Bigg] \le C_X^J, \tag{4.5}$$

with $C_X^J := C_X^J(s, X_0, L_0, L_1, L_2, L(1), \eta, q_2, \lambda)$.

Remark 6 (**Moment bound**)**.** The moment in (4.5) only depends on η and q_2 because the diffusion coefficient g in (1.2) (due to (1.1)) is superlinearly bounded whereas the the jump coefficient γ in (4.2) is not, see Assumption 3.1.1 and 4.1.1. Furthermore, since the moment in (4.5) is the same as the one in (3.6), Assumption 3.1.2 in Chapter 3 can be and is assumed directly in this chapter.

Now we show the proof of Lemma (4.1.2).

Proof. Let us first define the stopping time $\pi_R := \inf\{t \ge 0 : \|X^J(t)\| > R\} \wedge T$, and notice that $\|X^J(t^-)\| \le R$ for $0 \le t \le \pi_R$. By Itô's formula in (2.18), we have almost surely for any $t \in [0, \bar{t}\,]$, $\bar{p} = p - 2q_2 - 2$ and $p \in (2, \eta)$ that

$$\begin{aligned}
\|X^J(t)\|^{\bar{p}} =& \|X_0^J\|^{\bar{p}} + \bar{p}\int_0^t \|X^J(s)\|^{\bar{p}-2}\big\langle X^J(s), f(X^J(s))\big\rangle ds \\
&+ \bar{p}\sum_{i=1}^m \int_0^t \|X^J(s)\|^{\bar{p}-2}\big\langle X^J(s), g_i(X^J(s))\big\rangle dW_i(s) \\
&+ \frac{\bar{p}(\bar{p}-2)}{2}\sum_{i=1}^m \int_0^t \|X^J(s)\|^{\bar{p}-4}\big\|g_i(X^J(s))^T X^J(s)\big\|^2 ds \\
&+ \int_0^t \int_Z \Big\{\big\|X^J(s^-) + \gamma(z, X^J(s^-))\big\|^{\bar{p}} - \|X^J(s^-)\|^{\bar{p}}\Big\} J_\nu(dz \times ds).
\end{aligned} \tag{4.6}$$

By Taylor's theorem we expand $\left\|X^{\mathtt{J}}(s^-)+\gamma(z,X^{\mathtt{J}}(s^-))\right\|^{\bar p}$ around $\|X^{\mathtt{J}}(s^-)\|^{\bar p}$, and by (2.34) and Young's inequality, we have

$$\begin{aligned}
&\left\|X^{\mathtt{J}}(s^-)+\gamma(z,X^{\mathtt{J}}(s^-))\right\|^{\bar p}-\|X^{\mathtt{J}}(s^-)\|^{\bar p}\\
&=\bar p\int_0^1\left\|X^{\mathtt{J}}(s^-)+\epsilon\gamma(z,X^{\mathtt{J}}(s^-))\right\|^{\bar p-1}\gamma(z,X^{\mathtt{J}}(s^-))d\epsilon\\
&\le 2^{\bar p-2}\bar p\Big(\left\|X^{\mathtt{J}}(s^-)\right\|^{\bar p-1}\left\|\gamma(z,X^{\mathtt{J}}(s^-))\right\|+\left\|\gamma(z,X^{\mathtt{J}}(s^-))\right\|^{\bar p}\Big)\\
&\le 2^{\bar p-2}\bar p\Big(\frac{\bar p-1}{\bar p}\left\|X^{\mathtt{J}}(s^-)\right\|^{\bar p}+\frac{1}{\bar p}\left\|\gamma(z,X^{\mathtt{J}}(s^-))\right\|^{\bar p}+\left\|\gamma(z,X^{\mathtt{J}}(s^-))\right\|^{\bar p}\Big)\\
&\le 2^{\bar p-2}(\bar p+1)\Big(\left\|X^{\mathtt{J}}(s^-)\right\|^{\bar p}+\left\|\gamma(z,X^{\mathtt{J}}(s^-))\right\|^{\bar p}\Big).
\end{aligned}\tag{4.7}$$

Substituting (4.7) back into (4.6), we have

$$\begin{aligned}
\|X^{\mathtt{J}}(t)\|^{\bar p}\le\|X^{\mathtt{J}}_0\|^{\bar p}&+\bar p\int_0^t\|X^{\mathtt{J}}(s)\|^{\bar p-2}\Big\{\big\langle X^{\mathtt{J}}(s),f(X^{\mathtt{J}}(s))\big\rangle+\frac{\bar p-2}{2}\left\|g(X^{\mathtt{J}}(s))\right\|^2_{\mathbf{F}}\Big\}ds\\
&+\bar p\sum_{i=1}^m\int_0^t\|X^{\mathtt{J}}(s)\|^{\bar p-2}\big\langle X^{\mathtt{J}}(s),g_i(X^{\mathtt{J}}(s))\big\rangle dW_i(s)\\
&+2^{\bar p-2}(\bar p+1)\int_0^t\int_Z\Big(\left\|X^{\mathtt{J}}(s^-)\right\|^{\bar p}+\left\|\gamma(z,X^{\mathtt{J}}(s^-))\right\|^{\bar p}\Big)J_\nu(dz\times ds).
\end{aligned}\tag{4.8}$$

For the second term on the RHS, since $\eta\ge\bar p+1$ by (4.3) and Young's inequality we have

$$\begin{aligned}
&\bar p\int_0^t\|X^{\mathtt{J}}(s)\|^{\bar p-2}\Big\{\big\langle X^{\mathtt{J}}(s),f(X^{\mathtt{J}}(s))\big\rangle+\frac{\bar p-2}{2}\left\|g(X^{\mathtt{J}}(s))\right\|^2_{\mathbf{F}}\Big\}ds\\
\le&\bar p\int_0^t\|X^{\mathtt{J}}(s)\|^{\bar p-2}\Big\{L_0+L_1\|X^{\mathtt{J}}(s)\|^2\Big\}ds\\
\le&\frac{2}{\bar p}L_0^{\bar p/2}+\bar p\left(\frac{\bar p-2}{\bar p}+L_1\right)\int_0^t\|X^{\mathtt{J}}(s)\|^{\bar p}ds.
\end{aligned}\tag{4.9}$$

Substituting (4.9) back into (4.8) we have

$$\begin{aligned}
\|X^{\mathtt{J}}(t)\|^{\bar p}\le&\|X^{\mathtt{J}}_0\|^{\bar p}+\frac{2}{\bar p}L_0^{\bar p/2}+\bar p\left(\frac{\bar p-2}{\bar p}+L_1\right)\int_0^t\|X^{\mathtt{J}}(s)\|^{\bar p}ds\\
&+\bar p\sum_{i=1}^m\int_0^t\|X^{\mathtt{J}}(s)\|^{\bar p-2}\big\langle X^{\mathtt{J}}(s),g_i(X^{\mathtt{J}}(s))\big\rangle dW_i(s)\\
&+2^{\bar p-2}(\bar p+1)\int_0^t\int_Z\Big(\left\|X^{\mathtt{J}}(s^-)\right\|^{\bar p}+\left\|\gamma(z,X^{\mathtt{J}}(s^-))\right\|^{\bar p}\Big)J_\nu(dz\times ds).
\end{aligned}$$

Taking the supremum over $[0,\bar t\wedge\pi_R]$ for $\bar t\in[0,T]$ and expectations on the both sides we have

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\le t\le \bar{t}\wedge\pi_R}\|X^{\mathrm{J}}(t)\|^{\bar{p}}\right]\\
\le&\|X_0^{\mathrm{J}}\|^{\bar{p}}+\frac{2}{\bar{p}}L_0^{\bar{p}/2}+\bar{p}\left(\frac{\bar{p}-2}{\bar{p}}+L_1\right)\mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\|X^{\mathrm{J}}(s)\|^{\bar{p}}ds\right]\\
&+\bar{p}\sum_{i=1}^{m}\mathbb{E}\left[\sup_{0\le t\le \bar{t}\wedge\pi_R}\left|\int_0^t\|X^{\mathrm{J}}(s)\|^{\bar{p}-2}\big\langle X^{\mathrm{J}}(s),g_i(X^{\mathrm{J}}(s))\big\rangle dW_i(s)\right|\right]\\
&+2^{\bar{p}-2}(\bar{p}+1)\mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\int_Z\left(\left\|X^{\mathrm{J}}(s^-)\right\|^{\bar{p}}+\left\|\gamma(z,X^{\mathrm{J}}(s^-))\right\|^{\bar{p}}\right)J_\nu(dz\times ds)\right].
\end{aligned}
\tag{4.10}
$$

For the fourth term on the RHS, we use the Burkholder-Davis-Gundy inequality (see [40, Thm. 7.3]) and (2.31) to have

$$
\begin{aligned}
&\bar{p}\sum_{i=1}^{m}\mathbb{E}\left[\sup_{0\le t\le \bar{t}\wedge\pi_R}\left|\int_0^t\|X^{\mathrm{J}}(s)\|^{\bar{p}-2}\big\langle X^{\mathrm{J}}(s),g_i(X^{\mathrm{J}}(s))\big\rangle dW_i(s)\right|\right]\\
\le&\bar{p}\sum_{i=1}^{m}\mathbb{E}\left[\left(\int_0^{\bar{t}\wedge\pi_R}\|X^{\mathrm{J}}(s)\|^{2(\bar{p}-1)}\|g_i(X^{\mathrm{J}}(s))\|^2ds\right)^{1/2}\right]\\
\le&\bar{p}\sum_{i=1}^{m}\mathbb{E}\left[\left(\sup_{0\le t\le \bar{t}\wedge\pi_R}\|X^{\mathrm{J}}(t)\|^{\bar{p}}\int_0^{\bar{t}\wedge\pi_R}\|X^{\mathrm{J}}(s)\|^{\bar{p}-2}\|g_i(X^{\mathrm{J}}(s))\|^2ds\right)^{1/2}\right]\\
\le&\frac{1}{2}\mathbb{E}\left[\sup_{0\le t\le \bar{t}\wedge\pi_R}\|X^{\mathrm{J}}(t)\|^{\bar{p}}\right]+\frac{\bar{p}^2}{2}\sum_{i=1}^{m}\mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\|X^{\mathrm{J}}(s)\|^{\bar{p}-2}\|g_i(X^{\mathrm{J}}(s))\|^2ds\right].
\end{aligned}
\tag{4.11}
$$

Then with the setting that $\bar{p}=p+2q_2-2$, and by (3.4), Young's inequality and Lemma 4.1.1 we have the second term in (4.11) as

$$
\begin{aligned}
&\frac{\bar{p}^2}{2}\sum_{i=1}^{m}\mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\|X^{\mathrm{J}}(s)\|^{\bar{p}-2}\|g_i(X^{\mathrm{J}}(s))\|^2ds\right]\\
\le&\ \frac{\bar{p}^2c^2m}{2}\mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\|X^{\mathrm{J}}(s)\|^{p-2(q_2+2)}\big(1+\|X^{\mathrm{J}}(s)\|^{q_2+2}\big)^2ds\right]\\
\le&\ \bar{p}^2c^2m\mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\Big(\frac{p-2(q_2+2)}{p}\|X^{\mathrm{J}}(s)\|^p+\frac{2(q_2+2)}{p}1^{p/2(q_2+2)}+\|X^{\mathrm{J}}(s)\|^p\Big)ds\right]\\
\le&\ \bar{p}^2c^2m\left(\frac{2(q_2+2)}{p}+\frac{2p-2(q_2+2)}{p}\sup_{0\le s\le T}\mathbb{E}\big[\|X^{\mathrm{J}}(s)\|^p\big]\right)|T\wedge\pi_R-0|\\
=:&\ C_1.
\end{aligned}
\tag{4.12}
$$

Substituting (4.12) back into (4.11) and returning to (4.10) we have

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\le t\le \bar{t}\wedge\pi_R}\|X^{\mathsf{J}}(t)\|^{\bar{p}}\right]\\
\le&\|X_0^{\mathsf{J}}\|^{\bar{p}}+\frac{2}{\bar{p}}L_0^{\bar{p}/2}+\bar{p}\left(\frac{\bar{p}-2}{\bar{p}}+L_1\right)\int_0^{\bar{t}\wedge\pi_R}\mathbb{E}\big[\|X^{\mathsf{J}}(s)\|^{\bar{p}}\big]ds\\
&+2^{\bar{p}-2}(\bar{p}+1)\mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\int_Z\left(\left\|X^{\mathsf{J}}(s)\right\|^{\bar{p}}+\left\|\gamma(z,X^{\mathsf{J}}(s))\right\|^{\bar{p}}\right)\nu(dz)ds\right]\\
&+\frac{1}{2}\mathbb{E}\left[\sup_{0\le t\le \bar{t}\wedge\pi_R}\|X^{\mathsf{J}}(t)\|^{\bar{p}}\right]+C_1. \qquad (4.13)
\end{aligned}
$$

Note for 3rd term on the RHS of (4.13), by (4.4) we have

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\int_Z\left(\left\|X^{\mathsf{J}}(s^-)\right\|^{\bar{p}}+\left\|\gamma(z,X^{\mathsf{J}}(s^-))\right\|^{\bar{p}}\right)J_\nu(dz\times ds)\right]\\
=&\ \mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\int_Z\left(\left\|X^{\mathsf{J}}(s^-)\right\|^{\bar{p}}+\left\|\gamma(z,X^{\mathsf{J}}(s^-))\right\|^{\bar{p}}\right)\nu(dz)ds\right]\\
\le&\ \mathbb{E}\left[\int_0^{\bar{t}\wedge\pi_R}\left(\lambda\left\|X^{\mathsf{J}}(s^-)\right\|^{\bar{p}}+2^{\bar{p}/2-1}L_2^{\bar{p}/2}+2^{\bar{p}/2}L(1)^{\bar{p}/2}\|X^{\mathsf{J}}(s^-)\|^{\bar{p}}\right)ds\right]\\
\le&\ \left(\lambda+2^{\bar{p}/2}L(1)^{\bar{p}/2}\right)\int_0^{\bar{t}\wedge\pi_R}\mathbb{E}\left[\left\|X^{\mathsf{J}}(s^-)\right\|^{\bar{p}}\right]ds+C_2, \qquad (4.14)
\end{aligned}
$$

where $C_2=2^{\bar{p}/2-1}L_2^{\bar{p}/2}|\bar{t}\wedge\pi_R-0|$. Since it holds that

$$
\mathbb{E}\left[\sup_{0\le t\le s\wedge\pi_R}\left\|X^{\mathsf{J}}(s^-)\right\|^{\bar{p}}\right]\le\mathbb{E}\left[\sup_{0\le t\le s\wedge\pi_R}\left\|X^{\mathsf{J}}(s))\right\|^{\bar{p}}\right],
$$

substituting (4.14) back into (4.13) we have

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0\le t\le \bar{t}\wedge\pi_R}\|X^{\mathsf{J}}(t)\|^{\bar{p}}\right]\le&\ \frac{1}{2}\mathbb{E}\left[\sup_{0\le t\le \bar{t}\wedge\pi_R}\|X^{\mathsf{J}}(t)\|^{\bar{p}}\right]\\
&+C_4\mathbb{E}\left[\int_0^{\bar{t}}\sup_{0\le t\le s\wedge\pi_R}\left\|X^{\mathsf{J}}(s)\right\|^{\bar{p}}ds\right]+C_3, \qquad (4.15)
\end{aligned}
$$

where

$$
\begin{aligned}
C_3=&\|X_0\|^{\bar{p}}+\frac{2}{\bar{p}}L_0^{\bar{p}/2}+C_1+2^{\bar{p}-1}(\bar{p}+1)C_2,\\
C_4=&\bar{p}\left(\frac{\bar{p}-2}{\bar{p}}+L_1\right)+2^{\bar{p}-1}(\bar{p}+1)\left(\lambda+2^{\bar{p}/2}L(1)^{\bar{p}/2}\right).
\end{aligned}
$$

Rearranging (4.15) and for all $\bar{t} \in [0,T]$ using Gronwall's inequality we have

$$\mathbb{E}\left[\sup_{0\le t\le \bar{t}\wedge\pi_R}\|X^{\mathtt{J}}(t)\|^{\bar{p}}\right] \le 2C_4\mathbb{E}\left[\int_0^{\bar{t}}\sup_{0\le t\le s\wedge\pi_R}\left\|X^{\mathtt{J}}(s)\right\|^{\bar{p}}ds\right] + 2C_3 \le C_{\mathtt{X}}^{\mathtt{J}} < \infty,$$

where $C_{\mathtt{X}}^{\mathtt{J}} := 2C_3\exp\left(2sC_4\right)$. We have (3.6) holds and the proof is complete. □

4.2 Jump-adapted adaptive time-stepping

Consider a d-dimensional SDE driven by Poisson random measure in (1.2) or (2.17). Let solutions be defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge0}, \mathbb{P})$. Let $\{\mathcal{G}_t\}_{t\ge0}$ be the natural filtration of W, $\{\mathcal{H}_t\}_{t\ge0}$ be the natural filtration of J, and $\mathcal{F}_t = \sigma\big(\mathcal{G}_t\cup\mathcal{H}_T\big)$, for all $t\ge0$.

We build JAAM based on the jump-adapted Milstein method as defined in Definition 2.3.3 and the adaptive Milstein method as defined in Definition 3.2.2. Therefore, the settings of JAAM in this chapter are similar to the ones for adaptive time-stepping strategies for SDEs with no jumps in Chapter 3. The corresponding class of jump-adapted time-stepping strategies is defined in Definition 4.2.2. Here we refer to Section 3.2 for detailed definitions and assumptions, with the updates made for jumps specified below.

Let $\{h_{n+1}^{\mathtt{J}}\}_{n\in\mathbb{N}}$ satisfy Assumption 3.2.1 and be a sequence of strictly positive random timesteps with corresponding random times $\{t_n := \sum_{i=1}^n h_i^{\mathtt{J}}\}_{n\in\mathbb{N}\setminus\{0\}}$, where $t_0 = 0$. Lemma 3.2.1 stands consequently i.e. the mesh points $\{t_n\}_{n\in\mathbf{N}}$ for jump-adapted Milstein method in this chapter are also $\{\mathcal{F}_t\}$-stopping times, except that the filtration $\{\mathcal{F}_t\}_{t\ge0}$ now contains information from both W and J. Notice that for $n\in\mathbb{N}$, each jump time τ_{n+1} (defined in Section 2.2) is $\mathcal{H}_T$-measurable so that $\mathcal{F}_{\tau_n}$-measurable. In practice, all jump times and sizes can be precomputed.

Let Definition 3.2.1 be satisfied with $N_{\min}^{(t)} := \lfloor t/h_{\max}\rfloor$ and $N_{\max}^{(t)} := \lceil t/h_{\min} + J(t)\rceil$ where $J(t)$ is the number of jumps on $[0,t]$ that we assume to be a finite random variable, (see Remark 2). Notice that $N_{\max}^{(t)}$ is $\mathcal{H}_T$-measurable so that a $(\mathcal{F}_t)_{t\ge0}$-measurable random variable.

Next, we introduce the JAAM scheme. Similar to Definition 3.2.2, it is characterised by the sequence of tuples $\left\{\big(\widetilde{Y}^{\mathtt{J}}(s)\big)_{s\in[t_n,t_{n+1}]}, h_{n+1}^{\mathtt{J}}\right\}_{n\in\mathbb{N}}$. With an indicator function separating the event $\{h_{n+1}^{\mathtt{J}} \le h_{\min}\}$, we use map φ for the backstop case, and map θ for the other (explicit Milstein).

Definition 4.2.1 (JAAM Scheme). Let $\{h_{n+1}^{\mathtt{J}}\}_{n\in\mathbb{N}}$ satisfy Assumption 3.2.1. Following the structures of the jump-adapted Milstein scheme in Definitions 2.3.3 and the adaptive Milstein scheme in Definition 3.2.2, we define $\widetilde{Y}^{\mathtt{J}}(s)$ as the continuous form of a *JAAM scheme* associated with $\{h_{n+1}^{\mathtt{J}}\}_{n\in\mathbb{N}}$ that

$$\widehat{\mathtt{Y}}^{\mathtt{J}}(s) := \theta\left(\widetilde{Y}^{\mathtt{J}}(t_n),\, t_n,\, s - t_n\right) \cdot \mathbf{1}_{\{h_{\min} < h_{n+1}^{\mathtt{J}} \le h_{\max}\}} + \varphi\left(\widetilde{Y}^{\mathtt{J}}(t_n),\, t_n,\, s - t_n\right) \cdot \mathbf{1}_{\{h_{n+1}^{\mathtt{J}} \le h_{\min}\}}, \tag{4.16}$$

$$\widetilde{Y}^{\mathtt{J}}(s) := \widehat{\mathtt{Y}}^{\mathtt{J}}(s) + \int_Z \gamma\big(z, \widehat{\mathtt{Y}}^{\mathtt{J}}(s)\big) J_\nu(dz \times \{s\}) \tag{4.17}$$

for $s \in [t_n, t_{n+1}]$, $n \in \mathbb{N}$, where $\widetilde{Y}^{\mathtt{J}}(0) = X(0)$ and the measure $J_\nu(dz \times \{s\})$ is defined in (2.16). The process $\widehat{\mathtt{Y}}^{\mathtt{J}}(s)$ in (4.16) stands for the approximation until the left-hand-limit of t_{n+1}, it follows the same structure of (3.10) but with $\widetilde{Y}^{\mathtt{J}}(t_n)$ as the input value and $h_{n+1}^{\mathtt{J}}$ as the timestep. Over each step, the map $\theta : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^d$ in (4.17) follows the explicit Milstein method as defined in (3.9), and the backstop map $\varphi : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^d$ satisfies a mean-square consistency requirement with deterministic step $h_{n+1}^{\mathtt{J}} \le h_{\min}$ with the form of (3.11) for $n \in \mathbb{N}$.

We inherit the notation in Definition 3.2.3 as $Y_\theta^{\mathtt{J}}(s)$ to denote $\widehat{\mathtt{Y}}^{\mathtt{J}}(s)$ in (4.16) when event $\{h_{\min} < h_{n+1}^{\mathtt{J}} \le h_{\max}\}$ occurs so that $\widehat{\mathtt{Y}}^{\mathtt{J}}(s)$ takes map θ, for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$. Next, we define the jump-adapted version of Definition 3.2.4.

Definition 4.2.2 (Jump-adapted path-bounded time-stepping strategies). Let the tuple $\big\{\widetilde{Y}^{\mathtt{J}}(t_n), h_{n+1}^{\mathtt{J}}\big\}_{n\in\mathbb{N}}$ given by (4.17) be a numerical approximation for (1.2), associated with a timestep sequence $\{h_{n+1}^{\mathtt{J}}\}_{n\in\mathbb{N}}$ satisfying Assumption 3.2.1. We say that $\{h_{n+1}^{\mathtt{J}}\}_{n\in\mathbb{N}}$ is a jump-adapted path-bounded time-stepping strategy for JAAM in (4.16)-(4.17) if there exists a real non-negative constant $R > 0$ such that whenever $h_{\min} < h_{n+1}^{\mathtt{J}} \le h_{\max}$,

$$\big\|\widetilde{Y}^{\mathtt{J}}(t_n)\big\| < R, \quad n = 0, \dots, N-1. \tag{4.18}$$

We now provide a specific example of such a strategy in (4.19). It is similar to the path-bounded strategy for adaptive Milstein method in (3.16), but with the extra constraint that is the distance to the next jump, so that the path-bounded strategy is also jump-adapted.

Lemma 4.2.1. *Let* $\left\{\left(\widetilde{Y}^J(s)\right)_{s\in[t_n,t_{n+1}]}, h_{n+1}^J\right\}_{n\in\mathbb{N}}$ *be JAAM scheme given in Definition 4.2.1. Then* $\{h_{n+1}^J\}_{n\in\mathbb{N}}$ *is jump-adapted path-bounded time-stepping strategy (as in Definition 4.2.2) if for each* $n = 0, \ldots, N-1$ *and* $\kappa > 0$*,*

$$h_{n+1}^J = \underbrace{\left(h_{\min} \vee \left(\frac{h_{\max}}{\left\|\widetilde{Y}^J(t_n)\right\|^{1/\kappa}} \wedge h_{\max}\right)\right)}_{\textit{adaptive } h_{n+1}} \wedge \underbrace{\left(\tau_{J(t_n)+1} - t_n\right)}_{\substack{\textit{time to the} \\ \textit{next jump}}}. \tag{4.19}$$

Proof. By Definition 4.2.2, if (4.18) is satisfied when $h_{\min} < h_{n+1}^{\mathsf{J}} \le h_{\max}$, we call the sequence $\{h_{n+1}^{\mathsf{J}}\}_{n\in\mathbb{N}}$ a jump-adapted path-bounded time-stepping strategy. When $h_{n+1}^{\mathsf{J}} \le h_{\min}$ we instead switch to the backstop method with (3.11) holds. Here, we denote the core function of the strategy in (4.19) by $\mathsf{F} := h_{\max}/\left\|\widetilde{Y}^{\mathsf{J}}(t_n)\right\|^{1/\kappa}$, and the distance to next jump time by $\mathsf{D} := \tau_{J(t_n)+1} - t_n$. Therefore, here we only focus on the event $\{h_{\min} < h_{n+1}^{\mathsf{J}} \le h_{\max}\}$, which has 3 outcomes on h_{n+1}^{J}: **case 1** $\left(h_{n+1}^{\mathsf{J}} = \mathsf{F}\right)$, **case 2** $\left(h_{n+1}^{\mathsf{J}} = \mathsf{D}\right)$, and **case 3** $\left(h_{n+1}^{\mathsf{J}} = h_{\max}\right)$.

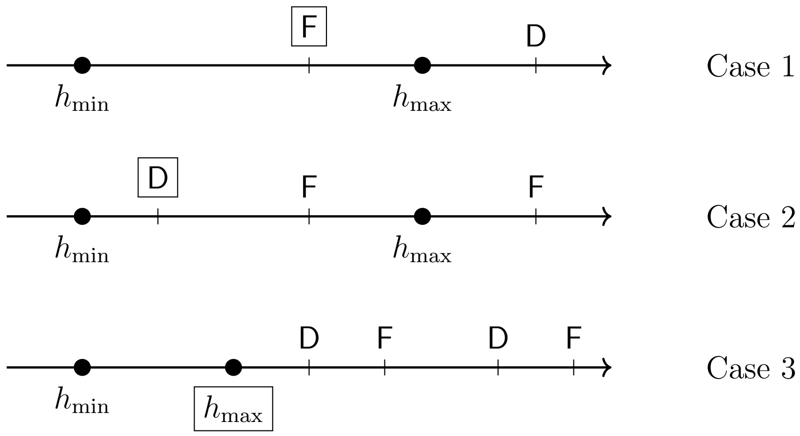

Figure 4.1: Possibilities of the position of the elements in the jump-adapted path-bounded time-stepping strategy

In Figure 4.1, we display all the possibilities of the position of the elements in the jump-adapted path-bounded time-stepping strategy in (4.19), by the 3 cases. To make the event $\{h_{\min} < h_{n+1}^{\mathsf{J}} \le h_{\max}\}$ hold, case 1 takes value F (in box) between $h_{\min}$ and $h_{\max}$, and all together smaller than D, as shown in the first line of Figure 4.1. Similarly, in case 2, D is taken and F can be either smaller or greater than $h_{\max}$. Lastly, in case 3 $h_{\max}$ is taken, so D and F can have any order when both greater than $h_{\max}$. For all possibilities shown, the following holds

$$\mathsf{D} = \frac{h_{\max}}{\left\|\widetilde{Y}^{\mathsf{J}}(t_n)\right\|^{1/\kappa}} > h_{\min}. \tag{4.20}$$

By rearranging (4.20) and with $\rho h_{\min} = h_{\max}$ we have

$$\left\|\widetilde{Y}^{\mathsf{J}}(t_n)\right\| < \left(\frac{h_{\max}}{h_{\min}}\right)^{\kappa} < \left(\frac{h_{\max}}{h_{\min}}\right)^{\kappa} = \rho^{\kappa},$$

so (4.18) is satisfied with $R = \rho^{\kappa}$ and $\{h^{\mathsf{J}}_{n+1}\}_{n\in\mathbb{N}}$ is a jump-adapted path-bounded time-stepping strategy for (4.17). □

Following Lemma 4.2.1, we see in Figure 4.2 the comparison of a non-jump-adapted mesh and a jump-adapted mesh. Assuming that a jump occurs at τ_i after t_n, non-jump-adapted mesh steps it over and lands at t_{n+1}, with the input value of this step taking at t_n (emphasised in box). On the other hand, the jump-adapted mesh stops at the jump time τ_i first and then goes to the next stop, with the approximation at τ_i valued from τ_i^-. Non-jump-adapted mesh might proceed faster as it does not have to track jumps, but small time steps are necessary to prevent from missing too many jumps, and so the accuracy can be ensured. Jump-adapted mesh could achieve better accuracy as every jump is tracked and approximated right before the jump time, but at the cost of more computational time.

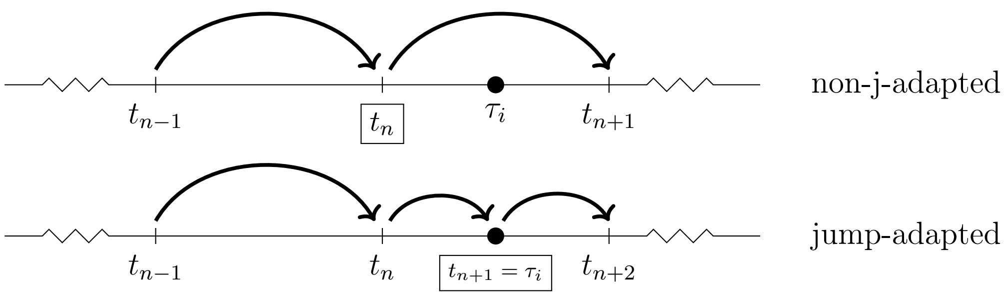

Figure 4.2: Comparison of non-jump-adapted and jump-adapted mesh

Next, we discuss how we update the jump-adapted mesh so that we preserve its advantage i.e. accuracy, and optimize its disadvantage i.e. efficiency. One way is to have adaptive steps rather than fixed steps between the jumps, so that the extra computational time consumed by tracking down each jump can be compensated.

Remark 7 (**Jump-adapted fixed vs Jump-adapted adaptive**)**.** The jump-adapted method introduced in [5] is applied over a mesh that is a superposition of jump times and equidistant time steps. However for JAAM at each step, either the next adaptive landing point or the next jump time is taken, whichever is sooner. Whenever there is a jump, the coefficient γ takes the left-hand limit of jump point as the input value. Additionally, [5] was restricted by the globally Lipschitz conditions on f and g_i but obtained strong convergence results for numerical methods of any order, whereas JAAM requires a monotone condition on f and g_i but only achieved strong convergence of order one for Milstein method.

To explain (4.19) and Remark 7 better, we first show in Figure 4.3 an example of the jump-adapted fixed-step mesh, which is a superposition of the jump times and the equidistant time steps.

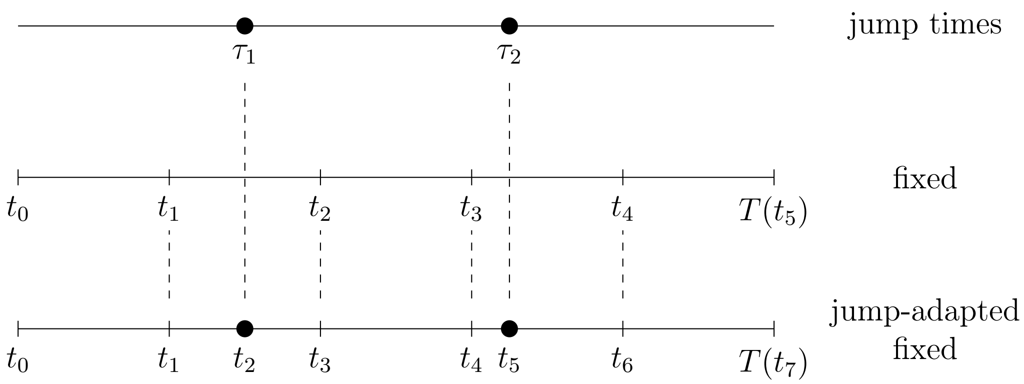

Figure 4.3: Illustration of jump-adapted fixed-step mesh (superposition)

Further, we illustrate the jump-adapted adaptive mesh in Figure 4.4. We can see that the jump times, which are precomputable, are on the first line. Waiting times π_i and jump sizes ζ_i for $i = 1, \ldots, J(T)$ are defined in Section 2.2. In the second line, we see that t_1 and t_2 are adaptive steps without being restricted by jump times, but obviously t_2 is stepping over the first time time τ_1, so we bring back t_2 to the position of τ_2, as shown in the third line. Notice that t_3 is newly calculated based on the value on t_2 (or τ_1) unlike the "t_3" in Figure 4.3, this is one difference compared to the superposition. Similarly, adaptive time t_4 in third time is crossing over the second jump time τ_2, so we bring it back to the jump time, as shown in the fourth line. Again, t_5 is newly calculated based on the approximation value at t_4 unlike the "t_6" in Figure 4.3. This process continues until it passes the last jump time and reaches the terminal time T.

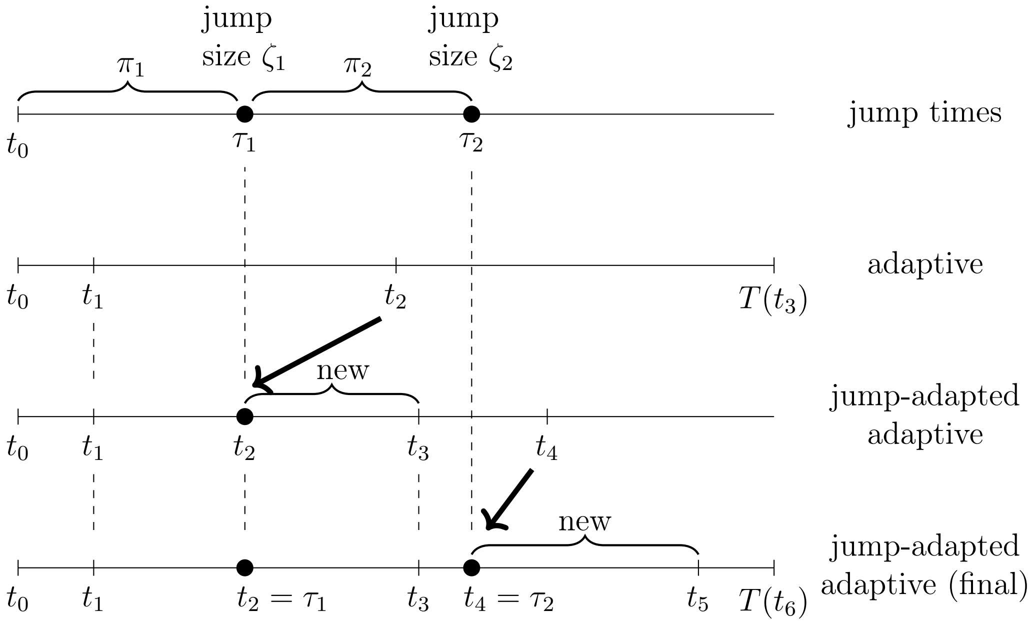

Figure 4.4: Illustration of jump-adapted adaptive mesh

4.3 Result: Theorem 4.3.1

The main result which is the strong convergence of order one is shown in Theorem 4.3.1 below, following by its proof in subsection 4.3.1.

Theorem 4.3.1 (Strong Convergence)**.** *Let* $(X^J(t))_{t\in[0,T]}$ *be a solution of* (1.2) *with initial value* $X^J(0) = X_0^J \in \mathbb{R}^d$*. Suppose that the conditions of Assumptions 3.1.1, 3.1.2 and 4.1.1 hold. Let* $\{(\widetilde{Y}^J(s))_{s\in[t_n,t_{n+1}]}, h_{n+1}^J\}_{n\in\mathbb{N}}$ *be the jump-adapted adaptive Milstein scheme given in Definition 4.2.1 with* $\widetilde{Y}_0^J = X_0^J$ *and* h_1^J *the second item in the jump-adapted path-bounded time-stepping strategy* $\{h_{n+1}^J\}_{n\in\mathbb{N}}$ *satisfying the conditions of Definition 4.2.2 for some* $R < \infty$*. Then there exists a constant* $C^J(R,\rho,T) > 0$ *with its explicit form in* (4.46) *such that*

$$\max_{t\in[0,T]} \left(\mathbb{E}\left[\|X^J(t) - \widetilde{Y}^J(t)\|^2\right]\right)^{1/2} \leq C^J(R,\rho,T)\, h_{\max}.$$

The proof of Theorem 4.3.1 is based on the jump-adapted mesh, so for convenience we rewrite (1.2) as for $s \in [t_n, t_{n+1}]$, $n \in \mathbb{N}$

$$\widehat{\mathtt{X}}^{\mathtt{J}}(s) := X^{\mathtt{J}}(t_n) + \int_{t_n}^{s} f(X^{\mathtt{J}}(r))dr + \sum_{i=1}^{m} \int_{t_n}^{s} g_i(X^{\mathtt{J}}(r))dW_i(r) \tag{4.21}$$

$$X^{\mathtt{J}}(s) = \widehat{\mathtt{X}}^{\mathtt{J}}(s) + \int_Z \gamma\big(z, \widehat{\mathtt{X}}^{\mathtt{J}}(s)\big) J_\nu(dz \times \{s\}). \tag{4.22}$$

Next, we define the one-step error function first. Notice that $\widetilde{Y}^{\mathtt{J}}(s)$, defined in JAAM method (4.17), takes either the Milstein map θ structured in (3.9) or the backstop map φ structured in (3.11), depending on the value of $h^{\mathtt{J}}_{n+1}$. Thus, $\widehat{\mathtt{X}}^{\mathtt{J}}(s)$ and $X^{\mathtt{J}}(s)$ defined in (4.21) and (4.22), similar to the error function for adaptive Milstein as in (3.34) we define the error function for JAAM by

$$\widehat{\mathtt{E}}^{\mathtt{J}}(s) := \widehat{\mathtt{X}}^{\mathtt{J}}(s) - \widehat{\mathtt{Y}}^{\mathtt{J}}(s) = E^{\mathtt{J}}_{\theta}(s) + E^{\mathtt{J}}_{\varphi}(s), \tag{4.23}$$

$$\begin{aligned}\widetilde{E}^{\mathtt{J}}(s) :=& X^{\mathtt{J}}(s) - \widetilde{Y}^{\mathtt{J}}(s)\\ =& \widehat{\mathtt{E}}^{\mathtt{J}}(s) + \int_Z \Delta\gamma\Big(z, \widehat{\mathtt{X}}^{\mathtt{J}}(s), \widehat{\mathtt{Y}}^{\mathtt{J}}(s)\Big) J_\nu(dz \times \{s\})\end{aligned} \tag{4.24}$$

for $s \in [t_n, t_{n+1}]$, $n \in \mathbb{N}$ and the jump coefficient difference defined as

$$\Delta\gamma\Big(z, \widehat{\mathtt{X}}^{\mathtt{J}}(s), \widehat{\mathtt{Y}}^{\mathtt{J}}(s)\Big) := \gamma\Big(z, \widehat{\mathtt{X}}^{\mathtt{J}}(s)\Big) - \gamma\Big(z, \widehat{\mathtt{Y}}^{\mathtt{J}}(s)\Big). \tag{4.25}$$

Here $E^{\mathtt{J}}_{\varphi}(s)$ in (4.24) is the error by taking the backstop map, i.e.

$$E^{\mathtt{J}}_{\varphi}(s) := \Big(\widehat{\mathtt{X}}^{\mathtt{J}}(s) - \varphi\Big(\widetilde{Y}^{\mathtt{J}}(t_n),\, t_n,\, s - t_n\Big)\Big)\mathbf{1}_{\{h^{\mathtt{J}}_{n+1} \le h_{\min}\}}. \tag{4.26}$$

By having $Y^{\mathtt{J}}_{\theta}(s)$ following the form in (3.12) but with $\widetilde{Y}^{\mathtt{J}}(t_n)$ as the input, the error taking the explicit Milstein method is given by

$$\begin{aligned}E^{\mathtt{J}}_{\theta}(s) :=& \Big(\widehat{\mathtt{X}}^{\mathtt{J}}(s) - Y^{\mathtt{J}}_{\theta}(s)\Big)\mathbf{1}_{\{h^{\mathtt{J}}_{n+1} \le h_{\min}\}}\\ =& \Bigg(\widetilde{E}^{\mathtt{J}}(t_n) + \int_{t_n}^{s} \Delta f\Big(X^{\mathtt{J}}(r), \widetilde{Y}^{\mathtt{J}}(t_n)\Big)dr\\ &+ \sum_{i=1}^{m}\int_{t_n}^{s} \Delta g_i\Big(r, X^{\mathtt{J}}(r), \widetilde{Y}^{\mathtt{J}}(t_n)\Big)dW_i(r)\Bigg)\mathbf{1}_{\{h^{\mathtt{J}}_{n+1} \le h_{\min}\}},\end{aligned} \tag{4.27}$$

with the differences from drift and diffusion coefficients: $\Delta f\Big(X^{\mathtt{J}}(r), \widetilde{Y}^{\mathtt{J}}(t_n)\Big)$ and $\Delta g_i\Big(r, X^{\mathtt{J}}(r), \widetilde{Y}^{\mathtt{J}}(t_n)\Big)$ follow the same structure of definitions in (3.35) and (3.36), respectively.

Since the first step of the jump-adaptive method is to approximate (1.2) when $\gamma = 0$, we restate the result of one-step error bound from Lemma 3.3.9 below.

Lemma 4.3.2. *Let f, g satisfy Assumption 3.1.1 and 3.1.2. Let $X^J(s)$ be a solution of* (1.2) *with $\gamma = 0$. Suppose $\widetilde{E}^J(s)$ is a solution of* (4.23) *with $E^J_{\theta}(s)$ defined in* (4.27) *for $s \in [t_n, t_{n+1}]$, $n \in \mathbb{N}$. There exists a constant C_E defined in* (3.45) *and an*

$\mathcal{F}_{t_n}$*-measurable random variable* $C^J_M(R, X^J)$ *such that*

$$\mathbb{E}\Big[\Big\|E^J_\theta(t_{n+1}))\Big\|^2\Big|\mathcal{F}_{t_n}\Big] \leq \;\|\widetilde{E}^J(t_{n+1})\|^2 + C_E(R)\int_{t_n}^{t_{n+1}}\mathbb{E}\left[\|E^J_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\right]dr \\ + C^J_M(R, X^J)\,(h^J_{n+1})^3, \quad a.s. \quad (4.28)$$

For proof see Section 3.3.3. The X dependence in $C^{\mathtt{J}}_M(R, X^{\mathtt{J}})$ means that it requires certain finite moments of the SDE (1.2) when $\gamma = 0$ to have finite expectation (see Remark 5), which is ensured by Assumption 3.1.2. In this chapter, we denote $\mathbb{E}\left[C^{\mathtt{J}}_M(R, X^{\mathtt{J}})\right]$ by $C^{\mathtt{J}}_M(R)$ which has the form of (3.76) in Chapter 3 with $C_{\mathtt{X}}$ being replaced by $C^{\mathtt{J}}_{\mathtt{X}}$ from Lemma 4.1.2.

4.3.1 Proof of Theorem 4.3.1 on strong convergence.

The mesh of JAAM is a combination of jump times and adaptive steps. The jump amplitude function γ takes path values at the left-hand limit of each jump point. Therefore, between each jump, the time-steps are adaptive for the approximation of (1.2) when $\gamma = 0$. By the jump-adapted adaptive time-stepping strategy in (3.16), each step is bounded by their distance to the next jump time so that the mesh is guaranteed to have the jump times included.

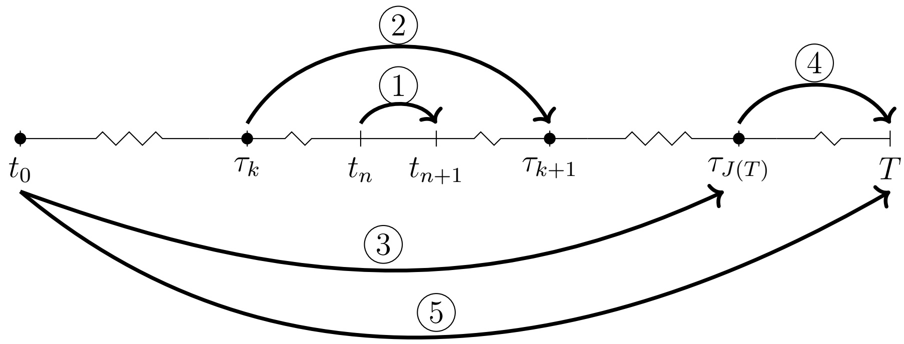

Figure 4.5: Steps of proof for Theorem 4.3.1

The proof is divided into five parts visualised in Figure 4.5: **Step 1**: the one-step error bound combined with backstop method for adaptive steps, that is (1.2) when $\gamma = 0$; **Step 2**: the error bound from k^{th} jump to $(k+1)^{\text{th}}$ jump for $k \in [0, J(t)-1]$, by adding up all the adaptive steps in-between from Step 1, that is from time τ_k to time τ_{k+1}; **Step 3**: the error bound from the beginning to the last jump by adding up all the jumps in-between from Step 2, that is from time $t_0 = 0$ to $\tau_{J(T)}$; **Step 4**:

the error bound from the last jump $\tau_{J(T)}$ to the terminal time T, which consists of all adaptive steps; **Step 5**: the error bound from t_0 to T by adding up Steps 3 and 4.

Proof. ***Step 1 – One-step error bound.*** By Lemma 4.3.2, we have the one-step error bound for approximating (1.2) when $\gamma = 0$. By combining the backstop map φ as in (4.26) and (3.11), for $n \in [N^{(\tau_k)}, N^{(\tau_{k+1})} - 1]$ from Definition 3.2.1 we have the combined one-step error bound as

$$\mathbb{E}\Big[\big\|\widetilde{E}^{\mathtt{J}}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \le \big\|\widetilde{E}^{\mathtt{J}}(t_n)\big\|^2 + \Gamma_1(R)\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\big\|\widetilde{E}^{\mathtt{J}}(r)\big\|^2\Big|\mathcal{F}_{t_n}\Big]dr + \Gamma_2^{\mathtt{J}}(R, X^{\mathtt{J}})h_{\max}^3, \quad (4.29)$$

where $\Gamma_1(R) := C_E(R) + C_{B_1}$ and $\Gamma_2^{\mathtt{J}}(R, X^{\mathtt{J}}) := C_M^{\mathtt{J}}(R, X^{\mathtt{J}}) + C_{B_2}$. By Lemma 4.3.2, we can also define $\Gamma_2^{\mathtt{J}}(R) := \mathbb{E}\big[\Gamma_2^{\mathtt{J}}(R, X^{\mathtt{J}})\big] = C_M^{\mathtt{J}}(R) + C_{B_2}$.

Step 2 – Jump-to-jump error bound. In this step, we calculate the error bound from k^{th} jump to $(k+1)^{\text{th}}$ jump, that is from time τ_k to τ_{k+1}, for $k \in [0, J(t)]$. Firstly, by (4.23) and (4.24) we have the error bound of the last step approaching τ_{k+1} as

$$\widetilde{E}^{\mathtt{J}}(\tau_{k+1}) = \widehat{\mathtt{E}}^{\mathtt{J}}(\tau_{k+1}) + \int_Z \Delta\gamma\Big(z, \widehat{\mathtt{X}}^{\mathtt{J}}(\tau_{k+1}), \widehat{\mathtt{Y}}^{\mathtt{J}}(\tau_{k+1})\Big) J_\nu(dz \times \{\tau_{k+1}\}). \quad (4.30)$$

Since τ_{k+1} is a jump time, by (2.30) we have

$$\int_Z \Delta\gamma\Big(z, \widehat{\mathtt{X}}^{\mathtt{J}}(\tau_{k+1}), \widehat{\mathtt{Y}}^{\mathtt{J}}(\tau_{k+1})\Big) J_\nu(dz \times \{\tau_{k+1}\}) = \Delta\gamma\Big(\zeta_{k+1}, \widehat{\mathtt{X}}^{\mathtt{J}}(\tau_{k+1}), \widehat{\mathtt{Y}}^{\mathtt{J}}(\tau_{k+1})\Big).$$

Taking norm squared on the both sides of (4.30) and by (2.34) we have

$$\Big\|\widetilde{E}^{\mathtt{J}}(\tau_{k+1})\Big\|^2 \le 2\Big\|\widehat{\mathtt{E}}^{\mathtt{J}}(\tau_{k+1})\Big\|^2 + 2\Big\|\Delta\gamma\Big(\zeta_{k+1}, \widehat{\mathtt{X}}^{\mathtt{J}}(\tau_{k+1}), \widehat{\mathtt{Y}}^{\mathtt{J}}(\tau_{k+1})\Big)\Big\|^2. \quad (4.31)$$

Taking expectation on the both sides of (4.31) conditioned on $\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}$, and by (4.2) we have

$$\mathbb{E}\left[\Big\|\widetilde{E}^{\mathtt{J}}(\tau_{k+1})\Big\|^2\Big|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right] \le 2\mathbb{E}\left[\Big\|\widehat{\mathtt{E}}^{\mathtt{J}}(\tau_{k+1})\Big\|^2\Big|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right] + 2\mathbb{E}\left[\Big\|\Delta\gamma\Big(\zeta_{k+1}, \widehat{\mathtt{X}}^{\mathtt{J}}(\tau_{k+1}), \widehat{\mathtt{Y}}^{\mathtt{J}}(\tau_{k+1})\Big)\Big\|^2\Big|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right]. \quad (4.32)$$

For the 2nd term on the RHS, by (4.2) we have

$$\mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right] \leq \mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right] + (1+2L(1))\mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right]. \quad (4.33)$$

By setting $t_n = t_{N^{(\tau_{k+1})}-1}$ and $t_{n+1} = \tau_{k+1}$ in (4.29) we have last step error bound reaching τ_{k+1}, that is $\mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right]$. Substituting it back into (4.33) to replace the 1st term on the RHS, we have

$$\begin{aligned}\mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right] &\leq \left\|\widetilde{E}^{\mathrm{J}}(t_{N^{(\tau_{k+1})}-1})\right\|^2 \\ &+ \Gamma_1(R)\int_{t_{N^{(\tau_{k+1})}-1}}^{\tau_{k+1}} \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(r)\right\|^2\middle|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right]dr \\ &+ \Gamma_2^{\mathrm{J}}(R,X^{\mathrm{J}})h_{\max}^3 + (1+2L(1))\mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right]. \quad (4.34)\end{aligned}$$

Multiplying by the indicator function on the both sides of (4.29) and (4.34), and since $\widehat{\mathrm{E}}^{\mathrm{J}}(t_{n+1}) = \widetilde{E}^{\mathrm{J}}(t_{n+1})$ for $n \in [N^{(\tau_k)}, N^{(\tau_{k+1})}-2]$, we sum up all the steps from τ_k to τ_{k+1} to have

$$\left.\sum_{n=N^{(\tau_k)}}^{N^{(\tau_{k+1})}-1}\left(\mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t_{n+1})\right\|^2\middle|\mathcal{F}_{t_n}\right] - \left\|\widetilde{E}^{\mathrm{J}}(t_n)\right\|^2\right)\mathcal{I}_{N^{(\tau_{k+1})}>n}\right\} =: \mathrm{LHS} \quad (4.35)$$

$$\begin{aligned}&\leq \Gamma_1(R)\sum_{n=N^{(\tau_k)}}^{N^{(\tau_{k+1})}-1}\int_{t_n}^{t_{n+1}} \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(r)\right\|^2\middle|\mathcal{F}_{t_n}\right]dr\,\mathcal{I}_{N^{(\tau_{k+1})}>n} \\ &+ \Gamma_2^{\mathrm{J}}(R,X^{\mathrm{J}})\sum_{n=N^{(\tau_k)}}^{N^{(\tau_{k+1})}-1} h_{\max}^3\mathcal{I}_{N^{(\tau_{k+1})}>n} \\ &+ (1+2L(1))\mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right].\end{aligned}$$

For the RHS, by Definition 3.2.1 we have $n = N^{(r)} - 1$, so the filtration $\mathcal{F}_{t_n}$ could be written as $\mathcal{F}_{t_{N^{(r)}-1}}$. By bounding the number of steps between the k^{th} and $(k+1)^{\text{th}}$ steps by $\pi_{k+1}/h_{\min} + 1$, and all the indicator functions by 1 we have

$$\begin{aligned}\mathrm{LHS} \leq\ &\Gamma_1(R)\int_{\tau_k}^{\tau_{k+1}} \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(r)\right\|^2\middle|\mathcal{F}_{t_{N^{(r)}-1}}\right]dr + \Gamma_2^{\mathrm{J}}(R,X^{\mathrm{J}})\left(\rho\pi_{k+1}+1\right)h_{\max}^2 \\ &+ (1+2L(1))\mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{F}_{t_{N^{(\tau_{k+1})}-1}}\right], \quad (4.36)\end{aligned}$$

which is the error bound between the k^{th} and the $(k+1)^{\text{th}}$ jump, with LHS defined in (4.35).

Step 3 – t_0 ***to last jump error bound.*** Summing up (4.36) over k for all jumps, we have

$$\sum_{k=0}^{J(t)-1} \text{LHS} \leq \Gamma_1(R) \sum_{k=0}^{J(t)-1} \int_{\tau_k}^{\tau_{k+1}} \mathbb{E}\left[\left\| \widetilde{E}^{\text{J}}(r) \right\|^2 \middle| \mathcal{F}_{t_{N(r)-1}} \right] dr + \Gamma_2^{\text{J}}(R, X^{\text{J}}) \sum_{k=0}^{J(t)-1} \left(\rho\pi_{k+1} + 1\right) h_{\max}^2 + (1 + 2L(1)) \sum_{k=0}^{J(t)-1} \mathbb{E}\left[\left\| \widehat{\text{E}}^{\text{J}}(\tau_{k+1}) \right\|^2 \middle| \mathcal{F}_{t_{N(\tau_{k+1})-1}} \right]. \quad (4.37)$$

Notice that we set $\tau_0 = t_0$ so that (4.37) includes the time from t_0 to the first jump τ_1. Taking expectation on the both sides of (4.37), conditioned on $\mathcal{H}_T$ (see Section 4.2) that only contains the information of all jump times and jump sizes known by time T, we have

$$\mathbb{E}\left[\sum_{k=0}^{J(t)-1} \text{LHS} \middle| \mathcal{H}_T \right] \leq \Gamma_1(R) \mathbb{E}\left[\sum_{k=0}^{J(t)-1} \int_{\tau_k}^{\tau_{k+1}} \mathbb{E}\left[\left\| \widetilde{E}^{\text{J}}(r) \right\|^2 \middle| \mathcal{F}_{t_{N(r)-1}} \right] dr \middle| \mathcal{H}_T \right] + \mathbb{E}\left[\Gamma_2^{\text{J}}(R, X^{\text{J}}) \sum_{k=0}^{J(t)-1} \left(\rho\pi_{k+1} + 1\right) \middle| \mathcal{H}_T \right] h_{\max}^2 + (1 + 2L(1)) \mathbb{E}\left[\sum_{k=0}^{J(t)-1} \mathbb{E}\left[\left\| \widehat{\text{E}}^{\text{J}}(\tau_{k+1}) \right\|^2 \middle| \mathcal{F}_{t_{N(\tau_{k+1})-1}} \right] \middle| \mathcal{H}_T \right].$$

Since $J(t)$ is $\mathcal{H}_T$-measurable, we can take out the summation on the RHS out of the conditional expectation. By $\mathcal{H}_T \subseteq \mathcal{F}_{t_{N(r)-1}}$ for $r \in [0, T]$, with tower property and $\tau_{J(t)} = \sum_{i=1}^{J(t)} \pi_i$ we have

$$\mathbb{E}\left[\sum_{k=0}^{J(t)-1} \text{LHS} \middle| \mathcal{H}_T \right] \leq \Gamma_1(R) \int_0^{\tau_{J(t)}} \mathbb{E}\left[\left\| \widetilde{E}^{\text{J}}(r) \right\|^2 \middle| \mathcal{H}_T \right] dr + \Gamma_2^J(R) \left(\rho\tau_{J(t)} + 1\right) h_{\max}^2 + (1 + 2L(1)) \sum_{k=0}^{J(t)-1} \mathbb{E}\left[\left\| \widehat{\text{E}}^{\text{J}}(\tau_{k+1}) \right\|^2 \middle| \mathcal{H}_T \right], \quad (4.38)$$

where $\Gamma_2^J(R)$ is defined in the statement following (4.28).

Step 4 – $\tau_{J(t)}$ ***to*** t ***error bound.*** The period reaching t after the last jump at $\tau_{j(t)}$ consist of errors only for diffusion process, so that $\widehat{\mathtt{E}}^{\mathtt{J}}(t_{n+1}) = \widetilde{E}^{\mathtt{J}}(t_{n+1})$ for $n \in [N^{(\tau_{J(t)})}, t_{N^{(t)}-1}]$. Notice that in the case that the last jump lands at the target time, i.e. $\tau_{J(t)} = t$, the whole Step 4 is not needed. Since $t \in \left[t_{N^{(t)}-1}, t_{N^{(t)}}\right]$, by replacing t_n, t_{n+1} with $t_{N^{(t)}-1}$ and t respectively in (4.29), we have the last step reaching t as

$$\mathbb{E}\left[\left\|\widetilde{E}^{\mathtt{J}}(t)\right\|^2 \middle| \mathcal{F}_{t_{N^{(t)}-1}}\right] \le \left\|\widetilde{E}^{\mathtt{J}}(t_{N^{(t)}-1})\right\|^2 + \Gamma_1(R)\int_{t_{N^{(t)}-1}}^{t} \mathbb{E}\left[\left\|\widetilde{E}^{\mathtt{J}}(r)\right\|^2 \middle| \mathcal{F}_{t_{N^{(t)}-1}}\right] dr + \Gamma_2^{\mathtt{J}}(R, X^{\mathtt{J}})\left|t - t_{N^{(t)}-1}\right|^3. \tag{4.39}$$

Multiplying the indicator function on the both sides of (4.29), summing up to t with the last step (4.39) and taking the expectation conditioned on $\mathcal{H}_T$ we have

$$\left.\begin{aligned} &\mathbb{E}\Bigg[\sum_{n=N^{(\tau_{J(t)})}}^{N^{(t)}-2}\left(\mathbb{E}\left[\left\|\widetilde{E}^{\mathtt{J}}(t_{n+1})\right\|^2 \middle| \mathcal{F}_{t_n}\right] - \left\|\widetilde{E}^{\mathtt{J}}(t_n)\right\|^2\right)\mathbf{1}_{\{N^{(t)}>n+1\}} \\ &\qquad + \mathbb{E}\left[\left\|\widetilde{E}^{\mathtt{J}}(t)\right\|^2 \middle| \mathcal{F}_{t_{N^{(t)}-1}}\right] - \left\|\widetilde{E}^{\mathtt{J}}(t_{N^{(t)}-1})\right\|^2 \Bigg| \mathcal{H}_T\Bigg] \end{aligned}\right\} =: \mathrm{LHS}_{\mathrm{last}} \tag{4.40}$$

$$\begin{aligned} \le\ &\Gamma_1(R)\mathbb{E}\Bigg[\sum_{n=N^{(\tau_{J(t)})}}^{N^{(t)}-2}\int_{t_n}^{t_{n+1}}\mathbb{E}\left[\left\|\widetilde{E}^{\mathtt{J}}(r)\right\|^2 \middle| \mathcal{F}_{t_n}\right]\mathbf{1}_{\{N^{(t)}>n+1\}}dr \\ &\qquad + \int_{t_{N^{(t)}-1}}^{t}\mathbb{E}\left[\left\|\widetilde{E}^{\mathtt{J}}(r)\right\|^2 \middle| \mathcal{F}_{t_{N^{(t)}-1}}\right]dr \Bigg| \mathcal{H}_T\Bigg] \\ &+\mathbb{E}\left[\Gamma_2^{\mathtt{J}}(R, X^{\mathtt{J}})\left(\sum_{n=N^{(\tau_{J(t)})}}^{N^{(t)}-2}(h_{n+1}^{\mathtt{J}})^3\mathbf{1}_{\{N^{(t)}>n+1\}} + \left|t - t_{N^{(t)}-1}\right|^3\right)\middle| \mathcal{H}_T\right] \end{aligned}$$

For the RHS, by changing the notation in the filtration and bounding indicator functions by 1 we have

$$\mathrm{LHS}_{\mathrm{last}} \le \Gamma_1(R)\int_{\tau_{J(t)}}^{t}\mathbb{E}\left[\left\|\widetilde{E}^{\mathtt{J}}(r)\right\|^2 \middle| \mathcal{H}_T\right]dr + \Gamma_2^{J}(R)\big(\rho(t-\tau_{J(t)})+1\big)h_{\max}^2. \tag{4.41}$$

Step 5** – t_0 **to** t **error bound. Adding the error bound after the last jump

(4.41) to the error bound at the last jump (4.38) we have

$$\mathbb{E}\left[\sum_{k=0}^{J(t)-1} \mathrm{LHS}\middle|\mathcal{H}_T\right] + \mathrm{LHS}_{\mathrm{last}} \leq \Gamma_1(R)\int_0^t \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(r)\right\|^2\middle|\mathcal{H}_T\right]dr$$
$$+\Gamma_2^J(R)\,(\rho t+1)\,h_{\max}^2 + (1+2L(1))\sum_{k=0}^{J(t)-1}\mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{H}_T\right], \quad (4.42)$$

where LHS and $\mathrm{LHS}_{\mathrm{last}}$ are defined in (4.35) and (4.40), respectively. Then for the left hand side of (4.42), we first combine the two sums on k and n to one sum on n.

$$\mathbb{E}\left[\sum_{k=0}^{J(t)-1} \mathrm{LHS}\middle|\mathcal{H}_T\right] + \mathrm{LHS}_{\mathrm{last}}$$
$$= \mathbb{E}\Bigg[\sum_{k=0}^{J(t)-1}\Bigg(\sum_{n=N^{(\tau_k)}}^{N^{(\tau_{k+1})}-1}\Big(\mathbb{E}\Big[\big\|\widetilde{E}^{\mathrm{J}}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] - \big\|\widetilde{E}^{\mathrm{J}}(t_n)\big\|^2\Big)\mathcal{I}_{N^{(\tau_{k+1})}>n}$$
$$+\sum_{n=N^{(\tau_{J(t)})}}^{N^{(t)}-2}\Big(\mathbb{E}\Big[\big\|\widetilde{E}^{\mathrm{J}}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] - \big\|\widetilde{E}^{\mathrm{J}}(t_n)\big\|^2\Big)\mathbf{1}_{\{N^{(t)}>n+1\}}$$
$$+\mathbb{E}\Big[\big\|\widetilde{E}^{\mathrm{J}}(t)\big\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big] - \big\|\widetilde{E}^{\mathrm{J}}(t_{N^{(t)}-1})\big\|^2\Bigg|\mathcal{H}_T\Bigg]$$
$$= \mathbb{E}\Bigg[\sum_{n=0}^{N^{(t)}-2}\Big(\mathbb{E}\Big[\big\|\widetilde{E}^{\mathrm{J}}(t_{n+1}))\big\|^2\Big|\mathcal{F}_{t_n}\Big] - \big\|\widetilde{E}^{\mathrm{J}}(t_n)\big\|^2\Big)\mathcal{I}_{N^{(t)}>n+1}$$
$$+\mathbb{E}\Big[\big\|\widetilde{E}^{\mathrm{J}}(t)\big\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big] - \big\|\widetilde{E}^{\mathrm{J}}(t_{N^{(t)}-1})\big\|^2\Bigg|\mathcal{H}_T\Bigg].$$

Since $N_{\max}^{(t)}$ is $\mathcal{H}_T$-measurable, we bound $N^{(t)}$ by $N_{\max}^{(t)}$ and move the sum out of the conditional expectation. Then, by Section 4.2 that $\mathcal{H}_T$ is a sub-σ-algebra of $\mathcal{F}_{t_n}$ for $n \in \left[0, N_{\max}^{(t)}-1\right]$, we apply tower property. Finally, from the telescoping sum with

$\widetilde{E}_0^{\mathrm{J}} = 0$ we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{k=0}^{J(t)-1} \mathrm{LHS}\middle|\mathcal{H}_T\right] + \mathrm{LHS}_{\mathrm{last}} \\
&= \sum_{n=0}^{N_{\max}^{(t)}-2} \mathbb{E}\Bigg[\mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t_{n+1})\right\|^2 \mathcal{I}_{N^{(t)}>n+1}\middle|\mathcal{F}_{t_n}\right] - \left\|\widetilde{E}^{\mathrm{J}}(t_n)\right\|^2 \mathcal{I}_{N^{(t)}>n+1} \\
&\qquad + \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t)\right\|^2\middle|\mathcal{F}_{t_{N^{(t)}-1}}\right] - \left\|\widetilde{E}^{\mathrm{J}}(t_{N^{(t)}-1})\right\|^2 \Bigg|\mathcal{H}_T\Bigg] \\
&= \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t_{N^{(t)}-1})\right\|^2\middle|\mathcal{H}_T\right] - \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t_0)\right\|^2\middle|\mathcal{H}_T\right] \\
&\qquad + \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t)\right\|^2\middle|\mathcal{H}_T\right] - \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t_{N^{(t)}-1})\right\|^2\middle|\mathcal{H}_T\right] \\
&= \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t)\right\|^2\middle|\mathcal{H}_T\right].
\end{aligned} \tag{4.43}$$

where $N_{\max}^{(t)} := \lceil t/h_{\min} + J(t)\rceil$. Substituting (4.43) back to (4.42), we have

$$\begin{aligned}
\mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t)\right\|^2\middle|\mathcal{H}_T\right] &\le \Gamma_1(R)\int_0^t \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(r)\right\|^2\middle|\mathcal{H}_T\right]dr + \Gamma_2^J(R)\,(\rho t+1)\,h_{\max}^2 \\
&\quad +(1+2L(1))\sum_{k=0}^{J(t)-1} \mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{H}_T\right].
\end{aligned} \tag{4.44}$$

By (2.15), we write the last jump term in the integral form and taking the final expectation to have

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=0}^{J(t)-1} \mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(\tau_{k+1})\right\|^2\middle|\mathcal{H}_T\right]\right] &= \mathbb{E}\left[\int_0^t \int_Z \mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(r)\right\|^2\middle|\mathcal{H}_T\right] J_\nu(dz\times dr)\right] \\
&\le \int_0^t \int_Z \mathbb{E}\left[\left\|\widehat{\mathrm{E}}^{\mathrm{J}}(r)\right\|^2\right]\nu(dz)dr \\
&\le \lambda\int_0^t \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(r)\right\|^2\right]dr.
\end{aligned} \tag{4.45}$$

Substituting (4.45) back into (4.44) and take a final expectation on the both sides we have

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(t)\right\|^2\middle|\mathcal{H}_T\right]\right] &\le \Gamma_1(R)\mathbb{E}\left[\int_0^t \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(r)\right\|^2\middle|\mathcal{H}_T\right]dr\right] + \Gamma_2^J(R)\,(\rho t+1)\,h_{\max}^2 \\
&\quad +\lambda(1+2L(1))\int_0^t \mathbb{E}\left[\left\|\widetilde{E}^{\mathrm{J}}(r)\right\|^2\right]dr.
\end{aligned}$$

Simplifying with tower property we have

$$\mathbb{E}\Big[\big\|\widetilde{E}^{\mathrm{J}}(t)\big\|^2\Big] \leq \Big(\Gamma_1(R) + \lambda(1+2L(1))\Big)\int_0^t \mathbb{E}\Big[\big\|\widetilde{E}^{\mathrm{J}}(r)\big\|^2\Big]dr + \Gamma_2^J(R)\,(\rho t+1)\,h_{\max}^2.$$

For all $t \in [0,T]$, by Gronwall's inequality we have

$$\Big(\mathbb{E}\Big[\big\|\widetilde{E}^{\mathrm{J}}(t)\big\|^2\Big]\Big)^{1/2} \leq C^{\mathrm{J}}(R,\rho,t)\,h_{\max}.$$

Taking the maximum over t on the both sides, the proof follows with

$$C^{\mathrm{J}}(R,\rho,T) := \sqrt{\Gamma_2^J(R)\,(\rho T+1)\exp\Big(T\big(\Gamma_1(R)+\lambda(1+2L(1))\big)\Big)}, \qquad (4.46)$$

where $\Gamma_1(R)$ and $\Gamma_2^J(R)$ are defined in (4.29).

□

Chapter 5

Numerical experiments

In this chapter, we illustrate the strong convergence result and computational efficiency of adaptive Milstein and JAAM in sections 5.1 and 5.2, respectively. Both 1D and 2D test models are implemented with one existing telomere SDE model applied with adaptive Misltein method in Section 5.1.2.

Without the explicit forms of the solutions to the test models that we implement in this chapter, a reference solution is taken to play the role of the true solution. We choose one existing fixed-step method that was developed based on Assumptions 3.1.1 and 4.1.1, then approximate the target model with timesteps that are small enough. Wiener increments generated by the reference solution are stored so that the different methods can be compared based on the same set of noises, with $W(t) \approx Z\sqrt{t}$ for $Z \sim \mathcal{N}(0,1)$. The L_2 strong convergence e.g. in (3.88) is the root-mean-square error, we use the sample mean of the Monte Carlo (MC) realisations to approximate the expectation at the terminal time T. Explicitly, for a d-dimensional system with M number of MC realisations and X being the true solution we have

$$\left(\mathbb{E}\left[\left\|X(T)-Y(T)\right\|^2\right]\right)^{1/2} \approx \left(\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{d}\left(Y_i^{\text{ref}}(T)-Y_i(T)\right)^2\right)^{1/2},$$

where $Y_i^{\text{ref}}(T)$ stands for the i^{th} element in the reference solution $Y^{\text{ref}} \in \mathbb{R}^d$ at terminal time T, and $Y \in \mathbb{R}^d$ stands for any numerical method that is tested.

Next, we run the adaptive method through the MC realisations and store all the adaptive steps on each path of each realisation, then take the average which gives

$$h_{\text{mean}} := \frac{1}{M}\sum_{m=1}^{M}\frac{T}{N_m},$$

with N_m denoting the number of adaptive steps taken on the m^{th} sample path to reach T. h_{mean} is used as the fixed stepsize for all other fixed-step methods that we want to compare.

For efficiency comparison, we re-run the adaptive method with other fixed-step methods separately, with independent Wiener increments and without a reference solution. The CPU time consumed for each method on each MC realisation is recorded as the measurement of the computational time they require. Finally, we take the sample mean of all the CPU times to approximate the expected efficiency.

Remark 8 ($h_{n+1} < h_{\min}$)**.** We ensure that we reach the final time by taking $h_N = T - t_{N-1}$ as our final step, and use the backstop method if $h_N < h_{\min}$. For adaptive Milstein method in Chapter 3, the last step is the only occasion of the backstop method taking on a step that might be smaller than $h_{\min}$. However, for JAAM in Chapter 4 this can occur at any step during the process, because according to the jump-adapted adaptive time-stepping strategy in (4.19) the lower bound of each step is the distance to the next jump time rather than the $h_{\min}$ that we choose, which might be smaller than $h_{\min}$.

Remark 9 (**Codes**)**.** All core MATLAB codes for generating the results in this chapter are in: `https://github.com/Gabriel-Lord/Fandi-Sun-Thesis.git`

5.1 Adaptive Milstein

In the numerical experiments below, we set the *adaptive Milstein scheme* (`AMil`) as in (3.10) with (3.16) as the choice of h_{n+1}. *Projected Milstein* (`PMil`) in [2, Eq. (24)] is set to be the backstop method of `AMil` and the reference method of all models. Then we compare the strong convergence looking at the root-mean-square (RMS) error, and efficiency by comparing the CPU time, of `AMil` and `PMil`, *Split-Step Backward Milstein* method (`SSBM`) [2, Eq. (25)], the *new variant of Milstein* (`TMil`) in [34], and the *Tamed Stochastic Runge-Kutta of order* 1.0 (`TSRK1`) method [15, Eq. (3.8) (3.9)].

In more detail, we write out explicitly the schemes that we use in this section. With a middle step in function l, the `PMil` method with parameter $\alpha = 0.25$ has the form of the following

$$l\big(Y_n^{\text{PMil}}\big) := \left(1 \wedge \frac{1}{\Delta t^{\alpha}\big\|Y_n^{\text{PMil}}\big\|}\right) Y_n^{\text{PMil}}, \tag{5.1}$$

$$\begin{aligned} Y_{n+1}^{\text{PMil}}\big(Y_n^{\text{PMil}}\big) := {}& l\big(Y_n^{\text{PMil}}\big) + \Delta t\, f\Big(l\big(Y_n^{\text{PMil}}\big)\Big) + \sum_{i=1}^{m} g_i\Big(l\big(Y_n^{\text{PMil}}\big)\Big)\Delta W_n^i \\ & + \sum_{i,j=1}^{m} \mathbf{D}g_i\Big(l\big(Y_n^{\text{PMil}}\big)\Big) g_j\Big(l\big(Y_n^{\text{PMil}}\big)\Big) \int_{t_n}^{t_{n+1}}\int_{t_n}^{r} dW_j(p)dW_i(r). \end{aligned} \tag{5.2}$$

The proposed method `AMil` at t_{n+1} with `PMil` being the backstop takes in the approximation value at t_n and uses *explicit Milstein* (`EMil`) if $h_{\min} < h_{n+1} \leq h_{\max}$ and `PMil` otherwise. It has the form of

$$Y_{n+1}^{\text{AMil}} := Y_{n+1}^{\text{EMil}}\big(Y_n^{\text{AMil}}\big)\mathbf{1}_{\{h_{\min}<h_{n+1}\leq h_{\max}\}} + Y_{n+1}^{\text{PMil}}\big(Y_n^{\text{AMil}}\big)\mathbf{1}_{\{h_{n+1}\leq h_{\min}\}}, \tag{5.3}$$

where Y_n^{PMil} are defined in (5.1)-(5.2), and Y_n^{EMil} follows directly from (2.21) that

$$\begin{aligned} Y_{n+1}^{\text{EMil}}\big(Y_n^{\text{EMil}}\big) := {}& Y_n^{\text{EMil}} + \Delta t\, f\Big(Y_n^{\text{EMil}}\Big) + \sum_{i=1}^{m} g_i\Big(Y_n^{\text{EMil}}\Big)\Delta W_n^i \\ & + \sum_{i,j=1}^{m} \mathbf{D}g_i\Big(Y_n^{\text{EMil}}\Big) g_j\Big(Y_n^{\text{EMil}}\Big) \int_{t_n}^{t_{n+1}}\int_{t_n}^{r} dW_j(p)dW_i(r). \end{aligned}$$

Similar to `PMil`, `SSBM` also has a middle step but here we do not write it as a function of the previous approximation. It yields as

$$\begin{aligned} \overline{Y_n}^{\text{SSBM}} &:= Y_n^{\text{SSBM}} + \Delta t\ f\Big(Y_n^{\text{SSBM}}\Big), \\ Y_{n+1}^{\text{SSBM}} &:= \overline{Y_n}^{\text{SSBM}} + \sum_{i=1}^{m} g_i\Big(\overline{Y_n}^{\text{SSBM}}\Big)\Delta W_n^i \\ &\quad + \sum_{i,j=1}^{m} \mathbf{D}g_i\Big(\overline{Y_n}^{\text{SSBM}}\Big) g_j\Big(\overline{Y_n}^{\text{SSBM}}\Big) \int_{t_n}^{t_{n+1}}\int_{t_n}^{r} dW_j(p)dW_i(r). \end{aligned}$$

`TMil` with $2\rho\theta = 4$ has the form of

$$\begin{aligned} Y_{n+1}^{\text{TMil}} := {}& Y_n^{\text{TMil}} + \Big(\Delta t\, f\big(Y_n^{\text{TMil}}\big) + \sum_{i=1}^{m} g_i\big(Y_n^{\text{TMil}}\big)\Delta W_n^i \\ & + \sum_{i,j=1}^{m} \mathbf{D}g_i\big(Y_n^{\text{TMil}}\big) g_j\big(Y_n^{\text{TMil}}\big) \int_{t_n}^{t_{n+1}}\int_{t_n}^{r} dW_j(p)dW_i(r)\Big) \Big/ \Big(1 + \Delta t\,\big\|Y_n^{\text{TMil}}\big\|^{2\rho\theta}\Big). \end{aligned}$$

Lastly, `TSRK1` yields as

$$Y_{n+1}^{\mathrm{TSRK1}} := Y_n^{\mathrm{TSRK1}} + \frac{f\big(Y_n^{\mathrm{TSRK1}}\big)\Delta t}{1+\Delta t\big\|f\big(Y_n^{\mathrm{TSRK1}}\big)\big\|} + \sum_{i=1}^{m}\frac{g_i\big(Y_n^{\mathrm{TSRK1}}\big)\Delta W_n^i}{1+\Delta t\big\|g_i\big(Y_n^{\mathrm{TSRK1}}\big)\big\|}$$
$$-\frac{1}{2}\sum_{i=1}^{m}\frac{g_i\big(H_{n+1}^{(2,i)}\big)\sqrt{\Delta t}}{1+\Delta t\big\|g_i\big(H_{n+1}^{(2,i)}\big)\big\|} + \frac{1}{2}\sum_{i=1}^{m}\frac{g_i\big(H_{n+1}^{(3,i)}\big)\sqrt{\Delta t}}{1+\Delta t\big\|g_i\big(H_{n+1}^{(3,i)}\big)\big\|},$$

where

$$H_{n+1}^{(2,i)} := Y_n^{\mathrm{TSRK1}} - \sum_{j=1}^{m} g_j\big(Y_n^{\mathrm{TSRK1}}\big)\Delta W_n^i \Delta W_n^j/(2\sqrt{\Delta t}) + g_i\big(Y_n^{\mathrm{TSRK1}}\big)\sqrt{\Delta t}/2;$$
$$H_{n+1}^{(3,i)} := Y_n^{\mathrm{TSRK1}} + \sum_{j=1}^{m} g_j\big(Y_n^{\mathrm{TSRK1}}\big)\Delta W_n^i \Delta W_n^j/(2\sqrt{\Delta t}) - g_i\big(Y_n^{\mathrm{TSRK1}}\big)\sqrt{\Delta t}/2.$$

5.1.1 One-dimensional test systems

In order to demonstrate strong convergence of order one for a scalar test equation with non-globally Lipschitz drift, consider

$$dX(t) = (X(t) - 3X(t)^3)dt + G(X(t))dW(t), \quad t \in [0,1]. \tag{5.4}$$

For illustrating both the multiplicative and additive noise cases, we estimate the RMS error by a Monte Carlo method using $M = 1000$ trajectories for $h_{\max} = [2^{-14}, 2^{-12}, 2^{-10}, 2^{-8}, 2^{-6}]$, $\rho = 2^2$, $\kappa = 1$, and use as a reference solution `PMil` over a mesh with uniform stepsizes $h_{\mathrm{ref}} = 2^{-18}$.

For model with additive noise we set $G(x) = \sigma$ in (5.4), and for model with multiplicative noise we set $G(x) = \sigma(1 - x^2)$ with $\sigma = 0.2$ and $X(0) = 11$ in both cases. Strong convergence of order one is displayed by all methods in Figure 5.1 part (a) and (c) for the additive and multiplicative cases respectively, with the efficiency displayed in parts (b) and (d).

Finally, consider Theorem 3.3.2. We illustrate that the probability of our time-stepping strategy selecting $h_{\min}$, and therefore triggering an application of the back-stop method, can be made arbitrarily small at every step by an appropriate choice of ρ (with fixed $\kappa = 1$). Consider (5.4) again with $G(x) = \sigma(1 - x^2)$, this time with $X(0) = 100$, $\kappa, T = 1$, $h_{\max} = 2^{-20}$ and $\rho = [2, 4, \ldots, 16]$. In Figure 5.2 (e), we plot two paths of h when $\rho = 2, 6$. Observe that when $\rho = 2$ the backstop is triggered only for the first 10^5 steps approximately, whereas once ρ is increase to 6

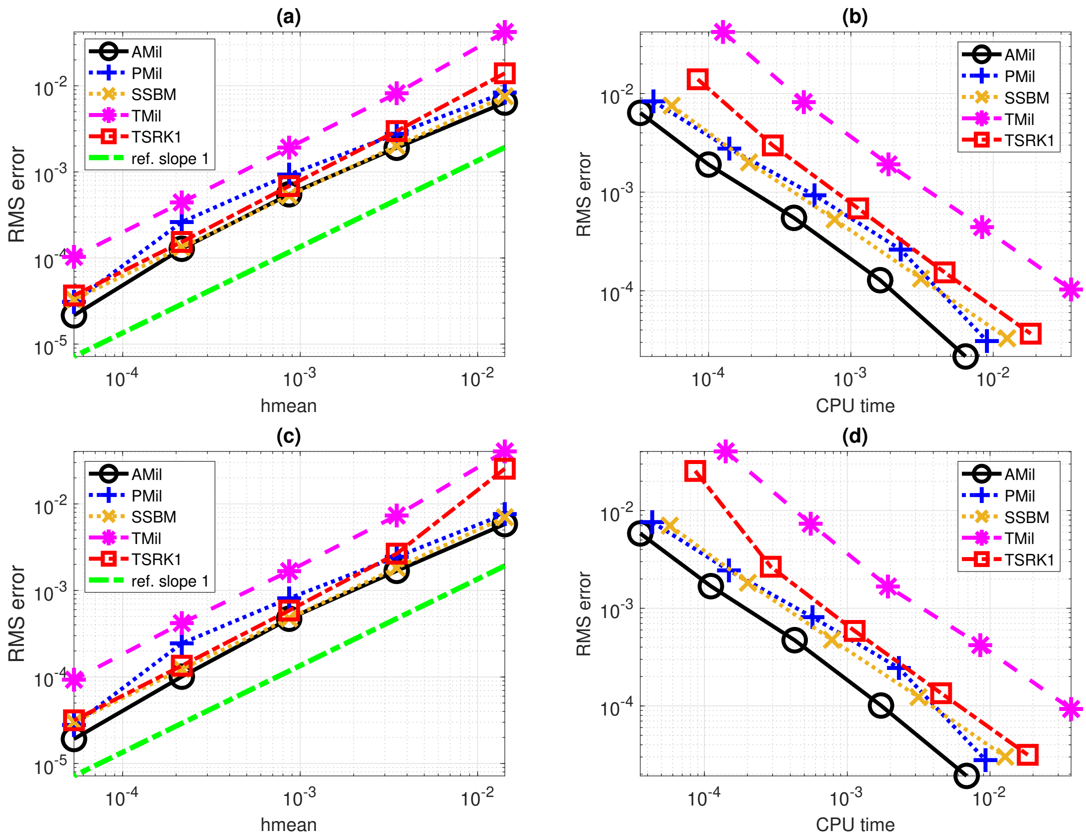

Figure 5.1: Strong convergence and efficiency of adaptive Milstein for approximating (5.4) with (a) and (b) for additive noise; (c) and (d) for multiplicative noise.

this is reduced to the first 2×10^4 steps approximately. Estimated probabilities of using $h_{\min}$ are plotted on a log-log scale as a function of ρ in Figure 5.2 (f) (with $M = 100$ realizations). The estimated probability of using $h_{\min}$ declines to zero as ρ increases. We observe a rate close to -1, matching that in (3.20) with $\kappa = 1$.

5.1.2 One-dimensional model of telomere shortening

In molecular biology, the telomere is a short region of highly repeated nucleotide sequence that caps the ends of eukaryotic chromosomes. They protect chromosomes from losing core DNA fragments due to the end-replication problem [3]. When the division number of a normal human cell reaches the Hayflick limit, or when the telomere is shortened to a certain critical level, the cell ceases to divide and enters the senescence phase [22]. In addition to the end-replication problem, factors such as oxidative stress also contribute to telomere loss [56] and significantly to cell senescence [1]. However, the enzyme telomerase, which appears in germ cells, some stem cells, and most cancer cells, extends telomere by transcribing it reversely [3].

One stochastic model was built in [18] for telomere length dynamics, where the

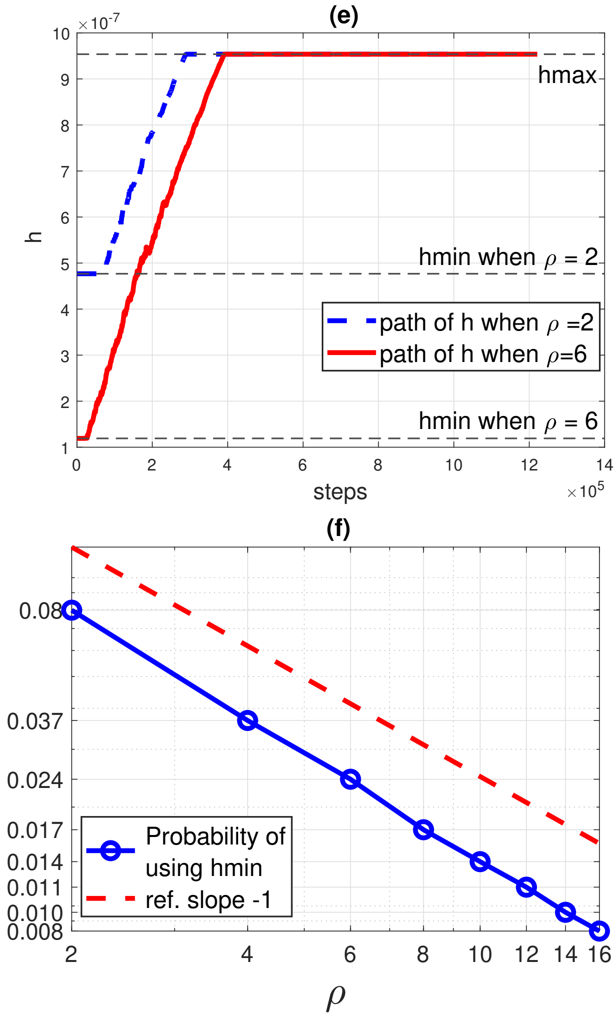

Figure 5.2: Two paths of timestep h for $\rho = 2, 6$ in (e) and the estimated probability of using $h_{\min}$ for the multiplicative noise model with $M = 100$ realizations in (f).

loss from replication was assumed to be a constant decrease with rate c and the impact of oxidative stress assumed to be random breaks that uniformly strike at the telomere. The activity of telomerase was ignored. With $L(t)$ denoting for the length of telomere at time t, the number of the breaks striking within a given time interval was modelled to follow a Poisson distribution with parameter $aL(t)$. Therefore, the following one-dimensional SDE model was given in [18, Eq. (A6)] for modelling the shortening over time of telomere length L in DNA replication

$$dL(t) = -\big(c + aL(t)^2\big)dt + \sqrt{\frac{1}{3}aL(t)^3}dW(t). \tag{5.5}$$

The parameter c determines the underlying decay rate of the length and a controls the intensity at which random breaks occur in the telomere; we take $(a, c) = (0.41 \times 10^{-6}, 7.5)$ as in [18]. In this example we fix $\rho = 4$, instead adjusting the parameter

κ in (3.16) to control use of the backstop method. Individual paths are shown in Figure 5.3 where we take $h_{\max} = 2^{-18}$, and $h = 2^{-20}$ for the fixed-step methods.

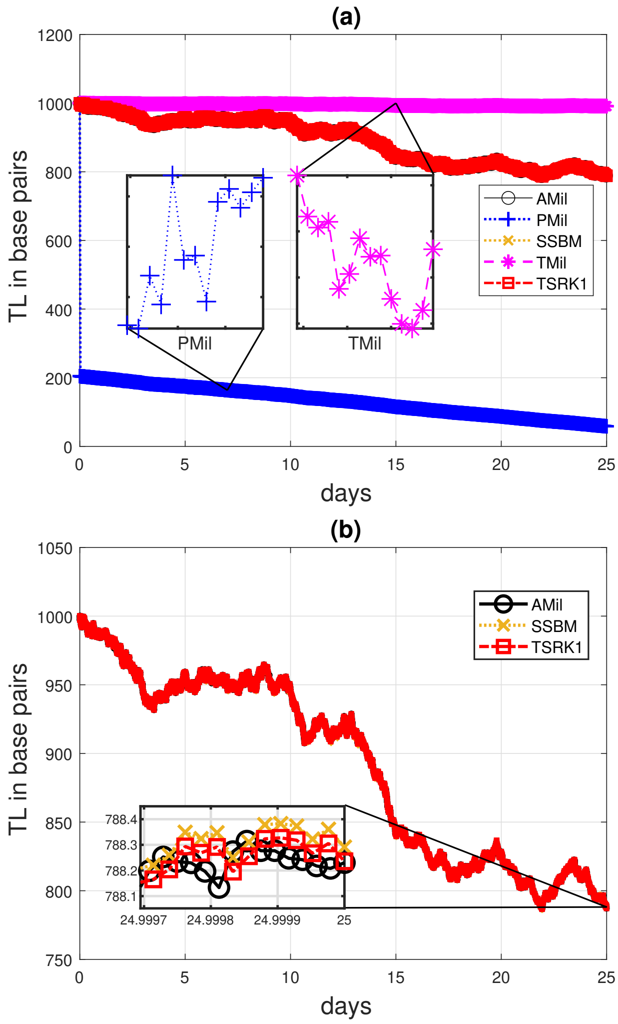

Figure 5.3: Single paths of the Telomere length SDE (5.5) solved over 25 days. (b) shows a detailed plot from (a).

We set $L(0) = 1000$, noting from [18] that initial values could be as high as (say) $L(0) = 6000$ and remain physically realistic. The end of the interval of valid simulation is determined by the first time at which trajectories reach zero, and is therefore random. However this is not observed to occur in the timescale (25 days) we consider here.

By design `PMil` projects the data onto a ball of radius determined in part by the growth of the drift term. We see in Figure 5.3 (a) that (`PMil`) immediately is reduced to approximately 200.

Contrarily, the design of `TMil` scales both drift and diffusion terms by $1/(1 + h|L|^2)$ for this model. When $h|L|^2$ is large this scaling can damp out changes from step to step, and in Figure 5.3 (a) we see that `TMil` shows as (spuriously) almost constant. The paths of the other methods, `AMil`, `SSBM` and `TSRK1` are close together as shown in Figure 5.3 (a) and in high detail in (b).

Notice that we used $\kappa = 8$ in (3.16) for `AMil` method to reduce the chance of requiring the backstop method `PMil` while keeping $\rho = 4$. We avoid setting $\kappa = 1$ in this case because $L(0) = 1000$ and so the adaptive step h_{n+1} would too frequently require the backstop method.

5.1.3 Two-dimensional test systems

We now consider three ($i = 1, 2, 3$) different SDEs:

$$dX(t) = F(X(t))dt + G_i(X(t))dW(t), \quad t \in [0, 1], \quad X(0) = [7, 9]^T, \tag{5.6}$$

with $W(t) = [W_1(t), W_2(t)]^T$, where W_1 and W_2 are independent scalar Wiener processes, $X(t) = [X_1(t), X_2(t)]^T$, $F(x) = [x_2 - 3x_1^3, x_1 - 3x_2^3]^T$, and

$$G_1(x) = \sigma \begin{pmatrix} x_1^2 & 0 \\ 0 & x_2^2 \end{pmatrix}, \; G_2(x) = \sigma \begin{pmatrix} x_2^2 & x_2^2 \\ x_1^2 & x_1^2 \end{pmatrix}, \; G_3(x) = \sigma \begin{pmatrix} 1.5x_1^2 & x_2 \\ x_2^2 & 1.5x_1 \end{pmatrix}. \tag{5.7}$$

G_1 is an example of diagonal noise, G_2 commutative noise, and G_3 non-commutative noise.

For G_1 and G_2 we use $h_{\max} = [2^{-14}, 2^{-12}, 2^{-10}, 2^{-8}, 2^{-6}]$, $h_{\text{ref}} = 2^{-18}$, $\rho = 4$ and $\kappa = 1$. In Figure 5.4 (a) and (c), we see order one strong convergence for all methods. Parts (b) and (d) show the efficiency of the adaptive method.

For the non-commutative noise case take $h_{\max} = [2^{-8}, 2^{-7}, 2^{-6}, 2^{-5}, 2^{-4}]$, $h_{\text{ref}} = 2^{-11}$, $\rho = 2^2$ and $X(0) = [3, 4]^T$. To simulate the Lévy areas we follow the method in [23, Sec. 4.3], which is based on the Euler approximation of a system of SDEs. Again, we observe order one convergence for all methods in Figure 5.4 (e) and that `AMil` is the most efficient in (f). Note that as `TSRK1` is only supported theoretically for commutative noise we do not consider it here.

5.2 Jump-adapted adaptive Milstein (JAAM)

In this section, we demonstrate JAAM by approximating (1.2) in both 1D and 2D with $f(x) = x - 3x^3$ and $\gamma(z, x) = z\,x$. Based on Definition 4.2.1 and (5.3), JAAM

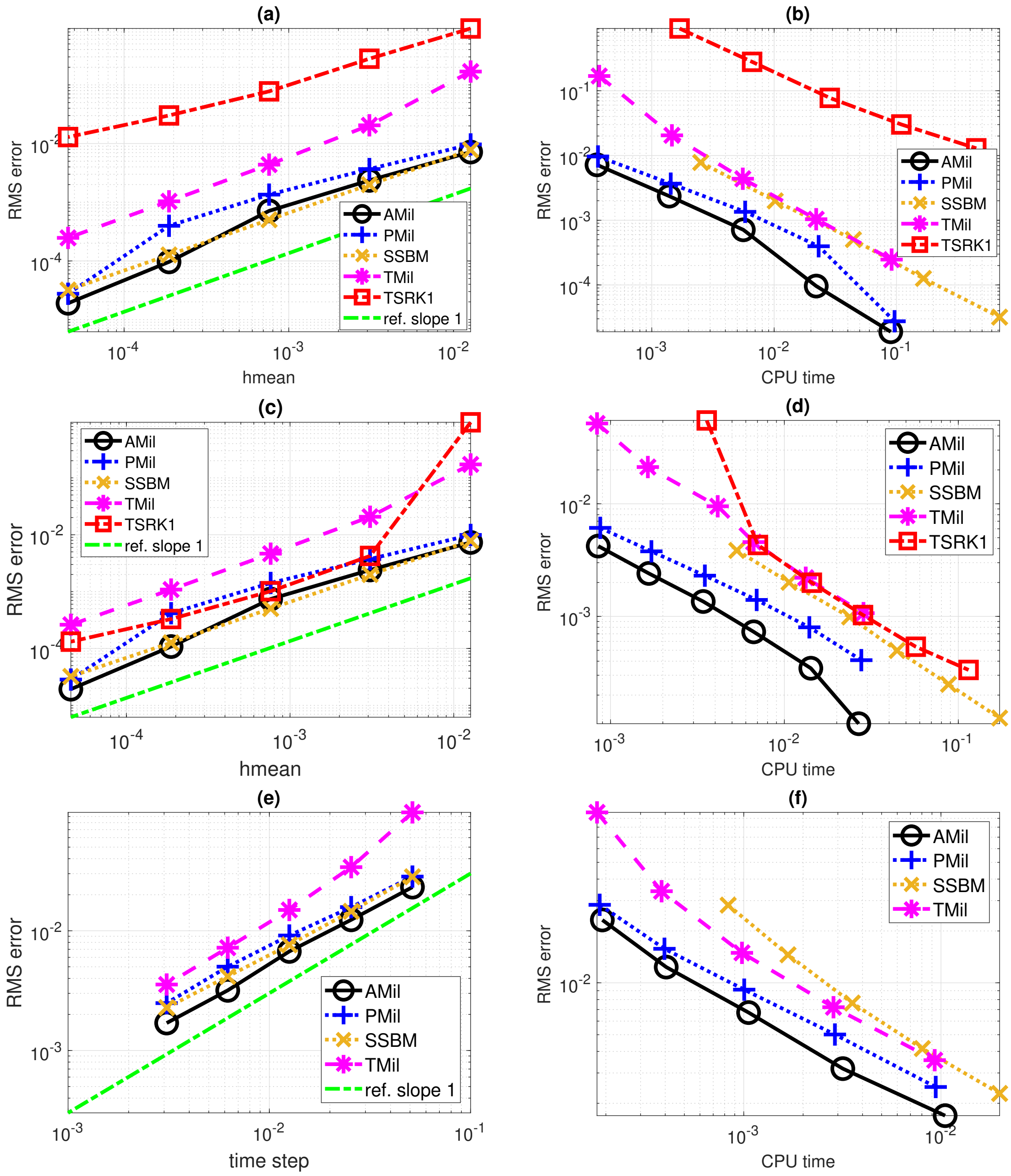

Figure 5.4: Adaptive Milstein method for approximating the two-dimensional system (5.6). (a) and (b) show the strong convergence and efficiency for diagonal noise, (c) and (d) with commutative and (e) and (f) for non-commutative noise. We choose $a = 3$, $\sigma = 0.2$ and $b = 1.5$.

with `PMil` being the backstop method has the form of

$$\begin{aligned}\widehat{\mathrm{Y}}_{n+1}^{\mathrm{AMil}} &= Y_{n+1}^{\mathrm{EMil}}\big(Y_n^{\mathrm{AMil}}\big)\mathbf{1}_{\{h_{\min}<h_{n+1}\le h_{\max}\}} + Y_{n+1}^{\mathrm{PMil}}\big(Y_n^{\mathrm{AMil}}\big)\mathbf{1}_{\{h_{n+1}\le h_{\min}\}},\\ Y_{n+1}^{\mathrm{AMil}} &= \widehat{\mathrm{Y}}_{n+1}^{\mathrm{AMil}} + \widehat{\mathrm{Y}}_{n+1}^{\mathrm{AMil}}\zeta_{t_{n+1}}\mathbf{1}_{\{t_{n+1}=\tau_{J(t_{n+1})}\}},\end{aligned}$$

where h_{n+1} is chosen by the strategy (4.19). The indicator function and the jump size ζ are both 0 when t_{n+1} is not a jump time.

5.2.1 One-dimensional test systems

We have the diffusion coefficients g the same as the ones in Section 5.1.1 for both additive and multiplicative noises. For the reference line and the comparison method, we use `PMil` with jumps on a deterministic mesh that is a superposition of the equidistant steps and jump times. In case of an approximation step landing between two reference steps, we use Brownian Bridge to split the Wiener increment. Together with `SSBM` `TMil` and `TSRK1`, all four comparison methods are updated to jumps based on their structure in Section 5.1. One example is jump-adapted projected Milstein (`JA-PMil`) with the form

$$\begin{aligned}l\big(Y_n^{\text{JA-PMil}}\big) &= \left(1\wedge\frac{1}{\Delta t^{\alpha}\big\|Y_n^{\text{JA-PMil}}\big\|}\right)Y_n^{\text{JA-PMil}},\\ \widehat{\mathrm{Y}}_{n+1}^{\text{JA-PMil}}\big(Y_n^{\text{JA-PMil}}\big) &= l\big(Y_n^{\text{JA-PMil}}\big) + \Delta t\, f\Big(l\big(Y_n^{\text{JA-PMil}}\big)\Big)\\ &\quad + \sum_{i=1}^{m} g_i\Big(l\big(Y_n^{\text{JA-PMil}}\big)\Big)\Delta W_n^i\\ &\quad + \sum_{i,j=1}^{m}\mathbf{D}g_i\Big(l\big(Y_n^{\text{JA-PMil}}\big)\Big)g_j\Big(l\big(Y_n^{\text{JA-PMil}}\big)\Big)\int_{t_n}^{t_{n+1}}\int_{t_n}^{r} dW_j(p)dW_i(r),\\ Y_{n+1}^{\text{JA-PMil}}\big(Y_n^{\text{JA-PMil}}\big) &= \widehat{\mathrm{Y}}_{n+1}^{\text{JA-PMil}}\big(Y_n^{\text{JA-PMil}}\big)\\ &\quad + \widehat{\mathrm{Y}}_{n+1}^{\text{JA-PMil}}\big(Y_n^{\text{JA-PMil}}\big)\zeta_{J(t_{n+1})}\mathbf{1}_{\{t_{n+1}=\tau_{J(t_{n+1})}\}}.\end{aligned}$$

For both additive (Figure 5.5 (a)-(b)) and multiplicative (Figure 5.5 (c)-(d)) noises, we have initial value 5, terminal time $T=1$, reference stepsize 2^{-18}, list of $h_{\max} = [2^{-14}, 2^{-13}, 2^{-12}, 2^{-11}, 2^{-10}]$, $\rho = 2^2$, jump intensity $\lambda = 2$, jump sizes follow the standard normal distribution $\mathcal{N}(0,1)$, 1000 Monte Carlo simulation. We can see in Figure 5.5 that in both cases, JAAM shows advantages in error and CPU time.

Further, we demonstrate the performance of JAAM when the jump intensity

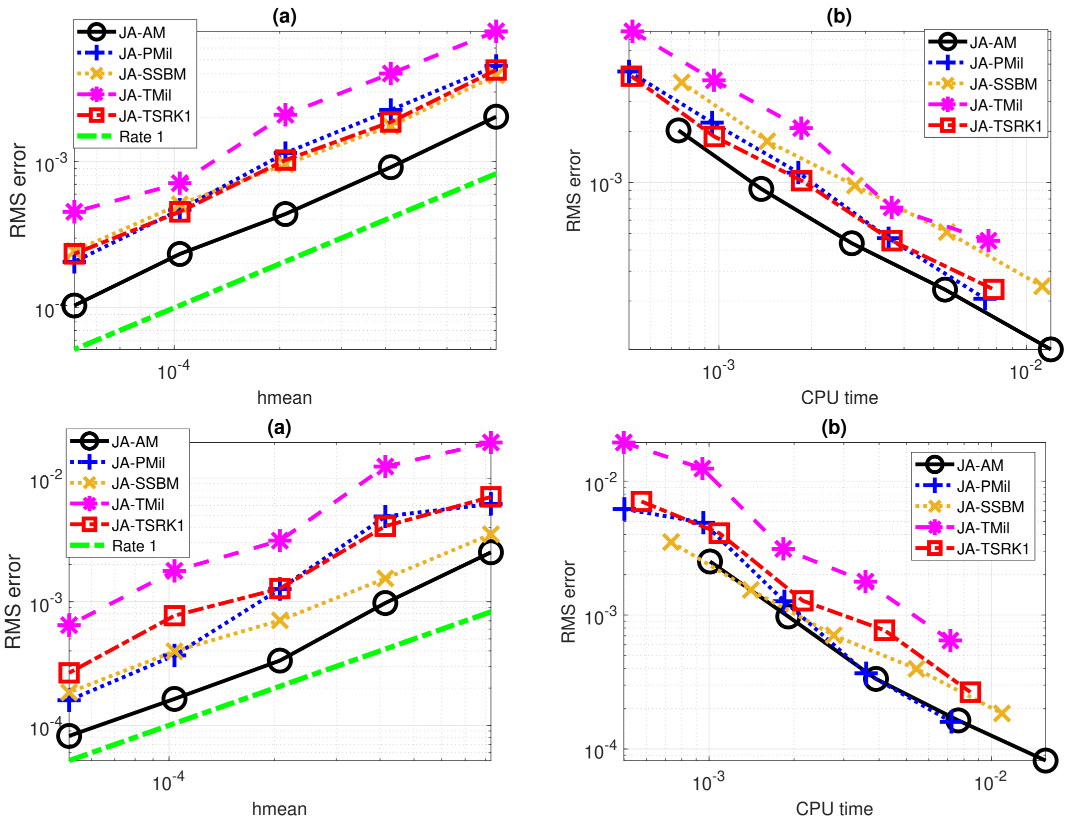

Figure 5.5: Strong convergence and efficiency of JAAM for approximating the one-dimensional system with (a) and (b) for additive noise; (c) and (d) for multiplicative noise.

increases in Figure 5.6. With the same settings for 1D multiplicative model, we see that JAAM outperforms when the jump intensity increases from 4 to 20.

5.2.2 Two-dimensional test systems

For 2D models, with the same structure of drift and jump coefficients as in (5.7), we have the same diagonal noise ($G_1(x)$), commutative noise ($G_2(x)$) and non-commutative noise, $G_3(x)$. The jump coefficient is $\gamma = z\,x$ with $z \sim \mathcal{N}(0,1)$.

For diagonal (Figure 5.7 (a)-(b)) and commutative (Figure 5.5 (c)-(d)) noises, we have initial value $[5,7]^T$, terminal time $T = 1$, reference stepsize 2^{-18}, list of $h_{\max} = [2^{-9}, 2^{-8}, 2^{-7}, 2^{-6}, 2^{-5}]$, $\rho = 2^2$, jump intensity $\lambda = 4$, jump sizes follow $\mathcal{N}(0,1)$, 1000 Monte Carlo simulation. For non-commutative noise (Figure 5.7 (e)-(f)), we have initial value $[15,17]^T$, reference stepsize 2^{-9}, list of $h_{\max} = [2^{-7}, 2^{-6.5}, 2^{-6}, 2^{-5.5}, 2^{-5}]$, 100 Monte Carlo simulation, and with rest settings the same. We can see in Figure 5.7 that in all 3 cases, `JAAM` shows advantages in error and CPU time.

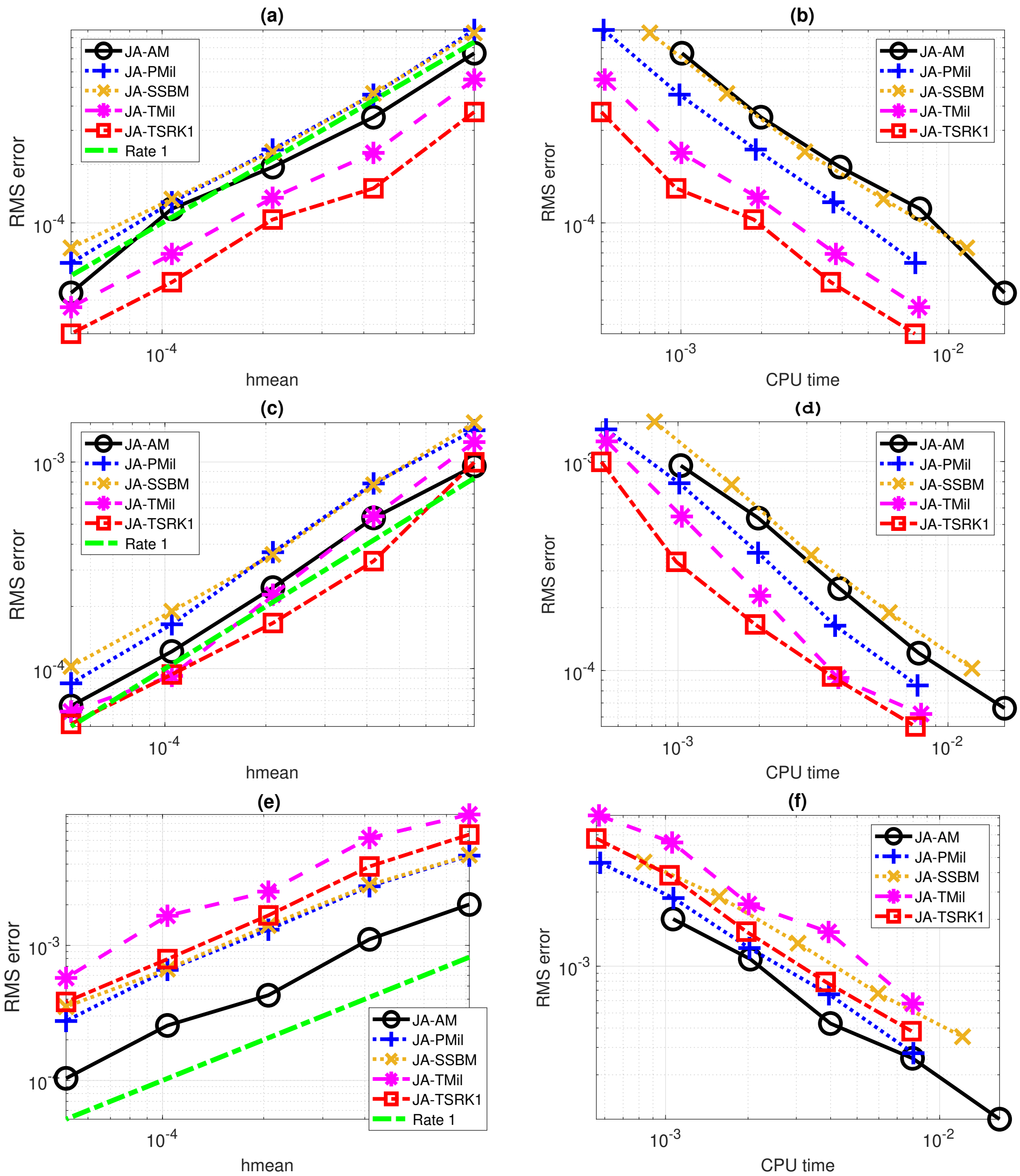

Figure 5.6: Strong convergence and efficiency of JAAM for approximating the one-dimensional system with multiplicative noise, when jump intensity increases. (a) and (b) with jump intensity 4; (a) and (b) with jump intensity 10; (a) and (b) with jump intensity 20.

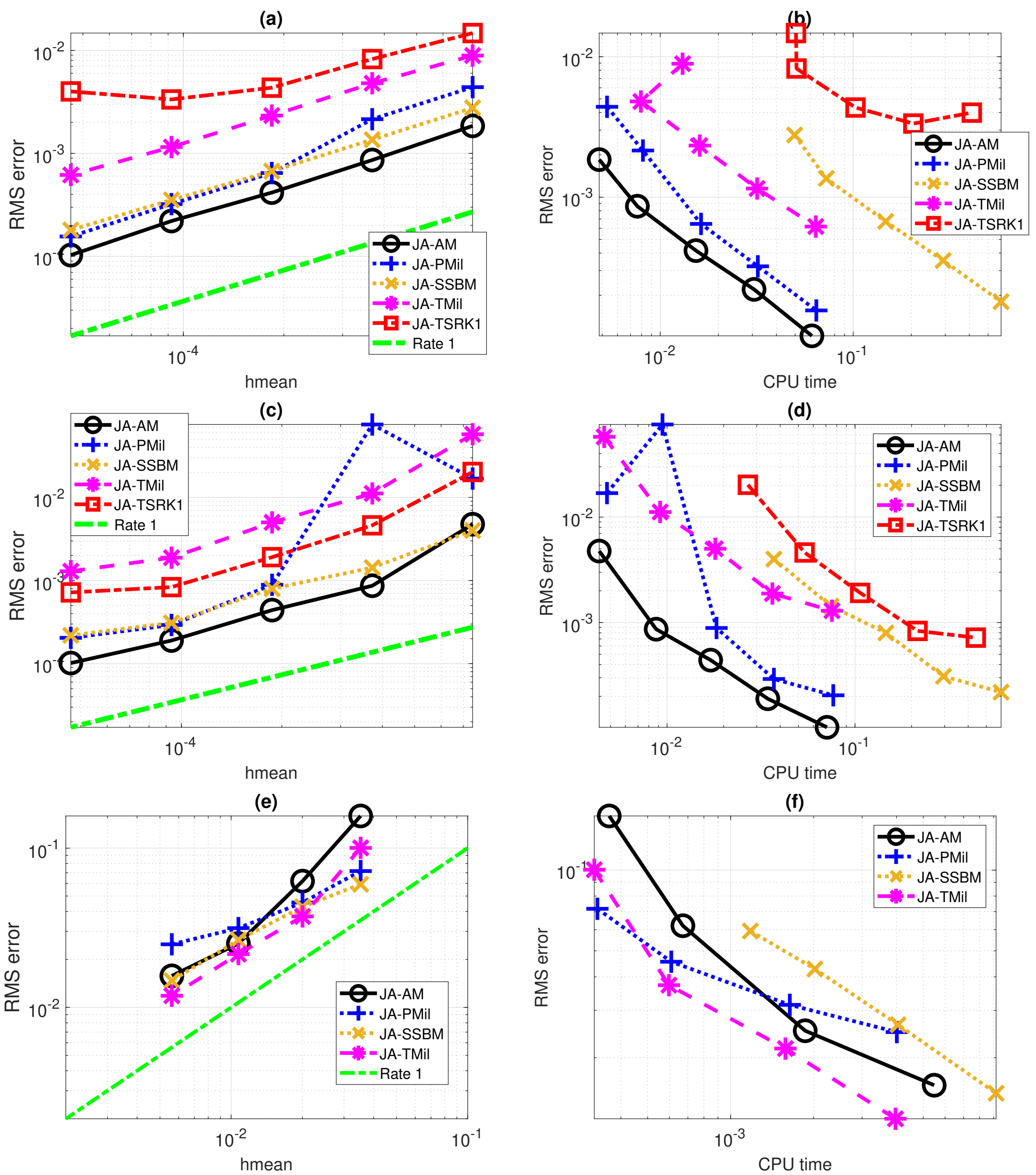

Figure 5.7: Strong convergence and efficiency of JAAM for approximating the two-dimensional system with (a) and (b) for diagonal noise; (c) and (d) for commutative noise; (e) and (f) for non-commutative noise.

Chapter 6

Summary

We proposed the adaptive time-stepping strategies of the explicit Milstein method in Chapter 3 for approximating the SDEs with non-globally Lipschitz coefficients, see Assumptions 3.1.1 and 3.1.2. With a suitable backstop method it reaches strong L_2 convergences of order one with proof in Section 3.3.3. The probability of using a backstop method can be made arbitrarily small with proof in Section 3.3.4. We then extended the time-stepping strategies to be jump-adapted for approximating SDEs driven by Poisson random measure in Chapter 4 with globally Lipschitz jump coefficient in Assumption 4.1.1. We showed that the jump-adapted adaptive Milstein method achieves strong L_2 convergences of order one in Section 4.3.1. Finally, we compared both variants of strategies with fixed-step methods in computer and showed their convergence and efficiency in Chapter 5.

For future work, the removal of the use of the backstop method in both the adaptive Milstein method and the JAAM could be one direction. Then, based on [5], the jump-adapted adaptive time-stepping introduced for the Milstein method in Chapter 4 could be extended to numerical schemes with convergence order $\alpha \in \{0.5, 1, 1.5, 2, \dots\}$. Finally, the Poisson random measure in (1.2) could be extended to be the Lévy measure, that is to include the infinite activity of jumps.

Bibliography

[1] K. B. Beckman and B. N. Ames. The free radical theory of aging matures. *Physiological reviews*, 1998.

[2] W.-J. Beyn, E. Isaak, and R. Kruse. Stochastic C-stability and B-consistency of explicit and implicit Milstein-type schemes. *Journal of Scientific Computing*, 70(3):1042–1077, 2017.

[3] E. H. Blackburn. Structure and function of telomeres. *Nature*, 350(6319):569–573, 1991.

[4] S. Blythe, X. Mao, and X. Liao. Stability of stochastic delay neural networks. *Journal of the Franklin Institute*, 338(4):481–495, 2001.

[5] N. Bruti-Liberati and E. Platen. Strong approximations of stochastic differential equations with jumps. *Journal of Computational and Applied Mathematics*, 205(2):982–1001, 2007.

[6] P. M. Burrage, R. Herdiana, and K. Burrage. Adaptive stepsize based on control theory for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 171(1-2):317–336, 2004.

[7] S. Campbell and G. Lord. Adaptive time-stepping for stochastic partial differential equations with non-Lipschitz drift. *arXiv preprint arXiv:1812.09036*, 2018.

[8] K. Dareiotis, C. Kumar, and S. Sabanis. On tamed Euler approximations of SDEs driven by Lévy noise with applications to delay equations. *SIAM Journal on Numerical Analysis*, 54(3):1840–1872, 2016.

[9] S. Deng, W. Fei, W. Liu, and X. Mao. The truncated EM method for stochastic differential equations with Poisson jumps. *Journal of Computational and Applied Mathematics*, 355:232–257, 2019.

[10] S. Dineen. *Probability Theory in Finance: a mathematical guide to the Black-Scholes Formula*. Graduate studies in mathematics; v.70. American Mathematical Society, Universities Press, 2011.

[11] W. Fang and M. B. Giles. Adaptive Euler–Maruyama method for SDEs with non-globally Lipschitz drift. In *International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 217–234. Springer, 2016.

[12] W. Fang and M. B. Giles. Adaptive Euler–Maruyama method for SDEs with non-globally Lipschitz drift. *The Annals of Applied Probability*, 30(2):526–560, 2020.

[13] A. Friedman. *Stochastic differential equations and applications, Volume 1*. Academic Press, New York, 1975.

[14] J. Gaines and T. Lyons. Variable step size control in the numerical solution of stochastic differential equations. *SIAM J. on Applied Math.*, 57(5):1455–1484, 1997.

[15] S. Gan, Y. He, and X. Wang. Tamed Runge-Kutta methods for SDEs with super-linearly growing drift and diffusion coefficients. *Applied Numerical Mathematics*, 152:379–402, 2020.

[16] H. Geman and M. Yor. Stochastic time changes in catastrophe option pricing. *Insurance: Mathematics and economics*, 21(3):185–193, 1997.

[17] E. Ghysels, A. C. Harvey, and E. Renault. 5 stochastic volatility. *Handbook of statistics*, 14:119–191, 1996.

[18] J. Grasman, H. Salomons, and S. Verhulst. Stochastic modeling of length-dependent telomere shortening in Corvus monedula. *Journal of Theoretical Biology*, 282(1):1–6, 2011.

[19] Q. Guo, W. Liu, X. Mao, and R. Yue. The truncated Milstein method for stochastic differential equations with commutative noise. *Journal of Computational and Applied Mathematics*, 338:298 – 310, 2018.

[20] I. Gyöngy and N. V. Krylov. On stochastic equations with respect to semimartingales I. *Stochastics: An International Journal of Probability and Stochastic Processes*, 4(1):1–21, 1980.

[21] R. Hasminskii. *Stochastic Stability of Differential Equations*. Sijthoff & Noordhoff, 1980.

[22] L. Hayflick. The limited in vitro lifetime of human diploid cell strains. *Experimental cell research*, 37(3):614–636, 1965.

[23] D. J. Higham and P. E. Kloeden. Maple and MATLAB for stochastic differential equations in finance. In *Programming Languages and Systems in Computational Economics and Finance*, pages 233–269. Springer, 2002.

[24] D. J. Higham and P. E. Kloeden. Convergence and stability of implicit methods for jump-diffusion systems. *Int. J. Numer. Anal. Model*, 3(2):125–140, 2006.

[25] D. J. Higham, X. Mao, and L. Szpruch. Convergence, non-negativity and stability of a new Milstein scheme with applications to finance. *Discrete and Continuous Dynamical Systems B*, pages 2083–2100, 2013.

[26] N. Hofmann, T. Müller-Gronbach, and K. Ritter. The optimal discretization of stochastic differential equations. *Journal of Complexity*, 17(1):117 – 153, 2001.

[27] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society of*

London A: Mathematical, Physical and Engineering Sciences, 467(2130):1563–1576, 2011.

[28] S. Ilie, K. R. Jackson, and W. H. Enright. Adaptive time-stepping for the strong numerical solution of stochastic differential equations. *Numer. Algorithms*, 68(4):791–812, 2015.

[29] J. Jang. Jump diffusion processes and their applications in insurance and finance. *Insurance: Mathematics and Economics*, 41(1):62–70, 2007.

[30] C. Kelly and G. J. Lord. Adaptive time-stepping strategies for nonlinear stochastic systems. *IMA Journal of Numerical Analysis*, 38(3):1523–1549, 2018.

[31] C. Kelly and G. J. Lord. Adaptive Euler methods for stochastic systems with non-globally Lipschitz coefficients. *Numerical Algorithms*, pages 1–27, 2021.

[32] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, Berlin, Heidelberg, 1 edition, 1992.

[33] C. Kumar and S. Sabanis. On tamed Milstein schemes of SDEs driven by Lévy noise. *Discrete and Continuous Dynamical Systems - B*, 22(2):421–463, 2017.

[34] C. Kumar and S. Sabanis. On Milstein approximations with varying coefficients: the case of super-linear diffusion coefficients. *BIT Numerical Mathematics*, 59(4):929–968, 2019.

[35] H. Lamba, J. C. Mattingly, and A. M. Stuart. An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *IMA J. Numer. Anal.*, 27(3):479–506, 2007.

[36] P. Lévy. Wiener's random function, and other Laplacian random functions. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. The Regents of the University of California, 1951.

[37] M. Li, C. Huang, and Z. Chen. Compensated projected Euler-Maruyama method for stochastic differential equations with superlinear jumps. *Applied Mathematics and Computation*, 393:125760, 2021.

[38] G. J. Lord, C. E. Powell, and T. Shardlow. *An introduction to computational stochastic PDEs*, volume 50. Cambridge University Press, 2014.

[39] S. Malham and A. Wiese. Efficient almost-exact Lévy area sampling. *Statistics & Probability Letters*, 88:50–55, 2014.

[40] X. Mao. *Stochastic differential equations and applications*. Woodhead Publishing, Cambridge, 2 edition, 2007.

[41] X. Mao. The truncated Euler–Maruyama method for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 290:370–384, 2015.

[42] X. Mao and C. Yuan. *Stochastic differential equations with Markovian switching*. Imperial college press, 2006.

[43] S. Mauthner. Step size control in the numerical solution of stochastic differential equations. *J. of Comp. and Applied Math.*, 100(1):93–109, 1998.

[44] A. Papoulis and S. U. Pillai. *Probability, random variables, and stochastic processes*. McGraw-Hill, New York, 4 edition, 2002.

[45] E. Platen and N. Bruti-Liberati. *Numerical solution of stochastic differential equations with jumps in finance*, volume 64. Springer Science & Business Media, 2010.

[46] C. Reisinger and W. Stockinger. An adaptive Euler-Maruyama scheme for Mckean-Vlasov SDEs with super-linear growth and application to the mean-field FitzHugh-Nagumo model. *arXiv preprint arXiv:2005.06034*, 2020.

[47] Q. Ren and H. Tian. Mean-square convergence and stability of two-step Milstein methods for stochastic differential equations with Poisson jumps. *Computational and Applied Mathematics*, 41(3):1–18, 2022.

[48] S. Rong. *Theory of stochastic differential equations with jumps and applications: mathematical and analytical techniques with applications to engineering*. Springer Science & Business Media, 2006.

[49] T. Shardlow and P. Taylor. On the pathwise approximation of stochastic differential equations. *BIT Numerical Mathematics*, 56(3):1101–1129, 2016.

[50] A. Shiryaev. *Probability*. Springer, Berlin, 2 edition, 1996.

[51] S. E. Shreve et al. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer, 2004.

[52] X. Sun, J. Bao, and Y. Shao. Mathematical modeling of therapy-induced cancer drug resistance: connecting cancer mechanisms to population survival rates. *Scientific reports*, 6(1):1–12, 2016.

[53] P. Tankov and R. Cont. *Financial modelling with jump process*. CHAPMAN and HALLI/CRC, New York, 2004.

[54] S. J. Taylor. Modeling stochastic volatility: A review and comparative study. *Mathematical finance*, 4(2):183–204, 1994.

[55] M. V. Tretyakov and Z. Zhang. A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications. *SIAM Journal on Numerical Analysis*, 51(6):3135–3162, 2013.

[56] T. Von Zglinicki. Oxidative stress shortens telomeres. *Trends in biochemical sciences*, 27(7):339–344, 2002.

[57] X. Wang and S. Gan. The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients. *Journal of Difference Equations and Applications*, 19(3):466–490, 2013.

[58] E. Wong. Stochastic neural networks. *Algorithmica*, 6(1):466–478, 1991.

[59] J. Yao and S. Gan. Stability of the drift-implicit and double-implicit Milstein schemes for nonlinear SDEs. *Applied Mathematics and Computation*, 339:294–301, 12 2018.

[60] C. Zhou. A jump-diffusion approach to modeling credit risk and valuing defaultable securities. *Available at SSRN 39800*, 1997.