Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
Wydział Matematyki Stosowanej
Katedra Analizy Matematycznej, Matematyki Obliczeniowej
i Metod Probabilistycznych

Rozprawa doktorska

Optimal algorithms for solving stochastic initial-value problems with jumps

Andrzej Kałuża

Promotor: dr hab. Paweł Przybyłowicz, prof. AGH

Kraków 2020

I dedicate this work to my wife Joanna and my daughter Maja for their support, understanding and patience.

To my supervisor Professor Paweł Przybyłowicz for the guidance, direction, encouragement, and advice.

Contents

Streszczenie

W rozprawie zajmujemy się problemem aproksymacji stochastycznych równań różniczkowych następującej postaci

$$\begin{cases} \mathrm{d}X(t) = a(t, X(t))\mathrm{d}t + b(t, X(t))\mathrm{d}W(t) + c(t, X(t-))\mathrm{d}N(t), \quad t \in [0, T], \\ X(0) = x_0, \end{cases}$$

gdzie $T > 0$, $N = \{N(t)\}_{t\in[0,T]}$ jest jednowymiarowym niejednorodnym procesem Poissona z intensywnością λ, $W = \{W(t)\}_{t\in[0,T]}$ jest m_{w}-wymiarowym procesem Wienera. Rozprawa składa się z trzech głównych części.

W pierwszej części rozważamy problem skalarny z jednowymiarowym procesem Wienera. Analizujemy w niej algorytm oparty na adaptacyjnej kontroli długości kroku całkowania. Bazując na kawałkami liniowej interpolacji wartości schematu Milsteina obliczonego w punktach wyznaczonej siatki, otrzymujemy aproksymację rozwiązania. W tej części rozprawy analizujemy również błąd metody nie używającej wartości pochodnych cząstkowych współczynnika dyfuzji. Dla obu metod wyznaczamy dokładne tempo zbieżności wraz z postacią stałych asymptotycznych. Ponadto uzyskane wyniki implikują optymalność zdefiniowanych algorytmów w rozważanych klasach metod.

W kolejnej części rozprawy rozważane są układy stochastycznych równań różniczkowych ze skokami w przypadku wielowymiarowego procesu Wienera. Jak w poprzedniej części rozprawy do aproksymacji rozwiązania wykorzystujemy interpolację kawałkami liniową wartości schematu Milsteina obliczonego w punktach siatki jednostajnej. Ponownie pokazujemy dokładne tempo zbieżności zdefiniowanego algorytmu wraz z postacią stałej asymptotycznej. Udowadniamy ponadto odpowiednio oszacowania z dołu na błąd, z których wynika optymalność skonstruowanej metody.

W trzeciej części pracy prezentujemy krótkie wprowadzenie do języka programowania CUDA C wraz z efektywną implementacją algorytmu optymalnego z drugiej części rozprawy. Przedstawiamy również wyniki przeprowadzonych eksperymentów numerycznych.

Słowa kluczowe

Analityczna złożoność obliczeniowa, stochastyczne równania różniczkowe ze skokami, informacja standardowa, n-ty bł ąd minimalny, asymptotycznie optymalana metoda, CUDA C

Abstract

In the thesis we study the problem of approximation of solutions of stochastic differential equations of the form

$$\begin{cases} \mathrm{d}X(t) = a(t, X(t))\mathrm{d}t + b(t, X(t))\mathrm{d}W(t) + c(t, X(t-))\mathrm{d}N(t), & t \in [0, T], \\ X(0) = x_0, \end{cases}$$

where $T > 0$, and $N = \{N(t)\}_{t\in[0,T]}$ is a one-dimensional non–homogeneous Poisson process, with intensity function λ, and $W = \{W(t)\}_{t\in[0,T]}$ is a m_{w}-dimensional Wiener process. The thesis consists of three main parts.

In the first part of thesis we investigate the scalar problem with $m_{\mathrm{w}} = 1$. We analyze algorithm based on path-independent adaptive step-size control. The method computes the adaptive discretization and next it uses a piecewise linear interpolation of the classical Milstein steps performed at the computed sampling points. We also analyze derivative-free version of this method. For the both methods we investigate the exact rate of convergence of the nth errors together with the asymptotic constants. Moreover, it turns out that the both methods are asymptotically optimal in certain class of algorithms.

In the second part of the thesis we investigate the systems of SDEs with $m_{\mathrm{w}} \geqslant 1$. We provide a construction of a suitable algorithm that is based on equidistant discretization. At the sampling points the method uses a piecewise linear interpolation of the classical Milstein steps. Again we show the exact rate of convergence of the defined method together with the asymptotic constants. We also provide corresponding sharp lower bounds which imply that the constructed method is asymptotically optimal.

In the third part of thesis we present introduction to CUDA C programming language together with efficient implementation of the optimal algorithm from the part two of the thesis. We also show numerical results that confirm our theoretical findings.

Key words

Information-Based Complexity, stochastic differential equation with jumps, standard information, nth minimal error, asymptotically optimal method, CUDA C

Introduction

Over the last years the number of publications devoted to stochastic problems, including the approximation of solutions of stochastic differential equations (SDEs) with jumps, has increased dramatically. One of the possibility which causes this behavior is the fact that the demand for this type of modeling is rapidly increasing. The areas where such SDEs problems find applications are for example, financial mathematics, physics, biology, and engineering, see [11, 19, 39, 61, 78]. The discussed equations often do not have analytical solutions and the use of efficient approximate methods is a necessity.

The first monograph which investigates to the topic of approximation of SDEs is [31] (new release [32]). Authors describe construction of algorithms based on Itô-Taylor expansions. Additionally, the authors investigate rate of convergence of the proposed algorithms for the strong approximation (where we approximate trajectories of solutions) and the weak approximation (where we approximate moments of solutions). Another main reference which investigates stochastic problems is [39]. Authors, apart from the results known from the monograph [31], investigate approximation of deterministic problems using probabilistic methods. They also investigate stochastic differential equations in presence of small noise and stochastic Hamiltonian systems. In both monographs authors focused on finding upper bounds for error of considered algorithms and the stability of the considered methods. The optimality of presented schemes was not discussed. Another main monograph dealing with SDEs with jumps is [61]. Authors concentrate on designing and analysing of discrete-time approximations for SDEs with jumps. Authors present theoretical background for SDEs with jumps motivated by several application from finance. They analyze stochastic expansion for a different order of schemes. They also investigate strong and weak approximation and derivative-free schemes.

Information Based Complexity (IBC) is a branch of numerical analysis, which deals with complexity of problems where information is partial, priced, and sometimes noisy. Partial means that multiple problems may share the same information, priced means that the cost of an algorithm is directly connected with the number and precision of observations, and noisy corresponds to some corruption for the observed values. One of the main tasks of IBC is giving answers about the minimal cost that is needed for solving a problem with the error at most ε, such minimal cost is called ε-complexity. Similarly, the problem of the nth minimal error in a given class of algorithms is also considered. The nth minimal error is defined as a minimal error of an algorithm that can be reached in a class of algorithms with a cost at most n. In this work we are interested in finding essential sharp lower and upper bounds for the nth minimal error in the context of stochastic problems. It should be stressed that the nth minimal error corresponds to the problem, not to particular algorithm.

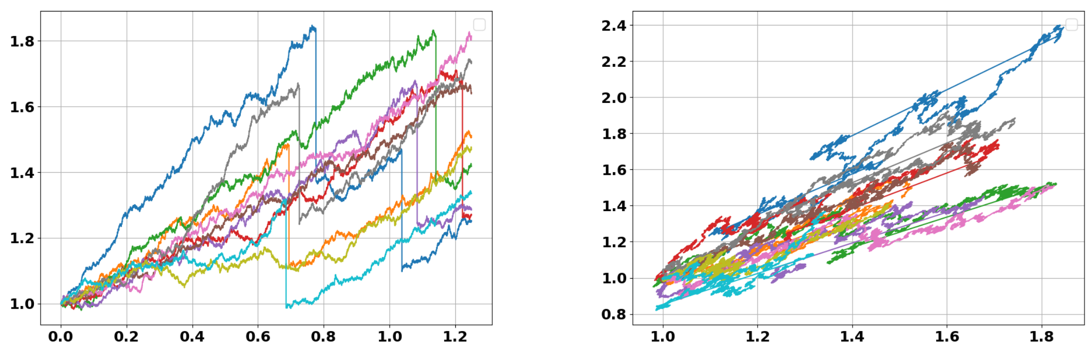
As a cornerstone and a kind of determinant that still determines the paradigms of studying computational problems in terms of their computational complexity, we can mention two books [80] and [79]. As a continuity of those, we can distinguish [46], where authors consider problems in the multidimensional case and analyzing the impact of the dimension on the complexity of a problem. The main problems considered in IBC are finding methods for solving mathematical problems such as approximation of functions (for example [53, 55–57, 60]), integration (for example [54, 55, 58]), optimal approximation of ordinary (for example [20–22]), partial (for example [12, 59, 85]), integral differential equations (for example [13, 84]), stochastic integration (for example [10, 14, 23, 63–65]), approximation of stochastic differential equations (for example [7, 8, 16, 30, 33, 43–45, 66–73, 75, 76]). We can highlight diffrent types of model of computations, the worst-case, asymptotic, average, randomized, and quantum settings. The problems with noisy information are also considered, for example [23, 40–42, 50–53].

In parallel to the development of theory there is a huge development of hardware which allows to prepare suitable algorithms which can compute solutions in acceptable time. Parallel computation, during the several decades, has been more and more popular in the world of computations. There are also a lot of problems which need parallel computation to get the solution in a reasonable time. Primary goal of parallel computation is to improve application's performance. Mathematical problems, for example approximation of stochastic differential equations, require simulations of huge number of independent trajectories, and it makes this type of problems computationally costly. Multiprocessing is a natural tool which can be applied to solve this issues. By employing CUDA technology and dedicated programming language CUDA C, we can create applications of high performance, which solve mathematical problems efficiently, e.g. matrix multiplication or approximation of stochastic problems. For example the documents [4, 28, 47, 74] contain a lot of information about the CUDA C programming language, together with examples, which can help to create applications.

In the thesis we deal with the global approximation of solutions of systems of stochastic differential equations (SDEs) of the following form

$$\begin{cases} \mathrm{d}X(t) = a(t, X(t))\mathrm{d}t + b(t, X(t))\mathrm{d}W(t) + c(t, X(t-))\mathrm{d}N(t), \quad t \in [0, T], \\ X(0) = x_0, \quad x_0 \in \mathbb{R}^d \end{cases} \tag{1}$$

where $T > 0$, and $N = \{N(t)\}_{t\in[0,T]}$ is a one-dimensional non–homogeneous Poisson process and $W = \{W(t)\}_{t\in[0,T]}$ is a m_{w}-dimensional Wiener process. There are a lot of positions in the literature which consider optimal approximation of solutions of SDEs driven only by the Wiener process. In that case both upper and lower bounds on error were established for the strong approximation, see, for example, [17, 18, 44, 67, 68].



(a) 1-dimensional case (b) 2-dimensional case

Figure 1: Examples of SDEs trajectories.

For a more complex problems, which also contain the jump term, suitable approximation schemes were provided, and upper bounds on their errors discussed. For example, the monograph [61] and in the articles [9, 15, 16, 34, 35] authors deal with the jump-diffusion SDEs. However, according to our best knowledge, till now there are only few papers that establish asymptotic lower bounds and exact rate of convergence of the minimal errors for the global approximation of the scalar SDEs with jumps, see [24, 69, 70, 72], and there are no articles addresing this problem in multidimensional case. In [69] the author considers the pure jump SDEs (1), i.e., $b \equiv 0$ and $c = c(t)$, while in [70] the general multiplicative case (1) is investigated. In [72] author provides a construction of a method based on path-dependent adaptive step-size control for global approximation of jump–diffusion SDEs. The discretization points and their number are chosen in adaptive way with respect to trajectories of the driving Poisson and Wiener processes. We also refer to [7], where the authors investigate the optimal rate of convergence for the problem of approximating stochastic integrals of regular functions with respect to a homogeneous Poisson process. In [70, 71] a suitable method has been defined and showed to be optimal. However, the optimal non-uniform discretization of the interval $[0, T]$ is defined in a non-constructive way. Therefore, the practical use of the method is highly limited. In the paper [24] authors show an implementable method based on path-independent adaptive step-size control that still preserves optimality properties. Such methods were constructed in pure Wiener case in several papers [18, 44]. However those methods were hard to implement.

In this thesis we present results based on [24] for the scalar case with $m_{\mathrm{w}} = 1$ and also not yet published results for the multi-dimensional case where with $m_{\mathrm{w}} \geqslant 1$. In both cases we assume that diffusion and jump coefficients satisfy the *jump commutative conditions* (see page 20. or 39.). Method constructed for the one-dimensional case is based on the path-independent adaptive step-size control. The method assumes that the step-size is adjusted at each step, but the adjustment is done independently of behavior of particulat trajectory. Roughly speaking it is adapted to the mean behavior of W and N. In a multidimensional case we analyze the exact rate of convergence of piecewise linear interpolation of the classical Milstein steps performed at equidistant discretization points. The main contributions of the thesis are

- construction and analysis of method based on path-independent adaptive step-size control for scalar SDEs with jumps driven by Wiener and Poisson processes,
- construction and analysis of method based on equidistant discretization for system of SDEs with jumps driven by Wiener and Poisson processes,
- establishing optimality of the considered methods,
- implementation of developed algorithms in CUDA C programming language.

The structure of the thesis is organized as follows. In Chapter 1 we show a short introduction to the computational model. In Chapter 2 we present definition of algorithm based on path-independent adaptive step-size control. The method computes the adaptive discretization and next it uses a piecewise linear interpolation of the classical Milstein steps performed at the computed sampling points. The construction of algorithm is computer implementable. We denote it by $\bar{X}^{Lin-M*}$. We also investigate a derivative free version $\bar{X}^{df-Lin-M*} = \{\bar{X}_{k_n}^{df-Lin-M*}\}$ of this algorithm. Both methods compute the adaptive discretization and then use a piecewise linear interpolation of the classical Milstein steps performed at the computed sampling points. Moreover, by the results of [70], the algorithms $\bar{X}^{Lin-M*}$ and $\bar{X}^{df-Lin-M*} = \{\bar{X}^{df-Lin-M*}_{k_n}\}$ are asymptotically optimal.

The main results of this chapter are Theorem 2.4 and Theorem 2.6 which states that for the method $\bar{X}^{Lin-M*}$ and $\bar{X}^{df-Lin-M*}$ we have that error behaves like

$$\lim_{n\to+\infty} k_n^{1/2} \cdot \left(\mathbb{E}\int_0^T \left|X(t) - \bar{X}^{Lin-M*}_{k_n}(t)\right|^2 \mathrm{d}t\right)^{1/2} = \frac{1}{\sqrt{6}} \int_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t, \tag{2}$$

$$\lim_{n\to+\infty} k_n^{1/2} \cdot \left(\mathbb{E}\int_0^T \left|X(t) - \bar{X}^{df-Lin-M*}_{k_n}(t)\right|^2 \mathrm{d}t\right)^{1/2} = \frac{1}{\sqrt{6}} \int_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t, \tag{3}$$

where $\mathcal{Y}(t) = |b(t, X(t))|^2 + \lambda(t) \cdot |c(t, X(t))|^2$ and k_n is the number of evaluations of the Poisson and Wiener processes. The number k_n is also adapted to the diffusion and jump coefficients, and to the intensity function λ. For the both methods we investigate the exact rate of convergence of the nth errors together with the asymptotic constants. Moreover, it turns out that the both methods are asymptotically optimal in certain class of algorithms. It means that nth minimal error behaves like $\Theta(n^{-1/2})$ in the considered class of algorithms (see Theorem 2.8).

Chapter 3 is dedicated to analyzis of the classical Milstein algorithm based on equidistant discretization for system of SDEs with jumps with multidimensional Wiener process. We construct an implementable algorithm, denoted by $\bar{X}^{Lin-M*} = \{\bar{X}^{Lin-M*}_n\}$ and we stress its ease in implementation. The method uses a piecewise linear interpolation of the classical Milstein steps performed at the sampling points. The main results of this chapter are Theorem 3.1 and Theorem 3.2, which imply the optimality of method $\bar{X}^{Lin-M*}$ in some class of algorithms (Theorem 3.4). By the Theorems we have that for the method $\bar{X}^{Lin-M*}$ the following estimations hold

$$\lim_{n\to+\infty} n^{1/2} \cdot \left(\mathbb{E}\int_0^T \left\|X(t) - \bar{X}^{Lin-M*}_{n}(t)\right\|^2 \mathrm{d}t\right)^{1/2} = \sqrt{\frac{T}{6}} \left(\int_0^T \mathbb{E}(\mathcal{Y}(t)) \mathrm{d}t\right)^{1/2}, \tag{4}$$

where $\mathcal{Y}(t) = \|b(t, X(t))\|_F^2 + \lambda(t) \cdot \|c(t, X(t))\|_F^2$, n is the number of evaluations of the Poisson and Wiener processes. For method we investigate the exact rate of convergence of the nth errors together with the asymptotic constants. Moreover, it turns out that method is asymptotically optimal and the nth minimal error behaves like $\Theta(n^{-1/2})$ in a considered class of algorithms.

In Chapter 4 we present simple notation and basic information about technology of CUDA and CUDA C programming language. We show simple introduction to CUDA C, which allows reader to understand the implementation of algorithm form Chapter 2. At the end of this section we show results from numerical experiments performed for algorithms from Chapter 2 and 3, which confirm theoretical results.

In Chapter 5 we simply conclude results and define open problem corresponding to considered problems.

Appendix A contains a theoretical background about random variables, stochastic processes, martingales, Itô integration with the respect to semi-martingales, stochastic differential equations, and other useful facts.

Finally Appendix B contains proofs of main Theorems and Lemmas, which are useful in proving of main results of thesis presented in Chapter 2 and 3. Most of given facts in this section were provided by us. As a main result in this section we can listed proofs of Theorem B.1 and Theorem B.13 which say about boundary and convergence of Milstein approximation in space $\mathfrak{L}^2(\Omega\times[0,T])$ for Time Continuous Milstein Scheme and derivative free version. A similar result has been justified in Theorem 6.4.1 in [61], however, under slightly stronger assumptions. In particular, in this thesis we do not assume the existence of continuous partial derivative $\partial f/\partial t$ for $f \in \{a,b,c\}$ and we do not assume any Lipschitz conditions for the second order partial derivatives of $f = f(t,y), f \in \{a,b,c\}$, with respect to y. Moreover, we consider non-homogeneous Poisson process, while in [61] in Theorem 6.4.1 has been shown only for homogeneous counting processes.

Symbols

$\mathbb{N} = \{1, 2, 3, \ldots\}$	:	set of natural numbers
$\mathbb{N}_0 = \mathbb{N} \cup \{0\}$	:	set of natural numbers with zero
$\mathbb{R} = (-\infty, +\infty)$	:	set of real numbers
$\mathbb{R}_+ = (0, +\infty)$	:	set of non negative real numbers
$\mathbb{R}^d$	:	d-dimensional euclidean space
$(a, b]^d$	:	d-dimensional interval given by $(a, b] \times \ldots \times (a, b]$
$\lvert \cdot \rvert$	:	absolute value
$\lVert \cdot \rVert_2$	:	second euclidean norm
$\lVert \cdot \rVert_F$	:	Frobenius matrix norm
$x_n \uparrow \infty$	:	increasing to infinity sequence of x_n
$max(a, b) = a \vee b$	:	maximum of $a, b \in \mathbb{R}$
$min(a, b) = a \wedge b$	:	minimum of $a, b \in \mathbb{R}$
$y = (y_1, \ldots, y_d)^T$	:	column vector $y \in \mathbb{R}^d$ with ith component x_i
$e_j = (0, \ldots, 0, 1, 0, \ldots, 0)^T$	:	$e_j \in \mathbb{R}^d$, $j \in \{1, \ldots, d\}$ vector where non-zero element is on the jth place
$\alpha \cdot y = y \cdot \alpha$	=	$(\alpha y_1, \ldots, \alpha y_d)^T$ for $y \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$
$A = [a^{i,j}]_{i,j=1}^{d,k} = [a_i]_{i=1}^{d} = [a^j]_{j=1}^{k}$	:	$(k \times d)$-matrix A with ijth component $a^{i,j}$, ith row a_i and jth column a^j

$\mathcal{F}, \mathcal{G}, \mathcal{H}$	:	collections of events, σ-algebras
$(\Omega, \mathcal{F}, \mathbb{P})$	:	probability space
X, Y	:	real valued random variable
$\mathbb{E}(Y)$	:	expected value of Y
$\mathbb{E}(Y \mid \mathcal{G})$	:	conditional expectation of Y under $\mathcal{G}$
τ, σ	:	stopping time
$\{\mathcal{F}_t\}_{t \geq 0}$	:	filtration
$\sigma(Y)$	:	σ-algebra generated by random variable Y
$\sigma(\mathcal{A})$	:	σ-algebra generated by collection $\mathcal{A}$
$\mathcal{F} \vee \mathcal{G}$	=	$\sigma(\mathcal{F} \cup \mathcal{G})$

$\mathcal{F} \otimes \mathcal{G}$	$=$	$\sigma(\{F \times G \ : \ F \in \mathcal{F}, G \in \mathcal{G}\})$
$\mathfrak{L}^2(\Omega, \mathcal{F}, \mathbb{P})$	$=$	$\mathfrak{L}^2(\Omega)$ space of square integrable random variables
$\|X\|_{\mathfrak{L}^2(\Omega)} := \left(\mathbb{E}\|X\|^2\right)^{1/2}$	:	norm of X in $\mathfrak{L}^2(\Omega)$
$\mathfrak{L}^2(\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}([0,T]), \mathbb{P} \otimes \lambda_1)$	$=$	$\mathfrak{L}^2(\Omega \times [0,T])$ space of square integrable stochastic processes
$\|Y\|_{\mathfrak{L}^2(\Omega \times [0,T])} := \left(\mathbb{E}\int_0^T \|Y(t)\|^2 \mathrm{d}t\right)^{1/2}$	:	norm of X in $\mathfrak{L}^2(\Omega \times [0,T])$
$\mathtt{N}(\mu, \sigma)$	:	normal distribution with mean μ and standard deviation σ
$\mathtt{Poiss}(\lambda)$	:	Poisson distribution with intensivity λ
$m(t)$	$=$	$\int_0^t \lambda(s)\mathrm{d}s$, for $t \geqslant 0$
$\Lambda(t,s)$	$=$	$m(t) - m(s)$ for $t, s \in [0,T]$
$\\| \cdot \\|_\infty$	:	supremum norm of a function
$\bar{\omega}(f,\delta)$	$=$	$\sup_{t,s\in[0,T], \|t-s\|\leq\delta} \|f(t) - f(s)\|, \ \delta \in [0,+\infty)$ modulus of continuity for a continuous function $f : [0,T] \to \mathbb{R}$,
$\frac{\partial^{\|\alpha\|} f}{\partial y^\alpha}$	:	$\alpha \in \mathbb{N}_0^d$ where $\|\alpha\| = \sum_{i=1}^d \alpha_i$ $\partial y^\alpha = \partial y_1^{\alpha_1} \dots \partial y_d^{\alpha_d}$
a.s.	:	almost surely

For a function $f, g : \mathbb{R} \to \mathbb{R}$

$f(x) = \mathcal{O}(g(x))$	:	$\exists_{x_0\in\mathbb{R}} \exists_{c_1>0} \forall_{x\geqslant x_0} \ \|f(x)\| \leqslant C\|g(x)\|$
$f(x) = \Omega(g(x))$	:	$\exists_{x_0\in\mathbb{R}} \exists_{c_1>0} \forall_{x\geqslant x_0} \ \|f(x)\| \geqslant C\|g(x)\|$
$f(x) = \Theta(g(x))$	:	$f(x) = \mathcal{O}(g(x))$ and $f(x) = \Omega(g(x))$

For a function $f:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$, $h>0$

$$
\begin{aligned}
\frac{\partial f}{\partial x_i}(t,x) &= \left(\frac{\partial f_1}{\partial x_i}(t,x),\dots,\frac{\partial f_d}{\partial x_i}(t,x)\right)^T\\
\frac{\partial^2 f}{\partial x_i\partial x_k}(t,x) &= \left(\frac{\partial^2 f_1}{\partial x_i\partial x_k}(t,x),\dots,\frac{\partial^2 f_d}{\partial x_i\partial x_k}(t,x)\right)^T\\
\nabla_x f(t,x) &= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(t,x) & \frac{\partial f_1}{\partial x_2}(t,x) & \dots & \frac{\partial f_1}{\partial x_d}(t,x)\\ \frac{\partial f_2}{\partial x_1}(t,x) & \frac{\partial f_2}{\partial x_2}(t,x) & \dots & \frac{\partial f_2}{\partial x_d}(t,x)\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial f_d}{\partial x_1}(t,x) & \frac{\partial f_d}{\partial x_2}(t,x) & \dots & \frac{\partial f_d}{\partial x_d}(t,x), \end{pmatrix}\\
\widetilde{\nabla}_{x,h} f(t,x) &= \begin{pmatrix} \frac{f_1(t,x+h\cdot e_1)-f_1(t,x)}{h} & \frac{f_1(t,x+h\cdot e_2)-f_1(t,x)}{h} & \dots & \frac{f_1(t,x+he_d)-f_1(t,x)}{h}\\ \frac{f_2(t,x+h\cdot e_1)-f_2(t,x)}{h} & \frac{f_2(t,x+h\cdot e_2)-f_2(t,x)}{h} & \dots & \frac{f_2(t,x+h\cdot e_d)-f_1(t,x)}{h}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{f_d(t,x+h\cdot e_1)-f_d(t,x)}{h} & \frac{f_d(t,x+h\cdot e_2)-f_d(t,x)}{h} & \dots & \frac{f_d(t,x+h\cdot e_d)-f_d(t,x)}{h} \end{pmatrix}
\end{aligned}
$$

Let $f:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$, $b:[0,T]\times\mathbb{R}^d\to\mathbb{R}^{d\times m_{\mathrm{w}}}$ and $c:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$. For $k\in\{1,\dots,m_{\mathrm{w}}\},(t,x)\in[0,T]\times\mathbb{R}^d,\ h>0$ we use the following notation

$$
\begin{aligned}
L_k f(t,x) &= \nabla_x f(t,x)\cdot b^k(t,x)\\
\mathcal{L}_{k,h} f(t,x) &= \widetilde{\nabla}_{x,h} f(t,x)\cdot b^k(t,x)\\
L_{-1} f(t,x) &= f(t,x+c(t,x))-f(t,x).
\end{aligned}
$$

If $d=m_{\mathrm{w}}=1$ we write

$$
\begin{aligned}
L_1 f(t,x) &= b(t,x)\cdot\frac{\partial f}{\partial x}(t,x)\\
\mathcal{L}_{1,h} f(t,x) &= \frac{f(t,x+h)-f(t,x)}{h}\cdot b(t,x)
\end{aligned}
$$

Additionally, all constants that appear in the estimations will depend only on the parameters of the problem and T, unless it is clearly stated otherwise. Moreover, to simplify nomenclature and numbers of different symbols we assume, that the same symbol can be used to indicate different constants. As we consider only asymptotic case, the exact value of constants is not investigated.

Chapter 1

General description of the problem and aim of the thesis

The aim of this thesis is to present a construction of optimal algorithms for the global approximation of solutions of d-dimensional system of stochastic differential equations (SDEs) of the following form

$$\begin{cases} \mathrm{d}X(t) = a(t, X(t))\mathrm{d}t + b(t, X(t))\mathrm{d}W(t) + c(t, X(t-))\mathrm{d}N(t), \quad t \in [0, T], \\ X(0) = x_0, \end{cases} \tag{1.1}$$

where $T > 0$, and $N = \{N(t)\}_{t\in[0,T]}$ is a one-dimensional non–homogeneous Poisson process, with intensity function λ, and $W = \{W(t)\}_{t\in[0,T]}$ is a m_{w}-dimensional Wiener process.

First, we will investigate the problem of approximating solutions of scalar stochastic differential equations (1.1) where $d = m_{\mathrm{w}} = 1$. Then, we will focus on d-dimensional system of stochastic differential equations (1.1), where $d \geqslant 1$ driven by m_{w} ($m_{\mathrm{w}} \geqslant 1$) independent Wiener processes.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space for both stochastic processes (see Appendix A.2). Both problems will be considered in special classes of functions a, b, c, λ. Let us know that our problem can be defined as a five-elements vector (a, b, c, λ, x_0).

Information

In our model of computation, we assume that we do not have the complete knowledge about realizations of Wiener and Poisson processes on considered interval

[0,T]. Instead, we can use only partial and standard information of evaluations of the Poisson and Wiener processes $\mathcal{N}_n(N, W)$, where $\mathcal{N}_n(N, W) : \Omega \to \mathbb{R}^{n\cdot(m_w+1)}$ is given as vector of evaluation of processes in given sampling points.

$$\mathcal{N}_n(N, W) := \Big[N(t_{1,n}), N(t_{2,n}), \ldots, N(t_{n,n}), W(t_{1,n}), W(t_{2,n}), \ldots, W(t_{n,n})\Big], \tag{1.2}$$

where points $t_{i,n}$ for $i \in \{0, 1, \ldots, n\}$ belong to partition of interval $[0, T]$ given by

$$\Delta_n = \{t_{0,n}, t_{1,n}, \ldots, t_{n,n}\}, \quad 0 = t_{0,n} < t_{1,n} < \ldots < t_{n,n} = T. \tag{1.3}$$

We denote by

$$\mathcal{N}(N, W) = \big\{\mathcal{N}_n(N, W)\big\}_{n\in\mathbb{N}} \tag{1.4}$$

the sequence of vectors $\mathcal{N}_n(N, W)$, where each provides standard information of the Poisson and Wiener processes.

For a single process $Z \in \{N, W, W_1, \ldots, W_{m_w}\}$ we use the notation

$$\mathcal{N}_n(Z) := \big[Z(t_{1,n}), Z(t_{2,n}), \ldots, Z(t_{n,n})\big].$$

It is important to know that $N(0) = 0$ and $W(0) = \bar{0}$. The information used to solve a problem may be *non-adaptive* or *adaptive*. We say that information is *non-adaptive* when we choose the points in advance (a priori). We say that the information is *adaptive* when discretization points are not given in advance and every next point is calculated using previous computations/observations. Especially the sequences of discretizations $\bar{\Delta} = \{\Delta_n\}_{n\in\mathbb{N}}$ may depend on functions a, b, c, λ and on initial value x_0. We also assume, that discretization does not depend on trajectories of the processes N and W. Information (1.2) uses the same evaluation points for all trajectories of the Poisson and Wiener processes. Therefore, the information (1.4) about the processes N and W is non-adaptive.

Algorithm

After computing the information $\mathcal{N}_n(N, W)$, we approximate solutions of our problem by an element in solution space $\mathfrak{L}^2\big([0, T]; \mathbb{R}^d\big)$. We apply the *algorithm* which is represented by Borel measurable mapping

$$\varphi_n : \mathbb{R}^{n\cdot(m_w+1)} \to \mathfrak{L}^2([0, T]; \mathbb{R}^d), \tag{1.5}$$

in order to obtain the nth approximation $\bar{X}_n = \big\{\bar{X}_n(t)\big\}_{t\in[0,T]}$ in the following way

$$\bar{X}_n = \varphi_n(\mathcal{N}_n(N, W)). \tag{1.6}$$

It is important to have a tool which allows us to compare exact solutions and approximations given by algorithms. In Appendix A.5 we discuss how to move between spaces of solutions and approximation.

So any approximation method $\bar{X} = \{\bar{X}_n\}_{n\in\mathbb{N}}$ can be defined by two sequences $\bar{\varphi} = \{\varphi_n\}_{n\in\mathbb{N}}$, $\bar{\Delta} = \{\Delta_n\}_{n\in\mathbb{N}}$.

The nth cost of the method $\bar{X}$ is defined as the total number of evaluations of N and W used by the nth approximation $\bar{X}_n$. In literature the cost of algorithm is also named cardinality of information. For considered in the thesis problem we define nth cost in the following way

$$cost_n(\bar{X}) = \begin{cases} (m_{\mathrm{w}}+1)\cdot n, & \text{if } b \not\equiv 0 \text{ and } c \not\equiv 0, \\ m_{\mathrm{w}}\cdot n, & \text{if } b \not\equiv 0 \text{ and } c \equiv 0, \\ n, & \text{if } b \equiv 0 \text{ and } c \not\equiv 0, \\ 0, & \text{if } b \equiv 0 \text{ and } c \equiv 0. \end{cases}$$

Cost calculation does not include combinatoric cost, which is defined as a total number of arithmetic operations used to calculate approximation. According to literature, we assume that the cost of information is greater than the cost of arithmetic operation. The class of all methods $\bar{X} = \{\bar{X}_n\}_{n\in\mathbb{N}}$, defined as above, is denoted by χ^{noneq}. Moreover, we consider the following subclass of χ^{noneq} defined as

$$\chi^{\mathrm{eq}} = \Big\{\bar{X} \in \chi^{\mathrm{noneq}} \;\Big|\; \exists_{n_0^*=n_0^*(\bar{X})\in\mathbb{N}} : \forall_{n\geqslant n_0^*}\ \Delta_n = \big\{iT/n \ : \ i = 0,1,\ldots,n\big\}\Big\}.$$

Methods based on the sequence of equidistant discretizations (1.3) belong to the class χ^{eq}, while methods that evaluate N and W at the same, possibly non-uniform, sampling points belong to the class χ^{noneq}. Of course, we have that $\chi^{\mathrm{eq}} \subset \chi^{\mathrm{noneq}}$.

The nth minimal error

To measure and compare the quality of algorithms we need to define specific criteria. The *nth error of a method* $\bar{X} = \{\bar{X}_n\}_{n\in\mathbb{N}}$ is defined as

$$e_n(\bar{X}) = \|X - \bar{X}_n\|_{\mathfrak{L}^2(\Omega\times[0,T])} = \Big(\mathbb{E}\int_0^T \big\|X(t) - \bar{X}_n(t)\big\|^2 \mathrm{d}t\Big)^{1/2}.$$

It is an average error of approximation taken over the whole possible trajectories dependent on realization of stochastic processes W and N. *The nth minimal error* (see, for example, [80]), in the respective class of methods under consideration, is defined by

$$e^{\diamond}(n) = \inf_{\bar{X}\in\chi^{\diamond}} e_n(\bar{X}), \ \diamond \in \{\mathrm{eq},\ \mathrm{noneq}\}. \tag{1.7}$$

Hence, (1.7) is the minimal possible error among all algorithms (from respective class) that use n evaluation of N and W.

Chapter 2

Global approximation of solutions of scalar SDEs with jumps

In this chapter we consider the problem of approximation of solutions of scalar stochastic differential equations of the form (1.1), where $T > 0$, and $N = \{N(t)\}_{t\in[0,T]}$ is a one-dimensional non–homogeneous Poisson process, and $W = \{W(t)\}_{t\in[0,T]}$ is a one-dimensional Wiener process. This Chapter is based on the article [24].

2.1. The setting

Let $T > 0$ be a given real number and $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. We consider on it two independent processes a one-dimensional Wiener process

$$W = \big\{W(t)\big\}_{t\in[0,T]},$$

and a one–dimensional non–homogeneous Poisson process

$$N = \big\{N(t)\big\}_{t\in[0,T]},$$

with continuous intensity function $\lambda = \lambda(t)$. Let us denote by $\big\{\mathcal{F}_t\big\}_{t\in[0,T]}$ the complete filtration, generated by the driving processes N and W.

Now we specify the assumptions about functions which build the problem (1.1). For a given function $f \in \{a, b, c\}$, $f : [0, T] \times \mathbb{R} \to \mathbb{R}$, we assume that f satisfy the following conditions

(A) $f \in C^{0,2}([0, T] \times \mathbb{R})$.

(B) There exists $K > 0$ that for all $t, s \in [0,T]$ and all $y, z \in \mathbb{R}$,

$(B1)$ $|f(t,y) - f(t,z)| \leqslant K|y-z|$,

$(B2)$ $|f(t,y) - f(s,y)| \leqslant K\big(1+|y|\big)|t-s|$,

$(B3)$ $\left|\frac{\partial f}{\partial y}(t,y) - \frac{\partial f}{\partial y}(t,z)\right| \leqslant K|y-z|$.

(C) In addition, there exists $K > 0$ such that for a function $f \in \{b, c\}$ for all $t \in [0,T]$ and $y, z \in \mathbb{R}$

$$\big|L_1 f(t,y) - L_1 f(t,z)\big| \leqslant K|y-z|.$$

We will also assume that functions b and c satisfy *the jump commutation condition* (assumption (D)).

(D) For all $(t,y) \in [0,T] \times \mathbb{R}$,

$$L_{-1} b(t,y) = L_1 c(t,y). \tag{2.1}$$

This condition will allow the calculation of stochastic integrals defined in (B.4). More details about why we use this condition will be given in the next section where the algorithm will be analyzed. We also refer to Chapter 6.3 in [61] where the condition (2.1) is widely discussed.

Moreover for the intensity function $\lambda : [0,T] \to (0, +\infty)$ we assume that

(E) $\lambda \in C([0,T])$.

By Appendix A.4 and the fact that a, b, c and λ satisf $(B1)$, $(B2)$ and (E) the problem (1.1) has a unique strong solution $X = \{X(t)\}_{t\in[0,T]}$ that is adapted to filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ and has càdlàg paths.

The following result characterizes the *local mean square smoothness* of the solution X in terms of the following process

$$\mathcal{Y}(t) = \big|b(t,X(t))\big|^2 + \lambda(t)\cdot\big|c(t,X(t))\big|^2, \quad t \in [0,T]. \tag{2.2}$$

Proposition 2.1 ([70])**.** Let us assume that the functions a, b, c and λ satisfy the assumptions $(B1)$, $(B2)$ and (E). Then, we have for the solution X of problem (1.1) for all $t \in [0,T)$ that

$$\lim_{h\to 0+} \frac{\|X(t+h) - X(t)\|_{\mathfrak{L}^2(\Omega)}}{h^{1/2}} = \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2}.$$

Proposition 2.1 describes local mean square smoothness of the solution X. This local smoothness reflects in the exact rate of convergence of minimal errors established in [70] and will be used for the construction of optimal methods based on path-independent adaptive step-size control.

In order to characterize asymptotic lower bounds we define

$$C^{\mathrm{noneq}} = \frac{1}{\sqrt{6}} \int_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t,$$

$$C^{\mathrm{eq}} = \sqrt{\frac{T}{6}} \cdot \left(\int_0^T \mathbb{E}(\mathcal{Y}(t)) \mathrm{d}t\right)^{1/2},$$

where the process $\{\mathcal{Y}(t)\}_{t\in[0,T]}$ is defined in (2.2). We have that

(i) $0 \leqslant C^{\mathrm{noneq}} \leqslant C^{\mathrm{eq}}$,

(ii) $C^{\mathrm{noneq}} = C^{\mathrm{eq}}$ iff there exists $\gamma \geqslant 0$ such that for all $t \in [0,T]$

$$\mathbb{E}\big(\mathcal{Y}(t)\big) = \gamma,$$

(iii) $C^{\mathrm{eq}} = 0$ iff $C^{\mathrm{noneq}} = 0$ iff $b(t, X(t)) = 0 = c(t, X(t))$ for all $t \in [0,T]$ and almost surely.

2.2. Algorithm based on path-independent adaptive step-size control

In this section we present an implementable and asymptotically optimal algorithm in the class χ^{noneq}, which is based on the idea of adaptive step-size control. The step-size control will use the same sampling points for every trajectory of stochastic processes W and N, which means that it will be path-independent. Moreover, selection of mesh points will be based on the local Hölder regularity (see Proposition 2.1). Because of the fact that we do not know the precise value of $\mathbb{E}(\mathcal{Y}(t))$ for $t \in [0,T]$, we have to use suitable approximations. In addition, the adaptive sampling will be adjusted to the regularity of the intensity function λ, described in the terms of its modulus of continuity.

2.2.1. Description of the method and its asymptotic performance

We define the adaptive path-independent step-size control as follows.

STEP 0 Take an arbitrary strictly positive sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that

$$\lim_{n\to+\infty} \varepsilon_n = \lim_{n\to+\infty} \left(n^{1/2}\cdot\varepsilon_n\right)^{-1} = \lim_{n\to+\infty} \varepsilon_n^{-1}\cdot\bar{\omega}(\lambda, T\cdot(n\cdot\varepsilon_n)^{-1}) = 0, \tag{2.3}$$

where $\bar{\omega}$ is the modulus of continuity for λ (see Remark 2.2).

STEP 1 Take any $n\in\mathbb{N}$ and let $\hat{t}_{0,n} = 0$, $\bar{X}^M(\hat{t}_{0,n}) = x_0$. Set $i := 0$.

STEP 2 If $\hat{t}_{i,n}\in[0,T)$ and $\bar{X}^M(\hat{t}_{i,n})$ are given then compute

$$\hat{t}_{i+1,n} = \hat{t}_{i,n} + \frac{T}{n\cdot\max\Big\{\varepsilon_n, \Big(\mathbb{E}(\mathcal{Y}^M(\hat{t}_{i,n}))\Big)^{1/2}\Big\}}, \tag{2.4}$$

where

$$\mathcal{Y}^M(\hat{t}_{i,n}) = \left|b(\hat{t}_{i,n}, \bar{X}^M(\hat{t}_{i,n}))\right|^2 + \lambda(\hat{t}_{i,n})\cdot\left|c(\hat{t}_{i,n}, \bar{X}^M(\hat{t}_{i,n}))\right|^2.$$

If $\hat{t}_{i+1,n} < T$ then compute

$$\begin{aligned}
\bar{X}^M(\hat{t}_{i+1,n}) &= \bar{X}^M(\hat{t}_{i,n}) + a(\hat{t}_{i,n}, \bar{X}^M(\hat{t}_{i,n}))\cdot(\hat{t}_{i+1,n} - \hat{t}_{i,n})\\
&+ b(\hat{t}_{i,n}, \bar{X}^M(\hat{t}_{i,n}))\cdot(W(\hat{t}_{i+1,n}) - W(\hat{t}_{i,n}))\\
&+ c(\hat{t}_{i,n}, \bar{X}^M(\hat{t}_{i,n}))\cdot(N(\hat{t}_{i+1,n}) - N(\hat{t}_{i,n}))\\
&+ L_1 b(\hat{t}_{i,n}, \bar{X}^M(\hat{t}_{i,n}))\cdot I_{\hat{t}_{i,n},\hat{t}_{i+1,n}}(W,W)\\
&+ L_{-1} c(\hat{t}_{i,n}, \bar{X}^M(\hat{t}_{i,n}))\cdot I_{\hat{t}_{i,n},\hat{t}_{i+1,n}}(N,N)\\
&+ L_{-1} b(\hat{t}_{i,n}, \bar{X}^M(\hat{t}_{i,n}))\cdot\big(I_{\hat{t}_{i,n},\hat{t}_{i+1,n}}(W,N) + I_{\hat{t}_{i,n},\hat{t}_{i+1,n}}(N,W)\big),
\end{aligned} \tag{2.5}$$

take $i := i+1$ and GOTO **STEP 2**.

Else compute $\bar{X}^M(T)$ by (2.5) with $\hat{t}_{i+1,n}$ replaced by T.

STOP

Remark 2.2. *If $\lambda : [0,T]\to(0,+\infty)$ is a Hölder function with the exponent $\varrho\in(0,1]$ then in* **STEP 0** *we can take $\varepsilon_n = n^{-\varrho/(2(\varrho+1))}$.*

Now we analyze the algorithm and define the stopping criterion. Then we prove that the algorithm stops in a finite number of steps. Let us define

$$k_n = \min\left\{i\in\mathbb{N} \mid \hat{t}_{i,n}\geqslant T\right\}, \quad n\in\mathbb{N},$$

which is the total number of computed discretization points greater than $\hat{t}_{0,n} = 0$. The end point T is attainable, since we have for all $n\in\mathbb{N}$ that

$$k_n \leqslant \lceil n(\varepsilon_n + \hat{C})\rceil, \tag{2.6}$$

for some $\hat{C} < +\infty$, where existence of $\hat{C}$ follows from the Fact 2.3.

Fact 2.3. *Let $n \in \mathbb{N}$ and let us assume that there exists $M_n \in \mathbb{N}$ such that $\hat{t}_{j,n} \in [0,T)$ for all $j = 0,1,\ldots,M_n - 1$. Then*

$$\hat{t}_{M_n,n} \geqslant M_n \cdot \frac{T}{n(\varepsilon_n + \hat{C})},$$

where $\hat{C} = K_1\big(1+\bar{C}\big)\cdot\big(1+\|\lambda\|_\infty^{1/2}\big)$ and $\bar{C}$ is the constant from Theorem B.1.

Proof. Let us consider time-continuous Milstein approximation $\big\{\tilde{X}^M_{M_n}(t)\big\}_{t\in[0,T]}$ based on the mesh $0 = \hat{t}_{0,n} < \hat{t}_{1,n} < \ldots < \hat{t}_{M_n-1,n} < T$. Since $\bar{X}^M(0) = \tilde{X}^M_{M_n}(0) = x_0$, we have that

$$\bar{X}^M(\hat{t}_{j,n}) = \tilde{X}^M_{M_n}(\hat{t}_{j,n}), \quad j = 0,1,\ldots,M_n-1.$$

Hence, by Theorem B.1 we have that

$$\max_{0\leq j\leq M_n-1} \mathbb{E}\big|\bar{X}^M(\hat{t}_{j,n})\big|^2 \leqslant \bar{C}^2,$$

which yields for $j = 0,1,\ldots,M_n-1$ that

$$\max\Big\{\varepsilon_n, \Big(\mathbb{E}\big(\mathcal{Y}^M(\hat{t}_{j,n})\big)\Big)^{1/2}\Big\} \leqslant \varepsilon_n + \hat{C}. \tag{2.7}$$

Hence, by (2.4) and (2.7)

$$\hat{t}_{M_n,n} = \sum_{j=0}^{M_n-1} (\hat{t}_{j+1,n} - \hat{t}_{j,n}) \geqslant M_n \cdot \frac{T}{n(\varepsilon_n+\hat{C})},$$

which ends the proof. ∎

Hence, if for a given $n \in \mathbb{N}$ we have that $M_n = \lceil n(\varepsilon_n + \hat{C})\rceil$ then by Fact 2.3 we get $\hat{t}_{M_n,n} \geqslant T$. This implies (2.6) and the fact that algorithm stops in a finite number of steps.

Now, running n through the natural numbers, we obtain the sequence of discretizations $\hat{\Delta} = \big\{\hat{\Delta}_{k_n}\big\}_{n\in\mathbb{N}}$, where each $\hat{\Delta}_{k_n}$ is defined as

$$\hat{\Delta}_{k_n} = \big\{\hat{t}_{0,n}, \hat{t}_{1,n}, \ldots, \hat{t}_{k_n,n}\big\}, \quad n \in \mathbb{N}.$$

We have that $\hat{t}_{i,n} < T$ for all $i = 0,1,\ldots,k_n-1$ and $\hat{t}_{k_n,n} \geqslant T$. Since we can observe the Poisson and the Wiener processes only in interval $[0,T]$, we define the final sequence of discretizations $\hat{\Delta}^* = \big\{\hat{\Delta}^*_{k_n}\big\}_{n\in\mathbb{N}}$ by

$$\hat{\Delta}^*_{k_n} = \big(\hat{\Delta}_{k_n} \setminus \{\hat{t}_{k_n,n}\}\big) \cup \{T\} = \big\{\hat{t}^*_{0,n}, \hat{t}^*_{1,n}, \ldots, \hat{t}^*_{k_n,n}\big\}, \quad n \in \mathbb{N},$$

where $\hat{t}^*_{i,n} = \hat{t}_{i,n} < T$ for all $i = 0, 1, \ldots, k_n - 1$ and $\hat{t}^*_{k_n,n} = T \leqslant \hat{t}_{k_n,n}$. So now we observe processes only in given interval $[0, T]$.

By $\bar{X}^{cM*} = \{\bar{X}^{cM*}_{k_n}\}_{n\in\mathbb{N}}$ we denote the *conditional Milstein method* based on the sequence of discretizations $\hat{\Delta}^*$, which is defined as

$$\bar{X}^{cM*}_{k_n}(t) = \mathbb{E}\big(\tilde{X}^{M*}_{k_n}(t) \mid \mathcal{N}^*_{k_n}(N, W)\big), \quad t \in [0, T],$$

where $\{\tilde{X}^{M*}_{k_n}\}_{n\in\mathbb{N}}$ is a sequence of the time-continuous Milstein approximations (B.2) – (B.3) based on $\{\hat{\Delta}^*_{k_n}\}_{n\in\mathbb{N}}$ and information

$$\begin{aligned} \mathcal{N}^*_{k_n}(N, W) \;=\; & \big[N(\hat{t}^*_{1,n}), N(\hat{t}^*_{2,n}), \ldots, N(\hat{t}^*_{k_n,n}), \\ & W(\hat{t}^*_{1,n}), W(\hat{t}^*_{2,n}), \ldots, W(\hat{t}^*_{k_n,n})\big]. \end{aligned} \tag{2.8}$$

We also denote information with the respect to given process $Z \in \{N, W\}$ as

$$\mathcal{N}^*_{k_n}(Z) \;=\; \big[Z(\hat{t}^*_{1,n}), Z(\hat{t}^*_{2,n}), \ldots, Z(\hat{t}^*_{k_n,n})\big].$$

Following Lemma B.24, Lemma B.26, and Lemma B.27 (see also [70]) we can write that

$$\begin{aligned} \bar{X}^{cM*}_{k_n}(t) = \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n}) \;&+\; a(\hat{t}^*_{i,n}, \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n})) \cdot (t - \hat{t}^*_{i,n}) \\ &+\; b(\hat{t}^*_{i,n}, \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n})) \cdot \Delta W^*_{i,n} \cdot \frac{t - \hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}} \\ &+\; c(\hat{t}^*_{i,n}, \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n})) \cdot \Delta N^*_{i,n} \cdot \frac{\Lambda(t, \hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n}, \hat{t}^*_{i,n})} \\ &+\; L_1 b(\hat{t}^*_{i,n}, \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n})) \cdot I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(W, W) \cdot \left(\frac{t - \hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}}\right)^2 \\ &+\; L_{-1} b(\hat{t}^*_{i,n}, \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n})) \cdot \Delta N^*_{i,n} \cdot \Delta W^*_{i,n} \\ &\qquad\qquad \times \frac{\Lambda(t, \hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n}, \hat{t}^*_{i,n})} \cdot \frac{t - \hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}} \\ &+\; L_{-1} c(\hat{t}^*_{i,n}, \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n})) \cdot I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(N, N) \cdot \left(\frac{\Lambda(t, \hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n}, \hat{t}^*_{i,n})}\right)^2, \end{aligned}$$

for $t \in [\hat{t}^*_{i,n}, \hat{t}^*_{i+1,n}]$, $i = 0, 1, \ldots, k_n - 1$, $\bar{X}^{cM*}_{k_n}(0) = x_0$ and

$$\Delta W^*_{i,n} = W(t^*_{i+1,n}) - W(t^*_{i,n}),$$

$$\Delta N^*_{i,n} = N(t^*_{i+1,n}) - N(t^*_{i,n}).$$

Note that $\bar{X}^{cM*}_{k_n}$ has continuous trajectories and coincides with $\tilde{X}^{M*}_{k_n}$ at the discretization points. The disadvantage of this algorithm is the use of the values of Λ. Hence, we also define the piece-wise linear interpolation $\bar{X}^{Lin-M*}_{k_n}$ of the classical Milstein steps by

$$\bar{X}^{Lin-M*}_{k_n}(t) = \frac{\tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n})(\hat{t}^*_{i+1,n} - t) + \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i+1,n})(t - \hat{t}^*_{i,n})}{\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}},$$

for $t \in [\hat{t}^*_{i,n}, \hat{t}^*_{i+1,n}]$, $i = 0, 1, \ldots, k_n - 1$. In general, the method $\bar{X}^{cM*}_{k_n}$ is not equal to $\bar{X}^{Lin-M*}_{k_n}$. In discretization points we have that values of methods are equal, it means that

$$\bar{X}^{Lin-M*}_{k_n}(\hat{t}^*_{i,n}) = \bar{X}^{cM*}_{k_n}(\hat{t}^*_{i,n}) = \tilde{X}^{M}_{k_n}(\hat{t}^*_{i,n}) = \bar{X}^{M}(\hat{t}^*_{i,n}).$$

However, as in [70] it is convenient to use the method $\bar{X}^{cM*} = \{\bar{X}^{cM*}_{k_n}\}_{n\in\mathbb{N}}$ in order to investigate the error of $\bar{X}^{Lin-M*} = \{\bar{X}^{Lin-M*}_{k_n}\}_{n\in\mathbb{N}}$. We show in the sequel that they behave asymptotically in the same way. Moreover, for a fixed discretization $\Delta^*_{k_n}$ the method $\bar{X}^{Lin-M*}_{k_n}$ does not evaluate Λ and it is implementable. If $b \neq 0, c \neq 0$ then the both methods $\bar{X}^{cM*}_{k_n}$ and $\bar{X}^{Lin-M*}_{k_n}$ use $2k_n$ values of the processes N and W at the same time points.

The Theorem 2.4 states the asymptotic performance of the methods $\bar{X}^{cM*}$ and $\bar{X}^{Lin-M*}$. The error is expressed as a function of the number k_n of evaluations of the processes W and N.

Theorem 2.4. *Let us assume that the functions a, b, c and λ satisfy the assumptions (A) – (E) and let $\bar{X}^* \in \{\bar{X}^{cM*}, \bar{X}^{Lin-M*}\}$.*

(i) We have that

$$\lim_{n\to+\infty} \frac{k_n}{n} = \frac{1}{T}\int_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t. \tag{2.9}$$

(ii) If $b \not\equiv 0$ and $c \not\equiv 0$ then

$$\lim_{n\to+\infty} (2k_n)^{1/2} \cdot e_{k_n}(\bar{X}^*) = \sqrt{2} \cdot C^{\text{noneq}}, \tag{2.10}$$

else

$$\lim_{n\to+\infty} k_n^{1/2} \cdot e_{k_n}(\bar{X}^*) = C^{\text{noneq}}. \tag{2.11}$$

Proof. First note that for all $n \in \mathbb{N}$

$$\tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n}) = \bar{X}^{M}(\hat{t}^*_{i,n}), \quad i = 0, 1, \ldots, k_n,$$

and

$$\mathcal{Y}^M(\hat{t}^*_{i,n}) = \mathcal{Y}^M(\hat{t}_{i,n}) = |b(U^*_{i,n})|^2 + \lambda(\hat{t}^*_{i,n}) \cdot |c(U^*_{i,n})|^2, \quad i = 0, 1, \ldots, k_n - 1, \tag{2.12}$$

where $U^*_{i,n} := (\hat{t}^*_{i,n}, \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n}))$. Let us define

$$S_{j,n} := \sum_{i=0}^{k_n-1} \max\Big\{ \Big(\mathbb{E}(\mathcal{Y}^M(\hat{t}_{i,n}))\Big)^{j/2}, \varepsilon_n^j \Big\} \cdot (\hat{t}_{i+1,n} - \hat{t}_{i,n})^j, \tag{2.13}$$

$$S^*_{j,n} := \sum_{i=0}^{k_n-1} \max\Big\{ \Big(\mathbb{E}(\mathcal{Y}^M(\hat{t}^*_{i,n}))\Big)^{j/2}, \varepsilon_n^j \Big\} \cdot (\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n})^j, \tag{2.14}$$

for $j \in \{1, 2\}$, $n \in \mathbb{N}$.

Firstly we prove (2.9). By definition of step in algorithm given by (2.4) we have that

$$T \leqslant \sum_{i=0}^{k_n-1} (\hat{t}_{i+1,n} - \hat{t}_{i,n}) = \frac{T}{n} \sum_{i=0}^{k_n-1} \frac{1}{\max\Big\{ \Big(\mathbb{E}(\mathcal{Y}^M(\hat{t}_{i,n}))\Big)^{1/2}, \varepsilon_n \Big\}} \leqslant T \cdot (n \cdot \varepsilon_n)^{-1} \cdot k_n,$$

which gives

$$k_n \geqslant n \cdot \varepsilon_n,$$

for all $n \in \mathbb{N}$. Hence, from (2.3)

$$\lim_{n \to +\infty} k_n = +\infty.$$

Since for all $n \in \mathbb{N}$

$$\{\hat{t}_{0,n}, \hat{t}_{1,n}, \ldots, \hat{t}_{k_n-1,n}\} \subset \hat{\Delta}_{k_n} \cap \hat{\Delta}^*_{k_n},$$

by (2.4) and (2.12) we have that

$$|S_{j,n} - S^*_{j,n}| \leqslant 2 \max\Big\{ \Big(\mathbb{E}(\mathcal{Y}^M(\hat{t}_{k_n-1,n}))\Big)^{j/2}, \varepsilon_n^j \Big\} \cdot (\hat{t}_{k_n,n} - \hat{t}_{k_n-1,n})^j \leqslant 2(T/n)^j, \tag{2.15}$$

for $j \in \{1, 2\}$. Furthermore, we have that for all $n \in \mathbb{N}$

$$\max_{0 \le i \le k_n-1} (\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}) \leqslant \max_{0 \le i \le k_n-1} (\hat{t}_{i+1,n} - \hat{t}_{i,n}) \leqslant T \cdot (n \cdot \varepsilon_n)^{-1}, \tag{2.16}$$

and, from (2.3),

$$\lim_{n \to +\infty} \max_{0 \le i \le k_n-1} (\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}) = 0. \tag{2.17}$$

Let

$$\tilde{S}^*_{j,n} := \sum_{i=0}^{k_n-1} \Big(\mathbb{E}(\mathcal{Y}^M(\hat{t}^*_{i,n}))\Big)^{j/2} \cdot (\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n})^j, \quad j \in \{1, 2\}.$$

We can write that

$$\begin{aligned}\tilde{S}^*_{1,n} &= \sum_{i=0}^{k_n-1}\Big(\mathbb{E}(\mathcal{Y}(\hat{t}^*_{i,n}))\Big)^{1/2}\cdot(\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n})+\tilde{R}^*_{1,n},\\ \tilde{R}^*_{1,n} &:= \sum_{i=0}^{k_n-1}\Bigg(\Big(\mathbb{E}(\mathcal{Y}^M(\hat{t}^*_{i,n}))\Big)^{1/2}-\Big(\mathbb{E}(\mathcal{Y}(\hat{t}^*_{i,n}))\Big)^{1/2}\Bigg)\cdot(\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n}).\end{aligned}$$

By the Fact B.29 we have that $[0,T]\ni t\to\mathbb{E}(\mathcal{Y}(t))$ is continuous, and by (2.17) it follows that

$$\lim_{n\to+\infty}\sum_{i=0}^{k_n-1}\Big(\mathbb{E}(\mathcal{Y}(\hat{t}^*_{i,n}))\Big)^{1/2}\cdot(\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n})=\int_0^T\Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2}\mathrm{d}t.$$

Then by the fact that for all $x,y\in\mathbb{R}$, it holds that $\big||x|^{1/2}-|y|^{1/2}\big|\leqslant\big|x-y\big|^{1/2}$, we have that

$$\begin{aligned}|\tilde{R}^*_{1,n}| &= \left|\sum_{i=0}^{k_n-1}\Big(\big(\mathbb{E}(\mathcal{Y}^M(\hat{t}^*_{i,n}))\big)^{1/2}-\big(\mathbb{E}(\mathcal{Y}(\hat{t}^*_{i,n}))\big)^{1/2}\Big)\cdot(\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n})\right|\\ &\leqslant \sum_{i=0}^{k_n-1}\Big|\mathbb{E}(\mathcal{Y}^M(\hat{t}^*_{i,n}))-\mathbb{E}(\mathcal{Y}(\hat{t}^*_{i,n}))\Big|^{1/2}\cdot(\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n}).\end{aligned}$$

By Lemma B.32 and Theorem B.1 we have that for $i=0,1,\ldots,k_n-1$

$$\begin{aligned}\Big|\mathbb{E}(\mathcal{Y}^M(\hat{t}^*_{i,n}))-\mathbb{E}(\mathcal{Y}(\hat{t}^*_{i,n}))\Big| &\leqslant \Big|\mathbb{E}|b(U^*_{i,n})|^2-\mathbb{E}|b(\hat{t}^*_{i,n},X(\hat{t}^*_{i,n}))|^2\Big|\\ &\quad+\|\lambda\|_\infty\cdot\Big|\mathbb{E}|c(U^*_{i,n})|^2-\mathbb{E}|c(\hat{t}^*_{i,n},X(\hat{t}^*_{i,n}))|^2\Big|\\ &\leqslant C\cdot(1+\|\lambda\|_\infty)\cdot\sup_{t\in[0,T]}\big\|\tilde{X}^{M*}_{k_n}(t)-X(t)\big\|_{\mathfrak{L}^2(\Omega)}\\ &\quad\times\Big(1+\sup_{t\in[0,T]}\big\|\tilde{X}^{M*}_{k_n}(t)\big\|_{\mathfrak{L}^2(\Omega)}+\sup_{t\in[0,T]}\big\|X(t)\big\|_{\mathfrak{L}^2(\Omega)}\Big)\\ &\leqslant C_1\cdot\max_{0\le i\le k_n-1}(\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n})\leqslant C_1T\cdot(n\cdot\varepsilon_n)^{-1}.\end{aligned}$$

We obtain

$$|\tilde{R}^*_{1,n}|\leqslant C_2\cdot\varepsilon_n^{1/2}\cdot(n^{1/2}\cdot\varepsilon_n)^{-1},$$

and, by (2.3),

$$\lim_{n\to+\infty}|\tilde{R}^*_{1,n}|=0.$$

Hence,

$$\lim_{n\to+\infty}\tilde{S}^*_{1,n}=\int_0^T\Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2}\mathrm{d}t. \tag{2.18}$$

By (2.14) we have

$$\tilde{S}^*_{1,n} \leqslant S^*_{1,n} \leqslant \tilde{S}^*_{1,n} + T \cdot \varepsilon_n,$$

which, together with (2.3) and (2.18), implies

$$\lim_{n\to+\infty} S^*_{1,n} = \lim_{n\to+\infty} \tilde{S}^*_{1,n} = \int\limits_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t.$$

Moreover, by (2.15) we have that

$$\lim_{n\to+\infty} S_{1,n} = \lim_{n\to+\infty} S^*_{1,n} = \int\limits_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t. \tag{2.19}$$

Since

$$S_{j,n} = k_n \cdot \left(\frac{T}{n}\right)^j, \quad j \in \{1,2\},$$

by (2.19) we obtain

$$\lim_{n\to+\infty} k_n \cdot \frac{T}{n} = \lim_{n\to+\infty} S_{1,n} = \int\limits_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t < +\infty, \tag{2.20}$$

which gives (2.9). □

Now, we go to the proof of (2.10) and (2.11). By (2.20) we also have that

$$\lim_{n\to+\infty} k_n \cdot n^{-2} = 0. \tag{2.21}$$

Hence, from (2.15) with $j = 2$, (2.20) and (2.21) we obtain

$$\lim_{n\to+\infty} k_n \cdot S^*_{2,n} = \lim_{n\to+\infty} k_n \cdot S_{2,n} = \lim_{n\to+\infty} \Big(k_n \cdot \frac{T}{n}\Big)^2 = \left(\int\limits_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t\right)^2. \tag{2.22}$$

From (2.14) it follows that

$$k_n \cdot S^*_{2,n} - \frac{k_n}{n} \cdot \varepsilon_n \cdot T^2 \leqslant k_n \cdot S^*_{2,n} - k_n \cdot \varepsilon_n^2 \cdot \sum_{i=0}^{k_n-1} (\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n})^2 \leqslant k_n \cdot \tilde{S}^*_{2,n} \leqslant k_n \cdot S^*_{2,n}. \tag{2.23}$$

Hence, from (2.3), (2.20), (2.22) and (2.23) we obtain

$$\lim_{n\to+\infty} k_n \cdot \tilde{S}^*_{2,n} = \left(\int\limits_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t\right)^2. \tag{2.24}$$

By decomposition (B.61), estimation (B.67) and (2.16) we have that

$$\Big|\|\tilde{X}^{M*}_{k_n} - \bar{X}^{cM*}_{k_n}\|_{\mathfrak{L}^2(\Omega\times[0,T])} - \|\tilde{H}^{M*}_{k_n}\|_{\mathfrak{L}^2(\Omega\times[0,T])}\Big| \leqslant \|\tilde{R}^{M*}_{k_n}\|_{\mathfrak{L}^2(\Omega\times[0,T])} \leqslant C(n\cdot\varepsilon_n)^{-1}.$$

Let us define

$$\hat{Z}_n^*(t) := Z(t) - \mathbb{E}\big(Z(t) \mid \hat{\mathcal{N}}_{k_n}^*(Z)\big), \quad Z \in \{N, W\}.$$

Then we have that

$$\begin{aligned}
\|\tilde{H}_{k_n}^{M*}\|^2_{\mathfrak{L}^2(\Omega\times[0,T])} &= \mathbb{E}\left(\int_0^T |\tilde{H}_{k_n}^{M*}(t)|^2 \mathrm{d}t\right) = \sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}^*}^{\hat{t}_{i+1,n}^*} \mathbb{E}|\tilde{H}_{k_n}^{M*}(t)|^2 \mathrm{d}t \\
&= \sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}^*}^{\hat{t}_{i+1,n}^*} \mathbb{E}|b(U_{i,n}^*) \cdot \hat{W}_n^*(t) + c(U_{i,n}^*) \cdot \hat{N}_n^*(t)|^2 \mathrm{d}t \\
&= \sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}^*}^{\hat{t}_{i+1,n}^*} \Big(\mathbb{E}|b(U_{i,n}^*) \cdot \hat{W}_n^*(t)|^2 + \ \mathbb{E}|c(U_{i,n}^*) \cdot \hat{N}_n^*(t)|^2 \\
&\qquad + 2 \cdot \mathbb{E}|b(U_{i,n}^*) \cdot c(U_{i,n}^*) \cdot \hat{W}_n^*(t) \cdot \hat{N}_n^*(t)|\Big) \mathrm{d}t.
\end{aligned}$$

We have that $b(U_{i,n}^*), c(U_{i,n}^*)$ are $\mathcal{F}_{\hat{t}_{i,n}^*}$-measurable. For all $t \in [\hat{t}_{i,n}^*, \hat{t}_{i+1,n}^*]$ the process $\hat{W}_n^*(t), \hat{N}_n^*(t)$ are independent of $\mathcal{F}_{\hat{t}_{i,n}^*}$. This imply that

$$\mathbb{E}|b(U_{i,n}^*) \cdot \hat{W}_n^*(t)|^2 = \mathbb{E}|b(U_{i,n}^*)|^2 \cdot \mathbb{E}|\hat{W}_n^*(t)|^2, \tag{2.25}$$

$$\mathbb{E}|c(U_{i,n}^*) \cdot \hat{N}_n^*(t)|^2 = \mathbb{E}|c(U_{i,n}^*)|^2 \cdot \mathbb{E}|\hat{N}_n^*(t)|^2. \tag{2.26}$$

By the fact that $b(U_{i,n}^*) \cdot c(U_{i,n}^*)$ are $\mathcal{F}_{\hat{t}_{i,n}^*}$-measurable and $\mathbb{E}|b(U_{i,n}^*) \cdot c(U_{i,n}^*)| < +\infty$ (by Hölder inequality and Theorem B.1) together with the fact that for all $t \in [\hat{t}_{i,n}^*, \hat{t}_{i+1,n}^*]$ $\hat{W}_n^*(t) \cdot \hat{N}_n^*(t)$ are independent of $\mathcal{F}_{\hat{t}_{i,n}^*}$ and $\hat{W}_n^*(t), \hat{N}_n^*(t)$ are independent we have that

$$\mathbb{E}|b(U_{i,n}^*) \cdot c(U_{i,n}^*) \cdot \hat{W}_n^*(t) \cdot \hat{N}_n^*(t)| = \mathbb{E}|b(U_{i,n}^*) \cdot c(U_{i,n}^*)| \cdot \mathbb{E}|\hat{W}_n^*(t)| \cdot \mathbb{E}|\hat{N}_n^*(t)| = 0. \tag{2.27}$$

Finally, by (2.25), (2.26) and (2.27) we obtain

$$\|\tilde{H}_{k_n}^{M*}\|^2_{\mathfrak{L}^2(\Omega\times[0,T])} = \left(\sum_{i=0}^{k_n-1} \int_{\hat{t}_{i,n}^*}^{\hat{t}_{i+1,n}^*} \mathbb{E}|b(U_{i,n}^*)|^2 \cdot \mathbb{E}|\hat{W}_n^*(t)|^2 + \mathbb{E}|c(U_{i,n}^*)|^2 \cdot \mathbb{E}|\hat{N}_n^*(t)|^2 \mathrm{d}t\right)^{1/2}.$$

By Lemma B.21 we can calculate that

$$\int_{\hat{t}_{i,n}^*}^{\hat{t}_{i+1,n}^*} \mathbb{E}|\hat{W}_n^*(t)|^2 \mathrm{d}t = \int_{\hat{t}_{i,n}^*}^{\hat{t}_{i+1,n}^*} \frac{(t_{i+1,n}^* - t)(t - t_{i,n}^*)}{(t_{i+1,n}^* - t_{i,n}^*)} \mathrm{d}t = \frac{1}{6}(\hat{t}_{i+1,n}^* - \hat{t}_{i,n}^*)^2. \tag{2.28}$$

Then for $i = 0, 1, \ldots, k_n - 1$ and $t \in (\hat{t}^*_{i,n}, \hat{t}^*_{i+1,n})$ we define

$$H_{i,n}(t) = \frac{\Lambda(t, \hat{t}^*_{i,n}) \cdot \Lambda(\hat{t}^*_{i+1,n}, t)}{(\hat{t}^*_{i+1,n} - t)(t - \hat{t}^*_{i,n})}.$$

Of course $H_{i,n} \in C((\hat{t}^*_{i,n}, \hat{t}^*_{i+1,n}))$ and it can be continuously extended to $[\hat{t}^*_{i,n}, \hat{t}^*_{i+1,n}]$, since

$$H(\hat{t}^*_{i,n}+) = \lambda(\hat{t}^*_{i,n}) \cdot \Lambda(\hat{t}^*_{i+1,n}\hat{t}^*_{i,n})/(\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n})$$

and

$$H(\hat{t}^*_{i+1,n}-) = \lambda(\hat{t}^*_{i+1,n}) \cdot \Lambda(\hat{t}^*_{i+1,n}, \hat{t}^*_{i,n})/(\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n})$$

are finite. Therefore, by Lemma B.22 and from the mean value theorem we have that

$$\begin{aligned}
\int\limits_{\hat{t}^*_{i,n}}^{\hat{t}^*_{i+1,n}} \mathbb{E}|\hat{N}^*_n(t)|^2 \mathrm{d}t &= \Lambda(\hat{t}^*_{i+1,n}, \hat{t}^*_{i,n})^{-1} \cdot \int\limits_{\hat{t}^*_{i,n}}^{\hat{t}^*_{i+1,n}} H_{i,n}(t) \cdot (\hat{t}^*_{i+1,n} - t) \cdot (t - \hat{t}^*_{i,n}) \mathrm{d}t \\
&= \Lambda(\hat{t}^*_{i+1,n}, \hat{t}^*_{i,n})^{-1} \cdot H_{i,n}(\hat{d}^*_{i,n}) \cdot \int\limits_{\hat{t}^*_{i,n}}^{\hat{t}^*_{i+1,n}} (\hat{t}^*_{i+1,n} - t) \cdot (t - \hat{t}^*_{i,n}) \mathrm{d}t \\
&= \frac{1}{6} \frac{\lambda(\hat{\alpha}^*_{i,n})\lambda(\hat{\beta}^*_{i,n})}{\lambda(\hat{\gamma}^*_{i,n})} \cdot (\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n})^2, \qquad (2.29)
\end{aligned}$$

for some $\hat{d}^*_{i,n}, \hat{\alpha}^*_{i,n}, \hat{\beta}^*_{i,n}, \hat{\gamma}^*_{i,n} \in [\hat{t}^*_{i,n}, \hat{t}^*_{i+1,n}]$, $i = 0, 1, \ldots, k_n - 1$. Now by (2.28) and (2.29) we define

$$\hat{S}^*_{2,n} = \sum_{i=0}^{k_n - 1} \left(\mathbb{E}|b(U^*_{i,n})|^2 + \mathbb{E}|c(U^*_{i,n})|^2 \cdot \frac{\lambda(\hat{\alpha}^*_{i,n})\lambda(\hat{\beta}^*_{i,n})}{\lambda(\hat{\gamma}^*_{i,n})} \right) \cdot (\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n})^2,$$

and of course we have that

$$k_n^{1/2} \cdot \left\| \tilde{H}^{M*}_{k_n} \right\|_{\mathfrak{L}^2(\Omega \times [0,T])} = \left(k_n \cdot \sum_{i=0}^{k_n - 1} \int\limits_{\hat{t}^*_{i,n}}^{\hat{t}^*_{i+1,n}} \mathbb{E}|\tilde{H}^{M*}_{k_n}(t)|^2 \mathrm{d}t \right)^{1/2} = \left(\frac{k_n}{6} \cdot \hat{S}^*_{2,n} \right)^{1/2}. \qquad (2.30)$$

Furthermore,

$$\begin{aligned}
|k_n \cdot \hat{S}^*_{2,n} - k_n \cdot \tilde{S}^*_{2,n}| &\leqslant k_n \cdot \sum_{i=0}^{k_n - 1} \mathbb{E}|c(U^*_{i,n})|^2 \cdot \left| \frac{\lambda(\hat{\alpha}^*_{i,n})\lambda(\hat{\beta}^*_{i,n})}{\lambda(\hat{\gamma}^*_{i,n})} - \lambda(\hat{t}^*_{i,n}) \right| \cdot (\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n})^2 \\
&\leqslant C \cdot \frac{k_n}{n} \cdot \varepsilon_n^{-1} \cdot \bar{\omega}(\lambda, T \cdot (n \cdot \varepsilon_n)^{-1}). \qquad (2.31)
\end{aligned}$$

Hence, from (2.3), (2.20), (2.24) and (2.31) we obtain

$$\lim_{n \to +\infty} k_n \cdot \hat{S}^*_{2,n} = \lim_{n \to +\infty} k_n \cdot \tilde{S}^*_{2,n} = \left(\int\limits_0^T \Big(\mathbb{E}(\mathcal{Y}(t)) \Big)^{1/2} \mathrm{d}t \right)^2. \qquad (2.32)$$

Therefore, by (2.30) and (2.32) we obtain

$$\lim_{n\to+\infty} k_n^{1/2}\cdot \big\|\tilde{H}^{M*}_{k_n}\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} = C^{\text{noneq}}.$$

Since from (2.3) and (2.20) it follows that

$$\lim_{n\to+\infty} k_n^{1/2}\cdot (n\cdot\varepsilon_n)^{-1} = \lim_{n\to+\infty}\left(\frac{k_n}{n}\right)^{1/2}\cdot (n^{1/2}\cdot\varepsilon_n)^{-1} = 0, \tag{2.33}$$

and we get

$$\lim_{n\to+\infty} k_n^{1/2}\cdot \big\|\tilde{X}^{M*}_{k_n} - \bar{X}^{cM*}_{k_n}\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} = \lim_{n\to+\infty} k_n^{1/2}\cdot \big\|\tilde{H}^{M*}_{k_n}\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} = C^{\text{noneq}}. \tag{2.34}$$

Next, from Theorem B.1

$$\begin{aligned}\Big|e_{k_n}(\bar{X}^{cM*}) - \big\|\tilde{X}^{M*}_{k_n} - \bar{X}^{cM*}_{k_n}\big\|_{\mathfrak{L}^2(\Omega\times[0,T])}\Big| \leqslant e_{k_n}(\tilde{X}^{M*}) &\leqslant C\cdot \max_{0\le i\le k_n-1}(\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}) \\ &\leqslant CT(n\cdot\varepsilon_n)^{-1}. \end{aligned} \tag{2.35}$$

Hence, from (2.33), (2.34) and (2.35) we have that

$$\lim_{n\to+\infty} k_n^{1/2}\cdot e_{k_n}(\bar{X}^{cM*}) = \lim_{n\to+\infty} k_n^{1/2}\cdot \big\|\tilde{X}^{M*}_{k_n} - \bar{X}^{cM*}_{k_n}\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} = C^{\text{noneq}},$$

which ends the proof in the case when $\bar{X}^* = \bar{X}^{cM*}$. □

Now we analyze the error of $\bar{X}^* = \bar{X}^{Lin-M*}_{k_n}$. Note that

$$\begin{aligned}\bar{R}^{M*}_{k_n}(t) &:= \bar{X}^{cM*}_{k_n}(t) - \bar{X}^{Lin-M*}_{k_n}(t) \\ &= c(U^*_{i,n})\cdot\Delta N^*_{i,n}\cdot\left(\frac{\Lambda(t,\hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n},\hat{t}^*_{i,n})} - \frac{t-\hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n}}\right) \\ &\quad + L_1 b(U^*_{i,n})\cdot I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(W,W)\cdot\frac{(t-\hat{t}^*_{i,n})\cdot(t-\hat{t}^*_{i+1,n})}{(\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n})^2} \\ &\quad + L_{-1} b(U^*_{i,n})\cdot\Delta N^*_{i,n}\cdot\Delta W^*_{i,n}\cdot\frac{\hat{t}^*_{i,n}-t}{\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n}}\cdot\frac{\Lambda(\hat{t}^*_{i+1,n},t)}{\Lambda(\hat{t}^*_{i+1,n},\hat{t}^*_{i,n})} \\ &\quad + L_{-1} c(U^*_{i,n})\cdot I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(N,N)\cdot\left(\left(\frac{\Lambda(t,\hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n},\hat{t}^*_{i,n})}\right)^2 - \frac{t-\hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n}}\right),\end{aligned}$$

for $t\in[\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}]$, $i=0,1,\ldots,k_n-1$. By the fact that $c(U^*_{i,n})$, $L_1 b(U^*_{i,n})$, $L_{-1}b(U^*_{i,n})$, $L_{-1}c(U^*_{i,n})$ are $\mathcal{F}_{\hat{t}^*_{i,n}}$-measurable and $\Delta N^*_{i,n}, \Delta W^*_{i,n}, I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(W,W), I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(N,N)$

are independent of $\mathcal{F}_{\hat{t}^*_{i,n}}$ and $\Delta W^*_{i,n}, \Delta N^*_{i,n}$ are also independent. Together with Lemma B.28 we have that

$$\begin{aligned}\mathbb{E}|\bar{R}^{M*}_{k_n}(t)|^2 \leqslant\ & \mathbb{E}|c(U^*_{i,n})|^2 \cdot \mathbb{E}|\Delta N^*_{i,n}|^2 \cdot \left| \frac{\Lambda(t,\hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n},\hat{t}^*_{i,n})} - \frac{t-\hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n}} \right|^2 \\ &+ \mathbb{E}|L_1 b(U^*_{i,n})|^2 \cdot \mathbb{E}|I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(W,W)|^2 \cdot \left| \frac{(t-\hat{t}^*_{i,n})\cdot(t-\hat{t}^*_{i+1,n})}{(\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n})^2} \right|^2 \\ &+ \mathbb{E}|L_{-1} b(U^*_{i,n})|^2 \cdot \mathbb{E}|\Delta N^*_{i,n}|^2 \cdot \mathbb{E}|\Delta W^*_{i,n}|^2 \cdot \left| \frac{\hat{t}^*_{i,n}-t}{\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n}} \cdot \frac{\Lambda(\hat{t}^*_{i+1,n},t)}{\Lambda(\hat{t}^*_{i+1,n},\hat{t}^*_{i,n})} \right|^2 \\ &+ \mathbb{E}|L_{-1} c(U^*_{i,n})|^2 \cdot \mathbb{E}|I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(N,N)|^2 \cdot \left| \left(\frac{\Lambda(t,\hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n},\hat{t}^*_{i,n})} \right)^2 - \frac{t-\hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n}} \right|^2,\end{aligned}$$

In addition, by (2.16) and Fact B.31 we have that

$$\left| \frac{\Lambda(t,\hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n},\hat{t}^*_{i,n})} - \frac{t-\hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n}-\hat{t}^*_{i,n}} \right| \leqslant C_1 \cdot \sup_{t,s\in[\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}]} |\lambda(t)-\lambda(s)| \leqslant C_1 \cdot \bar{\omega}(\lambda, T\cdot(n\cdot\varepsilon_n)^{-1}), \tag{2.36}$$

for $t \in [\hat{t}^*_{i,n}, \hat{t}^*_{i+1,n}]$, $i = 0,1,\ldots,k_n-1$.

By the Lemma B.28, Lemma B.2 and (2.36) we obtain the following estimation.

$$\begin{aligned}\mathbb{E}|\bar{R}^{M*}_{k_n}(t)|^2 \leqslant\ & C_1 \cdot \big(\bar{\omega}(\lambda, T/(n\cdot\varepsilon_n))\big)^2 \cdot \big(1+(n\cdot\varepsilon_n)^{-1}\big) \cdot (n\cdot\varepsilon_n)^{-1} \\ &+ C_2 \cdot (n\cdot\varepsilon_n)^{-2} \cdot \big(1+(n\cdot\varepsilon_n)^{-1}\big).\end{aligned} \tag{2.37}$$

Since, from (2.37), (2.21) and (2.33) we have

$$\begin{aligned}\left| k_n^{1/2}\cdot e_{k_n}(\bar{X}^{Lin-M*}) - k_n^{1/2}\cdot e_{k_n}(\bar{X}^{cM*}) \right| &\leqslant k_n^{1/2} \cdot \big\| \bar{R}^{M*}_{k_n} \big\|_{\mathfrak{L}^2(\Omega\times[0,T])} \\ &\leqslant C_1 \cdot \varepsilon_n^{-1} \cdot \bar{\omega}(\lambda, T/(n\cdot\varepsilon_n)) \cdot \big(1+(n\cdot\varepsilon_n)^{-1}\big)^{1/2} \cdot (k_n/n)^{1/2} \cdot \varepsilon_n^{1/2} \\ &\quad + C_2 \cdot k_n^{1/2}/(n\cdot\varepsilon_n) \cdot \big(1+(n\cdot\varepsilon_n)^{-1}\big)^{1/2},\end{aligned}$$

we get (2.10) and (2.11) for $\bar{X}^* = \bar{X}^{Lin-M*}$. This ends the proof. ■

2.2.2. Derivative-free version of the path-independent adaptive step-size control

In this section we present the derivative-free version of the Milstein scheme, which can be used for the path-independent adaptive step-size control and achieves asymptotically the same rate of convergence as $\bar{X}^{Lin-M*}$.

STEP 0 Take an arbitrary strictly positive sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that

$$\lim_{n\to+\infty}\varepsilon_n=\lim_{n\to+\infty}(n^{1/2}\cdot\varepsilon_n)^{-1}=\lim_{n\to+\infty}\varepsilon_n^{-1}\cdot\bar{\omega}(\lambda,T\cdot(n\cdot\varepsilon_n)^{-1})=0,$$

where $\bar{\omega}$ is the modulus of continuity for λ (see Remark 2.2).

STEP 1 Take any $n\in\mathbb{N}$ and let $\hat{t}_{0,n}=0$, $\bar{X}^M(\hat{t}_{0,n})=x_0$. Set $i:=0$.

STEP 2 If $\hat{t}_{i,n}\in[0,T)$ and $\bar{X}^{df-M}(\hat{t}_{i,n})$ are given then compute

$$\hat{t}_{i+1,n}=\hat{t}_{i,n}+\frac{T}{n\cdot\max\Big\{\varepsilon_n,\Big(\mathbb{E}(\mathcal{Y}^{df-M}(\hat{t}_{i,n}))\Big)^{1/2}\Big\}},\tag{2.38}$$

where

$$\mathcal{Y}^{df-M}(\hat{t}_{i,n})=|b(\hat{t}_{i,n},\bar{X}^{df-M}(\hat{t}_{i,n}))|^2+\lambda(\hat{t}_{i,n})\cdot|c(\hat{t}_{i,n},\bar{X}^{df-M}(\hat{t}_{i,n}))|^2.$$

If $\hat{t}_{i+1,n}<T$, $\hat{h}_{i,n}=\hat{t}_{i+1,n}-\hat{t}_{i,n}$ then compute

$$\begin{aligned}\bar{X}^{df-M}(\hat{t}_{i+1,n}) &= \bar{X}^{df-M}(\hat{t}_{i,n})+a(\hat{t}_{i,n},\bar{X}^{df-M}(\hat{t}_{i,n}))\cdot(\hat{t}_{i+1,n}-\hat{t}_{i,n})\\ &+ b(\hat{t}_{i,n},\bar{X}^{df-M}(\hat{t}_{i,n}))\cdot(W(\hat{t}_{i+1,n})-W(\hat{t}_{i,n}))\\ &+ c(\hat{t}_{i,n},\bar{X}^{df-M}(\hat{t}_{i,n}))\cdot(N(\hat{t}_{i+1,n})-N(\hat{t}_{i,n}))\\ &+ \mathcal{L}_{1,\hat{h}_i}b(\hat{t}_{i,n},\bar{X}^{df-M}(\hat{t}_{i,n}))\cdot I_{\hat{t}_{i,n},\hat{t}_{i+1,n}}(W,W)\\ &+ L_{-1}c(\hat{t}_{i,n},\bar{X}^{df-M}(\hat{t}_{i,n}))\cdot I_{\hat{t}_{i,n},\hat{t}_{i+1,n}}(N,N)\\ &+ L_{-1}b(\hat{t}_{i,n},\bar{X}^{df-M}(\hat{t}_{i,n}))\cdot\big(I_{\hat{t}_{i,n},\hat{t}_{i+1,n}}(N,W)+I_{\hat{t}_{i,n},\hat{t}_{i+1,n}}(W,N)\big),\end{aligned}\tag{2.39}$$

take $i:=i+1$ and GOTO **STEP 2**.

Else compute $\bar{X}^{df-M}(T)$ by (2.39) with $\hat{t}_{i+1,n}$ replaced by T.

STOP

The stopping criterion is defined in the same way as in previous algorithm and

$$k_n=\min\big\{i\in\mathbb{N}\mid\hat{t}_{i,n}\geqslant T\big\},\quad n\in\mathbb{N},$$

which is the total number of computed discretization points greater than $\hat{t}_{0,n}=0$. The end point T is attainable, since we have for all $n\in\mathbb{N}$ that

$$k_n\leqslant\lceil n(\varepsilon_n+\hat{C})\rceil,\tag{2.40}$$

for some $\hat{C}<+\infty$, where existence of $\hat{C}$ follows from the Fact 2.5.

Fact 2.5. *Let $n \in \mathbb{N}$ and let us assume that there exists $M_n \in \mathbb{N}$ such that $\hat{t}_{j,n} \in [0,T)$ for all $j = 0,1,\ldots,M_n - 1$. Then*

$$\hat{t}_{M_n,n} \geqslant M_n \cdot \frac{T}{n(\varepsilon_n + \hat{C})},$$

where $\hat{C} = K_1\big(1+\bar{C}\big) \cdot \big(1 + \|\lambda\|_\infty^{1/2}\big)$ and $\bar{C}$ is the constant from Theorem B.13.

The proof of Fact 2.5 goes analogously as proof of Fact 2.3, so we skip it. Hence, if for a given $n \in \mathbb{N}$ we have that $M_n = \lceil n(\varepsilon_n + \hat{C})\rceil$ then by Fact 2.5 we get $\hat{t}_{M_n,n} \geqslant T$. This implies (2.40) and the fact that algorithm stops in a finite number of steps.

Again, we obtain two sequences of discretizations $\hat{\Delta} = \big\{\Delta_{k_n}\big\}_{n\in\mathbb{N}}$ and $\hat{\Delta}^* = \big\{\Delta^*_{k_n}\big\}_{n\in\mathbb{N}}$ and we define the conditional derivative-free Milstein method $\bar{X}^{df-cM*} = \big\{\bar{X}^{df-cM*}_{k_n}\big\}_{n\in\mathbb{N}}$ as

$$\bar{X}^{df-cM*}_{k_n}(t) = \mathbb{E}\big(\tilde{X}^{df-M*}_{k_n}(t) \mid \mathcal{N}^*_{k_n}(N,W)\big), \quad t \in [0,T],$$

where $\big\{\tilde{X}^{df-M*}_{k_n}\big\}_{n\in\mathbb{N}}$ is the sequence of time-continuous derivative-free Milstein approximations (B.70) – (B.71) based on $\big\{\hat{\Delta}^*_{k_n}\big\}_{n\in\mathbb{N}}$ and vector of information $\mathcal{N}^*_{k_n}(N,W)$ is as in (2.8). Followed by Lemma B.24, Lemma B.26, Lemma B.27 (see also [70]) it follows that

$$\begin{aligned}
\bar{X}^{df-cM*}_{k_n}(t) = \tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i,n}) \quad &+ \quad a(\hat{t}^*_{i,n}, \tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i,n})) \cdot (t - \hat{t}^*_{i,n}) \\
&+ \quad b(\hat{t}^*_{i,n}, \tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i,n})) \cdot \Delta W^*_{i,n} \cdot \frac{t - \hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}} \\
&+ \quad c(\hat{t}^*_{i,n}, \tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i,n})) \cdot \Delta N^*_{i,n} \cdot \frac{\Lambda(t, \hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n}, \hat{t}^*_{i,n})} \\
&+ \quad \mathcal{L}_{1,\hat{h}^*_i} b(\hat{t}^*_{i,n}, \tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i,n})) \cdot I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(W,W) \\
&\qquad \times \Big(\frac{t - \hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}}\Big)^2 \\
&+ \quad L_{-1} b(\hat{t}^*_{i,n}, \tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i,n})) \cdot \Delta N^*_{i,n} \cdot \Delta W^*_{i,n} \\
&\qquad \times \frac{\Lambda(t, \hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n}, \hat{t}^*_{i,n})} \cdot \frac{t - \hat{t}^*_{i,n}}{\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}} \\
&+ \quad L_{-1} c(\hat{t}^*_{i,n}, \tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i,n})) \cdot I_{\hat{t}^*_{i,n},\hat{t}^*_{i+1,n}}(N,N) \\
&\qquad \times \Big(\frac{\Lambda(t, \hat{t}^*_{i,n})}{\Lambda(\hat{t}^*_{i+1,n}, \hat{t}^*_{i,n})}\Big)^2,
\end{aligned}$$

for $t \in [\hat{t}^*_{i,n}, \hat{t}^*_{i+1,n}]$, $i = 0,1,\ldots,k_n - 1$, $\hat{h}^*_i = \hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}$ and $\bar{X}^{df-cM*}_{k_n}(0) = x_0$,

$$\Delta W^*_{i,n} = W(\hat{t}^*_{i+1,n}) - W(\hat{t}^*_{i,n}),$$

$$\Delta N^*_{i,n} = N(\hat{t}^*_{i+1,n}) - N(\hat{t}^*_{i,n}).$$

We also take the piecewise linear interpolation $\bar{X}^{Lin-M*}_{k_n}$ of the derivative-free Milstein steps

$$\bar{X}^{df-Lin-M*}_{k_n}(t) = \frac{\tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i,n})(\hat{t}^*_{i+1,n} - t) + \tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i+1,n})(t - \hat{t}^*_{i,n})}{\hat{t}^*_{i+1,n} - \hat{t}^*_{i,n}},$$

for $t \in [\hat{t}^*_{i,n}, \hat{t}^*_{i+1,n}]$, $i = 0, 1, \ldots, k_n - 1$.

Due to Theorem B.13, (B.95) and the decomposition (B.90) we can repeat argumentation from the proof of Fact 2.3 and Theorem 2.4 in order to obtain the same asymptotic result for $\bar{X}^{df-cM*}$ and $\bar{X}^{df-Lin-M*}$.

Theorem 2.6. *Let us assume that the functions a, b, c and λ satisfy the assumptions (A) – (E) and let $\bar{X}^* \in \{\bar{X}^{df-cM*}, \bar{X}^{df-Lin-M*}\}$.*

(i) We have that

$$\lim_{n\to+\infty} \frac{k_n}{n} = \frac{1}{T}\int_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t. \tag{2.41}$$

(ii) If $b \not\equiv 0$ and $c \not\equiv 0$ then

$$\lim_{n\to+\infty} (2k_n)^{1/2} \cdot e_{k_n}(\bar{X}^*) = \sqrt{2} \cdot C^{\text{noneq}}, \tag{2.42}$$

else

$$\lim_{n\to+\infty} k_n^{1/2} \cdot e_{k_n}(\bar{X}^*) = C^{\text{noneq}}. \tag{2.43}$$

Proof. The proof of (2.41) goes almost exactly in the same way as proof of (2.9). The main change in the proof goes as follows. Change:

$\bar{X}^{cM}_{k_n}$	into	$\bar{X}^{df-cM}_{k_n}$,
$\tilde{X}^{M*}_{k_n}$	into	$\tilde{X}^{df-M*}_{k_n}$,
$\mathcal{Y}^{M}(\hat{t}^*_{i,n})$	into	$\mathcal{Y}^{df-M}(\hat{t}^*_{i,n})$,
$U^*_{i,n} := (\hat{t}^*_{i,n}, \tilde{X}^{M*}_{k_n}(\hat{t}^*_{i,n}))$	into	$U^{df*}_{i,n} := (\hat{t}^*_{i,n}, \tilde{X}^{df-M*}_{k_n}(\hat{t}^*_{i,n}))$.

Then we use definition of step given by (2.38) and Theorem B.13 instead of (2.4) and Theorem B.1. Hence, finally we obtain

$$\lim_{n\to+\infty} k_n \cdot \frac{T}{n} = \int_0^T \Big(\mathbb{E}(\mathcal{Y}(t))\Big)^{1/2} \mathrm{d}t < +\infty,$$

which gives (2.41). Using the same argumentation as in proofs of (2.10) and (2.11) we have that

$$\lim_{n\to+\infty} k_n^{1/2}\cdot e_{k_n}(\bar{X}^{df-cM*}) = \lim_{n\to+\infty} k_n^{1/2}\cdot \left\|\tilde{X}_{k_n}^{df-M*} - \bar{X}_{k_n}^{df-cM*}\right\|_{\mathfrak{L}^2(\Omega\times[0,T])} = C^{\mathrm{noneq}},$$

which ends the proof in the case when $\bar{X}^* = \bar{X}^{df-cM*}$. □

Now we analyze the error of $\bar{X}^* = \bar{X}_{k_n}^{df-Lin-M*}$. Note that in this case

$$\begin{aligned}\bar{R}_{k_n}^{df-M*}(t) &:= \bar{X}_{k_n}^{df-cM*}(t) - \bar{X}_{k_n}^{df-Lin-M*}(t)\\ &= c(U_{i,n}^{df*})\cdot\Delta N_{i,n}^*\cdot\left(\frac{\Lambda(t,\hat{t}_{i,n}^*)}{\Lambda(\hat{t}_{i+1,n}^*,\hat{t}_{i,n}^*)} - \frac{t-\hat{t}_{i,n}^*}{\hat{t}_{i+1,n}^*-\hat{t}_{i,n}^*}\right)\\ &\quad + \mathcal{L}_{1,\hat{h}_{i,n}}b(U_{i,n}^{df*})\cdot I_{\hat{t}_{i,n}^*,\hat{t}_{i+1,n}^*}(W,W)\cdot\frac{(t-\hat{t}_{i,n}^*)\cdot(t-\hat{t}_{i+1,n}^*)}{(\hat{t}_{i+1,n}^*-\hat{t}_{i,n}^*)^2}\\ &\quad + L_{-1}b(U_{i,n}^{df*})\cdot\Delta N_{i,n}^*\cdot\Delta W_{i,n}^*\cdot\frac{\hat{t}_{i,n}^*-t}{\hat{t}_{i+1,n}^*-\hat{t}_{i,n}^*}\cdot\frac{\Lambda(\hat{t}_{i+1,n}^*,t)}{\Lambda(\hat{t}_{i+1,n}^*,\hat{t}_{i,n}^*)}\\ &\quad + L_{-1}c(U_{i,n}^{df*})\cdot I_{\hat{t}_{i,n}^*,\hat{t}_{i+1,n}^*}(N,N)\cdot\left(\left(\frac{\Lambda(t,\hat{t}_{i,n}^*)}{\Lambda(\hat{t}_{i+1,n}^*,\hat{t}_{i,n}^*)}\right)^2 - \frac{t-\hat{t}_{i,n}^*}{\hat{t}_{i+1,n}^*-\hat{t}_{i,n}^*}\right),\end{aligned}$$

for $t\in[\hat{t}_{i,n}^*,\hat{t}_{i+1,n}^*]$, $i=0,1,\ldots,k_n-1$. By the fact that $c(U_{i,n}^{df*})$, $\mathcal{L}_{1,\hat{h}_{i,n}}b(U_{i,n}^{df*})$, $L_{-1}b(U_{i,n}^{df*})$, $L_{-1}c(U_{i,n}^{df*})$ are $\mathcal{F}_{\hat{t}_{i,n}^*}$-measurable and $\Delta N_{i,n}^*$, $\Delta W_{i,n}^*$, $I_{\hat{t}_{i,n}^*,\hat{t}_{i+1,n}^*}(W,W)$, $I_{\hat{t}_{i,n}^*,\hat{t}_{i+1,n}^*}(N,N)$ are independent of $\mathcal{F}_{\hat{t}_{i,n}^*}$ and $\Delta W_{i,n}^*,\Delta N_{i,n}^*$ are also independent. Using the same logic as in the case of (2.37) the only change is that we use $\mathcal{L}_{1,\hat{h}_{i,n}}b(U_{i,n}^{df*})$ instead of $L_1b(U_{i,n}^{df*})$, which satisfies exactly the same assumptions, by Theorem B.13 and (B.77) we have that

$$\begin{aligned}\mathbb{E}|\bar{R}_{k_n}^{df-M*}(t)|^2 &\leqslant C_1\cdot(\bar{\omega}(\lambda,T/(n\cdot\varepsilon_n)))^2\cdot(1+(n\cdot\varepsilon_n)^{-1})\cdot(n\cdot\varepsilon_n)^{-1}\\ &\quad + C_2\cdot(n\cdot\varepsilon_n)^{-2}\cdot(1+(n\cdot\varepsilon_n)^{-1}).\end{aligned}\tag{2.44}$$

Since, from (2.44), $\lim\limits_{n\to+\infty} k_n\cdot n^{-2}=0$ and

$$\lim_{n\to+\infty} k_n^{1/2}\cdot(n\cdot\varepsilon_n)^{-1} = \lim_{n\to+\infty}\left(\frac{k_n}{n}\right)^{1/2}\cdot(n^{1/2}\cdot\varepsilon_n)^{-1}=0,$$

then

$$\begin{aligned}\left|k_n^{1/2}\cdot e_{k_n}(\bar{X}^{df-Lin-M*}) - k_n^{1/2}\cdot e_{k_n}(\bar{X}^{df-cM*})\right| &\leqslant k_n^{1/2}\cdot\left\|\bar{R}_{k_n}^{df-M*}\right\|_{\mathfrak{L}^2(\Omega\times[0,T])}\\ &\leqslant C_1\cdot\varepsilon_n^{-1}\cdot\bar{\omega}(\lambda,T/(n\cdot\varepsilon_n))\cdot(1+(n\cdot\varepsilon_n)^{-1})^{1/2}\cdot(k_n/n)^{1/2}\cdot\varepsilon_n^{1/2}\\ &\quad + C_2\cdot(k_n^{1/2}/(n\cdot\varepsilon_n))\cdot(1+(n\cdot\varepsilon_n)^{-1})^{1/2},\end{aligned}$$

we obtain (2.42) and (2.43) for $\bar{X}^* = \bar{X}^{df-Lin-M*}$. This ends the proof. ■

2.3. Lower Bounds

We have the following result.

Theorem 2.7 ([72]). *Let us assume that the mappings a, b, c, and λ satisfy the assumptions (A) – (E). Let $\bar{X}$ be an arbitrary method from χ^{noneq}. Then*

$$\liminf_{n\to+\infty} \left(cost_n(\bar{X})\right)^{1/2} \cdot e_n(\bar{X}) \geqslant C^{\mathrm{noneq}}.$$

From Theorems 2.4, 2.6, and 2.7 we can obtain the following main result of this chapter.

Theorem 2.8. *We have that*

$$\lim_{n\to+\infty} n^{1/2} \cdot e^{\mathrm{noneq}}(n) = C^{\mathrm{noneq}},$$

and the methods $\bar{X}^{Lin-M*}$ and $\bar{X}^{df-Lin-M*}$ are asymptotically optimal in the class χ^{noneq} .

Chapter 3

Global approximation of solutions of multidimensional SDEs with jumps

In this section, not yet published, we consider the problem of approximation of stochastic differential equations given by (1.1) where $T > 0$ and $N = \{N(t)\}_{t\in[0,T]}$ is a one-dimensional non-homogeneous Poisson process, with intensity function λ, and $W = \{W(t)\}_{t\in[0,T]}$ is a m_{w}-dimensional Wiener process. According to our best knowledge these are first results in the case of multidimensional jump-diffusion SDEs.

3.1. The setting

Let $T > 0$ be a given real number, parameters $d, m_{\mathrm{w}} \in \mathbb{N}$, and $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. We consider on this space independent processes a m_{w}-dimensional Wiener process

$$W = \{W(t)\}_{t\in[0,T]},$$

$$W(t) = \begin{pmatrix} W_1(t) \\ W_2(t) \\ \vdots \\ W_{m_{\mathrm{w}}}(t) \end{pmatrix},$$

and a one–dimensional non–homogeneous Poisson process

$$N = \{N(t)\}_{t\in[0,T]}, \tag{3.1}$$

with intensity function $\lambda = \lambda(t)$. Let us denote by $\{\mathcal{F}_t\}_{t\in[0,T]}$ the complete filtration generated by the driving processes N and W.

To simplify notation we use the same symbol $\|\cdot\|$ for both Frobenius and Euclidean norm, and the meaning is clear from the context. In this chapter we always use Frobenius norm for matrices and Euclidean norm for vectors. Let for function $f \in \{a, b, c\}$, exists $K > 0$ such that

(A_{MD}) $f \in C^{0,2}\big([0,T] \times \mathbb{R}^d\big)$.

(B_{MD}) For all $t, s \in [0,T]$ and all $y, z \in \mathbb{R}^d$

- $(B1_{\mathrm{MD}})$ $\|f(t,y) - f(t,z)\| \leqslant K\|y - z\|$,
- $(B2_{\mathrm{MD}})$ $\|f(t,y) - f(s,y)\| \leqslant K\big(1 + \|y\|\big)|t - s|$,
- $(B3_{\mathrm{MD}})$ $\left\|\frac{\partial f}{\partial y_j}(t,y) - \frac{\partial f}{\partial y_j}(t,z)\right\| \leqslant K\|y - z\|$ for all $j \in \{1, \ldots, d\}$.

(C_{MD}) There exists $K > 0$ such that for $f \in \{b^1, \ldots, b^{m_{\mathrm{w}}}, c\}$, for all $t \in [0,T]$, $y, z \in \mathbb{R}^d$, $j \in \{1, 2, \ldots, m_{\mathrm{w}}\}$ we have

$$\|L_j f(t,y) - L_j f(t,z)\| \quad \leqslant \quad K\|y - z\|.$$

The diffusion and the jump coefficients satisfy the following *jump commutativity conditions.*

(D_{MD}) For all $(t,y) \in [0,T] \times \mathbb{R}^d$, all $j_1, j_2 \in \{1, 2, \ldots, m_{\mathrm{w}}\}$,

$$L_{j_1} b^{j_2}(t,y) \quad = \quad L_{j_2} b^{j_1}(t,y), \tag{3.2}$$

$$L_{j_1} c(t,y) \quad = \quad L_{-1} b^{j_1}(t,y). \tag{3.3}$$

(E_{MD}) For the intensity function $\lambda : [0,T] \to \mathbb{R}_+$ we assume that $\lambda \in C\big([0,T]\big)$.

By Appendix A.4 and the fact that a, b, c and λ satisfying $(B1_{\mathrm{MD}})$, $(B2_{\mathrm{MD}})$, and (E_{MD}) the equation (1.1) has a unique strong solution $X = \{X(t)\}_{t\in[0,T]}$ that is adapted to $\{\mathcal{F}_t\}_{t\in[0,T]}$ and has càdlàg paths.

Condition (D_{MD}) will allow the calculation of stochastic integrals defined in (B.4). More details about why we use this condition will be given in the next section, where the algorithm will be analyzed. (We refer to Chapter 6.3 in [61], where the conditions (3.2) and (3.3) are widely discussed.)

In order to characterize asymptotic behavior we define constant

$$C^{\mathrm{eq}}_{\mathrm{MD}} = \sqrt{\frac{T}{6}} \cdot \left(\int_0^T \mathbb{E}(\mathcal{Y}(t))\mathrm{d}t\right)^{1/2},$$

where the process $\{\mathcal{Y}(t)\}_{t\in[0,T]}$ is defined as

$$\mathcal{Y}(t) = \|b(t, X(t))\|^2 + \|c(t, X(t))\|^2 \cdot \lambda(t), \quad t \in [0, T].$$

Jump Commutative conditions

Before we present algorithm we would like to show examples of problems which satisfy *jump commutative conditions.*

Let us consider the problem of the following form

$$\begin{cases} \mathrm{d}X_i(t) = rX_i(t)\mathrm{d}t + \sum_{j=1}^{m_{\mathrm{w}}} \sigma^{i,j} X_i(t)\mathrm{d}W_j(t) + c_i X_i(t)\mathrm{d}N(t), \quad t \in [0, T], \\ X_i(0) > 0, \quad i = 1, \ldots, d. \end{cases}$$

In that case for $x \in \mathbb{R}^d$ we define the functions

$$a(t, x) = r \cdot x = r \cdot (x_1, \ldots, x_d)^T, \tag{3.4}$$

$$b(t, x) = [\sigma^{i,j} x_i]_{i,j=1}^{d,m_{\mathrm{w}}}, \tag{3.5}$$

$$c(t, x) = (c^1 x_1, \ldots, c^d x_d)^T. \tag{3.6}$$

Let $\sigma^{i,j} = \sigma^{i,j}(t)$ and $c^i = c^i(t)$ for all $j \in \{1, 2, \ldots, m_{\mathrm{w}}\}$ and $i \in \{1, 2, \ldots, d\}$. Let $y \in \mathbb{R}^d$, then the functions defined by (3.5) and (3.6) satisfy condition (D_{MD}).

We now justify the claim above. For all $j_1, j_2 \in \{1, 2, \ldots, m_{\mathrm{w}}\}$ let us check the equality $L_{j_1} b^{j_2}(t, y) = L_{j_2} b^{j_1}(t, y)$

$$L_{j_1} b^{j_2}(t, y) = \begin{pmatrix} \sigma^{1,j_2}(t) & 0 & \cdots & 0 \\ 0 & \sigma^{2,j_2}(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^{d,j_2}(t) \end{pmatrix} \cdot \begin{pmatrix} \sigma^{1,j_1}(t) y_1 \\ \sigma^{2,j_1}(t) y_2 \\ \vdots \\ \sigma^{d,j_1}(t) y_d \end{pmatrix} = \begin{pmatrix} \sigma^{1,j_2}(t)\sigma^{1,j_1}(t) y_1 \\ \sigma^{2,j_2}(t)\sigma^{2,j_1}(t) y_2 \\ \vdots \\ \sigma^{d,j_2}(t)\sigma^{d,j_1}(t) y_d \end{pmatrix},$$

$$L_{j_2} b^{j_1}(t, y) = \begin{pmatrix} \sigma^{1,j_1}(t) & 0 & \cdots & 0 \\ 0 & \sigma^{2,j_1}(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^{d,j_1}(t) \end{pmatrix} \cdot \begin{pmatrix} \sigma^{1,j_2}(t) y_1 \\ \sigma^{2,j_2}(t) y_2 \\ \vdots \\ \sigma^{d,j_2}(t) y_d \end{pmatrix} = \begin{pmatrix} \sigma^{1,j_1}(t)\sigma^{1,j_2}(t) y_1 \\ \sigma^{2,j_1}(t)\sigma^{2,j_2}(t) y_2 \\ \vdots \\ \sigma^{d,j_1}(t)\sigma^{d,j_2}(t) y_d \end{pmatrix}.$$

This implies that $L_{j_1} b^{j_2}(t,y) = L_{j_2} b^{j_1}(t,y)$. For all $j_1 \in \{1,2,\ldots,m_{\rm w}\}$ let us check the equality $L_{j_1} c(t,y) = L_{-1} b^{j_1}(t,y)$

$$L_{j_1} c(t,y) = \begin{pmatrix} c^1(t) & 0 & \cdots & 0 \\ 0 & c^2(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c^d(t) \end{pmatrix} \cdot \begin{pmatrix} \sigma^{1,j_1}(t) y_1 \\ \sigma^{2,j_1}(t) y_2 \\ \vdots \\ \sigma^{d,j_1}(t) y_d \end{pmatrix} = \begin{pmatrix} c^1(t)\sigma^{1,j_1}(t) y_1 \\ c^2(t)\sigma^{2,j_1}(t) y_2 \\ \vdots \\ c^d(t)\sigma^{d,j_1}(t) y_d \end{pmatrix},$$

$$L_{-1} b^{j_1}(t,y) = \begin{pmatrix} \sigma^{1,j_1}(t)(y_1 + c^1(t)) y_1) \\ \sigma^{2,j_1}(t)(y_2 + c^2(t)) y_2) \\ \vdots \\ \sigma^{d,j_1}(t)(y_d + c^d(t)) y_d) \end{pmatrix} - \begin{pmatrix} \sigma^{1,j_1}(t) y_1 \\ \sigma^{2,j_1}(t) y_2 \\ \vdots \\ \sigma^{d,j_1}(t) y_d \end{pmatrix} = \begin{pmatrix} c^1(t)\sigma^{1,j_1}(t) y_1 \\ c^2(t)\sigma^{2,j_1}(t) y_2 \\ \vdots \\ c^d(t)\sigma^{d,j_1}(t) y_d \end{pmatrix}.$$

This implies that $L_{j_1} c(t,y) = L_{-1} b^{j_1}(t,y)$. That ends the proof of fact that problem defined by functions (3.4) – (3.6) satisfy jump commutative conditions.

If

$$\begin{aligned} a(t,x) &= r \cdot x = r \cdot (x_1,\ldots,x_d)^T, \\ b(t,x) &= [\sigma^{i,j}]_{i,j=1}^{d,m_{\rm w}}, \\ c(t,x) &= (c^1,\ldots,c^d)^T. \end{aligned}$$

It is easy to see that for all $j_1, j_2 \in \{1,2,\ldots,m_{\rm w}\}$ and for all $(t,y) \in [0,T] \times \mathbb{R}^d$ $L_{j_1} b^{j_2}(t,y) = 0$, $L_{-1} b^{j_1}(t,y) = 0$, and $L_{j_1} c(t,y) = 0$. It means that in this case the condition ($D_{\rm MD}$) is also satisfied. For more examples see also page 227 in [61].

3.2. Algorithm based on equidistant mesh

In this section we present an implementable and asymptotically optimal algorithm in the class $\chi^{\rm eq}$, which is based on equidistant mesh.

3.2.1. Description of the method and its asymptotic performance

We define the algorithm based on equidistant mesh.

STEP 1 Take any $n \in \mathbb{N}$ and let $t_{0,n} = 0$, $\bar{X}^M(t_{0,n}) = x_0$. Set $i := 0$.

STEP 2 If $t_{i,n} \in [0,T)$ and $\bar{X}^M(t_{i,n})$ are given, then compute

$$t_{i+1,n} = t_{i,n} + \frac{T}{n}.$$

If $t_{i+1,n} \leqslant T$, then compute

$$\begin{aligned}
\bar{X}^M(t_{i+1,n}) &= \bar{X}^M(t_{i,n}) + a(t_{i,n}, \bar{X}^M(t_{i,n})) \cdot (t_{i+1,n} - t_{i,n}) \\
&+ b(t_{i,n}, \bar{X}^M(t_{i,n})) \cdot (W(t_{i+1,n}) - W(t_{i,n})) \\
&+ c(t_{i,n}, \bar{X}^M(t_{i,n})) \cdot (N(t_{i+1,n}) - N(t_{i,n})) \\
&+ \frac{1}{2} \sum_{j_1,j_2=1}^{m_{\mathrm{w}}} L_{j_1} b^{j_2}(t_{i,n}, \bar{X}^M(t_{i,n})) \\
&\qquad\qquad \times \Big(I_{t_{i,n},t_{i+1,n}}(W_{j_1}, W_{j_2}) + I_{t_{i,n},t_{i+1,n}}(W_{j_2}, W_{j_1}) \Big) \\
&+ L_{-1} c(t_{i,n}, \bar{X}^M(t_{i,n})) \cdot I_{t_{i,n},t_{i+1,n}}(N, N) \\
&+ \sum_{j_1=1}^{m_{\mathrm{w}}} L_{-1} b^{j_1}(t_{i,n}, \bar{X}^M(t_{i,n})) \cdot \Big(I_{t_{i,n},t_{i+1,n}}(W_{j_1}, N) + I_{t_{i,n},t_{i+1,n}}(N, W_{j_1}) \Big),
\end{aligned}$$

take $i := i + 1$ and GOTO **STEP 2**.

STOP

Running n through the natural numbers, we obtain the sequence of equidistant discretizations $\Delta = \{\Delta_n\}_{n\in\mathbb{N}}$, where

$$\Delta_n = \{t_{0,n}, t_{1,n}, \ldots, t_{n,n}\}, \quad n \in \mathbb{N}.$$

We have that $t_{i,n} \leqslant T$ for all $i = 0, 1, \ldots, n$. So we observe the Poisson and the Wiener processes only in $[0, T]$. Here the definition of discretizations satisfies the assumptions of the calculation model. By $\bar{X}^{cM*} = \{\bar{X}^{cM}_n\}_{n\in\mathbb{N}}$ we denote the *conditional Milstein method* based on the sequence of discretizations Δ, which is defined as

$$\bar{X}^{cM}_n(t) = \mathbb{E}\big(\tilde{X}^M_n(t) \mid \mathcal{N}_n(N, W)\big), \quad t \in [0, T],$$

where $\{\tilde{X}^M_n\}_{n\in\mathbb{N}}$ is a sequence of the time-continuous Milstein approximations (B.2) – (B.3) based on discretization $\{\Delta_n\}_{n\in\mathbb{N}}$.

By Lemma B.24 – B.27 we can calculate that $\bar{X}_n^{cM}(t)$ is given by

$$\begin{aligned}
\bar{X}_n^{cM}(t) &= \tilde{X}_n^M(t_{i,n}) + a(t_{i,n}, \tilde{X}_n^M(t_{i,n})) \cdot (t - t_{i,n}) \\
&\quad + b(t_{i,n}, \tilde{X}_n^M(t_{i,n})) \cdot \Delta W_{i,n} \cdot \frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \\
&\quad + c(t_{i,n}, \tilde{X}_n^M(t_{i,n})) \cdot \Delta N_{i,n} \cdot \frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} \\
&\quad + \frac{1}{2} \sum_{j_1, j_2 = 1}^{m_{\mathrm{w}}} L_{j_1} b^{j_2}(t_{i,n}, \tilde{X}_n^M(t_{i,n})) \\
&\qquad\quad \times \Big(I_{t_{i,n}, t_{i+1,n}}(W_{j_1}, W_{j_2}) + I_{t_{i,n}, t_{i+1,n}}(W_{j_2}, W_{j_1}) \Big) \cdot \left(\frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \right)^2 \\
&\quad + \sum_{j_1 = 1}^{m_{\mathrm{w}}} L_{-1} b^{j_1}(t_{i,n}, \tilde{X}_n^M(t_{i,n})) \cdot \Big(I_{t_{i,n}, t_{i+1,n}}(W_{j_1}, N) + I_{t_{i,n}, t_{i+1,n}}(N, W_{j_1}) \Big) \\
&\qquad\quad \times \frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} \cdot \frac{t - t_{i,n}}{t_{i+1,n} - t_{i,n}} \\
&\quad + L_{-1} c(t_{i,n}, \tilde{X}_n^M(t_{i,n})) \cdot I_{t_{i,n}, t_{i+1,n}}(N, N) \cdot \left(\frac{\Lambda(t, t_{i,n})}{\Lambda(t_{i+1,n}, t_{i,n})} \right)^2,
\end{aligned}$$

for $t \in [t_{i,n}, t_{i+1,n}]$, $i = 0, 1, \ldots, n-1$ and $\bar{X}_n^{cM}(0) = x_0$,

$$\Delta W_{i,n} = W(t_{i+1,n}) - W(t_{i,n}),$$

$$\Delta N_{i,n} = N(t_{i+1,n}) - N(t_{i,n}).$$

Note that $\bar{X}_n^{cM}$ has continuous trajectories and coincides with $\tilde{X}_n^M$ at the discretization points. The disadvantage of this algorithm is the usage of the values of Λ. Hence, we also define the piecewise linear interpolation $\bar{X}_n^{Lin-M}$ of the classical Milstein steps by

$$\bar{X}_n^{Lin-M}(t) = \frac{\tilde{X}_n^M(t_{i,n}) \cdot (t_{i+1,n} - t) + \tilde{X}_n^M(t_{i+1,n}) \cdot (t - t_{i,n})}{t_{i+1,n} - t_{i,n}},$$

for $t \in [t_{i,n}, t_{i+1,n}]$, $i = 0, 1, \ldots, n-1$. In general, the method $\bar{X}_n^{cM}$ is not equal to $\bar{X}_n^{Lin-M}$, but in discretization points we have that for all $i = 0, 1, \ldots, n-1$

$$\bar{X}_n^{Lin-M}(t_{i,n}) = \bar{X}_n^{cM}(t_{i,n}) = \tilde{X}_n^M(t_{i,n}) = \bar{X}^M(t_{i,n}).$$

However, it is convenient as in the scalar case to use the method $\bar{X}^{cM} = \left\{\bar{X}_n^{cM}\right\}_{n\in\mathbb{N}}$ in order to investigate the error of $\bar{X}^{Lin-M} = \left\{\bar{X}_n^{Lin-M}\right\}_{n\in\mathbb{N}}$. We show that they behave asymptotically in the same way. Moreover, the method $\bar{X}_n^{Lin-M}$ for any fixed discretization Δ_n does not evaluate Λ, and it is implementable. Both methods $\bar{X}_n^{cM}$ and $\bar{X}_n^{Lin-M}$ use $(m_{\mathrm{w}}+1)\cdot n$ values of the processes N and W, when $b \not\equiv 0$ and $c \not\equiv 0$.

We also have the following results. The proof of the following results go by using the extension of the technique proposed by author of [44].

Theorem 3.1. *Let us assume that the mappings a, b, c and λ satisfy the assumptions (A_{MD}) – (E_{MD}). Let $\bar{X}$ be an arbitrary method from χ^{eq}. Then we have the following upper bounds.*

(i) If $b \not\equiv 0$ and $c \not\equiv 0$ then

$$\liminf_{n\to+\infty} \big(cost_n(\bar{X})\big)^{1/2} \cdot e_n(\bar{X}) \geqslant (m_{\mathrm{w}}+1)^{1/2} \cdot C^{\mathrm{eq}}_{\mathrm{MD}}. \tag{3.7}$$

(ii) If $b \not\equiv 0$ and $c \equiv 0$ then

$$\liminf_{n\to+\infty} \big(cost_n(\bar{X})\big)^{1/2} \cdot e_n(\bar{X}) \geqslant (m_{\mathrm{w}})^{1/2} \cdot C^{\mathrm{eq}}_{\mathrm{MD}}. \tag{3.8}$$

(iii) If $b \equiv 0$ and $c \not\equiv 0$ then

$$\liminf_{n\to+\infty} \big(cost_n(\bar{X})\big)^{1/2} \cdot e_n(\bar{X}) \geqslant C^{\mathrm{eq}}_{\mathrm{MD}}. \tag{3.9}$$

Proof. We start with showing (3.7) in the case when $b \not\equiv 0$ and $c \not\equiv 0$. Let $\bar{X} = \{\bar{X}_n\}_{n\in\mathbb{N}} \in \chi^{\mathrm{eq}}$ be a method based on sequence of uniform discretizations $\bar{\Delta} = \{\Delta_n^{\mathrm{eq}}\}_{n\in\mathbb{N}}$, where each $\Delta_n^{\mathrm{eq}} = \{jT/n \mid j = 0,1,\ldots,n\}$. Therefore, we have that for all $i \in \{0,1,\ldots,n-1\}$,

$$t_{i+1,n} - t_{i,n} = \frac{T}{n}. \tag{3.10}$$

We denote by $\mathcal{N}(N,W) = \{\mathcal{N}_n(N,W)\}_{n\in\mathbb{N}}$, where each vector $\mathcal{N}_n(N,W)$ consists of the values of N and W at Δ_n^{eq}, i.e.,

$$\begin{aligned}\mathcal{N}_n(N,W) \;=\; & \big[N(t_{1,n}), N(t_{2,n}), \ldots, N(t_{n,n}),\\ & W(t_{1,n}), W(t_{2,n}), \ldots, W(t_{n,n})\big].\end{aligned} \tag{3.11}$$

Every $\bar{X}_n$ uses information (3.11) about the processes N and W. Let us denote by $\{\tilde{X}_n^M\}_{n\in\mathbb{N}}$ the sequence of continuous Milstein approximations (B.2) – (B.3) based on the sequence of discretizations $\bar{\Delta}$ and which use the information $\mathcal{N}(N,W)$ about the processes N and W. From Theorem B.1 and fact that we consider equidistant mesh we have that

$$\|X - \tilde{X}_n^M\|_{\mathfrak{L}^2(\Omega\times[0,T])} \leqslant C \cdot n^{-1}, \tag{3.12}$$

where the positive constant C does not depend on n. Moreover, let

$$\hat{Z}_n(t) = Z(t) - \mathbb{E}\big(Z(t) \mid \mathcal{N}_n(Z)\big),$$

for $Z \in \big\{N, W, W_1, \ldots, W_{m_{\mathrm{w}}}\big\}$ and $t \in [0,T]$. Note that for any $t \in [t_{i,n}, t_{i+1,n}]$ the random variable $\hat{Z}_n(t)$ is a convex combination of $Z(t) - Z(t_{i,n})$ and $-(Z(t_{i+1,n}) - Z(t))$. Hence, $\hat{Z}_n(t)$ is independent of $\mathcal{F}_{t_{i,n}}$ for all $t \in [t_{i,n}, t_{i+1,n}]$ and the processes $\big\{\hat{N}_n(t)\big\}_{t\in[0,T]}$, $\big\{\hat{W}_n(t)\big\}_{t\in[0,T]}$ are independent. By the definition of Wiener process we also have that $\mathbb{E}\big(\hat{W}_{j,n}(t)\big) = 0$. Moreover, random variable $\hat{N}_n(t) \cdot \hat{W}_{j,n}(t)$ for $t \in [t_{i,n}, t_{i+1,n}]$ and $j \in \{1, \ldots, m_{\mathrm{w}}\}$ are independent of $\mathcal{F}_{t_{i,n}}$. For almost all $t \in [0,T]$ we have that

$$\mathbb{E}\big\|\tilde{X}_n^M(t) - \bar{X}_n^M(t)\big\|^2 \geqslant \mathbb{E}\big\|\tilde{X}_n^M(t) - \mathbb{E}(\tilde{X}_n^M(t) \mid \mathcal{N}_n(N,W))\big\|^2.$$

From (3.10), (3.12), (B.61) and Lemma B.11 we have that

$$\begin{aligned}
e_n(\bar{X}) &\geqslant \big\|\bar{X}_n - \tilde{X}_n^M\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} - \big\|X - \tilde{X}_n^M\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} \\
&\geqslant \big\|\tilde{X}_n^M - \mathbb{E}\big(\tilde{X}_n^M \mid \mathcal{N}_n(N,W)\big)\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} - C\cdot n^{-1} \\
&\geqslant \big\|\tilde{H}_n^M + \tilde{R}_n^M\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} - C\cdot n^{-1} \\
&\geqslant \big\|\tilde{H}_n^M\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} - C\cdot n^{-1}.
\end{aligned} \tag{3.13}$$

Let $U_{i,n} = (t_{i,n}, \tilde{X}_n^M(t_{i,n}))$, then we have that

$$\begin{aligned}
\big\|\tilde{H}_n^M\big\|^2_{\mathfrak{L}^2(\Omega\times[0,T])} &= \int_0^T \mathbb{E}\big\|\tilde{H}_n^M(t)\big\|^2 \mathrm{d}t \\
&= \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} \mathbb{E}\big\|b(U_{i,n})\cdot\hat{W}_n(t) + c(U_{i,n})\cdot\hat{N}_n(t)\big\|^2 \mathrm{d}t.
\end{aligned}$$

For all $i \in \{0,1,\ldots,n-1\}$ and $t \in [t_{i,n}, t_{i+1,n}]$ we have that

$$\begin{aligned}
&\mathbb{E}\big\|b(U_{i,n})\cdot\hat{W}_n(t) + c(U_{i,n})\cdot\hat{N}_n(t)\big\|^2 \\
&\quad= \mathbb{E}\sum_{k=1}^{d}\left(\sum_{j=1}^{m_{\mathrm{w}}} b^{k,j}(U_{i,n})\cdot\hat{W}_{j,n}(t) + c^k(U_{i,n})\cdot\hat{N}_n(t)\right)^2 \\
&\quad= \mathbb{E}\sum_{k=1}^{d}\left(\sum_{j=1}^{m_{\mathrm{w}}} b^{k,j}(U_{i,n})\cdot\hat{W}_{j,n}(t)\right)^2 \\
&\qquad+ 2\,\mathbb{E}\sum_{k=1}^{d}\sum_{j=1}^{m_{\mathrm{w}}} b^{k,j}(U_{i,n})\cdot\hat{W}_{j,n}(t)\cdot c^k(U_{i,n})\cdot\hat{N}_n(t) \\
&\qquad+ \mathbb{E}\sum_{k=1}^{d}\Big(c^k(U_{i,n})\cdot\hat{N}_n(t)\Big)^2.
\end{aligned} \tag{3.14}$$

By the fact that for all $k \in \{1,\dots,d\}$ and $j \in \{1,\dots,m_{\mathrm{w}}\}$, we have that $b^{k,j}(U_{i,n}) \cdot c^k(U_{i,n})$ are $\mathcal{F}_{t_{i,n}}$ measurable. Then by the Hölder inequality and Theorem B.1 we have that $\mathbb{E}\big|b^{k,j}(U_{i,n}) \cdot c^k(U_{i,n})\big| < \infty$. So we obtain

$$\begin{aligned}\mathbb{E}\big(b^{k,j}(U_{i,n}) \cdot \hat{W}_{j,n}(t) \cdot c^k(U_{i,n}) \cdot \hat{N}_n(t)\big) &= \mathbb{E}\big(b^{k,j}(U_{i,n}) \cdot c^k(U_{i,n})\big) \cdot \mathbb{E}\big(\hat{W}_{j,n}(t) \cdot \hat{N}_n(t)\big) \\ &= \mathbb{E}\big(b^{k,j}(U_{i,n}) \cdot c^k(U_{i,n})\big) \cdot \mathbb{E}\big(\hat{W}_{j,n}(t)\big) \cdot \mathbb{E}\big(\hat{N}_n(t)\big) = 0. \end{aligned} \tag{3.15}$$

By (3.14) and (3.15) we have that

$$\big\|\tilde{H}_n^M\big\|^2_{\mathfrak{L}^2(\Omega\times[0,T])} = \sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} \Big(\mathbb{E}\big\|b(U_{i,n}) \cdot \hat{W}_n(t)\big\|^2 + \mathbb{E}\big\|c(U_{i,n}) \cdot \hat{N}_n(t)\big\|^2\Big)\mathrm{d}t. \tag{3.16}$$

By the fact that for all $k \in \{1,\dots,d\}$ and $j \in \{1,\dots,m_{\mathrm{w}}\}$, $b^{k,j}(U_{i,n})$ and $c(U_{i,n})$ are $\mathcal{F}_{t_{i,n}}$-measurable. Processes $\hat{N}_n, \hat{W}_{j,n}$ are independent of $\mathcal{F}_{t_{i,n}}$, so we have that

$$\begin{aligned}\mathbb{E}\big\|b(U_{i,n}) \cdot \hat{W}_n(t)\big\|^2 &= \mathbb{E}\left(\sum_{k=1}^{d}\Big(\sum_{j=1}^{m_{\mathrm{w}}} b^{k,j}(U_{i,n}) \cdot \hat{W}_{j,n}(t)\Big)^2\right) \\ &= \sum_{k=1}^{d}\sum_{j=1}^{m_{\mathrm{w}}} \mathbb{E}\Big(b^{k,j}(U_{i,n}) \cdot \hat{W}_{j,n}(t)\Big)^2 \\ &= \sum_{k=1}^{d}\sum_{j=1}^{m_{\mathrm{w}}} \mathbb{E}\big(b^{k,j}(U_{i,n})\big)^2 \cdot \mathbb{E}\big(\hat{W}_{j,n}(t)\big)^2 \\ &= \sum_{j=1}^{m_{\mathrm{w}}} \mathbb{E}\big\|b^j(U_{i,n})\big\|^2 \cdot \mathbb{E}\big(\hat{W}_{j,n}(t)\big)^2, \end{aligned} \tag{3.17}$$

and

$$\mathbb{E}\big\|c(U_{i,n}) \cdot \hat{N}_n(t)\big\|^2 = \mathbb{E}\|c(U_{i,n})\|^2 \cdot \mathbb{E}\big(\hat{N}_n(t)\big)^2. \tag{3.18}$$

Finally, by (3.16) – (3.18) we have that

$$\begin{aligned}\big\|\tilde{H}_n^M\big\|^2_{\mathfrak{L}^2(\Omega\times[0,T])} = &\left(\sum_{i=0}^{n-1} \int_{t_{i,n}}^{t_{i+1,n}} \Big(\sum_{j=1}^{m_{\mathrm{w}}} \mathbb{E}\big\|b^j(U_{i,n})\big\|^2 \cdot \mathbb{E}\big(\hat{W}_{j,n}(t)\big)^2\right. \\ &\left. + \ \mathbb{E}\big\|c(U_{i,n})\big\|^2 \cdot \mathbb{E}\big(\hat{N}_n(t)\big)^2\Big)\mathrm{d}t\right)^{1/2}. \end{aligned} \tag{3.19}$$

Now, we analyze the asymptotic behavior of the first term in (3.19). From Theorem B.20 we have that for all $j \in \{1,2,\dots,m_{\mathrm{w}}\}$ (analogously like in (2.28)) it follows that

$$\int_{t_{i,n}}^{t_{i+1,n}} \mathbb{E}\big(\hat{W}_{j,n}(t)\big)^2\mathrm{d}t = \frac{1}{6}(t_{i+1,n} - t_{i,n})^2. \tag{3.20}$$

From Lemma B.32 we have that

$$\begin{aligned}
&\left|\sum_{i=0}^{n-1}\mathbb{E}\big\|b(U_{i,n})\big\|^2\cdot\frac{T}{n}-\sum_{i=0}^{n-1}\mathbb{E}\big\|b(t_{i,n},X(t_{i,n}))\big\|^2\cdot\frac{T}{n}\right|\leqslant\\
&\qquad\leqslant\ \sum_{i=0}^{n-1}\Big|\mathbb{E}\big\|b(U_{i,n})\big\|^2-\mathbb{E}\big\|b(t_{i,n},X(t_{i,n}))\big\|^2\Big|\cdot\frac{T}{n}\\
&\qquad\leqslant\ \sum_{i=0}^{n-1}\sum_{j=0}^{m_{\mathrm{w}}}\Big|\mathbb{E}\big\|b^j(U_{i,n})\big\|^2-\mathbb{E}\big\|b^j(t_{i,n},X(t_{i,n}))\big\|^2\Big|\cdot\frac{T}{n}\\
&\qquad\leqslant\ \sum_{i=0}^{n-1}C\cdot\Big(1+\sup_{t\in[0,T]}\big\|\tilde{X}_n^M(t)\big\|_{\mathfrak{L}^2(\Omega)}+\sup_{t\in[0,T]}\big\|X(t)\big\|_{\mathfrak{L}^2(\Omega)}\Big)\\
&\qquad\qquad\qquad\times\ \big\|\tilde{X}_n^M(t_{i,n})-X(t_{i,n})\big\|_{\mathfrak{L}^2(\Omega)}\cdot\frac{T}{n}\\
&\qquad\leqslant\ C_1/n.
\end{aligned}\tag{3.21}$$

By (3.21) and by Fact B.30 we have that $[0,T]\ni t\to\mathbb{E}\big\|b(t,X(t))\big\|^2$ is continuous, we arrive that

$$\lim_{n\to+\infty}\sum_{i=0}^{n-1}\mathbb{E}\big\|b(U_{i,n})\big\|^2\cdot\frac{T}{n}=\int_0^T\mathbb{E}\big\|b(t,X(t))\big\|^2\mathrm{d}t\tag{3.22}$$

The asymptotic behavior of the second term in (3.19) goes from the following consideration. Analogously like in (2.29) we have that

$$\int_{t_{i,n}}^{t_{i+1,n}}\mathbb{E}\big(\hat{N}_n(t)\big)^2\mathrm{d}t=\frac{1}{6}\frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})}\cdot(t_{i+1,n}-t_{i,n})^2,\tag{3.23}$$

for some $\alpha_{i,n},\beta_{i,n},\gamma_{i,n}\in[t_{i,n},t_{i+1,n}]$, $i=0,1,\ldots,n-1$.

From Lemma B.32 we have that

$$\begin{aligned}
&\left|\sum_{i=0}^{n-1}\frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})}\cdot\mathbb{E}\big\|c(U_{i,n})\big\|^2\cdot\frac{T}{n}-\sum_{i=0}^{n-1}\lambda(t_{i,n})\cdot\mathbb{E}\big\|c(t_{i,n},X(t_{i,n}))\big\|^2\cdot\frac{T}{n}\right|\leqslant\\
&\qquad\leqslant\sum_{i=0}^{n-1}\left|\frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})}\cdot\mathbb{E}\big\|c(U_{i,n})\big\|^2-\lambda(t_{i,n})\cdot\mathbb{E}\big\|c(t_{i,n},X(t_{i,n}))\big\|^2\right|\cdot\frac{T}{n}\\
&\qquad\leqslant\sum_{i=0}^{n-1}\left|\frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})}\right|\cdot\left|\mathbb{E}\big\|c(U_{i,n})\big\|^2-\mathbb{E}\big\|c(t_{i,n},X(t_{i,n}))\big\|^2\right|\cdot\frac{T}{n}\\
&\qquad\qquad+\sum_{i=0}^{n-1}\left|\frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})}-\lambda(t_{i,n})\right|\cdot\mathbb{E}\big\|c(t_{i,n},X(t_{i,n}))\big\|^2\cdot\frac{T}{n}\\
&\qquad\leqslant\sum_{i=0}^{n-1}C\Big(1+\sup_{t\in[0,T]}\big\|\tilde{X}_n^M(t)\big\|_{\mathfrak{L}^2(\Omega)}+\sup_{t\in[0,T]}\big\|X(t)\big\|_{\mathfrak{L}^2(\Omega)}\Big)\\
&\qquad\qquad\qquad\times\big\|\tilde{X}_n^M(t_{i,n})-X(t_{i,n})\big\|_{\mathfrak{L}^2(\Omega)}\cdot\frac{T}{n}\\
&\qquad\qquad+C_2\cdot\|1/\lambda\|_\infty\cdot\|\lambda\|_\infty\cdot\sum_{i=0}^{n-1}\Big(|\lambda(\beta_{i,n})-\lambda(t_{i,n})|\\
&\qquad\qquad\qquad+|\lambda(\alpha_{i,n})-\lambda(\gamma_{i,n})|\Big)\cdot\frac{T}{n}\\
&\qquad\leqslant C_1/n+C_3\cdot\bar{\omega}(\lambda,T/n).
\end{aligned}\tag{3.24}$$

By (3.24) and Fact B.30 we have that $[0,T]\ni t\to\lambda(t)\cdot\mathbb{E}\big\|c(t,X(t))\big\|^2$ is continuous and we arrive that

$$\lim_{n\to+\infty}\sum_{i=0}^{n-1}\frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})}\cdot\mathbb{E}\|c(U_{i,n})\|^2\cdot\frac{T}{n}=\int_0^T\lambda(t)\cdot\mathbb{E}\|c(t,X(t))\|^2\mathrm{d}t.\tag{3.25}$$

By (3.13), (3.20), (3.23), (3.22), and (3.25) we have that

$$\begin{aligned}
\liminf_{n\to+\infty}n^{1/2}\cdot e(\bar{X})\ &\geqslant\ \liminf_{n\to+\infty}n^{1/2}\big\|\tilde{H}_n^M\big\|_{\mathfrak{L}^2(\Omega\times[0,T])}\\
&=\ \liminf_{n\to+\infty}n^{1/2}\Bigg(\sum_{i=0}^{n-1}\sum_{j=1}^{m_{\mathrm{w}}}\mathbb{E}\big\|b^j(U_{i,n})\big\|^2\cdot\frac{T^2}{6n^2}\\
&\qquad\qquad+\ \mathbb{E}\|c(U_{i,n})\|^2\cdot\frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})}\cdot\frac{T^2}{6n^2}\Bigg)^{1/2}\\
&=\ \liminf_{n\to+\infty}\sqrt{\frac{T}{6}}\Bigg(\sum_{i=0}^{n-1}\mathbb{E}\big\|b(U_{i,n})\big\|^2\cdot\frac{T}{n}\\
&\qquad\qquad+\ \sum_{i=0}^{n-1}\mathbb{E}\big\|c(U_{i,n})\big\|^2\cdot\frac{\lambda(\alpha_{i,n})\lambda(\beta_{i,n})}{\lambda(\gamma_{i,n})}\cdot\frac{T}{n}\Bigg)^{1/2}
\end{aligned}$$

$$= \sqrt{\frac{T}{6}}\left(\int_0^T \mathbb{E}\Big(\big\|b(t,X(t))\big\|^2 + \big\|c(t,X(t))\big\|^2 \cdot \lambda(t)\Big)\mathrm{d}t\right)^{1/2}$$

$$= \sqrt{\frac{T}{6}}\left(\int_0^T \mathbb{E}\big(\mathcal{Y}(t)\big)\mathrm{d}t\right)^{1/2}. \tag{3.26}$$

Therefore, by (3.26) we obtain

$$\liminf_{n\to+\infty} \big(cost_n(\bar{X})\big)^{1/2} \cdot e_n(\bar{X}) \geqslant (m_{\mathrm{w}}+1)^{1/2} \cdot C^{\mathrm{eq}}_{\mathrm{MD}},$$

which ends the proof of (3.7) in the case when $b \not\equiv 0$ and $c \not\equiv 0$.
If $b \not\equiv 0$ and $c \equiv 0$ then $cost_n(\bar{X}) = m_{\mathrm{w}} \cdot n$ which yield

$$\liminf_{n\to+\infty} \big(cost_n(\bar{X})\big)^{1/2} \cdot e_n(\bar{X}) \geqslant (m_{\mathrm{w}})^{1/2} \cdot C^{\mathrm{eq}}_{\mathrm{MD}}.$$

If $b \equiv 0$ and $c \not\equiv 0$ then $cost_n(\bar{X}) = n$, which yield

$$\liminf_{n\to+\infty} \big(cost_n(\bar{X})\big)^{1/2} \cdot e_n(\bar{X}) \geqslant C^{\mathrm{eq}}_{\mathrm{MD}}.$$

For $b \equiv 0$ and $c \equiv 0$ we obtain a trivial lower bound. That ends the proof. ■

Theorem 3.2. *Let us assume that the mappings $a, b, c,$ and λ satisfy the assumptions (A_{MD}) – (E_{MD}), then for $\bar{X} \in \{\bar{X}^{cM}, \bar{X}^{Lin-M}\}$ we have the following upper bounds.*

(i) If $b \not\equiv 0$ and $c \not\equiv 0$ then

$$\limsup_{n\to+\infty} \big(cost_n(\bar{X})\big)^{1/2} \cdot e_n(\bar{X}) \leqslant (m_{\mathrm{w}}+1)^{1/2} \cdot C^{\mathrm{eq}}_{\mathrm{MD}}. \tag{3.27}$$

(ii) If $b \not\equiv 0$ and $c \equiv 0$ then

$$\limsup_{n\to+\infty} \big(cost_n(\bar{X})\big)^{1/2} \cdot e_n(\bar{X}) \leqslant (m_{\mathrm{w}})^{1/2} \cdot C^{\mathrm{eq}}_{\mathrm{MD}}. \tag{3.28}$$

(iii) if $b \equiv 0$ and $c \not\equiv 0$ then

$$\limsup_{n\to+\infty} \big(cost_n(\bar{X})\big)^{1/2} \cdot e_n(\bar{X}) \leqslant C^{\mathrm{eq}}_{\mathrm{MD}}. \tag{3.29}$$

Proof. Firstly we prove (3.27) – (3.29) for $\bar{X}^{cM}$. We have that

$$\begin{aligned} e_n(\bar{X}^{cM}) &= \big\|X - \bar{X}^{cM}_n\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} \\ &\leqslant \big\|X - \bar{X}^{M}_n\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} + \big\|\bar{X}^{M}_n - \mathbb{E}\big(\bar{X}^{M}_n \mid \mathcal{N}_n(W,N)\big)\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} \\ &\leqslant Cn^{-1} + \big\|\tilde{H}^{M}_n\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} + \big\|\tilde{R}^{M}_n\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} \\ &\leqslant Cn^{-1} + \big\|\tilde{H}^{M}_n\big\|_{\mathfrak{L}^2(\Omega\times[0,T])}. \end{aligned} \tag{3.30}$$

Follow by steps analogously as in proof of Theorem 3.1 we have that

$$\limsup_{n\to+\infty} n^{1/2}\cdot e_n(\bar{X}^{cM}) \quad\leqslant\quad \limsup_{n\to+\infty} n^{1/2}\cdot \left\|\tilde{H}_n^M\right\|_{\mathfrak{L}^2(\Omega\times[0,T])} = C_{\text{MD}}^{eq}. \tag{3.31}$$

And then by (3.30) and (3.31) we have that

$$\limsup_{n\to+\infty} \left(cost_n(\bar{X}^{cM})\right)^{1/2}\cdot e_n(\bar{X}^{cM}) \leqslant (m_{\text{w}}+1)^{1/2}\cdot C_{\text{MD}}^{eq}.$$

That ends the proof of (3.27) in the case when $b\not\equiv 0$ and $c\not\equiv 0$.
If ($b\not\equiv 0$ and $c\equiv 0$) then $cost_n(\bar{X}^{eq}) = m_{\text{w}}\cdot n$ which yield

$$\limsup_{n\to+\infty} \left(cost_n(\bar{X}^{cM})\right)^{1/2}\cdot e_n(\bar{X}^{cM}) \leqslant (m_{\text{w}})^{1/2}\cdot C_{\text{MD}}^{eq}.$$

If ($b\equiv 0$ and $c\not\equiv 0$) then $cost_n(\bar{X}^{eq}) = n$, which yield

$$\limsup_{n\to+\infty} \left(cost_n(\bar{X}^{cM})\right)^{1/2}\cdot e_n(\bar{X}^{cM}) \leqslant C_{\text{MD}}^{eq}.$$

That ends the proof of (3.27) – (3.29) in a case of $\bar{X} = \bar{X}_n^{cM}$. □

Now we analyze the error of $\bar{X} = \bar{X}_n^{Lin-M}$. Note that

$$\begin{aligned}
\bar{R}_n^M(t) \quad &:= \quad \bar{X}_n^{cM}(t) - \bar{X}_n^{Lin-M}(t)\\
&= \quad c(U_{i,n})\cdot\Delta N_{i,n}\cdot\left(\frac{\Lambda(t,t_{i,n})}{\Lambda(t_{i+1,n},t_{i,n})} - \frac{t-t_{i,n}}{t_{i+1,n}-t_{i,n}}\right)\\
&\quad + \frac{1}{2}\sum_{j_1,j_2=1}^{m_{\text{w}}} L_{j_1}b^{j_2}(U_{i,n})\cdot\left(I_{t_{i,n},t_{i+1,n}}(W_{j_1},W_{j_2}) + I_{t_{i,n},t_{i+1,n}}(W_{j_2},W_{j_1})\right)\\
&\qquad\qquad \times\frac{(t-t_{i,n})\cdot(t-t_{i+1,n})}{(t_{i+1,n}-t_{i,n})^2}\\
&\quad + \sum_{j_1=1}^{m_{\text{w}}} L_{-1}b^{j_1}(U_{i,n})\cdot\left(I_{t_{i,n},t_{i+1,n}}(W_{j_1},N) + I_{t_{i,n},t_{i+1,n}}(N,W_{j_1})\right)\\
&\qquad\qquad \times\frac{t_{i,n}-t}{t_{i+1,n}-t_{i,n}}\cdot\frac{\Lambda(t_{i+1,n},t)}{\Lambda(t_{i+1,n},t_{i,n})}\\
&\quad + L_{-1}c(U_{i,n})\cdot I_{t_{i,n},t_{i+1,n}}(N,N)\cdot\left(\left(\frac{\Lambda(t,t_{i,n})}{\Lambda(t_{i+1,n},t_{i,n})}\right)^2 - \frac{t-t_{i,n}}{t_{i+1,n}-t_{i,n}}\right),
\end{aligned}$$

for $t\in[t_{i,n},t_{i+1,n}]$, $i = 0,1,\ldots,k_n-1$. By the fact that $c(U_{i,n})$, $L_{j_1}b^{j_2}(U_{i,n})$, $L_{-1}b^{j_1}(U_{i,n})$, $L_{-1}c(U_{i,n})$ are $\mathcal{F}_{t_{i,n}}$-measurable for all $j_1,j_2\in\{1,\ldots,m_{\text{w}}\}$ and

$\Delta N_{i,n}$, $I_{t_{i,n},t_{i+1,n}}(W_{j_1},W_{j_2})$, $I_{t_{i,n},t_{i+1,n}}(N,N)$, $I_{t_{i,n},t_{i+1,n}}(N,W_{j_1})$, $I_{t_{i,n},t_{i+1,n}}(W_{j_1},N)$ are independent of $\mathcal{F}_{t_{i,n}}$. Together with Lemma B.28 we have that

$$\begin{aligned}
\mathbb{E}\big\|\bar{R}_n^M(t)\big\|^2 \leqslant\ & \mathbb{E}\|c(U_{i,n})\|^2 \cdot \mathbb{E}|\Delta N_{i,n}|^2 \cdot \left|\frac{\Lambda(t,t_{i,n})}{\Lambda(t_{i+1,n},t_{i,n})} - \frac{t-t_{i,n}}{t_{i+1,n}-t_{i,n}}\right|^2 \\
&+ \frac{1}{2}\sum_{j_1,j_2=1}^{m_{\mathrm{w}}} \mathbb{E}\|L_{j_1}b^{j_2}(U_{i,n})\|^2 \cdot \mathbb{E}\big|I_{t_{i,n},t_{i+1,n}}(W_{j_1},W_{j_2}) + I_{t_{i,n},t_{i+1,n}}(W_{j_2},W_{j_1})\big|^2 \\
&\qquad \times \left|\frac{(t-t_{i,n})\cdot(t-t_{i+1,n})}{(t_{i+1,n}-t_{i,n})^2}\right|^2 \\
&+ \sum_{j_1=1}^{m_{\mathrm{w}}} \mathbb{E}\|L_{-1}b^{j_1}(U_{i,n})\|^2 \cdot \mathbb{E}\big|I_{t_{i,n},t_{i+1,n}}(W_{j_1},N) + I_{t_{i,n},t_{i+1,n}}(N,W_{j_1})\big|^2 \\
&\qquad \times \left|\frac{t_{i,n}-t}{t_{i+1,n}-t_{i,n}} \cdot \frac{\Lambda(t_{i+1,n},t)}{\Lambda(t_{i+1,n},t_{i,n})}\right|^2 \\
&+ \mathbb{E}\|L_{-1}c(U_{i,n})\|^2 \cdot \mathbb{E}\big|I_{t_{i,n},t_{i+1,n}}(N,N)\big|^2 \\
&\qquad \times \left|\left(\frac{\Lambda(t,t_{i,n})}{\Lambda(t_{i+1,n},t_{i,n})}\right)^2 - \frac{t-t_{i,n}}{t_{i+1,n}-t_{i,n}}\right|^2,
\end{aligned}$$

In addition, by Fact B.31 we have that

$$\left|\frac{\Lambda(t,t_{i,n})}{\Lambda(t_{i+1,n},t_{i,n})} - \frac{t-t_{i,n}}{t_{i+1,n}-t_{i,n}}\right| \leqslant C_1 \cdot \sup_{t,s\in[t_{i,n},t_{i+1,n}]} |\lambda(t)-\lambda(s)| \leqslant C_1 \cdot \bar{\omega}(\lambda, T/n), \quad (3.32)$$

for $t \in [t_{i,n}, t_{i+1,n}]$, $i = 0,1,\ldots,n-1$. By the Lemmas B.2 and B.28, and (3.32) we have that

$$\begin{aligned}
\mathbb{E}\big\|\bar{R}_n^M(t)\big\|^2 \ \leqslant\ & C_1 \cdot \big(\bar{\omega}(\lambda,T/n)\big)^2 \cdot (1+n^{-1}) \cdot n^{-1} \\
&+ C_2 \cdot n^{-2} \cdot (1+n^{-1}). \qquad (3.33)
\end{aligned}$$

Since, from (3.33) we have

$$\begin{aligned}
\Big|n^{1/2}\cdot e_n(\bar{X}^{Lin-M}) - n^{1/2}\cdot e_n(\bar{X}^{cM})\Big| &\leqslant n^{1/2}\cdot \big\|\bar{R}_n^M\big\|_{\mathfrak{L}^2(\Omega\times[0,T])} \\
&\leqslant C_1 \cdot \bar{\omega}(\lambda,T/n)\cdot \big(1+n^{-1}\big)^{1/2} \\
&\quad + C_2 \cdot n^{-1/2} \cdot \big(1+n^{-1}\big)^{1/2}.
\end{aligned}$$

We obtain the same asymptotic behavior for $\bar{X}^{Lin-M}$ like for the method $\bar{X}^{cM}$. So we have that

$$\limsup_{n\to+\infty} \big(cost_n(\bar{X}^{Lin-M})\big)^{1/2} \cdot e_n(\bar{X}^{Lin-M}) \leqslant (m_{\mathrm{w}}+1)^{1/2} \cdot C^{eq}_{\mathrm{MD}}.$$

This ends the proof of (3.27) in the case when $b \not\equiv 0$ and $c \not\equiv 0$.
If $b \not\equiv 0$ and $c \equiv 0$ then $cost_n(\bar{X}^{Lin-M}) = m_{\mathrm{w}} \cdot n$ which yield

$$\limsup_{n\to+\infty} \left(cost_n(\bar{X}^{Lin-M})\right)^{1/2} \cdot e_n(\bar{X}^{cM}) \leqslant m_{\mathrm{w}}^{1/2} \cdot C^{eq}_{\mathrm{MD}}.$$

If ($b \equiv 0$ and $c \not\equiv 0$) then $cost_n(\bar{X}^{Lin-M}) = n$, which yield

$$\limsup_{n\to+\infty} \left(cost_n(\bar{X}^{Lin-M})\right)^{1/2} \cdot e_n(\bar{X}^{Lin-M}) \leqslant C^{eq}_{\mathrm{MD}}.$$

This ends the proof. ■

From Theorems 3.1 and 3.2 we obtain the following results on the asymptotic performance of the methods $\bar{X}^{cM}$ and $\bar{X}^{Lin-M}$.

Theorem 3.3. *Let us assume that the functions a, b, c and λ satisfy the assumptions (A_{MD}) – (E_{MD}) and let $\bar{X} \in \left\{\bar{X}^{cM}, \bar{X}^{Lin-M}\right\}$. Then we have the following estimations.*

(i) If $b \not\equiv 0$ and $c \not\equiv 0$ then

$$\lim_{n\to+\infty} \left((m_{\mathrm{w}}+1)\cdot n\right)^{1/2} \cdot e_n(\bar{X}) = (m_{\mathrm{w}}+1)^{1/2} \cdot C^{\mathrm{eq}}_{\mathrm{MD}}.$$

(ii) If $b \not\equiv 0$ and $c \equiv 0$ then

$$\lim_{n\to+\infty} (m_{\mathrm{w}}\cdot n)^{1/2} \cdot e_n(\bar{X}) = (m_{\mathrm{w}})^{1/2} \cdot C^{\mathrm{eq}}_{\mathrm{MD}}.$$

(iii) If $b \equiv 0$ and $c \not\equiv 0$ then

$$\lim_{n\to+\infty} n^{1/2} \cdot e_n(\bar{X}) = C^{\mathrm{eq}}_{\mathrm{MD}}.$$

Finally, by Theorems 3.1 and 3.2 we get the asymptotic behavior of the nth minimal error in the class χ^{eq}.

Theorem 3.4. *We have that*

$$\lim_{n\to+\infty} n^{1/2} \cdot e^{\mathrm{eq}}(n) = C^{\mathrm{eq}}_{\mathrm{MD}},$$

and the methods $\bar{X}^{Lin-M}$ and $\bar{X}^{cM}$ are asymptotically optimal in the class χ^{eq}.

Chapter 4

Basics information about CUDA C programming language and numerical experiments

Approximation of solutions of SDEs requires simulations of many independent trajectories, what is computationally very demanding. Luckily, parallel computations are becoming more and more popular and the architecture of the Graphics Processor Unit (GPU) allows to significantly decrease the time of computations. That was the primary motivation to develop a CUDA C library. The library is dedicated to parallel simulations on GPUs of many independent trajectories of solutions of system stochastic differential equations with jumps. We named it by cuSTOCH. There is ongoing effort to develop and document a stable code version which can be released. Code and implementation are not the main subject of this thesis, but we mention it due to the fact that the results of simulations presented in Section 4.3 were obtained based on algorithms developed in CUDA C (then incorporated as a part of the library). In order to better understand the code shown in Section 4.2, the Section 4.1 has been developed. We show in Section 4.1 short introduction and basic tools of the CUDA C programming language.

4.1. An introduction to CUDA C programming language

In this section we present the programming tools of CUDA C, which are employed in the library and simulations. We also assume that the reader has a basic knowledge about programming, e.g. C/C++. For more details and specification we recommend [74] and documentation of CUDA C [47].

4.1.1. Basic notation and definitions

We start with introduction to CUDA C notations used in this work. We will consider CPU and GPU (central processor unit and graphic processor unit), which we can divide into two parts *host* and *device*. *Host* represents standard CPU together with dedicated CPU memory RAM. By *device* we mean GPU which consists of several parts. In the Figure 4.1 we show example of discussed architecture, where we have one CPU and one GPU. As we can see in Figure 4.1, GPU consists of multiple parts: various memory types and multiple processing units. Our description starts with CUDA application components as a kind of abstract structure, then we describe the processing unit. Finally, we describe types of memory.

As we simply describe CUDA C concept we should discuss three abstract structures which build programs which are run on GPU. There are *thread, block of threads (or block) and grid*. *Thread* is the smallest part of that structure. It is a single unit which performs given operations. Threads build a *block of threads* which is an independent copy of the kernel and is placed in the same stream processor. Locations in the same block give threads possibility to communicate with each other by a special part of memory. Multiply blocks are combined to bigger form in this structure, this form is named *grid*. All blocks in the same grid have the same number of threads. More details about definition of thread and block are presented in Section 4.1.3.

To understand the roots of the architecture, we present such a brief of history. Graphics processing unit (GPU) is a type of computer chip that rapidly performs specific mathematical calculations. The primary usage is for the purpose of rendering images. In the early days of computing, CPU was performing these calculations but when more graphics-intensive applications were developed, their demand put strain on the CPU and degraded performance. GPUs were developed as a way to offload those tasks from CPUs and free up processing power. Nowadays, graphics processors are being adapted to share the work of CPUs and for example solve computation

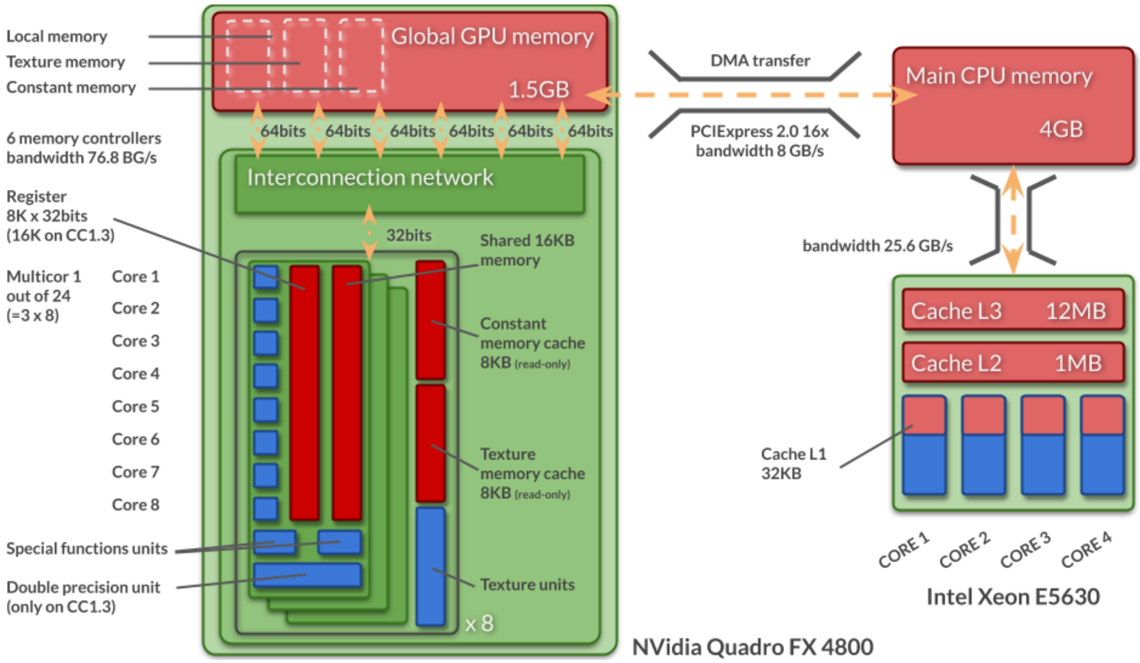


Figure 4.1: CPU + GPU structure.

consuming tasks like molecular chemistry simulations or training deep neural networks for AI applications. A GPU may be easily integrated with a CPU on the same circuit, on a graphics card or in the motherboard in both personal computer or server. The major players in the GPU market are NVIDIA, AMD, Intel, and ARM (recently announced to be acquired by NVIDIA).

The last part that should be explained is the memory structure. Graphic cards are equipped with several types of memory. Each of them are designed to perform different tasks. We summarize memory specification in Tables 4.1, 4.2 we show comparison of memory sizes. In Table 4.3 we present how to define variables in selected types of memory. We have the following memory types:

- **register memory**
  It is the fastest type of memory available on the graphic card. Register memory is used to store data by the threads. Stored data is accessible only by the thread which wrote it into memory, and is available only for the lifetime of that thread.
- **shared memory**
  This type of memory is more complex than register memory. Data stored there are accessible to all threads within a block of threads. Variables are available until the end of the block existence. This type of memory allows threads from one block of threads to communicate and to share data between each other.

- **constant memory**
  For the data that will not change over the lifetime of kernel execution we can use constant memory. Available for a grid during lifetime of application.
- **texture memory**
  It is one of the special types of memory. It is a variety of read-only memory on device which is used to reduce memory traffic in specific situations. Available for grid during lifetime of application.
- **local memory**
  Local memory has the same scope as registers, but the difference is that local memory is slower than register memory. It is the largest part of available global memory.
- **global memory**
  This type of memory is the most universal. Device as well as host processes can manage data stored here. Data is available during the lifetime of the host application. This type of memory takes the biggest part of whole memory.

Type	**Read/write**	**Scope**	**Lifetime**	**Speed**
Global	read/write	grid	application	slow, but cached
Texture	read only	grid	application	cache optimized for 2D/3D access pattern
Constant	read only	grid	application	where constants and kernel arguments are stored
Shared	read/write	block	block	fast
Local	read/write	thread	thread	used when it does not fit in to registers part of global memory slow but cached
Registers	read/write	thread	thread	fast

Table 4.1: Summary of properties of particular types of variables

Type	RTX 2080 Ti	Titan V	GTX 950M
Total global memory	10989 MB	12037 MB	2004 MB
Texture alignment	512 B	512 B	512 B
Total constant memory	64 KB	64 KB	64 KB
Total shared memory per block	48 KB	48 KB	48 KB
Total registers per block	64 KB	64 KB	64 KB

Table 4.2: Comparison of memory size

Type	Declaration
Global	__device__ int globalV
Texture	see [82]
Constant	__constant__ int constantV
Shared	__shared__ int sharedV
Local	int vArray[10]
Registers	int v

Table 4.3: Declaration of individual types of variables

4.1.2. Differences between C/C++ and CUDA C

Now we discuss a simple example, which shows us the main differences between C/C++ and CUDA C programming language.

```
1 #include <stdio.h>
2 #include <cstdio>
3 #include <cstdlib>
4 #include <iostream>
5
6
7 __host__ void CPUFunction(){
8   printf("This function is invoke on CPU and run on the CPU");
9 }
10
11
12 __device__ void GPUFunction(){
13   printf("This function is invoke on GPU and run on the GPU");
14 }
15
16
17 __global__ void kernelFunction(){
18   printf("This function is invoke globally and run on the GPU");
19   GPUFunction();
20 }
21
22
23 int main(){
24   CPUFunction();
25   kernelFunction<<<1,1>>>();
26   cudaDeviceSynchronize();
27 }
28
29 //RESULTS:
30 //
31 //  This function is invoke on CPU and run on the CPU
32 //  This function is invoke globally and run on the GPU
33 //  This function is invoke on GPU and run on the GPU
34 //
```

Listing 4.1.1: Comparision between C/C++ and CUDA C – main code and results

In the example we highlight fragments, which are especially important in the context of CUDA C programming language. Firstly, we can see that CUDA C requires qualifiers to the standard C/C++ function (lines 6, 10, 14 in Listing 4.1.1). There are three different qualifiers *__global__*, *__host__*, and *__device__*. It is a mechanism which informs the compiler to know where a function will be compiled to run. Now we describe each of them.

- **__host__ void CPUFunction()**
  Function marked with a qualifier __*host*__ can be invoked only by host functions and their executions is also on host. It means that the whole processes and declarations included in the function are done in a host context.
- **__device__ void GPUFunction()**
  The __*device*__ function can be invoked both by other __*device*__ and __*global*__ function. Execution of this function is purely on the device. It means that the whole processes and declarations included in the function are done in a device context.
- **__global__ void kernelFunction()**
  The __*global*__ function named also as *kernel* can be invoked (we say also that 'kernel is *launched*') by host code (for example, in main() function) but their execution is on device. It means that the host timeline launches the kernel function, but the whole calculations are performed on the device. The kernel function has access to __*device*__ functons but does not have access to __*host*__ functions.

Another modification is a special way of calling kernel function (line 21 in Listing 4.1.1)

- **kernelFunction**$<<< 1, 1 >>>>$**();**
  In this line we specify *execution configuration* parameters. Every time we launch a kernel we have to define the number of blocks as well as number of threads per block. This configuration looks like

  $$<<< numberOfBlocks,\ numberOfThreadsPerBlock >>>$$

  where these two parameters in one-dimensional case are of type **int**. This tool allows to define hierarchy of threads for the launched kernel. Configuration as well as threads hierarchy will be discussed in the Section 4.1.3 in more details.

The last novelty introduced in the example (line 22 in Listing 4.1.1) is a special function which comes from **run_time_api.h** library (see [47]) .

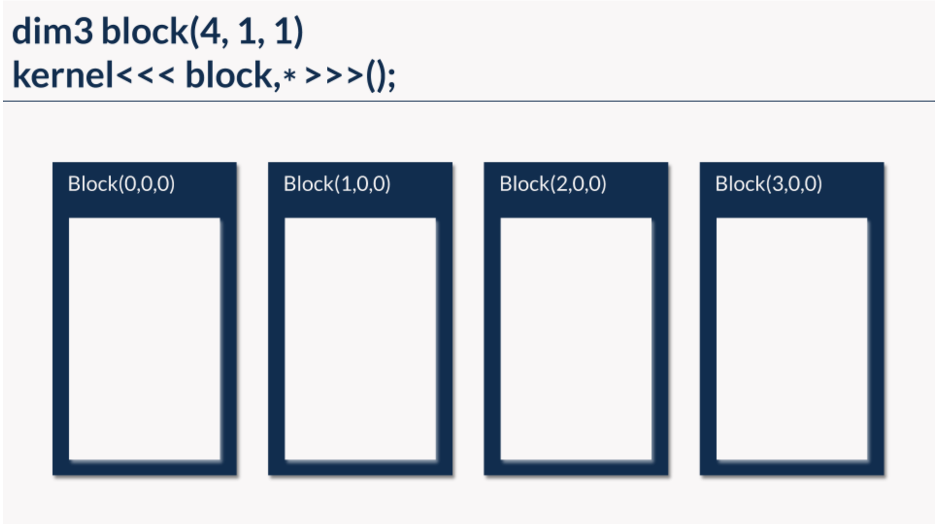
- **cudaDeviceSynchronize();**
  In a standard timeline generated by C/C++, kernels are launched as asynchronous processes. It means that the host code continues execution without waiting for the kernel to complete the tasks. It is needed to inform the compiler that the execution should wait for the kernel to be completed. We inform the compiler that it will wait with the function *cudaDeviceSynchronize()* used after the definition of the kernel. What is more, calling this function may occur after several independent kernels.

4.1.3. CUDA thread hierarchy

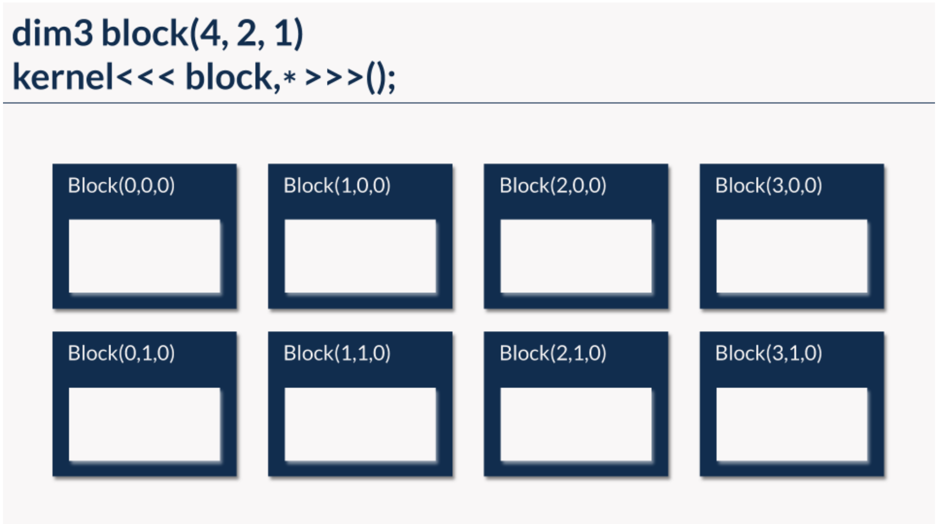
Definition of thread hierarchy can be specified by user and it should be matched to the problem. The grid and the thread blocks can be 1, 2 or 3-dimensional. The choice of dimension depends on the problem that we want to solve. For example, two-dimensional blocks of threads are used in matrix multiplication and three-dimensional blocks of thread are always used in graphical simulations, where the dimensions correspond to the definition of color representation, i.e. RGB (Red, Green, Blue). It is also important to know that a 1-dimensional case means that the size of the second and third dimension is equal to one. In our work we mostly use one-dimensional case so we show more details about this case.

We present different definitions of execution configuration and we illustrate how we can imagine these blocks and threads. To define size of a block and number of threads in each dimension we have to use structure which allows us to represent three-dimensional vector type. The most common structure to define the grid and block dimensions in a kernel invocation is type **dim3**. It is an integer vector type object that comes from **vector_types.h** (for more see [47]). In Figures 4.2 and 4.3 we show the structure of the generated grid and blocks. For transparency, we use white rectangle in Figure 4.2 to represent threads in blocks. We can replace it by one of the thread structures from Figure 4.3.

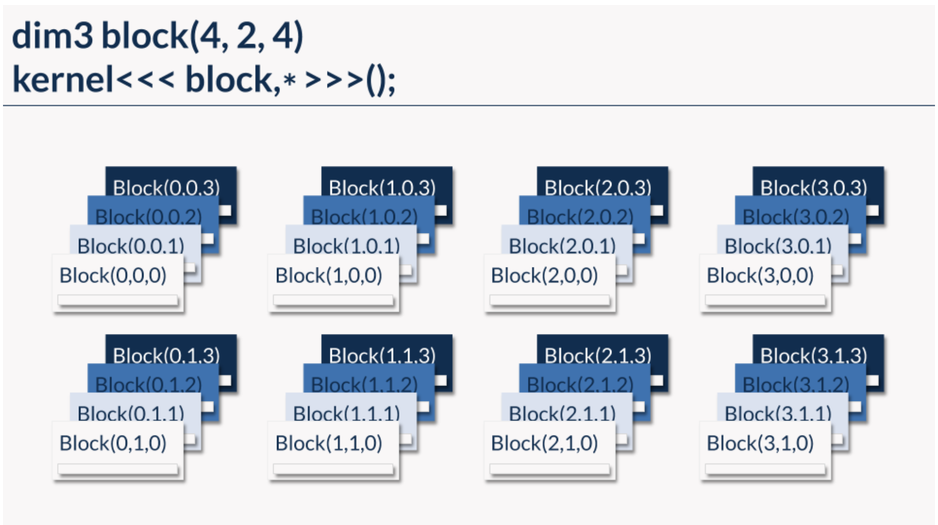
The structure of the launched kernel is not arbitrary. There are certain limits on the number of blocks and threads in each block. First limit follows from the architectures of available devices and should be less than $2^{16}-1$, and the number of threads in block is bounded by $2^{10}-1$. We calculate the size of the grid as well as blocks by multiplying individual dimensions. For example, we can run $(1023, 1, 1)$ blocks of threads, but $(1024, 1, 1)$ is not available.



(a) One-dimensional grid with size (4,1,1).

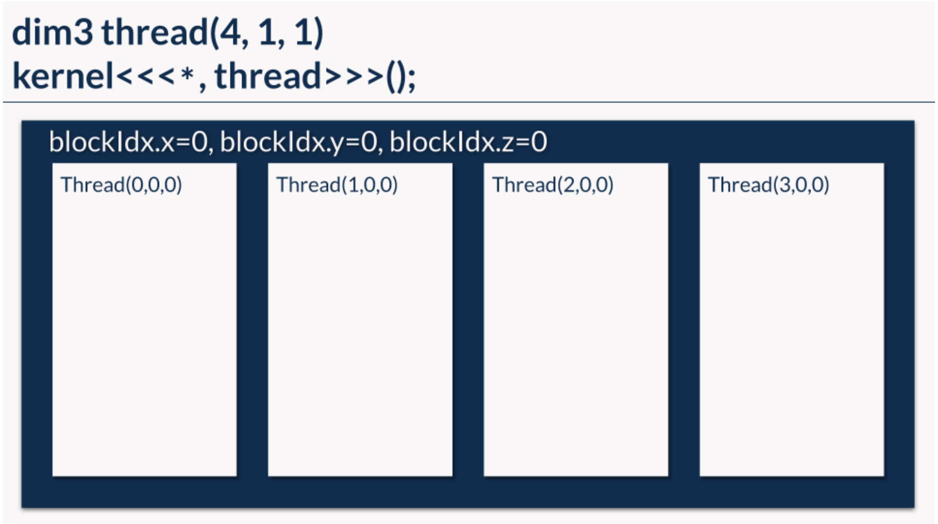


(b) Two-dimensional grid with size (4,2,1).

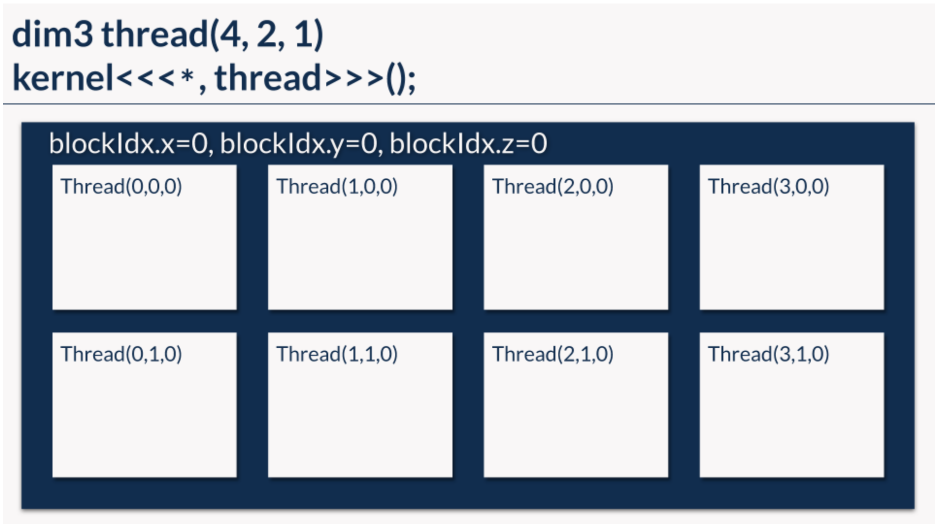


(c) Three-dimensional grid with size (4,2,4).

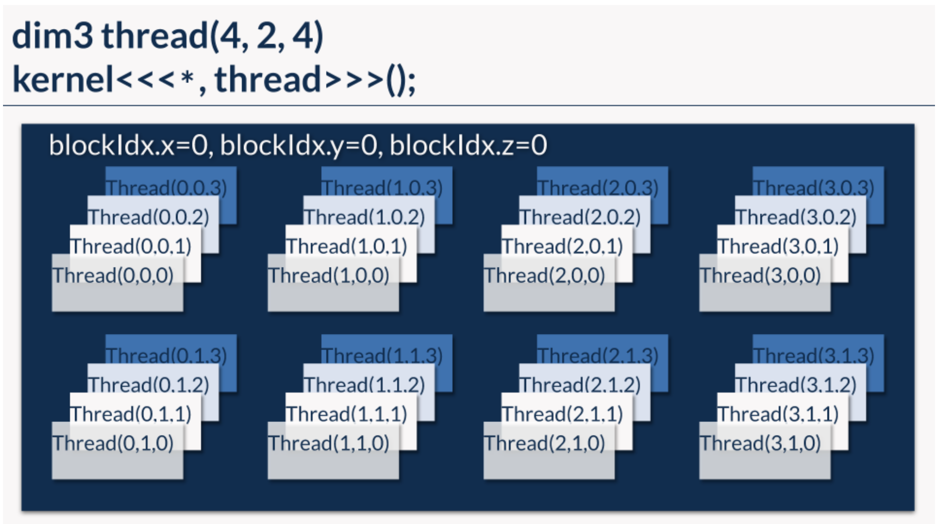
Figure 4.2: Example of grid visualization.



(a) One-dimensional thread with size (4,1,1).



(b) two-dimensional thread with size (4,2,1).



(c) Three-dimensional thread with size (4,2,4).

Figure 4.3: Example of threads visualization.

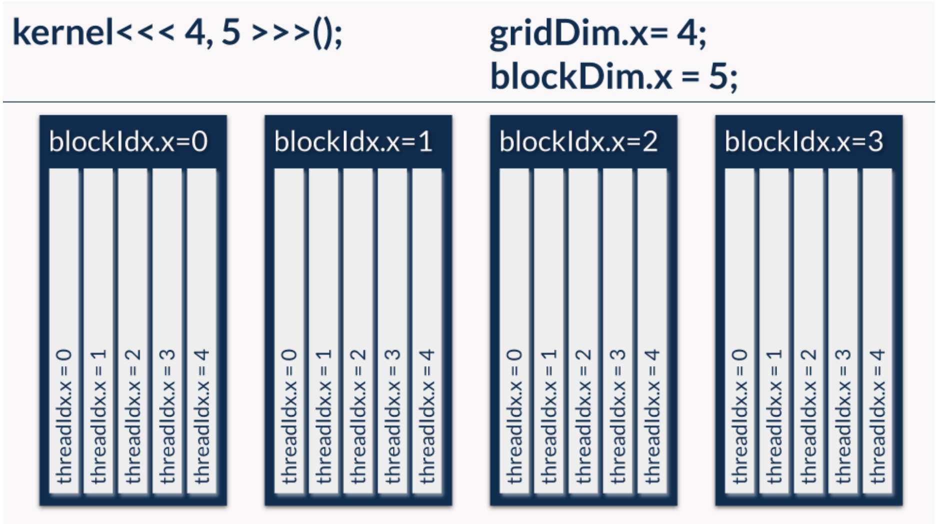


Figure 4.4: CUDA thread hierarchy identification of value of build in variables in grid.

We know how to define a thread hierarchy. Another important topic is managing the threads and blocks. CUDA technology gives us a built-in system which allows us to manage threads and blocks during writing the program code. When we launch a kernel in our program, this system generates built-in variables. We have:

- $gridDim.*$ – returns total number of block in $*$-axis
- $blockIdx.*$ – returns block ID in the $*$-axis of the block that is executing the given blocks of code,
- $blockDim.*$ – returns the *block's dimension* (i.e., the number of threads in a the block in the $*$-axis),
- $threadIdx.*$ – returns the thread ID in the $*$-axis of the thread that is being executed by the particular block,

where $* \in \{x, y, z\}$ is a dimension coordination. In the Figure 4.4 we present specification of these variables when we launch a kernel with configuration $<<< 4, 5 >>>$.

4.1.4. Management of parallel threads

Each thread from one-dimensional block with one-dimensional structure of threads can be identified by the following expression

$$tid = threadIdx.x + blockIdx.x \cdot blockDim.x. \tag{4.1}$$

So for the configuration from Figure 4.4 ($<<< 4, 5 >>>$) we have that the *tid*

- in block 0 has a number from range 0 – 4,
- in block 1 has a number from range 5 – 9,
- in block 2 has a number from range 10 – 14,
- in block 3 has a number from range 15 – 19.

In Listing 4.1.2 we present code where we have three functions. Based on it we show how to manage threads and blocks. We also present how to deal with the following three cases

1. when we use less threads than defined in the grid,
2. when we use the same number of threads as defined in the grid,
3. when we perform more tasks than threads which are defined in the grid.

```
#include <stdio.h>
#include <cstdio>
#include <cstdlib>
#include <iostream>

__global__ void identifyThread(){
  int tid = threadIdx.x + blockIdx.x * blockDim.x;
  printf("Block ID: %d, Thread ID %d, TID: %d\n",  threadIdx.x, blockIdx.x, tid);
}

__global__ void lessThanThread(int N){
  int tid = threadIdx.x + blockIdx.x * blockDim.x;
  if(tid < N){
    printf("Block ID: %d, Thread ID %d, TID: %d\n", threadIdx.x, blockIdx.x, tid);
  }
}

__global__ void moreThanThread(int N){
  int tid = threadIdx.x + blockIdx.x * blockDim.x;
  while(tid < N){
    printf("Block ID: %d, Thread ID %d, TID: %d\n", threadIdx.x, blockIdx.x, tid);
    tid += blockDim.x * gridDim.x;
  }
}

int main(){
  identifyThread<<<4,5>>>();
  cudaDeviceSynchronize();

  lessThanThread<<<4,5>>>(3);
  cudaDeviceSynchronize();

  lessThanThread<<<4,5>>>(40);
```

```
    cudaDeviceSynchronize();

    moreThanThread<<<4,5>>>(40);
    cudaDeviceSynchronize();
}
```

Listing 4.1.2: Identification and managing with possible scenarios – code.

In Listing 4.1.2 main function consists of evaluation of four functions, i.e.

1. $identifyThread <<< 4, 5 >>> ();$
   This function prints information about *blockId* and *threadId*, and calculates identification numbers based on the equation (4.1).
2. $lessThanThreadNumber <<< 4, 5 >>> (3);$
   The second function checks if the variable *tid* is less than a given parameter N in this case $N = 3$. When condition is true, the function displays identification parameters.
3. $lessThanThreadNumber <<< 4, 5 >>> (40);$
   Here the kernel was launched with configuration $<<< 4, 5 >>>$ which means that we generate 20 threads in a grid. As a parameter we put number 40. This means that each generated thread displays its message.

But what if we would like to run jobs with parameter larger than 20? Next function gives us a solution to this problem.

4. $moreThanThreadNumber <<< 4, 5 >>> (40);$
   In the last function from Listing 4.1.2 we have *'while'* instead of *'if'* condition. At the end of the while loop we increase the *tid* number by the total number of threads in blocks. This allows us to call as many tasks as we give as a function parameter in the case when we have less threads than the number of these tasks. We use this method to generate cycles of work length. Moreover, the number of operations that we want to perform, is not limited by the available number of threads. In the Figure 4.5 we present a simple example where each task performed by the same thread was colored by the thread's color.

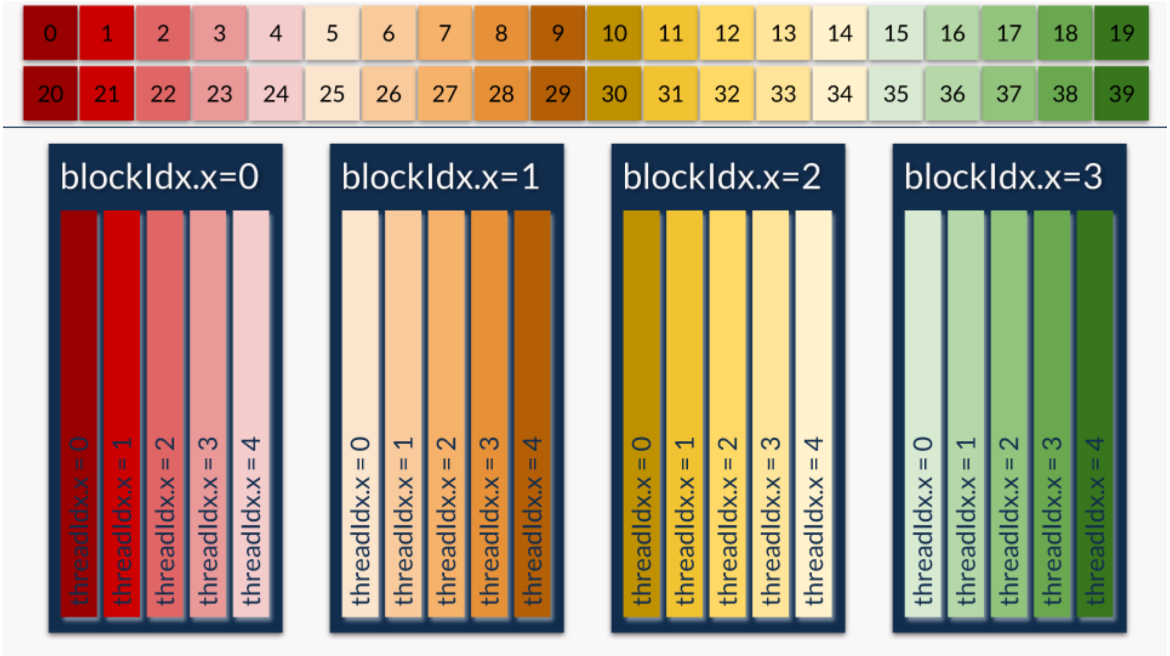


Figure 4.5: CUDA thread hierarchy identification of value of build in variables in grid and performed tasks.

4.1.5. Memory allocation

When we conduct processing on GPU, but the used variables are kept in host's memory, we spend much more time with connection between host and device. In C to allocate memory we apply *malloc*() function. In CUDA C we use dedicated functions *cudaMalloc()* and *cudaMallocManaged()*, which behave very similar to function *malloc()*.

- *cudaMalloc(void** devPtr, size_t size)*
  This function behaves similar to the standard C function *malloc()*. The first argument of function *void** devPtr* is a pointer to the allocated device memory. The size in bytes of the allocated memory is given as a second parameter. Variable type *size_t* is the unsigned integer type. The return type of this function is *cudaError_t* for more details see [47].
- *cudaMallocManaged(void** devPtr, size_t size, unsigned int flags = cudaMemAttachGlobal)*
  Allocates memory, that will be automatically managed by the Unified Memory system. Unified Memory is a single memory address space accessible from any processor in a system. This hardware/software technology allows applications to

allocate data that can be read or written from code running on either CPUs or GPUs. First and second parameters are the same as in the previous function. Parameter flags must be either *cudaMemAttachGlobal* or *cudaMemAttachHost*. The former allows any stream device to access memory. However, the latter does not allow any stream on any device to access memory. The return type as well as in the previous function is *cudaError_t*.

To support cudaError_t, we define a simple function *checkCuda()*. This function simply detects that the call has returned an error, and prints the associated error message.

When we would like to use a function *cudaMalloc*() it is important to know a function which allows us to copy our variables (part of memory) between different locations.

- *cudaMemcpy (void* dst, const void* src, size_t count, cudaMemcpyKind kind)* This function as a first parameter takes the memory address. Source memory address is given as the second parameter. As a third parameter we give size of memory to copy in bytes. As a last parameter we give type of transfer which can be one of directions
  — ***HostToDevice*** – when we want to copy from host to device.
  — ***DeviceToHost*** – when we want to copy from device to host.
  — ***DeviceToDevice*** – when we want to copy from one device to another device.
  — ***HostToHost*** – when we want to copy from one host to another host.
  This function also returns cudaError_t type which we can handle. For more specification see [47].

In Listing 4.1.3 we present a simple example where we show how to use the presented function to manage memory in the program.

```
#include <stdio.h>
#include <assert.h>

inline cudaError_t checkCuda(cudaError_t result){
  if (result != cudaSuccess) {
    fprintf(stderr, "CUDA Runtime Error: %s\n", cudaGetErrorString(result));
    assert(result == cudaSuccess);
  }
  return result;
}

__global__ void initArrayWithValue(float num, float *a, int N){
```

```
  int tid = threadIdx.x + blockIdx.x * blockDim.x;
  if(tid < N)
    a[tid] = num;
}

void displayArray(float *a, int N){
  for(int i = 0; i < N; ++i){
    printf("%f, ", a[i]);
  }
  printf("\n");
}


int main()
{
  // Number of bytes of an N - values vector
  const int N = 5;
  size_t size = N * sizeof(float);

  // Allocate memory for two vectors
  float *a;
  float *b;
  checkCuda(cudaMallocManaged(&a, size));
  checkCuda(cudaMalloc(&b, size));

  // Definition of block and thread size
  size_t threadsPerBlock = 256;
  size_t numberOfBlocks = (N + threadsPerBlock - 1) / threadsPerBlock;

  // Call GPU version of vector initalization
  initArrayWithValue<<<numberOfBlocks, threadsPerBlock>>>(3, a, N);
  initArrayWithValue<<<numberOfBlocks, threadsPerBlock>>>(4, b, N);

  // Wait for the GPU to finish before proceeding
  checkCuda(cudaGetLastError());
  checkCuda(cudaDeviceSynchronize());

  // Dispaly array which was allocated by cudaMallocManaged()
  displayArray(a, N);

  /*
   * // displayArray(b, N);
   *
   * This execution of displayArray(b,N) caused Memory error. It is
   * becouse of fact that host do not have acces to device memory
   * in this situation
   */

  // Dispaly array which was allocated by cudaMalloc()
  float *cpu_b;
```

```
  cpu_b = (float *)malloc(size);
  checkCuda(cudaMemcpy((void *) cpu_b, b, size , cudaMemcpyDeviceToHost));

  displayArray(cpu_b, N);

  // Wait for the GPU to finish before proceeding
  checkCuda(cudaGetLastError());
  checkCuda(cudaDeviceSynchronize());

  // Free all our allocated memory
  checkCuda(cudaFree(a));
  checkCuda(cudaFree(b));
  free(cpu_b);
}
//RESULTS:
//  3.000000, 3.000000, 3.000000, 3.000000, 3.000000,
//  4.000000, 4.000000, 4.000000, 4.000000, 4.000000,
```

Listing 4.1.3: Managing with memory between CPU and GPU

4.1.6. Examples from numerical linear algebra

To illustrate how to easily use CUDA C for known problems we present the following example of linear algebra. In Listing 4.1.4 we show the program code which is connected with matrix multiplication. Here large speed up can be observed. In this example we present code for both CPU and GPU implementation of algorithms. At the end of the program we also check performance in both cases.

```
#include <stdio.h>
#include <assert.h>

inline cudaError_t checkCuda(cudaError_t result){
  if (result != cudaSuccess) {
    fprintf(stderr, "CUDA Runtime Error: %s\n", cudaGetErrorString(result));
    assert(result == cudaSuccess);
  }
  return result;
}

__global__ void matrixMulGPU(int * a, int * b, int * c , int N){
  int val = 0;
  int row = blockIdx.x * blockDim.x + threadIdx.x;
  int col = blockIdx.y * blockDim.y + threadIdx.y;
  if(row < N && col < N){
    for (int k = 0; k < N; k++){
      val += a[row * N + k] * b[k * N + col];
    }
```

```
      c[row * N + col] = val;
    }
}

void matrixMulCPU(int * a, int * b, int * c, int N){
  int val = 0 ;
  for(int row = 0; row < N; row++){
    for(int col = 0; col < N; col++){
      val = 0;
      for(int k = 0; k < N; ++k) {
        val += a[row * N + k] * b[k * N + col];
      }
      c[row * N + col] = val;
    }
  }
}

int main() {
  // Initialize pointer to array
  int *a, *b, *c_cpu, *c_gpu;

  // Number of bytes of an N x N matrix
  int N = 64;
  int size = N * N * sizeof (int);

  // Allocate memory
  checkCuda(cudaMallocManaged(&a, size));
  checkCuda(cudaMallocManaged(&b, size));
  checkCuda(cudaMallocManaged(&c_cpu, size));
  checkCuda(cudaMallocManaged(&c_gpu, size));

  // Initialize both matrix with values and zeros for results matrices
  for(int row = 0; row < N; row++){
    for(int col = 0; col < N; col++){
      a[row * N + col] = row;
      b[row * N + col] = col + 2;
      c_cpu[row * N + col] = 0;
      c_gpu[row * N + col] = 0;
    }
  }

  // Definition of block and thread size
  dim3 threads_per_block(16, 16, 1); // A 16 x 16 block threads
  dim3 number_of_blocks((N /threads_per_block.x) + 1, (N /threads_per_block.y) + 1,1);

  // Call GPU version of matrix multiplication
  matrixMulGPU<<<number_of_blocks, threads_per_block>>>(a, b, c_gpu, N);

  // Wait for the GPU to finish before proceeding
  checkCuda(cudaGetLastError());
```

```
  checkCuda(cudaDeviceSynchronize());


  // Call the CPU version to check our work
  matrixMulCPU(a, b, c_cpu, N);

  // Compare the two answers to make sure they are equal
  bool error = false;
  for(int row = 0; row < N && !error; ++row){
    for(int col = 0; col < N && !error; ++col){
      if (c_cpu[row * N + col] != c_gpu[row * N + col]){
        printf("FOUND ERROR at c[%d][%d]\n", row, col);
        error = true;
        break;
      }
    }
  }

  if (!error) {
    printf("SUCCESS! Matrix are multiplied correctly.\n");
  }

  // Free all our allocated memory
  checkCuda(cudaFree(a));
  checkCuda(cudaFree(b));
  checkCuda(cudaFree(c_cpu));
  checkCuda(cudaFree(c_gpu));
}
```

Listing 4.1.4: Matrix multiplication

4.2. Implementation of algorithm $\bar{X}_{k_n}^{Lin-M}$ in CUDA C

In this section we present the full code of one of the algorithms considered in Chapter 2. We divide full code into smallest part. It is because beter understanding of problem.

Definition of problem Specification

```
double X0 = 0.1;
__device__ double X000 = 0.1;
__device__ double MI = 0.5;
__device__ double SIGMA = 1.0;
__device__ double PC = 1;
__device__ double LAMBDA = 1;

__device__ double a(double t, double x){ return MI * x; }

__device__ double b(double t, double x){ return SIGMA * x; }

__device__ double c(double t, double x){ return PC * x; }

__device__ double a_(double t, double x){ return MI; }

__device__ double b_(double t, double x){ return SIGMA; }

__device__ double c_(double t, double x){ return PC; }

__device__ double L1b(double t, double x){ return b(t,x) * b_(t,x); }

__device__ double L1c(double t, double x){ return b(t,x) * c_(t,x); }

__device__ double L_1b(double t, double x){ return b(t,x + c(t,x)) - b(t,x); }

__device__ double L_1c(double t, double x){ return c(t,x + c(t,x)) - c(t,x); }

__device__ double lambda(double t){ return LAMBDA * t; }

__device__ double intLambda(double t_1, double t_2){ return LAMBDA * (t_2 - t_1); }
```

Listing 4.2.1: Definition of problem Specification.

Additional Functions

```
inline cudaError_t checkCuda(cudaError_t result) {
    if(result != cudaSuccess) {
      fprintf(stderr, "CUDA Runtime Error: %s\n", cudaGetErrorString(result));
      assert(result == cudaSuccess) ;
    }
    return result;
}

__global__ void initState(float seed, curandState_t* states, int size) {
    int tid = blockIdx.x * blockDim.x + threadIdx.x;
    while(tid < size){
        curand_init(seed, tid, 0, &states[tid]);
```

```
        tid += blockDim.x * gridDim.x;
    }
}


void compensate(double *App, double results, int size){
    results=0;
    for(int i=0; i < size; i++){
        results += App[i];
    }
}

void saveToFileArray(ofstream &o, double t, double *array, int size){
    o << t << ";";
    for(int i = 0; i < size; i++){
        o << array[i] << ";";
    }
    o << "\n";
}
```

Listing 4.2.2: Additional Functions.

Main Algorithm

```
__device__ double milsteinCommutative(double t, double x, double dt,
                                       double dw, double dn) {
    double res = x + a(t,x) * dt + b(t,x) * dw + c(t,x) * dn +
                 L1b(t,x) * (dw * dw - dt) / 2 + L_1c(t,x) * (dn * (dn - 1)) / 2 +
                 L_1b(t,x) * dw * dn;
    return res;
}

__global__ void calculateApproximation(double t, double t_prev,
                                        curandState_t* states_normal,
                                        curandState_t* states_poisson,
                                        double *Xalg, int size){
    int tid = blockIdx.x * blockDim.x + threadIdx.x;
    register double step = (double) (t - t_prev);
    while(tid < size){
        double DW;
        int DN;
        DW = sqrt(step) * curand_normal_double(&states_normal[tid]);
        DN = curand_poisson(&states_poisson[tid],intLambda(t_prev, t));
        Xalg[tid] = milsteinCommutative(t_prev, Xalg[tid], step, DW, DN);
        tid += blockDim.x * gridDim.x;
    }
}

```

```
__global__ void calculateApp(int n, double t, double *gpuXLinM, double *gpuApp,
                             int size){
    int tid = blockIdx.x * blockDim.x + threadIdx.x;
    while(tid < size) {
        gpuApp[tid] = b(t, gpuXLinM[tid]) * b(t, gpuXLinM[tid]) +
                      lambda(t) * c(t, gpuXLinM[tid]) * c(t, gpuXLinM[tid]);
        tid += blockDim.x * gridDim.x;
    }
}

double calculateNextT(double eps, double app, double T, int n){
    if (eps < app)
        return T / (n * app);
    else
        return T / (n * eps);
}

void oneDimMilsteinStepSize(int numberOfTrajectories, int numberOfSteps,
                  double cModule, double T, string fileName){
    size_t threadsPerBlock = 256;
    size_t numberOfBlocks = (numberOfTrajectories + threadsPerBlock - 1)
                            / threadsPerBlock;
    double eps = pow((double) numberOfSteps, cModule);

    // allocate space on the GPU for the wienner random states
    curandState_t* states_normal;
    curandState_t* states_poisson;
    checkCuda(cudaMallocManaged((void**) &states_normal,
                              numberOfTrajectories * sizeof(curandState_t)));
    checkCuda(cudaMallocManaged((void**) &states_poisson,
                              numberOfTrajectories * sizeof(curandState_t)));

    // initiate states for both proceses
    initState<<<numberOfBlocks, threadsPerBlock>>>(time(NULL), states_poisson,
                                                  numberOfTrajectories);
    checkCuda(cudaDeviceSynchronize());
    initState<<<numberOfBlocks, threadsPerBlock>>>(time(NULL), states_normal,
                                                  numberOfTrajectories);
    checkCuda(cudaDeviceSynchronize());

    /* allocate space on the GPU for all needed lists */
    double *App;
    double *XLinM;
    checkCuda(cudaMallocManaged((void**) &App, numberOfTrajectories*sizeof(double)));
    checkCuda(cudaMallocManaged((void**) &XLinM,numberOfTrajectories*sizeof(double)));

    double *t;
    double *t_prev;
    double *results;
    checkCuda(cudaMallocManaged((void**) &t, sizeof(double)));
```

```
    checkCuda(cudaMallocManaged((void**) &t_prev, sizeof(double)));
    checkCuda(cudaMallocManaged((void**) &results, sizeof(double))) ;

    int kn = 0;

    // set all starting values for elements in arrays
    for(int i = 0; i < numberOfTrajectories; i++){
        XLinM[i] = X0;
    }
    memset(t, 0, sizeof(double));
    memset(t_prev, 0, sizeof(double));
    checkCuda(cudaDeviceSynchronize());

    //open file to save results and save first point
    ofstream plik;
    plik.open(fileName);


    //main part of algorithm
    while (*t < T){
        saveToFileArray(plik, *t, XLinM, numberOfTrajectories);
        kn += 1;
        *t_prev = *t;
        calculateApp<<<numberOfBlocks, threadsPerBlock>>>(numberOfSteps, *t, XLinM,
                                                    App, numberOfTrajectories);
        checkCuda(cudaGetLastError());
        checkCuda(cudaDeviceSynchronize());
        compensate(App, *results, numberOfTrajectories);
        *t += calculateNextT(eps, *results, T, numberOfSteps);
        calculateApproximation<<<numberOfBlocks, threadsPerBlock>>>(*t, *t_prev,
                                                            states_normal,
                                                            states_poisson,
                                                            XLinM,
                                                        numberOfTrajectories);
        checkCuda(cudaGetLastError());
        checkCuda(cudaDeviceSynchronize());
    }
    *t = T;
    calculateApproximation<<<numberOfBlocks, threadsPerBlock>>>(*t, *t_prev,
                                                        states_normal,
                                                        states_poisson, XLinM,
                                                        numberOfTrajectories);
    checkCuda(cudaGetLastError());
    checkCuda(cudaDeviceSynchronize());
    saveToFileArray(plik, *t, XLinM, numberOfTrajectories);

    // free allocated memory
    cudaFree(states_normal);
    cudaFree(states_poisson);
    cudaFree(App);
```

```
    cudaFree(XLinM);
    cudaFree(t);
    cudaFree(t_prev);
    cudaFree(results);
}
```

Listing 4.2.3: Main Algorithm.

4.3. Numerical experiments

4.3.1. Problems

Scalar Problem

First, let us consider the following linear scalar SDE used in the Merton's model

$$\begin{cases} \mathrm{d}X(t) = \mu X(t)\mathrm{d}t + \sigma X(t)\mathrm{d}W(t) + cX(t-)\mathrm{d}N(t), \quad t \in [0,T], \\ X(0) = x_0, \end{cases} \tag{4.2}$$

where $c > -1$ and $x_0\sigma > 0$, $\mu \in \mathbb{R}$. The exact solution of problem (4.2) has the following form

$$X(t) = x_0 \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right) \cdot (1+c)^{N(t)}.$$

We have $\mathbb{E}(\mathcal{Y}(t)) = x_0^2\,(\sigma^2 + c^2 \cdot \lambda(t)) \exp\Big(2(\mu + \sigma^2/2)t + c(c+2)m(t)\Big)$ for $t \in [0,T]$.

Multidimensional Problem

For multidimensional problem we consider the case when

$$\begin{cases} \mathrm{d}X(t) = \mu X(t)\mathrm{d}t + \begin{pmatrix} \sigma^{1,1}X_1(t) & \sigma^{1,2}X_1(t) & \cdots & \sigma^{1,m_{\mathrm{w}}}X_1(t) \\ \sigma^{2,1}X_2(t) & \sigma^{2,2}X_2(t) & \cdots & \sigma^{2,m_{\mathrm{w}}}X_2(t) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{d,1}X_d(t) & \sigma^{d,2}X_d(t) & \cdots & \sigma^{d,m_{\mathrm{w}}}X_d(t) \end{pmatrix} \mathrm{d}W(t) + cX(t-)\mathrm{d}N(t), \\ X(0) = x_0, \quad t \in [0,T] \end{cases} \tag{4.3}$$

where $c > -1$ and $\sigma^{i,j} > 0$ for $i \in \{1,\dots,d\}$, $j \in \{1,\dots,m_{\mathrm{w}}\}$, $\mu \in \mathbb{R}$, $x_0 \in \mathbb{R}^d_+$.

The exact solution of problem (4.3) has the following form

$$X_i(t) = X_i(0) \cdot \exp\left(\left(\mu - \frac{1}{2}\Big(\sum_{j=1}^{m_{\mathrm{w}}} \sigma^{ij}\Big)^2\right)t + \sum_{j=1}^{m_{\mathrm{w}}} \sigma^{ij} W_j(t)\right) \cdot (1+c)^{N(t)}.$$

By the considerations presented at page 40 the problem (4.3) satisfies jump commutative conditions (D_{MD}). On the Figure 4.6 we show sample of approximations of trajectories for solutions of problem (4.3) in one and two-dimensional case and on Figure 4.7 we present three dimensional case.

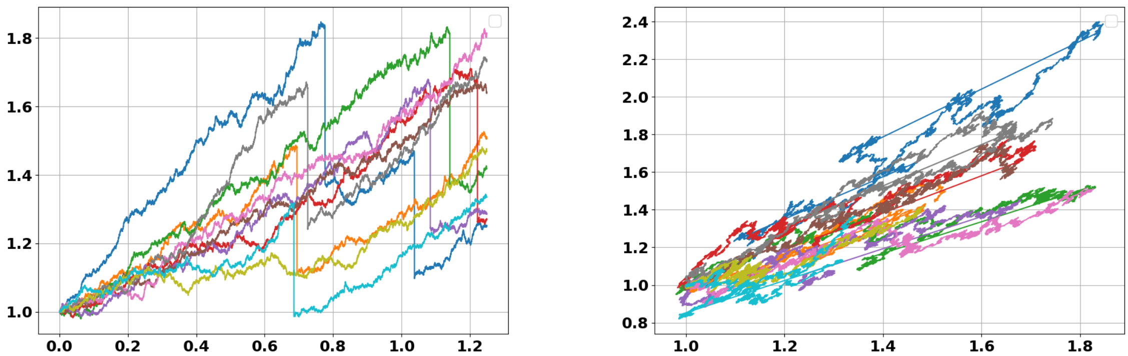


Figure 4.6: Examples of SDEs trajectories.

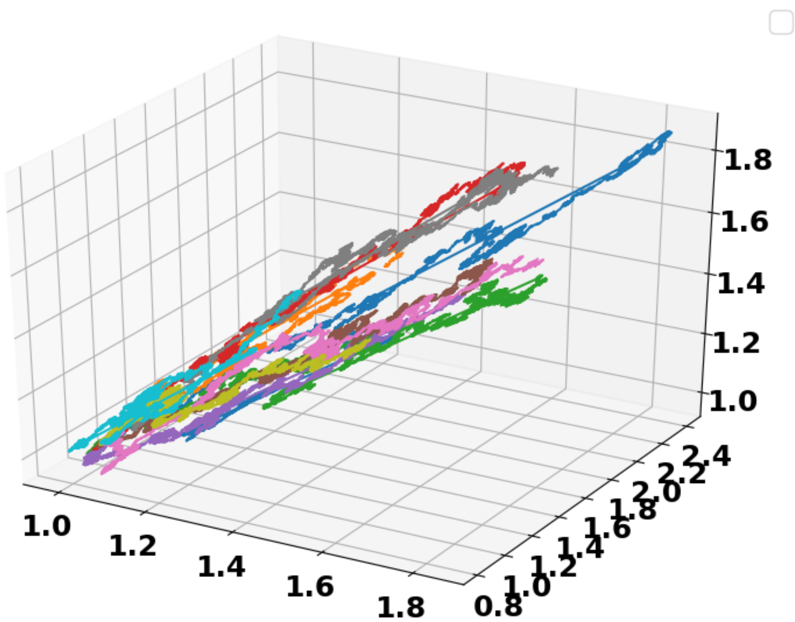


Figure 4.7: Examples of SDEs trajectories.

4.3.2. Error criterion

Now we define the way how we calculate error. We take as an estimator of the error of $\|X - \bar{X}_n\|_{\mathfrak{L}^2(\Omega\times[0,T])}$, for one-dimensional case

$$\varepsilon_K(\bar{X}_{k_n}) = \left(\frac{1}{K}\sum_{j=1}^{K} Q_n\big(|X_{(j)} - \bar{X}_{j,k_n}|^2\big)\right)^{1/2},$$

where $\bar{X}_{k_n} \in \big\{\bar{X}_{k_n}^{Lin-M*}, \bar{X}_{k_n}^{Lin-M-eq}, \bar{X}_{k_n}^{df-Lin-M*}, \bar{X}_{k_n}^{df-Lin-M-eq}\big\}$. For multidimensional case we use the same estimations, but only for method $\bar{X}_n^{Lin-M-eq}$

$$\varepsilon_K(\bar{X}_n^{Lin-M-eq}) = \left(\frac{1}{K}\sum_{j=1}^{K} Q_n\big(\|X_{(j)} - \bar{X}_{(j),n}^{Lin-M-eq}\|^2\big)\right)^{1/2}.$$

In both methods of estimations Q_n is the composite Simpson quadrature based on the knots $\big\{\hat{t}_{0,n}^*, \hat{t}_{1,n}^*, \ldots, \hat{t}_{k_n,n}^*\big\} \cup \big\{(\hat{t}_{i,n}^* + \hat{t}_{i+1,n}^*)/2\big\}_{i=0,1,\ldots,k_n-1}$ for a one-dimensional case. For multidimensional case, when we consider only equidistant mesh we use $\big\{t_{0,n}, t_{1,n}, \ldots, t_{n,n}\big\} \cup \big\{(t_{i,n} + t_{i+1,n})/2\big\}_{i=0,1,\ldots,n-1}$. We assume that $X_{(j)}$ is the jth (simulated) trajectory of the solution both problems (4.2) and (4.3) and $\bar{X}_n^{Lin-M-eq}$ is the piecewise linear interpolation of the classical Milstein steps performed at the equidistant discretization $t_{i,n}^{\mathrm{eq}} = iT/n$, for $i = 0, 1, \ldots, n$ for jth trajectories. (Hence, we use the same number of steps for $\bar{X}_{k_n}^{Lin-M*}$ and $\bar{X}_{k_n}^{Lin-M-eq}$.) In one-dimensional case we also compare the error of the method $\bar{X}_{k_n}^{Lin-M*}$ with the error of $\bar{X}_{k_n}^{Lin-M-eq}$ performed at equidistant points $t_i = iT/k_n$, $i = 0, 1, \ldots, k_n$. The improvement, observed in the numerical experiments, is defined by

$$\mathrm{imp}_{K,k_n} = \varepsilon_K(\bar{X}_{k_n}^{Lin-M*})/\varepsilon_K(\bar{X}_{k_n}^{Lin-M-eq}), \tag{4.4}$$

$$\mathrm{imp}_{K,k_n}^{df} = \varepsilon_K(\bar{X}_{k_n}^{df-Lin-M*})/\varepsilon_K(\bar{X}_{k_n}^{df-Lin-M-eq}). \tag{4.5}$$

4.3.3. Results of numerical experiments

Numerical experiments for method $\bar{X}_{k_n}^{Lin-M*}$

We have performed numerical experiments for the following particular cases of (4.2) for regular method.

1. $\mu = 0.08,\ \sigma = 0.4,\ c = -0.03,\ \lambda(t) = 2,\ x_0 = 5,\ T = 3,\ K = 60000.$

n	k_n	$\varepsilon_K(\bar{X}_{k_n}^{Lin-M*})$	$\varepsilon_K(\bar{X}_{k_n}^{Lin-M-eq})$	imp_{K,k_n}
3	7	2.37	2.36	1.004 42
10	23	6.30×10^{-1}	6.34×10^{-1}	$9.939\,68 \times 10^{-1}$
32	70	1.87×10^{-1}	1.88×10^{-1}	$9.962\,56 \times 10^{-1}$
102	230	5.50×10^{-2}	5.70×10^{-2}	$9.647\,49 \times 10^{-1}$
326	757	1.67×10^{-2}	1.75×10^{-2}	$9.543\,49 \times 10^{-1}$
1043	2433	5.15×10^{-3}	5.31×10^{-3}	$9.698\,94 \times 10^{-1}$
3338	7842	1.61×10^{-3}	1.63×10^{-3}	$9.896\,44 \times 10^{-1}$
10682	24871	5.07×10^{-4}	5.61×10^{-4}	$9.033\,90 \times 10^{-1}$
34182	87332	1.70×10^{-4}	1.80×10^{-4}	$9.459\,98 \times 10^{-1}$

Table 4.4: Results of calculated error and improvement.

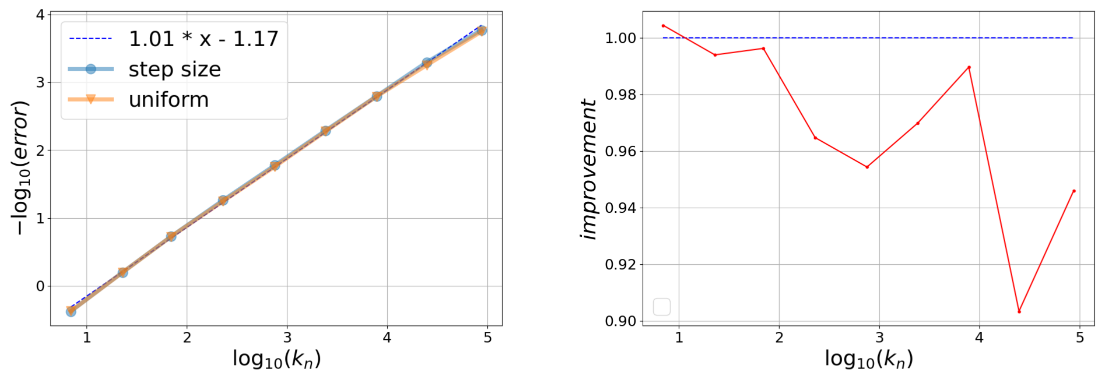
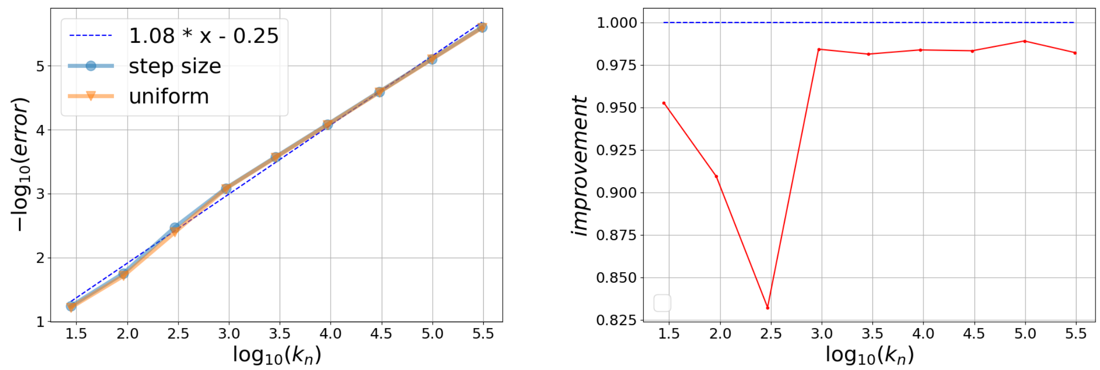


Figure 4.8: Left figure – comparison with theoretical rate of convergence, right – figure improvement calculated by (4.4).

2. $\mu = 1$, $\sigma = 0.7$, $c = 1$, $\lambda(t) = 0.1$, $x_0 = 10$, $T = 0.25$, $K = 60000$.

n	k_n	$\varepsilon_K(\bar{X}_{k_n}^{Lin-M*})$	$\varepsilon_K(\bar{X}_{k_n}^{Lin-M-eq})$	imp_{K,k_n}
3	28	5.77×10^{-2}	6.06×10^{-2}	9.52741×10^{-1}
10	92	1.77×10^{-2}	1.94×10^{-2}	9.09378×10^{-1}
32	294	3.36×10^{-3}	4.04×10^{-3}	8.32343×10^{-1}
102	932	8.26×10^{-4}	8.40×10^{-4}	9.84158×10^{-1}
326	2877	2.67×10^{-4}	2.72×10^{-4}	9.81255×10^{-1}
1043	9253	8.33×10^{-5}	8.47×10^{-5}	9.83717×10^{-1}
3338	30287	2.54×10^{-5}	2.58×10^{-5}	9.83227×10^{-1}
10682	98321	7.85×10^{-6}	7.94×10^{-6}	9.88980×10^{-1}
34182	307098	2.50×10^{-6}	2.55×10^{-6}	9.82182×10^{-1}

Table 4.5: Results of calculated error and improvement.



(a)

Figure 4.9: Left figure – comparison with theoretical rate of convergence, right figure – improvement calculated by (4.4).

3. $\mu = 10,\ \sigma = 3,\ c = -0.9,\ \lambda(t) = 1.5,\ x_0 = 5.75,\ T = 0.25,\ K = 60000.$

n	k_n	$\varepsilon_K(\bar{X}^{Lin-M*}_{k_n})$	$\varepsilon_K(\bar{X}^{Lin-M-eq}_{k_n})$	imp_{K,k_n}
3	434	5.88×10^{1}	4.90×10^{1}	$1.199\,12$
10	1620	6.76×10^{1}	1.22×10^{1}	$5.525\,69$
32	4976	1.99	4.79	$4.168\,43 \times 10^{-1}$
102	19780	3.32×10^{-1}	2.32	$1.431\,87 \times 10^{-1}$
326	47558	1.05×10^{-1}	1.64×10^{-1}	$6.382\,89 \times 10^{-1}$
1043	169508	1.68×10^{-2}	5.02×10^{-2}	$3.350\,85 \times 10^{-1}$
3338	1065420	4.08×10^{-3}	7.41×10^{-3}	$5.508\,01 \times 10^{-1}$
10682	1405330	2.88×10^{-3}	2.05×10^{-2}	$1.405\,99 \times 10^{-1}$

Table 4.6: Results of calculated error and improvement.

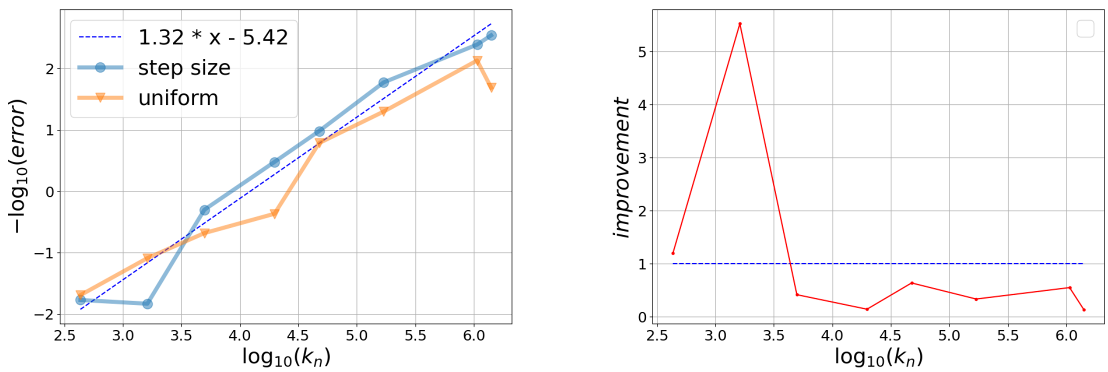
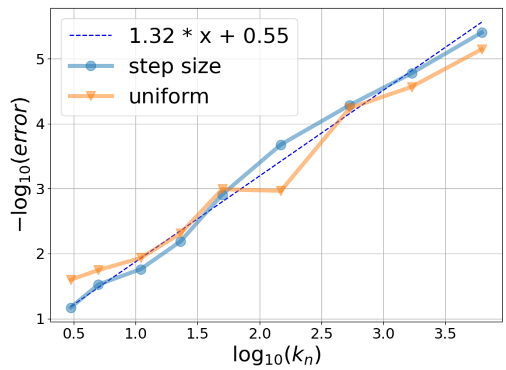


Figure 4.10: Left figure – comparison with theoretical rate of convergence, right figure – improvement calculated by (4.4).

4. $\mu = -1,\ \sigma = 1.5,\ c = 0,\ \lambda(t) = 0,\ x_0 = 0.1,\ T = 1,\ K = 60000.$

n	k_n	$\varepsilon_K(\bar{X}_{k_n}^{Lin-M*})$	$\varepsilon_K(\bar{X}_{k_n}^{Lin-M-eq})$	imp_{K,k_n}
3	3	6.76×10^{-2}	2.52×10^{-2}	2.675 37
10	5	3.00×10^{-2}	1.80×10^{-2}	1.668 50
32	11	1.73×10^{-2}	1.17×10^{-2}	1.485 60
102	23	6.50×10^{-3}	4.91×10^{-3}	1.324 75
326	50	1.23×10^{-3}	1.02×10^{-3}	1.206 75
1043	149	2.10×10^{-4}	1.08×10^{-3}	$1.935\,79 \times 10^{-1}$
3338	536	5.17×10^{-5}	5.76×10^{-5}	$8.969\,19 \times 10^{-1}$
10682	1696	1.67×10^{-5}	2.70×10^{-5}	$6.176\,00 \times 10^{-1}$
34182	6239	3.94×10^{-6}	7.13×10^{-6}	$5.524\,70 \times 10^{-1}$

Table 4.7: Results of calculated error and improvement.



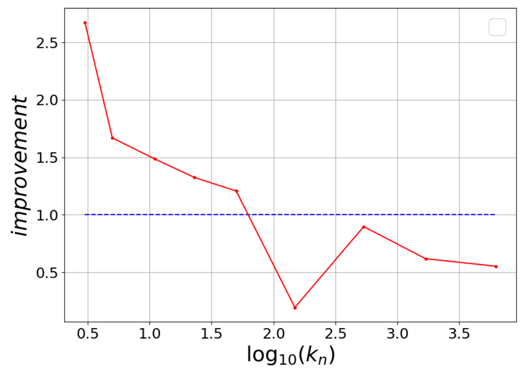


Figure 4.11: Left figure – comparison with theoretical rate of convergence, right figure – improvement calculated by (4.4).

Numerical experiments for method $\bar{X}^{df-Lin-M*}_{k_n}$

1. $\mu = 0.08$, $\sigma = 0.4$, $c = -0.03$, $\lambda(t) = 2$, $x_0 = 1$, $T = 5$, $K = 60000$.

n	k_n	$\varepsilon_K(\bar{X}^{df-Lin-M*}_{k_n})$	$\varepsilon_K(\bar{X}^{df-Lin-M-eq}_{k_n})$	imp^{df}_{K,k_n}
3	3	2.53	1.79	1.410 49
10	6	8.25×10^{-1}	8.04×10^{-1}	1.025 82
32	17	2.13×10^{-1}	2.53×10^{-1}	$8.396\,74 \times 10^{-1}$
102	52	5.07×10^{-2}	5.55×10^{-2}	$9.145\,76 \times 10^{-1}$
326	169	1.69×10^{-2}	1.76×10^{-2}	$9.591\,50 \times 10^{-1}$
1043	568	4.11×10^{-3}	4.61×10^{-3}	$8.929\,29 \times 10^{-1}$
3338	1676	1.42×10^{-3}	1.73×10^{-3}	$8.247\,52 \times 10^{-1}$
10682	5902	4.05×10^{-4}	4.45×10^{-4}	$9.099\,42 \times 10^{-1}$
34182	17044	1.41×10^{-4}	1.60×10^{-4}	$8.790\,91 \times 10^{-1}$

Table 4.8: Results of calculated error and improvement.

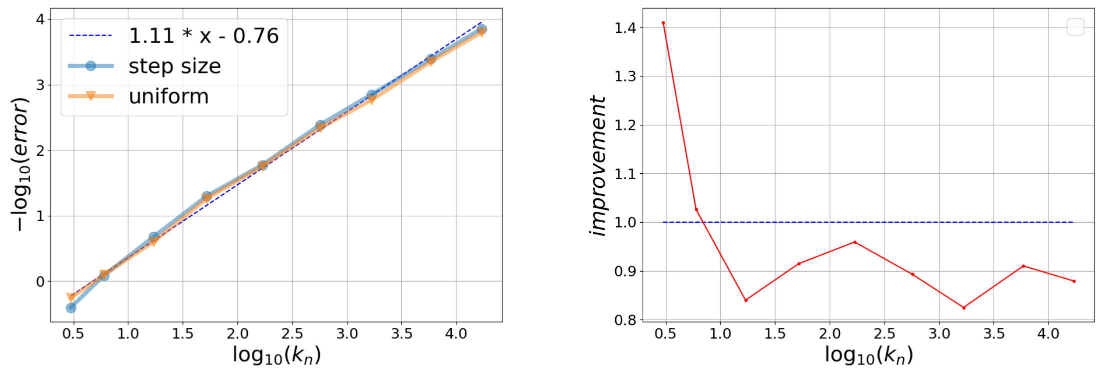
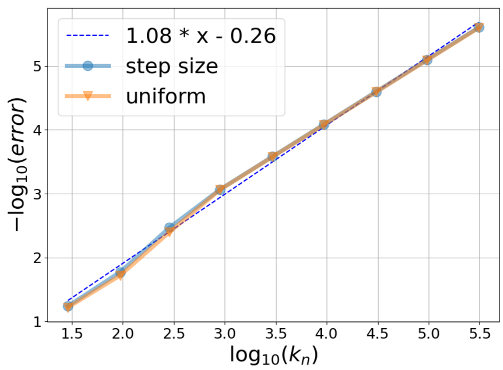


Figure 4.12: Left figure – comparison with theoretical rate of convergence, right figure – improvement calculated by (4.5).

2. $\mu = 1,\ \sigma = 0.7,\ c = 1,\ \lambda(t) = 0.1,\ x_0 = 10,\ T = 0.25,\ K = 60000.$

n	k_n	$\varepsilon_K(\bar{X}^{df-Lin-M*}_{k_n})$	$\varepsilon_K(\bar{X}^{df-Lin-M-eq}_{k_n})$	$\mathrm{imp}^{df}_{K,k_n}$
3	29	5.76×10^{-2}	6.05×10^{-2}	$9.522\,42 \times 10^{-1}$
10	94	1.69×10^{-2}	1.92×10^{-2}	$8.820\,98 \times 10^{-1}$
32	287	3.39×10^{-3}	4.02×10^{-3}	$8.444\,03 \times 10^{-1}$
102	898	8.61×10^{-4}	8.70×10^{-4}	$9.894\,51 \times 10^{-1}$
326	2946	2.62×10^{-4}	2.65×10^{-4}	$9.872\,63 \times 10^{-1}$
1043	9397	8.22×10^{-5}	8.32×10^{-5}	$9.884\,45 \times 10^{-1}$
3338	30675	2.51×10^{-5}	2.56×10^{-5}	$9.825\,92 \times 10^{-1}$
10682	96359	7.99×10^{-6}	8.12×10^{-6}	$9.846\,98 \times 10^{-1}$
34182	309389	2.49×10^{-6}	2.54×10^{-6}	$9.824\,08 \times 10^{-1}$

Table 4.9: Results of calculated error and improvement.



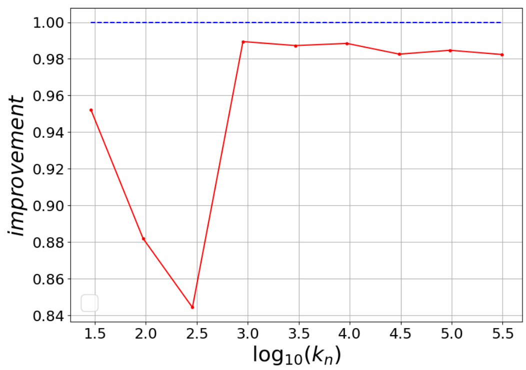


Figure 4.13: Left figure – comparison with theoretical rate of convergence, right figure – improvement calculated by (4.5).

3. $\mu = 0.08$, $\sigma = 0.4$, $c = -0.03$, $\lambda(t) = 2$, $x_0 = 10$, $T = 3$, $K = 60000$.

n	k_n	$\varepsilon_K(\bar{X}_{k_n}^{df-Lin-M*})$	$\varepsilon_K(\bar{X}_{k_n}^{df-Lin-M-eq})$	imp_{K,k_n}^{df}
3	15	3.89	3.90	$9.980\,08 \times 10^{-1}$
10	49	1.07	1.09	$9.805\,73 \times 10^{-1}$
32	154	3.30×10^{-1}	3.38×10^{-1}	$9.758\,61 \times 10^{-1}$
102	490	1.00×10^{-1}	1.05×10^{-1}	$9.555\,14 \times 10^{-1}$
326	1560	3.16×10^{-2}	3.34×10^{-2}	$9.458\,77 \times 10^{-1}$
1043	5066	9.75×10^{-3}	1.03×10^{-2}	$9.479\,37 \times 10^{-1}$
3338	16492	3.15×10^{-3}	3.29×10^{-3}	$9.570\,43 \times 10^{-1}$
10682	56293	1.02×10^{-3}	1.09×10^{-3}	$9.315\,08 \times 10^{-1}$
34182	175734	3.45×10^{-4}	3.84×10^{-4}	$8.975\,61 \times 10^{-1}$

Table 4.10: Results of calculated error and improvement.

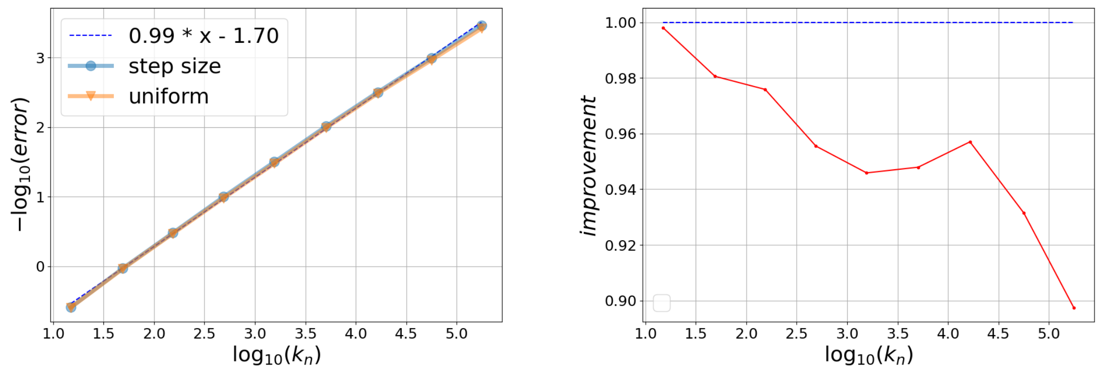


Figure 4.14: Left figure – comparison with theoretical rate of convergence, right figure – improvement calculated by (4.5).

Numerical experiments for method $\bar{X}_n^{Lin-M-eq}$ in multidimensional case

1. $a(t,x) = 0.5\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix},$

$$b(t,x) = \begin{pmatrix} 0.075x_1 & 0.22x_1 & 0.741x_1 & 0.172x_1 & 0.01x_1 \\ 0.254x_2 & 0.634x_2 & 0.925x_2 & 0.901x_2 & 0.943x_2 \\ 0.109x_3 & 0.333x_3 & 0.825x_3 & 0.273x_3 & 0.256x_3 \\ 0.027x_4 & 0.577x_4 & 0.9x_4 & 0.461x_4 & 0.867x_4 \\ 0.623x_5 & 0.097x_5 & 0.438x_5 & 0.275x_5 & 0.682x_5 \end{pmatrix},\ c(t,x) = 1.25\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix},$$

$$\lambda(t) = 1.1245t,\ x_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},\ T = 5,\ K = 60000.$$

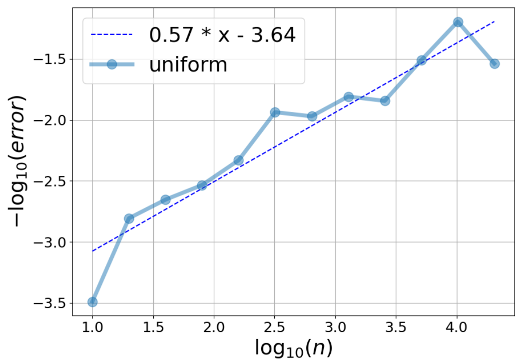


Figure 4.15: Comparison with theoretical rate of convergence.

n	$\varepsilon_K(\bar{X}_n^{Lin-M-eq})$
10	3.08×10^3
20	6.40×10^2
40	4.50×10^2
80	3.42×10^2
160	2.14×10^2
320	8.65×10^1
640	9.34×10^1
1280	6.44×10^1
2560	6.99×10^1
5120	3.24×10^1
10240	1.57×10^1
20480	3.46×10^1

Table 4.11: Results of calculated error.

2. $a(t,x) = 0.5\begin{pmatrix} x_1 \end{pmatrix}$, $b(t,x) = \begin{pmatrix} 0.075x_1 & 0.22x_1 & 0.741x_1 & 0.172x_1 & 0.01x_1 \end{pmatrix}$, $c(t,x) = 1.25\begin{pmatrix} x_1 \end{pmatrix}$, $\lambda(t) = 5$, $x_0 = \begin{pmatrix} 1 \end{pmatrix}$, $T = 2.25$, $K = 60000$.

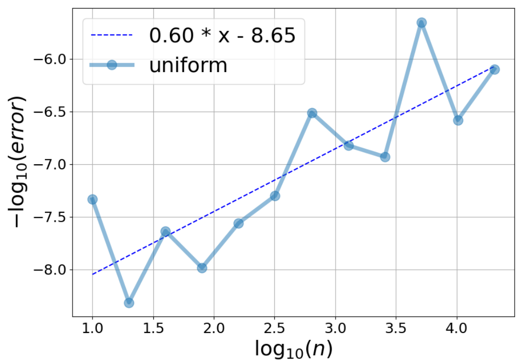


Figure 4.16: Comparison with theoretical rate of convergence.

n	$\varepsilon_K(\bar{X}_n^{Lin-M-eq})$
10	2.16×10^7
20	2.06×10^8
40	4.31×10^7
80	9.64×10^7
160	3.64×10^7
320	2.00×10^7
640	3.24×10^6
1280	6.64×10^6
2560	8.51×10^6
5120	4.47×10^5
10240	3.81×10^6
20480	1.25×10^6

Table 4.12: Results of calculated error.

3. $a(t,x)\mu = 0.5\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix},$

$$b(t,x) = \begin{pmatrix} 0.075x_1 & 0.22x_1 & 0.741x_1 & 0.172x_1 & 0.01x_1 \\ 0.254x_2 & 0.634x_2 & 0.925x_2 & 0.901x_2 & 0.943x_2 \\ 0.109x_3 & 0.333x_3 & 0.825x_3 & 0.273x_3 & 0.256x_3 \\ 0.027x_4 & 0.577x_4 & 0.9x_4 & 0.461x_4 & 0.867x_4 \\ 0.623x_5 & 0.097x_5 & 0.438x_5 & 0.275x_5 & 0.682x_5 \end{pmatrix}, \; c(t,x) = 1.25\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix},$$

$$\lambda(t) = 5, \; x_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \; T = 2.25, \; K = 60000.$$

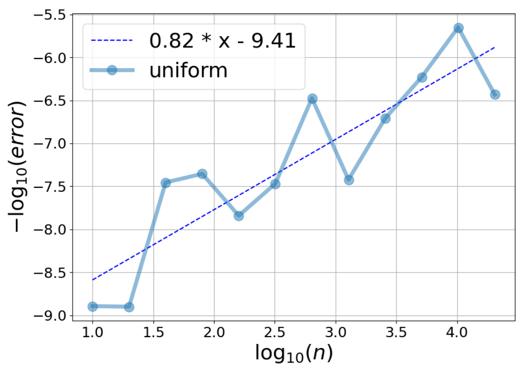


Figure 4.17: Comparison with theoretical rate of convergence.

n	$\varepsilon_K(\bar{X}_n^{Lin-M-eq})$
10	7.82×10^8
20	7.93×10^8
40	2.85×10^7
80	2.26×10^7
160	6.94×10^7
320	2.94×10^7
640	2.98×10^6
1280	2.64×10^7
2560	5.14×10^6
5120	1.70×10^6
10240	4.48×10^5
20480	2.71×10^6

Table 4.13: Results of calculated error.

4. $a(t,x) = 0.5\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$, $b(t,x) = \begin{pmatrix} 0.075x_1 \\ 0.254x_2 \\ 0.109x_3 \\ 0.027x_4 \\ 0.623x_5 \end{pmatrix}$, $c(t,x) = 1.25\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$,

$\lambda(t) = 1.1245t$, $x_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, $T = 2.25$, $K = 60000$.

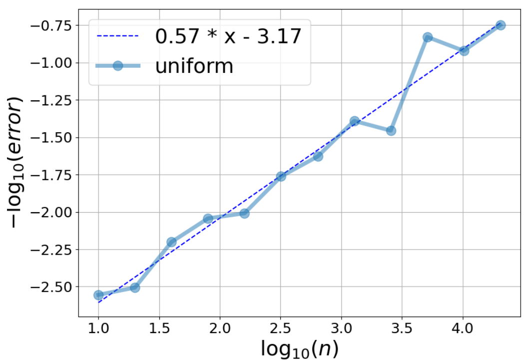


Figure 4.18: Comparison with theoretical rate of convergence.

n	$\varepsilon_K(\bar{X}_n^{Lin-M-eq})$
10	3.59×10^2
20	3.21×10^2
40	1.59×10^2
80	1.10×10^2
160	1.02×10^2
320	5.78×10^1
640	4.25×10^1
1280	2.46×10^1
2560	2.86×10^1
5120	6.73
10240	8.34
20480	5.61

Table 4.14: Results of calculated error.

Chapter 5

Conclusions and future work

In this section we shortly summarize results presented in the thesis. We also identify some open problems corresponding to the equation (1.1).

5.1. Summary of results

In the thesis we considered problem of optimal strong approximation of stochastic differential equation with jumps. In the first part of thesis we investigated scalar problem driven by one-dimensional Wiener and Poisson processes. We analyzed algorithms based on the path-independent adaptive step-size control. We proved that these algorithms are asymptotically optimal in considered class of methods.

In the second part of the thesis we investigated systems of SDEs driven by multi-dimensional Wiener process and one-dimensional Poison process. We considered piecewise linear interpolation of the classical Milstein scheme based on equidistant mesh.

In the last part of the thesis we discussed CUDA C programming language and its application to simulation of stochastic processes. We also presented results of numerical experiments preformed on GPUs.

5.2. Open problems

(OP1) In the future work we would like to investigate algorithms based on path-independent and path-dependent adaptive step-size control in the case when the driving Wiener and Poisson processes are multidimensional.

(OP2) Analysis of (OP1) without assuming jump commutative conditions.

In [23] we investigated the problem of optimal approximation of stochastic integrals $\mathcal{I}(X,W) = \int\limits_0^T X(t)\mathrm{d}W(t)$, where $T > 0$ and $W = \{W(t)\}_{t\geq 0}$ is a standard one-dimensional Wiener process. We were aiming at methods that were based only on discrete values of X and W which were, additionally, corrupted with some noise. Hence, it is natural to investigate the following problem:

(OP3) Investigation of (1.1) in the case when the coefficients a, b, c, as well as the driving processes N and W, are corrupted with some noise.

Inspired by a practical applications we plan a further development of the cuSTOCH library. We also think about application of DNN (deep neural network) into stochastic problems. In [2] we proposed the first solution which was the hybrid model. It combined appropriate methodology for performing fast Monte Carlo simulations on GPUs with application of DNNs to approximating prices of some financial derivative instruments. We plan to go deeper into that topic in the future.

Appendix A

Basic information on stochastic processes and stochastic differential equations

In this section we present basic definitions about random variables, stochastic processes, stochastic integration, stochastic differential equations, and auxiliary results. We have collected here the most important information about the topic discussed in this thesis.

A.1. Random variables and conditional expectation

Definition A.1 ([77]). Let X be a set, 2^X represents a power set of X. The subset $\mathcal{F} \subset 2^X$ is called *σ-algebra* if

- $X \in \mathcal{F}$ and $\emptyset \in \mathcal{F}$,
- for all $A \in \mathcal{F}$, $X \setminus A \in \mathcal{F}$,
- for all $A_1, A_2, \ldots \in \mathcal{F}$, $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Definition A.2 ([77]). Let Ω be a set, and $\mathcal{A}$ be a family of subsets of Ω (i.e. $\mathcal{A} \subset 2^{\Omega}$). The smallest in terms of inclusions σ-algebra contains sets from $\mathcal{A}$ is called σ-algebra generated by family $\mathcal{A}$. We denote it by $\sigma(\mathcal{A})$.

We say that pair $(X, \mathcal{F})$ is a *measurable space*. If X is a topological space, then the σ-algebra generated by all open sets in X (we denote it by $Top(X)$) is called the

Borel σ-algebra on X and we denote it by $\mathcal{B}(X)$. The element $A \in \mathcal{B}$ is called *Borel sets*.

Definition A.3 ([77]). Let $(F, \mathcal{F}_F)$, $(G, \mathcal{F}_G)$ be a measurable space. The product $\mathcal{F}_F \otimes \mathcal{F}_G$ of σ-algebras $\mathcal{F}_F$ and $\mathcal{F}_G$ on $F \times G$ is defined as

$$\mathcal{F}_F \otimes \mathcal{F}_G := \sigma\big(\{A \times B : \ A \in \mathcal{F}_F, B \in \mathcal{F}_G\}\big).$$

Definition A.4 ([77]). Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{G}, \mathcal{H}$ be a sub-σ-algebra in $\mathcal{F}$. The sum of $\mathcal{G}, \mathcal{H}$ is defined as

$$\mathcal{G} \vee \mathcal{H} := \sigma(\mathcal{G} \cup \mathcal{H}).$$

Definition A.5 ([77]). Let $(\Omega, \mathcal{F})$ be a measurable space, mapping $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is called *probability measure* if

- $0 \leqslant \mathbb{P}(A) \leqslant 1$, for all $A \in \mathcal{F}$,
- $\mathbb{P}(\Omega) = 1$,
- for all $A_1, A_2, \ldots \in \mathcal{F}$, $A_i \cap A_j = \emptyset$ for $i \neq j$ then

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Definition A.6 ([77]). Let $(\Omega, \mathcal{F})$ be a measurable space and function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ be a probability measure defined on $\mathcal{F}$. A triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called *probability space*.

Here Ω is a sample space, a set $A \in \mathcal{F}$ is an event and $\mathbb{P}(A)$ is a probability of event A.

Definition A.7 ([1]). The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *complete probability space* if for all $A \subset B$ such that $B \in \mathcal{F}$ and $\mathbb{P}(B) = 0$ we have that $A \in \mathcal{F}$ and $\mathbb{P}(A) = 0$.

Definition A.8 ([5]). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{A_i\}_{i \in I}$ be an indexed family of events. The events A_i, $i \in I$ are called independent if for each finite subset $I_0 \subset I$ we have

$$\mathbb{P}\Big(\bigcap_{i \in I_0} A_i\Big) = \prod_{i \in I_0} \mathbb{P}(A_i).$$

Definition A.9 ([1]). Let $(F, \mathcal{F}_F)$, $(G, \mathcal{F}_G)$ be measurable spaces. Mapping $f : (F, \mathcal{F}_F) \to (G, \mathcal{F}_G)$ is called a *$\mathcal{F}_F/\mathcal{F}_G$-measurable* if for all $A \in \mathcal{F}_G$ we have that

$$f^{-1}(A) \in \mathcal{F}_F.$$

If G is a topological space equipped with the Borel σ-algebra $\mathcal{F}_G = \mathcal{B}(Top(G))$, where $Top(G)$ is a collection of all open set in G, we say that f is *Borel measurable.*

Definition A.10 ([37])**.** Let $(X, \mathcal{F}, \mu)$ be a space with σ-finite measure and a Banach space E with a norm $\|\cdot\|$. A function $f : (X, \mathcal{F}) \to (E, \|\cdot\|)$ is called *strongly measurable* (or *Bochner measurable*) if there exists a sequence of simple functions $\{f_n\}_{n\in\mathbb{N}}$ such that

$$\|f_n(x) - f(x)\| \to 0$$

for almost all $x \in X$.

Based on [27] if space E is separable then every E-valued Borel measurable function f is strongly measurable. It follows from the fact that a subset of separable metric space is itself separable. Moreover, if $f : X \to E$ is Borel measurable then mapping

$$X \in x \to \|f(x)\| \in (0, \infty)$$

is $\mathcal{F}/\mathcal{B}(\mathbb{R}^d)$-measurable, we also write $\mathcal{F}$-measurable in a case of Borel set when $E = \mathbb{R}^d$.

Definition A.11 ([1])**.** Let $(\Omega, \mathcal{F})$, $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be measurable spaces. Mapping $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is called *random variable* if it is a $\mathcal{F}$-measurable. We write $X : \Omega \to \mathbb{R}^d$.

Definition A.12 ([1])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}^d$ is a random variable defined on it. The σ-algebra generated by X is given by

$$\sigma(X) := \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R}^d)\}.$$

Definition A.13 ([5])**.** Let $\{X_i\}_{i\in I}$ be an indexed family of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$, and with values in the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The random variables X_i, $i \in I$, are called independent if for each choice of sets A_i in $\mathcal{B}(\mathbb{R}^d)$, $i \in I$, the events $X_i^{-1}(A_i)$ are independent.

Definition A.14 ([5])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{\mathcal{F}_i\}_{i\in I}$ be an indexed family of sub-σ-algebras of $\mathcal{F}$. The σ-algebras $\{\mathcal{F}_i\}_{i\in I}$ are independent if for each choice of sets $A_i \in \mathcal{F}_i$, where $i \in I$, the events A_i are independent.

Fact A.15 ([5])**.** *If $\{X_i\}_{i\in I}$ is an indexed family of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the random variables X_i, where $i \in I$, are independent if and only if the σ-algebras $\{\sigma(X_i)\}_{i\in I}$ are independent.*

Definition A.16. Let $(X, \mathcal{F}, \mu; E)$ be a measured space where E is a separable Banach space with norm $\|\cdot\|$, and let $p \in [1, \infty)$. We define space

$$\mathfrak{L}^p(X, \mathcal{F}, \mu; E) = \Big\{ f : X \to E \mid f - \text{Borel measurable}, \int_X \|f(x)\|^p \mu(\mathrm{d}x) < \infty \Big\}.$$

If we identify functions which are equal μ-almost everywhere, then $\mathfrak{L}^p(X, \mathcal{F}, \mu; E)$ is Banach space with norm

$$\|f\|_{\mathfrak{L}^p(X,\mathcal{F},\mu;E)} = \left(\int_X \|f(x)\|^p \mu(\mathrm{d}x) \right)^{1/p}.$$

We use the following notation for the $\mathfrak{L}^p$ spaces.

Definition	Shortcut Norm
$\mathfrak{L}^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$	$\mathfrak{L}^2(\Omega; \mathbb{R}^d)$ or $\mathfrak{L}^2(\Omega)$ $\|X\|_{\mathfrak{L}^2(\Omega;\mathbb{R}^d)} = \mathbb{E}\big(\|X\|^2\big)^{1/2}$
$\mathfrak{L}^2([0,T], \mathcal{B}([0,T]), \lambda_1; \mathbb{R}^d)$	$\mathfrak{L}^2([0,T]; \mathbb{R}^d)$ $\|f\|_{\mathfrak{L}^2([0,T];\mathbb{R}^d)} = \big(\int_0^T \|f(x)\|^2 \mathrm{d}x\big)^{1/2}$
$\mathfrak{L}^2(\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}([0,T]), \mathbb{P} \times \lambda_1; \mathbb{R}^d)$	$\mathfrak{L}^2(\Omega \times [0,T]; \mathbb{R}^d)$ or $\mathfrak{L}^2(\Omega \times [0,T])$ $\|f\|_{\mathfrak{L}^2(\Omega\times[0,T];\mathbb{R}^d)} = \Big(\mathbb{E}\int_0^T \|f(x)\|^2 \mathrm{d}x\Big)^{1/2}$
$\mathfrak{L}^2(\Omega \times [0,T], \sigma(\mathcal{N}_n(W,N)) \otimes \mathcal{B}([0,T]), \mathbb{P} \times \lambda_1; \mathbb{R}^d)$	$\mathfrak{L}^2(\Omega \times [0,T]; \mathbb{R}^d)$ or $\mathfrak{L}^2(\Omega \times [0,T])$ $\|f\|_{\mathfrak{L}^2(\Omega\times[0,T];\mathbb{R}^d)} = \Big(\mathbb{E}\int_0^T \|f(x)\|^2 \mathrm{d}x\Big)^{1/2}$

Definition A.17 ([36])**.** Let X be an integrable random variable defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ it means that $\mathbb{E}|X| < \infty$. Suppose $\mathcal{G}$ is a σ-algebra and $\mathcal{G} \subset \mathcal{F}$. *The conditional expectation* of X given $\mathcal{G}$ is defined to be the unique random variable Y (up to $\mathbb{P}$-measure one) satisfying the following conditions:

1. Y is $\mathcal{G}$-measurable,
2. $\int_A X \mathrm{d}\mathbb{P} = \int_A Y \mathrm{d}\mathbb{P}$ for all $A \in \mathcal{G}$.

We use $\mathbb{E}(X \mid \mathcal{G})$ to denote the conditional expectation of X given $\mathcal{G}$. We recall that the notion $\mathbb{E}(X \mid Y) = \mathbb{E}(X \mid \sigma(Y))$ formally refers to conditioning given σ-algebra generated by the random variable Y.

Proposition A.18 ([29, 49])**.** Let X, Y be integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, and a, b be real numbers. Then:

a) $\mathbb{E}(aX + bY \mid \mathcal{G}) = a \cdot \mathbb{E}(X \mid \mathcal{G}) + b \cdot \mathbb{E}(Y \mid \mathcal{G})$,

b) $\big|\mathbb{E}(X \mid \mathcal{G})\big| \leqslant \mathbb{E}\big(|X| \mid \mathcal{G}\big)$,

c) if X is $\mathcal{G}$-measurable then

$$\mathbb{E}(X \mid \mathcal{G}) = X,$$

d) if $\mathcal{G}, \mathcal{H}$ are σ-algebras such that $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, then

$$\mathbb{E}\big(\mathbb{E}(X \mid \mathcal{H}) \mid \mathcal{G}\big) = \mathbb{E}\big(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}\big) = \mathbb{E}(X \mid \mathcal{G}),$$

and in particular $\mathbb{E}\big(\mathbb{E}(X \mid \mathcal{G})\big) = \mathbb{E}(X)$.

Proposition A.19 ([29]). Let X be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}X^2 < \infty$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$. Then $\mathbb{E}(X \mid \mathcal{G})$ is the orthogonal projection of X on $\mathfrak{L}^2(\Omega, \mathcal{G}, \mathbb{P})$. That is, for any $\mathcal{G}$-measurable random variable Y with $\mathbb{E}Y^2 < \infty$,

$$\mathbb{E}\big(X - Y\big)^2 \geqslant \mathbb{E}\big(X - \mathbb{E}(X \mid \mathcal{G})\big)^2$$

with the equality if and only if $Y = \mathbb{E}(X \mid \mathcal{G})$.

Lemma A.20 ([6]). *Let X be integrable random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\mathcal{G}, \mathcal{F}$ be a sub-σ-algebras such that $\mathcal{H} \perp\!\!\!\perp \sigma(\sigma(X) \cup \mathcal{G})$ then*

$$\mathbb{E}(X \mid \mathcal{G} \vee \mathcal{H}) = \mathbb{E}(X \mid \mathcal{G}).$$

Definition A.21 ([6]). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ be a three sub-σ-algebras of $\mathcal{F}$. $\mathcal{F}_1$ and $\mathcal{F}_3$ are called *conditionally independent given* $\mathcal{F}_2$ if

$$\mathbb{E}(Y_1 Y_3 \mid \mathcal{F}_2) = \mathbb{E}(Y_1 \mid \mathcal{F}_2) \cdot \mathbb{E}(Y_3 \mid \mathcal{F}_2),$$

where Y_1, Y_3 denote positive random variables measurable with a respect to the corresponding σ-algebras $\mathcal{F}_1, \mathcal{F}_3$. We will mark it as $\mathcal{F}_1 \perp\!\!\!\perp_{\mathcal{F}_2} \mathcal{F}_3$.

Theorem A.22 ([6]). *Let $\mathcal{F}_{12} = \mathcal{F}_1 \vee \mathcal{F}_2$. Then $\mathcal{F}_1$ and $\mathcal{F}_3$ are conditionally independent given $\mathcal{F}_2$ if and only if*

$$\mathbb{E}(Y_3 \mid \mathcal{F}_{12}) = \mathbb{E}(Y_3 \mid \mathcal{F}_2), \ a.s.$$

for every $\mathcal{F}_3$-measurable and integrable random variable Y_3.

Proposition A.23. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ be a three sub-σ-algebras of $\mathcal{F}$ which satisfy $\mathcal{F}_1 \perp\!\!\!\perp_{\mathcal{F}_2} \mathcal{F}_3$. Let $Y_1, Y_3 : \Omega \to \mathbb{R}$ be a random variables and $\sigma(Y_1) \subset \mathcal{F}_1$, $\sigma(Y_3) \subset \mathcal{F}_3$. We assume that $\mathbb{E}|Y_1| < +\infty$, $\mathbb{E}|Y_3| < +\infty$ and $\mathbb{E}|Y_1 Y_3| < +\infty$. Then we have that

$$\mathbb{E}(Y_1 Y_3 \mid \mathcal{F}_2) = \mathbb{E}(Y_1 \mid \mathcal{F}_2) \cdot \mathbb{E}(Y_3 \mid \mathcal{F}_2) \ a.s. \tag{A.1}$$

Proof. Let $Y_i = Y_i^+ - Y_i^-$ for $i = 1, 3$. We have that $0 \leqslant Y_i^+, Y_i^-$ and $0 \leqslant Y_i^{+/-} \leqslant |Y_i|$. $\sigma(Y_i^{+/-}) \subset \sigma(Y_i) \subset \mathcal{F}_i$ for $i = 1, 3$. We also have that $0 \leqslant Y_1^{+/-} Y_3^{+/-} \leqslant Y_1 Y_3$. So random variables $Y_1^+, -Y_1^-, Y_3^+, -Y_3^-$ are positive and integrable. From assumption about conditional independence $\mathcal{F}_1 \perp\!\!\!\perp_{\mathcal{F}_2} \mathcal{F}_3$ and integrability of defined random variables, we have

$$\begin{aligned}
\mathbb{E}(Y_1 Y_3 \mid \mathcal{F}_2) &= \mathbb{E}\big((Y_1^+ - Y_1^-) \cdot (Y_3^+ - Y_3^-) \mid \mathcal{F}_2\big) \\
&= \mathbb{E}(Y_1^+ Y_3^+ \mid \mathcal{F}_2) - \mathbb{E}(Y_1^+ Y_3^- \mid \mathcal{F}_2) - \mathbb{E}(Y_1^- Y_3^+ \mid \mathcal{F}_2) + \mathbb{E}(Y_1^- Y_3^- \mid \mathcal{F}_2) \\
&= \mathbb{E}(Y_1^+ \mid \mathcal{F}_2) \cdot \mathbb{E}(Y_3^+ \mid \mathcal{F}_2) - \mathbb{E}(Y_1^+ \mid \mathcal{F}_2) \cdot \mathbb{E}(Y_3^- \mid \mathcal{F}_2) \\
&\qquad - \mathbb{E}(Y_1^- \mid \mathcal{F}_2) \cdot \mathbb{E}(Y_3^+ \mid \mathcal{F}_2) + \mathbb{E}(Y_1^- \mid \mathcal{F}_2) \cdot \mathbb{E}(Y_3^- \mid \mathcal{F}_2) \\
&= \mathbb{E}(Y_1^+ \mid \mathcal{F}_2) \cdot \Big(\mathbb{E}(Y_3^+ \mid \mathcal{F}_2) - \mathbb{E}(Y_3^- \mid \mathcal{F}_2)\Big) \\
&\qquad + \mathbb{E}(Y_1^- \mid \mathcal{F}_2) \cdot \Big(\mathbb{E}(Y_3^+ \mid \mathcal{F}_2) - \mathbb{E}(Y_3^- \mid \mathcal{F}_2)\Big) \\
&= \Big(\mathbb{E}(Y_1^+ \mid \mathcal{F}_2) - \mathbb{E}(Y_1^- \mid \mathcal{F}_2)\Big) \cdot \Big(\mathbb{E}(Y_3^+ \mid \mathcal{F}_2) - \mathbb{E}(Y_3^- \mid \mathcal{F}_2)\Big) \\
&= \mathbb{E}(Y_1 \mid \mathcal{F}_2) \cdot \mathbb{E}(Y_3 \mid \mathcal{F}_2).
\end{aligned}$$

That ends the proof. ∎

A.2. Basic fact from the theory of stochastic processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. Let $\mathcal{B}([0, +\infty))$ be a σ-algebra of Borel sets defined on $[0, \infty)$. Now, we recall the definitions of a filtration and a stochastic process.

Definition A.24 ([38]). A *filtration* is a family $\{\mathcal{F}_t\}_{t \geq 0}$ of increasing sub-σ-algebras of $\mathcal{F}$, i.e. $\mathcal{F}_t \subset \mathcal{F}_s \subset \mathcal{F}$ for all $0 \leqslant t < s$. The filtration is called *right continuous* if $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ for all $t \geqslant 0$. For a complete probability space the filtration is said to satisfy *the usual conditions* if it is right continuous and $\mathcal{F}_0$ contains all sets of zero measure.

Let us define $\mathcal{F}_\infty = \sigma\Big(\bigcup_{t\geq 0}\mathcal{F}_t\Big)$.

Definition A.25 ([38]). A family $X = \big\{X(t)\big\}_{t\geq 0}$ of $\mathbb{R}^d$-valued random variables is called a *stochastic process* with the parameter set $\mathbb{R}_+$ and the state space Ω.

For any parameter $t \in [0, +\infty)$ we have a random variable

$$\Omega \ni \omega \mapsto X(\omega, t) \in \mathbb{R}^d.$$

For a fixed state $\omega \in \Omega$, a function

$$[0, +\infty) \ni t \mapsto X(\omega, t),$$

is called a *sample path* of the process.

Definition A.26 ([38]). A stochastic process X is *continuous* if for almost all $\omega \in \Omega$ the function $X(\omega, \,\cdot\,)$ is continuous on $[0, +\infty)$.

Definition A.27 ([38]). A stochastic process X is *càdlàg* if the process has right continuous paths and left limits almost everywhere. A stochastic process X is *càglàd* if the process has left continuous paths and right limits almost everywhere.

Definition A.28 ([38]). We say that a process X is *adapted to filtration* $\big\{\mathcal{F}_t\big\}_{t\geq 0}$ if for all $t \geqslant 0$ the random variable $X(t)$ is $\mathcal{F}_t$-measurable.

Definition A.29 ([38]). Let $\mathfrak{O}$ (resp. $\mathfrak{P}$) denote the smallest σ-algebra on $\Omega \times \mathbb{R}_+$ with respect to every càdlàg adapted process (resp. càglàd) is a measurable function of (ω, t). A stochastic process is said to be *optional* (resp. *predictable*) if the process regarded as a function of (ω, t) is $\mathfrak{O}$-measurable (resp. $\mathfrak{P}$-measurable).

Theorem A.30 ([38]). *Every càglàd and adapted to filtration* $\big\{\mathcal{F}_t\big\}_{t\geq 0}$ *stochastic process* X *is predictable.*

Definition A.31 ([38]). We say that a stochastic process X is *measurable* if the process regarded as a function of two variables (ω, t) from $\Omega \times [0, +\infty) \to \mathbb{R}^d$ is $\mathcal{F} \otimes \mathcal{B}([0, +\infty)])$-measurable.

Definition A.32 ([38]). A stochastic process $X : \Omega \times [0, +\infty) \to \mathbb{R}^d$ is called *progressively measurable* if for every $t > 0$ the function $X|_{\Omega\times[0,t]}$ is $\mathcal{F}_t \otimes \mathcal{B}([0, t])/\mathcal{B}(\mathbb{R}^d)$-measurable.

Theorem A.33 ([38])**.** *Every $\{\mathcal{F}_t\}_{t\geq 0}$ progressively measurable process is measurable and adapted to filtration $\{\mathcal{F}_t\}_{t\geq 0}$.*

Definition A.34 ([38])**.** A *natural filtration* $\{\mathcal{F}_t^X\}_{t\geq 0}$ of a process X is defined as

$$\mathcal{F}_t^X = \sigma\big(X(s)\colon 0 \leqslant s \leqslant t\big).$$

Any stochastic process is adapted to its natural filtration.

Definition A.35 ([38])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$. A standard *one-dimensional Brownian motion* is a real-valued continuous and $\{\mathcal{F}_t\}_{t\geq 0}$-adapted process $\{W(t)\}_{t\geq 0}$ with the following properties:

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

(ii) for $0 \leqslant s < t$, the increment $W(t) - W(s) \sim \mathtt{N}(0, t-s)$,

(iii) for $0 \leqslant s < t$, the increment $W(t) - W(s)$ is independent of $\mathcal{F}_s$.

Any Brownian motion is adapted to its natural filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$. Moreover, if $\{\mathcal{F}_t\}_{t\in[0,T]}$ is (in terms of inclusion) 'larger' filtration, i.e. $\mathcal{F}_t^W \subset \mathcal{F}_t$ for all $t \geqslant 0$, and $W(t) - W(s)$ independent of $\mathcal{F}_s$ whenever $0 \leqslant s < t < \infty$, then $W(t)$ is a Brownian motion with respect to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$.

The σ-algebra

$$\mathcal{F}_t = \sigma\big(\mathcal{F}_t^W \cup \{A \in \mathcal{F}\colon \mathbb{P}(A) = 0\}\big).$$

It is called an augmentation under $\mathbb{P}$ of the natural filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$ generated by Brownian motion W. The augmentation is a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition A.36 ([38])**.** An m_{w}-dimensional process $\{W(t) = \big(W_1(t), \ldots, W_{m_{\mathrm{w}}}(t)\big)^T\}_{t\geq 0}$ is called an m_{w}-dimensional Brownian motion if every $\{W_i(t)\}_{t\geq 0}$ is a one-dimensional Brownian motion and $\{W_1(t)\}_{t\geq 0}, \ldots, \{W_{m_{\mathrm{w}}}(t)\}_{t\geq 0}$ are independent.

Definition A.37 ([61])**.** Stochastic process $\{N(t)\}_{t\geq 0}$ is called *the non-homogeneous Poisson point process*, with intensity function $\lambda(t) > 0$ and $\int_0^+ \infty\lambda(t)\mathrm{d}t < +\infty$ when it satisfies the following conditions

(i) $N(0) = 0$,

(ii) has independent increments,

(iii) $N(t) - N(s) \sim \mathtt{Poiss}(\Lambda(t, s))$,

where

$$m(t) = \int_0^t \lambda(s)\mathrm{d}s,$$

$$\Lambda(t,s) = m(t) - m(s),\ t,s \in [0,T].$$

By the Definition A.37 we have the following properties.

Proposition A.38 ([61]). For the homogeneous Poisson point process $\{N(t)\}_{t\geq 0}$ we have for all $t > 0$ that

(i) $\mathbb{E}\big(N(t)\big) = m(t)$,

(ii) $\mathbb{P}\big(N(t) = n\big) = \frac{(m(t))^n}{n!}e^{-m(t)}$.

Definition A.39 ([61]). The *compensated Poisson process* $\tilde{N} = \{\tilde{N}(t)\}_{t\in[0,T]}$ is defined as follows

$$\tilde{N}(t) = N(t) - m(t), \quad t \in [0,T]. \tag{A.2}$$

Fact A.40 ([61]). *The compensated Poisson process $\tilde{N} = \{\tilde{N}(t)\}_{t\in[0,T]}$ is a martingale (see Definition A.41).*

A.3. Stochastic integration with respect to square integrable martingale

Definition A.41 ([38]). A real valued, $\{\mathcal{F}_t\}_{t\geq 0}$-adapted integrable process $\{M(t)\}_{t\geq 0}$ (i.e. $\mathbb{E}|M(t)| < \infty$ for all t) is called a *martingale with respect to* $\{\mathcal{F}_t\}_{t\geq 0}$ (or simply, *martingale*) if

$$\mathbb{E}\big(M(t) \mid \mathcal{F}_s\big) = M(s)\ a.s.,$$

for all $0 < s < t < \infty$.

Definition A.42 ([78]). Let $\{M(t)\}_{t\geq 0}$ be a martingale such that $M(0) = 0$, we say that $\{M(t)\}_{t\geq 0}$ is *square integrable martingale* if for all $t \geqslant 0$

$$\mathbb{E}\big(M^2(t)\big) < \infty.$$

By $\mathcal{M}^2$ we denote a space of square integrable martingale. If additionally $\{M(t)\}_{t\geq 0}$ is continuous we say that $\{M(t)\}_{t\geq 0}$ is *square integrable continuous martingale*, and by $\mathcal{M}^{2,c}$ we denote space of square integrable continuous martingale.

Definition A.43 ([38])**.** A random variable $\tau : \Omega \to [0, \infty]$ (it may take the value ∞) is called an $\{\mathcal{F}_t\}_{t\geq 0}$-*stopping time* (or simply, *stopping time*) if

$$\{\omega : \tau(\omega) \leqslant t\} \in \mathcal{F}_t,$$

for any $t > 0$.

Definition A.44 ([38])**.** A right continuous adapted process $M = \{M(t)\}_{t\geq 0}$ is called a *local martingale* if there exists a non-decreasing sequence $\{\tau_k\}_{k\geq 0}$ of stopping times with $\tau_k \uparrow \infty$ a.s. such that every $\{M(\tau_k \wedge t) - M(0)\}_{t\geq 0}$ is a martingale.

Definition A.45 ([78])**.** A right continuous adapted process $M = \{M(t)\}_{t\geq 0}$ is called a *locally square integrable martingale* if there exists a non-decreasing sequence $\{\tau_k\}_{k\geq 0}$ of stopping times with $\tau_k \uparrow \infty$ a.s. such that every $\{M(\tau_k \wedge t)\}_{t\geq 0} \in \mathcal{M}^2$.
By $\mathcal{M}^{2,loc}$ we denote a space of locally square integrable martingale. If additionally $\{M(t)\}_{t\geq 0}$ is continuous we say that it is *locally square integrable continuous martingale*, and by $\mathcal{M}^{2,loc,c}$ we denote space of locally square integrable continuous martingale.

Definition A.46 ([25])**.** An càdlàg adapted process X is said to be a *semi-martingale* if X can be decomposed into $X = M + A$ where M is an càdlàg local martingale and A is an càdlàg process whose paths have finite variation on $[0, T]$ for all $T < \infty$. We call this decomposition D-M (Doob-Meyer) decomposition.

Proposition A.47 ([78])**.** Let $\{M(t)\}_{t\geq 0}, \{\tilde{M}(t)\}_{t\geq 0} \in \mathcal{M}^2$ then

- $\{M^2(t)\}_{t\geq 0}$ has a unique D-M decomposition as follow

  $$M^2(t) = martingale + \langle M\rangle(t),$$

  where $\langle M\rangle(t)$ is a natural (predictable) integrable increasing process, and it is called *(predictable) characteristic process* for $M(t)$.
- $\{M(t) \cdot \tilde{M}(t)\}_{t\geq 0}$ has a unique D-M decomposition as follow

  $$M(t) \cdot \tilde{M}(t) = martingale + \langle M, \tilde{M}\rangle(t),$$

  where $\langle M, \tilde{M}\rangle(t)$ is a natural (predictable) integrable finite variational process, i.e. it is a difference of two natural (predictable) integrable increasing processes and it is called the cross predictable characteristic process (or (predictable) quadratic variational $\mathcal{F}_t$-adapted process) for $M(t)$ and $\tilde{M}(t)$.

Now we show partial construction of stochastic integral with the respect to square integrable martingale. For the full concept we refer to [78].

Definition A.48 ([78])**.** By $\mathfrak{L}^0(\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}([0,T]), \mathbb{P} \times \lambda_1; \mathbb{R}^d)$ we denote the space of all real-valued, $\{\mathcal{F}_t\}_{t\geq 0}$-adapted processes such that exists decomposition $0 = t_0 < t_1 < \ldots < t_n < \ldots \to +\infty$ and exists $\psi_i(\omega)$ $\mathcal{F}_{t_i}$-measurable such that $\sup_i(\text{ess sup}\,\|\psi_i(\omega,t)\|^2) < +\infty$ and we can write f as simply function

$$f(\omega,t) = \psi_0(\omega)\mathbb{1}_{t=0}(t) + \sum_{i=0}^{\infty} \psi_i(\omega)\mathbb{1}_{t\in(t_i,t_{i+1}]}(t).$$

Fact A.49 ([78])**.** *$\mathfrak{L}^0(\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}([0,T]), \mathbb{P} \times \lambda_1; \mathbb{R}^d)$ is dense in $\mathfrak{L}^2(\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}([0,T]), \mathbb{P} \times \lambda_1; \mathbb{R}^d)$ with the respect to complete norm*

$$\|f\|_{\mathfrak{L}^2} = \sum_{n=0}^{\infty} \frac{1}{2^n}\big(\|f\|^2_{\mathfrak{L}^2(\Omega\times[0,T])} \wedge 1\big).$$

Definition A.50 ([78])**.** We denote by $\mathfrak{L}^2_M$ the space of all real-valued, $\{\mathcal{F}_t\}_{t\geq 0}$-adapted processes $f = \{f(t,\omega)\}_{t\geq 0}$ such that for all $t > 0$ we have that

$$\|f\|^2_{\mathfrak{L}^2(\Omega\times[0,t]),M} = \mathbb{E}\left(\int_0^t |f(s)|^2 \mathrm{d}\langle M\rangle(t)\right) < +\infty.$$

Fact A.51 ([78])**.** *$\mathfrak{L}^0$ is dense in $\mathfrak{L}^2_M$ with the respect to complete norm*

$$\|f\|_{\mathfrak{L}^2,M} = \sum_{n=0}^{\infty} \frac{1}{2^n}\big(\|f\|^2_{\mathfrak{L}^2(\Omega\times[0,t]),M} \wedge 1\big).$$

Definition A.52 ([78])**.** Let $M \in \mathcal{M}^2$. For every $f \in \mathfrak{L}^0$ we define a *Itô integral* for $t_n < t \leqslant t_{n+1}$, $n = 0,1,\ldots$ as

$$I(f)(\omega,t) = \int_0^t f(\omega,s)\mathrm{d}M(\omega,s) = \sum_{i=0}^{n} f_i(\omega)\cdot\big(M(\omega,t_{i+1}) - M(\omega,t_i)\big),$$

and we can write as a infinite sum

$$I(f)(t) = \int_0^t f(s)\mathrm{d}M(s) = \sum_{i=0}^{\infty} f_i \cdot \big(M(t_{i+1}\wedge t) - M(t_i \wedge t)\big).$$

Definition A.53 ([78])**.** Let $M \in \mathcal{M}^2$ and let $f \in \mathfrak{L}^2_M$ and $\{f_n\}_{n=1}^{\infty}$ be a sequence of processes from $\mathfrak{L}^0$ such that $\|f - f_n\|_{\mathfrak{L}^2,M} \to 0$ when $n \to \infty$. We define a *Itô integral* as a limit for

$$I(f)(t) = \lim_{n\to\infty} I(f_n)(t),$$

and we write

$$I(f)(t) = \int_0^t f(s)\mathrm{d}M(s).$$

Fact A.54 ([78])**.** *If $f \in \mathfrak{L}^2$ then the Itô integral $I(f)$ belongs to $\mathcal{M}^{2,c}$.*

Theorem A.55 ([26,36,38,48,78,83])**.** *Let $M, \tilde{M} \in \mathcal{M}^2$, $f, g \in \mathfrak{L}^2_M$ and $h \in \mathfrak{L}^2_{\tilde{M}}$ then for all $\tau \geqslant \sigma$ σ, τ – stopping time, for all $t, s \geq 0$ we have that*

- $\mathbb{E}\left(\int_0^s f(t)\mathrm{d}M(t)\right) = 0,$
- $\mathbb{E}\left(\int_0^s f(t)\mathrm{d}M(t)\right)^2 = \mathbb{E}\int_0^s |f(t)|^2\mathrm{d}\langle M\rangle(t),$
- $\int_0^s f(t)\mathrm{d}M(t)$ *is $\mathcal{F}_s$-measurable,*
- $\mathbb{E}\left(\left(\int_0^{t\wedge\tau} f(u)\mathrm{d}M(u) - \int_0^{t\wedge\sigma} f(u)\mathrm{d}M(u)\right) \mid \mathcal{F}_\sigma\right) = 0$ *a.s.,*
- $\int_0^t f(u) + g(u)\mathrm{d}M(u) = \int_0^t f(u)\mathrm{d}M(u) + \int_0^t g(u)\mathrm{d}M(u)$ *a.s.,*
- $\mathbb{E}\left(\left(\int_0^{t\wedge\tau} f(u)\mathrm{d}M(u) - \int_0^{t\wedge\sigma} f(u)\mathrm{d}M(u)\right) \mid \mathcal{F}_\sigma\right) = 0$ *a.s.,*
- $\mathbb{E}\left(\left(\int_0^{t\wedge\tau} f(u)\mathrm{d}M(u) - \int_0^{t\wedge\sigma} f(u)\mathrm{d}M(u)\right) \cdot \left(\int_0^{t\wedge\tau} g(u)\mathrm{d}M(u) - \int_0^{t\wedge\sigma} g(u)\mathrm{d}M(u)\right) \mid \mathcal{F}_\sigma\right)$
  $= E\left(\left(\int_{t\wedge\sigma}^{t\wedge\tau} f(u)g(u)\mathrm{d}\langle M\rangle(u)\right| \mathcal{F}_\sigma\right)$ *a.s.,*
- $\mathbb{E}\left(\int_0^t f(u)\mathrm{d}M(u) - \int_0^s f(u)\mathrm{d}M(u)\right) \cdot \left(\int_0^t h(u)\mathrm{d}\tilde{M}(u) - \int_0^s h(u)\mathrm{d}\tilde{M}(u)\right) \mid \mathcal{F}_s\Big)$
  $= \mathbb{E}\left(\left(\int_s^t f(u)h(u)\mathrm{d}\langle M, \tilde{M}\rangle(u)\right| \mathcal{F}_s\right)$ *a.s.*
- *Let ξ is a real-valued $\mathcal{F}_s$-measurable random variable then*

$$\int_s^t \xi \cdot f(u)\mathrm{d}M(u) = \xi \int_s^t f(u)\mathrm{d}M(u) \text{ a.s.}$$

Stochastic integral defined in this section can be extended for more general stochastic processes (see [78]). It turns out that any càglàd processes are integrable with respect to a semi-martingale. It is because of the fact that it compensator is absolutely continuous. By this fact whole considered in thesis processes (N and W) satisfy all the necessary assumptions and can be integrated with the respect to semi-martingales.

Lemma A.56 ([38]). *Let W_1, W_2 be a one-dimensional Wiener processes and N be non-homogeneous Poisson process. Then we have that*

$$\langle W_1, W_1\rangle(t) = t,$$
$$\langle W_1, W_2\rangle(t) = 0,$$
$$\langle W_1, N\rangle(t) = 0.$$

In the thesis we also use multidimensional Wiener process so we have to define a *multidimensional Itô integral.*

Definition A.57 ([38]). Let $M_j \in \mathcal{M}^2$ for $j \in \{1, \dots, m_{\mathrm{w}}\}$. Using matrix notation, we define the multi-dimensional Itô integral for f

$$I(f)(t) = \int\limits_0^t f(s)\mathrm{d}M(s) = \int\limits_0^t \begin{pmatrix} f^{1,1}(s) & f^{1,2}(s) & \dots & f^{1,m_{\mathrm{w}}}(s) \\ f^{2,1}(s) & f^{2,2}(s) & \dots & f^{2,m_{\mathrm{w}}}(s) \\ \vdots & \vdots & \ddots & \vdots \\ f^{d,1}(s) & f^{d,2}(s) & \dots & f^{d,m_{\mathrm{w}}}(s) \end{pmatrix} \begin{pmatrix} \mathrm{d}M_1(s) \\ \mathrm{d}M_2(s) \\ \vdots \\ \mathrm{d}M_{m_{\mathrm{w}}}(s) \end{pmatrix},$$

where $f^{ij} \in \mathfrak{L}^2_{M_j}$, $i \in \{1, \dots, d\}$, $j \in \{1, \dots, m_{\mathrm{w}}\}$. Defined in that way stochastic integral $I(f)(t)$ is the d-dimensional column-vector-valued process whose ith component is the following sum of 1-dimensional Itô integrals,

$$\sum_{j=1}^{m_{\mathrm{w}}} \int\limits_0^t f^{i,j}(s)\mathrm{d}M_j(s).$$

Lemma A.58. *Let $f = (f^1, \dots, f^{m_{\mathrm{w}}})$ be a function where $f^j \in \mathfrak{L}^2_{W_j}$, $j \in \{1, \dots, m_{\mathrm{w}}\}$ and W_j for $j \in \{1, \dots, m_{\mathrm{w}}\}$ be a Brownian Motion. Then we have the following Itô isometry*

- $\mathbb{E}\left(\int\limits_0^s f(t)\mathrm{d}W(t)\right)^2 = \mathbb{E}\int\limits_0^s \|f(t)\|^2\mathrm{d}t.$

Proof.

$$
\begin{aligned}
\mathbb{E}\left(\int_0^s f(t)\mathrm{d}W(t)\right)^2 &= \mathbb{E}\Big(\sum_{j=1}^{m_{\mathrm{w}}}\int_0^t f^j(s)\mathrm{d}W_j(s)\Big)^2 \\
&= \sum_{j=1}^{m_{\mathrm{w}}}\mathbb{E}\Big(\int_0^t f^j(s)\mathrm{d}W_j(s)\Big)^2 \\
&\qquad + 2\mathbb{E}\sum_{i<j}\mathbb{E}\bigg(\Big(\int_0^t f^i(s)\mathrm{d}W_i(s)\Big)\cdot\Big(\int_0^t f^j(s)\mathrm{d}W_j(s)\Big)\bigg) \\
&= \sum_{j=1}^{m_{\mathrm{w}}}\mathbb{E}\int_0^t \Big(f^j(s)\Big)^2\mathrm{d}t = \mathbb{E}\int_0^t \|f(s)\|^2\mathrm{d}t.
\end{aligned}
$$

This ends the proof. ∎

In this thesis we consider also stochastic integrals with respect to the Poisson process N. The process is a semi-martingale and for any càglàd process f the stochastic integral with respect to N is defined as follows (see [83]) for all $s, t > 0$

$$
\int_s^t f(u)\mathrm{d}N(u) = \int_s^t f(u)\mathrm{d}\tilde{N}(u) + \int_s^t f(u)\lambda(u)\mathrm{d}u,
$$

where $\tilde{N}$ is defined in Definition A.39. Note that the integral with respect to $\tilde{N}$ is a stochastic integral with respect to the square-integrable martingale $\tilde{N}$. Moreover, due to the fact that the trajectories of N and $\tilde{N}$ are of finite variation, the above stochastic integrals with respect to N and $\tilde{N}$ are equivalent to Lebesgue-Stjeltjes integrals (for more see [62]).

Now we can show a multidimensional version of the Itô formula for semi-martingales with jumps, see, for example, [78] or [62].

Lemma A.59 ([19])**.** *Let us assume that the mappings a, b, c and λ satisfy the assumptions $(B1_{\mathrm{MD}}), (B2_{\mathrm{MD}})$, and (E_{MD}). Let a function $U : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ belongs* *to* $C^{1,2}([0,T]\times\mathbb{R}^d)$*. Then for the solution* X *of* (1.1) *it holds that*

$$
\begin{aligned}
U(t,X(t)) \;=\; & U(0,X(0)) + \int_0^t \Big(\frac{\partial}{\partial t}U(s,X(s)) + \nabla_x U(s,X(s))\cdot a(s,X(s)) \\
& + \frac{1}{2}\sum_{j_1,j_2=1}^{d}\frac{\partial^2}{\partial x_{j_1}\partial x_{j_2}}U(s,X(s))\cdot\big(b_{j_1}(s,X(s))\cdot b_{j_2}^T(s,X(s))\big)\Big)\mathrm{d}s \\
& + \int_0^t \nabla_x U(s,X(s))\cdot b(s,X(s))\mathrm{d}W(s) \\
& + \int_0^t \Big(U(s,X(s-)+c(s,X(s-))) - U(s,X(s-))\Big)\mathrm{d}N(s),
\end{aligned}
$$

and the kth component is given by

$$
\begin{aligned}
U_k(t,X(t)) = U_k(0,X(0)) \;+\; & \int_0^t \Big(\frac{\partial}{\partial t}U_k(s,X(s)) + \sum_{i=1}^{d}\frac{\partial}{\partial x_i}U_k(s,X(s))\cdot a_i(s,X(s)) \\
& + \frac{1}{2}\sum_{j_1,j_2=1}^{d}\frac{\partial^2}{\partial x_{j_1}\partial x_{j_2}}U_k(s,X(s)) \\
& \times \sum_{j=1}^{m_{\mathrm{w}}}\big(b^{j_1,j}(s,X(s))\cdot b^{j_2,j}(s,X(s))\big)\Big)\mathrm{d}s \\
+\; & \sum_{j=1}^{m_{\mathrm{w}}}\int_0^t\sum_{i=1}^{d}\frac{\partial}{\partial x_i}U_k(s,X(s))b^{i,j}(s,X(s))\mathrm{d}W_j(s) \\
+\; & \int_0^t \Big(U_k(s,X(s-)+c(s,X(s-))) - U_k(s,X(s-))\Big)\mathrm{d}N(s).
\end{aligned}
$$

A.4. Stochastic differential equations

Followed by [25] we show here basic theorem about existence and uniqueness of solutions of SDEs. At the beginning we have to describe notations used in this section.

Definition A.60 ([27])**.** Sequence f_n *converges to* f *in topology of uniform convergence on compact sets* (ucc topology) if

$$\sup_{t<T}\big\|f_n(t)-f(t)\big\| \to 0 \;\; \forall_{T<\infty}.$$

Definition A.61 ([27])**.** Let $(\mathcal{T}, \tau)$ be a topological space and (Y, d_Y) be a metric space. A sequence of functions $f_n : \mathcal{T} \to Y$, $n \in \mathbb{N}$, *converge compactly* as $n \to \infty$ to some function $f : \mathcal{T} \to Y$ if, for every compact set $K \subseteq \mathcal{T}$,

$$\lim_{n\to\infty} \sup_{x \in K} \mathrm{d}_Y\big(f_n(x), f(x)\big) = 0.$$

By $\mathbb{D}_d$ we define space of function with topology of uniform convergence defined as follows

$$\mathbb{D}_d = \mathbb{D}\big([0,\infty), \mathbb{R}^d\big) = \big\{f : [0,\infty) \to \mathbb{R}^d \mid f - is\ càdlàg\big\}.$$

Let $Y_1, Y_2, \ldots, Y_k$ be a càdlàg semi-martingales with the respect to filtration $\big\{\mathcal{F}_t\big\}_{t\geq 0}$. We will consider SDEs of the following form

$$\begin{cases} \mathrm{d}X(t) &= \tilde{z}(\,\cdot\,, t, X)\mathrm{d}Y(t) \\ X(0) &= \xi_0, \end{cases}$$

where the functional $\tilde{z}$ is given as follows. Let $\mathcal{B}(\mathbb{D}_d)$ be the smallest σ-algebra defined on $\mathbb{D}_d$ under which *coordinate mappings* θ_t given by

$$\theta_t(\gamma) = \gamma(t), \quad \gamma \in \mathbb{D}_d, \quad 0 \leqslant t < \infty,$$

are measurable, $(\mathcal{B}(\mathbb{D}_d) = \sigma(\theta_t : 0 \leqslant t < \infty))$.

Let

$$z : \Omega \times [0,\infty) \times \mathbb{D}_d \to \mathbb{R}^{d\times k},$$

be such that for all $t \in [0,\infty)$,

$$(\omega, \gamma) \mapsto z(\omega, t, \gamma)\ is\ \mathcal{F}_t \otimes \mathcal{B}(\mathbb{D}_d) - measurable. \tag{A.3}$$

For all $(\omega, \gamma) \in \Omega \times \mathbb{D}_d$,

$$t \mapsto z(\omega, t, \gamma)\ is\ an\ càdlàg\ mapping. \tag{A.4}$$

Suppose that there is an increasing càdlàg adapted process κ such that for all $\gamma, \gamma_1, \gamma_2 \in \mathbb{D}_d$,

$$\sup_{0\leq s\leq t} \|z(\omega, s, \gamma)\| \leqslant \kappa(\omega, t) \sup_{0\leq s\leq t} \big(1 + \|\gamma(s)\|\big), \tag{A.5}$$

$$\sup_{0\leq s\leq t} \|z(\omega, s, \gamma_1) - z(\omega, s, \gamma_2)\| \leqslant \kappa(\omega, t) \sup_{0\leq s\leq t} \|\gamma_1(s) - \gamma_2(s)\|. \tag{A.6}$$

Let $\tilde{z} : \Omega \times [0,\infty) \times \mathbb{D}_d \to \mathbb{R}^{d\times k}$ be given by

$$\tilde{z}(\omega, s, \gamma) = z(\omega, s-, \gamma). \tag{A.7}$$

The Theorem A.62 gives us knowledge about existence and uniqueness of the solutions of stochastic differential equations with respect to multidimensional semi-martingale.

Theorem A.62 ([25])**.** *Let* $Y_1, Y_2, \dots, Y_k$ *be a càdlàg semi-martingales with the respect to filtration* $\{\mathcal{F}_t\}_{t\geq 0}$*,* $Y = (Y_1, Y_2, \dots, Y_k)^T$*. Let* z *satisfy assumptions* (A.3) – (A.6) *and let* $\tilde{z}$ *be defined by* (A.7)*. Let* H *be an adapted càdlàg process. Then there exists an adapted càdlàg process* X *such that*

$$X(t) = H(t) + \int_{0+}^{t} \tilde{z}(\,\cdot\,, s, X)\mathrm{d}Y(s).$$

Now let us check that the definition in (1.1) under considered in thesis assumption has unique solutions. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Let $W(t) = (W_1(t), \dots, W_{m_{\mathrm{w}}}(t))^T$ be an m_{w}-dimensional Brownian motion defined on that space and $N(t)$ be one-dimensional Poisson process. Let $0 < T < +\infty$, $x_0 \in \mathbb{R}^d$. Let $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m_{\mathrm{w}}}$ and $c : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ be Borel measurable functions and satisfy the following assumptions.

For function $f \in \{a, b, c\}$, exists $K > 0$ such that

($\tilde{A}$) $f \in C([0,T] \times \mathbb{R}^d)$,

($\tilde{B}$) for all $t, s \in [0,T]$ and all $y, z \in \mathbb{R}^d$

  ($\tilde{B}1$) $\|f(t,y) - f(t,z)\| \leqslant K\|y - z\|$.

Consider the d-dimensional stochastic differential equation of Itô type

$$\mathrm{d}X(t) = a(t, X(t))\mathrm{d}t + b(t, X(t))\mathrm{d}W(t) + c(t, X(t-))\mathrm{d}N(t) \quad \text{on} \quad 0 \leqslant t \leqslant T,$$

with the initial value $X(0) = x_0$. This equation is the notion for the following stochastic integral equation

$$X(t) = x_0 + \int_0^t a(s, X(s))\mathrm{d}s + \int_0^t b(s, X(s))\mathrm{d}W(s) + \int_0^t c(s, X(s-))\mathrm{d}N(s) \quad t \in [0,T]. \tag{A.8}$$

We refer to a, b and c as to *drift*, *diffusion* and *jump* coefficients, respectively. Problem (A.8) can be rewritten as an SDE driven by the multidimensional semi-martingale $Y = (t, W, N)^T = (t, W_1, \dots, W_{m_{\mathrm{w}}}, N)^T$

$$X(t) = x_0 + \int_0^t F(s, X(s-))\mathrm{d}Y(s), \quad t \in [0,T], \tag{A.9}$$

where

$$F(t,y) = \Big(F^{i,j}(t,y)\Big)_{1\le i\le d,\ 1\le j\le 2+m_{\rm w}} = \begin{pmatrix} a_1 & b^{1,1} & b^{1,2} & \cdots & b^{1,m_{\rm w}} & c_1 \\ a_2 & b^{2,1} & b^{2,2} & \cdots & b^{2,m_{\rm w}} & c_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_d & b^{d,1} & b^{d,2} & \cdots & b^{d,m_{\rm w}} & c_d \end{pmatrix}(t,y). \tag{A.10}$$

Defined function $F : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times(2+m_{\rm w})}$. Under the assumptions $(\tilde{A})$, $(\tilde{B})$ we have that $F \in C([0,T]\times\mathbb{R}^d)$ and for all $t\in[0,T]$, $x,y\in\mathbb{R}^d$,

$$\|F(t,x) - F(t,y)\| \leqslant K\|x-y\|.$$

Let us define function $z : [0,\infty)\times\mathbb{D}_d \to \mathbb{R}^{d\times(2+m_{\rm w})}$

$$z(t,\gamma) = F(t,\gamma(t)),$$

and then

$$\tilde{z}(t,\gamma) = z(t-,\gamma) = F(t-,\gamma(t-)) = F(t,\gamma(t-)).$$

Now, let us check that the defined function z satisfies the conditions (A.3) – (A.6).

Let's start with explaining (A.3). For $t\leqslant T$ we have that mapping $\gamma\to F(t,\gamma(t))$ is $\mathcal{B}(\mathbb{D}_d)$-measurable. By assumption $(A_{\rm MD})$, $F(t,\ \cdot\ )$ is continuous and it is also $\mathcal{F}_t$-measurable. By the definition of $\mathcal{B}(\mathbb{D}_d)$ coordinate mapping is $\mathcal{B}(\mathbb{D}_d)$-measurable. Moreover, mapping $F(t,\gamma(t))$ does not depend on ω. Combining it together we have that $(\omega,\gamma)\to F(t,\gamma(t))$ is $\mathcal{F}_t\otimes\mathcal{B}(\mathbb{D}_d)$-measurable.

To show (A.4), let $(\omega,\gamma)\in\Omega\times\mathbb{D}_d$. We have that mapping $t\to F(t,\gamma(t))$ is càdlàg as a submission, because F is continuous and γ is càdlàg.

By Lemma B.2 we have that

$$\begin{aligned}\sup_{0\le s\le t}\|z(\omega,s,\gamma)\| &\leqslant \sup_{0\le s\le t}\big\|F(s,\gamma(s))\big\| \\ &\leqslant \sup_{0\le s\le t}\Big(\sum_{i=1}^{d}\sum_{j=1}^{2+m_{\rm w}}\|F^{i,j}(s,\gamma(s))\|^2\Big)^{1/2} \\ &\leqslant \kappa(t,\omega)\sup_{0\le s\le t}\big(1+\|\gamma(s)\|\big),\end{aligned}$$

which proofs that definition of z satisfies condition (A.5).

By assumption $(\tilde{B}1)$ we have that

$$\begin{aligned}\sup_{0\le s\le t}\|z(\omega,s,\gamma_1)-z(\omega,s,\gamma_2)\| &\leqslant \sup_{0\le s\le t}\big\|F(s,\gamma_1(s))-F(s,\gamma_2(s))\big\| \\ &\leqslant \kappa(t,\omega)\sup_{0\le s\le t}\big\|\gamma_1(s)-\gamma_2(s)\big\|,\end{aligned}$$

which proofs that definition of z satisfies condition (A.6).

In Chapters 2 and 3 we use stronger assumptions so the existence and uniqueness also holds. By the above considerations and Theorem A.62 there exists a unique solution $X(t)$ of the equation (A.8), and the solution belongs to $\mathcal{M}^2([0,T];\mathbb{R}^d)$.

We also have the following estimates for the solution X under the additional assumptions $(\tilde{B}2), (\tilde{E})$. For function $f \in \{a,b,c\}$, and λ exists $K>0$ such that

$(\tilde{B})$ for all $t,s \in [0,T]$ and all $y,z \in \mathbb{R}^d$,

$(\tilde{B}2)$ $\|f(t,y) - f(s,y)\| \leqslant K(1+\|y\|)|t-s|$,

$(\tilde{E})$ $\lambda \in C([0,T])$.

Lemma A.63 ([62])**.** *Let us assume that the functions a, b, c and λ satisfy the assumptions $(\tilde{A}), (\tilde{B}1), (\tilde{B}2)$ and $(\tilde{E})$. Then there exists positive constants C_1, C_2 such that*

$$\Big\| \sup_{t\in[0,T]} \|X(t)\| \Big\|_{\mathfrak{L}^4(\Omega)} \leqslant C_1,$$

and for all $t,s \in [0,T]$

$$\big\| X(t) - X(s) \big\|_{\mathfrak{L}^2(\Omega)} \leqslant C_2 |t-s|^{1/2}.$$

A.5. Random elements with values in Banach spaces

In stochastic analysis it is important to have tools which allow us to switch between two possible ways of looking at a stochastic process. Firstly, we can consider a stochastic process as a product measurable function

$$\Omega \times [0,+\infty) \ni (\omega,t) \to X(\omega,t) \in \mathbb{R}^d.$$

On the other hand, if almost all trajectories of process X belongs to $(E,\mathcal{E})$ where $(E,\mathcal{E})$ is some functional space equipped with a σ-algebra $\mathcal{E}$, we can consider mapping $\hat{X} : \Omega \to E$ defined in the following way

$$\hat{X}(\omega) = X(\omega, \,\cdot\,), \quad \omega \in \Omega. \tag{A.11}$$

If $\hat{X}$ is $\mathcal{F}/\mathcal{E}$-measurable then we say that the process X generates the random element $\hat{X}$ in $(E,\mathcal{E})$. Moreover, the law of μ of $\hat{X}$ is a probabilistic measure induced by $\hat{X}$ in the measurable space $(E,\mathcal{E})$.

Theorem A.64 ([11])**.** *Let $E = \mathfrak{L}^2([0,T])$, equipped with the norm $\|\cdot\|_{\mathfrak{L}^2([0,T])}$, and $\mathcal{E} = \mathcal{B}(E)$. If $X \in \mathfrak{L}^2(\Omega\times[0,T], \mathcal{F}\otimes\mathcal{B}([0,T]), \mathbb{P}\times\lambda_1)$, then it generates the random element $\hat{X}$ in $\mathfrak{L}^2([0,T])$.*

Having a random element $\bar{X}$ in some functional space E it is natural to ask if there is a product measurable stochastic process X satisfying (A.11). In the case when $E = \mathfrak{L}^2([0,T])$ the answer is provided by the Theorem A.65.

Theorem A.65 ([37] Proposition 2, page 741)**.** *Let $\bar{X}$ be a random element in $\mathfrak{L}^2([0,T])$. Then there exists a product measurable process X such that for almost all ω, the equality $X(\omega,t) = (\bar{X}(\omega))(t)$ holds almost everywhere on $[0,T]$.*

Note that $X \in \mathfrak{L}^2(\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}([0,T]), \mathbb{P} \times \lambda_1)$. Hence by Theorem A.64 the solution X of the SDE (1.1) generates their representation element $\hat{X} :\to \mathfrak{L}^2([0,T];\mathbb{R}^d)$ which is $\mathcal{F}/\mathcal{B}(\mathfrak{L}^2([0,T];\mathbb{R}^d))$-measurable. In the thesis we use tha same symbol X for the product measurable solution of SDE (1.1) as for it representation as a random element in $\mathfrak{L}^2([0,T];\mathbb{R}^d)$.

By (1.5) and (1.6) we have that $\bar{X}_n : \Omega \to \mathfrak{L}^2([0,T];\mathbb{R}^d)$ is a $\sigma(\mathcal{N}_n(W,N))/\mathcal{B}(\mathfrak{L}^2([0,T];\mathbb{R}^d))$-measurable random element in $\mathfrak{L}^2([0,T];\mathbb{R}^d)$.

By Theorem A.65 there exists $\sigma(\mathcal{N}_n(W,N)) \otimes \mathcal{B}([0,T])/\mathcal{B}(\mathbb{R}^d)$-measurable process $\hat{X}_n$ such that for almost all $\omega \in \Omega$

$$\hat{X}_n(\omega,t) = \big(\bar{X}_n(\omega)\big)(t)$$

holds for almost all $t \in [0,T]$. In particular, this implies that for almost all $t \in [0,T]$ the random variable

$$\Omega \ni \bar{X}_n(\,\cdot\,,t) \to \mathbb{R}^d$$

is $\sigma(\mathcal{N}_n(W,N))$-measurable. Again, we do not distinguish between $\bar{X}_n$ and $\hat{X}_n$.

A.6. Auxiliary results

A.6.1. Properties of Frobenius norm

Let $A = [a^{i,j}]_{i,j=1}^{d,k}$ be the $d \times k$ real matrix. Then *the Frobenius* norm of A is defined as

$$\|A\| = \left(\sum_{i=1}^{d}\sum_{j=1}^{k} |a^{i,j}|^2\right)^{1/2}.$$

In the special case, when x is a vector of length d,

$$\|x\| = \left(\sum_{i=1}^{d} |x_i|^2\right)^{1/2}$$

is just the Euclidean vector norm.

Now, if we denote jth column of A by a^j, and ith row of A by a_i , then the norm can be expressed as

$$\|A\| = \left(\sum_{j=1}^{m} \|a^j\|^2\right)^{1/2} = \left(\sum_{j=1}^{d} \|a_i\|^2\right)^{1/2}.$$

The Frobenius norm has a useful property of submultiplicity.

Lemma A.66 ([81]). *Let* $A = [a^{i,j}]_{i,j=1}^{d,m}$ *and* $B = [b^{i,j}]_{i,j=1}^{m,k}$ *be matrices of sizes* $d \times m$ *and* $m \times k$*, respectively. Then the the product* $C = [c^{i,j}]_{i,j=1}^{d,k}$ *of matrices* A *and* B *is an* $d \times k$ *matrix and*

$$\|C\| = \|AB\| \leqslant \|A\|\|B\|.$$

A.6.2. Grönwall's inequality

Theorem A.67 ([38]). *Let* $T > 0$ *and* $c \geqslant 0$*. Let* $u(\cdot)$ *be a Borel measurable, bounded, and nonnegative function on* $[0, T]$*, and let* $v(\cdot)$ *be a nonnegative integrable function on* $[0, T]$*. If*

$$u(t) \leqslant c + \int_0^t v(s)u(s)\mathrm{d}s \quad \textit{for all} \quad t \in [0, T],$$

then

$$u(t) \leqslant c \cdot \exp\left(\int_0^t v(s)\mathrm{d}s\right) \quad \textit{for all} \quad t \in [0, T].$$

Appendix B

Time-continuous Milstein approximation

In this section we show basic properties about time continuous Milstein approximation which is used in this thesis. We will provide definition of approximation in two cases, first which use information about first derivative, and second which corresponds to derivative free version of Milstein scheme. Then we prove main theorems which say about the rate of convergence of both algorithms. We also show here the useful Lemmas and Facts which help to prove Theorem B.1 and Theorem B.13.

B.1. Time-continuous Milstein approximation for system of SDEs

Let $m \in \mathbb{N}$ and

$$0 = t_0 < t_1 < \ldots < t_m = T, \tag{B.1}$$

be an arbitrary discretization of interval $[0, T]$. By

$$\Delta Z_i = Z(t_{i+1}) - Z(t_i),$$

we denote the increment of stochastic processes $Z \in \{N, W, W_1, \ldots, W_{m_{\mathrm{w}}}\}$, where $i = 0, 1, \ldots, m-1$, it is both a vector or a number depending on process structure.

Followed by [61] *the time-continuous Milstein approximation* $\tilde{X}_m^M = \big\{\tilde{X}_m^M(t)\big\}_{t\in[0,T]}$ based on the discretization (B.1) is defined as follows. We set

$$\tilde{X}_m^M(0) = x_0, \tag{B.2}$$

and for $t \in [t_i, t_{i+1}]$, $i = 0, 1, \ldots, m-1$,

$$\begin{aligned}\tilde{X}_m^M(t) = \tilde{X}_m^M(t_i) &+ a(U_i)\cdot(t-t_i) + b(U_i)\cdot\big(W(t)-W(t_i)\big)\\ &+ c(U_i)\cdot\big(N(t)-N(t_i)\big) + \sum_{j_1,j_2=1}^{m_{\mathrm{w}}} L_{j_1}b^{j_2}(U_i)\cdot I_{t_i,t}(W_{j_1},W_{j_2})\\ &+ \sum_{j_1=1}^{m_{\mathrm{w}}} L_{j_1}c(U_i)\cdot I_{t_i,t}(W_{j_1},N) + \sum_{j_1=1}^{m_{\mathrm{w}}} L_{-1}b^{j_1}(U_i)\cdot I_{t_i,t}(N,W_{j_1})\\ &+ L_{-1}c(U_i)\cdot I_{t_i,t}(N,N),\end{aligned} \tag{B.3}$$

where $U_i = \big(t_i, \tilde{X}_m^M(t_i)\big)$ and multiple stochastic integrals defined as

$$I_{t_i,t}(Y,Z) = \int_{t_i}^{t}\int_{t_i}^{s-} dY(u)dZ(s), \tag{B.4}$$

for $Y, Z \in \big\{N, W_1, \ldots, W_{m_{\mathrm{w}}}\big\}$. For more properties about multiple integration we refer to Appendix B.4, where we consider basic properties about multiple stochastic integrals in a way when partial information about processes are known.

We stress that for any $m \in \mathbb{N}$ the approximation $\big\{\tilde{X}_m^M(t)\big\}_{t\in[0,T]}$, in our model of computation (even under the commutative conditions (D_{MD})), is not an implementable numerical scheme, since computation of a trajectory of $\tilde{X}_m^M$ requires complete knowledge of a corresponding trajectories of N and W. However, if the conditions (D_{MD}) holds, by Lemma B.23, we can compute values of $\tilde{X}_m^M$ at the discrete points (B.1) using only function evaluations of W and N at (B.1).

For every $m \in \mathbb{N}$ the process $\big\{\tilde{X}_m^M(t)\big\}_{t\in[0,T]}$ is adapted to $\big\{\mathcal{F}_t\big\}_{t\in[0,T]}$ and has càdlàg paths. Furthermore, under the commutative conditions (D_{MD}) the random variables $\big\{\tilde{X}_m^M(t_i)\big\}_{i=0}^m$ are measurable with respect to the σ-algebra

$$\sigma\big(\mathcal{N}_m(N,W)\big) = \sigma\big(N(t_1), N(t_2), \ldots, N(t_m), W(t_1), W(t_2), \ldots, W(t_m)\big), \tag{B.5}$$

and the upper bound on the error of $\tilde{X}_m^M$ is given by Theorem B.1. We provide an auxiliary result concerning an upper bound on the error for the continuous Milstein approximation $\tilde{X}_m^M$. A similar result has been justified in Theorem 6.4.1 in [61], however, under slightly stronger assumptions. In particular, in this thesis we do not assume the existence of continuous partial derivative $\partial f/\partial t$ for $f \in \{a,b,c\}$ and we do not assume any Lipschitz conditions for the second order partial derivatives of $f = f(t,y), f \in \{a,b,c\}$, with respect to y. Moreover, we consider non-homogeneous Poisson process, while in [61] in Theorem 6.4.1 has been shown only for homogeneous counting processes.

Theorem B.1. *Let us assume that the mappings a, b, c and λ satisfy assumptions (A_{MD}) – (C_{MD}) and (E_{MD}). Let $m \in \mathbb{N}$ and let* (B.1) *be an arbitrary discretization of the interval* $[0,T]$. *Then for continuous Milstein approximation $\tilde{X}_m^M$, based on the mesh* (B.1) *we have that*

$$\sup_{t\in[0,T]} \big\|\tilde{X}_m^M(t)\big\|_{\mathfrak{L}^2(\Omega)} \leqslant C_1, \tag{B.6}$$

and

$$\sup_{t\in[0,T]} \big\|X(t) - \tilde{X}_m^M(t)\big\|_{\mathfrak{L}^2(\Omega)} \leqslant C_2 \max_{0\leq i\leq m-1}(t_{i+1}-t_i), \tag{B.7}$$

where $C_1, C_2 > 0$ do not depend on m.

As a proof of Theorem B.1 is long we decide to divide it into smaller parts. We also proof some lemmas, which are repeatable in the main proof. Firstly we show results following from the given assumptions (A_{MD}) – (C_{MD}).

Lemma B.2. *Let $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfy (A_{MD}) – (B_{MD}) then for all $(t,y) \in [0,T] \times \mathbb{R}^d$, exist $K_1 > 0$ depends only on $\|f(0,0)\|$, K and T such that*

$$\|f(t,y)\| \leqslant K_1\big(1 + \|y\|\big), \tag{B.8}$$

$$\Big\|\frac{\partial^{|\alpha|} f}{\partial y^\alpha}(t,y)\Big\| \leqslant K, \quad |\alpha| = 1,2, \tag{B.9}$$

where $\alpha \in \mathbb{N}_0^d$, and $|\alpha| = \sum_{k=1}^d \alpha_k$. We also have that

$$\big\|\nabla_x f(t,y)\big\| \leqslant d \cdot K. \tag{B.10}$$

Moreover if function f satisfy assumption (C_{MD}) we have that for all $(t,y) \in [0,T] \times \mathbb{R}^d$

$$\max\big\{\|L_{-1}f(t,y)\|, \|L_1 f(t,y)\|, \ldots, \|L_{m_{\mathrm{w}}} f(t,y)\|\big\} \leqslant K_2\big(1 + \|y\|\big), \tag{B.11}$$

with $K_2 = KK_1$. (For $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m_{\mathrm{w}}}$ the statement (B.8) *also holds.)*

Proof. Firstly we show (B.8). By Cauchy-Schwarz inequality and by assumption (B_{MD}) we have that

$$\begin{aligned} \|f(t,y)\| &\leqslant \|f(t,y)-f(0,y)\| + \|f(0,y)-f(0,0)\| + \|f(0,0)\| \\ &\leqslant K_1\big(1+\|y\|\big). \end{aligned}$$

Let $e_k = (0,\dots,0,1,0,\dots,0)^T$ for $k \in \{1,\dots,d\}$ be a d-dimensional vector where non-zero element is on ith position. We have that $\|e_i\| = 1$. Then we go to proof (B.9) in case when $|\alpha| = 1$. By assumptions ($B1_{\text{MD}}$) we have that

$$\left\| \frac{\partial f}{\partial y_k}(t,y) \right\| = \lim_{h\to 0} \left\| \frac{f(t,y+h\cdot e_k) - f(t,y)}{h} \right\| \leqslant \lim_{h\to 0} \frac{K\|h\cdot e_k\|}{|h|} \leqslant K.$$

Now, we go to proof of (B.9) when $|\alpha| = 2$. For $k_1, k_2 \in \{1,\dots,d\}$ it follow that

$$\left\| \frac{\partial^2 f}{\partial y_{k_1}\partial y_{k_2}}(t,y) \right\| = \lim_{h\to 0} \left\| \frac{\frac{\partial f}{\partial y_{k_1}}(t,y+h\cdot e_{k_2}) - \frac{\partial f}{\partial y_{k_1}} f(t,y)}{h} \right\| \leqslant \lim_{h\to 0} \frac{K\|h\cdot e_{k_2}\|}{|h|} \leqslant K.$$

The (B.10) is a natural consequence of (B.9). Finally we prove (B.11). Hence by ($B1_{\text{MD}}$) and (B.8) we have

$$\begin{aligned} \|L_{-1}f(t,y)\| &= \|f(t,y+c(t,y)) - f(t,y)\| \\ &\leqslant K\|y + c(t,y) - y\| \leqslant KK_1\big(1+\|y\|\big). \end{aligned}$$

Then, directly from (B.8) for $j \in \{1,\dots,m_{\text{w}}\}$ we have

$$\|L_j f(t,y)\| \leqslant \big\|\nabla_x f(t,y)\big\| \cdot \big\|b^j(t,y)\big\| \leqslant KK_1\big(1+\|y\|\big).$$

This ends the proof. ■

Let $f \in \{a, b^1, \dots, b^{m_{\text{w}}}, c\}$ and (B.1) be a discretization of interval $[0,T]$. Let $u \in [t_i, t_{i+1}]$ for $i \in \{1,\dots,m\}$, then we can define functions $\alpha_i, \beta_i, \gamma_i$ by

$$\begin{aligned} \alpha_i(f,u) &:= \nabla_x f(t_i, X(u)) \cdot a(u, X(u)) \\ &\quad + \frac{1}{2} \sum_{j_1,j_2=1}^{d} \frac{\partial f}{\partial x_{j_1}\partial x_{j_2}}(t_i, X(u)) \cdot \big(b_{j_1}(u,X(u)) \cdot b_{j_2}^T(u,X(u))\big), \quad \text{(B.12)} \\ \beta_i(f,u) &:= \nabla_x f(t_i, X(u)) \cdot b(u, X(u)), \quad \text{(B.13)} \\ \gamma_i(f,u) &:= f(t_i, X(u-) + c(u, X(u-))) - f(t_i, X(u)). \quad \text{(B.14)} \end{aligned}$$

By Lemma A.59 (Itô formula) applied to function $U(x) = f(t_i, x)$, by Definition A.39, (B.12) – (B.14) we can write that

$$f(t_i, X(s)) - f(t_i, X(t_i)) = \int_{t_i}^{s} \alpha_i(f,u)\mathrm{d}u + \int_{t_i}^{s} \beta_i(f,u)\mathrm{d}W(u) + \int_{t_i}^{s} \gamma_i(f,u)\mathrm{d}N(u). \tag{B.15}$$

Based on given assumptions about function f (A_{MD}) – (C_{MD}), we have the following estimations.

Lemma B.3. *For $i \in \{1, \ldots, m\}$, for all $u \in (t_i, t_{i+1}]$ we have that*

$$\big\|\alpha_i(f,u)\big\|^2 \leqslant C\big(1 + \|X(u)\|\big)^4. \tag{B.16}$$

Proof. By the Lemma A.66 we have that

$$\begin{aligned} \big\|\alpha_i(f,u)\big\|^2 &\leqslant \big\|\nabla_x f(t_i, X(u)) \cdot a(u, X(u))\big\|^2 \\ &+ \frac{1}{2}\, C \sum_{j_1,j_2=1}^{d} \Big\|\frac{\partial f}{\partial x_{j_1} \partial x_{j_2}}(u, X(u)) \cdot b_{j_1}(u, X(u)) \cdot b_{j_2}^T(t_i, X(u))\Big\|^2 \\ &\leqslant \|\nabla_x f(t_i, X(u))\|^2 \cdot \|a(u, X(u))\|^2 \\ &+ \frac{1}{2}\, C \sum_{j_1,j_2=1}^{d} \Big\|\frac{\partial f}{\partial x_{j_1} \partial x_{j_2}}(t_i, X(u))\Big\|^2 \cdot \big(b_{j_1}(u, X(u)) \cdot b_{j_2}^T(u, X(u))\big)^2. \end{aligned}$$

Now, by (B.8), (B.9) and assumption $(B2_{\text{MD}})$ we have (B.16) and this ends the proof. ■

Lemma B.4. *For $i \in \{1, \ldots, m\}$, for all $u \in (t_i, t_{i+1}]$ we have that*

$$\big\|\beta_i(f,u)\big\|^2 \leqslant C\big(1 + \|X(u)\|\big)^2. \tag{B.17}$$

Proof. By the Lemma A.66 we have that

$$\big\|\beta_i(f,u)\big\|^2 \leqslant \big\|\nabla_x f(t_i, X(u))\big\|^2 \cdot \|b(u, X(u))\|^2.$$

Now, by (B.8), (B.9) we have (B.17) and this ends the proof. ■

Lemma B.5. *For $i \in \{1, \ldots, m\}$, for all $u \in (t_i, t_{i+1}]$ we have that*

$$\big\|\gamma_i(f,u)\big\|^2 \leqslant C\big(1 + \|X(u-)\|\big)^2. \tag{B.18}$$

Proof. By assumption ($B1_{\text{MD}}$) we have that

$$\|\gamma_i(f,u)\|^2 \leqslant \|c(u,X(u-))\|^2.$$

Now, by (B.8) we have (B.18) and this ends the proof. ■

Lemma B.6. *For $i \in \{1,\ldots,m\}$, for all $u \in (t_i,t_{i+1}]$, for $f \in \{b^1,\ldots,b^{m_{\text{w}}},c\}$ for $k \in \{1,2,\ldots,m_{\text{w}}\}$, $U_i = \big(t_i,\tilde{X}^M_m(t_i)\big)$ we have that*

$$\mathbb{E}\big\|(\beta_i(f,u))^k - L_k f(U_i)\big\|^2 \leqslant C\,\Big(\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2 + (u-t_i) + (u-t_i)^2\Big).$$

Proof. By the assumption (C_{MD}), Theorem A.63, (B.8), we have that

$$\begin{aligned}
\mathbb{E}\big\|\big(\beta_i(f,u)\big)^k - L_k f(U_i)\big\|^2 &\leqslant \mathbb{E}\big\|\big(\beta_i(f,u)\big)^k - L_k f(t_i,X(u))\big\|^2 \\
&\qquad + \mathbb{E}\big\|L_k f(t_i,X(u)) - L_k f(U_i)\big\|^2 \\
&\leqslant \mathbb{E}\big(\|\nabla_x f(t_i,X(u))\|^2\cdot\|b^k(u,X(u)) - b^k(t_i,X(u))\|^2\big) \\
&\qquad + K\ \mathbb{E}\big\|X(u)-\tilde{X}^M_m(t_i)\big\|^2 \\
&\leqslant K\mathbb{E}\big(1+\|X(u)\|\big)^2\cdot|u-t_i|^2 \\
&\qquad + K\mathbb{E}\big\|X(u)-X(t_i)\big\|^2 + K\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2 \\
&\leqslant C\big(\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2 + (u-t_i) + (u-t_i)^2\big).
\end{aligned}$$

That ends the proof. ■

Lemma B.7. *For $i \in \{1,\ldots,m\}$, for all $u \in [t_i,t_{i+1}]$ we have that for $f \in \{b^1,\ldots,b^{m_{\text{w}}},c\}$ for $k \in \{1,2,\ldots,m_{\text{w}}\}$, $U_i = \big(t_i,\tilde{X}^M_m(t_i)\big)$*

$$\mathbb{E}\int_{t_i}^{s} \big\|\gamma_i(f,u) - L_{-1}f(U_i)\big\|^2 \mathrm{d}u \leqslant C\Big(\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2 + (t_{i+1}-t_i)^2\Big).$$

Proof. Firstly we show estimation for $\big\|\gamma_i(f,u) - L_{-1}f(U_i)\big\|^2$ for all $u \in [t_i,t_{i+1}]$ By the assumption ($B1_{\text{MD}}$), we have that

$$\begin{aligned}
\big\|\gamma_i(f,u) - L_{-1}f(U_i)\big\| &\leqslant \big\|f(t_i,X(u-)+c(u,X(u-))) - f(t_i,\tilde{X}^M_m(t_i)+c(U_i))\big\| \\
&\qquad + \big\|f(t_i,X(u)) - f(U_i)\big\| \\
&\leqslant C\big\|X(u-)-\tilde{X}^M_m(t_i)\big\| + C\big\|c(u,X(u-)) - c(U_i)\big\|
\end{aligned}$$

$$
\begin{aligned}
&\leqslant C\big\|X(u-) - X(t_i)\big\| + \big\|X(t_i) - \tilde{X}_m^M(t_i)\big\| \\
&\quad + C\Big(\mathbb{E}\big\|c(u, X(u-)) - c(t_i, X(u-))\big\| \\
&\quad + \big\|c(t_i, X(u-)) - c(U_i)\big\|\Big) \\
&\leqslant C\big(\big\|X(t_i) - \tilde{X}_m^M(t_i)\big\| \\
&\quad + \big\|X(u-) - X(t_i)\big\| + \big(1 + \|X(u-)\|\big) \cdot |u - t_i|\big).
\end{aligned}
$$

Then we have that

$$
\begin{aligned}
\mathbb{E}\int_{t_i}^{s} \big\|\gamma_i(f,u) - L_{-1}f(U_i)\big\|^2 \mathrm{d}u &\leqslant C\mathbb{E}\int_{t_i}^{s} \big\|X(u-) - X(t_i)\big\|^2 \mathrm{d}u \\
&\quad + C\mathbb{E}\big\|X(t_i) - \tilde{X}_m^M(t_i)\big\|^2 (s - t_i) \\
&\quad + C\mathbb{E}\int_{t_i}^{s} \big(1 + \|X(u-)\|^2\big) \cdot (u - t_i)^2 \mathrm{d}u \\
&\leqslant C\mathbb{E}\int_{t_i}^{s} (u - t_i)\mathrm{d}u + \; CT\mathbb{E}\big\|X(t_i) - \tilde{X}_m^M(t_i)\big\|^2 \\
&\quad + C\big(1 + \sup_{0\leq t\leq T} \mathbb{E}\|X(t)\|^2\big) \cdot \frac{1}{3}(s - t_i)^3.
\end{aligned}
$$

By Theorem A.63,

$$
\mathbb{E}\int_{t_i}^{s} \big\|\gamma_i(f,u) - L_{-1}f(U_i)\big\|^2 \mathrm{d}u \;\leqslant\; C(t_{i+1} - t_i)^2 + \; C\mathbb{E}\big\|X(t_i) - \tilde{X}_m^M(t_i)\big\|^2.
$$

This ends the proof. ■

The solutions of problem (1.1) given by $X = X(t)$ and time continuous Milstein approximation $\tilde{X}_m^M = \tilde{X}_m^M(t)$ can be decomposed into

$$
X(t) = x_0 + A(t) + B(t) + C(t), \tag{B.19}
$$

$$
\tilde{X}_m^M(t) = x_0 + \tilde{A}_m^M(t) + \tilde{B}_m^M(t) + \tilde{C}_m^M(t), \tag{B.20}
$$

where

$$A(t) = \int_0^t \sum_{i=0}^{m-1} a(s,X(s))\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s, \tag{B.21}$$

$$B(t) = \int_0^t \sum_{i=0}^{m-1} b(s,X(s))\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W(s), \tag{B.22}$$

$$C(t) = \int_0^t \sum_{i=0}^{m-1} c(s,X(s))\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}N(s), \tag{B.23}$$

and

$$\tilde{A}_m^M(t) = \int_0^t \sum_{i=0}^{m-1} a(U_i)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s, \tag{B.24}$$

$$\begin{aligned}\tilde{B}_m^M(t) = \sum_{j=1}^{m_{\mathrm{w}}}\Big(\int_0^t \sum_{i=0}^{m-1} \Big(b^j(U_i) + \sum_{k=1}^{m_{\mathrm{w}}}\int_{t_i}^{s} L_k b^j(U_i)\mathrm{d}W_k(u) \\ + \int_{t_i}^{s} L_{-1}b^j(U_i)\mathrm{d}N(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W_j(s)\Big),\end{aligned} \tag{B.25}$$

$$\begin{aligned}\tilde{C}_m^M(t) = \int_0^t \sum_{i=0}^{m-1} \Big(c(U_i) + \sum_{j=1}^{m_{\mathrm{w}}}\int_{t_i}^{s} L_j c(U_i)\mathrm{d}W_j(u) \\ + \int_{t_i}^{s} L_{-1}c(U_i)\mathrm{d}N(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}N(s).\end{aligned} \tag{B.26}$$

Lemma B.8. *Let us assume that the mappings a, b, c and λ satisfy assumptions* (A_{MD}) – (C_{MD}) *and* (E_{MD}). *Let $m \in \mathbb{N}$ and let* (B.1) *be an arbitrary discretization of the interval* $[0,T]$. *Let $A(t)$ and $\tilde{A}_m^M(t)$ are given by* (B.21), (B.24). *For all $t \in [0,T]$ we have that*

$$\mathbb{E}\big\|A(t)-\tilde{A}_m^M(t)\big\|^2 \leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}_m^M(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \tag{B.27}$$

Proof. W have that for all $t \in [0,T]$

$$\begin{aligned}\mathbb{E}\big\|A(t)-\tilde{A}_m^M(t)\big\|^2 &\leqslant \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\big(a(s,X(s))-a(U_i)\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2 \\ &\leqslant 3\Big(\mathbb{E}\big\|\tilde{A}_{m,1}^M(t)\big\|^2 + \mathbb{E}\big\|\tilde{A}_{m,2}^M(t)\big\|^2 + \mathbb{E}\big\|\tilde{A}_{m,3}^M(t)\big\|^2\Big),\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}\big\|\tilde{A}^M_{m,1}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\big(a(s,X(s))-a(t_i,X(s))\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2,\\
\mathbb{E}\big\|\tilde{A}^M_{m,2}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\big(a(t_i,X(s))-a(t_i,X(t_i))\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2,\\
\mathbb{E}\big\|\tilde{A}^M_{m,3}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\big(a(t_i,X(t_i))-a(t_i,\tilde{X}^M_m(t_i))\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2.
\end{aligned}$$

Now, by a Hölder inequality, Lemma A.63 and assumption ($B2_{\text{MD}}$) it follows that for all $t \in [0,T]$

$$\begin{aligned}
\mathbb{E}\big\|\tilde{A}^M_{m,1}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\big(a(s,X(s))-a(t_i,X(s))\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2\\
&\leqslant C\sum_{i=0}^{m-1}\mathbb{E}\int_0^t \big\|a(s,X(s))-a(t_i,X(s))\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\
&\leqslant C\sum_{i=0}^{m-1}\int_0^t K^2(s-t_i)^2\cdot\mathbb{E}\big(1+\|X(s)\|\big)^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\
&\leqslant C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \qquad \text{(B.28)}
\end{aligned}$$

By decomposition (B.15) and decomposition of N given by (A.2) we have the following estimation

$$\begin{aligned}
\mathbb{E}\big\|\tilde{A}^M_{m,2}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\big(a(t_i,X(s))-a(t_i,X(t_i))\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2\\
&\leqslant C\Big(\mathbb{E}\big\|\tilde{M}^M_{m,1}(t)\big\|^2+\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2+\mathbb{E}\big\|\tilde{M}^M_{m,3}(t)\big\|^2+\mathbb{E}\big\|\tilde{M}^M_{m,4}(t)\big\|^2\Big),
\end{aligned}$$

where

$$\mathbb{E}\big\|\tilde{M}^M_{m,1}(t)\big\|^2 = \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\Big(\int_{t_i}^s \alpha_i(a,u)\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2,$$

$$\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2 = \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\Big(\int_{t_i}^s \beta_i(a,u)\mathrm{d}W(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2,$$

$$\mathbb{E}\big\|\tilde{M}^M_{m,3}(t)\big\|^2 = \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\Big(\int_{t_i}^s \gamma_i(a,u)\mathrm{d}\tilde{N}(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2,$$

$$\mathbb{E}\big\|\tilde{M}^M_{m,4}(t)\big\|^2 = \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\Big(\int_{t_i}^s \gamma_i(a,u)\lambda(u)\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2.$$

From Hölder inequality and Lemma B.3 we have that

$$\begin{aligned}\mathbb{E}\big\|\tilde{M}^M_{m,1}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\Big(\int_{t_i}^s \alpha_i(a,u)\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2 \\ &\leqslant C\int_0^t \sum_{i=0}^{m-1}(s-t_i)\Big(\int_{t_i}^s \mathbb{E}\big(1+\|X(u)\|\big)^4\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s.\end{aligned}$$

Then, by Theorem A.63 we have that

$$\mathbb{E}\big\|\tilde{M}^M_{m,1}(t)\big\|^2 \leqslant C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \tag{B.29}$$

By the definition of Euclidean norm we have that

$$\begin{aligned}\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2 &= \mathbb{E}\sum_{k=1}^d\Big(\int_0^t \sum_{i=0}^{m-1}\Big(\int_{t_i}^s \big(\beta_i(a,u)\big)_k\mathrm{d}W(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big)^2 \\ &= \mathbb{E}\sum_{k=1}^d\Big(\int_0^t \sum_{i_1=0}^{m-1}\Big(\int_{t_{i_1}}^{s_1}\sum_{j_1=1}^{m_\mathrm{w}} \big(\beta_i(a,u)\big)^{k,j_1}\mathrm{d}W_{j_1}(u)\Big)\mathbb{1}_{(t_{i_1},t_{i_1+1}]}(s_1)\mathrm{d}s_1\Big) \\ &\quad\times\Big(\int_0^t \sum_{i_2=0}^{m-1}\Big(\int_{t_{i_2}}^{s_2}\sum_{j_2=1}^{m_\mathrm{w}} \big(\beta_i(a,u)\big)^{k,j_2}\mathrm{d}W_{j_2}(u)\Big)\mathbb{1}_{(t_{i_2},t_{i_2+1}]}(s_2)\mathrm{d}s_2\Big) \\ &= \mathbb{E}\sum_{k=1}^d\Big(\int_0^t\int_0^t \sum_{i_1=0}^{m-1}\sum_{i_2=0}^{m-1}\Big(\int_{t_{i_1}}^{s_1}\sum_{j_1=1}^{m_\mathrm{w}} \big(\beta_{i_1}(a,u)\big)^{k,j_1}\mathrm{d}W_{j_1}(u)\Big) \\ &\quad\times\Big(\int_{t_{i_2}}^{s_2}\sum_{j_2=1}^{m_\mathrm{w}} \big(\beta_{i_2}(a,u)\big)^{k,j_2}\mathrm{d}W_{j_2}(u)\Big)\cdot\mathbb{1}_{(t_{i_1},t_{i_1+1}]\times(t_{i_2},t_{i_2+1}]}(s_1,s_2)\mathrm{d}s_1\mathrm{d}s_2\Big).\end{aligned}$$

The multiplication is non zero only when $i_1 = i_2$, so we have that

$$\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2 = \sum_{k=1}^{d}\bigg(\int_0^t\int_0^t\sum_{i=0}^{m-1}\sum_{j_1,j_2=1}^{m_{\mathrm{w}}}\mathbb{E}\Big(\int_{t_i}^{s_1}\big(\beta_i(a,u)\big)^{k,j_1}\mathrm{d}W_{j_1}(u) \times\int_{t_i}^{s_2}\big(\beta_i(a,u)\big)^{k,j_2}\mathrm{d}W_{j_2}(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]^2}(s_1,s_2)\mathrm{d}s_1\mathrm{d}s_2\bigg).$$

By Theorem A.55 and Lemma A.56

$$\begin{aligned}\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2 &= \sum_{k=1}^{d}\Big(\int_0^t\int_0^t\sum_{i=0}^{m-1}\sum_{j=1}^{m_{\mathrm{w}}}\mathbb{E}\Big(\int_{t_i}^{s_1\wedge s_2}\Big(\big(\beta_i(a,u)\big)^{k,j}\Big)^2\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]^2}(s_1,s_2)\mathrm{d}s_1\mathrm{d}s_2\Big)\\ &= \int_0^t\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\Big(\int_{t_i}^{s_1\wedge s_2}\big\|\beta_i(a,u)\big\|^2\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]^2}(s_1,s_2)\mathrm{d}s_1\mathrm{d}s_2.\end{aligned}$$

Then, by Lemma B.4 and Theorem A.63

$$\begin{aligned}\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2 &\leqslant \int_0^t\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\Big(\int_{t_i}^{s_1\wedge s_2}C\big(1+\|X(u)\|\big)^2\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]^2}(s_1,s_2)\mathrm{d}s_1\mathrm{d}s_2\\ &\leqslant C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \end{aligned}\tag{B.30}$$

Analogously as previous by Theorem A.55, Lemma B.5, Theorem A.63 and assumption (E_{MD}) we have the following estimation

$$\begin{aligned}\mathbb{E}\big\|\tilde{M}^M_{m,3}(t)\big\|^2 &= \int_0^t\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\Big(\int_{t_i}^{s_1\wedge s_2}\big\|\gamma_i(a,u)\big\|^2\lambda(u)\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]^2}(s_1,s_2)\mathrm{d}s_1\mathrm{d}s_2\\ &\leqslant \int_0^t\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\Big(\int_{t_i}^{s_1\wedge s_2}C\big(1+\|X(u-)\|\big)^2\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]^2}(s_1,s_2)\mathrm{d}s_1\mathrm{d}s_2\\ &\leqslant C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \end{aligned}\tag{B.31}$$

By the Hölder inequality Lemma B.5, assumption (E_{MD}) and Theorem A.63 we have that

$$\begin{aligned}
\mathbb{E}\big\|\tilde{M}^M_{m,4}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\Big(\int_{t_i}^s \gamma_i(a,u)\lambda(u)\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\Big\|^2 \\
&\leqslant C\mathbb{E}\int_0^t \sum_{i=0}^{m-1}\Big\|\int_{t_i}^s \gamma_i(a,u)\lambda(u)\mathrm{d}u\Big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s \\
&\leqslant C\int_0^t \sum_{i=0}^{m-1}(s-t_i)\cdot\mathbb{E}\Big(\int_{t_i}^s \big(1+\|X(u-)\|\big)^2\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s \\
&\leqslant C\max_{0\leq i\leq m-1}(t_{i+1}-t_i)^2. \qquad\qquad \text{(B.32)}
\end{aligned}$$

By the (B.29) – (B.32) it follows that

$$\mathbb{E}\big\|\tilde{A}^M_{m,2}(t)\big\|^2 \leqslant C\max_{0\leq i\leq m-1}(t_{i+1}-t_i)^2. \qquad \text{(B.33)}$$

By the Hölder inequality, assumption (B_{MD}), Lemma B.5, and Theorem A.63 we have

$$\begin{aligned}
\mathbb{E}\big\|\tilde{A}^M_{m,3}(t)\big\|^2 &\leqslant CT\sum_{i=0}^{m-1}\mathbb{E}\int_0^t \big\|a(t_i,X(t_i))-a(t_i,\tilde{X}^M_m(t_i))\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s \\
&\leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s. \qquad \text{(B.34)}
\end{aligned}$$

Finally, from (B.28), (B.33), (B.34) we proof that (B.27) holds. That ends the proof of (B.27). ■

Lemma B.9. *Let us assume that the mappings a, b, c and λ satisfy assumptions* (A_{MD}) – (C_{MD}) *and* (E_{MD}). *Let $m \in \mathbb{N}$ and let* (B.1) *be an arbitrary discretization of the interval* $[0,T]$. *Let $B(t)$ and $\tilde{B}^M_m(t)$ be given by* (B.22), (B.25). *For all $t \in [0,T]$ we have the following estimation*

$$\mathbb{E}\big\|B(t)-\tilde{B}^M_m(t)\big\|^2 \leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C\max_{0\leq i\leq m-1}(t_{i+1}-t_i)^2. \qquad \text{(B.35)}$$

Proof. Let $U_i = \big(t_i, \tilde{X}^M_m(t_i)\big)$

$$\mathbb{E}\big\|B(t)-\tilde{B}^M_m(t)\big\|^2 \leqslant 3\Big(\mathbb{E}\big\|\tilde{B}^M_{m,1}(t)\big\|^2+\mathbb{E}\big\|\tilde{B}^M_{m,2}(t)\big\|^2+\mathbb{E}\big\|\tilde{B}^M_{m,3}(t)\big\|^2\Big),$$

where

$$\mathbb{E}\big\|\tilde{B}^M_{m,1}(t)\big\|^2 = \mathbb{E}\Big\|\sum_{j=1}^{m_{\rm w}}\int_0^t\sum_{i=0}^{m-1}\big(b^j(s,X(s))-b^j(t_i,X(s))\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W_j(s)\Big\|^2$$

$$\begin{aligned}\mathbb{E}\big\|\tilde{B}^M_{m,2}(t)\big\|^2 &= \mathbb{E}\Big\|\sum_{j=1}^{m_{\rm w}}\int_0^t\sum_{i=0}^{m-1}\Big(\, b^j(t_i,X(s))-b^j(t_i,X(t_i)),\\ &\quad-\sum_{k=1}^{m_{\rm w}}\int_{t_i}^s L_k b^j(U_i)\mathrm{d}W_k(u)-\int_{t_i}^s L_{-1}b^j(U_i)\mathrm{d}N(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W_j(s)\Big\|^2,\end{aligned}$$

$$\mathbb{E}\big\|\tilde{B}^M_{m,3}(t)\big\|^2 = \mathbb{E}\Big\|\sum_{j=1}^{m_{\rm w}}\int_0^t\sum_{i=0}^{m-1}\big(b^j(t_i,X(t_i))-b^j(U_i)\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W_j(s)\Big\|^2.$$

By the Itô isometry (see Theorem A.55), and assumption ($B_{\rm MD}$)

$$\begin{aligned}\mathbb{E}\big\|\tilde{B}^M_{m,1}(t)\big\|^2 &\leqslant C\sum_{j=1}^{m_{\rm w}}\mathbb{E}\int_0^t\sum_{i=0}^{m-1}\big\|b^j(s,X(s))-b^j(t_i,X(s))\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\mathbb{E}\int_0^t\sum_{i=0}^{m-1}K^2(s-t_i)^2\mathbb{E}(1+\|X(s)\|)^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \qquad \text{(B.36)}\end{aligned}$$

Then, by decomposition (B.15) applied to functions b^j for $j\in\{1,2,\dots,m_{\rm w}\}$ and (A.2) we have the following estimation

$$\mathbb{E}\big\|\tilde{B}^M_{m,2}(t)\big\|^2 \leqslant C\Big(\mathbb{E}\big\|\tilde{M}^M_{m,1}(t)\big\|^2+\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2+\mathbb{E}\big\|\tilde{M}^M_{m,3}(t)\big\|^2+\mathbb{E}\big\|\tilde{M}^M_{m,4}(t)\big\|^2\Big),$$

where

$$\mathbb{E}\big\|\tilde{M}^M_{m,1}(t)\big\|^2=\mathbb{E}\Big\|\sum_{j=1}^{m_{\rm w}}\int_0^t\sum_{i=0}^{m-1}\Big(\int_{t_i}^s\alpha_i(b^j,u)\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W_j(s)\Big\|^2,$$

$$\begin{aligned}\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2&=\mathbb{E}\Big\|\sum_{j=1}^{m_{\rm w}}\int_0^t\sum_{i=0}^{m-1}\Big(\sum_{k=1}^{m_{\rm w}}\int_{t_i}^s\Big(\big(\beta_i(b^j,u)\big)^k-L_kb^j(U_i)\Big)\mathrm{d}W_k(u)\Big)\\ &\qquad\times\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W_j(s)\Big\|^2,\end{aligned}$$

$$\mathbb{E}\big\|\tilde{M}^M_{m,3}(t)\big\|^2=\mathbb{E}\Big\|\sum_{j=1}^{m_{\rm w}}\int_0^t\sum_{i=0}^{m-1}\Big(\int_{t_i}^s\big(\gamma_i(b^j,u)-L_{-1}b^j(U_i)\big)\mathrm{d}\tilde{N}(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W_j(s)\Big\|^2,$$

$$\mathbb{E}\big\|\tilde{M}^M_{m,4}(t)\big\|^2=\mathbb{E}\Big\|\sum_{j=1}^{m_{\rm w}}\int_0^t\sum_{i=0}^{m-1}\Big(\int_{t_i}^s\big(\gamma_i(b^j,u)-L_{-1}b^j(U_i)\big)\lambda(u)\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W_j(s)\Big\|^2.$$

Next, by the Itô isometry, Lemma B.3 and Theorem A.63 we have that

$$\begin{aligned}\mathbb{E}\big\|\bar{M}^M_{m,1}(t)\big\|^2 &\leqslant C\sum_{j=1}^{m_\mathrm{w}}\mathbb{E}\int_0^t\sum_{i=0}^{m-1}\Big\|\int_{t_i}^s\alpha_i(b^j,u)\mathrm{d}u\Big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\int_0^t\sum_{i=0}^{m-1}(s-t_i)\int_{t_i}^s\mathbb{E}\big(1+\|X(s)\|\big)^4\mathrm{d}u\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \qquad\text{(B.37)}\end{aligned}$$

Then, by the Itô isometry, Lemma B.6 and Theorem A.63 it follow that

$$\begin{aligned}\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2 &\leqslant C\sum_{j=1}^{m_\mathrm{w}}\int_0^t\sum_{i=0}^{m-1}\sum_{k=1}^{m_\mathrm{w}}\mathbb{E}\Big\|\int_{t_i}^s\Big(\big(\beta_i(b^j,u)\big)^k-L_kb^j(U_i)\Big)\mathrm{d}W_k(u)\Big\|^2\\ &\qquad\times\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\sum_{j=1}^{m_\mathrm{w}}\int_0^t\sum_{i=0}^{m-1}\sum_{k=1}^{m_\mathrm{w}}\int_{t_i}^s\mathbb{E}\big\|\big(\beta_i(b^j,u)\big)^k-L_kb^j(U_i)\big\|^2\mathrm{d}u\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\sum_{j=1}^{m_\mathrm{w}}\int_0^t\sum_{i=0}^{m-1}\int_{t_i}^s\Big(\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\\ &\qquad+(u-t_i)+(u-t_i)^2\Big)\mathrm{d}u\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\qquad+C_2\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \qquad\text{(B.38)}\end{aligned}$$

By the Itô isometry, Lemma B.7 and fact that λ is continuous (assumption (E_{MD})) we have that

$$\begin{aligned}\mathbb{E}\big\|\tilde{M}^M_{m,3}(t)\big\|^2 &\leqslant C\sum_{j=1}^{m_\mathrm{w}}\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\Big\|\int_{t_i}^s\big(\gamma_i(b^j,u)-L_{-1}b^j(U_i)\big)\mathrm{d}\tilde{N}(u)\Big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\sum_{j=1}^{m_\mathrm{w}}\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\int_{t_i}^s\big\|\gamma_i(b^j,u)-L_{-1}b^j(U_i)\big\|^2\lambda(u)\mathrm{d}u\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\sum_{j=1}^{m_\mathrm{w}}\int_0^t\sum_{i=0}^{m-1}\Big(\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2+(t_{i+1}-t_i)^2\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\ &\leqslant C\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s+C_2\max_{0\le i\le m-1}(t_{i+1}-t_i)^2.\end{aligned}$$

(B.39)

Then, by the Itô isometry, Lemma B.7 by the assumption (E_{MD}) we have the following estimation

$$\begin{aligned}
\mathbb{E}\big\|\tilde{M}^M_{m,4}(t)\big\|^2 &\leqslant C\sum_{j=1}^{m_{\mathrm{w}}}\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\Big\|\int_{t_i}^{s}\big(\gamma_i(b^j,u)-L_{-1}b^j(U_i)\big)\lambda(u)\mathrm{d}u\Big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\
&\leqslant C\sum_{j=1}^{m_{\mathrm{w}}}\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\int_{t_i}^{s}\big\|\gamma_i(b^j,u)-L_{-1}b^j(U_i)\big\|^2\mathrm{d}u\\
&\qquad\qquad\times\int_{t_i}^{s}|\lambda(u)|^2\mathrm{d}u\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\
&\leqslant C\int_0^t\sum_{i=0}^{m-1}\Big(\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2+\ (t_{i+1}-t_i)^2\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\
&\leqslant C\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s+C_2\max_{0\le i\le m-1}(t_{i+1}-t_i)^2.
\end{aligned}\tag{B.40}$$

By the (B.37) – (B.40) we have that

$$\mathbb{E}\big\|\tilde{B}^M_{m,2}(t)\big\|^2 \leqslant C_1\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s+C_2\max_{0\le i\le m-1}(t_{i+1}-t_i)^2.\tag{B.41}$$

Then, by the Itô isometry and assumption ($B1_{\text{MD}}$) it follows that

$$\begin{aligned}
\mathbb{E}\big\|\tilde{B}^M_{m,3}(t)\big\|^2 &\leqslant C\sum_{j=1}^{m_{\mathrm{w}}}\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\big\|b^j(t_i,X(t_i))-b^j(U_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\\
&\leqslant C\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s.
\end{aligned}\tag{B.42}$$

Finally, by the estimations (B.36), (B.41), (B.42) we have that

$$\mathbb{E}\big\|B(t)-\tilde{B}^M_m(t)\big\|^2\leqslant C\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s+C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2.$$

This ends the proof of (B.35). ■

Lemma B.10. *Let us assume that the mappings a, b, c and λ satisfy assumptions* (A_{MD}) – (C_{MD}) *and* (E_{MD}). *Let $m \in \mathbb{N}$ and let* (B.1) *be an arbitrary discretization of the interval* $[0,T]$. *Let $C(t)$ and $\tilde{C}^M_m(t)$ be given by* (B.23), (B.26). *For all $t \in [0,T]$ it follow that*

$$\mathbb{E}\big\|C(t) - \tilde{C}^M_m(t)\big\|^2 \leqslant C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|X(t_i) - \tilde{X}^M_m(t_i)\big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C \max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \tag{B.43}$$

Proof. We start with decomposition

$$\mathbb{E}\big\|C(t) - \tilde{C}^M_m(t)\big\|^2 \leqslant 3\Big(\mathbb{E}\big\|\tilde{C}^M_{m,1}(t)\big\|^2 + \mathbb{E}\big\|\tilde{C}^M_{m,2}(t)\big\|^2 + \mathbb{E}\big\|\tilde{C}^M_{m,3}(t)\big\|^2\Big),$$

where

$$\begin{aligned}
\mathbb{E}\big\|\tilde{C}^M_{m,1}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1} \big(c(s,X(s)) - c(t_i,X(s))\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}N(s)\Big\|^2,\\
\mathbb{E}\big\|\tilde{C}^M_{m,2}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1} \Big(c(t_i,X(s)) - c(t_i,X(t_i))\\
&\quad - \sum_{j=1}^{m_{\text{w}}} \int_{t_i}^s L_j c(U_i)\mathrm{d}W_j(u) - \int_{t_i}^s L_{-1}c(U_i)\mathrm{d}N(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}N(s)\Big\|^2,\\
\mathbb{E}\big\|\tilde{C}^M_{m,3}(t)\big\|^2 &= \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1} \big(c(t_i,X(t_i)) - c(U_i)\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}N(s)\Big\|^2.
\end{aligned}$$

Now, by decomposition (A.2), Itô isometry, and Hölder inequality, Lemma A.63 and assumption $(B2_{\text{MD}})$ for all $t \in [0,T]$ we have the following estimation

$$\begin{aligned}
\mathbb{E}\big\|\tilde{C}^M_{m,1}(t)\big\|^2 &\leqslant C\Bigg(\mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1} \big(c(s,X(s)) - c(t_i,X(s))\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}\tilde{N}(s)\Big\|^2\\
&\quad + \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1} \big(c(s,X(s)) - c(t_i,X(s))\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s\Big\|^2\Bigg)\\
&\leqslant C\Bigg(\int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|c(s,X(s)) - c(t_i,X(s))\big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s\\
&\quad + \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|c(s,X(s)) - c(t_i,X(s))\big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s \cdot \int_0^t |\lambda(s)|^2\mathrm{d}s\Bigg)
\end{aligned}$$

$$\leqslant \; C\sum_{i=0}^{m-1}\int_0^t K^2(s-t_i)^2\cdot\mathbb{E}\big(1+\|X(s)\|\big)^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s$$
$$\leqslant \; C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \tag{B.44}$$

By decomposition (B.15) applied to function c we have that

$$\mathbb{E}\big\|\tilde{C}^M_{m,2}(t)\big\|^2 \;\leqslant\; C\Big(\mathbb{E}\big\|\tilde{M}^M_{m,1}(t)\big\|^2+\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2+\mathbb{E}\big\|\tilde{M}^M_{m,3}(t)\big\|^2\Big),$$

where

$$\mathbb{E}\big\|\tilde{M}^M_{m,1}(t)\big\|^2 \;=\; \mathbb{E}\Big\|\int_0^t\sum_{i=0}^{m-1}(\int_{t_i}^s\alpha_i(c,u)du\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}N(s)\Big\|^2,$$

$$\mathbb{E}\|\tilde{M}^M_{m,2}(t)\|^2 \;=\; \mathbb{E}\Big\|\int_0^t\sum_{i=0}^{m-1}(\sum_{k=1}^{m_{\mathrm{w}}}\int_{t_i}^s\big((\beta_i(c,u))^k-L_jc(U_i)\big)\mathrm{d}W_k(u)\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}N(s)\Big\|^2,$$

$$\mathbb{E}\|\tilde{M}^M_{m,3}(t)\|^2 \;=\; \mathbb{E}\Big\|\int_0^t\sum_{i=0}^{m-1}\Big(\int_{t_i}^s\big(\gamma_i(c,u)-L_{-1}c(U_i)\big)\mathrm{d}N(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}N(s)\Big\|^2.$$

Now, by decomposition of N given by (A.2), Itô isometry, and Hölder inequality, Lemma B.3, Lemma A.63, assumption (E_{MD}), we have that for all $t\in[0,T]$

$$\mathbb{E}\big\|\tilde{M}^M_{m,1}(t)\big\|^2 \;\leqslant\; C\mathbb{E}\Big\|\int_0^t\sum_{i=0}^{m-1}\Big(\int_{t_i}^s\alpha_i(c,u)\mathrm{d}u\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}\tilde{N}(s)\Big\|^2$$
$$+\,C\mathbb{E}\Big\|\int_0^t\sum_{i=0}^{m-1}\Big(\int_{t_i}^s\alpha_i(c,u)du\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s\Big\|^2$$
$$\leqslant\; C\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\Big\|\int_{t_i}^s\alpha_i(c,u)\mathrm{d}u\Big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s$$
$$+\,C\int_0^t\sum_{i=0}^{m-1}\mathbb{E}\Big\|\int_{t_i}^s\alpha_i(c,u)\mathrm{d}u\Big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\cdot\int_0^t|\lambda(s)|^2\mathrm{d}s$$
$$\leqslant\; C\int_0^t\sum_{i=0}^{m-1}(s-t_i)\cdot\int_{t_i}^s\mathbb{E}\big(1+\|X(u)\|\big)^4\mathrm{d}u\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s$$
$$\leqslant\; C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2, \tag{B.45}$$

and

$$
\begin{aligned}
\mathbb{E}\big\|\tilde{M}^M_{m,2}(t)\big\|^2 \leqslant C\mathbb{E}\Big\| \int_0^t \sum_{i=0}^{m-1} \Big(\sum_{k=1}^{m_{\mathrm{w}}} \int_{t_i}^{s} \Big(\big(\beta_i(c,u)\big)^k - L_k c(U_i)\Big) \mathrm{d}W_k(u)\Big) \\
\times \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}\tilde{N}(s)\Big\|^2 \\
+ C\mathbb{E}\Big\| \int_0^t \sum_{i=0}^{m-1} \Big(\sum_{k=1}^{m_{\mathrm{w}}} \int_{t_i}^{s} \Big(\big(\beta_i(c,u)\big)^k - L_k c(U_i)\Big) \mathrm{d}W_k(u)\Big) \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s\Big\|^2 \\
\leqslant C \int_0^t \sum_{i=0}^{m-1} \sum_{k=1}^{m_{\mathrm{w}}} \mathbb{E}\Big\| \int_{t_i}^{s} \Big(\big(\beta_i(c,u)\big)^k - L_k c(U_i)\Big) \mathrm{d}W_k(u)\Big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s \\
+ C \int_0^t \sum_{i=0}^{m-1} \sum_{k=1}^{m_{\mathrm{w}}} \mathbb{E}\Big\| \int_{t_i}^{s} \Big(\big(\beta_i(c,u)\big)^k - L_k c(U_i)\Big) \mathrm{d}W_k(u)\Big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s \\
\leqslant C \int_0^t \sum_{i=0}^{m-1} \sum_{k=1}^{m_{\mathrm{w}}} \int_{t_i}^{s} \mathbb{E}\big\| \big(\beta_i(c,u)\big)^k - L_k c(U_i)\big\|^2 \mathrm{d}u \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s \\
+ C \int_0^t \sum_{i=0}^{m-1} \sum_{k=1}^{m_{\mathrm{w}}} \int_{t_i}^{s} \mathbb{E}\big\| \big(\beta_i(c,u)\big)^k - L_k c(U_i)\big\|^2 \mathrm{d}u \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s \\
\leqslant C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|X(t_i) - \tilde{X}^M_m(t_i)\big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C_2 \max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \quad \text{(B.46)}
\end{aligned}
$$

Now, by decomposition of N given by (A.2), Itô isometry, and Hölder inequality, Lemma B.7, Lemma A.63, assumption (E_{MD}), we have that for all $t \in [0,T]$

$$
\begin{aligned}
\mathbb{E}\big\|\tilde{M}^M_{m,3}(t)\big\|^2 \leqslant\ & C\mathbb{E}\Big\| \int_0^t \sum_{i=0}^{m-1} \Big(\int_{t_i}^{s} \big(\gamma_i(c,u) - L_{-1}c(U_i)\big)\mathrm{d}N(u)\Big) \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}\tilde{N}(s)\Big\|^2 \\
& + C\mathbb{E}\Big\| \int_0^t \sum_{i=0}^{m-1} \Big(\int_{t_i}^{s} \big(\gamma_i(c,u) - L_{-1}c(U_i)\big)\mathrm{d}N(u)\Big) \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s\Big\|^2 \\
\leqslant\ & C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\Big\| \int_{t_i}^{s} \big(\gamma_i(c,u) - L_{-1}c(U_i)\big)\mathrm{d}N(u)\Big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s \\
& + C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\Big\| \int_{t_i}^{s} \big(\gamma_i(c,u) - L_{-1}c(U_i)\big)\mathrm{d}N(u)\Big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s. \quad \text{(B.47)}
\end{aligned}
$$

So analogously like (B.40) we have that

$$\mathbb{E}\big\|\tilde{M}^M_{m,3}(t)\big\|^2 \leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \tag{B.48}$$

Finally, by the (B.45) – (B.48) we have that

$$\mathbb{E}\big\|\tilde{C}^M_{m,2}(t)\big\|^2 \leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2. \tag{B.49}$$

Now, by decomposition of N given by (A.2), Itô isometry, and Hölder inequality, assumption $(B1_{\text{MD}}),(E_{\text{MD}})$ we have that for all $t\in[0,T]$

$$\begin{aligned}
\mathbb{E}\big\|\tilde{C}^M_{m,3}(t)\big\|^2 &\leqslant C\Big(\mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\big(c(t_i,X(t_i))-c(U_i)\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}\tilde{N}(s)\Big\|^2 \\
&\quad + \mathbb{E}\Big\|\int_0^t \sum_{i=0}^{m-1}\big(c(t_i,X(t_i))-c(U_i)\big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s\Big\|^2\Big) \\
&\leqslant C\Big(\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|c(t_i,X(t_i))-c(U_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s \\
&\quad + \int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|c(t_i,X(t_i))-c(U_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s\cdot\int_0^t|\lambda(s)|^2\mathrm{d}s\Big) \\
&\leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s. \qquad \text{(B.50)}
\end{aligned}$$

Finally, by the estimations (B.44), (B.49), (B.50) we have that

$$\mathbb{E}\big\|C(t)-\tilde{C}^M_m(t)\big\|^2 = C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}^M_m(t_i)\big\|^2\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2.$$

This ends the proof of (B.43). ■

B.1.1. Proof of Theorem B.1

Proof. of Theorem B.1 Let $U_i = \big(t_i, \tilde{X}_m^M(t_i)\big)$, we have that $L_{j_1} f(U_i)$ is $\mathfrak{F}_{t_i}$-measurable for $f \in \{b^1, \dots, b^{m_\mathrm{w}}, c\}$, $j_1 \in \{1, 2, \dots, m_\mathrm{w}\} \cup \{-1\}$ because $\tilde{X}_m^M(t_i)$ depends only of evaluation of processes until t_i. Firstly, we show that

$$\sup_{t\in[0,T]} \big\|\tilde{X}_m^M(t)\big\|_{\mathfrak{L}^2(\Omega)} < \infty. \tag{B.51}$$

At the beginning, we show a first step of induction for t_0 ($i = 0$). We have that

$$\big\|\tilde{X}_m^M(t_0)\big\|_{\mathfrak{L}^2(\Omega)} = \mathbb{E}\Big(\big\|\tilde{X}_m^M(t_0)\big\|^2\Big)^{1/2} = \|x_0\| < \infty. \tag{B.52}$$

Now we assume that for $l = 1, 2, \dots, i$, there is that $\big\|\tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)} < \infty$. We show that for $l = 1, 2, \dots, i$ and $t \in [t_l, t_{l+1}]$ we have the following estimation

$$\big\|\tilde{X}_m^M(t) - \tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)} \leqslant C\Big(1 + \big\|\tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)}\Big) \cdot (t - t_l)^{1/2}. \tag{B.53}$$

By Hölder inequality we have that

$$\begin{aligned}
\|\tilde{X}_m^M(t) - \tilde{X}_m^M(t_l)\|^2_{\mathfrak{L}^2(\Omega)} \leqslant C\Big(&\mathbb{E}\big\|a(U_l)\cdot(t-t_l)\big\|^2 + \mathbb{E}\big\|b(U_l)\cdot\big(W(t)-W(t_l)\big)\big\|^2 \\
&+ \mathbb{E}\big\|c(U_l)\cdot\big(N(t)-N(t_l)\big)\big\|^2 + \mathbb{E}\Big\|\sum_{j_1,j_2=1}^{m_\mathrm{w}} L_{j_1} b^{j_2}(U_l)\cdot I_{t_l,t}(W_{j_1}, W_{j_2})\Big\|^2 \\
&+ \mathbb{E}\Big\|\sum_{j_1=1}^{m_\mathrm{w}} L_{j_1} c(U_l)\cdot I_{t_l,t}(W_{j_1}, N)\Big\|^2 + \mathbb{E}\Big\|\sum_{j_1=1}^{m_\mathrm{w}} L_{-1} b^{j_1}(U_l)\cdot I_{t_l,t}(N, W_{j_1})\Big\|^2 \\
&+ \mathbb{E}\big\|L_{-1} c(U_l)\cdot I_{t_i,t}(N, N)\big\|^2\Big).
\end{aligned}$$

By (B.8) we have that

$$\mathbb{E}\big\|a(U_l)\cdot(t-t_l)\big\|^2 \leqslant K^2 T(t-t_l)\cdot\Big(1 + \big\|\tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)}\Big)^2. \tag{B.54}$$

By (B.8), and fact that all $b(U_l)$ and $\big(W(t) - W(t_l)\big)$ are independent, it holds because $b(U_l)$ is $\mathcal{F}_{t_l}$-measurable and $W(t) - W(t_l)$ is independent of $\mathcal{F}_{t_l}$, we have that

$$\begin{aligned}
\mathbb{E}\big\|b(U_l)\cdot\big(W(t)-W(t_l)\big)\big\|^2 &\leqslant \mathbb{E}\big\|b(U_l)\big\|^2\cdot\mathbb{E}\big\|W(t)-W(t_l)\big\|^2 \\
&\leqslant K^2 T(t-t_l)\cdot\Big(1 + \big\|\tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)}\Big)^2.
\end{aligned} \tag{B.55}$$

Analogously by (B.8), by assumptions (E_{MD}) and fact that all $c(U_i)$ and $N(t)-N(t_l)$ are independent, it holds because $c(U_l)$ is $\mathcal{F}_{t_l}$-measurable and $N(t) - N(t_l)$ is independent of $\mathcal{F}_{t_l}$, it follows that

$$\begin{aligned}
\mathbb{E}\big\|c(U_l)\cdot\big(N(t)-N(t_l)\big)\big\|^2 &= \mathbb{E}\big\|c(U_l)\big\|^2\cdot\mathbb{E}|N(t)-N(t_l)|^2 \\
&\leqslant K^2(t-t_l)\cdot\Big(1 + \big\|\tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)}\Big)^2.
\end{aligned} \tag{B.56}$$

For consider $j_1, j_2 \in \{1, \ldots, m_{\mathrm{w}}\}$, by (B.11), and Fact B.28 we have that $L_{j_1} b^{j_2}(U_l)$ and $I_{t_l,t}(W_{j_1}, W_{j_2})$ are independent, then we have that

$$\begin{aligned}
\mathbb{E}\Big\| \sum_{j_1,j_2=1}^{m_{\mathrm{w}}} L_{j_1} b^{j_2}(U_l) \cdot I_{t_l,t}(W_{j_1}, W_{j_2}) \Big\|^2 &\leqslant C \sum_{j_1,j_2=1}^{m_{\mathrm{w}}} \mathbb{E}\|L_{j_1} b^{j_2}(U_l)\|^2 \cdot \mathbb{E}|I_{t_l,t}(W_{j_1}, W_{j_2})|^2 \\
&\leqslant CT(t-t_l)^2 \cdot \Big(1 + \big\|\tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)}\Big)^2. \quad \text{(B.57)}
\end{aligned}$$

Analogously, we have that for $j_1 \in \{1, \ldots, m_{\mathrm{w}}\}$, by (B.11), and Fact B.28 the pairs $L_{j_1} c(U_l)$ and $I_{t_l,t}(W_{j_1}, N)$, $L_{-1} b^{j_1}(U_l)$ and $I_{t_l,t}(N, W_{j_1})$, and $L_{-1} c(U_l)$ and $N(t) - N(t_l)$ are independent. By assumptions (E_{MD}) we have following estimations

$$\begin{aligned}
\mathbb{E}\Big\| \sum_{j_1=1}^{m_{\mathrm{w}}} L_{j_1} c(U_l) \cdot I_{t_l,t}(W_{j_1}, N) \Big\|^2 &\leqslant \sum_{j_1=1}^{m_{\mathrm{w}}} \mathbb{E}\|L_{j_1} c(U_l)\|^2 \cdot \mathbb{E}|I_{t_l,t}(W_{j_1}, N)|^2 \\
&\leqslant C(t-t_l)^2 \cdot \Big(1 + \big\|\tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)}\Big)^2, \quad \text{(B.58)}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\Big\| \sum_{j_1=1}^{m_{\mathrm{w}}} L_{-1} b^{j_1}(U_l) \cdot I_{t_l,t}(N, W_{j_1}) \Big\|^2 &\leqslant \sum_{j_1=1}^{m_{\mathrm{w}}} \mathbb{E}\|L_{-1} b^{j_1}(U_l)\|^2 \cdot \mathbb{E}|I_{t_l,t}(N, W_{j_1})|^2 \\
&\leqslant C(t-t_l)^2 \cdot \Big(1 + \big\|\tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)}\Big)^2, \quad \text{(B.59)}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\big\| L_{-1} c(U_l) I_{t_l,t}(N, N) \big\|^2 &\leqslant \mathbb{E}\|L_{-1} c(U_l)\|^2 \cdot \mathbb{E}|I_{t_l,t}(N, N)|^2 \\
&\leqslant C(t-t_l)^2 \cdot \Big(1 + \big\|\tilde{X}_m^M(t_l)\big\|_{\mathfrak{L}^2(\Omega)}\Big)^2. \quad \text{(B.60)}
\end{aligned}$$

Hence by (B.54) – (B.60) we have that $\sup\limits_{t \in [t_l, t_{l+1}]} \|\tilde{X}_m^M(t)\|_{\mathfrak{L}^2(\Omega)} < +\infty$ and in particular, $\big\|\tilde{X}_m^M(t_{l+1})\big\|_{\mathfrak{L}^2(\Omega)} < +\infty$. Therefore, we get $\sup\limits_{0 \le i \le m} \|\tilde{X}_m^M(t_i)\|_{\mathfrak{L}^2(\Omega)} < +\infty$ and (B.6). This ends the first part of proof of (B.6). □

Now we justify (B.7). Then, by decomposition (B.19) and (B.20) we can write that

$$\begin{aligned}
&\big\|X(t) - \tilde{X}_m^M(t)\big\|_{\mathfrak{L}^2(\Omega)} = \Big(\mathbb{E}\big\|X(t) - \tilde{X}_n^M(t)\big\|^2\Big)^{1/2} \\
&\qquad \leqslant C\Big(\mathbb{E}\big\|A(t) - \tilde{A}_m^M(t)\big\|^2 + \mathbb{E}\big\|B(t) - \tilde{B}_m^M(t)\big\|^2 + \mathbb{E}\big\|C(t) - \tilde{C}_m^M(t)\big\|^2\Big)^{1/2}.
\end{aligned}$$

Then, by Lemma B.8, B.9, B.10 we have that

$$\mathbb{E}\big\|A(t) - \tilde{A}_m^M(t)\big\|^2 \leqslant C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|X(t_i) - \tilde{X}_m^M(t_i)\big\|^2 \mathbb{1}_{(t_i, t_{i+1}]}(s) \mathrm{d}s + C \max_{0 \le i \le m-1} (t_{i+1} - t_i)^2,$$

$$\mathbb{E}\big\|B(t) - \tilde{B}_m^M(t)\big\|^2 \leqslant C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|X(t_i) - \tilde{X}_m^M(t_i)\big\|^2 \mathbb{1}_{(t_i, t_{i+1}]}(s) \mathrm{d}s + C \max_{0 \le i \le m-1} (t_{i+1} - t_i)^2,$$

$$\mathbb{E}\big\|C(t)-\tilde{C}_m^M(t)\big\|^2 \leqslant C\int\limits_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|X(t_i)-\tilde{X}_m^M(t_i)\big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2.$$

Finally, if follows that for all $t\in[0,T]$

$$\begin{aligned}\sup_{0\le s\le t}\mathbb{E}\big\|X(s)-\tilde{X}_m^M(s)\big\|^2 \quad &\leqslant \quad C\int\limits_0^s \sum_{i=0}^{m-1}\sup_{0\le u\le t}\mathbb{E}\big\|X(u)-\tilde{X}_m^M(u)\big\|^2 \mathbb{1}_{(t_i,t_{i+1}]}(u)\mathrm{d}u \\ &\quad + C\max_{0\le i\le m-1}(t_{i+1}-t_i)^2.\end{aligned}$$

By Lemma A.63 and (B.6) mapping

$$[0,T]\ni t\to \sup_{0\le s\le t}\mathbb{E}\big\|X(t)-\tilde{X}_m^M(t)\big\|^2\in\mathbb{R}_+\cup\{0\},$$

is bounded and Borel measurable. Then, by the Theorem A.67 (Grönronwall's inequality) we have (B.7). This ends the proof. ■

B.2. Time-continuous Milstein approximation for system of SDEs under jump commutative condition

In this section we show the definition of time-continuous Milstein approximation under jump commutative condition. We discuss the most important properties about it. Let $U_i = \big(t_i,\tilde{X}_m^M(t_i)\big)$. Under jump commutative condition (D_{MD}) we have that (B.3) takes the following form

$$\begin{aligned}\tilde{X}_m^M(t) = \tilde{X}_m^M(t_i) &+ a(U_i)\cdot(t-t_i) + b(U_i)\cdot\big(W(t)-W(t_i)\big) + c(U_i)\cdot\big(N(t)-N(t_i)\big) \\ &+ \frac{1}{2}\sum_{j_1,j_2=1}^{m_{\text{w}}} L_{j_1}b^{j_2}(U_i)\cdot\Big(I_{t_i,t}(W_{j_1},W_{j_2})+I_{t_i,t}(W_{j_2},W_{j_1})\Big) \\ &+ \sum_{j_1=1}^{m_{\text{w}}} L_{j_1}c(U_i)\cdot\big(I_{t_i,t}(W_{j_1},N)+I_{t_i,t}(N,W_{j_1})\big) + \; L_{-1}c(U_i)\cdot I_{t_i,t}(N,N).\end{aligned}$$

Moreover, for all $i\in\{0,1,\dots,m-1\}$ and $t\in[t_i,t_{i+1}]$, we have the following decomposition

$$\tilde{X}_m^M(t)-\mathbb{E}\big(\tilde{X}_m^M(t)\mid \mathcal{N}_m(N,W)\big) = \tilde{H}_m^M(t)+\tilde{R}_m^M(t), \tag{B.61}$$

where by Fact B.19 we can write that

$$\begin{aligned}
\tilde{H}_m^M(t) &= b(U_i)\cdot\Big(W(t)-\mathbb{E}\big(W(t)\mid \mathcal{N}_m(W)\big)\Big)\\
&\quad + c(U_i)\cdot\Big(N(t)-\mathbb{E}\big(N(t)\mid \mathcal{N}_m(N)\big)\Big), \qquad (B.62)\\
\tilde{R}_m^M(t) &= \frac{1}{2}\sum_{j_1,j_2=1}^{m_{\mathrm{w}}} L_{j_1}b^{j_2}(U_i)\cdot\Big(I_{t_i,t}(W_{j_1},W_{j_2})+I_{t_i,t}(W_{j_2},W_{j_1})\\
&\qquad -\mathbb{E}\big(I_{t_i,t}(W_{j_1},W_{j_2})+I_{t_i,t}(W_{j_2},W_{j_1})\mid \mathcal{N}_m(W_{j_1},W_{j_2})\big)\Big)\\
&\quad + \sum_{j=1}^{m_{\mathrm{w}}} L_{j_1}c(U_i)\cdot\Big(I_{t_i,t}(N,W_j)+I_{t_i,t}(W_j,N)\\
&\qquad -\mathbb{E}\big(I_{t_i,t}(N,W_j)+I_{t_i,t}(W_j,N)\mid \mathcal{N}_m(W_j,N)\big)\Big)\\
&\quad + L_{-1}c(U_i)\cdot\Big(I_{t_i,t}(N,N)-\mathbb{E}\big(I_{t_i,t}(N,N)\mid \mathcal{N}_m(N)\big)\Big). \qquad (B.63)
\end{aligned}$$

Lemma B.11. *Let us assume that the mappings a, b, c and λ satisfy the assumptions $(A_{\mathrm{MD}})-(E_{\mathrm{MD}})$. For all $t\in[t_i,t_{i+1}], i=0,1,\ldots,m-1$, it follow that*

$$\mathbb{E}\big\|\tilde{R}_m^M(t)\big\|^2 \leqslant C(t_{i+1}-t_i)^2, \qquad (B.64)$$

where $C>0$ does not depend on m nor i.

Proof. From (B.11) and Theorem B.1 we have that for $f\in\{b^1,\ldots,b^{m_{\mathrm{w}}},c\}$ and $j\in\{-1,1,\ldots,m_{\mathrm{w}}\}$ we have the following estimation

$$\mathbb{E}\big\|L_j f(U_i)\big\|^2 \leqslant C, \qquad (B.65)$$

where $C>0$ does not depend on m nor i. Moreover, for $f\in\{b^1,\ldots,b^{m_{\mathrm{w}}},c\}$ and $j\in\{-1,1,\ldots,m_{\mathrm{w}}\}$ the random variable $L_jf(U_i)$ is $\mathcal{F}_{t_i}$-measurable. From Fact B.28 (ii) and by Lemma B.24 – B.27 we have that for $j_1,j_2\in\{1,\ldots,m_{\mathrm{w}}\}$ the random variables

$$\begin{aligned}
&I_{t_i,t}(N,N)-\mathbb{E}\big(I_{t_i,t}(N,N)\mid \mathcal{N}_m(N)\big),\\
&I_{t_i,t}(W_{j_1},W_{j_2})+I_{t_i,t}(W_{j_2},W_{j_1})-\mathbb{E}\big(I_{t_i,t}(W_{j_1},W_{j_2})+I_{t_i,t}(W_{j_2},W_{j_1})\mid \mathcal{N}_m(W_{j_1},W_{j_2})\big),\\
&I_{t_i,t}(N,W_{j_1})+I_{t_i,t}(W_{j_1},N)-\mathbb{E}\big(I_{t_i,t}(N,W_{j_1})+I_{t_i,t}(W_{j_1},N)\mid \mathcal{N}_m(N,W_{j_1})\big),
\end{aligned}$$

are independent of $\mathcal{F}_{t_i}$. Hence, by (B.63), Fact B.28 (i) and (B.65) we have

$$\begin{aligned}\big\|\tilde{R}^M_m(t)\big\|_{\mathfrak{L}^2(\Omega)} \leqslant\; & C_1 \sum_{j_1,j_2=1}^{m_{\mathrm{w}}} \big\|L_{j_1} b^{j_2}(U_i)\big\|_{\mathfrak{L}^2(\Omega)} \cdot \big\|I_{t_i,t}(W_{j_1},W_{j_2}) + I_{t_i,t}(W_{j_2},W_{j_1})\big\|_{\mathfrak{L}^2(\Omega)} \\ & + C_1 \sum_{j_1=1}^{m_{\mathrm{w}}} \big\|L_{j_1} c(U_i)\big\|_{\mathfrak{L}^2(\Omega)} \cdot \big\|I_{t_i,t}(N,W_{j_1}) + I_{t_i,t}(W_{j_1},N)\big\|_{\mathfrak{L}^2(\Omega)} \\ & + C_1 \big\|L_{-1} c(U_i)\big\|_{\mathfrak{L}^2(\Omega)} \cdot \big\|I_{t_i,t}(N,N)\big\|_{\mathfrak{L}^2\Omega)} \\ \leqslant\; & C(t-t_i), \qquad \text{(B.66)}\end{aligned}$$

for $t \in [t_i, t_{i+1}]$, so that ends the proof of (B.64). ■

From Lemma B.11 there exists a constant $C > 0$ such that for all $m \in \mathbb{N}$ and arbitrary discretization (B.1)

$$\sup_{t\in[0,T]} \mathbb{E}\big\|\tilde{R}^M_m(t)\big\|^2 \leqslant C \max_{0\leq i\leq m-1} (t_{i+1}-t_i)^2. \qquad \text{(B.67)}$$

B.3. Derivative free time-continuous Milstein approximation for system of SDEs under jump commutative conditions

In this section we discuss basic properties of the derivative free time-continuous Milstein approximation under jump commutative conditions. Note that the derivative-free version of the Milstein scheme has to be defined in a suitable way, since the operator $\mathcal{L}_{j,h} f$ approximates $L_j f$ but does not commute.

Let $m \in \mathbb{N}$ and

$$0 = t_0 < t_1 < \ldots < t_m = T, \qquad \text{(B.68)}$$

be an arbitrary discretization of interval $[0,T]$. By

$$\Delta Z_i = Z(t_{i+1}) - Z(t_i),$$

we denote the increment of stochastic processes $Z \in \{N, W, W_1, \ldots, W_{m_{\mathrm{w}}}\}$ where for $i \in \{0,1,\ldots,m-1\}$, ΔZ_i is either a vector or a scalar depending on process structure. For $h > 0$ and $f \in \{b^1, \ldots, b^{m_{\mathrm{w}}}, c\}$ we denote by

$$\mathcal{L}_{k,h} f(t,y) = \widetilde{\nabla}_{x,h} f(t,y) \cdot b^k(t,y), \quad (t,y) \in [0,T] \times \mathbb{R}^d.$$

Defined in that way operator approximate operator $L_k f$. Let us define another operator

$$\widetilde{\mathcal{L}}_{j_1,h} b^{j_2} := \begin{cases} \mathcal{L}_{j_1,h} b^{j_2}, & j_1 \leqslant j_2, \\ \mathcal{L}_{j_2,h} b^{j_1}, & j_1 > j_2. \end{cases} \qquad \text{(B.69)}$$

By the Lemma B.12 we have that operator $\widetilde{\mathcal{L}}_{j_1,h}$ commute.

Lemma B.12. *We have that for* $j_1, j_2 \in \{1, \ldots, m_{\mathrm{w}}\}$

$$\widetilde{\mathcal{L}}_{j_1,h} b^{j_2} = \widetilde{\mathcal{L}}_{j_2,h} b^{j_1}.$$

Proof. Without loss of generality $j_1 < j_2$ (when $j_1 = j_2$ it is trivial). By the definition of operator $\widetilde{\mathcal{L}}_{j_1,h}$ given by (B.69) we have that

$$\widetilde{\mathcal{L}}_{j_1,h} b^{j_2} = \mathcal{L}_{j_1,h} b^{j_2}, \quad \widetilde{\mathcal{L}}_{j_2,h} b^{j_1} = \mathcal{L}_{j_1,h} b^{j_2},$$

and that ends the proof. ∎

By the Lemma B.12, under the jump-commutativity condition (D_{MD}) the time-continuous derivative-free Milstein approximation $\tilde{X}_m^{df-M} = \{\tilde{X}_m^{df-M}(t)\}_{t\in[0,T]}$ based on the mesh (B.68) is defined as follows. We set

$$\tilde{X}_m^{df-M}(0) = x_0, \tag{B.70}$$

and

$$\begin{aligned}
\tilde{X}_m^{df-M}(t) \quad = \quad & \tilde{X}_m^{df-M}(t_i) + a(U_i^{df}) \cdot (t - t_i) \\
& + b(U_i^{df}) \cdot \big(W(t) - W(t_i)\big) + c(U_i^{df}) \cdot \big(N(t) - N(t_i)\big) \\
& + \frac{1}{2} \sum_{j_1,j_2=1}^{m_{\mathrm{w}}} \widetilde{\mathcal{L}}_{j_1,h_i} b^{j_2}(U_i^{df}) \cdot \big(I_{t_i,t}(W_{j_1}, W_{j_2}) + I_{t_i,t}(W_{j_2}, W_{j_1})\big) \\
& + \sum_{j_1=1}^{m_{\mathrm{w}}} L_{-1} b^{j_1}(U_i^{df}) \cdot \big(I_{t_i,t}(N, W_{j_1}) + I_{t_i,t}(W_{j_1}, N)\big) \\
& + L_{-1} c(U_i^{df}) \cdot I_{t_i,t}(N, N),
\end{aligned} \tag{B.71}$$

for $t \in [t_i, t_{i+1}]$, $i = 0, 1, \ldots, m-1$, where $U_i^{df} = (t_i, \tilde{X}_m^{df-M}(t_i))$ and $h_i = t_{i+1} - t_i$.

For every $m \in \mathbb{N}$ the process $\{\tilde{X}_m^{df-M}(t)\}_{t\in[0,T]}$ is adapted to $\{\mathcal{F}_t\}_{t\in[0,T]}$ and has càdlàg paths. Moreover, the random variables $\{\tilde{X}_m^{df-M}(t_i)\}_{i=0}^m$ are measurable with respect to the σ-algebra generated by vector of information $\mathcal{N}_m(N, W)$, it is that

$$\sigma\big(\mathcal{N}_m(N, W)\big) = \sigma\big(N(t_1), N(t_2), \ldots, N(t_m), W(t_1), W(t_2), \ldots, W(t_m)\big). \tag{B.72}$$

In [61] the authors proposed a derivative-free version of the Milstein scheme. However, the error was investigated under stronger assumptions than imposed in Theorem B.13.

Theorem B.13. *Let us assume that the mappings a, b, c and λ satisfy assumptions (A_{MD}) – (C_{MD}) and (E_{MD}). Let $m \in \mathbb{N}$ and let* (B.68) *be an arbitrary discretization of the interval $[0, T]$. Then for continuous Milstein approximation $\tilde{X}_m^{df-M}$, based on the mesh* (B.68) *we have that*

$$\sup_{t\in[0,T]} \left\|\tilde{X}_m^{df-M}(t)\right\|_{\mathfrak{L}^2(\Omega)} \leqslant C_1, \tag{B.73}$$

and

$$\sup_{t\in[0,T]} \left\|X(t) - \tilde{X}_m^{df-M}(t)\right\|_{\mathfrak{L}^2(\Omega)} \leqslant C_2 \max_{0\leq i\leq m-1}(t_{i+1} - t_i), \tag{B.74}$$

where $C_1, C_2 > 0$ do not depend on m.

Before we show proof of Theorem B.13, we will focus on important auxiliary lemmas, which help to prove the theorem. We start with additional results following from the given assumptions (A_{MD}) – (C_{MD}).

Lemma B.14. *Let $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfy (A_{MD}) – (C_{MD}) then for all $(t, y), (t, z) \in [0, T] \times \mathbb{R}^d$, exists $K_1 > 0$ depends only on $\|f(0, 0)\|$, K and T such that*

$$\left\|\widetilde{\nabla}_{x,h} f(t, x)\right\| \leqslant Kd, \tag{B.75}$$

$$\left\|\nabla_x f(t, x) - \widetilde{\nabla}_{x,h} f(t, x)\right\| \leqslant Kdh, \tag{B.76}$$

$$\left\|\mathcal{L}_{j,h} f(t, x)\right\| \leqslant KK_1 d\big(1 + \|x\|\big), \tag{B.77}$$

and

$$\left\|L_j f(t, x) - \mathcal{L}_{j,h} f(t, z)\right\| \leqslant K\|x - z\| + KK_1 d\big(1 + \|z\|\big)h, \tag{B.78}$$

$$\left\|L_{-1} f(t, x) - L_{-1} f(t, z)\right\| \leqslant 3K\|x - z\|. \tag{B.79}$$

Proof. By Lemma B.2 we have that

$$\left\|\widetilde{\nabla}_{x,h} f(t, x)\right\| = \left(\sum_{i_1,i_2=1}^{d} \left|\frac{f_{i_1}(t, x + h\cdot e_{i_2}) - f_{i_1}(t, x)}{h}\right|^2\right)^{1/2} \leqslant Kd,$$

and that end the proof of (B.75). □

We have that $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $f = (f_1, \dots, f_d)$. For $t \in [0, T], x \in \mathbb{R}^d$ we have that

$$f_i(t, x + h \cdot e_j) - f_i(t, x) = f_i(t, x_1, \dots, x_j + h, \dots, x_d) - f_i(t, x_1, \dots, x_j, \dots, x_d).$$

We can use a standard mean value theorem

$$
\begin{aligned}
f_i(t,x_1,\dots,x_j+h,\dots,x_d) &- f_i(t,x_1,\dots,x_j,\dots,x_d) \\
&= h\cdot\frac{\partial f_i}{\partial x_j}(t,x_1,\dots,x_{j-1},\xi_{i,j},x_{j+1},\dots,x_d)
\end{aligned}
$$

for some $\xi_{ij}\in[x_j,x_j+h]$. Let us define $\tilde{\xi}_{i,j}=(x_1,\dots,x_{j-1},\xi_{i,j},x_{j+1},\dots x_d)$. It is easy to see that $\|x-\tilde{\xi}_{i,j}\|\leqslant h$. So we have that

$$
\begin{aligned}
\big\|\nabla_x f(t,x)-\widetilde{\nabla}_{x,h}f(t,x)\big\|^2 &= \sum_{i,j=1}^{d}\Big|\frac{\partial f_i}{\partial x_j}(t,x)-\frac{f_i(t,x+h\cdot e_j)-f_i(t,x)}{h}\Big|^2 \\
&= \sum_{i,j=1}^{d}\Big|\frac{\partial f_i}{\partial x_j}(t,x)-\frac{\partial f_i}{\partial x_j}(t,\tilde{\xi}_{i,j})\Big|^2.
\end{aligned}
$$

By assumption (B_{MD}) we have that

$$
\big\|\nabla_x f(t,x)-\widetilde{\nabla}_{x,h}f(t,x)\big\| \leqslant \sum_{i,j=1}^{d}K\|x-\tilde{\xi}_{i,j}\|\leqslant Kdh.
$$

That end the proof of (B.76). □

Now we go to prove (B.77). By Lemma B.2 and assumption (B.75) we have that

$$
\begin{aligned}
\big\|\mathcal{L}_{j,h}f(t,x)\big\| &\leqslant \big\|\widetilde{\nabla}_{x,h}f(t,y)\big\|\cdot\|b^j(t,y)\| \\
&\leqslant KK_1\big(1+\|x\|\big).
\end{aligned}
$$

That ends the proof of (B.77). □

We go to prove (B.78). By Lemma B.2, (D_{MD}) and (B.76) we have that

$$
\begin{aligned}
\big\|L_jf(t,x)-\mathcal{L}_{j,h}f(t,z)\big\| &\leqslant \big\|L_jf(t,x)-L_jf(t,z)\big\|+\big\|L_jf(t,z)-\mathcal{L}_{j,h}f(t,z)\| \\
&\leqslant K\|x-z\|+\big\|\nabla_x f(t,z)-\widetilde{\nabla}_{x,h}f(t,z)\big\|\cdot\|b^k(t,z)\| \\
&\leqslant K\|x-z\|+KdKK_1\big(1+\|z\|\big)h.
\end{aligned}
$$

That ends the proof of (B.78). □

Finally, we go to prove (B.79). By Lemma B.2, assumption (D_{MD}) and (B.76) we have that

$$
\begin{aligned}
\|L_{-1}f(t,x)-L_{-1}f(t,z)\| &\leqslant \big\|f(t,x+c(t,x))-f(t,z+c(t,z))\big\|+\|f(t,z)-f(t,x)\| \\
&\leqslant K\|x-z\|+\|c(t,z)-c(t,x)\| \\
&\leqslant 3K\|x-z\|.
\end{aligned}
$$

That ends the proof. ■

B.3.1. Proof of Theorem B.13

***Proof.* of Theorem B.13** Since the functions a, b, c, $\widetilde{\mathfrak{L}}_{j_1,h_i}b^{j_2}$, $L_{-1}b^{j_2}$ and $L_{-1}c$ for $j_1, j_2 \in \{1, \ldots, m_{\mathrm{w}}\}$ satisfy linear growth condition, the estimate (B.73) follows from standard arguments, as in proof of Theorem B.1, so we skip it.

To proof (B.74), let us define $U_i^{df} = \big(t_i, \tilde{X}_m^{M-df}(t_i)\big)$, we have the following decomposition $\tilde{X}_m^{M-df}$ for all $t \in [0,T]$

$$\tilde{X}_m^{df-M}(t) = x_0 + \tilde{A}_m^{df-M}(t) + \tilde{B}_m^{df-M}(t) + \tilde{C}_m^{df-M}(t),$$

where

$$\begin{aligned}
\tilde{A}_m^{df-M}(t) &= \int_0^t \sum_{i=0}^{m-1} a(U_i^{df}) \cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s, \\
\tilde{B}_m^{df-M}(t) &= \sum_{j=1}^{m_{\mathrm{w}}} \Big(\int_0^t \sum_{i=0}^{m-1} \Big(b^j(U_i^{df}) + \sum_{k=1}^{m_{\mathrm{w}}} \int_{t_i}^s \widetilde{\mathfrak{L}}_{k,h_i} b^j(U_i^{df}) \mathrm{d}W_k(u) \\
&\qquad + \int_{t_i}^s L_{-1} b^j(U_i^{df}) \mathrm{d}N(u) \Big) \mathbb{1}_{(t_i,t_{i+1}]}(s) \mathrm{d}W_j(s) \Big), \\
\tilde{C}_m^{df-M}(t) &= \int_0^t \sum_{i=0}^{m-1} \Big(c(U_i^{df}) + \sum_{j=1}^{m_{\mathrm{w}}} \int_{t_i}^s L_{-1} b^j(U_i^{df}) \mathrm{d}W_j(u) \\
&\qquad + \int_{t_i}^s L_{-1} c(U_i^{df}) \mathrm{d}N(u) \Big) \mathbb{1}_{(t_i,t_{i+1}]}(s) \mathrm{d}N(s).
\end{aligned}$$

Let $U_i = (t_i, \tilde{X}_m^M(t_i))$. Moreover, from (B.20) we have for all $t \in [0,T]$

$$\tilde{X}_m^M(t) = x_0 + \tilde{A}_m^M(t) + \tilde{B}_m^M(t) + \tilde{C}_m^M(t), \tag{B.80}$$

where

$$\begin{aligned}
\tilde{A}_m^M(t) &= \int_0^t \sum_{i=0}^{m-1} a(U_i)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s,\\
\tilde{B}_m^M(t) &= \sum_{j=1}^{m_{\mathrm{w}}}\Big(\int_0^t \sum_{i=0}^{m-1}\Big(b^j(U_i) + \sum_{k=1}^{m_{\mathrm{w}}}\int_{t_i}^{s} L_k b^j(U_i)\mathrm{d}W_k(u)\\
&\qquad + \int_{t_i}^{s} L_{-1}b^j(U_i)\mathrm{d}N(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}W_j(s)\Big),\\
\tilde{C}_m^M(t) &= \int_0^t \sum_{i=0}^{m-1}\Big(c(U_i) + \sum_{j=1}^{m_{\mathrm{w}}}\int_{t_i}^{s} L_{-1}b^j(U_i)\mathrm{d}W_j(u)\\
&\qquad + \int_{t_i}^{s} L_{-1}c(U_i)\mathrm{d}N(u)\Big)\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}N(s).
\end{aligned}$$

From Hölder inequality and assumption ($B1_{\mathrm{MD}}$) we have for all $t \in [0,T]$ the following estimation

$$\mathbb{E}\big\|\tilde{A}_m^M(t) - \tilde{A}_m^{df-M}(t)\big\|^2 \leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|\tilde{X}_m^M(t_i) - \tilde{X}_m^{df-M}(t_i)\big\|^2\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s. \quad \text{(B.81)}$$

Now, by Itô isometry and Hölder inequality we have that

$$\mathbb{E}\|\tilde{B}_m^M(t) - \tilde{B}_m^{df-M}(t)\|^2 \leqslant 3\Big(B_{1,m}(t) + B_{2,m}(t) + B_{3,m}(t)\Big),$$

where

$$\begin{aligned}
B_{1,m}(t) &= \sum_{j=1}^{m_{\mathrm{w}}}\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|b^j(U_i) - b^j(U_i^{df})\big\|^2\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s,\\
B_{2,m}(t) &= \sum_{j=1}^{m_{\mathrm{w}}}\int_0^t \sum_{i=0}^{m-1}\sum_{k=1}^{m_{\mathrm{w}}}\mathbb{E}\Big\|\int_{t_i}^{s}\big(L_k b^j(U_i) - \widetilde{\mathfrak{L}}_{k,h_i}b^j(U_i^{df})\big)\mathrm{d}W_k(u)\Big\|^2\cdot\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s,\\
B_{3,m}(t) &= \sum_{j=1}^{m_{\mathrm{w}}}\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\Big\|\int_{t_i}^{s}\big(L_{-1}b^j(U_i) - L_{-1}b^j(U_i^{df})\big)\mathrm{d}N(u)\Big\|^2\cdot\mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s.
\end{aligned}$$

From assumption ($B1_{\mathrm{MD}}$) we have that

$$B_{1,n}(t) \leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|\tilde{X}_m^M(t_i) - \tilde{X}_m^{df-M}(t_i)\big\|^2\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s, \quad \text{(B.82)}$$

and by Itô isometry, (B.78) and (B.73),

$$\begin{aligned} B_{2,n}(t) &\leqslant C\int_0^t \sum_{i=0}^{m-1}\Big(\mathbb{E}\big\|\tilde{X}_m^M(t_i)-\tilde{X}_m^{df-M}(t_i)\big\|^2 \\ &\qquad\qquad + h_i^3\Big(1+\mathbb{E}\big\|\tilde{X}_m^{df-M}(t_i)\big\|^2\Big)\Big)\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s \\ &\leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|\tilde{X}_m^M(t_i)-\tilde{X}_m^{df-M}(t_i)\big\|^2\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C\max_{0\le i\le m-1} h_i^3. \end{aligned} \tag{B.83}$$

Then, by using the decomposition $N(t) = \tilde{N}(t) + m(t)$ together with martingale isometry for the compensated Poisson process and Lemma B.79 we have that

$$\begin{aligned} B_{3,n}(t) &\leqslant \sum_{j=1}^{m_{\mathrm{w}}}\int_0^t \sum_{i=0}^{m-1}\int_{t_i}^s \mathbb{E}\big\|L_{-1}b^j(U_i)-L_{-1}b^j(U_i^{df})\big\|^2\lambda(u)\mathrm{d}u\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s \\ &\quad + \int_0^t \sum_{i=0}^{m-1}\int_{t_i}^s \mathbb{E}\big\|L_{-1}b^j(U_i)-L_{-1}b^j(U_i^{df})\big\|^2\lambda^2(u)\mathrm{d}u\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s \\ &\leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|\tilde{X}_m^M(t_i)-\tilde{X}_m^{df-M}(t_i)\big\|^2\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s. \end{aligned} \tag{B.84}$$

Combine together (B.82) – (B.84)

$$\mathbb{E}\big\|\tilde{B}_m^M(t)-\tilde{B}_m^{df-M}(t)\big\|^2 \leqslant C\int_0^t \sum_{i=0}^{m-1}\mathbb{E}\big\|\tilde{X}_m^M(t_i)-\tilde{X}_m^{df-M}(t_i)\big\|^2\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s + C_9\max_{0\le i\le m-1} h_i^3. \tag{B.85}$$

To show estimation for part $\mathbb{E}\big\|\tilde{C}_m^M(t)-\tilde{C}_m^{df-M}(t)\big\|^2$ we again use the decomposition $N(t) = \tilde{N}(t) + m(t)$ together with the martingale isometry and assumption (D_{MD}). Then, it follows that

$$\mathbb{E}\big\|\tilde{C}_m^M(t)-\tilde{C}_m^{df-M}(t)\big\|^2 \leqslant 3\Big(C_{1,m}(t)+C_{2,m}(t)+C_{3,m}(t)\Big),$$

where

$$\begin{aligned} C_{1,m}(t) &= \int_0^t \sum_{i=0}^{m-1}\mathbb{E}\|c(U_i)-c(U_i^{df})\|^2\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s \\ &\quad + \int_0^t \sum_{i=0}^{m-1}\mathbb{E}\|c(U_i)-c(U_i^{df})\|^2\cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda^2(s)\mathrm{d}s, \end{aligned}$$

$$C_{2,m}(t) = \int_0^t \sum_{i=0}^{m-1} \sum_{k=1}^{m_{\mathrm{w}}} \mathbb{E}\Big\| \int_{t_i}^{s} \big(L_{-1}b^k(U_i) - L_{-1}b^k(U_i^{df})\big)\mathrm{d}W_k(u)\Big\|^2 \cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s,$$

$$+ \int_0^t \sum_{i=0}^{m-1} \sum_{k=1}^{m_{\mathrm{w}}} \mathbb{E}\Big\| \int_{t_i}^{s} \big(L_{-1}b^k(U_i) - L_{-1}b^k(U_i^{df})\big)\mathrm{d}W_k(u)\Big\|^2 \cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda^2(s)\mathrm{d}s,$$

$$C_{3,m}(t) = \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\Big\| \int_{t_i}^{s} \big(L_{-1}c(U_i) - L_{-1}c(U_i^{df})\big)\mathrm{d}N(u)\Big\|^2 \cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda(s)\mathrm{d}s$$

$$+ \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\Big\| \int_{t_i}^{s} \big(L_{-1}c(U_i) - L_{-1}c(U_i^{df})\big)\mathrm{d}N(u)\Big\|^2 \cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\lambda^2(s)\mathrm{d}s.$$

Proceeding analogously as for the term $\mathbb{E}\|\tilde{B}_m^M(t) - \tilde{B}_m^{df-M}(t)\|^2$ we arrive that

$$C_{1,m}(t) \leqslant C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|\tilde{X}_m^M(t_i) - \tilde{X}_m^{df-M}(t_i)\big\|^2 \cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s, \tag{B.86}$$

$$C_{2,m}(t) \leqslant C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|\tilde{X}_m^M(t_i) - \tilde{X}_m^{df-M}(t_i)\big\|^2 \cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s, \tag{B.87}$$

$$C_{3,m}(t) \leqslant C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|\tilde{X}_m^M(t_i) - \tilde{X}_m^{df-M}(t_i)\big\|^2 \cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s. \tag{B.88}$$

Finally by (B.86) – (B.88), we obtain that

$$\mathbb{E}\big\|\tilde{C}_m^M(t) - \tilde{C}_m^{df-M}(t)\big\|^2 \leqslant C \int_0^t \sum_{i=0}^{m-1} \mathbb{E}\big\|\tilde{X}_m^M(t_i) - \tilde{X}_m^{df-M}(t_i)\big\|^2 \cdot \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathrm{d}s. \tag{B.89}$$

Hence, by (B.81), (B.85) and (B.89) we have that for all $t \in [0,T]$

$$\mathbb{E}\big\|\tilde{X}_m^M(t) - \tilde{X}_m^{df-M}(t)\big\|^2 \leqslant C \int_0^t \sup_{0 \le u \le s} \mathbb{E}\big\|\tilde{X}_m^M(u) - X_m^{df-M}(u)\big\|^2 \mathrm{d}s + C \max_{0 \le i \le m-1} h_i^3,$$

and by the Grönwall's inequality (Lemma A.67) we get for all $t \in [0,T]$ that

$$\mathbb{E}\big\|\tilde{X}_m^M(t) - \tilde{X}_m^{df-M}(t)\big\|^2 \leqslant C \max_{0 \le i \le m-1} h_i^3.$$

This implies (B.74) and that ends the proof. ■

For all $t \in [t_i, t_{i+1}]$, $i = 0, 1, \ldots, m-1$ we have the following decomposition (analogous as for $\tilde{X}_m^M$),

$$\tilde{X}_m^{df-M}(t) - \mathbb{E}\big(\tilde{X}_m^{df-M}(t) \mid \mathcal{N}_m(N,W)\big) = \tilde{H}_m^{df-M}(t) + \tilde{R}_m^{df-M}(t), \tag{B.90}$$

where by Fact B.19

$$
\begin{aligned}
\tilde{H}_m^{df-M}(t) &= b(U_i^{df}) \cdot \Big(W(t) - \mathbb{E}\big(W(t) \mid \mathcal{N}_m(W)\big)\Big) \\
&\quad + c(U_i^{df}) \cdot \Big(N(t) - \mathbb{E}\big(N(t) \mid \mathcal{N}_m(N)\big)\Big), \qquad \text{(B.91)} \\
\tilde{R}_m^{df-M}(t) &= \frac{1}{2} \sum_{j_1,j_2=1}^{m_{\mathrm{w}}} \widetilde{\mathcal{L}}_{j_1,h_i} b^{j_2}(U_i^{df}) \cdot \Big(I_{t_i,t}(W_{j_1}, W_{j_2}) + I_{t_i,t}(W_{j_2}, W_{j_1}) \\
&\qquad - \mathbb{E}\big(I_{t_i,t}(W_{j_1}, W_{j_2}) + I_{t_i,t}(W_{j_2}, W_{j_1}) \mid \mathcal{N}_m(W_{j_1}, W_{j_2})\big)\Big) \\
&\quad + \sum_{j=1}^{m_{\mathrm{w}}} L_{-1} b^j(U_i^{df}) \cdot \Big(I_{t_i,t}(N, W_j) + I_{t_i,t}(W_j, N) \\
&\qquad - \mathbb{E}(I_{t_i,t}(N, W_j) + I_{t_i,t}(W_j, N) \mid \mathcal{N}_m(N, W))\Big) \\
&\quad + L_{-1} c(U_i^{df}) \cdot \Big(I_{t_i,t}(N, N) - \mathbb{E}\big(I_{t_i,t}(N, N) \mid \mathcal{N}_m(N)\big)\Big). \qquad \text{(B.92)}
\end{aligned}
$$

Since (B.77) holds, we can show the following estimation for $\tilde{R}_m^{df-M}(t)$.

Lemma B.15. *Let us assume that the mappings a, b, c and λ satisfy the assumptions (A_{MD}) – (E_{MD}). Let* (B.68) *be discretization of interval* $[0, T]$ *For all* $t \in [t_i, t_{i+1}]$, $i = 0, 1, \ldots, m-1$ *we have that*

$$
\mathbb{E}\|\tilde{R}_m^{df-M}(t)\|^2 \leqslant C(t_{i+1} - t_i)^2,
$$

where $C > 0$ does not depend on m nor i.

Proof. From (B.11), (B.77) and Theorem B.13 we have that for $f \in \{b^1, \ldots, b^{m_{\mathrm{w}}}, c\}$ and $j \in \{-1, 1, \ldots, m_{\mathrm{w}}\}$ the following estimations holds

$$
\mathbb{E}\|L_{-1} f(U_i^{df})\|^2 \leqslant C, \qquad \text{(B.93)}
$$

$$
\mathbb{E}\|\widetilde{\mathcal{L}}_{j_1,h_i} b^{j_2}(U_i^{df})\|^2 \leqslant C, \qquad \text{(B.94)}
$$

where $C > 0$ does not depend on n nor i. Moreover, for $f \in \{b^1, \ldots, b^{m_{\mathrm{w}}}, c\}$ and $j \in \{-1, 1, \ldots, m_{\mathrm{w}}\}$ the random variables $L_{-1} f(U_i)$ and $\widetilde{\mathcal{L}}_{j_1,h_i} b^{j_2}$ for $j_1, j_2 \in \{1, \ldots, m_{\mathrm{w}}\}$ are $\mathcal{F}_{t_i}$-measurable. Then from Fact B.28 (ii) and by Lemma B.24 - B.27 we have that for $j_1, j_2 \in \{1, \ldots, m_{\mathrm{w}}\}$, the following random variables

$$
\begin{aligned}
& I_{t_i,t}(N, N) - \mathbb{E}\big(I_{t_i,t}(N, N) \mid \mathcal{N}_m(N)\big), \\
& I_{t_i,t}(W_{j_1}, W_{j_2}) + I_{t_i,t}(W_{j_2}, W_{j_1}) - \mathbb{E}\big(I_{t_i,t}(W_{j_1}, W_{j_2}) + I_{t_i,t}(W_{j_2}, W_{j_1}) \mid \mathcal{N}_m(W_{j_1}, W_{j_2})\big), \\
& I_{t_i,t}(N, W_{j_1}) + I_{t_i,t}(W_{j_1}, N) - \mathbb{E}\big(I_{t_i,t}(N, W_{j_1}) + I_{t_i,t}(W_{j_1}, N) \mid \mathcal{N}_m(N, W_{j_1})\big),
\end{aligned}
$$

are independent of $\mathcal{F}_{t_i}$. Hence, by (B.92), Fact B.28 (i) and (B.93), (B.94) we have the following estimation

$$\begin{aligned}
&\|\tilde{R}_m^{df-M}(t)\|_{\mathfrak{L}^2(\Omega)} \\
&\quad\leqslant C \sum_{j_1,j_2=1}^{m_{\mathrm{w}}} \|\widetilde{\mathcal{L}}_{j_1,h_i} b^{j_2}(U_i^{df})\|_{\mathfrak{L}^2(\Omega)} \cdot \|I_{t_i,t}(W_{j_1},W_{j_2}) + I_{t_i,t}(W_{j_2},W_{j_1})\|_{\mathfrak{L}^2(\Omega)} \\
&\qquad + C \sum_{j_1=1}^{m_{\mathrm{w}}} \|L_{-1} b_1^j(U_i^{df})\|_{\mathfrak{L}^2(\Omega)} \cdot \|I_{t_i,t}(N,W_{j_1}) + I_{t_i,t}(W_{j_1},N)\|_{\mathfrak{L}^2(\Omega)} \\
&\qquad + C\|L_{-1} c(U_i^{df})\|_{\mathfrak{L}^2(\Omega)} \cdot \|I_{t_i,t}(N,N)\|_{\mathfrak{L}^2(\Omega)} \\
&\quad\leqslant C(t-t_i),
\end{aligned}$$

for $t \in [t_i, t_{i+1}]$, which ends the proof. ■

By the Lemma B.15 there exists a constant $C > 0$ such that for all $m \in \mathbb{N}$ and arbitrary discretization (B.68)

$$\sup_{t\in[0,T]} \mathbb{E}\|\tilde{R}_m^{df-M}(t)\|^2 \leqslant C \max_{0\le i\le m-1} (t_{i+1}-t_i)^2. \tag{B.95}$$

B.4. Properties of stochastic processes on given interval and discretization

Now we show basic properties about stochastic processes on a given interval $[0,T]$ and discretization points. Without loss of generality let $W = (W_1, W_2)^T$ and N be respectively two-dimensional Wiener process and one-dimensional Poisson process. Let $m \in \mathbb{N}$ and let

$$0 = t_0 < t_1 < \ldots < t_m = T,$$

be an arbitrary discretization of the interval $[0,T]$. We have a following vectors of information about processes:

$$\begin{aligned}
\mathcal{N}_m(W) &= [W_1(t_1), W_1(t_2), \ldots, W_1(t_m), W_2(t_1), W_2(t_2), \ldots, W_2(t_m)], \\
\mathcal{N}_m(W_1) &= [W_1(t_1), W_1(t_2), \ldots, W_1(t_m)], \\
\mathcal{N}_m(W_2) &= [W_2(t_1), W_2(t_2), \ldots, W_2(t_m)], \\
\mathcal{N}_m(N) &= [N(t_1), N(t_2), \ldots, N(t_m)], \\
\mathcal{N}_m(Z_1,Z_2) &= \mathcal{N}_m(Z_1) \cup \mathcal{N}_m(Z_2), \quad Z_1, Z_2 \in \{W, W_1, W_2, N\}.
\end{aligned}$$

Definition B.16. Let X be a square integrable random variable. *Conditional variance* by σ-algebra $\mathcal{G}$ is defined by

$$Var(X \mid \mathcal{G}) := \mathbb{E}\Big((X - \mathbb{E}(X \mid \mathcal{G}))^2 \mid \mathcal{G}\Big) = \mathbb{E}(X^2 \mid \mathcal{G}) - \big(\mathbb{E}(X \mid \mathcal{G})\big)^2.$$

Lemma B.17. *Let $Z \in \{W_1, W_2, N\}$ we have that for all $t \in [0, T]$*

$$Var\big(Z(t) - Z(t_i) \mid \mathcal{N}_m(Z)\big) = Var\Big(Z(t) \mid \mathcal{N}_m(Z)\Big).$$

Proof. By the Definition B.16 and fact that $\sigma(Z(t_i)) \subset \sigma(\mathcal{N}_m(Z))$ we have that

$$\begin{aligned} Var\big(Z(t) - Z(t_i) \mid \mathcal{N}_m(Z)\big) &= \mathbb{E}\Big(\big(Z(t) - Z(t_i) - \mathbb{E}(Z(t) - Z(t_i) \mid \mathcal{N}_m(Z))\big)^2 \mid \mathcal{N}_m(Z)\Big) \\ &= \mathbb{E}\Big(Z(t) - \mathbb{E}\big(Z(t) \mid \mathcal{N}_m(Z)\big)\Big)^2 \\ &= Var\big(Z(t) \mid \mathcal{N}_m(Z)\big). \end{aligned}$$

That ends the proof. ■

Lemma B.18. *Let X, Y be a stochastic processes on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $X, Y : \Omega \times [0, \infty) \to \mathbb{R}$ such that both are $\mathcal{F} \otimes \mathcal{B}([0, \infty))/\mathcal{B}(\mathbb{R})$-measurable and independent (i.e. $\mathcal{F}^X_\infty \perp\!\!\!\perp \mathcal{F}^Y_\infty$, where $\mathcal{F}^Z_\infty = \sigma\Big(\bigcup_{t \geq 0} \sigma\big(Z(t)\big)\Big)$ for $Z \in \{X, Y\}$). We assume that for all $t \geqslant 0$*

$$\mathbb{E}|X(t)| < \infty, \; \mathbb{E}|Y(t)| < \infty.$$

Let $m, n \in \mathbb{N}$ and points $t^X_1, \ldots, t^X_m, t^Y_1, \ldots, t^Y_n$ satisfy $0 \leqslant t^X_1 < t^X_2 < \ldots < t^X_m$, $0 \leqslant t^Y_1 < t^Y_2 < \ldots < t^Y_n$. Then we have that

$$\mathcal{F}^X_\infty \perp\!\!\!\perp_{\sigma(X(t^X_1), \ldots, X(t^X_m), Y(t^Y_1), \ldots, Y(t^Y_n))} \mathcal{F}^Y_\infty. \tag{B.96}$$

Proof. We set $\mathcal{F}_1 = \mathcal{F}^X_\infty$ and $\mathcal{F}_3 = \mathcal{F}^Y_\infty$ and we define

$$\mathcal{F}_2 := \sigma\big(X(t^X_1), \ldots, X(t^X_m), Y(t^Y_1), \ldots, Y(t^Y_n)\big).$$

Lets take random variables Y_3-integrable and $\mathcal{F}_3$-measurable. Then we have that

$$\mathbb{E}(Y_3 \mid \mathcal{F}_1 \vee \mathcal{F}_2) = \mathbb{E}\big(Y_3 \mid \underbrace{\mathcal{F}^X_\infty \vee \sigma\big(X(t^X_1), \ldots, X(t^X_m)\big)}_{\mathcal{H}} \vee \underbrace{\sigma\big(Y(t^Y_1), \ldots, Y(t^Y_n)\big)}_{\mathcal{G}}\big).$$

By the facts that

$$\mathcal{F}_\infty^X \vee \sigma\big(X(t_1^X), \ldots, X(t_m^X)\big) = \sigma(\mathcal{F}_\infty^X \cup \underbrace{\sigma\big(X(t_1^X), \ldots, X(t_m^X)\big)}_{\subset \mathcal{F}_\infty^X}) \subset \mathcal{F}_\infty^X,$$

$$\underbrace{\sigma(Y_3)}_{\subset \mathcal{F}_\infty^Y} \vee \underbrace{\sigma\big(Y(t_1^Y), \ldots, Y(t_n^Y)\big)}_{\subset \mathcal{F}_\infty^Y} \subset \mathcal{F}_\infty^Y.$$

From independence of $\mathcal{F}_\infty^X \perp\!\!\!\perp \mathcal{F}_\infty^Y$ we have that

$$\mathcal{F}_\infty^X \vee \sigma\big(X(t_1^X), \ldots, X(t_m^X)\big) \perp\!\!\!\perp \sigma(Y_3) \vee \sigma\big(Y(t_1^Y), \ldots, Y(t_n^Y)\big).$$

From Lemma A.20 we have that

$$\mathbb{E}(Y_3 \mid \mathcal{F}_{12}) = \mathbb{E}\big(Y_3 \mid \sigma\big(Y(t_1^Y), \ldots, Y(t_n^Y)\big)\big).$$

Then by $\sigma\big(X(t_1^X), \ldots, X(t_m^X)\big) \subset \mathcal{F}_\infty^X$, we have that

$$\sigma\big(X(t_1^X), \ldots, X(t_m^X)\big) \perp\!\!\!\perp \sigma(Y_3) \vee \sigma\big(Y(t_1^Y), \ldots, Y(t_n^Y)\big).$$

And again from Lemma A.20 we have that

$$\begin{aligned}\mathbb{E}(Y_3 \mid \mathcal{F}_{12}) &= \mathbb{E}\big(Y_3 \mid \sigma\big(X(t_1^X), \ldots, X(t_m^X)\big) \vee \sigma\big(Y(t_1^Y), \ldots, Y(t_m^Y)\big)\big) \\ &= \mathbb{E}(Y_3 \mid \mathcal{F}_2).\end{aligned}$$

So we prove (B.96) and that ends the proof. ■

Fact B.19. *We have that for all* $\alpha, \beta, t \in \mathbb{R}$, *such that* $0 \leqslant \alpha \leqslant t$, $0 \leqslant \beta \leqslant t$

$$\mathbb{E}\big(N(t) - N(\alpha) \mid \sigma(\mathcal{N}_m(W, N))\big) = \mathbb{E}\big(N(t) - N(\alpha) \mid \sigma(\mathcal{N}_m(N))\big), \tag{B.97}$$

$$\mathbb{E}\big(W_1(s) - W_1(\beta) \mid \sigma(\mathcal{N}_m(W, N))\big) = \mathbb{E}\big(W_1(s) - W_1(\beta) \mid \sigma(\mathcal{N}_m(W_1))\big), \tag{B.98}$$

$$\begin{aligned}&\mathbb{E}\big((N(t) - N(\alpha)) \cdot (W_1(s) - W_1(\beta)) \mid \sigma(\mathcal{N}_m(W_1, N))\big) \\ &\quad = \mathbb{E}\big(N(t) - N(\alpha) \mid \sigma(\mathcal{N}_m(N))\big) \cdot \mathbb{E}\big(W_1(s) - W_1(\beta) \mid \sigma(\mathcal{N}_m(W_1))\big),\end{aligned} \tag{B.99}$$

$$\begin{aligned}&\mathbb{E}\big((W_1(t) - W_1(\alpha)) \cdot (W_2(s) - W_2(\beta)) \mid \sigma(\mathcal{N}_m(W_1, W_2))\big) \\ &\quad = \mathbb{E}\big(W_1(t) - W_1(\alpha) \mid \sigma(\mathcal{N}_m(W_1))\big) \cdot \mathbb{E}\big(W_2(s) - W_2(\beta) \mid \sigma(\mathcal{N}_m(W_2))\big).\end{aligned} \tag{B.100}$$

Proof. The proof of (B.97) and (B.98) is a natural consequence of fact that N, W_1 are independent. The proof of (B.99) goes as follows. Lets substitute in Lemma B.18

$X = N$ and $Y = W_1$, X, Y are independent and both are integrable. From independence we have that $\mathbb{E}|X(t)Y(s)| < \infty$ for all $t \geqslant 0,\ s \geqslant 0$. Let

$$\begin{aligned}
\mathcal{F}_1 &= \mathcal{F}_\infty^N, \\
\mathcal{F}_3 &= \mathcal{F}_\infty^{W_1}, \\
\mathcal{F}_2 &= \sigma\big(X(t_1^X), \ldots, X(t_m^X), Y(t_1^Y), \ldots, Y(t_n^Y)\big),
\end{aligned}$$

and random variables

$$\begin{aligned}
Y_1 &= N(t) - N(\alpha), \qquad & t \geqslant \alpha \geqslant 0, \\
Y_3 &= W_1(s) - W_1(\beta), \qquad & s \geqslant \beta \geqslant 0.
\end{aligned}$$

Of course we have that $\sigma(Y_1) \subset \mathcal{F}_1$, $\sigma(Y_3) \subset \mathcal{F}_3$, and $\mathbb{E}|Y_1| < \infty$, $\mathbb{E}|Y_3| < \infty$ and from independence of N, W_1 By the (B.96) and Proposition A.23 we have that $\mathbb{E}|Y_1Y_3| = \mathbb{E}|Y_1| \cdot \mathbb{E}|Y_3| < \infty$ the conditional expectation

$$\begin{aligned}
&\mathbb{E}\big((N(t) - N(\alpha)) \cdot (W_1(s) - W_1(\beta)) \mid \sigma(\mathcal{N}_m(W_1, N))\big) \\
&\qquad = \mathbb{E}\big(Y_1Y_3 \mid \mathcal{F}_2\big) = \mathbb{E}\big(Y_1 \mid \mathcal{F}_2\big) \cdot \mathbb{E}\big(Y_3 \mid \mathcal{F}_2\big) \\
&\qquad = \mathbb{E}\big(N(t) - N(\alpha) \mid \sigma(\mathcal{N}_m(W_1, N))\big) \\
&\qquad\quad \times \mathbb{E}\big(W_1(s) - W_1(\beta) \mid \sigma(\mathcal{N}(W_1, N)\big),
\end{aligned}$$

we have that

$$\underbrace{\sigma\big(W_1(t_1^{W_1}), \ldots, W_1(t_m^{W_1})\big)}_{\subset \mathcal{F}_\infty^{W_1}} \;\perp\!\!\!\perp\; \sigma\big((N(t) - N(\alpha)\big) \vee \underbrace{\sigma\big(N(t_1^N), \ldots, N(t_m^N)\big)}_{\subset \mathcal{F}_\infty^N}, \tag{B.101}$$

so from Lemma A.20 and by (B.101) we have

$$\mathbb{E}\big(N(t) - N(\alpha) \mid \sigma(\mathcal{N}_m(W_1, N))\big) = \mathbb{E}\big(N(t) - N(\alpha) \mid \sigma(\mathcal{N}_m(N))\big),$$

similarly

$$\mathbb{E}(W_1(s) - W_1(\beta) \mid \sigma(\mathcal{N}_m(W_1, N))) = \mathbb{E}\big(W_1(s) - W_1(\beta) \mid \sigma(\mathcal{N}_m(W_1))\big).$$

And finally we have that

$$\begin{aligned}
&\mathbb{E}\big((N(t) - N(\alpha)) \cdot (W_1(s) - W_1(\beta)) \mid \sigma(\mathcal{N}_m(W_1, N))\big) \\
&\qquad = \mathbb{E}\big(N(t) - N(\alpha) \mid \sigma(\mathcal{N}_m(N))\big) \cdot \mathbb{E}\big(W_1(s) - W_1(\beta) \mid \sigma(\mathcal{N}_m(W_1))\big).
\end{aligned}$$

The proof of (B.100) goes analogously as (B.99). This ends the proof. ■

Theorem B.20 ([14])**.** *For fixed* $m \geqslant 1$*, conditioned with* $W_1(t_1), \ldots, W_1(t_m)$*,* $0 = t_0 < t_1 < \ldots < t_m$*, stochastic process* W_1 *is Gaussian with mean*

$$m_{t_i,t_{i+1}}(t) = \frac{(t - t_i) \cdot W_1(t_{i+1}) + (t_{i+1} - t) \cdot W_1(t_i)}{t_{i+1} - t_i},$$

and covariance function

$$r_{t_i,t_{i+1}}(s,t) = \frac{(t_{i+1} - s \vee t) \cdot (s \wedge t - t_i)}{t_{i+1} - t_i},$$

on the interval $[t_i, t_{i+1}]$ *for* $i = 0, \ldots, m-1$*, and with mean* $W_1(t_m)$ *and covariance function*

$$s \wedge t - t_m,$$

on $[t_m, \infty)$*.*

Lemma B.21. *For all* $i = 0, 1, \ldots, m-1$ *and* $t \in [t_i, t_{i+1}]$

(i)

$$\mathbb{E}\big(W_1(t) \mid \mathcal{N}_m(W_1)\big) = \frac{(t - t_i) \cdot W_1(t_{i+1}) + (t_{i+1} - t) \cdot W_1(t_i)}{t_{i+1} - t_i} \quad a.s.,$$

(ii)

$$\mathbb{E}\Big(\big|W_1(t) - \mathbb{E}\big(W_1(t) \mid \mathcal{N}_m(W_1)\big)\big|^2 \mid \mathcal{N}_m(W_1)\Big) = \frac{(t_{i+1} - t)(t - t_i)}{(t_{i+1} - t_i)} \quad a.s.,$$

and, in particular,

$$\mathbb{E}\big|W_1(t) - \mathbb{E}\big(W_1(t) \mid \mathcal{N}_m(W_1)\big)\big|^2 = \frac{(t_{i+1} - t)(t - t_i)}{(t_{i+1} - t_i)}.$$

Proof. Proof of this Lemma follow directly from Theorem B.20. ∎

Lemma B.22. *For all* $i = 0, 1, \ldots, m-1$ *and* $t \in [t_i, t_{i+1}]$ *we have that*

(i)

$$\mathbb{E}\big(N(t) \mid \mathcal{N}_m(N)\big) = \frac{N(t_{i+1}) \cdot \Lambda(t, t_i) + N(t_i) \cdot \Lambda(t_{i+1}, t)}{\Lambda(t_{i+1}, t_i)} \quad a.s., \tag{B.102}$$

(ii)

$$\mathbb{E}\Big(\big|N(t) - \mathbb{E}\big(N(t) \mid \mathcal{N}_m(N)\big)\big|^2 \mid \mathcal{N}_m(N)\Big) = \big(N(t_{i+1}) - N(t_i)\big) \cdot \frac{\Lambda(t_{i+1}, t) \cdot \Lambda(t, t_i)}{\big(\Lambda(t_{i+1}, t_i)\big)^2} \quad a.s., \tag{B.103}$$

and, in particular,

$$\mathbb{E}\big|N(t) - \mathbb{E}\big(N(t) \mid \mathcal{N}_m(N)\big)\big|^2 = \frac{\Lambda(t_{i+1}, t) \cdot \Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)}. \tag{B.104}$$

Proof. Let $t = t_i$, for $i = 0, 1, \dots, m$. In this case we get directly (B.102), (B.103) and (B.104). Now, let $t \in (t_i, t_{i+1})$ for $i = 0, 1, \dots, m-1$. From the fact that the process N has independent increments and is based on results from [3], we obtain that conditioned on $\mathcal{N}_m(N)$ the increment $N(t) - N(t_i)$ is a binomial random variable with the number of trials $N(t_{i+1}) - N(t_i)$ and with the probability of success in each trial equal to $\dfrac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)}$. It means that

$$\mathbb{E}\big(N(t) - N(t_i) \mid \mathcal{N}_m(N)\big) = \frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)} \cdot \big(N(t_{i+1}) - N(t_i)\big), \tag{B.105}$$

$$Var\big(N(t) - N(t_i) \mid \mathcal{N}_m(N)\big) = \frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)} \cdot \Big(1 - \frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)}\Big) \cdot \big(N(t_{i+1}) - N(t_i)\big). \tag{B.106}$$

We start with proof of (B.102), by (B.105) we have that

$$\begin{aligned}
\mathbb{E}\big(N(t) \mid \mathcal{N}_m(N)\big) &= \mathbb{E}\big(N(t) - N(t_i) \mid \mathcal{N}_m(N)\big) + N(t_i) \\
&= \big(N(t_{i+1}) - N(t_i)\big) \cdot \frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)} + N(t_i) \\
&= \frac{N(t_{i+1}) \cdot \Lambda(t, t_i) + N(t_i) \cdot \Lambda(t_{i+1}, t)}{\Lambda(t_{i+1}, t_i)},
\end{aligned}$$

which gives (B.102). Then, by Definition B.17 and (B.106) we have that

$$\begin{aligned}
&\mathbb{E}\Big(\big|N(t) - \mathbb{E}(N(t) \mid \mathcal{N}_m(N))\big|^2 \mid \mathcal{N}_m(N)\Big) \\
&\quad = \mathbb{E}\Big(\big|\big(N(t) - N(t_i)\big) - \mathbb{E}\big(N(t) - N(t_i) \mid \mathcal{N}_m(N)\big)\big|^2 \mid \mathcal{N}_m(N)\Big) \\
&\quad = Var\Big(\big(N(t) - N(t_i)\big) \mid \mathcal{N}_m(N)\Big) \\
&\quad = \big(N(t_{i+1}) - N(t_i)\big) \cdot \frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)} \cdot \Big(1 - \frac{\Lambda(t, t_i)}{\Lambda(t_{i+1}, t_i)}\Big) \\
&\quad = \big(N(t_{i+1}) - N(t_i)\big) \cdot \frac{\Lambda(t_{i+1}, t) \cdot \Lambda(t, t_i)}{\big(\Lambda(t_{i+1}, t_i)\big)^2},
\end{aligned}$$

which gives (B.103).

Since

$$\mathbb{E}\big(N(t_{i+1}) - N(t_i)\big) = \Lambda(t_{i+1}, t_i),$$

and

$$\mathbb{E}\big|N(t) - \mathbb{E}\big(N(t) \mid \mathcal{N}_m(N)\big)\big|^2 = \mathbb{E}\Big(\mathbb{E}\big(\big|N(t) - \mathbb{E}\big(N(t) \mid \mathcal{N}_m(N)\big)\big|^2 \mid \mathcal{N}_m(N)\big)\Big),$$

we have (B.104). ■

Let N, W_1, W_2 be an independent Poisson process and two one-dimensional Wiener processes. We define double Itô integrals of the following form

$$I_{a,b}(Y,Z) = \int_a^b \int_a^{v-} \mathrm{d}Y(u)\mathrm{d}Z(v),$$

for stochastic processes $Y, Z \in \{N, W_1, W_2\}$ and $a, b \in \mathbb{R}_+$.

Lemma B.23 ([61])**.** *Let $0 < s < t$, $\{\tau_i\}_{i=0}^{N(t)}$ be a sequence such that $\tau_0 = 0$ and $N(u) = N(v)$ for each $u, v \in [\tau_i, \tau_{i+1})$ then*

$$\begin{aligned}
I_{s,t}(W_1,W_1) &= \frac{1}{2}\Big(\big(W_1(t)-W_1(s)\big)^2 - (t-s)\Big), && \text{(B.107)}\\
I_{s,t}(W_1,W_2) &= \int_s^t \int_s^u \mathrm{d}W_1(z)\mathrm{d}W_2(u), && \text{(B.108)}\\
I_{s,t}(W_1,W_2) + I_{s,t}(W_2,W_1) &= \big(W_1(t)-W_1(s)\big)\big(W_2(t)-W_2(s)\big), && \text{(B.109)}\\
I_{s,t}(N,N) &= \frac{1}{2}\Big(\big(N(t)-N(s)\big)^2 - \big(N(t)-N(s)\big)\Big), && \text{(B.110)}\\
I_{s,t}(W_1,N) &= \sum_{j=N(s)+1}^{N(t)} W_1(\tau_j) - W_1(s)\big(N(t)-N(s)\big), && \text{(B.111)}\\
I_{s,t}(N,W_1) &= \big(W_1(t)-W_1(s)\big)\big(N(t)-N(s)\big) - I_{s,t}(W_1,N), && \text{(B.112)}\\
I_{s,t}(N,W_1) + I_{s,t}(W_1,N) &= \big(W_1(t)-W_1(s)\big)\big(N(t)-N(s)\big). && \text{(B.113)}
\end{aligned}$$

Lemma B.24. *For all $i = 0, 1, \ldots, m-1$ and $t \in [t_i, t_{i+1}]$ holds*

$$\mathbb{E}\big(I_{t_i,t}(W_1,W_1) \mid \mathcal{N}_m(W_1)\big) = \left(\frac{t-t_i}{t_{i+1}-t_i}\right)^2 \cdot I_{t_i,t_{i+1}}(W_1,W_1).$$

Proof. By Lemma B.21 we have that

$$\mathbb{E}\big(W_1(t) - W_1(t_i) \mid \mathcal{N}_m(W_1)\big) = \frac{t-t_i}{t_{i+1}-t_i} \cdot \big(W_1(t_{i+1}) - W_1(t_i)\big). \qquad \text{(B.114)}$$

By Lemma B.23, Lemma B.17 and Definition B.16 we have that

$$\begin{aligned}
\mathbb{E}\big(I_{t_i,t}(W_1,W_1) \mid \mathcal{N}_m(W_1)\big) &= \frac{1}{2}\,\mathbb{E}\Big(\big(W_1(t)-W_1(t_i)\big)^2 \mid \mathcal{N}_m(W_1)\Big) - \frac{1}{2}(t-t_i)\\
&= \frac{1}{2}\,Var\big(W_1(t) \mid \mathcal{N}_m(W_1)\big)\\
&\quad + \frac{1}{2}\,\Big(\mathbb{E}\big(W_1(t)-W_1(t_i) \mid \mathcal{N}_m(W_1)\big)\Big)^2 - \frac{1}{2}(t-t_i).
\end{aligned}$$

Finally, from Lemma B.21 and (B.114) we have that

$$\begin{aligned}\mathbb{E}\big(I_{t_i,t}(W_1,W_1)\mid \mathcal{N}_m(W_1)\big) &= \frac{1}{2}\Bigg(\frac{(t_{i+1}-t)(t-t_i)}{(t_{i+1}-t_i)}-(t-t_i)\\ &\qquad+\left(\frac{t-t_i}{t_{i+1}-t_i}\right)^2\cdot\big(W_1(t_{i+1})-W_1(t_i)\big)^2\Bigg)\\ &= \left(\frac{t-t_i}{t_{i+1}-t_i}\right)^2\cdot I_{t_i,t_{i+1}}(W_1,W_1).\end{aligned}$$

Which ends the proof. ∎

Lemma B.25. *For all $i=0,1,\ldots,m-1$ and $t\in[t_i,t_{i+1}]$ we have that*

$$\begin{aligned}&\mathbb{E}\big(I_{t_i,t}(W_1,W_2)+I_{t_i,t}(W_2,W_1)\mid \mathcal{N}_m(W_1,W_2)\big)\\ &\qquad= \left(\frac{t-t_i}{t_{i+1}-t_i}\right)^2\cdot\big(I_{t_i,t_{i+1}}(W_1,W_2)+I_{t_i,t_{i+1}}(W_2,W_1)\big).\end{aligned}$$

Proof. By Lemma B.109, Lemma B.21 (like in (B.114)), independence of W_1, W_2 and Fact B.19 it follows that

$$\begin{aligned}&\mathbb{E}\big(I_{t_i,t}(W_1,W_2)+I_{t_i,t}(W_2,W_1)\mid \mathcal{N}_m(W_1,W_2)\big)\\ &\quad= \mathbb{E}\Big(\big(W_1(t)-W_1(t_i)\big)\cdot\big(W_2(t)-W_2(t_i)\big)\mid \mathcal{N}_m(W_1,W_2)\Big)\\ &\quad= \frac{t-t_i}{t_{i+1}-t_i}\cdot\big(W_1(t_{i+1})-W_1(t_i)\big)\cdot\frac{t-t_i}{t_{i+1}-t_i}\cdot\big(W_2(t_{i+1})-W_2(t_i)\big)\\ &\quad= \left(\frac{t-t_i}{t_{i+1}-t_i}\right)^2\cdot\big(I_{t_i,t_{i+1}}(W_1,W_2)+I_{t_i,t_{i+1}}(W_2,W_1)\big).\end{aligned}$$

Which ends the proof. ∎

Lemma B.26. *For all $i=0,1,\ldots,m-1$ and $t\in[t_i,t_{i+1}]$ it follows that*

$$\begin{aligned}&\mathbb{E}\big(I_{t_i,t}(W_1,N)+I_{t_i,t}(N,W_1)\mid \mathcal{N}_m(W_1,N)\big)\\ &\quad= \frac{\Lambda(t,t_i)}{\Lambda(t_{i+1},t_i)}\cdot\frac{t-t_i}{t_{i+1}-t_i}\cdot\big(I_{t_i,t_{i+1}}(W_1,N)+I_{t_i,t_{i+1}}(N,W_1)\big).\end{aligned}$$

Proof. By Lemma B.22 we have that

$$\mathbb{E}\big(N(t)-N(t_i)\mid \mathcal{N}_m(N)\big)=\frac{\Lambda(t,t_i)}{\Lambda(t_{i+1},t_i)}\big(N(t_{i+1})-N(t_i)\big). \tag{B.115}$$

By Fact B.19, Lemma B.21, Lemma B.23, (like in (B.114)), (B.115) and independence of W_1, N, we have that

$$\begin{aligned} E\big(I_{t_i,t}(W_1,N) + I_{t_i,t}(N,W_1) \mid \mathcal{N}_m(W_1,N)\big) \\ &= \mathbb{E}\Big(\big(W_1(t) - W_1(t_i)\big) \cdot \big(N(t) - N(t_i)\big) \mid \mathcal{N}_m(W_1,N)\Big) \\ &= \mathbb{E}\big(W_1(t) - W_1(t_i) \mid \mathcal{N}_m(W_1)\big) \cdot \mathbb{E}\big(N(t) - N(t_i) \mid \mathcal{N}_m(N)\big) \\ &= \frac{t - t_i}{t_{i+1} - t_i} \cdot \big(W_1(t_{i+1}) - W_1(t_i)\big) \cdot \frac{\Lambda(t,t_i)}{\Lambda(t_{t_{i+1},t_i})} \cdot \big(N(t_{i+1}) - N(t_i)\big) \\ &= \frac{\Lambda(t,t_i)}{\Lambda(t_{t_{i+1},t_i})} \cdot \frac{t - t_i}{t_{i+1} - t_i} \cdot \big(I_{t_i,t_{i+1}}(W_1,N) + I_{t_i,t_{i+1}}(N,W_1)\big). \end{aligned}$$

Which ends the proof. ■

Lemma B.27. *For all $i = 0, 1, \dots, m-1$ and $t \in [t_i, t_{i+1}]$ we have that*

$$\mathbb{E}\big(I_{t_i,t}(N,N) \mid \mathcal{N}_m(N)\big) = \left(\frac{\Lambda(t,t_i)}{\Lambda(t_{i+1},t_i)}\right)^2 \cdot I_{t_i,t_{i+1}}(N,N).$$

Proof. By Lemma B.23 and by Definition B.16 we have that

$$\begin{aligned} \mathbb{E}\big(I_{t_i,t}(N,N) \mid \mathcal{N}_m(N)\big) &= \frac{1}{2}\,\mathbb{E}\Big(\big(N(t) - N(t_i)\big)^2 - \big(N(t) - N(t_i)\big) \mid \mathcal{N}_m(N)\Big) \\ &= \frac{1}{2}\,Var\big(N(t) \mid \mathcal{N}_m(N)\big) + \frac{1}{2}\,\Big(\mathbb{E}\big(N(t) - N(t_i) \mid \mathcal{N}_m(N)\big)\Big)^2 \\ &\quad - \frac{1}{2}\,\mathbb{E}\big(N(t) - N(t_i) \mid \mathcal{N}_m(N)\big). \end{aligned}$$

Finally, from Lemma B.22 and (B.115) we have that

$$\begin{aligned} \mathbb{E}\big(I_{t_i,t}(N,N) \mid \mathcal{N}_m(N)\big) &= \frac{1}{2}\left(\frac{\Lambda(t_{i+1},t)\cdot\Lambda(t,t_i)}{\big(\Lambda(t_{i+1},t_i)\big)^2} \cdot \big(N(t_{i+1}) - N(t_i)\big)\right. \\ &\quad + \left(\frac{\Lambda(t,t_i)}{\Lambda(t_{i+1},t_i)}\right)^2 \cdot \big(N(t_{i+1}) - N(t_i)\big)^2 \\ &\quad \left. - \frac{\Lambda(t,t_i)}{\Lambda(t_{i+1},t_i)} \cdot \big(N(t_{i+1}) - N(t_i)\big)\right). \end{aligned}$$

By the fact that

$\Lambda(t_{i+1},t) - \Lambda(t_{i+1},t_i) = m(t_{i+1}) - m(t) - m(t_{i+1}) - m(t_i) = -(m(t) - m(t_i)) = -\Lambda(t,t_i)$,

we get

$$\mathbb{E}\big(I_{t_i,t}(N,N) \mid \mathcal{N}_m(N)\big) = \left(\frac{\Lambda(t,t_i)}{\Lambda(t_{i+1},t_i)}\right)^2 \cdot I_{t_i,t_{i+1}}(N,N).$$

Which ends the proof. ■

Fact B.28. *(i) There exists $C > 0$ such that for all $0 \leqslant s \leqslant t \leqslant T$ and $Y, Z \in \{N, W_1, W_2\}$ we have*

$$\mathbb{E}\big|I_{s,t}(Y,Z)\big|^2 \leqslant C(t-s)^2. \tag{B.116}$$

(ii) For all $0 \leqslant s \leqslant t \leqslant T$ and $Y, Z \in \{N, W_1, W_2\}$ the stochastic integral $I_{s,t}(Y,Z)$ is independent of $\mathcal{F}_s$.

Proof. The proof of (i) can be straightforwardly delivered from (A.2), Lemma B.23, the isometry for stochastic integrals driven by martingales and by the independence of W_1, W_2 and N. To show methodology of proof, we present only the case when $(Y,Z) \in \big\{(W_1,W_2),(W_1,N)\big\}$. Other cases goes in the same way.

For stochastic integral $\mathbb{E}|I_{s,t}(W_1,W_2)|^2$ we have that

$$\mathbb{E}\big|I_{s,t}(W_1,W_2)\big|^2 = \mathbb{E}\Big|\int_s^t\int_s^{v-} \mathrm{d}W_1(u)\mathrm{d}W_2(v)\Big|^2 = \int_s^t \mathbb{E}\Big|\int_s^{v-} \mathrm{d}W_1(u)\Big|^2 \mathrm{d}v = \int_s^t\int_s^{v-} \mathrm{d}u\mathrm{d}v. \tag{B.117}$$

By the (B.117) we get (B.116) in the case when we consider multiple stochastic integrals for $(W_1,W_2),(W_2,W_1)$.

For stochastic integral $\mathbb{E}|I_{s,t}(W_1,N)|^2$ by Itô isometry for $\tilde{N}(t) = N(t) - \lambda(t)$, Hölder inequality assumption (E_{MD}) we have that

$$\begin{aligned}
\mathbb{E}\big|I_{s,t}(W_1,N)\big|^2 &= \mathbb{E}\Big|\int_s^t\int_s^{v-} \mathrm{d}W_1(u)\mathrm{d}N(v)\Big|^2 \\
&\leqslant \mathbb{E}\Big|\int_s^t\int_s^{v-} \mathrm{d}W_1(u)\mathrm{d}\tilde{N}(v)\Big|^2 + \mathbb{E}\Big|\int_s^t\int_s^{v-} \mathrm{d}W_1(u)\lambda(v)\mathrm{d}v\Big|^2 \\
&\leqslant \int_s^t \mathbb{E}\Big|\int_s^{v-} \mathrm{d}W_1(u)\Big|^2 \lambda(v)\mathrm{d}v + C\int_s^t \mathbb{E}\Big|\int_s^{v-} \mathrm{d}W_1(u)\Big|^2 \mathrm{d}v \\
&\leqslant C\int_s^t\int_s^{v-} \mathrm{d}u\mathrm{d}v.
\end{aligned} \tag{B.118}$$

By the (B.118) we get (B.116) in the case when we consider multiple stochastic integrals for $(W_1,N),(W_2,N)$. Other cases $(W_1,W_1),(W_2,W_2),(N,W_1),(N,W_2),(N,N)$ goes in analogous way as (B.117) and (B.118). That ends the first part of proof. □

For the proof of (ii) note that directly from (B.107) and (B.110) we have that $I_{s,t}(Y,Y)$, $Y \in \{N, W_1, W_2\}$, is independent of $\mathcal{F}_s$. So the only case of interest is when

$(Y,Z) \in \{(N,W_1),(W_1,N),(W_1,W_2)\}$. (Case $(N,W_2),(W_2,N)$ are exactly the same as considered $(N,W_1),(W_1,N)$).
Let fix $s,t \in [0,T]$, $s \leqslant t$, and let $\Delta_n = \{\alpha_{0,n},\alpha_{1,n},\ldots,\alpha_{m,n}\}$, $n \in \mathbb{N}$, be a sequence of discretization of $[s,t]$ such that $s = \alpha_{0,n} < \alpha_{1,n} < \ldots < \alpha_{m,n} = t$ and $\lim\limits_{n\to+\infty} \|\Delta_m\| = 0$, where $\|\Delta_m\| = \max\limits_{0\leq i\leq m-1}(\alpha_{i+1,n}-\alpha_{i,n})$.

Firstly we consider case (N,W_1). By the definition of Itô integral we have that

$$\int_s^t \int_s^{v-} \mathrm{d}N(u)\mathrm{d}W_1(v) = \int_s^t \big(N(v)-N(s)\big)\mathrm{d}W_1(v).$$

Let us define process

$$\bar{N}_n(u) = \sum_{i=0}^{m-1} \big(N(\alpha_{i,n})-N(s)\big)\cdot \mathbb{1}_{(\alpha_{i,n},\alpha_{i+1,n}]}(u),$$

We show that the defined process $\bar{N}_n(u)$ converges in space $\mathfrak{L}^2_{W_1}(\Omega)$ to process $N(u)-N(s)$.

$$\begin{aligned}
\mathbb{E}\int_s^t \big|\bar{N}_n(u) - \big(N(u)-N(s)\big)\big|^2\mathrm{d}u &= \sum_{i=0}^{m-1}\int_{\alpha_{i,n}}^{\alpha_{i+1,n}} \mathbb{E}\big|N(\alpha_{i,n})-N(u)\big|^2\mathrm{d}u \\
&= \sum_{i=0}^{m-1}\int_{\alpha_{i,n}}^{\alpha_{i+1,n}} \big(m(u)-m(\alpha_{i,n})\big) + \big(m(u)-m(\alpha_{i,n})\big)^2\mathrm{d}u \\
&\leqslant \sum_{i=0}^{m-1}\int_{\alpha_{i,n}}^{\alpha_{i+1,n}} \|\lambda\|_\infty\big(u-\alpha_{i,n}\big) + \|\lambda\|_\infty^2\big(u-\alpha_{i,n}\big)^2\mathrm{d}u \\
&\leqslant \frac{1}{2}\sum_{i=0}^{m-1}\|\lambda\|_\infty\big(\alpha_{i+1,n}-\alpha_{i,n}\big)^2 + \frac{1}{3}\sum_{i=0}^{m-1}\|\lambda\|_\infty^2\big(\alpha_{i+1,n}-\alpha_{i,n}\big)^3 \\
&\leqslant \frac{1}{2}\|\lambda\|_\infty\|\Delta_n\|\big(t-s\big) + \frac{1}{3}\|\lambda\|_\infty\|\Delta_n\|^2\big(t-s\big). \qquad \text{(B.119)}
\end{aligned}$$

By (B.119) and assumption $\lim\limits_{n\to+\infty}\|\Delta_n\| = 0$ we have that

$$\mathbb{E}\int_s^t \big|\bar{N}_n(u) - \big(N(u)-N(s)\big)\big|^2\mathrm{d}u \overset{n\to\infty}{\longrightarrow} 0.$$

So $\{\bar{N}_n(u)\}_{n\in\mathbb{N}}$ is a sequence of simple processes which approximate $N(u)-N(s)$ for $u \in [s,t]$ so by the definition of Itô integral for simple function we have that

$$\int_s^t \bar{N}_n(u)\mathrm{d}W_1(u) = \sum_{i=0}^{m-1}\big(N(\alpha_{i,n})-N(u)\big)\cdot\big(W_1(\alpha_{i+1,n})-W_1(\alpha_{i,n})\big).$$

So we can define

$$I^n_{s,t}(N,W_1) = \sum_{i=0}^{m-1}(N(\alpha_{i,n}) - N(s)) \cdot (W(\alpha_{i+1,n}) - W(\alpha_{i,n})).$$

We have that

$$I_{s,t}(N,W_1) = \lim_{n\to+\infty} I^n_{s,t}(N,W_1) \quad \text{in} \quad \mathfrak{L}^2(\Omega).$$

Therefore, the sequence $\{I^n_{s,t}(N,W_1)\}_{n\in\mathbb{N}}$ converges also in probability, and by the independence of the increments of N and W_1, every random variable $I^m_{s,t}(N,W_1)$ is independent of $\mathcal{F}_s$. Hence, the limit $I_{s,t}(N,W_1)$ is also independent of $\mathcal{F}_s$. By (B.111) we have that also $I_{s,t}(W,N)$ is independent of $\mathcal{F}_s$.

Let us define process

$$\bar{W}_{1,n}(u) = \sum_{i=0}^{m-1} \big(W_1(\alpha_{i,n}) - W_1(s)\big) \cdot \mathbb{1}_{(\alpha_{i,n},\alpha_{i+1,n}]}(u).$$

We show that the defined process $\bar{W}_{1,n}(u)$ converges in space $\mathfrak{L}^2_{W_2}(\Omega)$ to process $W_1(u) - W_1(s)$.

$$\begin{aligned}
\mathbb{E}\int_s^t \big|\bar{W}_{1,n}(u) - \big(W_1(u) - W_1(s)\big)\big|^2 \mathrm{d}u &= \sum_{i=0}^{m-1}\int_{\alpha_{i,n}}^{\alpha_{i+1,n}} \mathbb{E}\big|W_1(\alpha_{i,n}) - W_1(u)\big|^2 \mathrm{d}u \\
&= \sum_{i=0}^{m-1}\int_{\alpha_{i,n}}^{\alpha_{i+1,n}} (\alpha_{i,n} - u)\mathrm{d}u \\
&\leqslant \frac{1}{2}\|\Delta_n\|^2\big(t-s\big). \qquad \text{(B.120)}
\end{aligned}$$

By (B.120) and assumption $\lim\limits_{n\to+\infty} \|\Delta_n\| = 0$ we have that

$$\mathbb{E}\int_s^t \big|\bar{W}_{1,n}(u) - \big(W_1(u) - W_1(s)\big)\big|^2 \mathrm{d}u \overset{n\to\infty}{\longrightarrow} 0.$$

So $\{\bar{W}_1(u)\}_{n\in\mathbb{N}}$ is a sequence of simple processes which approximate $W_1(u) - W_1(s)$ for $u \in [s,t]$ so by the definition of Itô integral for simple function we have that

$$\int_s^t \bar{W}_{1,n}(u)\mathrm{d}W_2(u) = \sum_{i=0}^{m-1}\big(W_1(\alpha_{i,n}) - W_1(u)\big)\cdot\big(W_2(\alpha_{i+1,n}) - W_2(\alpha_{i,n})\big).$$

So we can define

$$I^n_{s,t}(W_1,W_2) = \sum_{i=0}^{m-1}\big(W_1(\alpha_{i,n}) - W_1(s)\big)\cdot\big(W_2(\alpha_{i+1,n}) - W_2(\alpha_{i,n})\big).$$

We have that

$$I_{s,t}(W_1, W_2) = \lim_{n\to+\infty} I^n_{s,t}(W_1, W_2) \quad \text{in} \quad \mathfrak{L}^2(\Omega).$$

Therefore, the sequence $\{I^n_{s,t}(W_1, W_2)\}_{n\in\mathbb{N}}$ converges also in probability and every random variable $I^m_{s,t}(W_1, W_2)$ is independent of $\mathcal{F}_s$. Hence, the limit $I_{s,t}(W_1, W_2)$ is also independent of $\mathcal{F}_s$. This ends the proof. ■

The proof of the following fact is straightforward.

Fact B.29 ([70])**.** *Let the mappings a, b, c and λ satisfy the assumptions $(B1)$, $(B2)$ and (E).*

(i) There exists a constant $C_1 > 0$ such that for all $f \in \{b, c\}$ and $t, s \in [0, T]$ we have

$$\Big|\mathbb{E}|f(t, X(t))|^2 - \mathbb{E}|f(s, X(s))|^2\Big| \leqslant C_1|t - s|^{1/2}.$$

(ii) The mapping

$$[0, T] \ni t \to \mathbb{E}(\mathcal{Y}(t)) \in \mathbb{R}_+ \cup \{0\},$$

is continuous.

(iii) There exists a constant $C_2 > 0$ such that

$$\mathbb{E}(\sup_{t\in[0,T]} \mathcal{Y}(t)) \leqslant C_2.$$

The proof of the following fact is straightforward.

Fact B.30. *Let the mappings a, b, c and λ satisfy the assumptions $(B1_{\text{MD}})$, $(B2_{\text{MD}})$ and (E_{MD}).*

(i) There exists a constant $C_1 > 0$ such that for all $f \in \{b^1, \ldots, b^{m_{\text{w}}}, c\}$ and $t, s \in [0, T]$ we have

$$\Big|\mathbb{E}\|f(t, X(t))\|^2 - \mathbb{E}\|f(s, X(s))\|^2\Big| \leqslant C_1|t - s|^{1/2}.$$

Proof. By the Jensen and Hölder inequalities we get,

$$\begin{aligned}\Big|\mathbb{E}\|f(t, X(t))\|^2 - \mathbb{E}\|f(s, X(s))\|^2\Big| &\leqslant \mathbb{E}\Big|\big(\|f(t, X(t))\| - \|f(s, X(s))\|\big) \\ &\qquad \times \big(\|f(t, X(t))\| + \|f(s, X(s))\|\big)\Big| \\ &\leqslant \Big(\mathbb{E}\Big|\|f(t, X(t))\| - \|f(s, X(s))\|\Big|^2\Big)^{1/2} \\ &\qquad \times \Big(\mathbb{E}\Big|\|f(t, X(t))\| + \|f(s, X(s))\|\Big|^2\Big)^{1/2}\end{aligned} \tag{B.121}$$

By the assumption $(B1_{\text{MD}})$, $(B2_{\text{MD}})$ and Fact A.63 we have that

$$\begin{aligned}\Big(\mathbb{E}\big\|f(t,X(t))-f(s,X(s))\big\|^2\Big)^{1/2} \\ &\leqslant K\Big(\mathbb{E}\big\|X(t)-X(s)\big\|^2\Big)^{1/2}+TK\Big(1+\big(\mathbb{E}\|X(s)\|^2\big)^{1/2}\Big)|t-s|^{1/2} \\ &\leqslant C|t-s|^{1/2}, \end{aligned} \tag{B.122}$$

$$\begin{aligned}\Big(\mathbb{E}\big\|f(t,X(t))\big\|^2\Big)^{1/2} &\leqslant C\Big(1+\big(\mathbb{E}\|X(t)\|^2\big)^{1/2}\Big) \\ &\leqslant C|t-s|^{1/2}. \end{aligned} \tag{B.123}$$

So by the (B.121), (B.122) and (B.123) we get that

$$\Big|\mathbb{E}\|f(t,X(t))\|^2-\mathbb{E}\|f(s,X(s))\|^2\Big| \leqslant C_1|t-s|^{1/2}.$$

This ends the proof. ■

Fact B.31. *Let $0\leqslant\alpha<\beta\leqslant T$, for all $\lambda\in C([0,T])$, $\lambda:[0,T]\to(0,+\infty)$ we have for all $t\in[\alpha,\beta]$ that*

$$\left|\frac{\Lambda(t,\alpha)}{\Lambda(\beta,\alpha)}-\frac{t-\alpha}{\beta-\alpha}\right|\leqslant\frac{1}{\inf\limits_{t\in[0,T}\lambda(t)}\cdot\sup_{t,s\in[\alpha,\beta]}|\lambda(t)-\lambda(s)|$$

Proof. From the fact that λ is continuous function on interval $[0,T]$ and it is separated from 0 we have that

$$\Lambda(\beta,\alpha)\geqslant\inf_{t\in[\alpha,\beta]}\lambda(t)(\beta-\alpha)>0, \tag{B.124}$$

and of course for all $t\in[\alpha,\beta]$

$$\frac{1}{\inf\limits_{t\in[0,T]}\lambda(t)}<+\infty. \tag{B.125}$$

By the mean value theorem we have that for $s,t\in[\alpha,\beta]$, $s<t$

$$\Lambda(t,s)=(t-s)\cdot\lambda(\xi), \tag{B.126}$$

where $\xi\in(s,t)$. By (B.124), (B.126) we have that

$$\begin{aligned}\left|\frac{\Lambda(t,\alpha)}{\Lambda(\beta,\alpha)}-\frac{t-\alpha}{\beta-\alpha}\right| &\leqslant \frac{\big|(\beta-\alpha)\cdot(t-\alpha)\cdot\lambda(\xi_1)-(\beta-\alpha)\cdot(t-\alpha)\cdot\lambda(\xi_2)\big|}{\inf\limits_{t\in[0,T]}\lambda(t)\cdot(\beta-\alpha)^2} \\ &\leqslant \frac{1}{\inf\limits_{t\in[0,T]}\lambda(t)}\cdot\frac{\big|(\beta-\alpha)\cdot(t-\alpha)\big|}{(\beta-\alpha)^2}\cdot|\lambda(\xi_1)-\lambda(\xi_2)| \\ &\leqslant \frac{1}{\inf\limits_{t\in[0,T]}\lambda(t)}\cdot\sup_{t,s\in[\alpha,\beta]}|\lambda(t)-\lambda(s)|,\end{aligned}$$

where $\xi_1 \in [\alpha, t]$, $\xi_2 \in [\alpha, \beta]$. ■

Lemma B.32. *For $f \in \{b^1, \dots, b^{m_{\mathrm{w}}}, c\}$ we have that for all $i \in \{0, 1, \dots, m\}$, it holds that*

$$\begin{aligned}\Big|\mathbb{E}\|f(t_i, \tilde{X}_m^M(t_i))\|^2 - \mathbb{E}\|f(t_i, X(t_i))\|^2\Big| \\ \leqslant C\Big(1 + \sup_{t\in[0,T]} \|\tilde{X}_m^M(t)\|_{\mathfrak{L}^2(\Omega)} + \sup_{t\in[0,T]} \|X(t)\|_{\mathfrak{L}^2(\Omega)}\Big) \\ \times \sup_{t\in[0,T]} \|\tilde{X}_m^M(t) - X(t)\|_{\mathfrak{L}^2(\Omega)}. \qquad \text{(B.127)}\end{aligned}$$

Proof. By the Jensen and Hölder inequalities we get,

$$\begin{aligned}&\Big|\mathbb{E}\|f(t_i, \tilde{X}_m^M(t_i))\|^2 - \mathbb{E}\|f(t_i, X(t_i))\|^2\Big| \\ &\leqslant \mathbb{E}\Big|\Big(\big\|f(t_i, \tilde{X}_m^M(t_i))\big\| - \big\|f(t_i, X(t_i))\big\|\Big) \\ &\quad\times \Big(\big\|f(t_i, \tilde{X}_m^M(t_i))\big\| + \big\|f(t_i, X(t_i))\big\|\Big)\Big| \\ &\leqslant \Big(\mathbb{E}\big\|f(t_i, \tilde{X}_m^M(t_i)) - f(t_i, X(t_i))\big\|^2\Big)^{1/2} \\ &\quad\times \Big(\mathbb{E}\Big(\big\|f(t_i, \tilde{X}_m^M(t_i))\big\| + \big\|f(t_i, X(t_i))\big\|\Big)^2\Big)^{1/2}. \qquad \text{(B.128)}\end{aligned}$$

By the assumption $(B1_{\text{MD}})$, $(B2_{\text{MD}})$ we have that

$$\begin{aligned}\Big(\mathbb{E}\big\|f(t_{i,n}, \tilde{X}_m^M(t_{i,n})) - f(t_{i,n}, X(t_{i,n}))\big\|^2\Big)^{1/2} &\leqslant K\Big(\mathbb{E}\big\|\tilde{X}_m^M(t_{i,n}) - X(t_{i,n}))\big\|^2\Big)^{1/2} \\ &\leqslant K \sup_{t\in[0,T]} \|\tilde{X}_m^M(t) - X(t)\|_{\mathfrak{L}^2(\Omega)}, \qquad \text{(B.129)}\end{aligned}$$

$$\begin{aligned}\Big(\mathbb{E}\big\|f(t_{i,n}, \tilde{X}_m^M(t_{i,n}))\big\|^2\Big)^{1/2} &\leqslant C\Big(1 + \big(\mathbb{E}\|\tilde{X}_m^M(t_{i,n})\|^2\big)^{1/2}\Big) \\ &\leqslant C\big(1 + \sup_{t\in[0,T]} \|\tilde{X}_m^M(t)\|_{\mathfrak{L}^2(\Omega)}\big), \qquad \text{(B.130)}\end{aligned}$$

and

$$\begin{aligned}\Big(\mathbb{E}\big\|f(t_{i,n}, X(t_{i,n}))\big\|^2\Big)^{1/2} &\leqslant C\Big(1 + \big(\mathbb{E}\|X(t_{i,n})\|^2\big)^{1/2}\Big) \\ &\leqslant C\big(1 + \sup_{t\in[0,T]} \|X(t)\|_{\mathfrak{L}^2(\Omega)}\big). \qquad \text{(B.131)}\end{aligned}$$

So by the (B.128) and estimations (B.129), (B.130), and (B.131) we get (B.127) and that ends the proof. ■

Lemma B.33. *For scalar function $f \in \{b, c\}$ we have that for all $i \in \{0, 1, \dots, k_n\}$, $\hat{t}_{i,n}$ defined by* (2.4) *and* (2.38) *it holds that*

$$\begin{aligned}\Big|\mathbb{E}|f(\hat{t}_{i,n}, X^{M*}_{k_n}(\hat{t}_{i,n}))|^2 - \mathbb{E}|f(\hat{t}_{i,n}, X(\hat{t}_{i,n}))|^2\Big| \\ \leqslant C(1 + \sup_{t\in[0,T]} \|\tilde{X}^{M*}_{k_n}(t)\|_{\mathfrak{L}^2(\Omega)} \sup_{t\in[0,T]} \|X(t)\|_{\mathfrak{L}^2(\Omega)}) \\ \times \sup_{t\in[0,T]} \|\tilde{X}^{M*}_{k_n}(t) - X(t)\|_{\mathfrak{L}^2(\Omega)}.\end{aligned}$$

The proof of Lemma B.33 goes analogously as proof of Lemma B.32 so we skip it.

Bibliography

[1] P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Statistics, Wiley, 1995.

[2] T. Bochacik, N. Czyżewska, A. Kałuża, D. Majchrowski, P. Morkisz, P. Przybyłowicz, and M. Studzińska-Wrona, *Hybrid option pricing through AI and GPU-powered SDEs solvers*, poster section at NVIDIA GTC 2020, https://www.nvidia.com/content/dam/en-zz/Solutions/gtc/conference-posters/gtc2020-posters/Algorithms_Numerical_Techniques_10_P22219_Pawel_Przybylowicz_Web.pdf.

[3] E. Bonet and D. Nualart Rodón, *Interpolation and forecasting in Poisson's processes.*, Stochastica, 2 (1977), pp. 36–40.

[4] J. Cheng, M. Grossman, and T. McKercher, *Professional CUDA C Programming*, EBL-Schweitzer, Wiley, 2014.

[5] D. Cohn, *Measure Theory: Second Edition*, Birkhäuser Advanced Texts Basler Lehrbücher, Springer New York, 2013.

[6] C. Dellacherie and P. Meyer, *Probabilities and Potential, B: Theory of Martingales*, ISSN, Elsevier Science, 2011.

[7] J. Dębowski and P. Przybyłowicz, *Optimal Approximation of Stochastic integrals with Respect to a Homogeneous Poisson Process*, Mediterranean Journal of Mathematics, 13 (2016).

[8] A. Friedman, Z. Birnbaum, and E. Lukacs, *Stochastic Differential Equations and Applications: Volume 1*, no. t. 1, Elsevier Science, 2014.

[9] A. Gardoń, *The Order of Approximations for Solutions of Itô-Type Stochastic Differential Equations with Jumps*, Stochastic Analysis and Applications, 22 (2004), pp. 679–699.

[10] C. Geiss and S. Geiss, *On approximation of a class of stochastic integrals and interpolation*, Stochastics and Stochastic Reports, 76 (2004), pp. 339–362.

[11] D. Gusak, A. Kukush, A. Kulik, Y. Mishura, and A. Pilipenko, *Theory of Stochastic Processes: With Applications to Financial Mathematics and Risk Theory*, Problem Books in Mathematics, Springer New York, 2010.

[12] S. Heinrich, *The quantum query complexity of elliptic PDE*, Journal of Complexity, 22 (2006), pp. 691 – 725. Special Issue: Information-Based Complexity Workshops FoCM Conference Santander, Spain, July 2005.

[13] S. Heinrich and P. Mathé, *The Monte Carlo Complexity of Fredholm Integral Equations*, Mathematics of Computation, 60 (1993), pp. 257–278.

[14] P. Hertling, *Nonlinear Lebesgue and Itô Integration Problems of High Complexity*, J. Complexity, 17 (2001), pp. 366–387.

[15] D. J. Higham and P. Kloeden, *Numerical methods for nonlinear stochastic differential equations with jumps*, Numerische Mathematik, 101 (2005), pp. 101–119.

[16] D. J. Higham and P. E. Kloeden, *Strong convergence rates for backward Euler on a class of nonlinear jump-diffusion problems*, Journal of Computational and Applied Mathematics, 205 (2007), pp. 949 – 956. Special issue on evolutionary problems.

[17] N. Hofmann, T. Müller-Gronbach, and K. Ritter, *Optimal approximation of stochastic differential equations by adaptive step-size control*, Mathematics of Computation, 69 (2000), pp. 1017–1034.

[18] N. Hofmann, T. Müller-Gronbach, and K. Ritter, *The Optimal Discretization of Stochastic Differential Equations*, Journal of Complexity, (2001), pp. 117–153.

[19] M. Jeanblanc, M. Yor, and M. Chesney, *Mathematical Methods for Financial Markets*, Springer Finance, Springer London, 2009.

[20] B. Kacewicz, *How to increase the order to get minimal-error algorithms for systems of ODE*, Numerische Mathematik, 45 (1984), pp. 93–104.

[21] B. Kacewicz, *Improved bounds on the randomized and quantum complexity of initial-value problems*, Journal of Complexity, 21 (2005), pp. 740 – 756.

[22] B. Kacewicz and P. Przybyłowicz, *Optimal adaptive solution of initial-value problems with unknown singularities*, Journal of Complexity, 24 (2008), pp. 455 – 476.

[23] A. Kałuża, P. Morkisz, and P. Przybyłowicz, *Optimal approximation of stochastic integrals in analytic noise model*, Applied Mathematics and Computation, 356 (2019), pp. 74–91.

[24] A. Kałuża and P. Przybyłowicz, *Optimal global approximation of jump-diffusion SDEs via path-independent step-size control*, Applied Numerical Mathematics, 128 (2018), pp. 24 – 42.

[25] R. Karandikar and B. Rao, *Introduction to Stochastic Calculus*, Indian Statistical Institute Series, Springer Singapore, 2018.

[26] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, Springer New York, 2014.

[27] J. Kelley, *General Topology*, Graduate Texts in Mathematics, Springer New York, 1975.

[28] D. Kirk and W. Hwu, *Programming Massively Parallel Processors: A Hands-on Approach*, Elsevier Science, 2016.

[29] A. Klenke, *Probability Theory: A Comprehensive Course*, Universitext, Springer London, 2007.

[30] P. Kloeden and A. Neuenkirch, *The Pathwise Convergence of Approximation Schemes for Stochastic Differential Equations*, LMS Journal of Computation and Mathematics, 10 (2007).

[31] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, 1992.

[32] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg, 2013.

[33] A. Kohatsu-Higa, A. Lejay, and K. Yasuda, *On Weak Approximation of Stochastic Differential Equations with Discontinuous Drift Coefficient*, Mathematical Economics, Ser. RIMS Kôkyûroku, 1788 (2011).

[34] C. Kumar, *Milstein-type Schemes of SDE Driven by L\'evy Noise with Super-linear Diffusion Coefficients*, arXiv preprint arXiv:1707.02343, (2017).

[35] C. Kumar and S. Sabanis, *On tamed milstein schemes of SDEs driven by Lévy noise*, Discrete & Continuous Dynamical Systems - B, 22 (2017), p. 421.

[36] H. Kuo, *Introduction to Stochastic Integration*, Universitext, Springer New York, 2006.

[37] B. Makarov and A. Podkorytov, *Real Analysis: Measures, Integrals and Applications*, Universitext, Springer London, 2013.

[38] X. Mao, *Stochastic Differential Equations and Applications*, Elsevier Science, 2007.

[39] G. Milstein and M. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Scientific Computation, Springer Berlin Heidelberg, 2004.

[40] P. Morkisz and L. Plaskota, *Complexity of approximating Hölder classes from information with varying Gaussian noise*, Journal of Complexity, 60 (2020), p. 101497.

[41] P. Morkisz and P. Przybyłowicz, *Optimal pointwise approximation of SDE's from inexact information*, Journal of Computational and Applied Mathematics, 324 (2017).

[42] P. Morkisz and P. Przybyłowicz, *Randomized derivative-free Milstein algorithm for efficient approximation of solutions of SDEs under noisy information*, Journal of Computational and Applied Mathematics, 383 (2020), p. 113112.

[43] T. Müller-Gronbach, *The optimal uniform approximation of systems of stochastic differential equations*, Ann. Appl. Probab., 12 (2002), pp. 664–690.

[44] T. Müller-Gronbach, *Strong Approximation of Systems of Stochastic Differential Equations*, 2002.

[45] T. Müller-Gronbach, *Optimal pointwise approximation of SDEs based on Brownian motion at discrete points*, Ann. Appl. Probab., 14 (2004), pp. 1605–1642.

[46] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems*, EMS tracts in mathematics, European Mathematical Society, 2008.

[47] NVIDIA, *NVIDIA Documentation*, 2020-11. `https://docs.nvidia.com/cuda/`.

[48] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Universitext, Springer Berlin Heidelberg, 2010.

[49] E. Pardoux and A. Răşcanu, *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*, Stochastic Modelling and Applied Probability, Springer International Publishing, 2014.

[50] L. Plaskota, *Average Complexity for Linear Problems in a Model with Varying Information Noise*, Journal of Complexity, 11 (1995), pp. 240 – 264.

[51] L. Plaskota, *How to Benefit from Noise*, Journal of Complexity, 12 (1996), pp. 175 – 184.

[52] L. Plaskota, *Worst Case Complexity of Problems with Random Information Noise*, Journal of Complexity, 12 (1996), pp. 416 – 439.

[53] L. Plaskota, *Average Case $L-\infty$-Approximation in the Presence of Gaussian Noise*, Journal of Approximation Theory, 93 (1998), pp. 501 – 515.

[54] L. Plaskota, *Automatic integration using asymptotically optimal adaptive Simpson quadrature*, Numerische Mathematik, 131 (2015), pp. 173–198.

[55] L. Plaskota, *On Linear Versus Nonlinear Approximation in the Average Case Setting*, Contemporary Computational Mathematics - A Celebration of the 80th Birthday of Ian Sloan, (2018), pp. 1035–1049.

[56] L. Plaskota and G. Wasilkowski, *Uniform Approximation of Piecewise r-Smooth and Globally Continuous Functions*, SIAM J. Numerical Analysis, 47 (2008), pp. 762–785.

[57] L. Plaskota, G. Wasilkowski, and Y. Zhao, *The power of adaption for approximating functions with singularities*, Math. Comput., 77 (2008), pp. 2309–2338.

[58] L. Plaskota and G. W. Wasilkowski, *Uniform Approximation of Piecewise r-Smooth and Globally Continuous Functions*, SIAM Journal on Numerical Analysis, 47 (2009), pp. 762–785.

[59] L. Plaskota, G. W. Wasilkowski, and H. Woźniakowski, *A New Algorithm and Worst Case Complexity for Feynman–Kac Path Integration*, Journal of Computational Physics, 164 (2000), pp. 335 – 353.

[60] L. Plaskota, G. W. Wasilkowski, and Y. Zhao, *An Adaptive Algorithm for Weighted Approximation of Singular Functions over* $\mathbb{R}$, SIAM Journal on Numerical Analysis, 51 (2013), pp. 1470–1493.

[61] E. Platen and N. Bruti-Liberati, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg, 2010.

[62] P. Protter, *Stochastic Integration and Differential Equations*, Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg, 2013.

[63] P. Przybyłowicz, *Linear information for approximation of the Itô integrals*, Numerical Algorithms, 52 (2009), pp. 677–699.

[64] P. Przybyłowicz, *Adaptive Itô-Taylor algorithm can optimally approximate the Itô integrals of singular functions*, Journal of Computational and Applied Mathematics, 235 (2010), p. 203–217.

[65] P. Przybyłowicz, *Optimal sampling design for approximation of stochastic Itô integrals with application to the nonlinear Lebesgue integration*, Journal of Computational and Applied Mathematics, 245 (2013), p. 10–29.

[66] P. Przybyłowicz, *Optimality of Euler-type algorithms for approximation of stochastic differential equations with discontinuous coefficients*, International Journal of Computer Mathematics, 91 (2014).

[67] P. Przybyłowicz, *Minimal asymptotic error for one-point approximation of SDEs with time-irregular coefficients*, Journal of Computational and Applied Mathematics, 282 (2015).

[68] P. Przybyłowicz, *Optimal global approximation of SDEs with time-irregular coefficients in asymptotic setting*, Applied Mathematics and Computation, 270 (2015), pp. 441–457.

[69] P. Przybyłowicz, *Optimal global approximation of stochastic differential equations with additive Poisson noise*, Numerical Algorithms, 73 (2016).

[70] P. Przybyłowicz, *Optimal sampling design for global approximation of jump diffusion SDEs*, `https://arxiv.org/abs/1701.08311`, (2017).

[71] P. Przybyłowicz, *Optimal sampling design for global approximation of jump diffusion stochastic differential equations*, Stochastics An International Journal of Probability and Stochastic Processes, 91 (2018).

[72] P. Przybyłowicz, *Efficient approximate solution of jump–diffusion SDEs via path-dependent adaptive step-size control*, Journal of Computational and Applied Mathematics, 350 (2019), pp. 396 – 411.

[73] P. Przybyłowicz and P. Morkisz, *Strong approximation of solutions of stochastic differential equations with time–irregular coefficients via randomized Euler algorithm*, Applied Numerical Mathematics, 78 (2013).

[74] J. Sanders and E. Kandrot, *CUDA by Example: An Introduction to General-Purpose GPU Programming*, NVIDIA, 01 2011.

[75] A. Semrau, *Euler's Approximations of Weak Solutions of Reflecting SDEs with Discontinuous Coefficients*, Bulletin of The Polish Academy of Sciences Mathematics, 55 (2007), pp. 171–182.

[76] A. Semrau, *Discrete Approximations of Strong Solutions of Reflecting SDEs with Discontinuous Coefficients*, Bulletin of The Polish Academy of Sciences Mathematics, 57 (2009), pp. 169–180.

[77] A. Shiryaev and S. Wilson, *Probability*, Graduate Texts in Mathematics, Springer New York, 2013.

[78] R. Situ, *Theory of Stochastic Differential Equations with Jumps and Applications: Mathematical and Analytical Techniques with Applications to Engineering*, Mathematical and Analytical Techniques with Applications to Engineering, Springer US, 2006.

[79] J. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms*, ACM monograph series, Academic Press, 1980.

[80] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*, Academic Press Professional, Inc., USA, 1988.

[81] L. Trefethen and D. Bau, *Numerical Linear Algebra*, Other Titles in Applied Mathematics, Society for Industrial and Applied Mathematics, 1997.

[82] UNKNOWN, *CUDA Programing Blog Spot*, 2020-11. `http://cuda-programming.blogspot.com/2013/02/texture-memory-in-cuda-what-is-texture.html`.

[83] H. von Weizsäcker and G. Winkler, *Stochastic Integrals: An Introduction*, Advanced Lectures in Mathematics, Vieweg+Teubner Verlag, 1990.

[84] A. Werschulz, *The Computational Complexity of Differential and Integral Equations: An Information-based Approach*, Oxford mathematical monographs, Oxford University Press, 1991.

[85] A. G. Werschulz, *Average case complexity of elliptic partial differential equations*, Journal of Complexity, 5 (1989), pp. 306 – 330.